Université de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

Présentée et soutenue par :

Jan HACKEL

Le 13 décembre 2017

Titre :

Molecular biogeography of grasses and tropical grasslands

ED SEVAB : Écologie, biodiversité et évolution

Unité de recherche :

UMR 5174 - Laboratoire Evolution & Diversité Biologique

Directeur(s) de Thèse :

Guillaume BESNARD

Rapporteurs :

Alex BAUMEL, Aix-Marseille Université
Yves VIGOUROUX, IRD Montpellier

Autre(s) membre(s) du jury :

Peter LINDER, Universität Zürich
Monique GARDES, UT3 Paul Sabatier

Molecular biogeography of grasses and tropical grasslands

Biogéographie moléculaire des graminées et des savanes tropicales

Jan Hackel

13 December 2017

Doctoral dissertation

Thèse de doctorat

Université Toulouse III – Paul Sabatier

Laboratoire Evolution et Diversité Biologique

Supervisor/*Directeur de thèse*:

Guillaume Besnard

Examination board/*Jury de thèse*:

Monique Gardes, Université Toulouse III – Paul Sabatier

Alex Baumel, Aix-Marseille Université (*rapporteur*)

Peter Linder, Universität Zürich

Yves Vigouroux, IRD Montpellier (*rapporteur*)

Acknowledgements

I would like to thank a number of people for accompanying me through these last three years. First of all, Guillaume Besnard was a great supervisor, always available for feedback, with this intuition for the curious details, and we spent hours in the afternoon heat of Madagascar sterilising grass leaves. Maria Vorontsova was involved in all parts of this dissertation. Pilar Catalán gave me the opportunity to participate in her and Miguel Minaya's work on Loliinae.

Colleagues at EDB made these three years a great time scientifically and personally. Sophie Manzi and Amaia Iribar-Pelozuelo helped me with lab work issues. Hervé Gryta was member of my PhD committee and had useful suggestions on several occasions. Lucie Zinger discussed metabarcoding and helped finding those little errors in my scripts. Patricia Jargeat provided fungal isolates. Christophe Thébaud lent me useful books on Madagascar. Antoine Fouquet critically read an article manuscript. The lab administration and the doctoral school made everything run smoothly. My office colleagues, fellow PhD students, and all those who went running, swimming or hiking with me helped me stay balanced and have a good time. Alexandre Meunier was a fast-learning MSc intern. And not to forget all the students who had to endure my first steps in teaching!

Several people outside Toulouse gave valuable input. P-A Christin, member of my PhD committee, was always available for questions. Matheus Bianconi shared great results of his parallel analyses on *Jansenella* and Co. Corinne Vacher, also member of my PhD committee, shared her expertise on the leaf microbiome. Several people answered questions or even shared data, including Elizabeth Kellogg, Caroline Strömberg and Jana U'Ren. Comments of anonymous reviewers greatly improved my papers. I also thank the countless contributors of open software, tutorials, Wikipedia articles and other free content on the internet; I believe they make science better and they certainly made my life easier. Finally, I thank Alex Baumel, Yves Vigouroux and Peter Linder for agreeing to examine my dissertation.

The 2016 trip to Madagascar was an incredible experience and I would like to thank all people involved. Special mention to Hélène Ralimanana, Jacqueline Razanatsoa, Cédrique Solofonanatra, Herison Andrianirina and Caroline Lehmann (and of course Maria), and the whole of the Kew Madagascar Conservation Centre in Antananarivo. I hope my little editing on the Flora of Madagascar on Wikipedia, facilitated by KMCC's library, is a small contribution to their work on plant documentation and conservation.

This dissertation was made possible through a contract funded by the French Government and through other public funds, including from two “Labex” (CEBA, ANR-10-LABX-25-01 and TULIP, ANR-10-LABX-0041) and an Idex UNITI student mobility grant for fieldwork in Madagascar (RECAC15EX110). The last two years in Europe and the world reminded me that the freedom of students and academics to move and exchange is far from a given and I am grateful for the opportunity to study and work in France.

My final thanks go to all my loved ones who supported me. Danke, merci!

Table of contents

1 Introduction

1.1 Patterns and processes in biogeography

The spatial distribution of species

Understanding the distribution of organisms on Earth has been a long-standing goal of biology. Why are some regions more diverse than others? Why are some species endemic and others widespread? How have groups with a common ancestor come to colonise widely separated ranges? Questions such as these form the core of the discipline of biogeography (Cox et al. 2016). Studying the geography of earth's biodiversity is an challenging area of fundamental research but has implications that go beyond pure scientific interest. Organisms interact with their physical environment and each other, forming dynamic ecosystems (Tansley 1935) which not least are the basis of human civilisation. In an epoch where humans are modifying the biosphere in unprecedented ways, both through physical alteration and the rearrangement of biotic communities (Waters et al. 2016), it is urgent to understand the fundamental processes behind the spatial distribution of biodiversity.

Biodiversity can be studied at various spatial and temporal scales, from biomes to genes, and from geological epochs to days, and these levels all have their own dynamics and influence each other (Levin 1992). One of the most commonly studied units, both in evolutionary biology and ecology, is the *species*. The existence of species appears an evident fact to most, but defining and delimiting species has been a major challenge in biology. Today, most biologists explicitly or implicitly share a species concept rooted in the theory of evolution, although they may be divided on how to detect and tell apart species in practice (Hey 2006, de Queiroz 2007). A species is seen as an evolutionary independent lineage of organisms that has split from another lineage, a “separately evolving metapopulation lineage” (de Queiroz 2007). This is the concept I also apply in this dissertation.

