University of Reading

Patterns amongst medicinally-used plants in Brazil

A thesis submitted by

Estevão do Nascimento Fernandes de Souza

for the degree of Doctor of Philosophy

School of Biological Sciences

University of Reading

September, 2017

To Arthur

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Estevão do Nascimento Fernandes de Souza

Reading, September, 2017

Abstract

Documentation of patterns in traditional knowledge is of paramount importance if we are to fully understand the relationship between people and their local environment. Investigations of medicinal plant use also contribute to discoveries of new drugs, by highlighting taxa which might have bioactivity of interest, and preservation of traditional knowledge, by identifying species important to local communities and describing how traditional knowledge is transmitted to future generations. Increasingly methods borrowed from ecology and community phylogenetics are used to understand patterns in medicinal plant use and inform bioprospecting. These methods are best applied to information-rich, large data sets, yet despite the growth in ethnobotanical research, data are scattered. This thesis exploits traditional knowledge documented in literature and herbaria, using newly compiled data for Brazil to explore patterns in medicinal plant use, in documentation of ethnomedicinal species, pharmacological investigation and distribution of plant use. Comparison of data from literature and herbarium sources, based on 1400 use reports of medicinal use of species of the family Leguminosae, showed herbaria to be a valuable source of data for ethnobotanical research. Despite the lower proportion of herbarium records (compared to literature records) with information on therapeutic application, modes of use and parts used, herbarium records validated data in the literature and added new information. These and additional data were used to demonstrate Ewé, a web-based database developed here, with tools to store and visualise ethnobotanical data. Research questions relating to pharmacological investigation of Brazilian plants used in traditional medicine, and to the spatial

distribution of traditional knowledge were addressed using these and other data sets compiled for this study. Phylogenetic methods were used to explore the diversity of Leguminosae species with medicinal use, revealing phylogenetic diversity of species used despite of lineages rich in species with medicinal use (hot nodes) accounting for 45% of the species used in local medicine. Hot nodes are better characterized pharmacologically. Finally, the spatial distribution of traditional knowledge in Brazil was investigated, using a novel generic-level phylogeny for the Brazilian flora and species-level herbarium data for Angiosperms in Brazil with medicinal uses. According to taxonomic and phylogenetic measures of beta diversity, spatial heterogeneity in medicinal plant use was identified, highlighting local adaptation of medicinal floras despite the preference for some plant groups revealed by clustering in deep nodes.

Acknowledgements

First of all, I would like to express my gratitude to my supervisor, Professor Julie Hawkins for all the support and motivation during my PhD study, and my stay in the UK. Since our first contact when I was looking for a supervisor for my PhD, Julie always helped and guided me, for my project and then for my thesis. Also, at the University of Reading, I would like to thank my friends and colleagues, Irene Teixidor-Toneu, Mazhani Binti, Marco Kreuzer, and Paul Wennekes, for the great discussions of our medicinal plants group.

This study was financially supported by the program Science Without Borders, from CAPES and the Brazilian Government, process number 7595-13-1. I would like to thank President Dilma Rouseff for making this program possible, supporting education and science in Brazil.

This project would not have been possible without the help of my friends in academia in Brazil, who contributed with ideas, supporting many initial steps of this thesis, Paulo Câmara, Micheline Carvalho-Silva, Carolyn Proença and Renata Martins.

My stay in the UK was a time for learning, and great experiences, and I would like to thank my friends and colleagues for being part of such experiences, Salwa, Yati, Natali, Hugh, Jason, David and Diego. Also, I am grateful to all my Brazilian friends, who even distant, helped me in difficult times, giving motivation and lots of support, Preca, Carol, Gabizinha, Ana Zotta, Aryane, Andre, Caio, Heloisa, Ellen, Alessandro, Bárbara, Danilo, Alan, Karla, Pedro, Carolzinha, and many who have not been named here but were very important during these four years.

Least, but not less important, I would like to express my gratitude to my wife, Elise, without whom this period of my life would have been much less interesting. Also my family in Brazil, my parents Ivana and Paulo (and Berninha), my great parents, Marta, Florinda, Ivan and Paulo, my cousins, uncles, and also Jandira, who became part of this big family.

Table of contents

List of Tables

List of Figures

Chapter 1 Introduction

"Cosi Ewé, Ewé Orixá, Ewé-ô, Ewé-ô Orixá
Sem folha não tem sonho
Sem folha não tem festa
Sem folha não tem vida
Sem folha não tem nada…"
(*Gerônimo / Ildásio Tavares*)

1.1 Medicinal plant use- history and perspectives

Gerônimo and Ilsádio Tavares' song expresses the relationship between followers of the Candomblé, an Afro-Brazilian religion, and plant species. The song says that without leaves or plants there is no dream, there is no joy, and there is no life. Candomblé is one of many religions where plants play a central role in worship and wider culture. Throughout human history plants have been sources of food, shelter, tools, and medicine (Gurib-Fakim, 2006). As in Candomblé, the relationship has sometimes encompassed supernatural beliefs of ritual use (Bussmann and Sharon, 2006; Dafni, 2006; Dhiman, 2003; Merlin, 2003; Naranjo, 1979; Prance, 1970). The medicinal properties of plant species have been the subject of empirical tests since ancient times (Gurib-Fakim, 2006; Halberstein, 2005; Leonti et al. 2013). Whether plant use is magic-symbolic, founded in relevant bioactivity, or both, is unknown for most plants species used medicinally (Fabricant and Farnsworth, 2001; Saslis-Lagoudakis et al. 2012a). Throughout history, medicinal plant knowledge has been transmitted mainly orally, either from older to younger generations (vertical transmission), or between non-parental pairs in the same generation (horizontal

transmission) (Hewlett and Cavalli-Sforza, 1986). Nevertheless, written pharmacopoeias and guides to medicinal plant use and the formulas and preparations can also be found from ancient times, for example in Sumerian or Chinese texts, and more recently in the form of many modern pharmacopoeias (Petrovska, 2012).

Medicinal plants are of great importance to human life, being used by almost 80% of the population of developing countries (Farnsworth et al. 1985; WHO, 2007), acting as a primary source of medicine for different societies (Alves and Rosa, 2007; Vandebroek et al. 2004). This relation of dependency on traditional medicine for healthcare is so strong that it is speculated that even when western medicine is available some traditional plant use will continue (Jäger, 2015). The significance of medicinal plant use raises concerns about conservation of both plant species and traditional knowledge (TK), since habitat and species loss together with erosion of traditional knowledge is reported (Alves and Rosa, 2007; Gupta et al. 2005; Shanley and Luz, 2003; Voeks and Leony, 2004).

Alongside their social and cultural importance, medicinal plants are an important source of drugs, and bioprospecting methods based on traditional knowledge are reported to contributed to discovery and development of new drugs (Cragg and Newman, 2013; Fabricant and Farnsworth, 2001; McChesney et al. 2007). Nevertheless, the number of plant species still to be investigated in bioprospecting is huge, and less than 20% of plant species have been accessed for bioactivity (Soejarto et al. 2005). The number of medicinal plant species globally is estimated at 50 000 species worldwide (Schippmann et al. 2002); exploiting these using and methods of

bioprospecting focusing on traditional knowledge might be expected to increase the number of commercial plant-based medicines. Although the rate of discovery is low (Firn, 2003), methodologies encompassing phylogenetic, ecological and meta-analytic methods to investigate traditional plant use may enhance drug discovery (Albuquerque and de Medeiros, 2012; de Albuquerque, 2010; Saslis-Lagoudakis et al. 2012a). The topic of this thesis "Patterns amongst medicinally-used plants in Brazil" combines a diversity of approaches to document and investigate medicinal plant use in Brazil in order to contribute to a better understanding of human-plant relationships.

1.2 Medicinal plant use in Brazil

With a high diversity of plant species, estimated at more than 40 000 species (Forzza et al. 2012), and rich in cultural manifestations, Brazil is a good source of traditional knowledge of plant species (Nogueira et al. 2010). Hotspots for plant communities, for example the Cerrado and the Atlantic Forest might become important for new medicinal plant discovery. Of the 10 000 species of the Cerrado (Mendonça et al. 2008) about 500 species have documented medicinal use (Guarim-Neto and Morais, 2003). With a wide variety of vegetation formations, Cerrado species have different adaptive strategies to biotic or abiotic stresses, resulting in secondary compounds that may be of great interest to bioprospecting for new drugs (Benko-Iseppon and Crovella, 2010). Surveys of Brazilian medicinal plants have been made throughout the country, but there are understudied regions remaining (de Medeiros et al. 2013). Brazilian traditional medicine incorporates not only its local flora, but also

many introduced species, as a result of human migration to the American continent (Bennett and Prance, 2000). Nevertheless, similarities between the Brazilian flora and others, such as the West-African flora, from where many Africans were brought to Brazil, facilitated the adaptation to the new environment and contributed to the increase in the diversity of traditionally used species (Voeks, 2013).

Although practiced in Brazil before the arrival of Europeans, traditional uses of Brazilian flora only started to be documented by naturalists that visited Brazil in the 19th century, such as Auguste P. Saint-Hilaire, Karl F.P von Martius, Johann E. Pohl, Gregory Langsdorffii (Brandão et al. 2008). The book "*Systema Materiae Medicae Vegetabilis Brasiliensis*" published by Karl Friedrich von Martius is one of the most important historical accounts of the Brazilian flora, documenting medicinal plants and their uses. Another important naturalist, Manuel Freire Allemão de Cyneiros, published several papers known as "*Materia Medica Brasileira"* between 1862 and 1864 (Lorenzi and Abreu Matos, 2002). Missionaries and priests also had an important role in documenting the Brazilian "*Materia medicas*", while spreading knowledge of edible and medicinal species, in particular the priests from the Jesuits Order (Anagnostou, 2007). Many of these "*Materia Medicas"* were compiled together with herbarium vouchers that can now be examined and used to investigate historical uses of medicinal plants in Brazil. Compilations of historical use data reported 203 species with past medicinal use in the Brazilian flora (Brandão et al. 2008) and since then, the number of reports has increased to 753 species with medicinal use from literature survey reported by de Medeiros et al. (2013). Despite this extensive list of medicinal plant species, only 57 plant species are present in the latest version of the official

Brazilian Pharmacopoeia (Anvisa, 2010). Although the number of plant species in the Brazilian Pharmacopoeia has fluctuated (Alkin et al. 2017), it has always represented a small proportion of the number of medicinal plants in use.

How plants are selected for medicinal use, and whether there are general patterns underpinning traditional use is a research area of interest (Ankli et al. 1999; Geck et al. 2017; Ladio et al. 2007; Leonti et al. 2002; Lucena et al. 2007; Saslis-Lagoudakis et al. 2014). Several studies have been directed towards a better understanding of the selection of Brazilian medicinal plants. Voeks (1996) showed that plant resources obtained from secondary forest were more important for local healers than the ones from primary forest in a study in the northeast of Brazil (Bahia State). Primary forest was preferred for timber gathering while secondary forest had more medicinal species. de Medeiros et al. (2013) showed that there is no difference between woody and non-woody habits in the Brazilian Traditional Medicine, but exotic (or introduced) plants have a higher relative importance than native species (which are more important considering only the number of species). Indigenous people, such as the Krahô tribe, would rather make use of only native plants (Rodrigues and Carlini, 2006), perhaps because of the isolation of their society. Despite increasing numbers of studies, many are focused on local communities. Studies using large data sets to identify patterns of medicinal plant use across Brazil are rare (Albuquerque and de Medeiros, 2012; de Medeiros et al. 2013).

1.3 Thesis organization

This thesis has three themes, the documentation of traditional knowledge of medicinal plants, the phylogenetic distribution of knowledge and the spatial distribution of knowledge. I aim to identify patterns relating to medicinal plant use in Brazil across all three themes.

1.3.1 The documentation of traditional knowledge

The perception that erosion of traditional knowledge was at an unprecedented rate, and the view that traditional knowledge of medicinal plants presents solutions to global health problems, has focussed ethnobotanical research on recording knowledge before it is lost (Cox, 2000). In light of this urgency, documentation of traditional use has often been entirely descriptive. Ethnographic research, taking in traditional use of plants, is traditionally focussed on definable groups of people. Many hold the view that each culture has its own unique relationship with natural resources (Albuquerque et al. 2007). This approach to documentation of TK has furnished a huge amount of data, and increasingly ethnobotanists are questioning whether these data can be used to test hypotheses (de Medeiros et al. 2013). Examples of questions addressed using published data describing medicinal plant use include whether "core" medicinal floras can be identified (Ellen and Puri, 2016), whether there are taxonomic or phylogenetic patterns in medicinal plant use (Saslis-Lagoudakis et al. 2012a; Saslis-Lagoudakis et al. 2011a; Weckerle et al. 2011), whether urbanisation influences plant use (de Medeiros et al. 2013) and whether socioeconomic and environmental factors influence plant use (de la Torre et al. 2012). Despite the potential of this research to contribute to the

understanding of relationship between humans and plants, the data that could be used are often very dispersed (Albuquerque and Medeiros, 2012). Compilation of existing data is important to determine the extent to which large-scale questions can be addressed, and to highlight where different information can be found.

In chapter 2 I present ethnobotanical data from different sources, namely literature and herbarium specimens, in order to quantify and qualify the information present for the medicinal uses of the Leguminosae in Brazil. Meta-analyses have used data published in the literature, but the documentation of traditional knowledge in herbaria is also noteworthy in providing data for ethnobotanical studies (Bedigian, 2004; de la Torre et al. 2012; Fantz, 1991; van Andel et al. 2014). The comparison of literature and herbarium data is in order to determine the extent to which herbarium and published sources are comparable. I also ask whether herbaria can fill gaps in our knowledge of plant use. Much of the exploration of the data described in chapter 2 depended on tools I developed and embedded in an online database. The primary data compiled for chapter 2 provide a core dataset to demonstrate the archiving, curation, efficient analysis and interpretation of ethnobotanical data using the new database. The new database, which I call Ewé, is described in chapter 3.

1.3.2   The phylogenetic distribution of knowledge

The potential of medicinal plant use in drug discovery and bioprospecting is reported in many studies (Cragg and Newman, 2013; Fabricant and Farnsworth, 2001; McChesney et al. 2007), and lately the use of phylogenetic methods together with ethnobotanical data is increasing in order to improve bioprospecting (Ernst et al. 2016; Halse-Gramkow et al. 2016; Lukhoba et al. 2006; Saslis-Lagoudakis et al. 2012a, 2015). Based on reports of overuse in some plant lineages (Amiguet et al. 2006; Bennett and Husby, 2008; Moerman et al. 1999) implying a phylogenetic signal in plant selection, phylogenetic investigation of medicinal plant use also relies on the presence of phylogenetic signal in the distribution of chemical constituents and shared biochemistry between closely related species (Agrawal et al. 2009; Champagne et al. 1993; Muellner et al. 2005; Rønsted et al. 2008; Wink, 2003, 2013; Wink and Mohamed, 2003). Despite the increase in number of studies targeting plant species for bioprospecting, the number of species that yield natural products that have been screened was reported to be of less than 10% (Fabricant and Farnsworth, 2001), and more recent estimates are needed.

In chapter 4 I make further use of the data compiled for chapter 2. Chapter 4 describes the investigation of phylogenetic structure amongst Leguminosae species used medicinally in Brazil. As well as describing the phylogenetic structure of use and for different therapeutic applications, the findings are reviewed relative to intensity of pharmcological investigation. Pharmacological data were sought and reported here, and I report the proportion of Brazilian plants screened. Whether the screening is

related to the phylogenetic distribution of the species used medicinally is explored, specifically with relations to lineages rich in traditional medicinal use.

1.3.3 The spatial distribution of knowledge

Several studies have sought to understand the effect of the composition of the local flora on species selection for medicinal purposes. Abundance is well known to influence plant selection, and weedy plants from disturbed environments are often preferred (Albuquerque & de Oliveira, 2007; Stepp and Moerman, 2001). Looking beyond ecology to floristic composition overall, there are few comparative studies which aim to tease out the extent to which the same plants are used in different cultural and floristic backgrounds. One such study (Saslis-Lagoudakis et al. 2014) considered all medicinal plant use of 12 people-groups in Nepal across three regions of Nepal. This study suggested the floristic environment was a more important influence on medicinal plant selection than cultural relatedness or proximity to other cultural groups. Another study (Cámara-Leret et al. 2017) showed a preference for widespread palm species in the Amazon, suggesting that TK transmission plays an important role in homogenising plant use across floristic environments. This view is supported by the study of migrations, which often highlight continuity in use of medicinal plants (van Andel and van't Klooster, 2007; Vandebroek et al. 2007; Volpato et al. 2008). Saslis-Lagoudakis et al. (2014) considered three regions of Nepal, but these did not correspond to natural floristic regions, whilst Cámara-Leret et al. (2017) compared communities considering the distribution of the study species but not the floristic

composition overall. More studies of spatial distribution of TK are needed to document and begin to understand complexity in traditional knowledge patterns.

In chapter 5 I set out to characterise spatial patterns in medicinal plant use in Brazil. Using data from herbaria for all Angiosperm species, I applied a set of methods borrowed from biogeography and ecology. This study used taxonomic and phylogenetic beta diversity to investigate the composition of species and medicinal floras between biomes, and Principal Components of Phylogenetic Structure (PCPS) to identify lineages contributing to biome differences. Brazil is an interesting contrast to Nepal, since it is relatively linguistically uniform, but floristically highly diverse yet well characterised. In Saslis-Lagoudakis et al. (2014) study the cultural identities were well-known, but floristic ones were not. In this study we expected to find relatively homogeneous plant use because of a lack of linguistic boundaries to TK transmission and a history of migration.

1.4 A synthesis

Each of the four chapters outlined above are presented as if for submission to a peer-reviewed scientific journal. Chapter 2 has been published at the time of submission of this thesis (Appendix 1), in the journal Economic Botany (Souza and Hawkins, 2017). Whilst each chapter is written to stand alone, there is clear cross-fertilisation between chapters. Chapters 2 and 3 depend on the same data set describing medicinal use of species of family Leguminosae in Brazil. Chapter 2 critically

evaluates the ethnobotanical data describing medicinal use of plants found on herbarium labels and makes comparison to data from written sources, whilst chapter 3 uses the data to populate and demonstrate the use of a new database. Chapter 4 also uses the same data set describing medicinal use of species of family Leguminosae in Brazil. In this chapter additional data to capture pharmacological research activity are collected. The novelty in this chapter lies in the comparison between traditional use and scientific screening, using a phylogenetic framework for comparison. Chapter 5 extends the data on medicinal plant use, compiling data from all Angiosperms. The compilation of these data from herbarium specimens is justified based on the findings presented in chapter 2. Phylogenetic methods are also applied in this chapter, but in this case they are methods from biogeography and ecology, rather than "hot node" analyses applied to the data in chapter 4. Thus the applications of plant phylogeny, emerging as tool to better understand how plants are used by people, are extended here.

Chapter 2 Comparison of herbarium label data and published medicinal use: herbaria as an underutilized source of ethnobotanical information

2.1   Introduction

Ethnobotanical research is crucial to understanding relationships between people and their biological environment (Thomas, 2003). Global or regional studies which compile and analyze data from multiple literature sources can lead to a general understanding about plant use (de la Torre et al. 2012; de Medeiros et al. 2013; Saslis-Lagoudakis et al. 2012a; Weckerle et al. 2011), but the availability of appropriate data is a limitation to broad-scale research (Albuquerque and de Medeiros, 2012). Herbaria are repositories of information in the form of vouchers, originally serving economic botany, and increasingly seen primarily as resources for plant taxonomy (Bebber et al. 2010; Van Andel et al. 2012). Today the wider value of herbaria is appreciated (Lavoie et al. 2013); herbaria worldwide house more than 300 million specimens collected over 400 years, and as such are a rich repository of specimen collection dates and localities (Thiers, 2014). Nevertheless, the extent to which herbarium specimens may contribute ethnobotanical data not captured in publications, filling gaps in our knowledge and providing data for analyses, remains only partially explored.

Ethnobotanists frequently and routinely collect and cite herbarium specimens; indeed, journals such as *Economic Botany* require voucher specimens to be cited alongside use reports. However, specimens which were not collected as vouchers to

support published studies may include use data. The most complete survey to assess the frequency of ethnobotanical information in herbarium collections was that of von Reis (1962; 1968). She reported almost 6,800 specimens citing medicinal uses amongst the 2,500,000 specimens of Harvard Herbarium, despite excluding those published or likely to be known already. Since von Reis enumerated the advantages of searching herbaria for novel reports of use, herbarium surveys have become a minor but established source of ethnobotanical data (Bedigian 2004; de la Torre et al. 2012; Fantz 1991; Jenks & Kim 2013; Krishna et al. 2014; Lampe 1986; Lira & Caballero 2002; Lukhoba et al. 2006; McKenna et al. 2011; Prakash 2011; Van Andel et al. 2014; Vickery 1990; Shinde and Prakash 2015). Since herbaria are rich in historical data, they have found particular use in documenting change (Nesbitt, 2014). For example, studies of historic herbaria have used the annotation on specimens to reveal changes in local names and uses (van Andel et al. 2012 - Hermann herbarium), or the changing species composition of pharmacopeia (Birch 2009 - Gideon Lincecum Herbarium; De Natale & Pollio 2012 - Trotter collection).

The Leguminosae (legumes) is one of the largest plants families (Lewis et al. 2005), over-utilized for medicine in some regions (Korea and Ecuador), and under-utilized in others (North America), and with many documented uses (Moerman et al.1999). In Brazil, the family comprises c. 2,800 species in more than 200 genera (Flora of Brazil 2020 in construction). The legumes of Brazil are distributed amongst the six biomes described for Brazil, Amazon Forest (1147 sp.), Atlantic Rainforest (997 sp.), Caatinga (620 sp.), Cerrado (1237 sp.), Pampas (138 sp.) and Pantanal (161 sp.). Ethnobotanical work in Brazil is increasing, particularly in the Northeast and Southeast

regions of Brazil, where there are active research groups. However, in the Amazon, Cerrado, Pantanal and Pampas there is a relative deficit of ethnobotanical data in the literature (Ritter et al. 2015).

The complementarity of herbarium and published data on medicinal use is yet to be formally assessed, and the proportion of herbarium specimens with medicinal use information is uncertain, since the von Reis study did not record the number of excluded specimens (von Reis 1962, 1968; Nesbitt 2014). Here we compare literature and herbaria as sources of ethnobotanical data, taking medicinal use of Brazilian legumes as a case study. We ask how often herbaria provide data on plant use that is not present in the literature, whether the data provided in the two sources is broadly comparable and if so, whether the data from herbarium specimens could be used to validate or augment poorly substantiated literature records. By addressing these questions, we seek to outline a role for herbarium data in ethnobotanical research.

2.2   Methods

2.2.1   Data collection

Published medicinal use data, and data from Brazilian herbarium specimens were compiled for the legumes of Brazil. Data were organized in a database with the following fields: genus and species as recorded in the publication or in the specimen database, currently accepted name, therapeutic application as recorded in the publication or in the specimen database, therapeutic application categorized according to WHO International Classification of Diseases 10 (World Health Organization 2016), parts used, mode of application and locality. Where necessary, Google maps and the Brazilian Geographic and Statistical Institute online data (IBGE) were used to get geographical coordinates for the localities cited on papers but with no cited coordinates. Generic and species names followed The Plant List (The Plant List 2013) and Missouri Botanical Garden's Tropicos database (Tropicos.org 2016), and were corrected to a currently accepted name using the Plantminer R script (Carvalho et al. 2010)

Publications citing medicinal uses of legumes species in Brazil were identified using Google Scholar and searches of the following journals: Acta Botanica Brasilica, *Economic Botany*, Fitoterapia, Flovet, Journal of Ethnobiology, Journal of Ethnobiology and Ethnomedicine, Journal of Ethnopharmacology, Journal of Medicinal Plants Research, Revista Brasileira de Biociencias, Revista Brasileira de Farmacognosia, Revista Brasileira de Plantas Medicinais and Rodriguesia. Search terms, which were used both in English and Portuguese, are presented in Table 2-1. From the subset of papers found which record medicinal plant use, a dataset was compiled to record

fields as above. The frequency of voucher specimens being cited on the selected publications was calculated.

Table 2-1.Combinations of words and terms used as search terms in the literature review. Each word or term in column 1 was combined with each in column 2. *The term "Brazilian Ethnobotany" was also searched alone.

Column 1	Column 2
Brazilian ethnobotany*	Cerrado
Ethnobotany	Atlantic rainforest
Ethnopharmacology	Caatinga
Medicinal Plants	Pantanal/ Wetlands
Medicinal Flora	Pampas/ Grasslands
Ethnobiology	Amazon forest

Herbarium data were extracted from the online list of herbarium and biological collections from Brazil, Species Link (http://splink.cria.org.br/). The file was exported to excel format and a search using the keywords "medicinal" and "uses" (in Portuguese) was conducted in order to filter only the specimens with medicinal information. Although search was limited to these two terms, a test search that included specific diseases or medical conditions, such as cold, influenza, migraines, heart conditions, abortifacients stomach-ache and diarrhea, rarely resulted in even one more use report that could be included in the analysis. Therefore, only the two previously cited terms were used to obtain the final list. Specimens were sourced from 15 herbaria, FEMACT, INPA, IPA, JBRJ, MBML, UEL, UEM, UEM, UFERSA, UFG, UFMS, UFPB, UFPE, UFRPE, UNESPSJRP, UNICAMP.

2.2.2 Data analysis

The data from the surveys were explored in several ways to draw out, for the different sources, how much data there were, whether they were comparable or complementary and their spatial and temporal characteristics.

Number of reports and presence/absence of ethnobotanical data: Numbers of use reports from the literature and from the herbarium were calculated, where a use report is the accepted species name plus all the associated data originating from one publication or one voucher specimen (i.e. entries or rows in Appendix 2). How frequently use reports, whether from publications or herbarium labels, include data on therapeutic application, plant part used or mode of application was also calculated as percentages.

Number of species described and comparison of species lists: The number of species reported from literature and from the herbarium sources was determined. By comparing species lists, species unique to either literature or herbarium sources were identified, and their proportions calculated based on the total number of species for each source. Similarly, the species with associated therapeutic indications/modes of application/plant part data were compared between the literature lists and the herbarium lists, and the proportions of species with these data were also calculated. To account for the possibility that the same study contributed both the herbarium and publication data, we scored the number of times authors and collectors were the same person, for records of the same content.

Comparison of information content: To test whether the information content of the data from the two sources was similar we compared numbers of reports by genus, for

each therapeutic application and in total. To compare therapeutic applications, we classified literature and herbarium reports following the WHO ICD-10 (http://www.who.int/classifications/icd/en/). Although data standardization is reported to be challenging, as researchers' (etic) classification does not always reflect properly on local communities' (emic) classification and understanding of diseases, for reasons of comparison and meta-analysis such standards are reported (Staub et al., 2015). This is discussed further in Chapter 3. Total numbers of each therapeutic application, plant part used and mode of application were recorded for both sources, so that the broad comparability of the data content could be assessed visually and using Spearman Rank-Order Correlation tests.

Spatial and temporal distribution of data: The number of reports and species, from both literature and herbaria, were calculated for each biome in Brazil. Their coordinates were plotted to compare spatial distribution of studies and collections. To investigate whether the number of reports in each biome from each source were correlated, Spearman Rank-Order Correlation was used. Dates of collection and publication between the two sources were also compared, and the changing proportions of specimens through time which report any information about therapeutic application, plant part used or mode of application was assessed.

Validation and augmentation of data sets: We searched for the species only reported once in the literature (whether data-rich, i.e. including mode of use, plant part or therapeutic application, or not) in the herbarium reports. We considered the literature report "validated" if there was an independently collected herbarium report of medicinal use for that species (not a voucher specimen collected by the authors of the

study). We also listed the species that were reported as medicinal in the literature, but where there were no literature reports of the mode of use, plant part or therapeutic application, and searched for these in the herbarium data. For these species we identified the number of herbarium reports of mode of use, plant part used or therapeutic application that were not in the literature; these reports we considered to augment the literature reports. Finally, considering only the species with reports from both literature and herbarium, we counted the total number of new use reports from each source.

2.3 Results

Number of reports and presence or absence of associated data describing use: Considering the data sourced from publications, excluding reports where identification was only to genus level and uses that were not medicinal, there were 938 reports from 104 publications. Of these a subset of 69% record therapeutic application, 49% provide information about mode of application and 61% provided information about plant part used (Table 2-2). 91% of publications cited voucher specimens. Considering the herbarium data, of the 240,000 specimens from 15 Brazilian herbaria searched, 462 (0.2%) indicated whether the plant was used medicinally. These represented 154 species in 62 genera. Of these specimens, 16% described therapeutic application, 4% provided information about mode of application and 6% provided information about plant part used (Table 2-2).

Number of species described and comparison of species lists: In total 264 species in 97 genera were reported as medicinal in the literature, of which 165 species (62%) were uniquely from literature. Of the 154 species in 62 genera known from the herbarium, 55 (36%) were known only from that source. The percentages of unique species from literature where therapeutic applications, plant parts used and modes of application were known were much higher, from 82 to 92%. Nevertheless, in the fewer cases from herbaria we found 14% of unique species with indicated plant parts used and 25% of unique species with therapeutic indications. (Table 2).

Table 2-2. Summary of data from literature and from herbarium specimens. The number of reports from literature or herbarium sources are indicated, and whether they include any further information about use (therapeutic indication, mode of application, plant parts). Species data is also presented. The total number of species with recorded medicinal use is given, also the numbers and proportions of species that are either unique to literature or unique to herbarium (those included in one data set but not the other). Species with uses are those where use for a species is described in that source; unique species are species for which information about that use (therapeutic indication, mode of application, plant parts) is derived only from that source.

	Literature	Herbaria
Number of use reports	938	462
Number of reports with therapeutic indications	654	76
Number of reports with modes of application	462	19
Number of reports with plant parts used	579	30
Number of reports without further information	284	385
Number of species (number/percentage of unique species)	264 (165/62%)	154 (55/36%)
Number of species with therapeutic indications (number/percentage of unique species)	203 (167/82%)	48 (12/25%)
Number of species with modes of application ((number/percentage of unique species)	162 (149/92%)	16 (3/19%)
Number of species with plant parts used (number/percentage of unique species)	184 (165/90%)	22 (3/14%)
Number of species without further information (number/percentage of unique species)	146(78/33%)	141(73/52%)

Comparison of information content: For publications, descriptions of use could be translated to 16 of the 19 WHO ICD-10 categories of use. For herbaria, therapeutic uses were allocated to 13 of the 19 WHO ICD-10 categories. Herbaria and literature reports showed a statistically significant association between the sets of ranks, when the most cited genera (p=<0,0001, rho=0.88), therapeutic applications (p=1.901e-07, rho=0.91), modes of application (p=0.002, rho 0.86) and plant parts (p=0.0003,

rho=0.94) were ranked and compared (Tables 2-3 and 2-4). The number of records which had the same information and for which collector and author were the same person was 35 (3.6% of records).