The distribution of species on the planet is highly uneven, but some generalities do exist. For example, most species in a given locality are rare (Preston 1948), the number of species increases with area in a repeatable manner (McGuinness 1984), and, on a global scale, the tropics have more species than higher latitudes in most groups of organisms (Hillebrand 2004). However, beneath these general observations, there is a diversity of patterns more complicated: the taxonomic composition of species assemblages differs between continents, creating distinct biogeographic regions recognised since more than a century (Sclater 1858, Engler 1879). Physically similar biomes, such as tropical forests, savannahs or high mountains, can be found on different continents, but the species, the biota, that compose them differ (Nelson 1978). Biogeography has traditionally addressed these patterns from two angles: an ecological one, focusing on the current-day, observable mechanisms that shape distribution patterns, and a historical–evolutionary one, studying the impact of past geological events and evolutionary history (Cox et al. 2016). These two have long existed as two parallel

disciplines, with little exchange, but there have been increasing efforts to merge them (Wiens & Donoghue 2004, Ricklefs & Jenkins 2011).

Ecological processes

Thinking in ecology has been strongly influenced by the idea that an organism is adapted to a particular set of environmental factors and other organisms it interacts with, which together form its *niche* (Grinnell 1917, Elton 1927, Hutchinson 1957). Thus, the niche of an organism would determine where it occurs, and the environment would "filter" species adapted to it (Weiher et al. 2011). This is evident in many cases, with e.g. plant species having adaptations to cold, arid, or humid climates, and being unable to survive under opposing conditions. Species interact with each other, and in a given location, different organisms with exactly the same niches are theoretically expected to exclude each other (Hardin 1960). Organisms may however co-exist when they have slightly different requirements and thus partition the niche space (MacArthur & Levins 1967). That different organisms co-exist despite competing for the same resources is evident for example in plant communities, where individuals directly compete for space and light.

The example of individuals in a plant community was also used to formalise an alternative explanation of species co-existence: individuals of different species may be ecologically equivalent (i.e. their niches overlap completely) but can co-exist in a dynamic equilibrium, where random ecological drift constantly opens vacant spaces that are filled by local or newly arriving species (Hubbell 2001). Hubbell's neutral theory highlights the role of new individuals arriving in a community from a regional source, and builds up on other theories that emphasise the movement of individuals and species, notably the theory of island biogeography (MacArthur & Wilson 1967). These theories underline that species assemblages are dynamic and change over time. Linked to this is the growing appreciation that local communities of species are not closed, but part of and exchanging with a regional "pool" of species (Ricklefs 1987, 2004). This regional pool is built up through evolutionary processes, i.e. the origin and disappearance of species. Historical contingency will thus influence both regional and local assemblages (Ricklefs 2004).

Evolutionary processes

Organismal lineages are linked through common ancestry. They emerge through speciation in a given area, can move to another area, and will eventually go extinct. These three processes, *speciation*, *dispersal* and *extinction*, thus fundamentally underlie species diversity and distribution (Wiens 2011). For example, differences in species richness between two regions can be explained by a higher speciation rate, or lower extinction rate, in one of them (Wiens 2011). High endemism in an area, or in a group of organisms, can be due to restricted dispersal and *in situ* speciation, for example on islands (Losos & Ricklefs 2009). One pattern that has received particular attention from evolutionary biogeographers is that of disjunct distributions (Cox et al. 2016): how has an organism, or a group of organisms, come to be present in widely separated areas?

Early biogeographers, influenced by Darwin and Wallace's recent theory of evolution, attempted to explain disjunct distributions with the area-of-origin concept: lineages would have originated in one area, and reached their current distributions through *dispersal* (Nelson 1978). A later arising, opposing view was that of panbiogeography, which explained disjunctions through *vicariance*, the separation of ancient widespread populations (defended mainly by Léon Croizat, e.g. Croizat 1958). The acceptance of plate tectonics in geology appeared to shift support towards this latter idea, providing a mechanism that could explain ancient continental connections (Cox et al. 2016). In the southern hemisphere, it was shown that the now widely separated continents were once joined in one supercontinent, Gondwana, from the Carboniferous until roughly the Cretaceous, around 150 Ma (million years ago); this could explain e.g. similarities of fossil floras from South America, Antarctica, Africa, Australia, and India (McLoughlin 2001). Biotas of continental islands, such as Madagascar or New Zealand, were explained as being predominantly remnants of an ancient Gondwanan fauna and flora that have since evolved in isolation (Leroy 1978, Craw 1989).

With the arrival of molecular phylogenetics and especially divergence time estimation (see next section), it became however increasingly clear that many lineages must have reached their current range through dispersal: many lineage divergences were much younger than the presumed vicariance event (de Queiroz 2005, Renner 2005). In the southern hemisphere, dispersal between formerly connected Gondwanan landmasses was found to have been more frequent than expected, especially in plants (Sanmartín & Ronquist 2004). The biotas of islands like Madagascar and New Zealand are now seen as largely dispersal-assembled (Winkworth et al. 2002, Yoder & Nowak 2006, Buerki et al. 2013). Even for groups with a well-established Gondwanan fossil record, such as the southern beeches (*Nothofagus*), it was suggested that some recent dispersal must have taken place (Cook & Crisp 2005). Some inferred dispersal events covered extremely large distances, such as that of cold-adapted crowberries (*Empetrum*) from the Arctic to the Antarctic (Popp et al. 2011).

It is important to distinguish how the term “dispersal” is applied in ecology and biogeography (Ricklefs & Jenkins 2011). In the ecological sense, dispersal describes the movement of individuals between habitat patches and populations (Matthysen 2012). In the evolutionary–biogeographical sense, which I apply in this section and the remainder of this dissertation, dispersal describes the macroevolutionary event of a lineage moving to a new area of distribution, with or without simultaneous speciation (Cox et al. 2016). Biogeographers usually, and implicitly, include both the actual dispersal event as well as successful establishment (“effective dispersal”, Nathan et al. 2003), without which there would be no lasting evolutionary imprint in the form of a new range or new lineage. Lineage dispersal was long seen as a rare, unlikely and elusive process, conveniently invoked when other explanations failed (e.g. Nelson 1978). Indeed, for plants it has been suggested that specific adaptations for long-distance dispersal may not exist, and that such dispersal events are often the result of rare drift through extreme climatic events (Nathan 2006). Nevertheless, lineage dispersal may have predictable components. This includes the directionality of dispersal vectors, e.g. ocean currents, prevailing winds, or bird migration routes (Gillespie et

al. 2012). Also, dispersal probability is expected to decrease in a predictable manner with distance, although it was suggested to be higher at very long distances than expected under an exponential decrease (Nathan et al. 2003).

Lineage dispersal is moreover not independent from local environmental factors. For plants, it has been shown that ecological niches, and the biomes they occur in, are relatively conserved in evolutionary history (Prinzing 2001, Crisp et al. 2009). This means that lineages will usually be confined to the conditions they evolved in, and a given area will preferentially be colonised by pre-adapted species (Donoghue 2008). However, there are examples where traits evolved multiple times in independent groups and facilitated adaptation, for example fire resistance in savannah trees (Simon et al. 2009, Maurin et al. 2014) or C_4 photosynthesis in grasses and other plants (see section 1.3). In the south-temperate danthonioid grasses, it was shown that climatic niches, i.e. cold tolerance, evolved among clades, specifically following dispersal events (Humphreys & Linder 2013, Wüest et al. 2015). Then again, an opinion paper suggested that the evolvability of traits is itself often conditioned by a clade's evolutionary and biogeographic history (Edwards & Donoghue 2013). The authors concluded that the ease of dispersal versus niche evolution must be studied considering the context of each particular group. Accordingly, an extensive survey of a relatively young tropical mountain has found recently that both adaptive shifts of lowland lineages as well as long-distance immigration of pre-adapted mountain lineages contributed to its biota (Merckx et al. 2015).

1.2 Molecular and analytical approaches to biogeography

Molecular phylogenetics

Phylogenetics, as the study of the relationships between organisms, derives directly from the theory of evolution (Delsuc et al. 2005). It uses the state of characters inherited from a common ancestor – homologous characters – to make hypotheses about branching order and evolutionary distance. Due to its large amount of characters, DNA is particularly useful for this (Delsuc et al. 2005). Typically, selected portions of genomes – e.g. particular genes or intergenic regions – are used to detect, through an *alignment*, homologous sites that differ and thus indicate an evolutionary change, i.e. a mutation (Anisimova et al. 2013). For a few years already, larger portions or even whole genomes have been used, marking the transition towards *phylogenomics* (Delsuc et al. 2005). This has been facilitated by the development of next-generation sequencing methods, which allow generating large amounts of sequences in parallel for a single sample (Metzker 2010). In plants for example, whole chloroplast genomes are assembled with relative ease even from short sequence reads due to their conserved structure and high copy number, and have increasingly been used to resolve relationships (Tonti-Filippini et al. 2017).

The possibility of using different parts of the genome also allows to detect reticulate patterns of evolution that are not captured in a bifurcating species tree. While organellar DNA such as that of chloroplasts is usually transmitted maternally (Greiner et al. 2015) and thus

reflects seed dispersal in plants, nuclear DNA, transmitted by both parents, may reflect additional paths of dispersal (i.e. through pollen). Also, nuclear DNA is often subject to horizontal transfer, for example through hybridisation (Kidwell 1993). The combination of nuclear and chloroplast DNA has revealed intricate biogeographic patterns where new species arose from nuclear and chloroplast lineages of distant origins (e.g. in the grass subfamily Danthonioideae, Pirie et al. 2009). Also, duplication of genes or whole genomes has often generated families of paralogous genes, particularly in plants (Adams & Wendel 2005). Study of such gene families has allowed for example to demonstrate independent origins of C_4 photosynthesis (e.g. Christin & Besnard 2009) or the evolution of cold tolerance in pooid grasses (Sandve et al. 2008).

Early methods for inferring phylogenetic relationships were based on distance measures or a parsimony method minimising the number of inferred changes (Felsenstein 2004). Today, probabilistic approaches that model substitution rates and site heterogeneity in a Maximum Likelihood or Bayesian framework are the methods of choice (Anisimova et al. 2013). The size of modern datasets combined with the complexity of phylogenetic calculations mean that analytical solutions – i.e. finding the best among all possible trees – are usually not feasible, and heuristics are used to search the space of trees (Felsenstein 2004). Probabilistic models have also allowed to integrate divergence time estimation into phylogenetics, which involves two main considerations: First, divergences have to be calibrated to absolute time by known ages (reviewed by Ho & Phillips 2009). These are most often fossils, which usually provide a minimum age for a given group. An exception is e.g. the appearance of pollen typical of flowering plants in the Early Cretaceous, which is assumed to provide a reasonable upper limit for their origin (Anderson et al. 2005). Early methods allowed setting a calibration point mainly as fixed age or hard lower or upper bound. More advanced methods allow specifying a parametric distribution, e.g. a normal distribution around a mean, or exponential or log-normal distributions which place most weight near the calibration age but have no hard bounds. Apart from the age uncertainty, the confidence with which one can attribute a fossil to a particular node of the phylogeny will also influence the analysis (Ho & Phillips 2009). The second important part is the model of rate variation applied: in contrast with early assumptions of a fixed global “molecular clock”, the rate of DNA sequence evolution varies more or less strongly between lineages (Bromham & Penny 2003). This is usually accounted for with “relaxed” molecular clocks, which allow rates to be correlated between lineages (Thorne et al. 1998), to belong to discrete classes of a specified distribution (Drummond et al. 2006), or to form clusters of randomly distributed “local clocks” (Drummond & Suchard 2010).