Table 2-3. Most cited genera in literature and herbarium. Counts represent the total number of reports for each genus, in the literature and in the herbarium.

Literature		Herbarium	
Genus	Count	Genus	Count
Senna	116	*Bauhinia*	77
Bauhinia	97	*Senna*	64
Hymenaea	66	*Mimosa*	27
Caesalpinia	65	*Caesalpinia*	23
Mimosa	50	*Hymenaea*	16
Anadenanthera	34	*Stryphnodendron*	15
Stryphnodendron	30	*Copaifera*	14
Copaifera	29	*Anadenanthera*	11
Amburana	26	*Amburana*	8
Bowdichia	23	*Bowdichia*	5
Desmodium	14	*Desmodium*	9
Libidibia	1	*Libidibia*	9

Table 2-4. Comparison of uses between literature and herbaria, showing the number of reports per therapeutic application, plant part used, and mode of application.

Use	Literature reports	Herbarium reports
Therapeutic Application		
Diseases of the respiratory system	193	21
Diseases of the digestive system	156	10
Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	151	13
Endocrine, nutritional and metabolic diseases	151	13
Diseases of the genitourinary system	64	4
Diseases of the femalegenito system	62	7
Diseases of the musculoskeletal system and connective tissue	56	5
Diseases of the circulatory system	48	5

Table 2.4 Comparison of uses between literature and herbaria, showing the number of reports per therapeutic application, plant part used, and mode of application (continuation).

Diseases of the nervous system		
	43	3
Certain infectious and parasitic diseases		
	41	5
Diseases of the skin and subcutaneous tissue		
	24	5
Neoplasms		
	14	1
Diseases of the eye and adnexa		
	6	0
Diseases of the ear and mastoid process		
	5	0
Pregnancy, childbirth and the puerperium		
	4	0
OTHERS	2	0
Mental and behavioural disorders		
	0	0
Plant Part Used		
Bark	288	14
Leaves	199	7
Root	74	12
Fruit	44	1
Seeds	38	1
Flowers	29	0
Whole Plant	16	0
Stem	8	0
Mode Of Application		
Decoction	153	3
Tea	131	11
Infusion	113	2
Syrup	50	2
Alcoholic Infusion	44	0
Bath	32	2
Powder	3	0
Oil	3	0
Wine	2	0

Spatial and temporal distribution of data: Figure 2-1 shows the spatial distribution of data. Considering the geographic origin of literature and herbarium reports combined, the Cerrado and Caatinga biomes were best represented, with 418 and 361 reports respectively, and the Pampas the most under-represented, with 12

reports (Table 2-5).
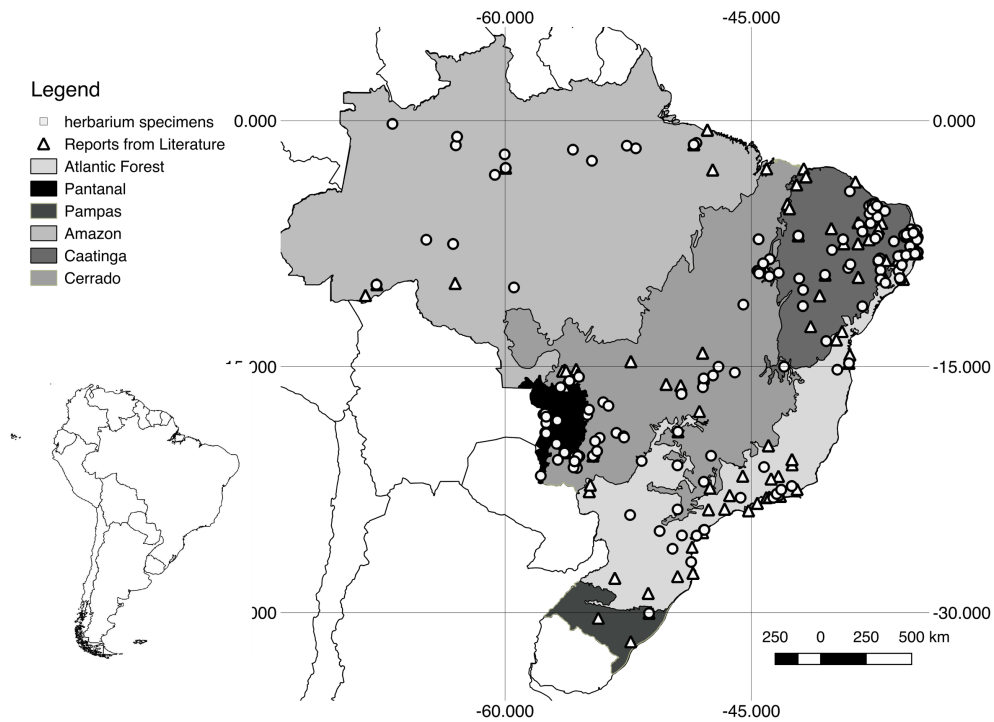
Figure 2-1. Geographic origin of data extracted from papers (triangles) and herbaria (circles). Triangles and circles shows the location of studied sites and can represent more than 1 record.

Table 2-5. Spatial comparison of data from literature and the herbaria. The distribution of reports and species from each source in the six Brazilian biomes is presented. The percentages of total reports and total species that each source contributes to each biome are shown.

Biomes	Herbarium		Literature		Total	
	Reports	Species	Reports	Species	Reports	Species
Amazon forest	32 (39%)	22 (39%)	50 (60%)	39 (68%)	82	57
Atlantic forest	74 (43%)	51 (56%)	96 (56%)	56 (61%)	170	91
Caatinga	159 (44%)	54 (55%)	202 (56%)	80 (82%)	361	97
Cerrado	96 (22%)	49 (34%)	322 (77%)	126 (88%)	418	142
Pampas	2 (16%)	2 (18%)	10 (83%)	8 (72%)	12	11
Pantanal	32 (58%)	22 (62%)	23 (41%)	18 (51%)	55	35

Biomes with more herbarium records also had more literature records (p=0.007,

rho=0.92). Dates of collection and publication comparison revealed herbaria as a

unique source of ethnobotanical data until 1980, and the 2000's as the decade with most reports (Table 2-6). The proportion of herbarium reports with medicinal information increased through time, (0.12%-0.36% from 1900-1910 to 2000-2009) (Figure 2-2). Data for 2010 to 2014 are not plotted: in this period, of 19910 specimens 23 (0.11%) included information about the medicinal use of the specimen.
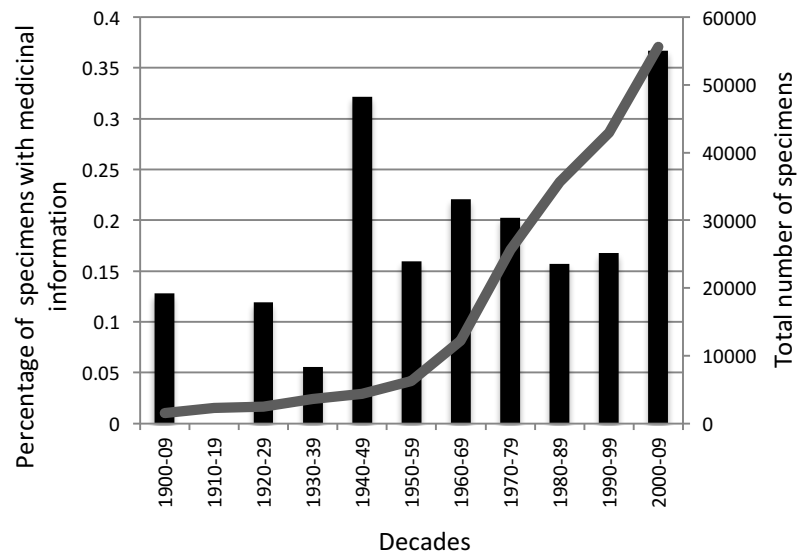
Figure 2-2. Temporal patterns in the recording of medicinal data in herbaria. The graph shows the total number of herbarium specimens collected in each period and now deposited in the herbaria surveyed (line), and the proportion of those specimens with any associated information indicating whether or how the plant is used medicinal (bars).

Table 2-6. Temporal comparison of data from literature and the herbaria. The total number of reports in each decade, and in the last four years, is reported.

Time period	Literature reports	Herbarium reports
1900-09	0	2
1910-19	0	0
1920-29	0	3
1930-39	0	2
1940-49	0	14
1950-59	0	7
1960-69	0	27
1970-79	0	51
1980-89	3	54
1990-99	20	70
2000-09	591	203
2010-2014	324	23

Validation and augmentation of data sets: From the literature, 123 species were reported once (hereafter "singletons"), excluding five species re-reported in later literature based on earlier original publication. Twenty-one (17%) of these species were validated by an independent herbarium report. Eighty-four singleton species were reported in the herbarium data (including eight species whose multiple reports originated from plant collections made in the same proximity by the same collector in the same year, since this is considered resulting from one observation). Of these 84, 39 (46%) were validated by a literature report. From the literature, there were 53 singleton species for which there was no information about how the species was used; use reports were available from the herbarium for 6 of these (1 for therapeutic application). From the herbarium there were 73 singleton species without use record; use reports were available from the literature for 27 of these (27 for therapeutic application, 24 for mode of use and 25 for plant part used). Considering the 94 shared species, herbarium data added 58 new therapeutic indications, 25 new plant parts used and 16 new modes of application, a total of 99 new uses when compared to literature.

2.4   Discussion

Our study shows that the modern (1900-2010) herbarium specimens we surveyed comprise a significant source of data. They are information-rich, often describe how the plant is used, and overall the information they contain appears broadly similar to the information recorded in the literature. We were surprised at how frequently specimens contained information about the specific applications of the plants used; we expected many specimens might simply report that a species was used medicinally. However, of those specimens citing any medicinal use at all, 69% also reported the therapeutic application of the species, 49% the mode of application and 61% the plant part used. Overall, the number of specimens with medicinal information is low, however. We show here that just 0.2% of the herbarium specimens we examined for this study contain some information about medicinal application of the species represented by the voucher.

Until critical studies of other herbaria are made it is not possible to determine whether our findings about the proportion of specimens recording ethnomedicinal data apply more generally. Our survey included all specimens, regardless of whether the species were previously known to have use, yet this proportion is less than the 0.27% reported in von Reis' studies even after she excluded species already known (von Reis 1962, 1968; Harvard Herbarium), and much less than another study reporting 3% of specimens with information about medicinal use (Shinde and Prakash 2015; Mumbai's 109 year old Blatter herbarium). Why this might be is unclear. Few of our herbarium specimens are as old as those in the Blatter herbarium and we also expect the age profile of the specimens we examine to be younger than the Harvard

specimens examined by von Reis. Her survey was carried out in the 1960s; fewer than 2% of the specimens we examined pre-dated 1900 (no pre-1900 specimens included medicinal information), and we don't include specimens collected before 1900. Very early collections were often made by botanists who were trained physicians or made in the context of colonial expansion through exploitation of plant resources, and thus were made to record plant use (Schiebinger 2004, Van Andel et al. 2012). Historical herbaria, and the significant national and institutional herbaria they have been bequeathed to and incorporated into, might therefore be richer in ethnobotanical information that the more modern herbaria we surveyed. Yet we also show that later herbarium specimens in our study are more likely to include medicinal information (Figure 2-2), suggesting a renaissance in Brazil in recording plant uses on herbarium labels. This might reflect the exponential growth in the number of individuals and institutions involved in ethnobotanical studies and the recent formal inclusion of ethnobotany in undergraduate and graduate courses in Brazil (Fonseca-Kruel et al. 2005). Outside of Brazil, we might also expect different herbaria, emphasizing specimens from different geographical regions, to include different proportions of specimens with label data reporting medicinal use. This might be because the label-making practices of collectors lodging their specimens in different herbaria differ, or because documentation of ethnobotanical uses is more incomplete in some areas (Cámara-Leret et al. 2014).

There are caveats about the quality of medicinal plant use data from herbarium specimens. Unlike publications, which report how the data were collected and from whom, the herbarium labels are silent about how the information was gathered. There

are no data about the ethnobotanical indices, such as informant consensus, recording the relative importance of the plants used or the distribution of knowledge in a community (Andrade-Cetto and Heinrich, 2011). It is even possible that the information about plant use is drawn from someone from outside of the community, such as a field assistant. Nevertheless, our tests to determine whether the overall content of herbarium labels shows that statistically the same plant parts are used in the same ways for the same therapeutic applications, whether data is sourced from herbarium or published sources. Specifically, the rank order of genera used, plant parts used and modes of application are not significantly different between sources. This statistical significance of the ranks of parts, modes and applications might hide more subtle differences between the ways in which ethnobotanists and general collectors report plant use. For example, the use of teas (extracts infused by pouring boiling water over them) is more commonly reported for the herbarium data than in literature reports, where decoctions (extracts concentrated by boiling) are more important. This might reflect a failure of more general collectors to distinguish between teas and decoctions. Similarly, the use of leaves may be under-reported in the herbarium labels, since leaves could be considered the "default" plant part, whereas in ethnobotanical literature that reports the plant parts used, use of leaves would be expected to be reported. Similar data content between publications and literature could be due to double-recording, if the authors of the publications have deposited specimens describing the uses they also document in their paper. We, as far as possible, rule out this explanation for shared patterns by showing that the collectors names on the specimens were the same as the author names of the publications only 3.6% of the time.

Whereas the genera recorded, and the way plants in general are used, are comparable between herbaria and literature, the geographical distribution of data are less similar. Ritter et al. (2015) reported in their survey of ethnobotanical research in Brazil, that the Atlantic Forest and Caatinga biomes are most studied, and that the Amazon, Pampa and Pantanal less frequently so. Our data from literature are not directly comparable, since we record numbers of reports, not articles, and the details of our research strategies differ. Nevertheless, we also find fewer literature reports from the Amazon, Pampa and Pantanal. Considering data from the herbarium, the same three biomes, Caatinga, Cerrado and Atlantic Forest, are also best known. However, the proportions of data from each source differ. For the less studied Pantanal, 58% of reports and 62% of the species known are known from the herbarium. Thus the herbarium goes some way towards meeting the deficit of information from this biome. This is not so for the Pampas, which lacks herbarium reports even more than literature reports (just 16% of reports and 18% of species from herbaria), nor for the Amazon biome (39% of reports and 39% of species from herbaria). For the Amazon, this might be because as Ritter et al. (2015) note, most Amazonian studies have been carried out by workers from outside of Brazil, and possibly related to governmental constraints on issuing permits for data collection. If these collections were made before a precedent for depositing Brazilian vouchers in Brazilian herbaria was established, the paucity of Brazilian specimens in our study with data could be because general collectors recording ethnobotany of the Amazon have not deposited their specimens in the herbaria we have surveyed. In general, these herbaria include fewer recent specimens from the Amazon, because of the logistical challenges of conducting fieldwork in remote areas and under challenging conditions, including amongst

communities less likely to speak Portuguese. Although the herbarium data are also few in areas with few literature reports, Figure 2-1 shows that data from otherwise unstudied sites are incorporated by including herbarium specimens in any study.

Beyond filling spatial gaps in our knowledge, herbarium data might validate and augment literature reports. We considered there was a need for validation of species with single reports. There are "legitimate" reasons a species might be cited only once; it might be genuinely rarely used, secret or specialist traditional knowledge or unshared uses (see Cámara-Leret et al. 2014). Alternatively, a single citation might result from misidentification. A second citation validates these reports. Misidentification is reported to be widespread in herbaria; it is estimated that more than 50% of tropical specimens are incorrectly named (Goodwin et al. 2015), although Goodwin et al.'s study was based on a taxonomic group known for having species hard to identify. In Brazil, continuous research effort in taxonomy, with updated species lists and floras being made (Flora do Brasil, 2020), can contribute to decreased misidentification. However, there is no reason to suppose that species identifications in ethnobotanists' studies are any more accurate; many identifications in the literature might be wrong since many ethnobotanists are not taxonomists. Herbarium specimens might be correctly determined when examined by experts at a later date, for this reason we might expect them to be more accurate. However, ethnobotanists frequently collect sterile material, whereas botanists almost never do, making it less likely that correct identification would take place as part of curation of herbarium collections. We found medicinal use of 21 of the 128 species known from only one report in the literature were substantiated from independently collected herbarium

specimens. Herbaria are shown here to add information to what is known from the literature. "Augmenting" literature reports in this way could therefore be another important use of herbaria. As we demonstrate herbaria can add significant data over what is known in the literature by specifying the use of plants.

Studies synthesizing regional or global plant use data may become more common as ethnobotanical research encompasses large-scale or meta-analysis (Albuquerque & de Medeiros 2012). Recent studies using data sets compiled from literature have tested whether there are global patterns of plant use between distantly related peoples (Saslis-Lagoudakis et al. 2012a), and determined the relative importance of different drivers of plant selection at national levels (de la Torre et al. 2012; de Medeiros et al. 2013; Saslis-Lagoudakis et al. 2014). Ethnopharmacological reviews also depend on the synthesis of data describing the therapeutic applications and modes of use of a taxon before interpreting use in the light of chemistry and pharmacological activity (Fernandes and Banu 2012; Seebaluck et al. 2015; Wang et al. 2013). Reviews can highlight knowledge gaps and prioritize communities and areas for field research (Uprety et al. 2012); nevertheless the scope and reach of many meta-analyses in ethnobotany demand more data than can be delivered by directed fieldwork programs. The availability of data and its quality poses a significant barrier to further study (de Medeiros et al 2013). This deficit of information, for certain types of study, might be overcome by using herbaria.

Herbaria are irreplaceable resources; the long-term capital investment in specimen acquisition and curation that they represent is increasingly rewarded across

many fields (Funk 2003). Our study, by explicitly quantifying ethnobotanical data from herbaria and literature, demonstrates that herbaria contain valuable information that can both supplement and complement literature reports. Visits to herbaria to check specimens for associated data would be extremely time consuming; our study was facilitated by fast, easy access to data from online herbaria. The ease in which data are accessed is set to increase as digitization projects progress (Haston et al. 2012), generating more and more accurate data. Ethnobotanical equivalents of the Darwin Core Biodiversity Data Standards (Wieczorek et al. 2012) might be envisaged, to maximize the value of digitisation efforts. We anticipate a wider role for herbarium data in the future.

Chapter 3 Ewé: a web-based ethnobotanical database for storing and analysing data

3.1   Introduction

The documentation of ethnobotanical information and associated knowledge (i.e. anthropology, botany, ecology, chemistry and pharmacology) is essential for a better understanding of the relations between humans and plants (Balick and Cox, 1996; Berlin, 1992), and to progress ethnobotanical and ethnopharmacological studies (Stepp and Thomas, 2010). The use of databases in ethnobotanical research can facilitate data management, analysis and also information sharing with other researchers (Ningthoujam et al. 2012). Nevertheless, database sustainability and collaboration can be challenging (Ningthoujam et al. 2012). Ethnobotanical databases found today may have an international focus (e.g. NAPRALERT which has a pharmaceutical focus (Loub et al. 1985)) or a regional one (e.g. Prelude database, which is focused on African medicinal plant use (Noe, 2012)). Others can have a different scope, such as the Medicinal Plant Names Services, from Kew Royal Botanical Gardens (available in http://mpns.kew.org/) that curates medicinal plants names, enabling correct identification of medicinal species in order to assure safety of use. The documentation and exploration of traditional knowledge (TK) are important contributors to bioprospecting (Helmstädter, 2017; Shaw, 2017). Documentation of TK can also contribute to the protection of intellectual property in the context of

bioprospecting (Downes, 2000), where traditional knowledge is identified as the property of communities that should be the recipients of benefit-sharing (Alexander et al. 2004), and to the preservation of traditional knowledge. Knowledge erosion is caused by cultural change, modernization, access to western medicine (Voeks and Leony, 2004), and a lack of interest amongst young people in learning about medicinal use of plants (Brouwer et al. 2005; Phillips and Gentry, 1993; Ugent, 2000). Finally, databases could highlight traditionally used species and direct research that might contribute to safety assessments (Calixto, 2000; Ekor, 2014; Jordan et al. 2010; Moreira et al. 2014). Here we present a web-based ethnobotanical database not only serving to store data, but also with tools to conduct simple analyses and visualise data.

Ethnobotany can be considered a recent discipline, being firstly defined and delimited in 1896 with the publication "The purposes of Ethnobotany" (Harshberger 1896). Since then, many studies concerning documentation of traditional knowledge have taken place. The increasing amount of data produced by such studies allows ethnobotanical research to address more comprehensive and comparative questions in order to better understand plant use and selection, yet data distribution and availability are barriers to this (Thomas et al. 2001). Despite the advances made so far in ethnobotany, more collaborative ethnobotanical research could be achieved by bridging the gap between ethnobotanists and modern bioinformatics (Thomas 2003). The development of ethnobotanical databases can bridge this gap in different ways. For example, the compilation of ethnobotanical data, quite often spread across different institutions or sources, such as publications, herbarium specimen labels or even unpublished theses and governmental reports that are of hard access, in a

35

database, can facilitate large-scale analysis by decreasing the time of data collection (Schalk and Oosterbroek 1996, Souza and Hawkins, 2017, Thomas 2003).

Until now, many different ethnobotanical databases have been developed, many with online access. Many of these databases were developed for particular user groups, academic or non-academic, and with different foci, geographic or cultural (Thomas, 2003). Without a central authority, or guidelines for data collection and storage, challenges may arise. Firstly, although the existence of such databases contributes towards data availability, data is not always comparable, and this is a barrier for collaborative and large-scale research (Thomas 2003). Secondly, the sustainability of such databases is associated with the continuity of a research group or project, long-term data availability is not ensured. These challenges can be added to others, such as data accuracy and intellectual property rights (Ningthoujam et al. 2012; Heinrich 2009).

Ethnobotanical databases can be important tools to both preserve traditional knowledge and advance in ethnobotanical research (Moerman, 2013, Heinrich et al., 2009). It has been reported that traditional knowledge, available publicly in databases can protect against biopiracy or misappropriation (Gupta 2004). Traditional knowledge holders can also benefit from ethnobotanical databases, when analysis and data generated and interpreted by the scientific community is returned to local populations so that it can both preserve their knowledge and also add new information or value to it. Examples of this can be seen when indigenous populations knowledge of foraging and phytomedicines are enhanced by scientific corroboration (Edwards and Heinrich,

2006)., Ethnopharmacological research can also add value, benefiting knowledge holders in the form of benefit-sharing as defined by the CBD. The "Access and Benefit Sharing" (ABS) mechanism was proposed in 1992 in the CBD and implemented in the Nagoya protocol since then (United Nations 2010). Although there is a deficit in studies to track the effectiveness of ABS, a few case-studies are currently available, including one in Brazil that shows a positive relationship between industry and traditional knowledge holders in supplying raw material (Laird and Wynberg 2008). On the other hand, it is argued that databases of traditional knowledge can, when knowledge is compiled as data or information, create barriers to the recognition of a proper heritage, particularly in the case of large well estabilished systems of Traditional Knowledge, such as Ayurveda, increasing claims of ownership (Reddy, 2006).

Ethnobotany and ethnopharmacology are both interdisciplinary subjects, and standardizing data from different studies is difficult when research questions can have different emphasis (Stepp and Thomas, 2010). Nevertheless, ethnobotanical studies have attempted to address broad questions using published data, an approach sometimes termed meta-analysis (de la Torre et al. 2012; de Medeiros et al. 2013). Meta-analyses will benefit from tools to summarize database entries, for example the diversity of species used for a certain therapeutic application, or a map of showing the geographical distribution of a species' use can be generated.

Over the last 30 years the number of publications describing ethnobotanical studies in Brazil has increased (Ritter et al. 2015). The amount of information

generated and recorded in peer-reviewed scientific publications has the potential to inform meta-analyses, but the information should be made available in a useful format. In Brazil, Plant-Database (Manhã et al. 2008) provides an extensive bibliographical data compilation with more than 2000 papers, many focused on pharmacological and drug discovery data. There is a need in Brazil for a database that provides easy access and visualization of species information. Open-source tools have been used to build the database, recognising that whilst funding for maintenance and development of databases is difficult to secure (Reiser et al. 2016), a community of users may contribute if the database can be modified to respond to their changing research needs.

3.2   Methods

3.2.1   Database development

A web-based app was developed using MEAN (MongoDB, ExpressJS, NodeJS and AngularJS, available at http://meanjs.org/). Information on the species was organized in a spreadsheet with the fields: family, genus, species, authority, synonym, common name, use category (according to WHO-ICD 10) use category (traditional, as recorded in the field), plant part used, mode of application, country, city, biome, latitude, longitude, reference (paper author or herbarium name), collector and origin (literature or herbarium). Subsequently it was exported in format JSON for indexing and searching. The database homepage is shown on Figure 3-1.

3.2.2   Data collection

Published medicinal use data, and data from Brazilian herbarium specimens were compiled for the Leguminosae of Brazil. Where necessary, geographical coordinates were assigned using Google maps and the Brazilian Geographic and Statistical Institute online data (IBGE). Generic and species names followed THE PLANT LIST (The Plant List, 2013) and Missouri Botanical Garden's Tropicos database (Tropicos.org, 2016), and were corrected using the Plantminer R script (Carvalho et al. 2010)
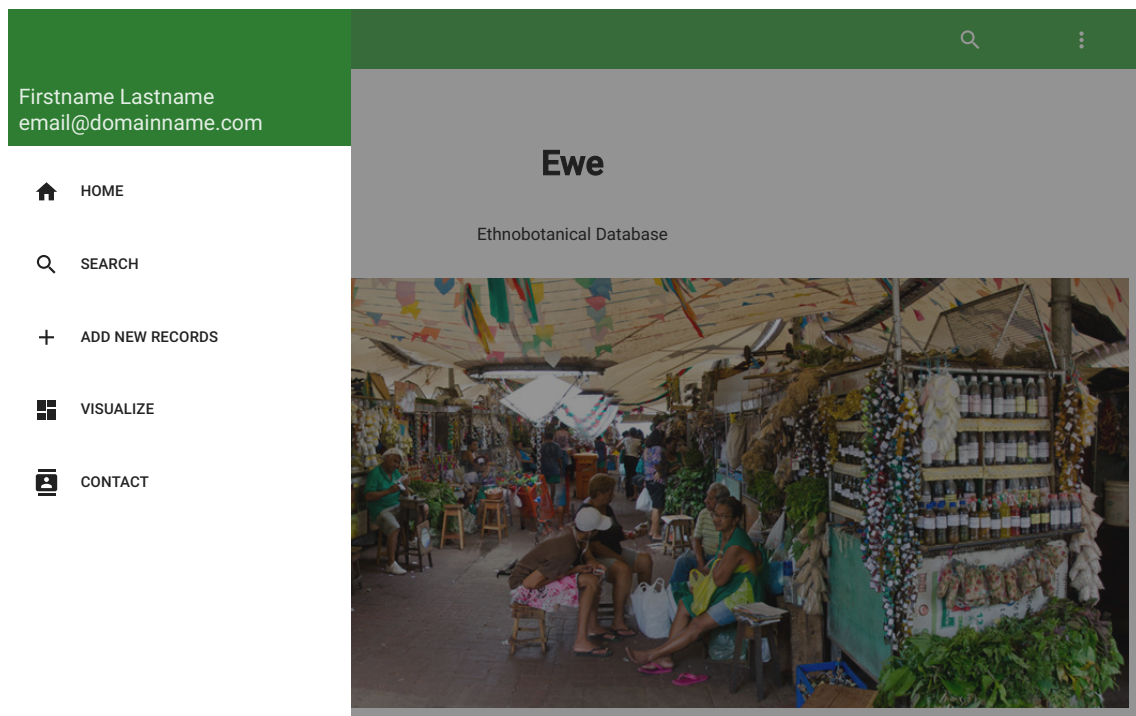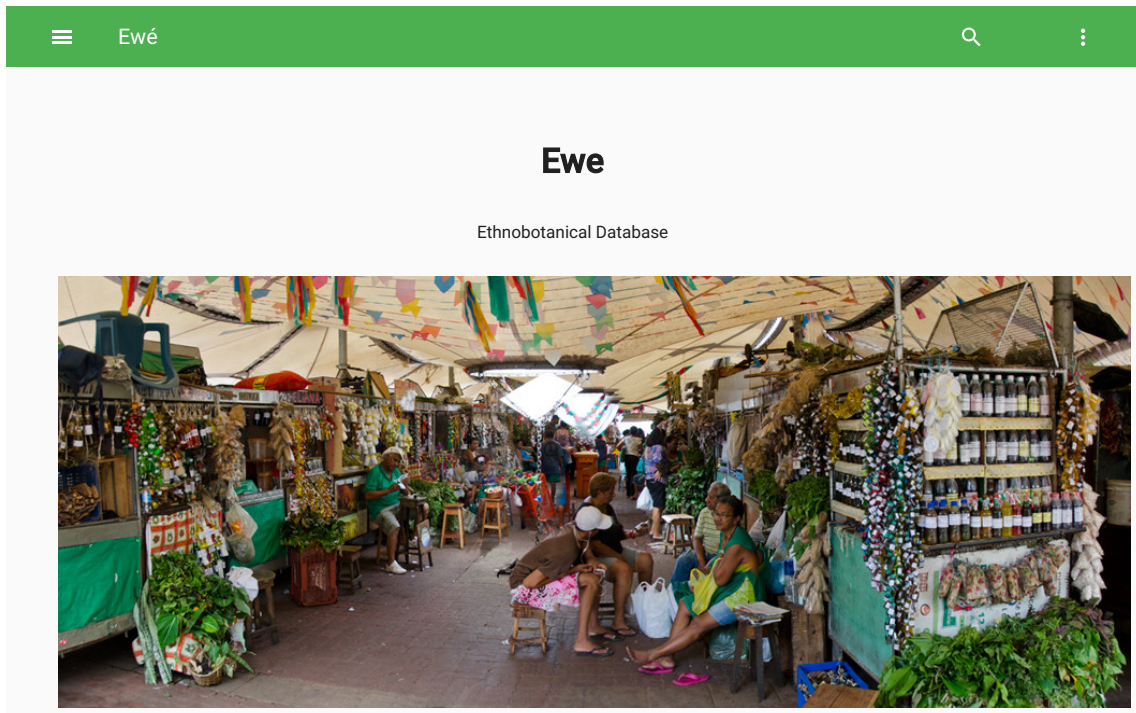
Figure 3-1 Homepage of Ewé (top) and side menu with search and visualization options (bottom).

Publications citing medicinal uses of Leguminosae species in Brazil were identified using Google Scholar and searches of following journals: Acta Botanica Brasilica, Economic Botany, Fitoterapia, Flovet, Journal of Ethnobiology, Journal of Ethnobiology and Ethnomedicine, Journal of Ethnopharmacology, Journal of Medicinal Plants Research, Revista Brasileira de Biociencias, Revista Brasileira de Farmacognosia, Revista Brasileira de Plantas Medicinais and Rodriguesia. Herbarium data were extracted from the online list of herbarium and biological collections from Brazil, Species Link (http://splink.cria.org.br/). The file was exported to Excel format and a search using the keywords "medicinal" and "uses" (in Portuguese) was conducted in order to filter only the specimens with medicinal information. The list of publications and herbaria sourced here are in Appendix 3

.

3.2.3   Data organization

Published medicinal use data, and data from Brazilian herbarium specimens were compiled for the Leguminosae of Brazil. Using the 1400 use reports from Chapter 2, and additional collected data, we generated 2 078 reports for databasing. When a species was reported in one publication to be used to treat for influenza and headaches, the two therapeutic applications generated two different reports, though the species is the same.

In order to standardize the traditional uses we adopted the World Health Organization international classification of diseases 10 (WHO ICD-10) as described in table 3-1, although there are deficiencies associated with standardisation (Staub et al. 2015). Biomes were assigned according to the geographical coordinates from each

record, plotted on a map and intersected with the Brazilian Biomes shapefile in QGIS

2.2.0. Mode of application data were classified according to Table 3-2.

Table 3-1. Codes for WHO-ICD 10 used to classify traditional information.

World Health Organization international classification of diseases 10	CODE
Certain conditions originating in the perinatal period	CCP
Congenital malformations, deformations and chromosomal abnormalities	CMC
Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	DBI
Diseases of the circulatory system	DCS
Diseases of the digestive system	DDS
Diseases of the eye and adnexa	DEA
Diseases of the ear and mastoid process	DEM
Diseases of the femalegenito system	DFS
Diseases of the genitourinary system	DGS
Diseases of the musculoskeletal system and connective tissue	DMC
Diseases of the nervous system	DNS
Diseases of the respiratory system	DRS
Diseases of the skin and subcutaneous tissue	DSS
Endocrine, nutritional and metabolic diseases	ENM
Certain infectious and parasitic diseases	IPD
Mental and behavioral disorders	MBD
Neoplasms	NEO
Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified	
Injury, poisoning and certain other consequences of external causes	O T H E R S
External causes of morbidity and mortality	
Factors influencing health status and contact with health services	
Codes for special purposes	
Pregnancy, childbirth and the puerperium	PCP

Table 3-2. Classification of different medicinal preparations into specific modes of application categories.

Terms used in database	Terms used in publications or specimens notes
Decoction	Decoction
Infusion	Tea, infusion, with water, with milk
Tincture	Wine, alcoholic infusion, alcoholic extraction
Maceration	Maceration
Paste	Topic use, plaster
Juice	Drink, juice
Wash	Bath, soap, gargle, mouthwash
Food	Edible, raw, *in natura*
Oil	Oil
Syrup	Syrup, with honey
Powder	Powder/ inhale

3.3   Results

3.3.1   Database content

Two thousand and seventy-eight records were compiled to date. These were sourced from 108 publications (1331 records) and 54 herbaria (747 records). Until now, EWÉ is focused on the Leguminosae from Brazil, with 322 species in 117 genera with medicinal uses. For all the databased records, 1165 (56%) indicate therapeutic applications (traditional/WHO uses), 948 (45%) indicate plant parts used and 790 (38%) indicate the modes of application (appendix 4). All 6 Brazilian biomes are indicated in our records, with Caatinga and the Cerrado the best represented.

Literature and herbaria differ in information content. Literature provided 87% of therapeutic applications, 95% of modes of application and 95% of parts used. On the other hand, herbaria contributed to 86% of cited geographical localities.

Senna and *Bauhinia* were the most cited genera, with medicinal use reported in all 6 biomes. For therapeutic applications, Diseases of the Respiratory System (DRS), Diseases of the Digestive System (DDS), Endocrine Nutritional and Metabolic Diseases (ENM), were the categories with most citations, bark and leaves the most commonly used plant parts, and infusion and decoction the most common modes of application (Figure 3-2).
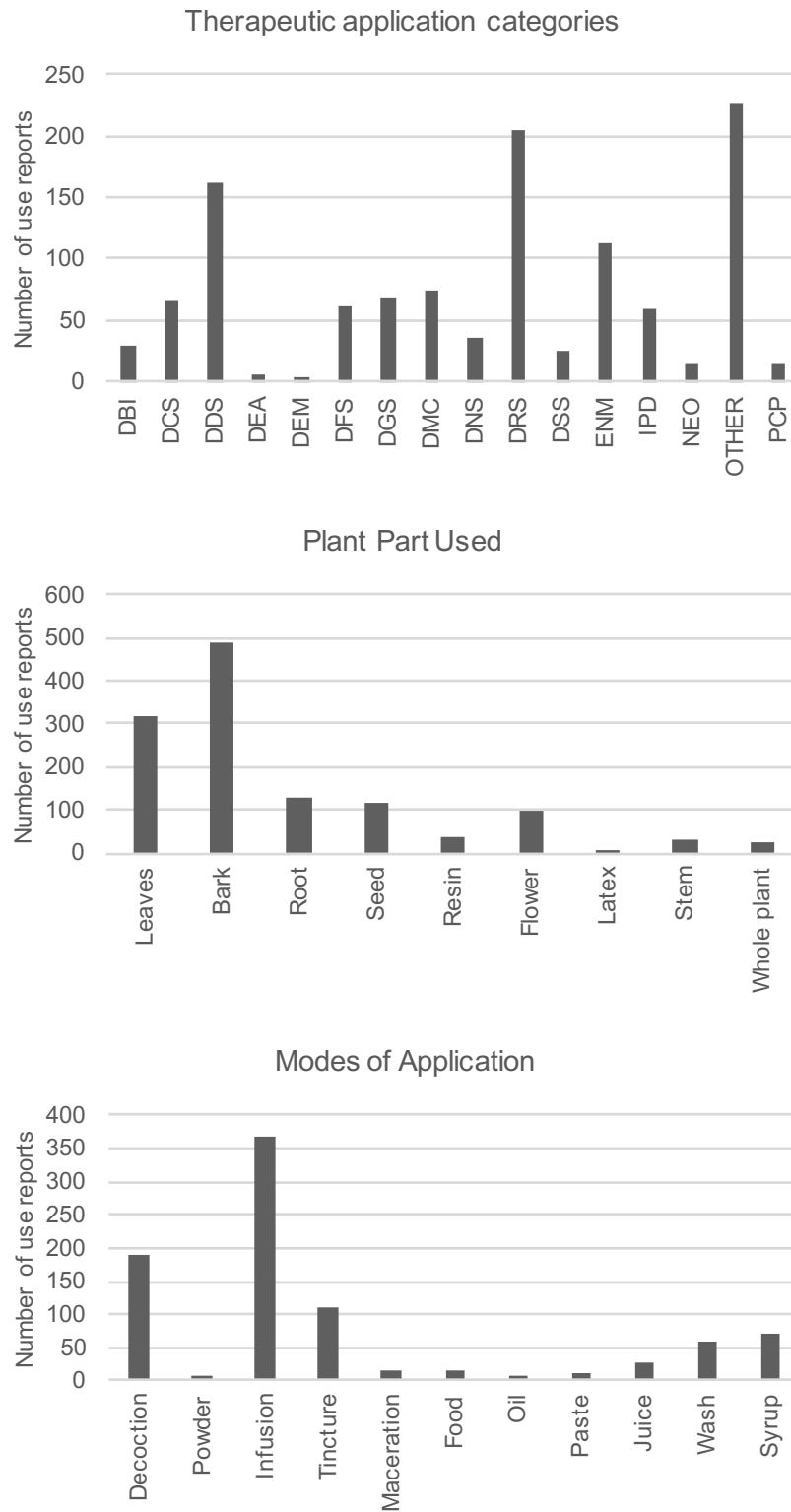
Figure 3-2. Number of use reports for each category in Therapeutic Applications, Plant Parts Used and Modes of Application.

Database search tools

When visualized in the search tool, records are shown as a list that can be organized according to any of the categories (Figure 3-3). A search field also allows the user to filter the search by any term shown in the list (Figure 3-4).
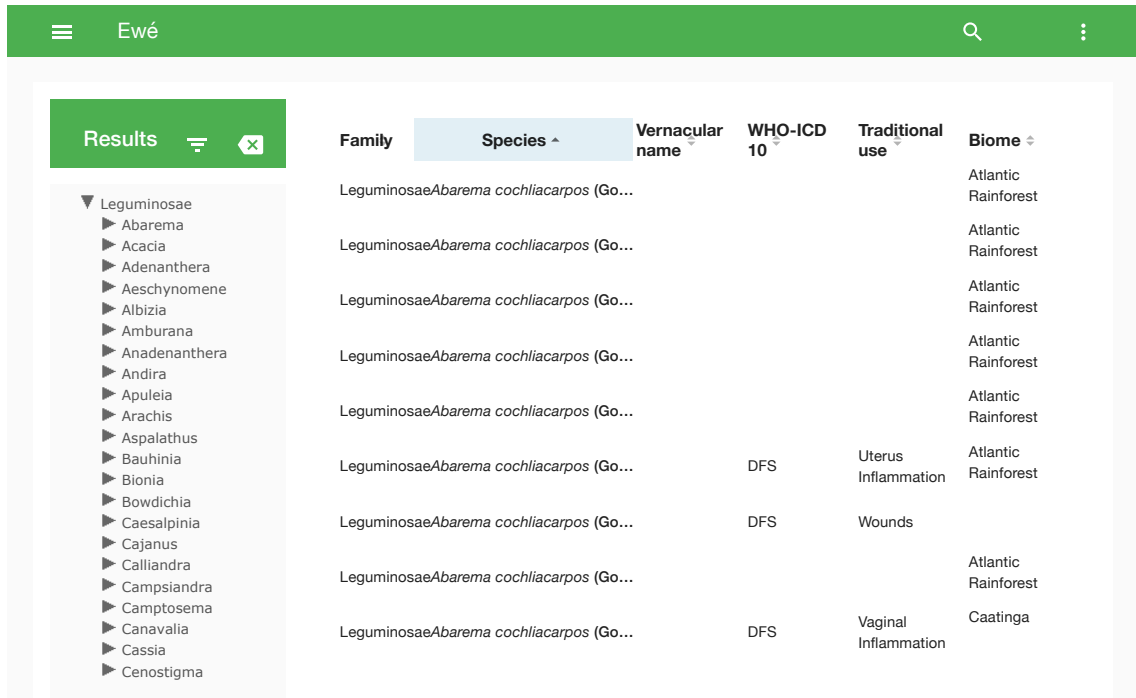
Figure 3-3. Search tool of Ewé, showing the number of records and data contents of each field of a chosen sample.
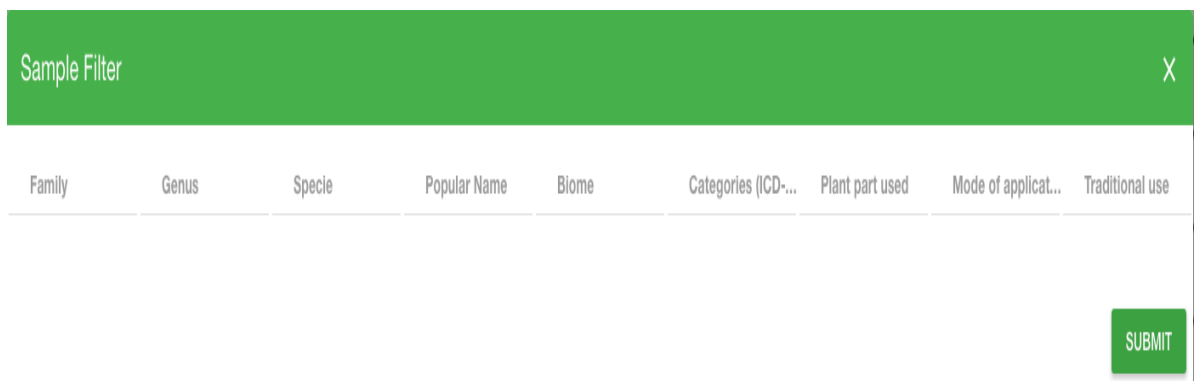
Figure 3-4. Detail of the filter tool from the search area.

Clicking on a record, opens the full record, including a map if the record has coordinates, below the records list (Figure 3-5).

Figure 3-5. Screen-shot of a full record (*Caesalpinia echinata* Lam.) illustrating its described uses, the part used, traditional use, forms of use (modes of application) and the paper reference from where it came.

3.3.2 Database visualization

Data are visualised as a tree map with three interactive levels, family, genus and species (Figure 3-6 top). For all levels the size of the square is proportional to the number of records (Figure 3-6 middle).
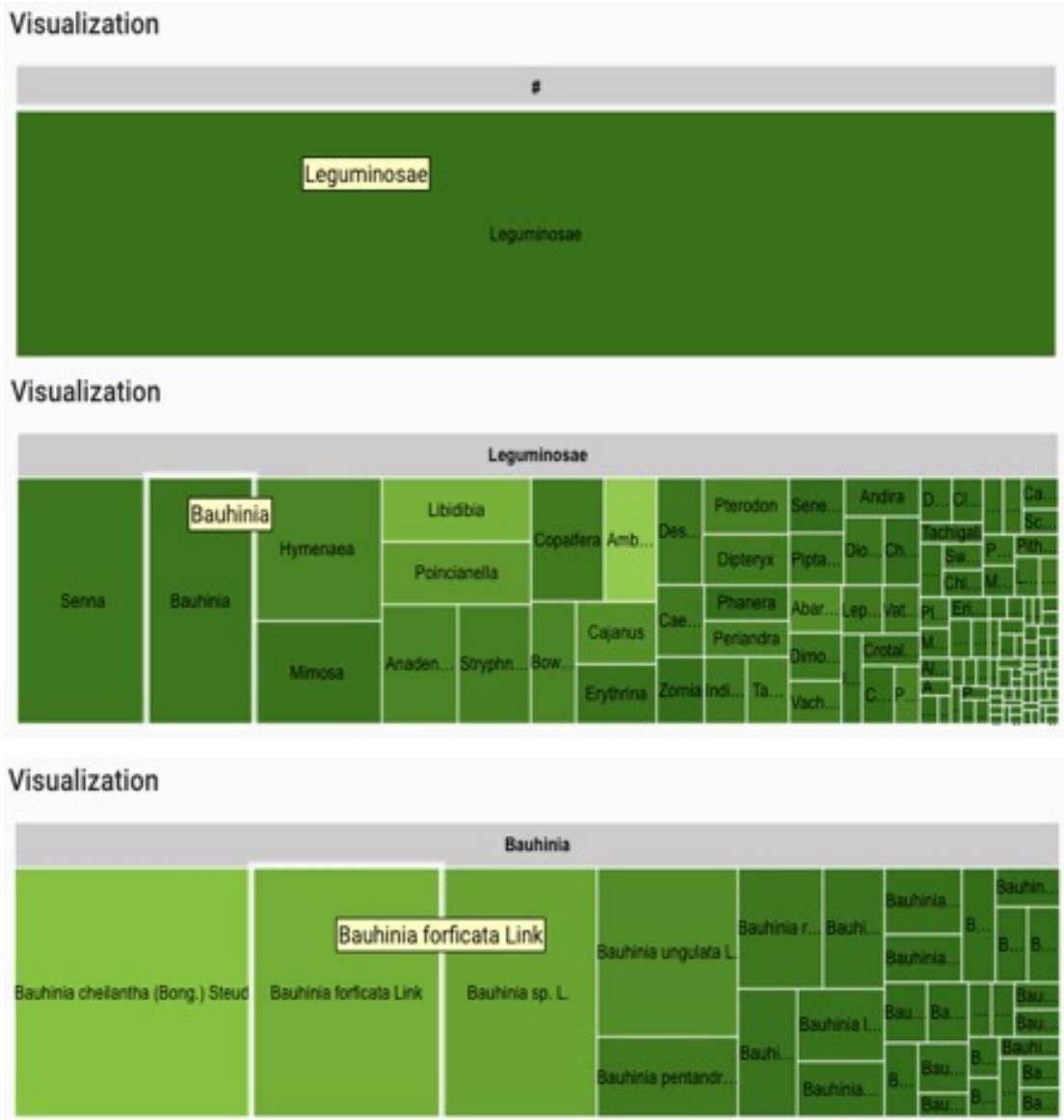
Figure 3-6. Detail of the visualization tool from EWÉ showing the tree map.

Below the tree map there are four graphs: therapeutic applications (top left), biomes distribution (top right), plant parts used (bottom left) and modes of application (bottom right). These four graphs change according to the taxonomic scale of the tree map. Also user selection of one bar in each histogram is possible (Figure 3-7).
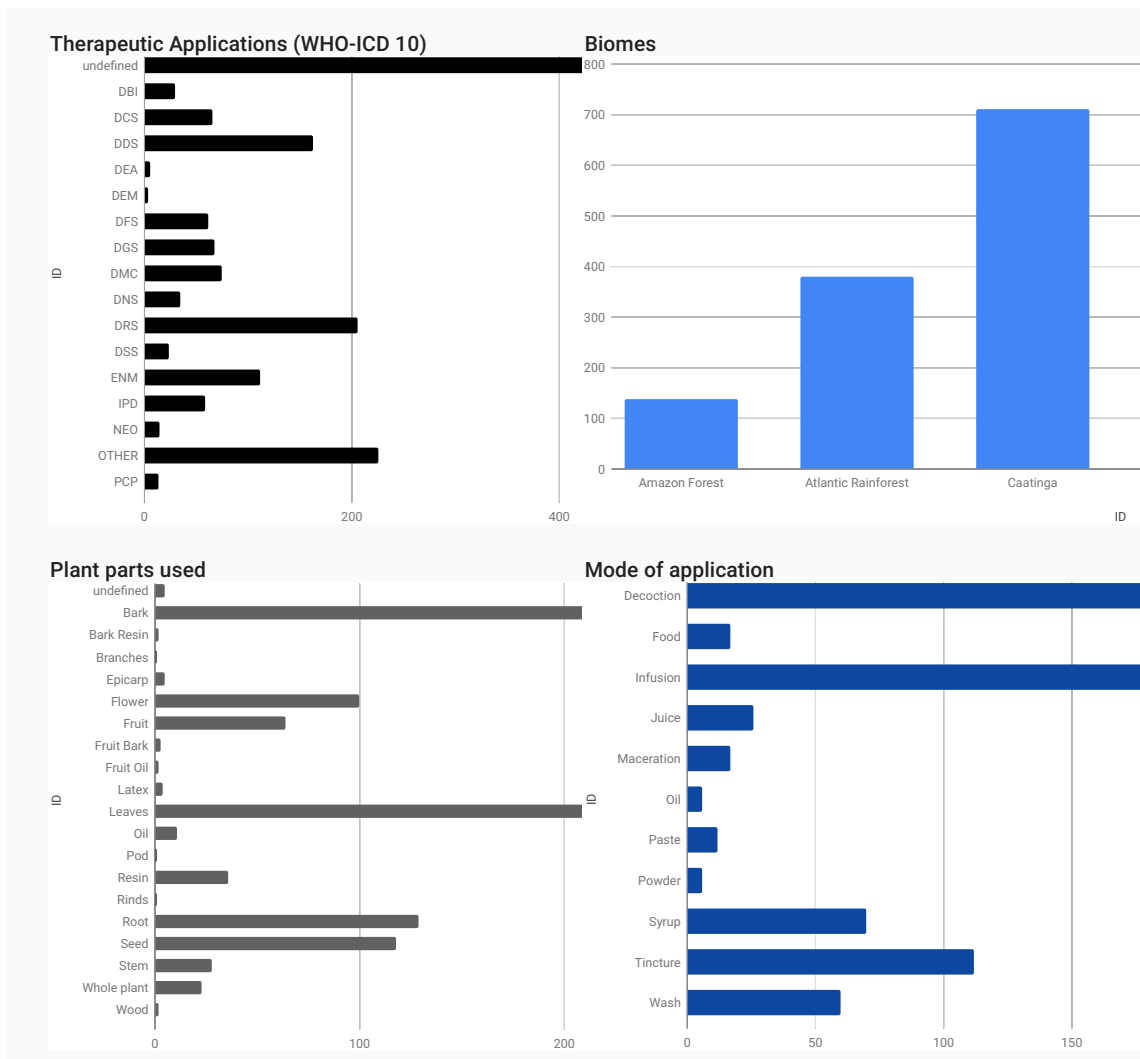
Figure 3-7. Detail of the visualization tool from EWÉ showing graphs with therapeutic applications, biomes, plant parts used and modes of applications.

3.4   Discussion

The internet has contributed to an explosion of data sharing amongst scientists, including ethnobotanists, thus contributing towards a better understanding of the relation of humans and plant use (Gaikwad et al. 2011; Albuquerque & Medeiros 2012). Nevertheless, much ethnobotanical data is still difficult to access, sometimes the result of inadequate data management (Gaikwad et al. 2011). Aside from storing a large dataset, we presented a database capable of data visualization permitting data exploration. To the best of our knowledge, Ewé is the first medicinal plants database to provide tools for data storage and visualization for Brazilian plants. Research can be facilitated using visualization tools, since hypothesis generation is supported by examining patterns and gaining insight into the data (Keim, 2001). Geographical areas poor in data can be identified, so future studies can be directed to areas of deficient knowledge.

Challenges in integrating data from different sources into a database, or even from different regional databases into a more comprehensive one are related to data standardisation and data management (Ningthoujam et al. 2012). Efforts to provide a common standard for ethnobotanical data were made by the International Working Group on Taxonomic Databases for Plant Sciences (TDWG), who presented the Economic Botany Data Collection Standard (EBDCS) (Cook, 1995). Although not all fields proposed by EBDCS are present in Ewé, many such as source of information on use, use, plant part used, organism, vernacular names, modes of application, are present here. Fields such as season of use, conservation data, ratings or popularity,

49

problems and potential, though important, can be hard to source from available material. In the future, Ewé can expand to include primary data, and information related to season of use and popularity can be included.

EBCDS also proposes categories of therapeutic applications and has been successfully applied by ethnobotanists when compiling traditional knowledge data, although it has also been argued that not all plants and therapeutic applications can follow this standard (Gruca et al. 2014). Categories used here were according to the WHO-ICD 10, many of them are very similar with the ones proposed by the EBDCS. The WHO-ICD 10 categories were used to facilitate data classification, so users can consult the WHO website and search for the appropriate term. This classification also allows information from Ewé to be compared with other studies that used, at least in part, the same classification (Baptista et al. 2013; Ernst et al. 2016; Leonti, 2011; Saslis-Lagoudakis et al. 2012a). Nevertheless, traditional medicinal use that were commonly cited such as pain, fever and inflammation could not be associated with any specific medical condition on the WHO ICD-10 standard, and therefore were assigned to "others". On the other hand, the original terms or classifications of therapeutic use are still provided, in order to contribute to the conservation of traditional knowledge.

Herbarium vouchers are potential sources of ethnobotanical information to populate databases. As efforts in digitisation for herbarium specimens increase (Beaman and Cellinese, 2012; Hill et al. 2012; Tegelberg et al. 2014; Thiers et al. 2016; van Oever and Gofferjé, 2012), again standardisation is needed and the Darwin Core standard is used to achieve this (Wieczorek et al. 2012). However, core descriptors do

not include any ethnobotanical fields, so data describing traditional knowledge present in herbarium vouchers is not captured. Herbarium vouchers have been described as important sources of ethnobotanical data (Eisenman et al. 2012; Nesbitt, 2014; Senchina, 2006; Souza and Hawkins, 2017). The inclusion of ethnobotanical information in the core descriptors for digitisation could be beneficial.

Database sustainability is recognised as a significant problem for the biological sciences (Attwood et al. 2015; Bastow and Leonelli, 2010; Merali and Giles, 2005; Ningthoujam et al. 2012). Being an open-source scheme, Ewé can be used on a private server as a research tool, and users can modify the code for their own needs. Furthermore, as an open-source scheme, the database can evolve to meet changing requirements. Also, Ewé could function as a collaborative database, where researchers can include their own data for private use but would be encouraged to contribute to the publically available data. Moreover, when Ewé is established in Brazil, projects to populate the database with information directly from the holders of TK, as a way to preserve their knowledge and assure rights to intellectual property, could be implemented.

As an online database, we expect Ewé to be a source of information and knowledge sharing between ethnobotanists in Brazil. Although at the present moment Ewé is focused on the medicinal use of the Leguminosae, more taxa could be included, perhaps the entire medicinal flora of Brazil. Nevertheless, data quality should be emphasized in order to increase the quality of the database.

Chapter 4 Phylogenetic exploration of plants used in ethnomedicine reveals significantly-used lineages are better characterized

4.1 Introduction

The search for patterns in plant use is the focus of extensive investigation in the fields of ethnobotany and ethnopharmacology (Albuquerque et al. 2007; Bennett and Husby, 2008; de la Torre et al. 2012; de Medeiros et al. 2013; Ladio and Lozada, 2004; Monteiro et al. 2006; Saslis-Lagoudakis et al. 2011a; Wayland and Walker, 2014; Zhu et al. 2011). At the present time, phylogenetic methods are emerging as powerful tools to investigate such patterns (Cámara-Leret et al. 2017; Ernst et al. 2016; Grace et al. 2015; Saslis-Lagoudakis et al. 2015; Saslis-Lagoudakis et al. 2012a; Saslis-Lagoudakis et al. 2011a; Zhu et al. 2011). It has long been known that plants with medicinal use are not distributed randomly in a flora, but are usually found more frequently in some groups than in others (Bennett and Husby, 2008; Moerman, 1991; Moerman et al. 1999; Saslis-Lagoudakis et al. 2011a; Weckerle et al. 2011). By identifying lineages rich in species with ethnomedicinal use or metabolites of interest, phylogenetic methods are currently being evaluated for their potential for predictive bioprospecting (Ernst et al. 2016; Grace et al. 2015; Rønsted et al. 2008; Saslis-Lagoudakis et al. 2012a; Yessoufou et al. 2015).

The use of phylogenetic methods to investigate phylogenetic structure of ethnomedicinal plants overcomes the need for taxonomic ranks (see Moerman et al. 1999, Bennett and Husby 2008, Weckerle et al. 2011), by measuring relatedness directly from a phylogeny (Saslis-Lagoudakis et al. 2015). Different tests are used to measure the relatedness of ethnomedicinal species, with the most common being the mean phylogenetic distance (MPD), mean nearest taxon distance (MNTD) and their related indices, net related index (NRI) and nearest taxon index (NTI) described by Webb et al. (2008). Besides the identification of phylogenetic structure, it is also possible to identify lineages in the phylogeny that are contributing to the structure. These lineages are called hot nodes (Saslis-Lagoudakis et al. 2012a), and are the nodes in a phylogeny whose descendent species include significantly more medicinal species than other nodes.

Closely-related species share similar secondary metabolites that could be of interest (Wink, 2003) and phylogenetic structure in the distribution of secondary metabolites has been well documented for the Leguminosae (Wink, 2013). The phylogenetic signal in phytochemistry has stimulated the development of new bioprospecting strategies (Halse-Gramkow et al. 2016; Rønsted et al. 2008). Despite the low number of species with ethnomedicinal use to have been investigated pharmacologically (Fabricant and Farnsworth 2001), plant material has been of great importance in bioprospecting, where ethnomedicinal use has guided investigation (Gurib-Fakim, 2006). In terms of phylogeny, describing lineages with congruent uses across different communities (Saslis-Lagoudakis et al. 2012a), thus highlighting such taxa, or identifying lineages with a strong phylogenetic signal in therapeutic

applications or properties of interest (Ernst et al. 2016), phylogenetic methods can be used to identify priority taxa for pharmacological study.

The Leguminosae (legumes) is the third largest plant family (Lewis et al. 2005). With many documented uses, the family is over-utilized for medicine in some regions (Korea and Ecuador) but under-utilized in others (North America) (Moerman et al. 1999). In Brazil, the family comprises c. 2,800 species in more than 200 genera (Flora of Brazil 2020 in construction). The species diversity, widespread distribution and many reported uses (Souza and Hawkins, 2017) and the availability of phylogenetic information (LPWG et al. 2013), combined with recognized pharmacological potential have prompted us to select the family as a case study. Drugs such as Danthron (used as a laxative, derived from *Cassia* sp.; Maridass and De Britto 2008) and Monocrotaline (an anti-tumor agent derived from *Crotolaria sessiliflora*; Wink, 2013) are drugs derived from species distributed in Brazil and used globally. Two further species are found in the official Farmacopeia Brasileira, *Senna auriculata* and *Stryphnodendron coriaceum* (Anvisa, 2010).

Fabricant and Farnsworth's study of 2001, showing more than 90% of ethnomedicinal plant species are pharmacologically unscreened, is still frequently cited to indicate our poor knowledge of the properties of plants used in ethnomedicine. The first objective of this study is to provide a contemporary estimate, at least for one plant group in one area. The description of the phylogenetic diversity of medicinal uses of the Leguminosae in Brazil enables identification of lineages rich in ethnomedicinal species. Whether "hot node" lineages have been the focus of screening efforts is

investigated in order to identify and explain patterns in ethnopharmacological research. Until now, phylogenetic investigations of ethnomedicinal plants have considered global floras (Halse-Gramkow et al. 2016), whole floras (e.g. Saslis-Lagoudakis et al. 2012a) or genera throughout their range (e.g. Saslis-Lagoudakis et al. 2011a, Ernst et al. 2016), and have used data based on presence/absence of use. As well as being the first study focused a family level, this study also takes into account the frequency of reported use per species, in order to shed more light on patterns of ethnomedicinal plants.

4.2 Materials and Methods

4.2.1 Ethnomedicinal use of Legumes in Brazil

Published ethnomedicinal use data were compiled for the Leguminosae of Brazil as shown in chapter 2 of this thesis. The information gathered was databased also as reported in chapter 3 of this thesis. Generic and species names followed the Flora of Brazil (Brazilian Flora, 2020), and were corrected (standardized) using the Plantminer R script (Carvalho et al. 2010).

4.2.2 Pharmacology of species used in Brazil

Pharmacological data associated with the Brazilian ethnomedicinal species, and for all genera, whether used or not, were sought online using "Google Scholar" and "Web of Science" (WoS). Search terms included the species name and commonly used synonyms, together with "pharmacolog*". The number of results from "WoS" for each species was recorded, together with publications related to pharmacological studies.