Analysis of range evolution

The phylogenetic relationships and divergence times inferred for a set of species, combined with their current distribution ranges, can be used to make hypotheses about the evolution of their distribution through time. The earliest such approaches built on the theory of plate tectonics and attempted to fit the sequence of species divergence to a sequence of area

divergence in so-called taxon–area cladograms (reviewed in Cox et al. 2016). Later techniques specifically attempted to infer events that led to changes in distribution, notably vicariance or dispersal (e.g. the influential dispersal–viariance analysis, DIVA, Ronquist 1997). Both kinds of approach applied a parsimony reasoning, attempting to infer the simplest history with the least necessary changes. Often, dispersal had to be invoked to explain disjunctions, but it was seen as random "noise" rather than an interesting process. Dispersal, like extinction, posed the problem of not being tractable from ancestor to descendant; both processes were thus minimised by applying penalties (Cox et al. 2016). However, when divergence time estimation showed that lineage dispersal must have played a more prominent role than previously thought, parsimony-based approaches like DIVA came under critique (Kodandaramaiah 2010).

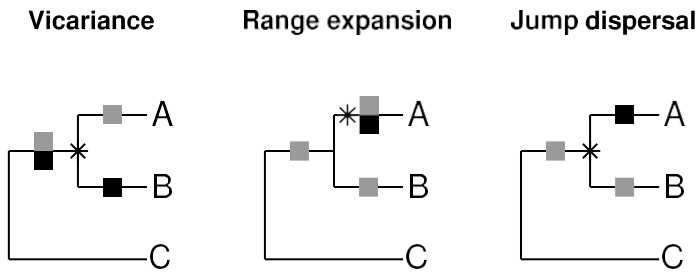

Figure 1. Phylogenetic patterns expected under processes of range evolution modelled by current methods. Grey and black represent two areas of distribution. The asterisk indicates where range evolution takes place: at speciation (cladogenetic) or on a branch (anagenetic). Full or partial range-copying (sympatry) is not shown. Note that a pattern found in a phylogeny does not imply that an event was actually possible. For example, a barrier could have emerged long before or after a presumed vicariance event.

Parsimony techniques use only the branching pattern of a phylogeny as evidence, but are indifferent to time. Thus, a dispersal event is considered the same way whether it occurred on a long or a short branch, i.e. in a long or a short time interval (Ree & Sanmartín 2009). This changed with the first parametric approach, the dispersal–extinction–cladogenesis (DEC) model (Ree et al. 2005, Ree & Smith 2008). This model specifically estimated rates of dispersal and extinction, which make those events more likely on longer branches. Parametric models also allowed to incorporate information on the timing of events that may have influenced the evolution of ranges, such as the appearance or disappearance of geological barriers (Ree & Sanmartín 2009). Furthermore, they estimate probabilities for each possible ancestral range, rather than simply reconstructing one most parsimonious scenario (Sanmartín 2016). DEC allows only one kind of dispersal – *range expansion* (Fig. 1). In contrast, a model developed for island biogeography (Sanmartín et al. 2008) allowed a lineage to instantly move to a new area, in a process called *jump dispersal* (Fig. 1). This was suggested to be closer to the reality of island lineages (Sanmartín et al. 2008). This was confirmed by a study which also developed a model integrating jump dispersal (or founder-event speciation) with DEC and a framework to compare different models (Matzke 2014). Parametric models are extensible and can incorporate more complex scenarios, e.g. dispersal rate varying with the carrying capacity of islands (Sanmartín et al. 2008) or with distance between areas (Van Dam & Matzke 2016).

DNA barcoding and metabarcoding

While phylogenetics is a tool to explore biodiversity in time, molecular methods have also been developed to study biodiversity in space. *DNA barcoding* was coined as a term to designate the automated identification of species by using a specific genomic marker, typically amplified by polymerase chain reaction (Hebert et al. 2003). The same new sequencing technologies that promoted the shift to phylogenomics also permitted the analysis of bulk samples that contain a mixture of DNA from different organisms (*metabarcoding*) (Taberlet et al. 2012b). This allowed studying realms of species diversity that were poorly explored before, for example micro-organisms in natural habitats. Such organisms, which previously had to be cultured – an impossibility for many of them – could now be identified directly from environmental samples (Taberlet et al. 2012a). Environmental metabarcoding has been applied to microbial communities from the human gut (Gill et al. 2006) over marine microplankton (Bucklin et al. 2016) to soil fungi at a global scale (Tedersoo et al. 2014) and allowed to confirm patterns found for macroorganisms such as biogeographic regionalisation and the latitudinal diversity gradient (e.g. Meiser et al. 2014). An increasing body of research has studied the importance of local environment versus dispersal and biogeographic contingency on structuring microbial communities (e.g. Östman et al. 2010, Lindström & Langenheder 2012).

Initial forages into environmental metabarcoding reported astonishing levels of diversity, for example in leaf-associated fungi (Jumpponen & Jones 2009). Some of these were later put into perspective when the frequency of errors in first high-throughput sequencing technologies was recognised (Quince et al. 2009). Newer technologies as well as bioinformatic methods can however minimise or correct for errors (Bálint et al. 2014). Other problems are related to what one considers as “species” in an environmental sample. Typically, *operational taxonomic units* (OTUs) are defined by DNA sequence similarity, but the idea of a fixed “barcoding gap” that would identify the limit between intra- and interspecific variability for a given genetic barcode was revised: rather, the amount of DNA divergence between species, or OTUs, is thought to vary among lineages (Puillandre et al. 2012). Moreover, to assign these OTUs to described species, an exhaustive and reliable reference sequence database is necessary (Collins & Cruickshank 2013). Such a database is typically not available for microorganisms, but methods exist for assigning them at least to higher-order taxa with existing reference sequences (Kõljalg et al. 2013, Lindahl et al. 2013). Metabarcoding relies heavily on computational tools in general for detecting errors, defining OTUs, and assigning them to taxa (Lindahl et al. 2013).