4.2.3 Phylogenetic reconstruction and analysis

In order to understand the distribution of ethnomedicinal species in the family, a phylogenetic tree for the Leguminosae in Brazil was reconstructed at genus level, with one species per genus being sampled. Sequences from the DNA marker MatK were compiled from the Tree of Legumes (LPWG, 2010) when present, and from publicly available sequences on Genbank (appendix 5). In total 217 genera were sampled (91% of the total flora). Sequences were aligned using MAFT alignment and adjustments were made manually. Sequence data were analyzed under the maximum-likelihood (ML) criterion, under the GRTGAMMA model as implemented in RAxML

7.2.8 (Stamatakis, 2014). Unsampled genera and species were added manually as polytomies using R, with the package PHYTOOLS and the function add.species.to.genus to generate a species tree (Revell, 2012). To make the tree ultrametric, we used the function chronos from the package APE (Paradis et al. 2004).

The phylogenetic structure of ethnomedicinal uses was investigated using the Net Relatedness Index (NRI) and Nearest Taxon Index (NTI) (Webb, 2000; Webb et al. 2002), with the functions ses.mpd and ses.mntd from the PICANTE package (Kembel et al. 2010). The chosen null model was "taxa.labels", which shuffles the distance matrix labels across all taxa included in distance matrix with 999 runs. The NRI and NTI values were obtained by multiplying the MPD and MNTD values by -1. The NRI value represents the average distance between a species and all other species with the same characteristic (in our case ethnomedicinal use, or a specific category of therapeutic application). Higher values of NRI (negative for MPD) indicates a phylogenetic clustering of a sample, while negative values (positive for MPD) show an overdispersion of the studied taxa. Similarly, NTI (and MNTD) values can indicate clustering or dispersion. According to Webb (2000), NRI values describe a deep relationship in the phylogeny, which here might be evidence for clustering or overdispersal at tribal or sub familial levels. On the other hand, NTI describes relatedness in terminal clades, in our case clustering or overdispersion of congeneric ethnomedicinal plants (Figure 4-1). Phylogenetic structure was investigated for the ethnomedicinal species compared to the whole Leguminosae flora in Brazil.

"Hot nodes" (as described on Saslis-Lagoudakis et al. 2012a) were sought using the NODESIG function in also in R (R Core Team, 2015) adapted by Abellán et al. (2016). We recorded the percentage of the total and ethnomedicinal flora included in the "hot nodes", as well as the "gain" of ethnomedicinal species in a hot node when compared to random samples, for overall ethnomedicinal use and for each specific category of therapeutic application (classified according to the WHO ICD-10).
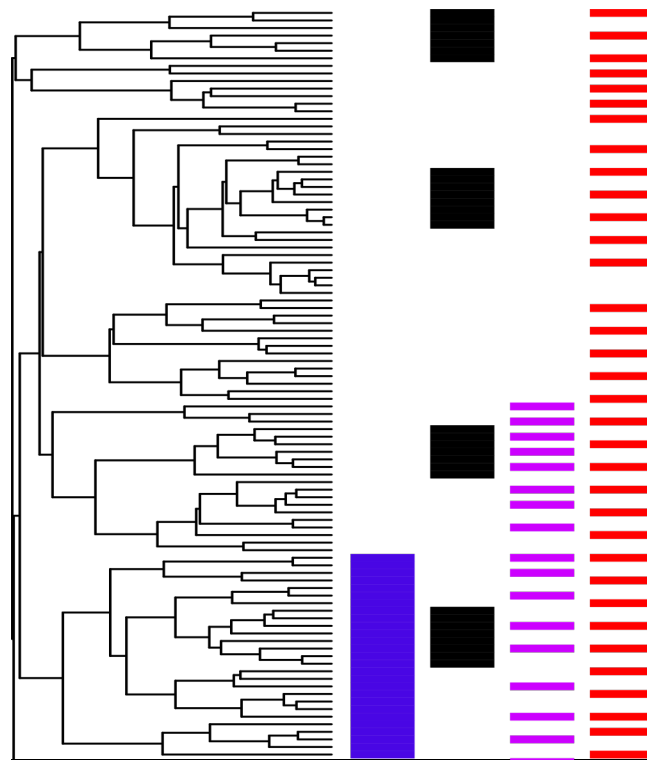
Figure 4-1. Schematic representation of scenarios of relatedness. Clustering and overdispersion are inferred from MPD and MNTD values (and their related NRI and NTI indices respectively). MPD describes relationships that occurred on deep nodes of the phylogeny, while MNDT reflects relationships at shallower nodes. Combined values can be used to infer different scenarios. Blue- Phylogenetic clustering, MPD and MNTD positive. Black- Overdispersion of clustering, MPD negative and MNTD positive. Pink- Clustering of overdispersion, MPD positive and MNTD negative. Red- Phylogenetic overdispersion, MPD negative and MNTD negative. (adapted from Mazel et al. 2016).

The relationship between pharmacological study and ethnomedicinal use was investigated. Firstly, we considered whether there are more species that have been the focus of pharmacological study within the lineages that are hot nodes for ethnomedicinal use. We also considered whether intensity of ethnomedicinal use of species for medicine was greater in the lineages that are hot nodes for ethnomedicinal use, and whether the intensity of research effort was greater within hot nodes. Intensity of ethnomedicinal use was estimated based on the number of citations/use reports in the database. Intensity of research effort was measured by counting the number of reported studies (publications) for each species. The intensity of research effort was compared for the species ethnomedicinally-used for medicine inside and outside of the hot nodes. A Wilcoxon test was used to determine whether there are differences in the number of pharmacological studies and ethnomedicinal uses between species in hot nodes and outside hot nodes than would be expected if pharmacological effort and ethnomedicinal use was homogeneous.

4.3 Results

4.3.1 Data description and comparison

A total of 286 species (10% of the total legume flora) in 104 genera (48% of the total legume flora) were reported to have ethnomedicinal use. The most cited therapeutic applications (classified according to WHO-IDC-10), were diseases of the digestive system (DDS) (14%), and diseases of the respiratory system (DRS) (18%). *Senna* (245 reports) and *Bauhinia* (157 reports) were the two genera most cited for ethnomedicinal use in Brazil.

The pharmacological investigation shows 143 species (50% of species with ethnomedicinal use, 5% of total flora) in 77 genera (72% of genera with ethnomedicinal use, 34% of total flora) had been the focus of pharmacological study (appendix 6). A much lower percentage of genera without any ethnomedicinal use had been studied (19%, 22 of the 118 genera). These studies were described in 3047 publications abstracted by the Web of Science. *Senna* and *Bauhinia* had the highest number of species with pharmacological studies (14 and 9) while 31 genera with ethnomedicinal use had not been the subject of any pharmacological study reported in the literature.

4.3.2 Phylogenetic analysis

The investigation of the phylogenetic structure of ethnomedicinal species for medicine shows an over-dispersed structure for overall ethnomedicinal use (NRI and NTI <0, p<0.05) (Figure 4-2). Eight of the ten therapeutic application categories for

60

which more than 10% of species were ethnomedicinal (>30 reports) showed an over-

dispersed structure with statistical significance (Table 4-1).
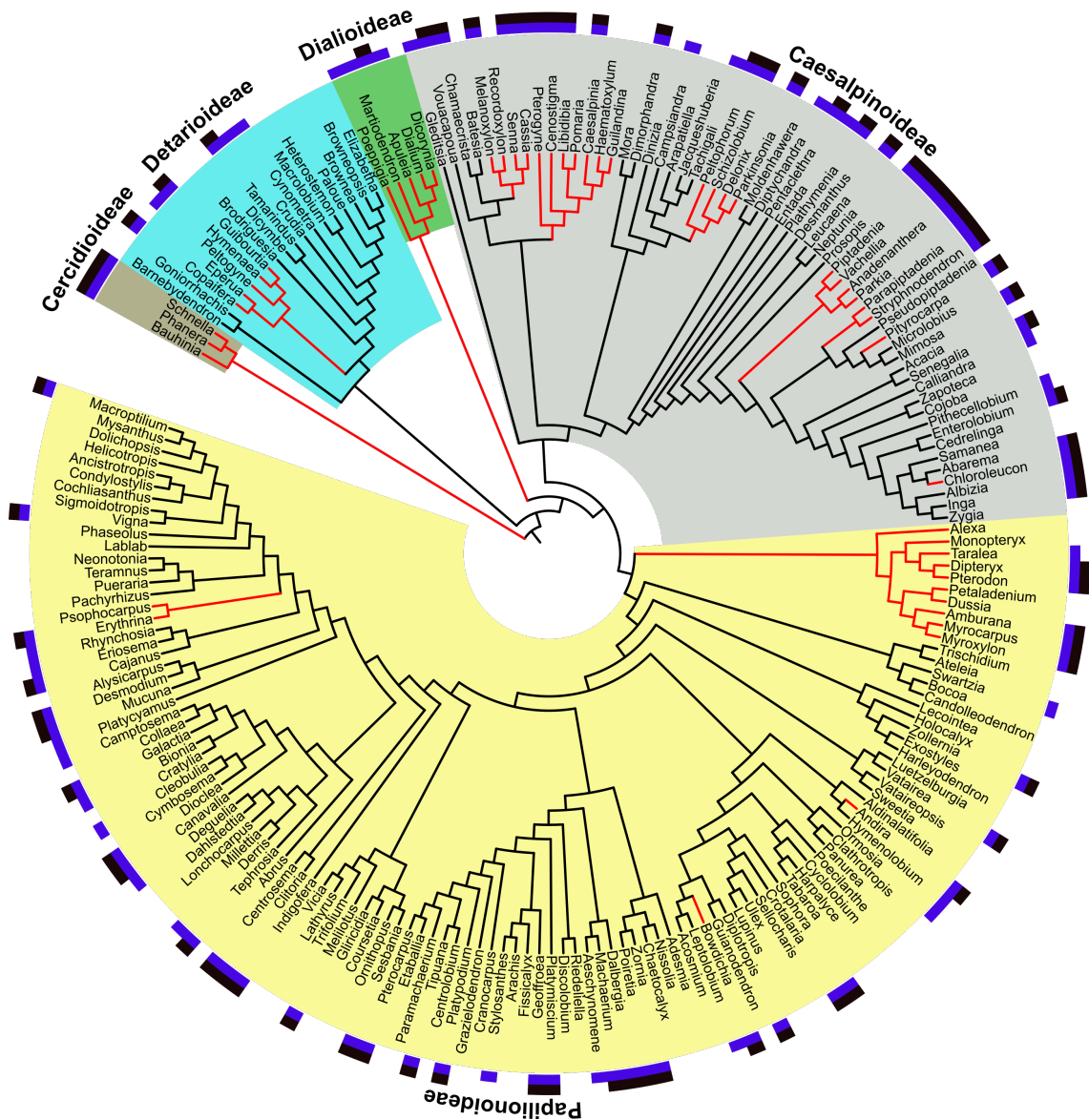
Figure 4-2. Distribution of medicinal uses and pharmacological studies in the Leguminosae phylogeny and "Hot nodes" for ethnomedicinal use. Block colours at the tips: purple = ethnomedicinal use for that genus; black = pharmacological studies for that genus. Branch colours: red branches = genera present in hot nodes for species with ethnomedicinal use. Clade colours: yellow, grey, green, light blue and brown indicate membership of subfamilies. Figure prepared using iTOL online (Letunic and Bork, 2016)

Table 4-1. MPD and MNTD values for therapeutic applications, plant parts used and modes of application of the ethnomedicinal medicinal uses of the Leguminosae in Brazil. DCS - Diseases of the circulatory system, DDS - Diseases of the digestive system; DFS Diseases of the femalegenito system; DGS - Diseases of the genitourinary system; DMC - Diseases of the musculoskeletal system and connective tissue; DRS - Diseases of the respiratory system; ENM - Endocrine, nutritional and metabolic diseases; IPD - Certain infectious and parasitic diseases; OTHER- not classified diseases such as fever, pain and inflammation.

| | ntaxa | mpd.obs | MPD | | mntd.obs | MNTD | |
			NRI	p-value		NTI	p-value
DCS	38	1.064	-1.886	<0.05	0.228	-1.301	ns
DDS	72	1.062	-2.065	<0.05	0.161	-1.279	ns
DFS	39	0.740	1.111	ns	0.115	0.322	ns
DGS	31	1.268	-3.342	<0.05	0.229	-0.634	ns
DMC	39	1.164	-2.756	<0.05	0.197	-0.603	ns
DRS	60	1.128	-2.412	<0.05	0.187	-1.314	ns
ENM	37	1.290	-3.235	<0.05	0.348	-2.317	<0.05
IPD	35	1.039	-1.712	ns	0.225	-0.979	ns
OTHER	105	1.057	-2.165	<0.05	0.142	-1.464	ns
Medicinal	276	1.172	-3.993	<0.05	0.147	-4.202	<0.05

4.3.3 Phylogenetic analysis: hot nodes

Nodesig analysis resulted in 45% of ethnomedicinal species and 16% of the total flora being found arising from hot nodes. The therapeutic application category with the highest percentage of ethnomedicinal species inside hot nodes was ENM, with 82% of ethnomedicinal species present, and the lowest was DGS where only 56% was found inside hot nodes (Table 4-2).

Table 4-2. Percentage of total and medicinal floras inside hot nodes, for categories of therapeutic applications and total medicinal use. DCS - Diseases of the circulatory system, DDS - Diseases of the digestive system; DFS - Diseases of the femalegenito system; DGS - Diseases of the genitourinary system; DMC - Diseases of the musculoskeletal system and connective tissue; DRS - Diseases of the respiratory system; ENM - Endocrine, nutritional and metabolic diseases; IPD - Certain infectious and parasitic diseases; OTHER- not classified diseases such as fever, pain and inflammation.

Hot nodes	total flora	Medicinal	% of total flora	% of medicinal flora	%gain
Medicinal	446	125	16%	45%	320%
DCS	320	26	12%	67%	1021%
DDS	661	48	24%	64%	1119%
DFS	305	23	11%	59%	1282%
DGS	336	18	12%	56%	1512%
DMC	542	33	20%	79%	1268%
DRS	397	41	14%	66%	777%
ENM	316	31	11%	82%	972%
IPD	457	23	16%	62%	1632%
OTHER	614	66	22%	61%	0%

4.3.4 Comparison with pharmacological studies

The number of species with ethnomedicinal use found inside the "hot nodes" was 126, representing 45.6% of all ethnomedicinal species. Of these 126 species with ethnomedicinal use, 74 were the focus of at least one pharmacological study. A lower number, 64 species, of the 150 ethnomedicinal species outside hot nodes had been screened. There is a preference for screening ethnomedicinal species that are in hot nodes (58% versus 42%) and when compared to total species, 16% of species inside hotnodes were subject to pharmacological studies, while this was true for only 3% of species outside hotnodes. Species with ethnomedicinal use present in the hot nodes were cited almost twice as often as the ones outside the hot nodes (p<0.05; table 4-3)

but no significant difference was found for the intensity of screening. Figure 4-3 shows

on a phylogeny how intensity of use and intensity of study are distributed.

Table 4-3. Characterization of "hot nodes" in terms of absolute numbers and proportions of species and studies. Proportions are reported according to the total number of species in hot nodes (446 species) and outside of hot nodes (2331), for species, and according to the total number of genera in hotnodes (50) and outside hotnodes (152). Citations means the number of reports a species have in the database for medicinal use.

	Hot nodes	Outside
Number of genera with ethnomedicinal study	37 (74%)	65 (42%)
Number of genera with pharmacological study	34 (68%)	60 (39%)
Number of species with ethnomedicinal use	126 (28%)	150 (6%)
Number of species with pharmacological study	74 (16%)	64 (3%)
Intensity of ethnomedicinal use (Number of citations)	1192	549
Intensity of pharmacological study (Number of studies)	1028	1665
Average number of studies per species	2.3	0.7
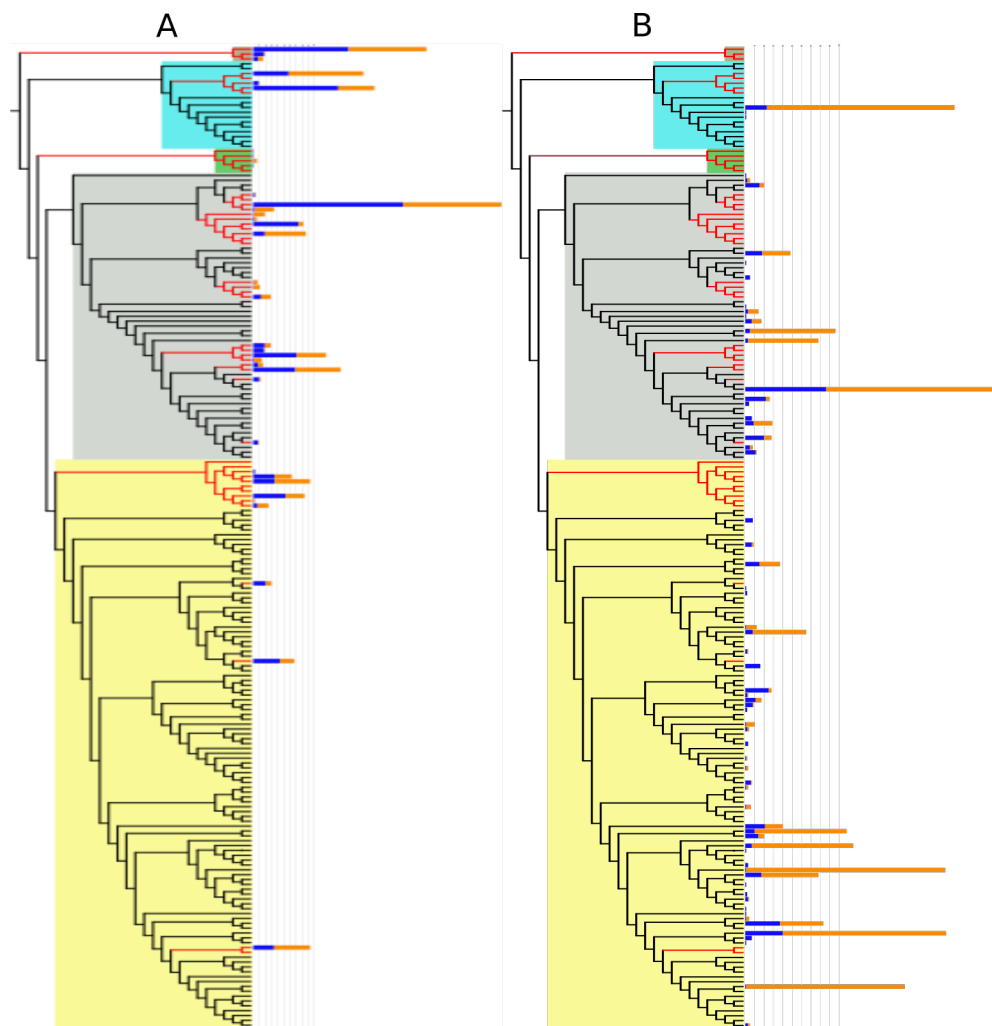
Figure 4-3. Comparison between hot nodes (A) and non-hot nodes (B) in relation to number of citations (blue) and number of pharmacological studies (orange). Hot nodes are calculated based on the distribution of recorded use of species. Species in hot nodes also have significantly more use reports per species (p<0.05). Figure prepared using iTOL online (Letunic and Bork, 2016)

4.4   Discussion

This study reveals that half of the species of Leguminosae used in ethnomedicine in Brazil have been the focus of at least one pharmacological study. This is a huge increase compared with Fabricant and Farnsworth's (2001) global estimate of 9% of plants screened. Fabricant and Farnsworth's study has become a citation classic, serving to highlight the paucity of knowledge of pharmacological properties of ethnomedicinal plants. Our study shows that at least for some plant lineages in some parts of the world, there is now considerably more knowledge of the pharmacological properties of plants used in ethnomedicine. Investment in natural products research informed by traditional knowledge has fluctuated (Koehn and Carter, 2005), and while there may be many factors at play, one factor sometimes cited is the challenge of negotiating benefit sharing with knowledge holders (Harvey et al. 2015). However, Brazil, like other BRIC countries, has invested in realizing the value of local and indigenous knowledge of indigenous plant resources through pharmacological research (Dutra et al. 2016; Saslis-Lagoudakis et al. 2012b). Given the apparent intensive efforts in ethnopharmacology focused on species found in Brazil, it is timely to review how this investment is directed. The consideration of the relationship between pharmacological study and intensity of ethnomedicinal use we present here is a first step towards a data-driven evaluation of how ethnomedicine might inform bioprospecting, or direct research into the health risks associated with local plant-based healthcare.

This study goes far beyond counting the number of species used in ethnomedicine and recording whether these plants are the focus of pharmacological studies. We use a phylogenetic framework to identify evolutionary lineages that are particularly important in ethnomedicine, and record whether species in those lineages are more often screened than other ethnomedicinal plants. A first step was the phylogenetic characterisation of ethnomedicinal species of Leguminosae in Brazil, and we revealed novel patterns with significant implications for bioprospecting. Previous phylogenetic studies reported clustering of species with ethnomedicinal use, both for whole floras (Saslis-Lagoudakis et al. 2012a), and when species-level phylogenies are used to interrogate the patterns of plant use in genera (Ernst et al. 2016; Saslis-Lagoudakis et al. 2011a). Our study is the first focused at the level of a plant family, and also the first to recover a signal of phylogenetic overdispersal of an ethnomedicinal flora belonging to the Angiosperms. (Table 4-1). We interpret the signal of overdispersion as indicative of multiple, independent lineages of ethnopharmacological interest and thus a diversity of potential leads. Phylogenetically-informed bioprospecting until now has been concerned with clustering, whether recovered using D-statistics or Pagel's lambda (Cámara-Leret et al. 2017; Rønsted et al. 2008), or community phylogenetic methods (Saslis-Lagoudakis et al. 2011a, 2012a, 2014, Yessoufou et al. 2015, Ernst et al. 2016, Halse-Gramkow et al. 2017). Overdispersal challenges the established use of phylogeny as a predictive tool to inform ethnopharmacological screening, but here we show that hot nodes can still be identified and lineages of interest highlighted (Figure 4-2). In the context of what might be overdispersed clusters (Figure 4-1) a fresh view of priorities for pharmacological research effort can be made.

Bioactivity of plant species is related to the presence of secondary metabolites. These mediate interactions of plant species with pathogens and herbivores, other plants and pollinators (Bourgaud et al. 2001, Edris, 2007; Heil, 2008; Huang et al. 2012). While some metabolites in the legumes, such as flavonoids, triterpenes and pinitol, are of broad occurrence and can be found in diverse groups, others are limited in their phylogenetic distribution (Wink, 2013). The pattern of overdispersal we find here might result from the adoption of multiple lineages with different secondary metabolites for ethnomedicine. Following this interpretation, each overdispersed cluster would represent a different target for bioprospecting. Although ethnomedicinal therapeutic applications may not relate to biomedical conditions (Staub et al. 2015) there may be phylogenetic patterns in application that relate to properties of the plants (Ernst et al. 2016). The overdispersed phylogenetic structure of therapeutic applications (Table 4-1) may highlight plants with different chemistries to treat each condition. With a huge floristic diversity and many different floristic environments, and also abiotic differences such as altitude, humidity, soil, a diversity of secondary metabolites in the Leguminosae in Brazil across species localities and occurrences is expected. The interplay of chemical and phylogenetic diversity also underscores that the hierarchical level of the study is important in devising and interpreting any study. Saslis-Lagoudakis et al. (2012a) indicated in a study across all Angiosperms that plants of the Leguminosae were included in cross-cultural hot nodes. However, in this study at the family level, overdispersal reveals phylogenetically non-random use, comprising overdispersed lineages that we hypothesise here are selected for their diverse

chemistries. Many more studies are needed to determine how general these emerging patterns are, and whether they apply to less species-rich families.

Our data show how pharmacological studies are distributed relative to hot nodes (Table 4-3). Although hot nodes comprise only 16% of total species they include almost half of the ethnomedicinal species. As well as being evolutionary hotspots for ethnomedicinal species, overall the hot nodes encompass significantly more use reports. These same hot nodes also include more than half of the species that have been the focus of pharmacological investigation. Therefore, we highlight here how just a small part of the phylogeny concentrates both traditional knowledge and pharmacological research, so that lineages outside of hot nodes are relatively poorly known. Our study corroborates the view that ethnomedicinal use guides pharmacological research (Gurib-Fakim, 2006), using a phylogenetic framework for Brazilian Leguminosae. Whether emphasis on these lineages benefits bioprospecting or protection of the health of communities using the plants is unclear. We consider this question from three perspectives, firstly the quality of the research, secondly the phylogenetic distribution of species of known efficacy, and thirdly the likely toxicity of the lineages.

Firstly, to contribute to bioprospecting or safety assessments, pharmacological and toxicological survey must be robust. Although many pharmacological studies were found for the ethnomedicinal species, the quality and significance of the studies were not easily assessed. Many low impact studies use meaningless bioassays, and report differences between in vitro and in vivo experiments that are a consequence of data

69

quality (Gertsch 2009). Secondly, there must be the expectation that the most important species for bioprospecting belong to hot nodes. At the level of whole floras, it is known that evolutionary lineages of plants that are rich in ethnomedicinal species also contain a significantly greater proportion of species with demonstrated efficacy (Saslis-Lagoudakis et al. 2012a). This suggests that the focus on hot nodes is well placed. However, despite the exponential growth of ethnopharmacological studies in the past years, few significant discoveries have been made (Gertsch, 2009). It is noteworthy that one of the two natural products derived from the Leguminosae and brought to market, monocrotaline originated from *Crotalaria sessilifolia*, is an anti-tumor agent with reported toxicity (Honorio Junior et al. 2010; Nogueira-Ferreira et al. 2015). *Crotalaria* species are used in Brazilian ethnomedicine, but with very specific applications, including as abortifacients, and are not part of a hotnode. Similarly, *Canavalia* includes ethnomedicinal species with few uses and is known to be toxic (Dang and Van Damme, 2015; Follmer et al. 2001; Oliveira et al. 1999), and is outwith the hot nodes. We suggest that toxic species are phylogenetic outliers in terms of ethnomedicinal use, and that the heavily used lineages and species are relatively benign. Plants having a mild effect are expected to be dominant in ethnopharmacopoeias, but novelty may be found in plants less closely related to these species which at least in Brazilian legumes are already well-known.

Whether or not novel leads remain to be discovered, there are compelling reasons to characterise plants that represent different lineages in ethnomedicine. Lack of knowledge may be problematic because health risks associated with ethnomedicinal plant use may be unreported (Jäger 2015). Even studies without significance for the

global pharmaceutical industry can still act as a start point to assure safety of use. In the case of Brazilian Leguminosae, two commonly used plants *Stryphnodendron* (Costa et al. 2013) and *Erythrina* (Dias et al. 2013), have been the focus of recent pharmacological and toxicological survey without substantiating concerns. However, *Crotalaria* species with known toxicity have been reported to be used, and approximately half of the ethnomedicinal species appear to be unscreened. Phylogenetic methods could to be used in tandem with ethnobotanical databases to predict the health risks associated with uncharacterized plant species. Here we show lineages rich in ethnomedicinal use are also rich in pharmacological studies, showing effort is directed towards characterizing closely-related, most-used species. Just as for bioprospecting purposes, it could be argued that phylogenetically-isolated species should be consciously included in future research.

This study shines a light on the phylogenetic distribution of ethnomedicinal plants in Leguminosae from Brazil. In highlighting an overdispersed structure for the first time it draws attention to the complex nature of plant selection for ethnomedicinal use. It also highlights a connection between lineages of importance for ethnomedicine and those that have become the focus of ethnopharmacological research, raising questions about the drivers of selection of species for screening in pharmacological studies. Much has been written about means devising rational, data-driven selection of plants for screening, for example by inferring useful phytochemical composition from ethnobotanical, chemosystematic or ecological information (Albuquerque et al. 2012; Albuquerque, 2010; Coley et al. 2003; Douwes et al. 2008; Saslis-Lagoudakis et al. 2011b; Tan et al. 2006), or from phylogeny (Ernst et al. 2016;

Grace et al. 2015; Rønsted et al. 2008; Saslis-Lagoudakis et al. 2012a; Yessoufou et al. 2015). Other factors are important, for example the availability of plentiful plant material, may be a strong driver (Atanasov et al. 2015). Widely distributed species might be more likely to be screened. Existing protocols and emerging interest may create a snowball effect, so that many studies are made of species of already proven interest, or of their close relatives. The species that are the most intensively screened often have food or fodder use (e.g. *Canavalia* species used as food for cattle; *Desmodium* species). Ultimately, our study demonstrates a method to identify the plant lineages that are most important ethnomedicinally. As research into natural products grows, this novel framework provides a timely insight into how knowledge of ethnomedicinal plants is distributed.

Chapter 5 Spatial structure in the distribution of traditional knowledge

5.1   Introduction

Cultural diversity is characterized by the relationships between biological diversity and traditional knowledge (Guglielmino et al. 1995; Pagel and Mace, 2004). Traditional knowledge of medicinal plant biodiversity is especially significant, because more than 50 000 species are estimated to have been used as medicine (Schippmann et al. 2006, 2002), and plant medicine continues to contribute to the health of millions of people (Farnsworth et al. 1985). The relationship between humans and plants, and consequently selection of plants for medicine, can be seen as a complex set of interactions, depending on plant species availability and demand for treatment or prophylaxis, leading to a so called "herbal landscape" (Sõukand and Kalle, 2010). How extensive such medicinal plant landscapes are can be difficult to assess. If plant use is the result of local experimentation, rather than sharing of traditional knowledge, we would expect traditional knowledge to be very strongly spatially patterned. However, transmission of knowledge of medicinal plants is well documented (Begossi et al. 2002; Ceolin et al. 2011; Leonti, 2011; Lozada et al. 2006), and migration is well-known to result in translocation of knowledge (Campos and Ehringhaus, 2003; de Medeiros et al. 2011; Vandebroek and Pieroni, 2007). Findings such as these challenge the view of medicinal plant use as strongly adapted to the local environment through

experimentation in situ over many generations (Schultes, 1960). If plant species used for medicine are readily adopted, flowing between peoples as a result of horizontal transfer of traditional knowledge, and in space as a result of peoples' migrations, then spatial patterns in traditional knowledge of medicinal plants might be limited.

Understanding the medicinal plant landscape is complicated by the distribution of plant species, and because of their relatedness. Plants comprise the most visible components of biological communities, and turn-over in species between biomes characterises many of these habitats. Medicinal plant use is shaped by the floristic environment (Saslis-Lagoudakis et al. 2014). In terms of relatedness, it is known that medicinal species are not randomly distributed in a flora, but can usually be found in certain lineages/families (Lukhoba et al. 2006; Moerman et al. 1999; Saslis-Lagoudakis et al. 2012a). This shared preference is usually attributed to independent discovery whilst shared use of closely related plants may point towards transmission of traditional knowledge (Hawkins and Teixidor-Toneu, 2017; Saslis-Lagoudakis et al. 2015; Saslis-Lagoudakis et al. 2012a).

Brazil has one of the world's richest floras, estimated to comprise more than 40 000 species and a high level of endemism (around 46.2%). This high biodiversity is partitioned into floristic and geographical environments called biomes. The six biomes comprise between 19 355 and 1 240 species, between 4 and 40% endemism, and two are important floristic hotspots, the Cerrado and the Atlantic Forest (Forzza et al. 2012). Brazil also has a rich cultural diversity, expressed as religious and ethnic diversity (Araújo et al. 2007; Deus, 2011; Fleury, 1999; Little, 2002), but linguistically it

is quite homogenous. Within Brazil, migration from the Atlantic coast inland was part of the colonization process, and since then significant migration has been from the rural Northeast to urban Southeast (Gonçalves, 2001).Migrations such as the settlements of descendants of African slaved people (Quilombos) in the Amazon have resulted in assimilation of the medicinal plants of the native Amazonian Indians (Voeks and Rashford, 2013). The widely-adopted dominant language and mobility presents opportunity for knowledge exchange.

In this study the relationship between human culture and resource use, shaped by the floristic environment, is explored in the context of evolutionary ethnobotany (Albuquerque and Ferreira Júnior, 2017; Saslis-Lagoudakis et al. 2015). The main objective is to test the null hypothesis that medicinal plant use in Brazil is homogeneous, aside from constraints of availability in local floras. The approach taken exploits 1. existing data describing the spatial distribution of plants 2. a modified hypothesis of plant phylogenetic relationships from published DNA sequence data and 3. a newly compiled traditional knowledge database, derived from herbarium specimen records. This attempt to critically evaluate the extent to which there is heterogeneity in medicinal plant use at a landscape scale depends on the application of methods from biogeography to traditional knowledge. These utility of these methods, including phylogenetic assessments of beta diversity, and Principal Components of Phylogenetic Structure (PCPS Duarte, 2011), is determined here, asking whether they could be used more widely to document turn-over in traditional knowledge of biodiversity.

5.2 Methodology

5.2.1 Data collection and standardization

Herbarium vouchers containing information on medicinal use were collected from online herbarium data available from CRIA Species Link (http://splink.cria.org.br/). A list of all angiosperms was downloaded from the website as a CSV file and species with medicinal use were identified by filtering and searching the field "description" with the key words "medicin*" and "uses". Species names were standardized with Plantminer (Carvalho et al. 2010) to names in THE PLANT LIST (The Plant List 2013) and Missouri Botanical Garden's Tropicos database (Tropicos.org 2016). Records were then filtered and only records with identification to species level and geographical coordinates were selected for further analysis. The floristic composition of each biome was extracted from the Brazilian flora (Brazilian Flora, 2020) and these species names also were standardized with Plantminer as described above. The information gathered was databased so that for every species/therapeutic application combination found there was a single entry in the database, which we call a species use report. The database had the following fields: genus and species as recorded in the publication, currently accepted name, therapeutic application as recorded in the publication, therapeutic application categorized according to the World Health Organization (WHO) International Classification of Diseases 10 (http://www.who.int/classifications/icd/en/), parts used and mode of application.

Species were plotted on a map using QGIS (Quantum GIS development team, 2017) with boundaries of Brazil and each of the six Brazilian biomes, and a matrix was extracted with each record being allocated to a biome.

5.2.2 Data analysis

A suite of analyses, from taxonomic and phylogenetic methods for investigating species composition were adopted here. The range of approaches and how possible outcomes could be interpreted in a phylogenetic context is presented in appendix 7.

5.2.2.1 Taxonomic Beta diversity

Taxonomic beta diversity and phylogenetic beta diversity were calculated to compare medicinal plant use between biomes. For the taxonomic beta diversity, the chosen dissimilarity index was beta-similarity due to its better fitness against a dataset of presence and absence, where:

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

Other dissimilarity metrics are also used for similar analysis, but can be strongly influenced by differences in species richness (see Kreft and Jetz, 2010). The dissimilarity between biomes was investigated in terms of medicinal plant use, and according to species composition at both species and generic level. Here, four different matrices were made, two for medicinal plant use containing only medicinal species/genera and two containing all species/genera in the Brazilian flora. To investigate the correlation between total flora and medicinal flora, a Mantel test based on "Pearson" distance was conducted between the dissimilarity matrices (medicinal

species vs total species and medicinal genera vs total genera) using R (R Core Team, 2015). A cluster analysis for the four different matrices was done also using R (R Core Team, 2015).

5.2.2.2 Phylogenetic structure and beta diversity

A phylogenetic tree at generic level was reconstructed for the Brazilian flora using the S.Phylomaker tool, (Qian and Jin, 2016), with one species per genus. S.Phylomaker uses PhytoPhylo megaphylogeny (an updated version of Zanne et al. (2014)) as a backbone, and matches a provided species list by the user to the species present in the phylogeny. When present, the species is selected for pruning, and when not available in the phylogeny, the species is added to the backbone within the genus of family. To get the final tree, all selected and added species are pruned from the megaphylogeny, and the phylogenetic tree is reconstructed according to three different scenarios. The selected phylogenetic scenario from S.Phylomaker was the third one, where genera not present in the megaphylogeny were added to their families using the same approach implemented in Phylomatic (Webb and Donoghue, 2005) and BLADJ (Webb et al. 2008) (Figure 5.1).

A   B
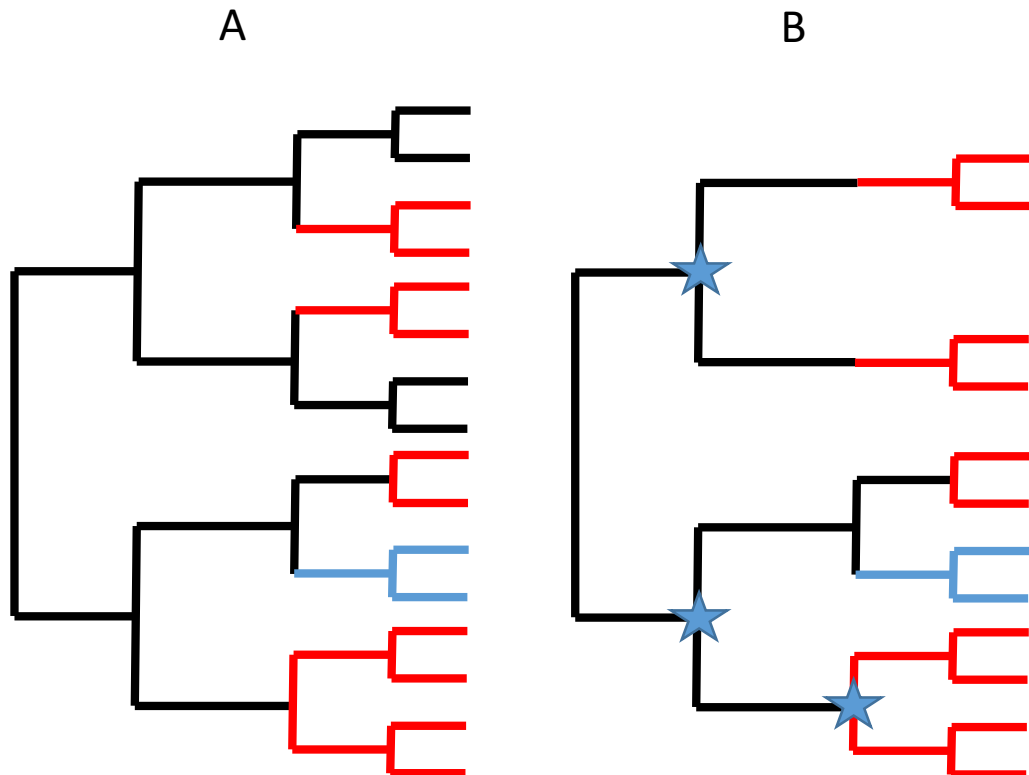
Figure 5-1. Graphic explanation of S.Phylomaker steps. A shows the first steps, where taxa provided by the user is matched to the taxa present in the megaphylogeny. Taxa present in both datasets are selected and marked for pruning (red clades). Taxa not present in the megaphylogeny but provided by the user is added to the backbone within the genus or the family (blue clade). Finally, the selected taxa (red and blue) are pruned, and the phylogeny is reconstructed, in this case following scenario 3, where the branch lengths are calculated according to information available from some nodes that were previously present in the phylogeny (stars) (following BLADJ-(Webb et al. 2008)).

The phylogenetic structure of traditional uses was investigated using the Net Relatedness Index (NRI) and Nearest Taxon Index (NTI) (Webb, 2000; Webb et al. 2002), with the functions ses.mpd and ses.mntd from the PICANTE package (Kembel et al. 2010), by multiplying the MPD and MNTD values by -1. The chosen null model was "taxa.labels", which shuffles the distance matrix labels across all taxa included in distance matrix with 999 runs. Phylogenetic structure of medicinal plant use was

investigated for each biome in relation to the total Brazilian flora. Possible outcomes and the relationships of NRI and NTI values were shown in chapter 4, figure 4.1.

Three different measures of phylogenetic beta diversity were used here. The first incorporates the βNRI and βNTI obtained from the functions "comdist" and "comdistnt" in Phylocom (Webb et al. 2008). "Comdist" measures the MPD for each taxon in a sample and measures the mean value to the average distance to all taxa in the other sample. "Comdistnt" does the same but for the MNTD values. Adding the abundance data (here abundance was the number of times a taxon was cited as medicinal in a biome), the average distance changes to be among two random individuals drawn from each of two samples (Webb et al. 2008).

The second phylogenetic index of beta diversity was calculated using the function "phyloSor" (Bryant et al. 2008) from the package Picante (Kembel et al. 2010) in R. This index ranges from 0 to 1, where 0 indicates a that two communities (here biomes) share no taxa, and 1 indicates that both biomes have the same taxa. This second index allows a comparison of the phylogenetic beta diversity and the taxonomic beta diversity to be made using a phylogenetic dissimilarity index, calculated as 1- phyloSor index (Souza-Neto et al. 2016). The taxonomic beta diversity was measured here using the function "betadiver" in R, and the chosen index was "Sorensen" due to its relation to phyloSor (Bryant et al. 2008; Feng et al. 2012; Swenson et al. 2011). This comparison is important because while the taxonomic beta diversity might indicate a high dissimilarity between biomes, the phylogenetic beta diversity dissimilarity might indicate some degree of relatedness between medicinal

plants used in both biomes (Figure 5-2). To test whether a species/generic turnover is dominant between two biomes, in other words whether beta-diversity is mostly due to different species/genera being used/present in different communities, the partitioning of betadiversity into nestedness and turnover components were calculated the function "beta.pair" from the package "Betapart", following Leprieur et al. (2012).

The third measure of phylogenetic beta diversity is based on the PCPS method (Pillar and Duarte, 2010). This method is based on decomposing the phylogenetic information for each community (here biome) using phylogenetic fuzzy weighting (Pillar and Duarte, 2010) into several orthogonal eigenvectors that can be then analysed as a PCA. Each eigenvector is a phylogenetic gradient for the communities studied and shows the variation across the whole phylogeny. Eigenvectors with the highest values describe deep relationships in the phylogenetic tree, while eigenvectors with low values describe shallower relationships in the phylogeny as changes in families or genera (Duarte et al. 2012). This method allows lineages influencing phylogenetic beta diversity differences when medicinal use is associated with specific biomes to be identified (Duarte et al. 2012).

5.2.2.3 Therapeutic applications

Brazilian biomes were analysed pair-wise with a Spearman rank-order correlation to test whether the proportion of therapeutic applications is homogeneous throughout Brazil.
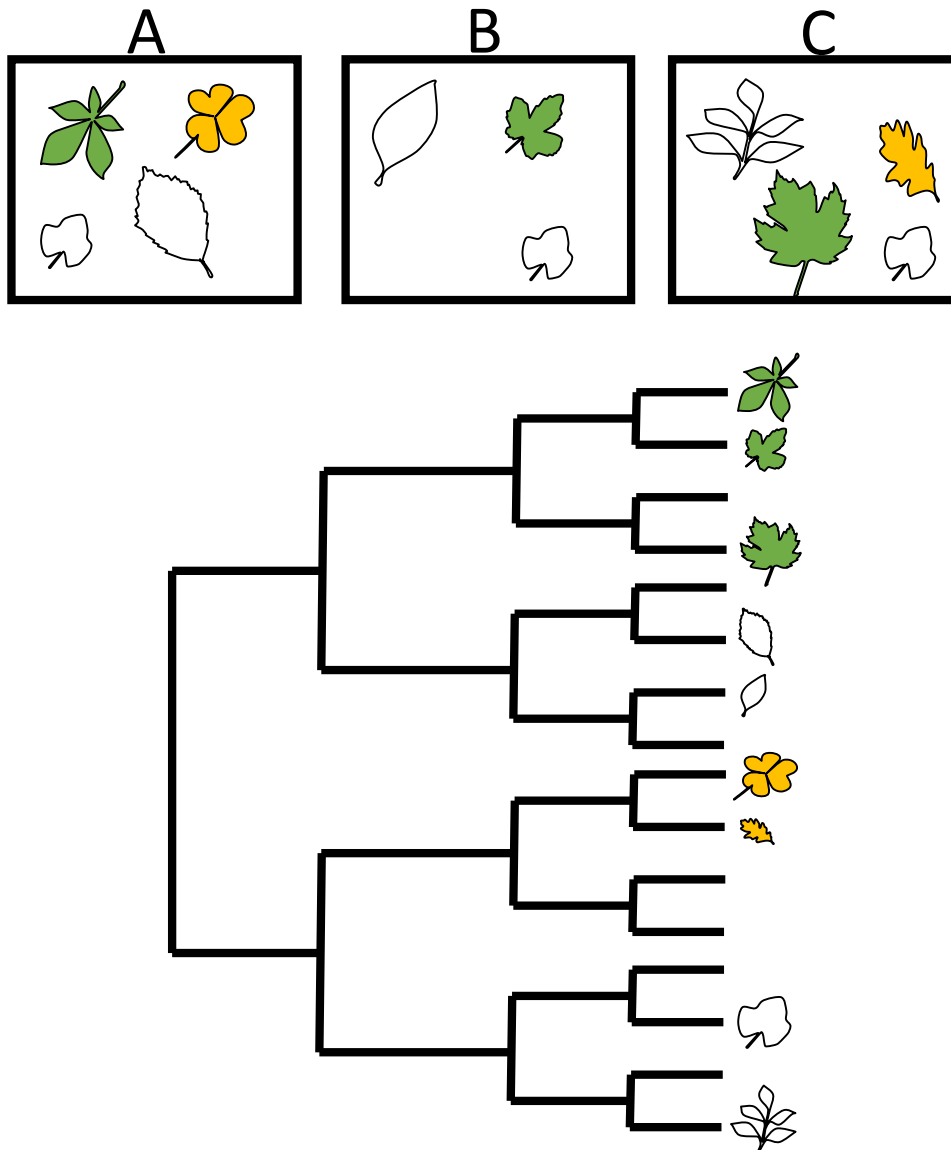
Figure 5-2. Comparison between taxonomic and phylogenetic beta diversity. A, B and C are different communities with different plant species, but sharing only one species. A taxonomic beta diversity analysis with such communities would result in a high value of dissimilarity, and distance between communities would be the same. On the other hand, a phylogenetic beta diversity analysis would result in a lower value of dissimilarity, since some of the species present are closely-related species, for example species coloured in green form a group of closely related species, and the same for species coloured in yellow. Considering the phylogenetic similarity, communities A and C would be more similar, than either are to B.

5.3 Results

5.3.1 Data collection

Of 5.38 million records online, 7 122 records referring to medicinal use were selected. These records included a total of 1 994 species in 880 genera and 209 families. Of these, 1 579 species in 733 genera and 199 families had geographical coordinates that allowed to match their occurrence to a specific biome. The biomes most represented here according to the number of records were the Atlantic Rainforest and the Cerrado, while the Pampas had the lowest number of records. Figure 5.3 shows the distribution of records and compare the number of species with medicinal use to the total number of species. Pantanal was the biome with the higher percentage of medicinal species when compared to the total flora (12%), followed by the Caatinga (8%), Cerrado (5%), Atlantic Forest (4%), Pampas (3%) and Amazon (2%).
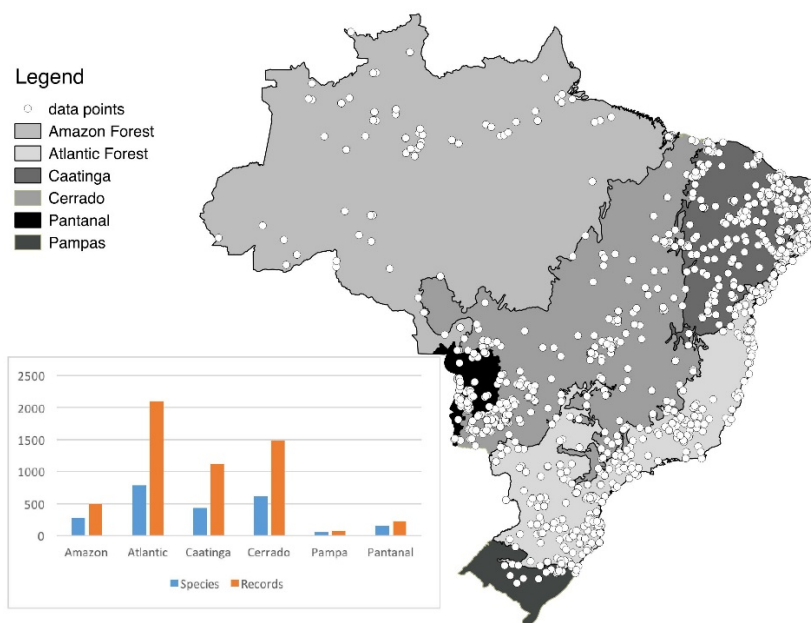
Figure 5-3. Distribution and number of species with ethnomedicinal use and records for each biome in Brazil.

5.3.2.1 Taxonomic beta diversity (or the relationship between medicinal flora and overall flora)

Taxonomic beta diversity analysis shows that changes in the species composition of medicinal floras between different biomes is overall more different than expected when compared to floristic changes. Analyses at generic level (appendix 8) show the same pattern, with the composition of medicinal floras at generic level more different than the composition of floras. When endemic species for each biome are excluded this result is still present, but the relation between medicinal flora and total flora changes between some pairs of biomes (Table 5-1).

Table 5-1 Dissimilarity matrices at species level for medicinal plant use (lower diagonal) and total flora (upper diagonal) in the Brazilian Biomes. Green colour indicates a smaller difference in the medicinal flora than in the total flora. Red colour indicates a larger difference.

A	Amazon	Atlantic	Caatinga	Cerrado	Pampa	Pantanal
Amazon		0.77	0.64	0.73	0.82	0.46
Atlantic	0.59		0.48	0.59	0.31	0.38
Caatinga	0.72	0.61		0.33	0.80	0.56
Cerrado	0.65	0.60	0.63		0.62	0.28
Pampa	0.84	0.46	0.81	0.58		0.74
Pantanal	0.82	0.65	0.74	0.42	0.89	

The Mantel test revealed a significant similarity between medicinal plant use and total flora at species level, for the Brazilian biomes (r=0.7101, p<0.05) and also at a generic level (r=0.7101, p<0.05) (Figure 5.4).
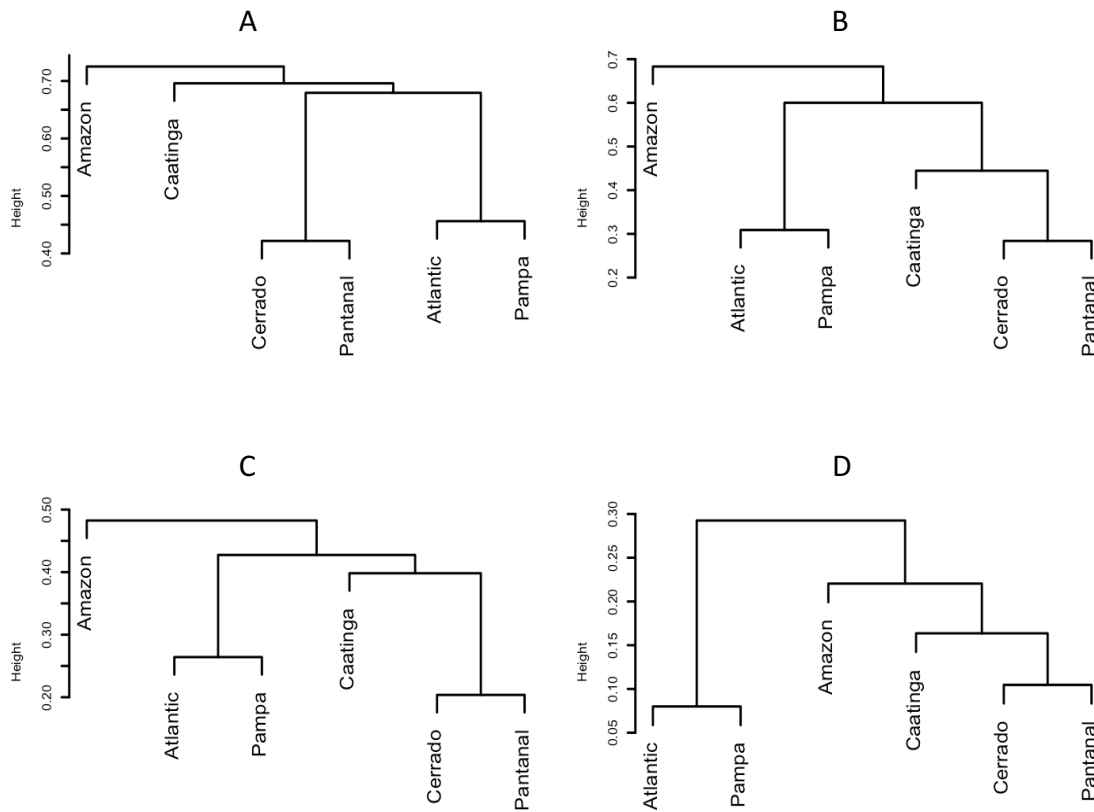
Figure 5-4. Clusters based on beta similarity for medicinal plant use and total flora in the Brazilian Biomes. A- medicinal plant use at species level. B- total flora at species level. C- medicinal plant use at generic level. D- total flora at generic level.

5.3.2.2 Phylogenetic structure and beta diversity

Overall, medicinal plants are phylogenetically clustered in Brazil, meaning that medicinal genera are more related to each other than expected by chance at both deep and shallow phylogenetic nodes (NRI and NTI). Considering the whole Brazilian flora, the five biomes Amazon Forest, Atlantic Forest, Caatinga, Cerrado and Pampas also showed significant phylogenetic clustering, but the Pantanal did not (Table 5-2).

Table 5-2. Phylogenetic structure for the Brazilian flora overall, according to medicinal plant use, and for each Brazilian biome.

| | ntaxa | mpd.obs | MPD | | mntd.obs | MNTD | |
			NRI	p-value		NTI	p-value
Brazil	710	261.7496	13.0461	0.001	58.7713	7.2529	0.001
Amazon	200	258.8319	8.2827	0.001	79.2895	3.5733	0.001
Atlantic	430	261.13	10.554	0.001	63.0911	6.3334	0.001
Caatinga	247	243.9496	11.9223	0.001	73.3218	4.5516	0.001
Cerrado	374	257.1855	10.5437	0.001	65.4044	5.77	0.001
Pampas	53	235.1347	5.619	0.001	94.9478	3.21	0.001
Pantanal	107	279.9804	1.9211	0.001	108.258	0.4272	0.217

Significant phylogenetic relatedness was observed between all pairs of biomes for medicinal plant use according to the βNRI index (both with and without abundance data). According to the βNTI index, with and without abundance, the Amazon Forest-Cerrado, Amazon Forest-Pampas, Atlantic Forest-Pampas and Caatinga-Pampas pairs were significantly related. One additional pair Cerrado-Pampas, was significantly clustered when only when abundance data is considered. (Table 5-3). Overall, there is an indication of medicinal families and genera used in Pantanal to be more distant to the ones used in other biomes (table 5-3A). Notably, the values for βNTI suggest some genera used in Pantanal find no close relatives in other biomes, but considering abundance data, there may be few records of such genera being used.

Table 5-3. Phylogenetic beta diversity indices from the "comdist" and "comdistnt" analysis in Phylocom. Upper diagonal= βNTI index. Lower diagonal= βNRI index. Values in bold have p>0.05. A- without abundance data. B- with abundance data

A	Amazon	Atlantic	Caatinga	Cerrado	Pampa	Pantanal
Amazon		2.98	2.62	**2.59**	**1.64**	-0.50
Atlantic	**10.71**		3.51	3.94	**1.49**	-2.20
Caatinga	**11.83**	**12.58**		3.65	**2.59**	-1.29
Cerrado	**10.61**	**11.58**	**12.56**		0.66	-2.04
Pampa	**8.29**	**8.12**	**8.99**	**7.98**		0.23
Pantanal	**5.41**	**5.36**	**6.32**	**5.33**	**5.01**	

B	Amazon	Atlantic	Caatinga	Cerrado	Pampa	Pantanal
Amazon		5.23	2.17	**4.23**	**4.35**	1.33
Atlantic	**8.32**		4.74	3.53	**7.60**	5.69
Caatinga	**8.06**	**9.14**		3.16	**5.03**	1.84
Cerrado	**7.73**	**9.10**	**8.92**		**7.14**	5.35
Pampa	**7.46**	**8.43**	**7.85**	**7.56**		0.72
Pantanal	**4.12**	**5.15**	**5.09**	**4.49**	**4.70**	

Taxonomic beta diversity is higher than phylogenetic beta diversity, indicating turnover between biomes within lineages (table 5-4).

Table 5-4. Beta diversity values for each pair of biomes, according to medicinal plant use at generic level. Upper diagonal- taxonomic beta diversity. Lower diagonal- Phylogenetic beta diversity.

	Amazon	Atlantic	Caatinga	Cerrado	Pampa	Pantanal
Amazon		0.60	0.59	0.56	0.83	0.67
Atlantic	0.44		0.54	0.44	0.84	0.72
Caatinga	0.42	0.40		0.49	0.84	0.65
Cerrado	0.40	0.32	0.34		0.82	0.64
Pampa	0.64	0.72	0.65	0.69		0.78
Pantanal	0.48	0.55	0.48	0.48	0.60	

Turnover component with a higher value at the taxonomic beta diversity suggests that close relatives are selected when species turn over between all pairs of biomes (Table 5-5).

Table 5-5. Turnover component of beta diversity indices for each pair of biomes, according to medicinal plant use at generic level. Upper diagonal- taxonomic beta diversity. Lower diagonal- Phylogenetic beta diversity.

	Amazon	Atlantic	Caatinga	Cerrado	Pampa	Pantanal
Amazon		0.37	0.54	0.37	0.58	0.53
Atlantic	0.24		0.37	0.40	0.26	0.31
Caatinga	0.38	0.24		0.36	0.55	0.42
Cerrado	0.23	0.29	0.21		0.26	0.20
Pampa	0.27	0.11	0.22	0.10		0.66
Pantanal	0.38	0.23	0.33	0.16	0.37	

The Principal Components of Phylogenetic Structure found with the first two being responsible for 72% of total variation (appendix 9). There is a higher similarity between Caatinga and Pampas and the Amazon-Atlantic-Cerrado, having the Pantanal as an outlier (Figure 5-5, appendix 10).
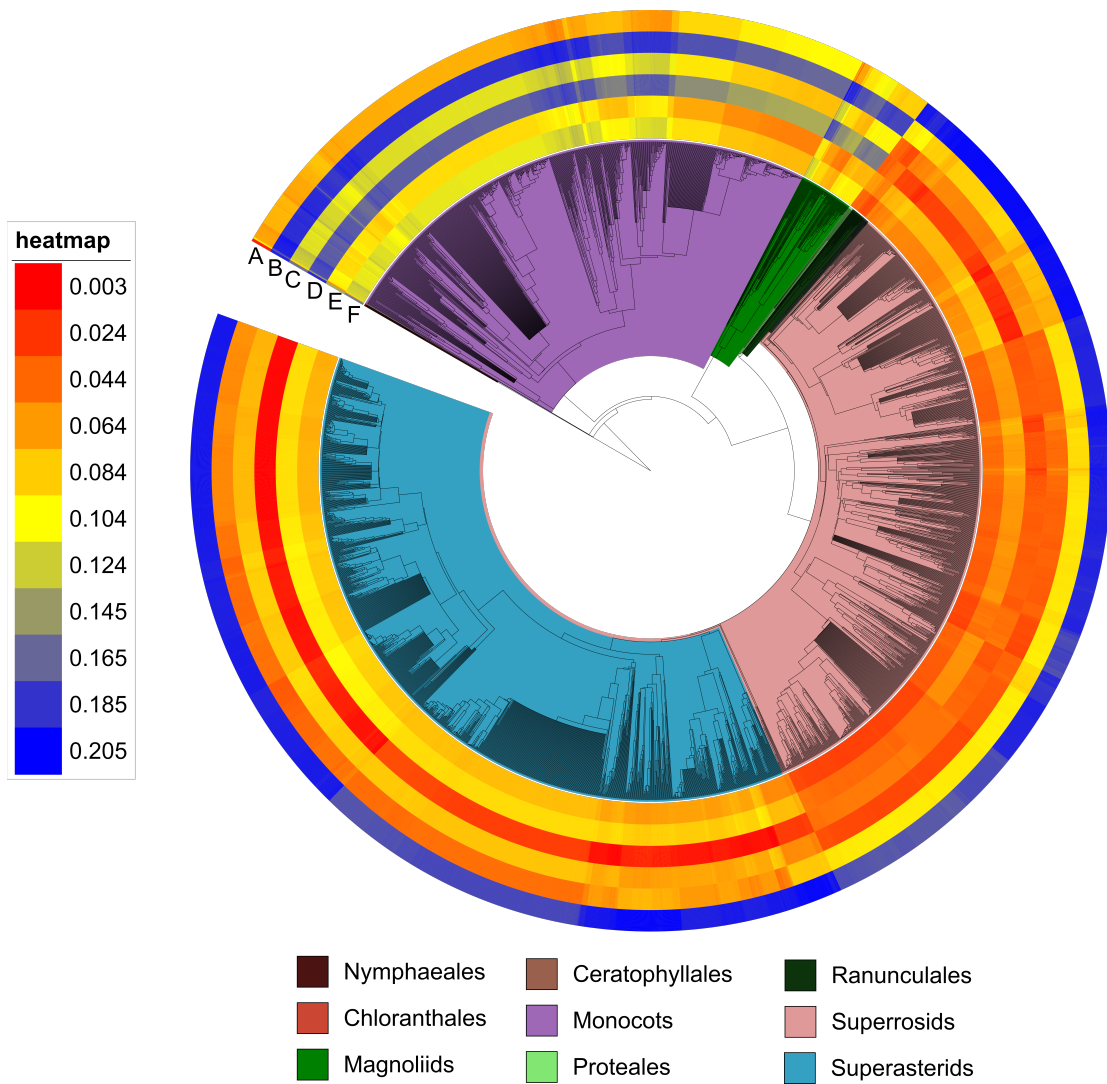
Figure 5-5. Heatmap showing the distances for each genera to a specific biome, obtained from the pcps analysis. A-Pantanal. B- Pampas. C- Cerrado. D- Caatinga. E-Atlantic Forest. F- Amazon Forest. Distances are shown as blue for smaller distances to red for higher distances. Figure prepared using iTOL online (Letunic and Bork, 2016).