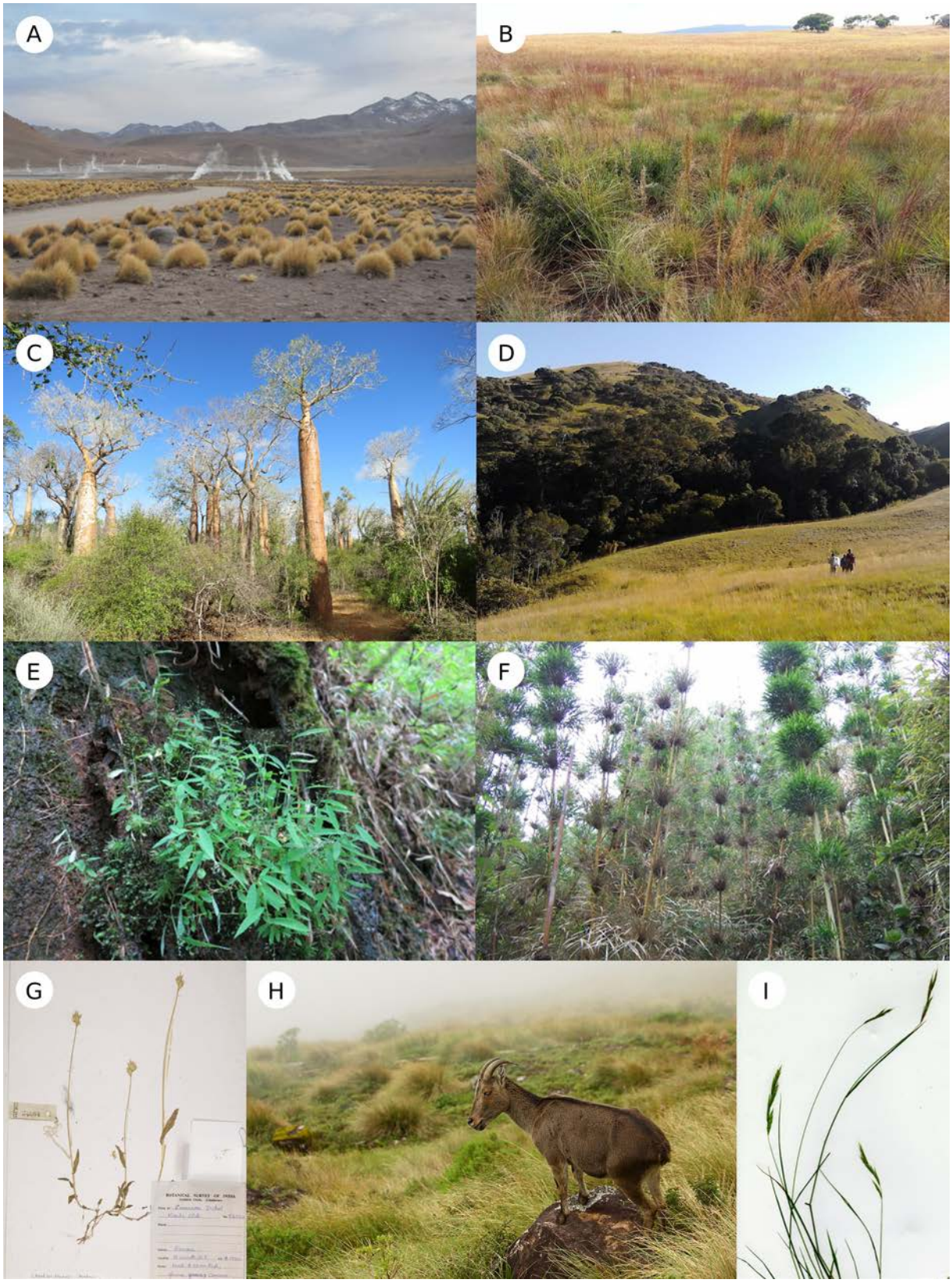
A
B
C
D
E
F
G
H
I

Figure 2 (previous page). Diversity of grasses and their habitats. The examples shown are related to this dissertation.
A – *Festuca orthophylla* occurs in the high Andes of South America. It belongs to the "American II" clade of fine-leaved Loliinae (subfamily Pooideae) that presumably immigrated via dispersal from Eurasia (chapter I). Loliinae are cool-adapted C_3 grasses mainly found in the northern hemisphere and are restricted to mountains at tropical latitudes.
B – Plateau grassland in the western part of Itremo protected area, Madagascar, dominated by C_4 grasses in Aristidoideae and Panicoideae: *Aristida tenuissima* (endemic), *Loudetia simplex* and *Schizachyrium sanguineum* (red patches). C_4 grasses immigrated repeatedly to Madagascar, presumably with the expansion of open habitats beginning in the Late Miocene (chapter II).
C – Spiny thicket in arid southwest Madagascar, assumed to be the most ancient biome on the island. It is rich in endemics such as baobabs but also grasses, one of which may be the oldest endemic C_4 grass lineage on the island (*Eragrostis chabouisii*, Chloridoideae, chapter II).
D – Plateau landscape at West Itremo, Madagascar, with open grassland in the foreground, dense gallery forest in the ravine, and open *tapia* woodland on the hills in the background. Grassland and tapia woodland have similar herbaceous plant communities dominated by C_4 grasses, while the forest understorey harbours almost exclusively C_3 grasses (chapters II and IV).
E – *Panicum vohitrense*, member of a C_3 "Madagascar shade clade" in subfamily Panicoideae that is mainly found in the forest understorey and diversified in Madagascar and nearby West Indian Ocean islands (chapter II).
F – *Nastus borbonicus*, an endemic bamboo (Bambusoideae) from Réunion Island. It belongs to the Hickeliinae, a woody bamboo clade that radiated in the wider Madagascar region and dispersed only once to the African mainland (chapter II).
G – *Chandrasekharania keralensis* (Panicoideae), type specimen at Royal Botanic Gardens, Kew. This monotypic genus and *Jansenella* form a species-poor tropical C_3 lineage mainly found in the Western Ghats of India (chapter III).
H – Mountain grassland in Eravikulum National Park, Western Ghats, India. The presence of endemic grazers such as the Nilgiri tahr (pictured) testifies to the old-growth nature of these grasslands. The Western Ghats may have been the origin of a very diverse pantropical C_4 grass lineage, Andropogoneae–Arundinelleae (chapter III).
I – *Brachypodium madagascariense* (Pooideae), an endemic C_3 grass from the forest understorey in highland Madagascar. It harbours symbiotic fungal endophytes of the genus *Epichloë* that presumably co-dispersed with their host plant (chapter IV).
Image sources: A – Yastay/Wikimedia Commons (CC BY-SA 4.0); F – B.navez/Wikimedia Commons (CC BY-SA 3.0); G – Bord of Trustees, Royal Botanic Gardens, Kew; H – Navaneeth Kishor/Wikimedia Commons (CC BY-SA 3.0); B, C, E and I – Maria S. Vorontsova (with permission); D – author.