5.3.2.3 Therapeutic applications

A total of 864 records with 514 species and 318 genera (15% of records) had information on therapeutic application. The most cited categories were Others, followed by DDS and DRS. The total number of records for each therapeutic application category for each biome is presented on table 5-6. A Spearman rank order correlation test showed there was no significant difference between almost all biomes in terms of the relative proportions of therapeutic applications, but the Amazon and the Pampas, and Caatinga and the Pampas (appendix 11).

Table 5-6. Distribution of therapeutic applications in each biome in Brazil, according to data collected from herbarium vouchers.

Therapeutic Application Categories	Amazon	Atlantic	Caatinga	Cerrado	Pampa	Pantanal
DBI	8	9	7	7	1	3
DCS	2	24	6	19	2	0
DDS	24	104	25	53	3	6
DEA	2	4	1	5	0	1
DFS	4	19	11	21	0	1
DGS	6	41	12	25	1	6
DMC	4	15	12	16	1	1
DNS	0	4	7	0	0	0
DRS	12	75	36	49	0	1
DSS	3	21	4	11	1	1
ENM	8	16	11	8	0	0
IPD	4	20	11	14	0	2
Neo	0	5	0	4	0	0
OTHERS	124	123	43	45	3	5
PCP	1	1	1	2	0	0

5.4 Discussion

5.4.1 Data mining from herbarium specimens: how many plant species used in medicine?

Meta-analyses have the potential to contribute to deeper understanding of patterns in plant use (de la Torre et al. 2012; de Medeiros et al. 2013; Saslis-Lagoudakis et al. 2012a; Weckerle et al. 2011). Existing meta-analyses have used data synthesized from the literature. Here, instead of using published sources, data were compiled from online herbarium vouchers, an approach validated through research carried out for this PhD study (Chapter 2; Souza and Hawkins, 2017) and provisioning a rich dataset, not only in terms of the high number of species with medicinal use, but also in terms of the spatial distribution of uses. The use of almost 2 000 species has been documented. Previously there have been medicinal plant checklists for regions (Agra et al. 2008, 2007; Guarim-Neto and Morais, 2003), but this is the first country-wide estimate of the number of medicinal plants in Brazil. Proportionally, 2 000 species represent only 6% of the Brazilian flora. Schippmann et al. (2006) estimated 17.1% of species used in medicine, based on figures for 15 countries, including Malaysia, the country with the lowest proportion of plant species used in medicine with 7.7% of species used. The State of the World's Plants report 2017 (Alkin et al. 2017) reported 8.3% of plant species used in medicine based on the number of names in the Medicinal Plants Names service database. Our data supposes Brazil has the lowest proportion of medicinal plants yet reported. Whether this might be due to poor data availability, or to a real deficit of medicinal plant use can be considered. All our data are drawn from digitized herbarium specimens, and this probably contributes to underestimation of

the number of medicinal plant species. A comparison of reports from digitized herbarium and literature sources showed 154 and 264 species respectively for Leguminosae used for medicine in Brazil (Souza and Hawkins, 2017; Chapter 2). If the same proportion (42%) of all plant species are not captured by herbarium survey, we might expect literature survey to identify as many as 3,400 species used in medicine. Though still less than 10% of the total flora, in this case our estimate for Brazil would fall within the range reported by Schippman et al. (2006).

Neither herbarium nor literature survey would be expected to record all uses, since there is a need for more fieldwork in ethnobotany (Cámara-Leret et al. 2014a; Souza and Hawkins, 2017). For the study of spatial heterogeneity in plant use, it may be more important that there are no biases, where under recording is one source of bias. Research using literature reports can infer spatial distribution from studies, but overall few study sites can be included. In chapter 3, 86% of total localities in our database of medicinal Leguminosae were from herbarium data, showing more complete distribution data from herbarium specimens.

5.4.2   An overview of the spatial distribution of uses: how good is the spatial data?

Data were assigned to the six Brazilian biomes, permitting spatial analysis at a regional scale. There is variation in the number of records between biomes, with the Atlantic Forest and the Cerrado with the highest number, and the Pampas and the Pantanal the lowest (Figure 5-3). This reflects the availability of digitized data in online sources, and could be related the activity of research groups in these areas. More

reports are from localities in the Caatinga, Cerrado and Atlantic Forest, where active ethnobotanical research programmes are associated with the largest universities in Brazil. Fewer are from the Amazon, which is well-known to be relatively under-documented (Cámara-Leret et al. 2014; Ritter et al. 2015; Souza and Hawkins, 2017; Chapter 2). However, it has also been suggested that people with relatively low acculturation may use fewer plant species for medicine (Vandebroek et al. 2004), and this might also contribute to real relatively low counts for medicinal plants in the Amazon. The relatively species-poor Pantanal and Pampas have a low number of records, though at 12% the Pantanal has the highest proportion of species with recorded medicinal use. Ultimately, the significant similarity between taxonomic and medicinal beta-diversity suggests that data are not spatially biased. However, the impact of removing endemic species points towards a bias in our data for recording "interesting" endemic species.

5.4.3 Taxonomic beta diversity: greater than expected

Beta diversity measures the similarity or dissimilarity in species composition amongst sites (Whittaker, 1972). We calculated taxonomic beta diversity, at species and genus level, recording beta diversity for both floristic compositions overall, and for the medicinal floras of biomes. Overall, floristic similarity predicted similarity of medicinal plants use ($p<0.05$), meaning the more similar the floras, the more similar the medicinal plants. However, although floristically similar biomes are more similar medicinally, the plants used are very different, with at least 42% and as many as 89% of species used differing between biomes. This difference is nearly always greater than

expected, given floristic differences between biomes. Cultivated species are included in this calculation (10% of medicinal plants). Nevertheless, significantly more (12/15; $p<0.05$ pairwise) comparisons showed greater difference. However, Brazil encompasses culturally diverse indigenous and rural communities (Araújo et al. 2007; Deus, 2011; Fleury, 1999; Little, 2002). That the medicinal plants are more heterogeneous than expected under our null model suggests interaction of culture with environment increases spatial heterogeneity in plant use. However, differing demand for treatment does not appear to be a driver of this differentiation, since therapeutic applications showed no difference between biomes according to a Spearman rank-order correlation test ($p<0.05$).

5.4.4 Phylogenetic structure: significant structure at deep and shallow levels

Phylogenetic analyses show that medicinal plant use in Brazil is significantly phylogenetically clustered at deep (MPD) and at tip (MNTD) levels, in other words medicinal plants are more closely-related to each other than expected by chance. This overall clustering was expected, since medicinal plant use is often reported as phylogenetically structured in other floras and taxa (Ernst et al. 2016; Forest et al. 2007; Grace et al. 2015; Lukhoba et al. 2006; Rønsted et al. 2008; Saslis-Lagoudakis et al. 2015; Saslis-Lagoudakis et al. 2012a). Only the Pantanal showed non-significant clustering of medicinal plant use, at tip (MNTD) level. Although one possible explanation could be because of the medicinal use in the Pantanal of a waterlily, a single species that is phylogenetically very distinctive, further tests that excluded this species still presented the same result. These tests indicate that, although the

medicinal use of a waterlily may contribute towards the overdispersion, it is not its sole cause, and that, the remaining medicinally used species in Pantanal are more distantly related than expected.

The tests for clustering are interpreted further in the light of the turn-over, or beta diversity, of species across biomes. Saslis-Lagoudakis et al. (2015), referred to "phylogenetic replacement", specifically in the context of migrant communities. They noted that many authors had observed that when there is a preference for a plant not available to a migrant community, a close relative is used in its place. Phylogenetic replacement, at the level of whole medicinal floras under comparison, can be identified using tests of phylogenetic beta diversity.

5.4.5 Phylogenetic beta diversity

Phylogenetic distances between species used in different biomes: whether or not abundance is considered, phylogenetic measures show medicinal plant use is more similar than expected for all deep node comparisons. Relatedness at deep nodes is attributed to shared preference for some of the lineages common across all biomes. Cross-cultural studies comparing floras from different floristic kingdoms using the same methodology have shown similar results (Saslis-Lagoudakis et al. 2012a). The regions compared here are in closer geographical proximity with greater opportunity for knowledge transfer. Nevertheless, only four of 15 comparisons are significant at shallow phylogenetic level though they all indicate clustering when abundance is considered. The measure used to assess shallow relationship is very stringent, only considering closely related species, and this study is at generic level, that could reduce

the signal of phylogenetic replacement. Yet still there is evidence of this phenomenon. When abundance is not considered, the medicinal plants of the Pantanal are more different than expected at tip nodes. We attribute this to the use of abundant phylogenetically isolated species adapted to this wetland biome, such as waterlilies used medicinally only in the Pantanal.

Phylogenetic beta diversity compared to taxonomic beta diversity: taxonomic beta diversity indices measure how species composition changes in space, and phylogenetic beta diversity shows how phylogenetic relatedness changes in space (Graham and Fine, 2008). The comparison of taxonomic and phylogenetic beta diversity, especially the turnover components of such indices, indicate whether the species differences are attributed to alternative species from closely-related lineages being represented in samples (Souza-Neto et al. 2016). Here, indices for phylogenetic beta diversity are lower than the ones for taxonomic beta diversity, providing evidence for phylogenetic replacement. Unlike the taxonomic beta diversity, phylogenetic beta diversity shows that the medicinal plants are more similar to each other than would be expected given phylogenetic measures of floristic differences. Despite the taxonomic heterogeneity of the medicinal plant landscape, phylogenetic replacement is a homogenizing force. Heterogeneity because of different species use is contrasted with phylogenetic continuity suggestive of relative uniformity of traditional knowledge across Brazil. The geographical extent of vernacular names in Portuguese to describe related Brazilian medicinal plants supports the view that traditional knowledge of related plants extends beyond the limits of individual species (Souza, unpublished). Generic complexes are a well-known phenomenon to ethnobotanists (Berlin, 1973),

but formal comparative studies of the spatial extent of names beyond species distributions are lacking.

Principal Component of Phylogenetic Structure (PCPS): PCPS has been used in ecology to identify lineages contributing to the phylogenetic differences between communities (Brum et al. 2012; Carlucci et al. 2017; Duarte et al. 2014, 2012; Gianuca et al. 2014; Pillar and Duarte, 2010). Applied here for the first time to medicinal floras, it is used to identify lineages contributing the differences between medicinal floras of the different biomes. The first two principal components in our study are responsible for 72% of the variation found between biomes, and correspond to deep lineage differences (appendices 8 and 9). Although significantly clustered at deep lineages, there are preferences for medicinal plants that differ between biomes. PCPS identifies the lineages which account for the relatively high phylogenetic distances between medicinal plants of Caatinga, Pampas and Pantanal biomes. Cerrado, Amazon and Atlantic medicinal plants are relatively similar at deep phylogenetic levels - Pampas and Caatinga differ from these because of the relatively few monocots+magnoliids used, whereas Pantanal medicinal floras differ from the Cerrado/Amazon/Atlantic cluster because of the over-use of monocots+magnoliids. These data, focused on deeper phylogenetic levels are not indicative of phylogenetic replacement but indicate deep lineages where phylogenetic replacement is more likely.

5.4.6   Conclusions

Medicinal plant knowledge has been characterized both as deep, the result of generations of experimentation in-situ, and as readily transmitted. The two processes exert opposite influences, the former driving towards strong spatial structure in knowledge, and the latter towards a more homogeneous landscape of medicinal plant use. This study shows that medicinal plant use is not homogeneous across Brazil. It is well-known that plant use can vary at small spatial scales where cultural differences are strong, and this has been documented for traditional knowledge of medicinal plants in Brazil (Shepard, 2004). At a landscape scale, our data also point towards local knowledge. Plant use in Brazil is more different than predicted by floristic differences at species level. This difference is despite a preference for some plant groups, significant migration and homogeneity in terms of language – three potential drivers of uniformity in plant use. In Brazil, the influence of the local environment in medicinal plant selection has been documented (Albuquerque et al. 2005; de Albuquerque and de Oliveira, 2007), through analysis of local, species level data. Using methods borrowed from ecological studies to understanding changes in plant composition in different environments, we shed light on distribution of medicinal plant knowledge in relation to floristic environment, but on a much larger scale. These Brazilian data might serve as benchmarks for future, more global studies of the distribution of traditional knowledge.

Chapter 6 General discussion and conclusions

6.1   Novel contributions

This thesis describes my exploration of ethnobotanical documentation together with phylogenetic and spatial patterns in species with medicinal use in Brazil. Brazil has one of the world's richest floras (Forzza et al. 2012), distributed in six different environments called biomes. The association of floristic environments with different cultural expressions is reflected in the rich, though not homogeneous, traditional knowledge present in Brazil. With a history of migration and colonization, medicinal plant use in Brazil that once was part of local indigenous knowledge, adapted and evolved to encompass traditional knowledge from other populations that moved to Brazil (Voeks, 2013). Plant knowledge began to be described and compiled in the 1800's (Brandão et al. 2012), and since then, the number of ethnobotanical investigations has increased significantly (Ritter et al. 2015). From its origins documenting traditional knowledge, ethnobotanical research has begun to address broader questions searching for patterns, such as similarities or differences in medicinal plant use, and seeking to explain them. Often research of this kind is based on published data (de Medeiros et al. 2013).

In this study, chapter 2 investigated the role of herbarium specimens in providing information for ethnobotanical research. By comparing information from

published sources and herbarium data I showed that herbaria are underutilized sources of ethnobotanical information. Despite the potential of herbaria to contribute to new species to lists, and new information for species already documented in the literature, the comparison I made here draws attention to arguments for better documentation of data associated with herbarium collections, especially in the light of increasing digitisation (Blagoderov et al. 2012; Thiers et al. 2016).

To manage the data compiled for species of Leguminosae with use medicinally, I developed a database, described in chapter 3. Besides storing data, Ewé also provided tools that supported some of the analyses that were used in my thesis research, such as identifying the most reported genera and their uses, and mapping the distribution of medicinal use of plant species. Although data present in Ewé to date relates to medicinal uses of Leguminosae in Brazil, there is potential to expand the database to include the whole medicinal flora of Brazil, and develop co-participation in populating and maintaining the database.

Medicinal plants are important not only in traditional health systems, but also as sources of biomedical drugs (Fabricant and Farnsworth, 2001; Gurib-Fakim, 2006). Bioprospecting could be informed by phylogeny (Saslis-Lagoudakis et al. 2012a), but in Brazil such methodology has never been applied. Using the data I compiled, it was possible to investigate patterns in the phylogenetic distribution of species with medicinal use, in order to identify lineages that could be targeted in bioprospecting. This work is described in chapter 4. By relating pharmacological study for the Leguminosae to use data, I found 50% of species used medicinally had been screened.

In a phylogenetic framework I showed lineages rich in species with medicinal use were better characterised. Although phylogenetic studies had not been used in Brazil to guide investment in pharmacological study, research effort was greater in groups of closely-related species that included significantly more plants used in medicine.

Chapter 5 explored the spatial distribution of traditional knowledge, also using data compiled from herbaria, but for all Angiosperms in Brazil. With more than 7 000 records distributed amongst the six Brazilian biomes, it was possible to identify spatial heterogeneity in plant use, characterised by a greater turn-over in plants with medicinal use than the floristic turn over. In this chapter, I used methods borrowed from ecology and biogeography to investigate species composition change between regions, anticipating that future research could be based on similar methodology to be able to compare results.

6.2   Data mining

Three sources of data were used for this thesis work. All were compiled from existing available data. Firstly, traditional knowledge related to medicinal plant use was sourced from the scientific literature and from herbarium vouchers. The intensity of research effort relating to pharmacological studies of the Leguminosae species of Brazil was estimated by searching the scientific literature. Thirdly, the phylogenetic frameworks used to interpret the pharmacological and the spatial data were reconstructed using DNA sequences from public databases, existing phylogenetic

hypotheses and inferences about relationship from existing taxonomies. Although much research is based on primary data collected in the field or laboratory, science today is frequently based on data mining as research in biology and health care becomes increasingly data intensive (Aronova et al. 2010; Birney, 2012; Liu et al. 2012; Thessen and Patterson, 2011). Data mining can be challenging, particularly when information is scattered in different places and data is presented in a complex, weakly-structured way, as is often the case in ethnobotany (Albuquerque and Medeiros, 2012). There is likely a reciprocity between research using primary and secondary data. The work presented here, based on secondary data, might influence what data is collected and where it is collected in the future.

The first task I took on in this project was data mining from literature and herbaria. Ethnobotanical literature is increasing in Brazil, remarkably since 2010 (Ritter et al. 2015). Although access to these papers was easy through tools like Google Scholar and Web of Knowledge, data collection and information quality assessment from each paper was time consuming. Despite increasing efforts to develop and benefit from text mining tools (Fan et al. 2006), there is a lack of a proper ontology in ethnobotanical data, and species names, therapeutic applications and preparations are presented in a diversity of ways (Christie, 2006; Daly et al. 2016; Ellen, 2016). Therefore, traditional knowledge data generally needs to be manually compiled and standardised prior to further comparison. Compilation of herbarium data is influenced by access to herbarium vouchers and availability of information on the vouchers labels. The digitisation of herbarium vouchers greatly benefited this thesis research, which used voucher information from CRIA-Species link, a Brazilian online biological

collection. Not all herbaria are yet digitised, and lack of uniformity in compilation of information from labels can be a barrier. CRIA-Species link transcribed label data, so I could search for key words. Generally labels are not transcribed in full, and a system of core descriptors are used by most major herbaria (Wieczorek et al. 2012). Most digitised herbarium collections cannot therefore be used to provision data for ethnobotanical meta-analysis. Also, as found in this study many times, information from herbarium vouchers can be incomplete if label text is only partially transcribed. This was true for both traditional use information and geographical coordinates, which were missing or gave a wrong point locality when plotted on a map.

My study draws attention to the need to establish a set of parameters and an ontology capable of assuring a better use of digitization efforts to extract as much information as possible (Beaman and Cellinese, 2012). Ontologies to describe traditional knowledge for all data collections, literature and herbaria, are needed to improve comparison power and address comprehensive questions or meta-analyses. Nevertheless, it is argued that such etic (scientific/outsider) standardisation distorts the emic (local) perception of local TK, discarding information (Staub et al. 2015). In this sense, I documented both standardised and traditional knowledge as presented in the literature, though the literature may already present an etic view. The use of secondary sources presents a risk of bias related to species identification, or sample reliability if used in meta-analyses (de Medeiros et al. 2013). Misleading interpretation of the emic view presents a further risk. For herbarium vouchers citing medicinal use, where there is no description of the source of such information or the methodologies used, it is harder to estimate whether these errors have been introduced.

6.3   The phylogenetic framework

To date, the phylogenetic framework applied to the interpretation of plants used in medicine, or potentially used in medicine, has been used in two ways. The first relates to describe hot nodes in traditional use and to identify closely-related lineages from different medicinal floras (Saslis-Lagoudakis et al. 2012a). This first framework is based on community phylogenetic metrics. The second framework is based on evolutionary metrics (Pagel's lambda), and relates to the identification of a phylogenetic signal in traits or medicinal chemistries of interest (Cámara-Leret et al. 2017; Ernst et al. 2016; Halse-Gramkow et al. 2016; Rønsted et al. 2008). Novelty in this study is two-fold. Phylogenetic frameworks are used, first to better understand research effort for the first time and secondly to characterise spatial distribution of traditional knowledge, again for the first time.

Phylogenetic frameworks have been used to inform the potential of lineages rich in medicinally used species as targets for bioprospecting (Saslis-Lagoudakis et al. 2012a). Nevertheless, this hypothesis is relatively new, and extensive evaluation and corroboration is needed. In chapter 4, although the phylogenetic framework was the same as the one used by Saslis-Lagoudakis et al. (2012a), by comparing lineages rich in medicinally used species and which species had been screened, it was possible to identify gaps in knowledge and lineages that could be targeted for future studies. It is not possible to affirm that pharmacological studies have been guided explicitly by the

search or closely-related species. Nevertheless, results shown in chapter 4 evidence the extent of pharmacological investigation for the Leguminosae, where lineages rich in medicinally used species are better characterised. At the same time, it draws attention to the need for further investigation of less characterised plant species, that still comprise an important part of traditional knowledge systems. The combination of methods used here might help on a better identification of new targets for bioprospecting.

The second novel phylogenetic framework presented here was the use of community ecology and biogeography to test turnover in medicinal species composition between biomes, in chapter 5. While many ethnobotanical studies have been concerned with listing plant species (an important task to preserve traditional knowledge), the use of more robust and comparable methodologies is of great interest for the theoretical foundations of ethnobotany (Albuquerque et al. 2016). Therefore, methodologies to investigate the relationships between humans and plant species are being "developed" or borrowed from other areas in science in order to address hypothesis-driven questions (Albuquerque and Hanazaki, 2009). In chapter 5, by using methods borrowed from biogeography and community phylogenetics, it was possible not only to visualise the spatial structure of traditional knowledge in different biomes, but also to quantify turnover in species composition. These methods have not been used in ethnobotanical research before, and I expect that future studies might use similar methods facilitating comparisons.

6.4   Phylogenies and evolutionary ethnobotany

Evolutionary ethnobotany is a relative new area being explored by ethnobotanists, and is reliant on phylogenetic frameworks. Phylogenetic investigations can test whether there are evolutionary relationships underlying traditional use, guiding drug discoveries (Saslis-Lagoudakis et al. 2015), or investigate trait similarities (Cámara-Leret et al. 2017; Ernst et al. 2016; Halse-Gramkow et al. 2016). As shown here in chapter 5, phylogenetic turnover can also be accounted for when investigating spatial distribution of traditional knowledge. Although studies to date revealed significant patterns in medicinal plant use, many were based on taxonomic scales ranging from floras (with generic level trees (Saslis-Lagoudakis et al. 2012a; Saslis-Lagoudakis et al. 2014; Yessoufou et al. 2015) to genus level (Ernst et al. 2016; Saslis-Lagoudakis et al. 2011a). Phylogenies in this thesis were reconstructed based on available sequences, and at generic level. A species-level tree for the Leguminosae, despite having almost all species could only be reconstructed if species were included as polytomies. Although community phylogenetics is often focused on deep nodes, with unaccounted or non-significant relationships at shallow nodes, better species-level trees are needed.

The selection of medicinal plant species has been investigated, and the idea of the presence of similar characteristics underlying species selection is explored in diverse ways. Casagrande (2000) explores the idea of medicinal prototypes, or groups of plant species with basic characteristics that identify species as medicinal. Ferreira Júnior and Albuquerque (2015) go further, and explore the existence of a "consensus

106

within a diversity", where species with great consensus represent an important core of a traditional medicine system, and are less likely to be replaced over time. Hawkins and Teixidor-Toneu (2017) and Garnatje et al. (2017) discuss concepts phylogenetic convergence, homology and shared traits medicinal species selection, and concepts of core, consensus species would benefit enormously from a well-resolved species level phylogeny. Species trees compared to a rich trait matrix, would permit the identification of generic complexes (Martin, 2004), and turnover and core lineages and phylogenetic replacement could be better characterised.

6.5 Future work

All analyses presented here were based on large-scale data distribution, related to the whole of Brazil (whether one family or the whole of the Angiosperms). Nevertheless, I believe future work on smaller scales, comparing local communities or investigating genera can benefit from the framework presented here. Much of the ethnobotanical information collected for this thesis had its origin in herbarium specimen labels (part of the Leguminosae in chapters 2, 3 and 4, and all of data for chapter 5). Although this work was facilitated by online availability of the data, not all herbaria in Brazil are digitised, and efforts to collect more data from this source could yield a more comprehensive picture of the ethnomedicinal flora. Nevertheless, for some herbarium this would be a huge task, for example in the herbarium UB, the 27% of specimens not yet digitised comprise 80000 specimens to be looked at (Proença-personal communication). Another challenge here would be to collect all information present in literature for Brazil.

Ewé database was designed not only to be used to support data storage and analysis for this thesis, but to be populated with ethnobotanical data for the whole Brazilian flora, and ideally, to be used by other ethnobotanists and ethnopharmacologists in Brazil. Ewe database development tools, being an open source scheme, can be useful in this expansion by allowing a collaborative development of the database by the users, allied to data being provided by other users, in order to expand and adequate to all research group's needs. Nevertheless, challenges here include funding and maintenance of the database system to assure its sustainability, that could be sought with Brazilian universities and science and technology agencies, and agreements on an ontology to guide classification of therapeutic applications, and taxonomic identifiers.

Another challenge encountered in this thesis is related to species names reported as currently accepted. In some chapters, species names followed the Brazilian flora while in others followed the Plant List and Trópicos. Although the latest version of the Brazilian Flora is still being worked on, many species from herbarium specimens were not reported in the Flora, but were reported as accepted names in the other species lists. A final list with all accepted species name in the future will help in overcoming issues like this one, and facilitate in the database organisation.

Finally, the phylogenetic framework reported in this thesis could be significantly improved, by addition of more species-level sequence data. Phylogenetic reconstruction, perhaps using Next Generation Sequencing, would benefit enormously

from cloud computing and computer clusters. Cloud computing would decrease analysis time and allow the possibility of working with larger and more comprehensive datasets, that could lead to better resolved phylogenies and an enhanced understanding of evolutionary relationships between humans and plant use.

References

Abellán, P., Carrete, M., Anadón, J.D., Cardador, L., Tella, J.L., 2016. Non-random patterns and temporal trends (1912-2012) in the transport, introduction and establishment of exotic birds in Spain and Portugal. Diversity and Distributions 22, 263–273.

Agra, M. de F., Silva, K.N., Basílio, I.J.L.D., Freitas, P.F. de, Barbosa-Filho, J.M., 2008. Survey of medicinal plants used in the region Northeast of Brazil. Brazilian Journal of Pharmacognosy 18, 472–508.

Agra, M.D.F., Freitas, P.F. De, Barbosa-Filho, J.M., 2007. Synopsis of the plants known as medicinal and poisonous in Northeast of Brazil. Revista Brasileira de Farmacognosia, 17(1), 114-140.

Agrawal, A.A., Salminen, J.-P., Fishbein, M., 2009. Phylogenetic trends in phenolic metabolism of milkweeds (*Asclepias*): evidence for escalation. Evolution 63, 663–673.

Albuquerque, U.P., Ferreira Júnior, W.S., 2017. What do we study in evolutionary ethnobiology? Defining the theoretical basis for a research program. Evolutionary Biology 44, 206–215.

Albuquerque, U.P., Ramos, M.A., Melo, J.G., 2012. New strategies for drug discovery in tropical forests based on ethnobotanical and chemical ecological studies. Journal of Ethnopharmacology 140, 197–201.

Albuquerque, U.P., 2010. Implications of ethnobotanical studies on bioprospecting strategies of new drugs in semi-arid regions. The Open Complementary Medicine Journal 2, 21–23.

Albuquerque, U.P., de Medeiros, P.M., 2012. Systematic reviews and meta-analysis applied to ethnobiological research. Ethnobiology and Conservation 1, 1–8.

Albuquerque, U.P., de Medeiros, P.M., Casas, A., 2016. Evolutionary Ethnobiology. Springer, Recife.

Albuquerque, U.P., de Oliveira, R.F., 2007. Is the use-impact on native Caatinga species in Brazil reduced by the high species richness of medicinal plants? Journal of Ethnopharmacology 113, 156–70.

Albuquerque, U.P., de Medeiros, P.M., de Almeida, A.L.S., Monteiro, J.M., Neto, E.M.F.L., de Melo, J.G., dos Santos, J.P., 2007. Medicinal plants of the Caatinga (semi-arid) vegetation of NE Brazil: a quantitative approach. Journal of Ethnopharmacology 114, 325–54.

Albuquerque, U.P., Hanazaki, N., 2009. Five problems in current ethnobotanical research—and some suggestions for strengthening them. Human Ecology 37, 653–661.

Alexander, M., Chamundeeswari, K., Kambu, A., Ruiz, M., & Tobin, B., 2004. The role of registers and databases in the protection of traditional knowledge: A comparative analysis. United Nations University Institute of Advanced Studies (UNU-IAS), Tokyo.

Alves, R.R.N., Rosa, I.M.L., 2007. Biodiversity, traditional medicine and public health: where do they meet? Journal of Ethnobiology and Ethnomedicine 3, 14.

Alkin, B., Patmore, K., Black, N., Booker, Canteiro, C., Dauncey, E., Edwards S., Forest, F., Giovannini, P., Howes, M-J., Hudson, A., Irving, J., Leon, C., Miliken, W., Lughadha, E.N., Schippmann, U., Simmonds., M. 2017. Useful plants- medicines, in: Willis, K.J. (Eds), State of the World Plants Report 2017, Royal Botanic Gardens, Kew.

Amiguet, V.T., Arnason, J.T., Maquin, P., Cal, V., Sánchez-Vindas, P., Alvarez, L.P., 2006. A regression analysis of Q'eqchi' Maya medicinal plants from southern Belize. Economic Botany 60, 24–38.

Anagnostou, S., 2007. The international transfer of medicinal drugs by the society of Jesus (sixteenth to eighteenth centuries) and connections with the work of Carolus Clusius, in: Egmond, F., Hoftijzer, P., Visser, R. (Eds.), Carolus Clusius. Royal Netherlands Academy of Arts and Sciences, Amsterdam, p. 293–313.

Andrade-Cetto, A., Heinrich, M., 2011. From the field into the lab: useful approaches to selecting species based on local knowledge. Frontiers in Pharmacology 2, 1–5.

Ankli, A., Sticher, O., Heinrich, M., 1999. Yucatec Maya medicinal plants versus nonmedicinal plants: indigenous characterization and selection. Human Ecology 27, 557–580.

Anvisa- Agência Nacional de Vigilância Sanitária, 2010. Farmacopéia Brasileira, 5th ed. Ed. Brasília, Brasília.

Araújo, E. de L., Castro, C.C., Albuquerque, U.P. de, 2007. Dynamics of Brazilian Caatinga – a review concerning the plants, environment and people. Functional Ecosystems and Communities 1, 15–28.