1.3 Grasses and grasslands

Grass diversity

In this dissertation, I focused on one particular group of plants, the grass family (Poaceae). Grasses are one of the largest known plant families, with around 11,000 species (Kellogg 2015). They have colonised all continents and are found from tropical latitudes to polar regions. The ecological amplitude they occupy is also evident in the habitats they occur in, ranging from moist tropical forests over savannahs to deserts and alpine tundra. Habitats where grasses dominate account for up to 40% of the Earth's terrestrial surface (Gibson 2009). The arguably most important cultivated plants – wheat, maize, millets, sugar cane and rice – belong to the family, and many more grass species have local to regional economic importance.

The grass family is thought to have initially evolved in the understorey of tropical forests (Edwards & Smith 2010). With subsequent diversification, different grass clades adapted to various ecological conditions (Fig. 2). Studies have identified two main lineages in the family, namely the BOP (or BEP) and the PACMAD clade (Fig. 3), named after the initials of the

subfamilies they comprise (GPWG 2001, GPWG II 2012, Soreng et al. 2015). These two clades each include roughly half of all grass species. Grasses in the BOP clade are mainly temperate and have most of their diversity in the high latitudes (Kellogg 2015). However, they also include species in the rice subfamily (Oryzoideae) occurring in tropical to subtropical wetlands and forest understorey as well as two bamboo tribes, Bambuseae and Olyreae, that are mainly tropical. The PACMAD clade, on the contrary, is most diverse at low latitudes, with exceptions e.g. in subfamily Danthonioideae. This group notably includes the grasses that dominate tropical savannahs and other grasslands (Kellogg 2015).

C_4 photosynthesis and tropical grasslands

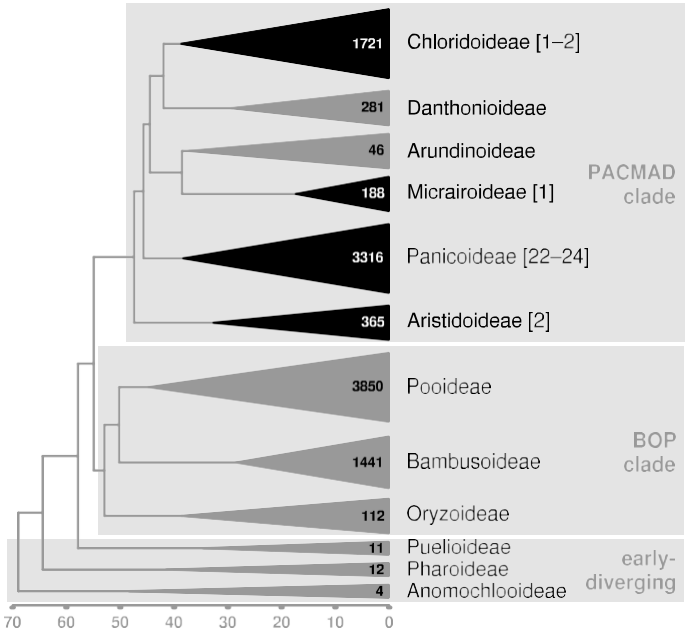

Figure 3. Relationships in the grass family (Poaceae). Species numbers are given for each subfamily. Subfamilies in which C_4 photosynthesis evolved are in black, with the number of independent origins given in square brackets. The scale is in millions of years (under external angiosperm calibration). Age-calibrated phylogeny taken from Christin et al. (2014), all other information from Kellogg (2015). Nomenclature follows Soreng et al. (2015, 2017).

Grasses, like most plants, evolved under an atmospheric CO_2 concentration much higher than today. Leaf anatomy and photosynthesis was adapted to efficiently assimilate carbon under such conditions (Sage et al. 2012). This efficiency was challenged when atmospheric CO_2 concentration dropped beginning around 34 million years ago (Ma), at the start of the Oligocene (Zachos et al. 2008). The leaf enzyme Rubisco which catalyses CO_2 assimilation will also metabolise oxygen at low relative CO_2 concentration, in a process called photorespiration which leads to a loss of CO_2 assimilation efficiency (reviewed in Sage et al. 2012). Photorespiration is further increased by factors that lower relative CO_2 concentration locally in the leaf or decrease Rubisco's specificity for CO_2, such as high temperatures and light, or hydraulic stress, e.g. on saline soil. Under this selective pressure, several lineages of plants independently evolved the C_4 pathway from the ancestral C_3 photosynthesis. This new anatomical and physiological syndrome involves pre-fixing CO_2 and concentrating it around Rubisco, thus effectively suppressing photorespiration. C_4 photosynthesis is beneficial in hot and seasonally dry climates but costly under high atmospheric CO_2 or in temperate climates (Sage et al. 2012). Grasses are the plant family with the highest number of C_4 lineages, counting at least 26 origins which all occurred in the PACMAD clade (Fig. 2; Kellogg 2015).

There is evidence from dated phylogenies that the Oligocene CO_2 decline triggered the evolution of C_4 photosynthesis in grasses (Christin et al. 2008, Vicentini et al. 2008), although older grass age estimates have challenged this (see below). In any case, the diversification and spread of C_4 grasses worldwide is estimated to have occurred several millions of years after their emergence (Strömberg 2005, 2011, Bouchenak-Khelladi et al. 2014). This means that, while low atmospheric CO_2 probably facilitated the switch to C_4 photosynthesis, other factors contributed to their ecological success and diversification (reviewed by Strömberg 2011). Indeed, the Earth's surface changed dramatically in the Miocene, with the climate getting drier and colder, forests retreating, and fires becoming more frequent. Grasslands expanded both at temperate and tropical latitudes. Fossil stable isotope evidence from paleosols and grazer tooth enamel indicate that in the tropics and warm-temperate regions, these were initially C_3 grasslands, which were then gradually colonised and eventually dominated by C_4 grasses (Strömberg 2011). This expansion of C_4 grassy biomes occurred in the Late Miocene, between 3 and 10 Ma and at variable pace in different locations (Edwards et al. 2010). The switch to C_4 photosynthesis and the expansion of C_4 grasslands led to an increase of diversification in the PACMAD clade, while concomitantly, the Pooideae subfamily in the BOP clade diversified in cool-temperate regions (Spriggs et al. 2014).

Tropical grasslands have thus been present since at least the Miocene, but in many places they are perceived as non-natural, created through human intervention such as the cutting of trees and the introduction of livestock (Veldman 2016). This is probably due to the fact that the climatic and soil niches of tropical grasslands and forests overlap and they can thus be found as mosaics in the same regions (Sankaran et al. 2005, Lehmann et al. 2011). Often, C_4 grassy biomes have even been classified as degraded forest, which they can resemble in structure (Veldman 2016). Their functioning is however fundamentally distinct from forests (reviewed by Ratnam et al. 2011): grasslands may feature an open tree canopy (then termed *savannah*) but the understorey is always dominated by grasses, which are predominantly C_4 grasses. The grasses are critical to the feedbacks that maintain a savannah, through their biomass accumulation in the wet season, feeding herbivores and fuelling fires in the dry season. Regular fires, along with the action of grazers, are indeed believed to maintain savannah and forest as alternative stable states under similar climatic and soil conditions (Lehmann et al. 2011, Hoffmann et al. 2012). Many tropical C_4 grasslands are threatened due to human encroachment and climate change but also because of wrong management such as fire suppression and afforestation, based on the misconception as degraded landscapes (Parr et al. 2014, Bond 2016). The presence of native, diverse assemblages of species and endemic radiations of fire-adapted taxa have been used as indicators for old-growth grasslands (Maurin et al. 2014, Veldman 2016).

Age and distribution

The age of the grass family is still a matter of debate. The first estimations of divergence times in Poales, using external monocot fossils for calibration, estimated that the grass family had already diversified at the end of the Cretaceous (Bremer 2002, Janssen & Bremer 2004).

Later analyses including more genetic markers and grass fossils (Christin et al. 2008, Vicentini et al. 2008) found more recent dates, with a BOP–PACMAD divergence at between 50 and 60 Ma and C_4 origins compatible with the Oligocene CO_2 decline (see above). These estimates were again challenged by the discovery of grass microfossils from the Late Cretaceous, which, when assigned to tribe Oryzeae, suggest early divergences in the family roughly 20 million years older (Prasad et al. 2011). An evaluation of different calibration approaches in grasses concluded that ages of macrofossils are more compatible with estimates based on external angiosperm ages, considered relatively well constrained (Christin et al. 2014). The authors concluded that ages based on old microfossils cannot be dismissed, but are best treated as alternative hypothesis at this time. Under both hypotheses however, the diversification of the large BOP and PACMAD clades would have occurred after the disintegration of Gondwana in the mid-Cretaceous. Thus, lineage dispersal between continents must have occurred repeatedly for the grasses to have reached their almost cosmopolitan distribution (Bouchenak-Khelladi et al. 2010). Indeed, a history of frequent dispersals has been inferred for clades with disjunct distributions, such as the temperate *Hordeum* (Blattner 2006), Loliinae (Inda et al. 2008) or the mainly south temperate Danthonioideae (Linder et al. 2013).

Symbiosis with endophytes

Throughout their distribution, grasses interact with other organisms. Like almost all plants, they have microbial symbionts, including fungi (Rodriguez et al. 2009). Endophytes – microbes growing within a plant host without harming it (Wilson 1995) – associating with grasses have attracted particular research interest. In the fungal family Clavicipitaceae (Ascomycota), endophytes of the genus *Epichloë* (synonym *Neotyphodium*, Leuchtmann et al. 2014) have specialised on grasses (Schardl et al. 2004). Growing in the leaves without causing any symptoms through most of their growth cycle, many species have evolved a close mutualistic relationship with their grass host: they protect it against herbivores through the production of alkaloids and can increase resistance to drought and certain pathogens (reviewed by Clay & Schardl 2002). Through these attributes, the fungi can even affect the structure of the food chain (Omacini et al. 2001) and the plant community (Clay & Holah 1999) their host belongs to. As the mutualistic *Epichloë* species are seedborne and transmitted vertically, they have co-evolved closely with their hosts, even leading to co-divergence of lineages (Clay & Schardl 2002).