Aronova, E., Baker, K.S., Oreskes, N., 2010. Big science and big data in biology: from the international geophysical year through the international biological program to the Long Term Ecological Research (LTER) network, 1957–present. Historical Studies in the Natural Sciences 40, 183–224.

Atanasov, A.G., Waltenberger, B., Pferschy-Wenzig, E.M., Linder, T., Wawrosch, C., Uhrin, P., Temml, V., Wang, L., Schwaiger, S., Heiss, E.H., Rollinger, J.M., Schuster, D., Breuss, J.M., Bochkov, V., Mihovilovic, M.D., Kopp, B., Bauer, R., Dirsch, V.M., Stuppner,

H., 2015. Discovery and resupply of pharmacologically active plant-derived natural products: a review. Biotechnology Advances 33, 1582–1614.

Attwood, T.K., Agit, B., Ellis, L.B.M., 2015. Longevity of biological databases. EMBnet.journal 21, e803.

Balick, M.J., Cox, P.A., 1996. Plants, People and Culture. The Science of Ethnobotany. Scientific American Library, New York.

Baptista, M.M., Ramos, M.A., de Albuquerque, U.P., Coelho-de-Souza, G., Ritter, M.R., 2013. Traditional botanical knowledge of artisanal fishers in southern Brazil. Journal of Ethnobiology and Ethnomedicine 9(1), 54.

Bastow, R., Leonelli, S., 2010. Sustainable digital infrastructure. EMBO Reports 11(10), 730–734.

Beaman, R.S., Cellinese, N., 2012. Mass digitization of scientific collections: new opportunities to transform the use of biological specimens and underwrite biodiversity science. ZooKeys 209, 7–17.

Bebber, D.P., Carine, M.A., Wood, J.R.I., Wortley, A.H., Harris, D.J., Prance, G.T., Davidse, G., Paige, J., Pennington, T.D., Robson, N.K.B., Scotland, R.W., 2010. Herbaria are a major frontier for species discovery. Proceedings of the National Academy of Sciences of the United States of America 107, 22169–71.

Bedigian, D., 2004. Slimy leaves and oily seeds: distribution and use of wild relatives of sesame in Africa. Economic Botany 58, S3–S33.

Begossi, A., Hanazaki, N., Tamashiro, J.Y., 2002. Medicinal plants in the Atlantic Forest (Brazil): knowledge, use, and conservation. Human Ecology 30, 281–299.

Benko-Iseppon, A.M., Crovella, S., 2010. Ethnobotanical bioprospection of candidates for potential antimicrobial drugs from Brazilian plants: state of art and perspectives. Current Protein & Peptide Science 11, 189–94.

Bennett, B.C., Husby, C.E., 2008. Patterns of medicinal plant use: an examination of the Ecuadorian Shuar medicinal flora using contingency table and binomial analyses. Journal of Ethnopharmacology 116, 422–30.

Bennett, B.C., Prance, G.T., 2000. Introduced plants in the indigenous pharmacopoeia of the northern South America. Economic Botany 54, 90–102.

Berlin, B., 1992. Ethnobiological Classification: Principles of Categorization Of Plants And Animals In Traditional Societies. Princeton University Press, Princeton.

Berlin, B., 1973. Folk systematics in relation to biological classification and nomenclature. Annual Review of Ecology and Systematics 4, 259–271.

Birch, J.L., 2009. A comparative analysis of nineteenth century pharmacopoeias in the Southern United States: a case study based on the Gideon Lincecum Herbarium. Economic Botany 63, 427–440.

Birney, E., 2012. The making of ENCODE: lessons for big-data projects. Nature 489, 49–51.

Blagoderov, V., Kitching, I.J., Livermore, L., Simonsen, T.J., Smith, V.S., 2012. No specimen left behind: industrial scale digitization of natural history collections. ZooKeys 133–46.

Bourgaud, F., Gravot, A., Milesi, S., Gontier, E., 2001. Production of plant secondary metabolites: a historical perspective. Plant Science 161, 839–851.

Brandão, M.G.L., Pignal, M., Romaniuc, S., Grael, C.F.F., Fagg, C.W., 2012. Useful Brazilian plants listed in the field books of the French naturalist Auguste de Saint-Hilaire (1779-1853). Journal of Ethnopharmacology 143, 488–500.

Brandão, M.G.L., Zanetti, N.N.S., Oliveira, P., Grael, C.F.F., Santos, A.C.P., Monte-Mór, R.L.M., 2008. Brazilian medicinal plants described by 19th century European naturalists and in the official pharmacopoeia. Journal of Ethnopharmacology 120, 141–8.

Brouwer, N., Liu, Q., Harrington, D., Kohen, J., Vemulpad, S., Jamie, J., Randall, M., Randall, D., 2005. An ethnopharmacological study of medicinal plants in New South Wales. Molecules 10, 1252–62.

Brum, F.T., Kindel, A., Hartz, S.M., Duarte, L.D.S., 2012. Spatial and phylogenetic structure drive frugivory in *Tyrannidae* birds across the range of Brazilian araucaria forests. Oikos 121, 899–906.

Bryant, J.A., Lamanna, C., Morlon, H., Kerkhoff, A.J., Enquist, B.J., Green, J.L., 2008. Microbes on mountainsides: contrasting elevational patterns of bacterial and plant diversity. Proceedings of the National Academy of Sciences of the United States of America 105 Supplement 1, 11505–11.

Bussmann, R.W., Sharon, D., 2006. Traditional medicinal plant use in Northern Peru: tracking two thousand years of healing culture. Journal of Ethnobiology and Ethnomedicine 2(1), 47.

Calixto, J.B., 2000. Efficacy, safety, quality control, marketing and regulatory guidelines for herbal medicines (phytotherapeutic agents). Brazilian Journal of Medical and Biological Research 33, 179–189.

Cámara-Leret, R., Faurby, S., Macía, M.J., Balslev, H., Göldel, B., Svenning, J.-C., Kissling, W.D., Rønsted, N., Saslis-Lagoudakis, C.H., 2017. Fundamental species traits explain provisioning services of tropical American palms. Nature Plants 3(2), 16220.

Cámara-Leret, R., Paniagua-Zambrana, N., Balslev, H., Macía, M.J., Borchsenius, F., Ferreira, E., Granville, J. de, Mejía, K., Millán, B., Moraes, M., Noblick, L., Stauffer, F., Kahn, F., Camara-Leret, R., Paniagua-Zambrana, N., Balslev, H., Macia, M.J., 2014a. Ethnobotanical knowledge is vastly under-documented in northwestern South America. PLoS ONE 9, e85794.

Cámara-Leret, R., Paniagua-Zambrana, N., Svenning, J.-C., Balslev, H., Macía, M.J., 2014b. Geospatial patterns in traditional knowledge serve in assessing intellectual property rights and benefit-sharing in northwest South America. Journal of Ethnopharmacology 158, 58–65.

Campos, M.T., Ehringhaus, C., 2003. Plant virtues are in the eyes of the beholders: a comparison of known palm uses among indigenous and folk communities of southwestern Amazonia. Economic Botany 57, 324–344.

Carlucci, M.B., Seger, G.D.S., Sheil, D., Amaral, I.L., Chuyong, G.B., Ferreira, L. V., Galatti, U., Hurtado, J., Kenfack, D., Leal, D.C., Lewis, S.L., Lovett, J.C., Marshall, A.R., Martin, E., Mugerwa, B., Munishi, P., Oliveira, Á.C.A., Razafimahaimodison, J.C., Rovero, F., Sainge, M.N., Thomas, D., Pillar, V.D., Duarte, L.D.S., 2017. Phylogenetic composition and structure of tree communities shed light on historical processes influencing tropical rainforest diversity. Ecography 40, 521–530.

Carvalho, G.H., Cianciaruso, M.V., Batalha, M.A., 2010. Plantminer: a web tool for checking and gathering plant species taxonomic information. Environmental Modelling and Software 25, 815–816.

Casagrande, D.G., 2000. Human taste and cognition in Tzeltal Maya medicinal plant use. Journal of Ecological Anthropology 4, 57–68.

Ceolin, T., Heck, R.M., Barbieri, R.L., Schwartz, E., Muniz, R.M., Pillon, C.N., 2011. Plantas medicinais: transmissão do conhecimento nas famílias de agricultores de base ecológica no sul do RS. Revista da Escola de Enfermagem da USP 45, 47–54.

Champagne, D.E., Isman, M.B., Downum, K.R., Towers, G.H.N., 1993. Insecticidal and growth-reducing activity of foliar extracts from Meliaceae. Chemoecology 4, 165–173.

Christie, M., 2006. Boundaries and accountabilities in computer-assisted ethnobotany. Research and Practice in Technology Enhanced Learning 1, 285–296.

Coley, P.D., Heller, M. V., Aizprua, R., Araúz, B., Flores, N., Correa, M., Gupta, M., Solis, P.N., Ortega-Barría, E., Romero, L.I., Gómez, B., Ramos, M., Cubilla-Rios, L., Capson, T.L., Kursar, T.A., 2003. Using ecological criteria to design plant collection strategies for drug discovery. Frontiers in Ecology and the Environment 1, 421–428.

Cook, F.E.M., 1995. Economic botany data collection standard. Royal Botanic Gardens, Kew.

Costa, M.A., Palazzo de Mello, J.C., Kaneshima, E.N., Ueda-Nakamura, T., Dias Filho, B.P., Audi, E.A., Nakamura, C.V., Dias Filho, B.P., Audi, E.A., Nakamura, C.V., 2013. Acute and chronic toxicity of an aqueous fraction of the stem bark of *Stryphnodendron adstringens* (Barbatimão) in rodents. Evidence-based Complementary and Alternative Medicine: eCAM 2013, 841580.

Cox, P.A., 2000. Will tribal knowledge survive the millennium? Science 287(5450), 44-45.

Cragg, G.M., Newman, D.J., 2013. Natural products: a continuing source of novel drug leads. Biochimica et Biophysica Acta - General Subjects 1830, 3670–3695.

Dafni, A., 2006. On the typology and the worship status of sacred trees with a special reference to the Middle East. Journal of Ethnobiology and Ethnomedicine 2, 26.

Daly, L., French, K., Miller, T.L., Nic Eoin, L., 2016. Integrating ontology into ethnobotanical research. Journal of Ethnobiology 36, 1–9.

Dang, L., Van Damme, E.J.M., 2015. Toxic proteins in plants. Phytochemistry 117, 51–64.

de la Torre, L., Cerón, C.E., Balslev, H., Borchsenius, F., 2012. A biodiversity informatics approach to ethnobotany: meta-analysis of plant use patterns in Ecuador. Ecology and Society 17, 15.

de Medeiros, P.M., Ladio, A.H., Albuquerque, U.P., 2013. Patterns of medicinal plant use by inhabitants of Brazilian urban and rural areas: a macroscale investigation based on available literature. Journal of Ethnopharmacology 150, 729–46.

de Medeiros, P.M. de, Soldati, G.T., Alencar, N.L., Vandebroek, I., Pieroni, A., Hanazaki, N., de Albuquerque, U.P., 2011. The use of medicinal plants by migrant people: adaptation, maintenance, and replacement. Evidence-Based Complementary and Alternative Medicine, 2012

De Natale, A., Pollio, A., 2012. A forgotten collection: the Libyan ethnobotanical exhibits (1912-14) by A. Trotter at the museum O. Comes at the university Federico II in Naples, Italy. Journal of Ethnobiology and Ethnomedicine 8, 4.

Deus, J.A.S. de, 2011. Identidade, etnicidade e paisagens culturais alternativas no vale do Rio Doce, Minas Gerais-Brasil. Revista Geográfica de América Central 2, 1–14.

Dhiman, A.K., 2003. Sacred Plants and Their Medicinal Uses. Daya Publishing House, Delhi.

Dias, S.A., Neves, A.E.O.O., Ferraz, A. de B.F.F., Picada, J.N., Pereira, P.P., 2013. Neuropharmacological and genotoxic evaluation of ethanol extract from *Erythrina*

falcata leaves, a plant used in Brazilian folk medicine. Brazilian Journal of Pharmacognosy 23, 335–341.

Douwes, E., Crouch, N.R., Edwards, T.J., Mulholland, D.A., 2008. Regression analyses of southern African ethnomedicinal plants: informing the targeted selection of bioprospecting and pharmacological screening subjects. Journal of Ethnopharmacology 119, 356–364.

Downes, D.R., 2000. How intellectual property could be a tool to protect traditional knowledge. Columbia Journal of Environmental Law 25, 253.

Duarte, L. da S., 2011. Phylogenetic habitat filtering influences forest nucleation in grasslands. Oikos 120, 208–215.

Duarte, L.D.S., Both, C., Debastiani, V.J., Carlucci, M.B., Gonçalves, L.O., Cappelatti, L., Seger, G.D.S., Bastazini, V.A.G., Brum, F.T., Salengue, E. V., Bernardo-Silva, J.S., 2014. Climate effects on amphibian distributions depend on phylogenetic resolution and the biogeographical history of taxa. Global Ecology and Biogeography 23, 213–222.

Duarte, L.D.S., Prieto, P. V., Pillar, V.D., 2012. Assessing spatial and environmental drivers of phylogenetic structure in Brazilian *Araucaria* forests. Ecography 35, 952–960.

Dutra, R.C., Campos, M.M., Santos, A.R.S., Calixto, J.B., 2016. Medicinal plants in Brazil: pharmacological studies, drug discovery, challenges and perspectives. Pharmacological Research 112, 4–29.

Edwards, S. E., & Heinrich, M.. 2006. Redressing cultural erosion and ecological decline in a far North Queensland aboriginal community (Australia): the Aurukun ethnobiology database project. Environment, Development and Sustainability, 8(4), 569-583.

Eisenman, S.W., Tucker, A.O., Struwe, L., 2012. Voucher Specimens are Essential for Documenting Source Material Used in Medicinal Plant Investigations. Journal of Medicinally Active Plants 1, 30–43.

Ekor, M., 2014. The growing use of herbal medicines: issues relating to adverse reactions and challenges in monitoring safety. Frontiers in Pharmacology 4, 177.

Ellen, R., 2016. Is there a role for ontologies in understanding plant knowledge systems? Journal of Ethnobiology 36, 10–28.

Ellen, R., Puri, R., 2016. Conceptualising "Core" medicinal floras: a comparative and methodological study of phytomedical resources in related Indonesian populations. Conservation and Society 14, 345.

Ernst, M., Saslis-Lagoudakis, C.H., Grace, O.M., Nilsson, N., Simonsen, H.T., Horn, J.W., Rønsted, N., 2016. Evolutionary prediction of medicinal properties in the genus *Euphorbia* L. Scientific Reports 6, 30531.

Fabricant, D.S., Farnsworth, N.R., 2001. The value of plants used in traditional medicine for drug discovery. Environmental Health Perspectives 109 Supplement, 69–75.

Fan, W., Wallace, L., Rich, S., Zhang, Z., 2006. Tapping the power of text mining. Communications of the ACM 49(9), 76-82.

Fantz, P.R., 1991. Ethnobotany of *Clitoria*. Economic Botany 45, 511–520.

Farnsworth, N.R., Akerele, O., Bingel, A.S., Soejarto, D.D., Guo, Z., 1985. Medicinal plants in therapy. Bulletin of the World Health Organization 63, 965–81.

Feng, G., Zhang, J., Pei, N., Rao, M., Mi, X., Ren, H., Ma, K., 2012. Comparison of phylobetadiversity indices based on community data from Gutianshan forest plot. Chinese Science Bulletin 57, 623–630.

Fernandes, G., Banu, J., 2012. Medicinal properties of plants from the genus *Cissus*: a review. Journal of Medicinal Plants Research 6, 3080–3086.

Ferreira Júnior, W.S., Albuquerque, U.P., 2015. "Consensus within diversity": an evolutionary perspective on local medical systems. Biological Theory 10, 363–368.

Firn, R.D., 2003. Bioprospecting – why is it so unrewarding? Biodiversity and Conservation 12, 207–216.

Fleury, M.T.L., 1999. The management of culture diversity: lessons from Brazilian companies. Industrial Management & Data Systems 993, 109–114.

Flora of Brazil 2020 in construction., 2016. Jardim Botânico do Rio de Janeiro. Available from: < http://floradobrasil.jbrj.gov.br/ >.

Follmer, C., Barcellos, G.B., Zingali, R.B., Machado, O.L., Alves, E.W., Barja-Fidalgo, C., Guimarães, J.A., Carlini, C.R., 2001. Canatoxin, a toxic protein from jack beans (*Canavalia ensiformis*), is a variant form of urease (EC 3.5.1.5): biological effects of urease independent of its ureolytic activity. The Biochemical Journal 360, 217–24.

Fonseca-Kruel, V.S., Silva, I.M., Pinheiro, C.U., 2005. O ensino acadêmico da etnobotânica no Brasil. Rodriguesia 56, 97–106.

Forest, F., Grenyer, R., Rouget, M., Davies, T.J., Cowling, R.M., Faith, D.P., Balmford, A., Manning, J.C., Procheş, Ş., van der Bank, M., Reeves, G., Hedderson, T.A.J., Savolainen, V., 2007. Preserving the evolutionary potential of floras in biodiversity hotspots. Nature 445, 757–760.

Forzza, R.C., et al. 2012. New Brazilian floristic list highlights conservation challenges. BioScience 62, 39–45.

Funk, V. A., 2003. The importance of herbaria. Plant Science Bulletin 49, 94–95.

Gaikwad, J., Wilson, K., Kohen, J., Vemulpad, S., Jamie, J., Ranganathan, S., 2011. Combining ethnobotany and informatics to discover knowledge from data, in: Ethnomedicinal plants: revitalization of traditional knowledge of herbs, Science Publishers, Enfield, New Hampshire, USA, 444-457.

Garnatje, T., Peñuelas, J., Vallès, J., 2017. Ethnobotany, phylogeny, and "omics" for human health and food security. Trends in Plant Science 22, 187–191.

Geck, M.S., Cabras, S., Casu, L., Reyes García, A.J., Leonti, M., 2017. The taste of heat: how humoral qualities act as a cultural filter for chemosensory properties guiding herbal medicine. Journal of Ethnopharmacology 198, 499–515.

Gertsch, J., 2009. How scientific is the science in ethnopharmacology? Historical perspectives and epistemological problems. Journal of Ethnopharmacology 122, 177–183.

Gianuca, A.T., Dias, R.A., Debastiani, V.J., Duarte, L.D.S., 2014. Habitat filtering influences the phylogenetic structure of avian communities across a coastal gradient in southern Brazil. Austral Ecology 39, 29–38.

Gonçalves, A.J., 2001. Migrações internas: evoluções e desafios. Estudos Avançados 15, 173–184.

Goodwin, Z.A., Harris, D.J., Filer, D., Wood, J.R.I., Scotland, R.W., 2015. Widespread mistaken identity in tropical plant collections. Current Biology 25, R1066–R1067.

Grace, O.M., Buerki, S., Symonds, M.R.E., Forest, F., van Wyk, A.E., Smith, G.F., Klopper, R.R., Bjorå, C.S., Neale, S., Demissew, S., Simmonds, M.S.J., Rønsted, N., 2015. Evolutionary history and leaf succulence as explanations for medicinal use in aloes and the global popularity of *Aloe vera*. BMC Evolutionary Biology 15, 29.

Graham, C.H., Fine, P.V.A., 2008. Phylogenetic beta diversity: linking ecological and evolutionary processes across space in time. Ecology Letters 11, 1265–1277.

Gruca, M., Cámara-Leret, R., Macía, M. J., & Balslev, H., 2014. New categories for traditional medicine in the Economic Botany Data Collection Standard. Journal of Ethnopharmacology, 155(2), 1388-1392.

Guarim-Neto, G., Morais, R.G., 2003. Recursos medicinais de espécies do Cerrado de Mato Grosso: um estudo bibliográfico. Acta Botanica Brasilica 17, 561–584.

Guglielmino, C. R., Viganotti, C., Hewlett, B., & Cavalli-Sforza, L. L., 1995. Cultural variation in Africa: role of mechanisms of transmission and adaptation. Proceedings of the National Academy of Sciences, 92, 7585-7589.

Gupta A. K. 2004. WIPO-UNEP study on the role of intellectual property rights in the sharing of benefits arising from the use of biological resources and associated traditional knowledge-Study no: 4. WIPO.

Gupta, M.P., Solís, P.N., Calderón, A.I., Guinneau-Sinclair, F., Correa, M., Galdames, C., Guerra, C., Espinosa, A., Alvenda, G.I., Robles, G., Ocampo, R., 2005. Medical ethnobotany of the Teribes of Bocas del Toro, Panama. Journal of Ethnopharmacology 96, 389–401.

Gurib-Fakim, A., 2006. Medicinal plants: traditions of yesterday and drugs of tomorrow. Molecular Aspects of Medicine 27, 1–93.

Halberstein, R.A., 2005. Medicinal Plants: historical and cross-cultural usage patterns. Annals of Epidemiology 15, 686–699.

Halse-Gramkow, M., Ernst, M., Rønsted, N., Dunn, R.R., Haris Saslis-Lagoudakis, C., 2016. Using evolutionary tools to search for novel psychoactive plants. Plant Genetic Resources: Characterization and Utilization 14, 246–255.

Harvey, A.L., Edrada-Ebel, R., Quinn, R.J., 2015. The re-emergence of natural products for drug discovery in the genomics era. Nature Publishing Group 14, 111–129.

Harshberger, J.W. 1896. The purposes of ethno-botany. Botanical Gazette 21:146-154.

Haston, E., Cubey, R., Pullan, M., Atkins, H., Harris, D.J., 2012. Developing integrated workflows for the digitisation of herbarium specimens using a modular and scalable approach. ZooKeys 209, 93–102.

Hawkins, J.A., Teixidor-Toneu, I., 2017. Defining "ethnobotanical convergence". Trends in Plant Science 22, 639–640.

Heinrich, M., Edwards, S., Moerman, D. E., Leonti, M..2009. Ethnopharmacological field studies: a critical assessment of their conceptual basis and methods. Journal of Ethnopharmacology, 124(1), 1-17.

Helmstädter, A., 2017. The botanical explorer's legacy: a promising bioprospecting tool. Drug Discovery Today 22, 757–760.

Hewlett, B.S., Cavalli-Sforza, L.L., 1986. Cultural Transmission among Aka Pygmies. American Anthropologist 88, 922–934.

Hill, A., Guralnick, R., Smith, A., Sallans, A., Rosemary Gillespie, Denslow, M., Gross, J., Murrell, Z., Tim Conyers, Oboyski, P., Ball, J., Thomer, A., Prys-Jones, R., de Torre, J., Kociolek, P., Fortson, L., 2012. The notes from nature tool for unlocking biodiversity records from museum records through citizen science. ZooKeys 219–33.

Honorio Junior, J.E.R., Soares, P.M., de Melo, C.L., Arruda Filho, A.C. V, Sena Filho, J.G., Barbosa Filho, J.M., Florenco Sousa, F.C., Franca Fonteles, M.M., de Almeida Leal, L.K., de Queiroz, M.G., Vasconcelos, S.M.M., 2010. Pharmacological activity of monocrotalina isolated from plants of the genus *Crotalaria*. Brazilian Journal of Pharmacognosy 20, 453–458.

IBGE (Instituto Brasileiro de Geografia e Estatística), Available from: http://www.ibge.gov.br/cidadesat

Jäger, A.K., 2015. Medicinal plant research: a reflection on translational tasks, in: Heinrich, M., Jäger, A.K. (Eds.), Ethnopharmacology. Wiley-Blackwell, Chichester, p. 11–16.

Jenks, A. A., Kim, S.C., 2013. Medicinal plant complexes of *Salvia* subgenus *Calosphace*: an ethnobotanical study of new world sages. Journal of Ethnopharmacology 146, 214–224.

Jordan, S.A., Cunningham, D.G., Marles, R.J., 2010. Assessment of herbal medicinal products: challenges, and opportunities to increase the knowledge base for safety assessment. Toxicology and Applied Pharmacology 243, 198–216.

Keim, D.A., 2001. Visual exploration of large data sets. Communications of the ACM 44(8), 38-44.

Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., Blomberg, S.P., Webb, C.O., 2010. Picante: R tools for integrating phylogenies and ecology. Bioinformatics 26, 1463–1464.

Koehn, F.E., Carter, G.T., 2005. The evolving role of natural products in drug discovery. Nature Reviews Drug Discovery 4, 206–220.

Kreft, H., Jetz, W., 2010. A framework for delineating biogeographical regions based on species distributions. Journal of Biogeography 37, 2029–2053.

Krishna, N.R., Saidulu, C., Hindumathi, A., 2014. Ethnomedicinal uses of some plant species by tribal healers in Adilabad district of Telangana state, India. World Journal of Pharmaceutical Research 3, 545–561.

Ladio, A., Lozada, M., Weigandt, M., 2007. Comparison of traditional wild plant knowledge between aboriginal communities inhabiting arid and forest environments in Patagonia, Argentina. Journal of Arid Environments 69, 695–715.

Ladio, A.H., Lozada, M., 2004. Patterns of use and knowledge of wild edible plants in distinct ecological environments: a case study of a Mapuche community from northwestern Patagonia. Biodiversity and Conservation 13, 1153–1173.

Laird, S., Wynberg, R..2008. Access and Benefit-Sharing in Practice: Trends in Partnerships Across Sectors, CBD Technical Series No. 38, Secretariat of the Convention on Biological Diversity.

Lampe, K.F., F., 1986. Dermatitis-producing Anacardiaceae of the Caribbean area. Clinics in Dermatology 4, 171–182.

Lavoie, C., Shah, M.A., Bergeron, A., Villeneuve, P., 2013. Explaining invasiveness from the extent of native range: new insights from plant atlases and herbarium specimens. Diversity and Distributions 19, 98–105.

Leonti, M., 2011. The future is written: Impact of scripts on the cognition, selection, knowledge and transmission of medicinal plant use and its implications for ethnobotany and ethnopharmacology. Journal of Ethnopharmacology, 134(3), 542-555.

Leonti, M., Cabras, S., Eugenia Castellanos, M., Challenger, A., Gertsch, J., Casu, L., 2013. Bioprospecting: evolutionary implications from a post-Olmec pharmacopoeia and the relevance of widespread taxa. Journal of Ethnopharmacology 147, 92–107.

Leonti, M., Sticher, O., Heinrich, M., 2002. Medicinal plants of the Popoluca, México: organoleptic properties as indigenous selection criteria. Journal of Ethnopharmacology 81, 307–15.

Leprieur, F., Albouy, C., De Bortoli, J., Cowman, P.F., Bellwood, D.R., Mouillot, D., 2012. Quantifying phylogenetic beta diversity: distinguishing between "true" turnover of lineages and phylogenetic diversity gradients. PLoS ONE 7, e42760.

Letunic, I., & Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic acids research, 44(W1), W242-W245.

Lewis, G., Schrire, B., Mackinder, B., Lock, M. (Eds.), 2005. Legumes of the World. Royal Botanic Gardens, Kew.

Lira, R., Caballero, J., 2002. Ethnobotany of the Mexican Curcubitaceae. Economic Botany 56, 380–398.

Little, P.E., 2002. Territórios sociais e povos tradicionais no Crasil: por uma antropologia da territorialidade. Serie Antropologia 322.

Liu, C.H., Wu, D.-Y., Pollock, J.D., 2012. Bioinformatic challenges of big data in non-coding RNA research. Frontiers in Genetics 3, 178.

Lorenzi, H., Abreu Matos, F.J., 2002. Plantas Medicinais do Brasil Nativas e Exóticas. Instituto Plantarum de Estudos da Flora LTDA, Nova Odessa.

Loub, W.D., Farnsworth, N.R., Soejarto, D.D., Quinn, M.L., 1985. NAPRALERT: computer handling of natural product research data. Journal of Chemical Information and Modeling 25, 99–103.

Lozada, M., Ladio, A., Weigandt, M., 2006. Cultural transmission of ethnobotanical knowledge in a rural community of northwestern Patagonia - Argentina. Economic Botany 60, 374–385.

Lucena, R.F.P. de, Araujo, E. de L., de Albuquerque, U.P., 2007. Does the local availability of woody Caatinga plants (northeastern Brazil) explain their use value? Economic Botany 61, 347–361.

Lukhoba, C.W., Simmonds, M.S.J., Paton, A.J., 2006. *Plectranthus*: a review of ethnobotanical uses. Journal of Ethnopharmacology 103, 1–24.

Manhã, E.M., Silva, M.C., Alves, M.G.C., Almeida, M.B., Brandão, M.G.L., 2008. PLANT - A bibliographic database about medicinal plants. Revista Brasileira de Farmacognosia 18, 614–617.

Maridass, M., De Britto, A.J., 2008. Origins of plant derived medicines. Ethnobotanical Leaflets 12, 373–387.

Martin, G.J., 2004. Ethnobotany: a methods manual. Earthscan, London.

Mazel, F., Davies, T.J., Gallien, L., Renaud, J., Groussin, M., Münkemüller, T., Thuiller, W., 2016. Influence of tree shape and evolutionary time-scale on phylogenetic diversity metrics. Ecography 39, 913–920.

McChesney, J.D., Venkataraman, S.K., Henri, J.T., 2007. Plant natural products: back to the future or into extinction? Phytochemistry 68, 2015–22.

McKenna, D.J., Ruiz, J.M., Hoye, T.R., Roth, B.L., Shoemaker, A.T., 2011. Receptor screening technologies in the evaluation of Amazonian ethnomedicines with potential applications to cognitive deficits. Journal of Ethnopharmacology 134, 475–492.

Mendonça, R. C., Felfili.; Walter, B. M. T.; Silva Junior, M. C.; Rezende, A. B.; Filgueiras, T. S.; Nogueira, P. E. & Fagg, C. W. 2008. Flora vascular do bioma Cerrado: checklist com 12.356 espécies, in: S. M. Sano; S. P. Almeida & J. F. Ribeiro (Eds.). Cerrado: Ecologia e Flora. v2. Embrapa Cerrados/Embrapa Informação Tecnológica. Brasília, p. 421-1279

Merali, Z., Giles, J., 2005. Databases in peril. Nature 435, 1010–1011.

Merlin, M.D., 2003. Archaeological evidence for the tradition of psychoactive plant use in the Old World. Economic Botany 57, 295–323.

Moerman, D.E., 1991. The medicinal flora of native North America: an analysis. Journal of Ethnopharmacology 31, 1–42.

Moerman, D.E., Pemberton, R.W., Kiefer, D., Berlin, B., 1999. A comparative analysis of five medicinal floras. Journal of Ethnobiology 19, 49–67.

Moerman, D.E. 2013. The Global Flora: Descriptive statistics with a commentary and an ethnobotanical example. Ethnobotany Research & Applications, 11, 109-119.

Monteiro, J.M., Albuquerque, U.P. de, Lins-Neto, E.M. de F., Araújo, E.L. de, Amorim, E.L.C. de, 2006. Use patterns and knowledge of medicinal species among two rural communities in Brazil's semi-arid northeastern region. Journal of Ethnopharmacology 105, 173–186.

Moreira, D. de L., Teixeira, S.S., Monteiro, M.H.D., De-Oliveira, A.C.A.X., Paumgartten, F.J.R., 2014. Traditional use and safety of herbal medicines. Revista Brasileira de Farmacognosia 24, 248–257.

Muellner, A.N., Samuel, R., Chase, M.W., Pannell, C.M., Greger, H., 2005. *Aglaia* (Meliaceae): an evaluation of taxonomic concepts based on DNA data and secondary metabolites. American Journal of Botany 92, 534–43.

Naranjo, P., 1979. Hallucinogenic plant use and related indigenous belief systems in the Ecuadorian Amazon. Journal of Ethnopharmacology 1, 121–145.

Nesbitt, M., 2014. Use of herbarium specimens in ethnobotany, in: Salick, J., Konchar, K., Nesbitt, M. (Eds.), Curating Biocultural Collections. Royal Botanic Gardens, Kew Publishing, p. 313–328.

Ningthoujam, S. S., Talukdar, A. D., Potsangbam, K. S., & Choudhury, M. D., 2012. Challenges in developing medicinal plant databases for sharing ethnopharmacological knowledge. Journal of Ethnopharmacology, 141(1), 9-32.

Noe N, Lehmann J (2012). Prelude Medicinal Plants Database. Belgian Biodiversity Platform. Checklist Dataset.

Nogueira-Ferreira, R., Vitorino, R., Ferreira, R., Henriques-Coelho, T., 2015. Exploring the monocrotaline animal model for the study of pulmonary arterial hypertension: a network approach. Pulmonary Pharmacology & Therapeutics 35, 8–16.

Nogueira, R.C., de Cerqueira, H.F., Soares, M.B.P., 2010. Patenting bioactive molecules from biodiversity: the Brazilian experience. Expert Opinion on Therapeutic Patents 20, 145–57.

Oliveira, A.E.A., Gomes, V.M., Sales, M.P., Fernandes, K.V.S., Carlini, C.R., Xavier-Filho, J., 1999. The toxicity of jack bean [*Canavalia ensiformis* (L.) DC.] canatoxin to plant pathogenic fungi. Revista Brasileira de Biologia 59, 59–62.

Pagel, M., Mace, R., 2004. The cultural wealth of nations. Nature 428, 275–278.

Paradis, E., Claude, J., Strimmer, K., 2004. APE: Analyses of phylogenetics and evolution in R language. Bioinformatics 20, 289–290.

Petrovska, B.B., 2012. Historical review of medicinal plants' usage. Pharmacognosy Reviews 6, 1–5.

Phillips, O., Gentry, A.H., 1993. The useful plants of Tambopata, Peru. Statistical hypotheses tests with a new quantitative technique. Economic Botany 47(1), 15-32.

Pillar, V.D., Duarte, L. D. S., 2010. A framework for metacommunity analysis of phylogenetic structure. Ecology Letters 13, 587–596.

Prakash, R., 2011. Documentation of ethnobotanical notes from herbarium specimens at the Natural History Museum, London (UK). Ethnobotany 23, 21–26.

Prance, G.T., 1970. Notes on the use of plant hallucinogens in Amazonian Brazil. Economic Botany 24, 62–68.

Qian, H., Jin, Y., 2016. An updated megaphylogeny of plants, a tool for generating plant phylogenies and an analysis of phylogenetic community structure. Journal of Plant Ecology 9, 233–239.

QGIS Development Team, 2009. QGIS Geographic Information System. Open Source Geospatial Foundation. Available from: http://qgis.osgeo.org

R Core Team, 2015. R: a language and environment for statistical computing.

Reddy, S. 2006. Making heritage legible: Who owns traditional medical knowledge?. International journal of cultural property, 13(2), 161-188.

Reiser, L., Berardini, T.Z., Li, D., Muller, R., Strait, E.M., Li, Q., Mezheritsky, Y., Vetushko, A., Huala, E., 2016. Sustainable funding for biocuration: The Arabidopsis Information Resource (TAIR) as a case study of a subscription-based funding model. Database, 2016, baw018

Revell, L.J., 2012. Phytools: a R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution 3, 217–223.

Ritter, M.R., Cristina Da Silva, T., De, E., Araújo, L., Albuquerque, U.P., 2015. Bibliometric analysis of ethnobotanical research in Brazil (1988–2013). Acta Botanica Brasilica 29, 113–119.

Rodrigues, E., Carlini, E.A., 2006. A comparison of plants utilized in ritual healing by two Brazilian cultures: Quilombolas and Krahô indians. Journal of Psychoactive Drugs 38, 37–41.

Rønsted, N., Savolainen, V., Mølgaard, P., Jäger, A.K., 2008. Phylogenetic selection of *Narcissus* species for drug discovery. Biochemical Systematics and Ecology 36, 417–422.

Saslis-Lagoudakis, C.H., Hawkins, J.A., Greenhill, S.J., Pendry, C.A., Watson, M.F., Tuladhar-douglas, W., Baral, S.R., Savolainen, V., 2014. The evolution of traditional knowledge: environment shapes medicinal plant use in Nepal. Procedures of the Royal Society B 281, 20132768.

Saslis-Lagoudakis, C.H., Klitgaard, B.B., Forest, F., Francis, L., Savolainen, V., Williamson, E.M., Hawkins, J.A., 2011a. The use of phylogeny to interpret cross-cultural patterns in plant use and guide medicinal plant discovery: an example from *Pterocarpus* (Leguminosae). PloS one 6, e22275.

Saslis-Lagoudakis, C.H., Rønsted, N., Clarke, A.C., Hawkins, J.A., 2015. Evolutionary approaches to ethnobiology, in: de Albuquerque, U.P., de Medeiros, P.M., Casas, A. (Eds.), Evolutionary Ethnobiology. Springer, Recife, p. 59–72.

Saslis-Lagoudakis, C.H., Savolainen, V., Williamson, E.M., Forest, F., Wagstaff, S.J., Baral, S.R., Watson, M.F., Pendry, C.A., Hawkins, J.A., 2012a. Phylogenies reveal predictive power of traditional medicine in bioprospecting. Proceedings of the National Academy of Sciences of the United States of America 109, 15835–40.

Saslis-Lagoudakis, C.H., Williamson, E.M., Savolainen, V., Hawkins, J.A., 2011b. Cross-cultural comparison of three medicinal floras and implications for bioprospecting strategies. Journal of Ethnopharmacology 135, 476–487.

Saslis-Lagoudakis, C.H., Williamson, E.M., Savolainen, V., Hawkins, J.A., 2012b. Cross-cultural comparisons of medicinal floras and bioprospecting revisited. Journal of Ethnopharmacology 139, 688–690.

Schalk P.H., and P. Oosterbroek. 1996. Interactive knowledge systems: Meeting the demand for disseminating up to-date biological information. Biodiversity Letters 3:119-123.

Schiebinger, L.L., 2004. Plants and empire: Colonial bioprospecting in the Atlantic World. Harvard University Press, Cambridge.

Schippmann, U., Leaman, D., Cunningham, A.B., 2006. A comparison of cultivation and wild collection of medicinal and aromatic plants under sustainability aspects. Frontis 17, 75–95.

Schippmann, U., Leaman, D.J., Cunningham, A.B., 2002. Impact of cultivation and gathering of medicinal plants on biodiversity: global trends and issues. Biodiversity and the ecosystem approach in agriculture, forestry and fisheries.

Schultes, R.E., 1960. Tapping our heritage of ethnobotanical lore. Economic Botany 14, 257–262.

Seebaluck, R., Gurib-Fakim, A., Mahomoodally, F., 2015. Medicinal plants from the genus *Acalypha* (Euphorbiaceae) - a review of their ethnopharmacology and phytochemistry. Journal of Ethnopharmacology 159, 137–57.

Senchina, D.., 2006. Utilizing herbaria in medical botany curricula. Vulpia 5, 1–13.

Shanley, P., Luz, L., 2003. The impacts of forest degradation on medicinal plant use and implications for health care in eastern Amazonia. BioScience 53, 573.

Shaw, J., 2017. The role of biodiversity centres in bioprospecting: a case study from Sarawak, in Patterson, R, Lima, N (Eds.), Bioprospecting. Springer International Publishing, p. 295–298.

Shepard Jr, G.H., 2004. A sensory ecology of medicinal plant therapy in two Amazonian societies. American Anthropologist 106, 252–266.

Shinde, R., Prakash, R., 2015. Phenological and ethnobotanical data extraction from herbarium specimens: a case study on Mumbai and Thane using Blatter herbarium. Xplore- The Xavier's Research Journal 6, 1–23.

Soejarto, D.D.D., Fong, H.H.S.H.S., Tan, G.T.T., Zhang, H.J.J., Ma, C.Y.Y., Franzblau, S.G.G., Gyllenhaal, C., Riley, M.C.C., Kadushin, M.R.R., Pezzuto, J.M.M., Xuan, L.T.T., Hiep, N.T.T., Hung, N.V. V, Vu, B.M.M., Loc, P.K.K., Dac, L.X.X., Binh, L.T.T., Chien, N.Q.Q., Hai, N.V. V, Bich, T.Q.Q., Cuong, N.M.M., Southavong, B., Sydara, K., Bouamanivong, S., Ly, H.M.M., Thuy, T. Van, Rose, W.C.C., Dietzman, G.R.R., 2005. Ethnobotany/ethnopharmacology and mass bioprospecting: issues on intellectual property and benefit-sharing. Journal of Ethnopharmacology 100, 15–22.

Sõukand, R., Kalle, R., 2010. Herbal landscape: the perception of landscape as a source of medicinal plants. Trames 14, 207–226.

Souza-Neto, A.C., Cianciaruso, M. V., Collevatti, R.G., 2016. Habitat shifts shaping the diversity of a biodiversity hotspot through time: insights from the phylogenetic structure of Caesalpinioideae in the Brazilian Cerrado. Journal of Biogeography 43, 340–350.

Souza, E.N.F., Hawkins, J.A., 2017. Comparison of herbarium label data and published medicinal use: herbaria as an underutilized source of ethnobotanical information. Economic Botany 71, 1–12.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.

Staub, P.O., Geck, M.S., Weckerle, C.S., Casu, L., Leonti, M., 2015. Classifying diseases and remedies in ethnomedicine and ethnopharmacology. Journal of Ethnopharmacology 174, 514–519.

Stepp, J., Thomas, M., 2010. Managing ethnopharmacological data: herbaria, relational databases, literature. Medical and Health Sciences 13.

Stepp, J.R., Moerman, D.E., 2001. The importance of weeds in ethnopharmacology. Journal of Ethnopharmacology 75, 19–23.

Swenson, N.G., Gaston, K., Ives, A., Thompson, J., Uriarte, M., 2011. Phylogenetic beta diversity metrics, trait evolution and inferring the functional beta diversity of communities. PLoS ONE 6, e21264.

Tan, G., Gyllenhaal, C., Soejarto, D., 2006. Biodiversity as a source of anticancer drugs. Current Drug Targets 7, 265–277.

Tegelberg, R., Mononen, T., Saarenmaa, H., 2014. High-performance digitization of natural history collections: automated imaging lines for herbarium and insect specimens. Taxon 63, 1307–1313.

The Legume Phylogeny Working Group, 2013. Legume phylogeny and classification in the 21st century: progress, prospects and lessons for other species-rich clades. Taxon 62, 217–248.

The Plant List, 2013.Version 1.1. available from: URL http://www.theplantlist.org/

Thessen, A.E., Patterson, D.J., 2011. Data issues in the life sciences. ZooKeys 15–51.

Thiers, B., 2014. Index Herbariorum: a global directory of public herbaria and associated staff. New York Botanical Garden's Virtual Herbarium.

Thiers, B.M., Tulig, M.C., Watson, K.A., 2016. Digitization of the New York Botanical Garden herbarium. Brittonia 68, 324–333.

Thomas, M.B., 2003. Emerging synergies between information technology and applied ethnobotanical research. Ethnobotany Research & Applications 1, 65–73.

Thomas, M. B., Lin, N., & Beck, H. H., 2001. A database Model for integrating and facilitating collaborative ethnomedicinal Research. Pharmaceutical biology, 39(sup1), 41-52.

Tropicos.org, 2016. Missouri Botanical Garden. Available from: http://www.tropicos.org

Ugent, D., 2000. Medicine, myths and magic the folk healers of a Mexican market. Economic Botany 54, 427–438.

United Nations, 2010. Nagoya protocol on Access to genetic resources and the fair and equitable sharing of benefits arising from their utilization to the Convention on Biological Diversity. Secretariat of the Convention on Biological Diversity, Montreal.

Uprety, Y., Asselin, H., Dhakal, A., Julien, N., 2012. Traditional use of medicinal plants in the boreal forest of Canada: review and perspectives. Journal of Ethnobiology and Ethnomedicine 8, 7.

van Andel, T., van't Klooster, C., 2007. Medicinal Plant Use by Surinamese Immigrants in Amsterdam, the Netherlands: Results of a Pilot Market Survey, in: Pieroni, A., Vandebroek, I. (Eds.), Traveling Cultures and Plants, the Ehnobiology and Ethnopharmacy of Human Migrations. Berghahn Books, New York, p. 122–144.

van Andel, T., Veldman, S., Maas, P., Thijsse, G., Eurlings, M., 2012. The forgotten Hermann herbarium: a 17th century collection of useful plants from Suriname. Taxon 61, 1296–1304.

van Andel, T.R., van 't Klooster, C.I.E.A., Quiroz, D., Towns, A.M., Ruysschaert, S., van den Berg, M., 2014. Local plant names reveal that enslaved Africans recognized substantial parts of the New World flora. Proceedings of the National Academy of Sciences of the United States of America 111, E5346-53.

van Oever, J.P., Gofferjé, M., 2012. "From pilot to production": large scale digitisation project at Naturalis Biodiversity Center. ZooKeys 87–92.

Vandebroek, I., Balick, M.J., Yukes, J., Duran, L., Kronenberg, F., Wade, C., Ososki, A.L., Cushman, L., Lantigua, R., Mejia, M., Robineau, L., 2007. Use of medicinal plants by Dominican immigrants in New York City for the treatment of common health conditions: a comparative analysis with literature data from the Dominican Republic, in: Pieroni, A., Vandebroek, I. (Eds.), Traveling Cultures and Plants, the Ehnobiology and Ethnopharmacy of Human Migrations. Berghahn Books, New York, p. 39–63.

Vandebroek, I., Pieroni, A., 2007. Traveling Cultures and Plants, the Ethnobiology and Ethnopharmacy of Human Migrations. Berghahn Books, New York, NY.

Vandebroek, I., Van Damme, P., Van Puyvelde, L., Arrazola, S., De Kimpe, N., 2004. A comparison of traditional healers' medicinal plant knowledge in the Bolivian Andes and Amazon. Social Science and Medicine 59, 837–849.

Vickery, A.., 1990. Ethnobotanical notes in the herbarium of Natural History Museum: 1- Rubiaceae. Ethnobotany 2, 25–30.

Voeks, R., 2013. Ethnobotany of Brazil's African diaspora: the role of floristic homogenization, in: African ethnobotany in the Americas. Springer New York, New York, p. 395–416.

Voeks, R.A., Leony, A., 2004. Forgetting the forest: assessing medicinal plant erosion in eastern brazil 58(sp1), S294-S306.

Voeks, R. A., 1996. Tropical forest healers and habitat preference. Economic Botany 50, 381–400.

Voeks, R., Rashford, J., 2013. African Ethnobotany in the Americas. Springer, New York.

Volpato, G., Godínez, D., Beyra, A., 2008. Migration and ethnobotanical practices: the case of tifey among Haitian immigrants in Cuba. Human Ecology 37, 43–53.

von Altschul, S., 1968. Unusual food plants in herbarium records. Economic Botany 22, 293–296.

Von Reis, S., 1962. Herbaria: sources of medicinal folklore. Economic Botany 16, 283–287.

Wang, D., Huang, L., Chen, S., 2013. *Senecio scandens* Buch.-Ham.: a review on its ethnopharmacology, phytochemistry, pharmacology, and toxicity. Journal of Ethnopharmacology 149, 1–23.

Wayland, C., Walker, L.S., 2014. Length of residence, age and patterns of medicinal plant knowledge and use among women in the urban Amazon. Journal of Ethnobiology and Ethnomedicine 10(1), 25.

Webb, C.O., 2000. Exploring the phylogenetic structure of ecological communities: an example for rain forest trees. The American Naturalist 156, 145–155.

Webb, C.O., Ackerly, D.D., Kembel, S.W., 2008. Phylocom: software for the analysis of phylogenetic community structure and trait evolution. Bioinformatics Applications Note 24, 2098–2100.

Webb, C.O., Ackerly, D.D., McPeek, M. A., Donoghue, M.J., 2002. Phylogenies and community ecology. Annual Review of Ecology and Systematics 33, 475–505.

Webb, C.O., Donoghue, Mi.J., 2005. Phylomatic: tree assembly for applied phylogenetics. Molecular Ecology Notes 5, 181–183.

Weckerle, C.S., Cabras, S., Castellanos, M.E., Leonti, M., 2011. Quantitative methods in ethnobotany and ethnopharmacology: considering the overall flora-hypothesis testing for over- and underused plant families with the Bayesian approach. Journal of Ethnopharmacology 137, 837–43.

Whittaker, R.H., 1972. Evolution and measurement of species diversity. Taxon 213, 213–251.

Wieczorek, J., et al. 2012. Darwin Core: an evolving community-developed biodiversity data standard. PLoS ONE 7, e29715.

Wink, M., 2013. Evolution of secondary metabolites in legumes (Fabaceae). South African Journal of Botany 89, 164–175.

Wink, M., 2003. Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. Phytochemistry 64, 3–19.

Wink, M., Mohamed, G.I.A., 2003. Evolution of chemical defense traits in the Leguminosae: mapping of distribution patterns of secondary metabolites on a molecular phylogeny inferred from nucleotide sequences of the rbcL gene. Biochemical Systematics and Ecology 31, 897–917.

World Health Organization, 2007. Health of indigenous peoples. Factsheets No 326, WHO. World Health Organization, Geneva, Switzerland.

World Health Organization International Classification of Diseases 10. Available from: URL http://www.who.int/classifications/icd/en

Yessoufou, K., Daru, B.H., Muasya, A.M., 2015. Phylogenetic exploration of commonly used medicinal plants in South Africa. Molecular Ecology Resources 15, 405–413.

Zanne, A.E., Tank, D.C., Cornwell, W.K., Eastman, J.M., Smith, S.A., FitzJohn, R.G., McGlinn, D.J., O'Meara, B.C., Moles, A.T., Reich, P.B., Royer, D.L., Soltis, D.E., Stevens, P.F., Westoby, M., Wright, I.J., Aarssen, L., Bertin, R.I., Calaminus, A., Govaerts, R., Hemmings, F., Leishman, M.R., Oleksyn, J., Soltis, P.S., Swenson, N.G., Warman, L., Beaulieu, J.M., 2014. Three keys to the radiation of angiosperms into freezing environments. Nature 506, 89–92.

Zhu, F., Qin, C., Tao, L., Liu, X., Shi, Z., Ma, X., Jia, J., Tan, Y., Cui, C., Lin, J., Tan, C., Jiang, Y., Chen, Y., 2011. Clustered patterns of species origins of nature-derived drugs and clues for future bioprospecting. Proceedings of the National Academy of Sciences of the United States of America 108, 12943–8.

Appendices

Appendices 1 to 6 are provided in digital format, and appendices 7 to 11 are provided here.

Appendix 1. Comparison of herbarium label and published medicinal use: herbaria as an underutilized source of ethnobotanical use (publication).

Appendix 2. Species, therapeutic applications, plants parts used, modes of application and source for reports of medicinal use of Legumes in Brazil.

Appendix 3. List of references and herbaria sourced for Ewé.

Appendix 4. Ewé data to date.

Appendix 5. Accession numbers for the Brazilian Leguminosae phylogeny.

Appendix 6. List of medicinally used Brazilian Leguminosae with pharmacological studies.

Appendix 7. Table of possible outcomes of results from phylogenetic frameworks in ethnobotanical investigation.

Appendix 8. Generic level taxonomic beta diversity

Appendix 9. PCPS values and eigenvectors

Appendix 10. PCPS scatterplot of the medicinal floras of the Brazilian Biomes, showing the relation of biomes according to the first two principal components

Appendix 11. Spearman rank order correlation for therapeutic application between each pair of biomes in Brazil. Numbers in bold =p<0.05.

Appendix 7. Table of possible outcomes of results from phylogenetic frameworks in ethnobotanical investigation

Comparison	Finding	Interpretation
Taxonomic beta diversity of medicinal plants at species level (TaxβMEDsp) v. Taxonomic beta diversity of the whole flora at species level (TaxβFLORsp)	TaxβMEDsp > TaxβFLORsp	Different species are used in different places, with more turnover than expected, given floristic turn-over; suggests high local adaptation/low transmission of knowledge in space
Taxonomic beta diversity of medicinal plants at species level (TaxβMEDsp) v. Taxonomic beta diversity of the whole flora at species level (TaxβFLORsp)	TaxβMEDsp < TaxβFLORsp	There is more shared use of species than would be expected given turn-over in species availability; suggests high transmission of knowledge, or preferential use of widespread species
Taxonomic beta diversity of medicinal plants at generic level (TaxβMEDgen) v. Taxonomic beta diversity of the whole flora at generic level (TaxβFLORgen)	TaxβMEDgen > TaxβFLORgen	Different genera are used in different places, with more turnover than expected, given floristic turn-over; suggests low transmission of knowledge in space
Taxonomic beta diversity of medicinal plants at generic level (TaxβMEDgen) v. Taxonomic beta diversity of the whole flora at generic level (TaxβFLORgen)	TaxβMEDgen < TaxβFLORgen	There is more shared use of genera than would be expected given turn-over in species availability; suggests high transmission of knowledge, or preferential use of widespread genera
Taxonomic beta diversity - Species level comparison v. Generic level comparison	(TaxβMEDsp > TaxβFLORsp) AND (TaxβMEDgen > TaxβFLORgen)	Diversity of genera and species high; suggests signal of species replacement within genera is not strong
Taxonomic beta diversity - Species level comparison v. Generic level comparison	(TaxβMEDsp > TaxβFLORsp) AND (TaxβMEDgen < TaxβFLORgen)	Diversity of genera low despite diversity of species; suggests species replacement within genera
Taxonomic beta diversity - Species level comparison v. Generic level comparison	(TaxβMEDsp < TaxβFLORsp) AND (TaxβMEDgen > TaxβFLORgen)	Not expected – high retention of species but turnover in genera

Taxonomic beta diversity - Species level comparison v. Generic level comparison	(TaxβMEDsp < TaxβFLORsp) AND (TaxβMEDgen < TaxβFLORgen)	Conservation of species use and use of genera; suggests use of widespread species and genera, and low local adaptation
Taxonomic beta diversity of medicinal plants at species level (TaxβMEDsp) v. Phylogenetic beta diversity of medicinal plants at species level (PhyloβMEDsp)	(TaxβMEDsp > PhyloβMEDsp)	Turn-over in use of species is high, but the species used are in shared lineages; suggests phylogenetic replacement
Taxonomic beta diversity of medicinal plants at species level (TaxβMEDsp) v. Phylogenetic beta diversity of the whole flora at species level (PhyloβMEDsp)	(TaxβMEDsp < PhyloβMEDsp)	Not expected – high retention of species but turnover in lineages
Taxonomic beta diversity of medicinal plants at generic level (TaxβMEDgeneric) v. Phylogenetic beta diversity of the whole flora at generic level (PhyloβMEDgen)	(TaxβMEDgen > PhyloβMEDgen)	Turn-over in use of genera is high, but the genera used as in shared lineages; suggests preferences of lineages above generic level, e.g. preferences for families
Taxonomic beta diversity of medicinal plants at generic level (TaxβMEDgeneric) v. Phylogenetic beta diversity of the whole flora at generic level (PhyloβMEDgen)	(TaxβMEDgen < PhyloβMEDgen)	Not expected – high retention of genera but turnover in lineages
Interpretation of statistics for ethnobotanical data		
Statistic	**Finding**	**Interpretation**
Phylogenetic structure of medicinal plants at shallow phylogenetic nodes using species/generic-level phylogeny (MED_NTIsp/MED_NTIgen)	+ve significant	Use of species/genera more closely-related to each other than expected from a random sample
Phylogenetic structure of medicinal plants at shallow phylogenetic nodes using species/generic-level phylogeny (MED_NTIsp/MED_NTIgen)	-ve significant	Use of species/genera more distantly related to each other than expected from a random sample
Phylogenetic structure of medicinal plants at deep phylogenetic nodes using species/generic-level phylogeny (MED_NRIsp/MED_NRIgen)	+ve significant	Use of species/genera belonging to the same deep lineages is greater than expected from a random sample
Phylogenetic structure of medicinal plants at deep phylogenetic	-ve significant	Use of species/genera belonging to different deep lineages is greater than

143

nodes using species/generic-level phylogeny (MED_NRIsp/MED_NRIgen)		expected from a random sample
Phylogenetic beta diversity of medicinal plants using NTI statistic and species/generic-level phylogeny (βMED_NTIsp/βMED_NTIgen)	+ve significant	Turn-over in species/genera is between closely-related species/genera; suggests "phylogenetic replacement"
Phylogenetic beta diversity of medicinal plants using NTI statistic and species/generic-level phylogeny (βMED_NTIsp/βMED_NTIgen)	-ve significant	Turn-over in species/genera is between distantly-related species/genera; suggestive of a driver to differentiate
Phylogenetic beta diversity of medicinal plants using NTI statistic and species/generic-level phylogeny (βMED_NRIsp/βMED_NRIgen)	+ve significant	Turn-over in species/genera is between species/genera in the same deep lineages; suggests preference of use for the same deep lineages
Phylogenetic beta diversity of medicinal plants using NTI statistic and species/generic-level phylogeny (βMED_NRIsp/βMED_NRIgen)	-ve significant	Turn-over in species/genera is between lineages; suggestive of a driver to differentiate

Appendix 8. Dissimilarity matrices at generic level for medicinal plant use (lower diagonal) and total flora (upper diagonal) in the Brazilian Biomes. in the Brazilian Biomes. Green color indicates a smaller difference in the medicinal flora than in the total flora. Red color indicates a larger difference.

B	Amazon	Atlantic	Caatinga	Cerrado	Pampa	Pantanal
Amazon		0.33	0.23	0.28	0.44	0.15
Atlantic	0.37		0.13	0.15	0.08	0.09
Caatinga	0.54	0.37		0.13	0.39	0.20
Cerrado	0.37	0.40	0.37		0.27	0.10
Pampa	0.58	0.26	0.55	0.26		0.54
Pantanal	0.54	0.31	0.43	0.20	0.66	

Appendix 9. PCPS values and eigenvectors for the PCPS analysis of the medicinally-used species of the Brazilian flora

A. Principal components for Brazilian Biomes as resulted from the PCPS analysis.

	pcps.1	pcps.2	pcps.3	pcps.4	pcps.5
Amazon	-0.013242374	0.04135311	0.001479729	-0.02800117	3.13E-02
Atlantic	0.009668027	0.04360697	0.044060126	0.04287837	-4.28E-05
Caatinga	-0.062669653	-0.01937558	-0.068328543	0.02326362	2.61E-03
Cerrado	-0.01584214	0.04526533	-0.007304452	-0.0253052	-3.14E-02
Pampas	-0.057178105	-0.07669566	0.045790136	-0.01048259	-2.18E-03
Pantanal	0.139264245	-0.03415416	-0.015696997	-0.00235302	-3.11E-04

B. Eigenvalues for the PCPS analysis on the medicinal flora of Brazilian Biomes

PCPS	Eigenvalues	Relative_eig	Cumulative_eig
1	0.027111156	0.49203888	0.4920389
2	0.013084742	0.23747427	0.7295131
3	0.009008761	0.16349952	0.8930127
4	0.003919591	0.07113643	0.9641491
5	0.001975371	0.03585091	1

Appendix 10. Principal components of phylogenetic structure for the medicinal floras of the Brazilian Biomes, showing the relation of biomes according to the first two principal components
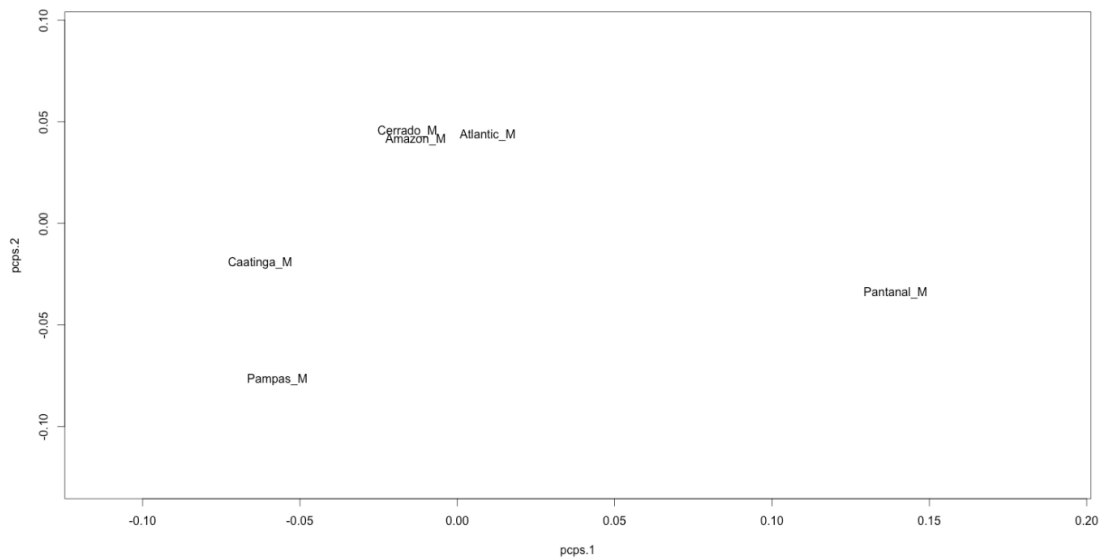
Appendix 11. Spearman rank order correlation for therapeutic application between each pair of biomes in Brazil. Numbers in bold =p<0.05.

Biomes	Amazon	Atlantic	Caatinga	Cerrado	Pampa
Atlantic	**0.73**				
Caatinga	**0.84**	**0.84**			
Cerrado	**0.77**	**0.92**	**0.81**		
Pampas	0.47	**0.63**	0.40	**0.56**	
Pantanal	**0.73**	**0.60**	**0.62**	**0.64**	**0.55**