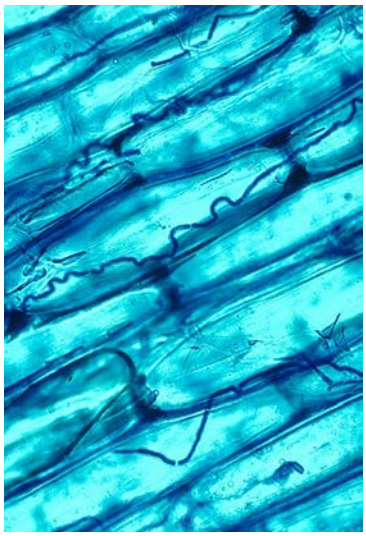

Figure 4. Hyphae of *Epichloë coenophiala* growing as endophyte between cells in leaf tissue of tall fescue, *Lolium arundinaceum*. Source: Nick Hill/USDA/ Wikimedia Commons (public domain).

The *Epichloë* endophytes are restricted to temperate grasses of subfamily Pooideae but are found on most continents (Clay & Schardl 2002). In a survey of tropical forest grasses in

Panama, they were absent (Higgins et al. 2011). Many more species of endophytes associate with grasses, although presumably more loosely than the *Epichloë* endophytes (Sánchez Márquez et al. 2012). Grass endophytes in tropical regions remain largely unknown, but the latitudinal diversity gradient with higher diversity in the tropics is also thought to apply to endophytes (Arnold et al. 2000). A mutualistic relationship has also been demonstrated between a panicoid grass, *Dichanthelium lanuginosum*, and a non-*Epichloë* endophyte (Redman et al. 2002), suggesting the existence of further unknown co-evolutionary systems, with potential community-wide impacts. Despite the importance of grasslands and herbivory in the tropics (see above), basic questions remain to be answered on tropical grass endophytes. These include whether dispersal limitation and regionalisation play an important role, and whether host and environmental factors are strong drivers of community structure, such as shown in tree phyllosphere fungi (Cordier et al. 2012, Zimmerman & Vitousek 2012, Bálint et al. 2013).

1.4 Dissertation outline

In the work I present here, I have used molecular tools to study the biogeography of grasses, the assembly of tropical grasslands, and the community structure of grass endophytes. All parts integrate the evolutionary history of grass lineages as a major determinant of current patterns. In the first chapter, I describe an analysis of dispersal history in a temperate clade of grasses found in isolated areas at tropical latitudes. Distance was found to be a major determinant of dispersal, while ecological differences between grass clades had a minor effect. In the second chapter, I present a study of the age and biogeographic history of the grass flora of Madagascar. Results show that it has predominantly African origins, features two diverse *in situ* radiations of bamboos and forest grasses, and is characterised by repeated immigration and speciation of C_4 grasses since the Miocene. In the third chapter, I show how phylogenomic methods allowed me to place three enigmatic Asian C_3 species in the grass phylogeny. Their position suggests that a very large, pantropical C_4 grass clade has its origin in India. In the fourth and last chapter, I describe a metabarcoding method to study communities of grass endophytes in Madagascar. Results are preliminary and highlight methodological limitations but attest the presence of Pooideae-associated *Epichloë* endophytes in Madagascar.

1.5 References

Adams K.L. & Wendel J.F. (2005): Polyploidy and genome evolution in plants. *Current Opinion in Plant Biology*, 8: 135–141.

Anderson C.L., Bremer K. & Friis E.M. (2005): Dating phylogenetically basal eudicots using *rbcL* sequences and multiple fossil reference points. *American Journal of Botany*, 92: 1737–1748.

Anisimova M., Liberles D.A., Philippe H., Provan J., Pupko T. & von Haeseler A. (2013): State-of the art methodologies dictate new standards for phylogenetic analysis. *BMC Evolutionary Biology*, 13: 161.

Arnold A.E., Maynard Z., Gilbert G.S., Coley P.D. & Kursar T.A. (2000): Are tropical fungal endophytes hyperdiverse? *Ecology Letters*, 3: 267–274.

Bálint M., Schmidt P.-A., Sharma R., Thines M. & Schmitt I. (2014): An Illumina metabarcoding pipeline for fungi. *Ecology and Evolution*, 4: 2642–2653.

Bálint M., Tiffin P., Hallström B., O'Hara R.B., Olson M.S. et al. (2013): Host genotype shapes the foliar fungal microbiome of balsam poplar (*Populus balsamifera*). *PLoS ONE*, 8: e53987.

Blattner F.R. (2006): Multiple intercontinental dispersals shaped the distribution area of *Hordeum* (Poaceae). *New Phytologist*, 169: 603–614.

Bond W.J. (2016): Ancient grasslands at risk. *Science*, 351: 120–122.

Bouchenak-Khelladi Y., Slingsby J.A., Verboom G.A. & Bond W.J. (2014): Diversification of C_4 grasses (Poaceae) does not coincide with their ecological dominance. *American Journal of Botany*, 101: 300–307.

Bouchenak-Khelladi Y., Verboom G.A., Savolainen V. & Hodkinson T.R. (2010): Biogeography of the grasses (Poaceae): a phylogenetic approach to reveal evolutionary history in geographical space and geological time. *Botanical Journal of the Linnean Society*, 162: 543–557.

Bremer K. (2002): Gondwanan evolution of the grass alliance of families (Poales). *Evolution*, 56: 1374–1387.

Bromham L. & Penny D. (2003): The modern molecular clock. *Nature Reviews Genetics*, 4: 216.

Bucklin A., Lindeque P.K., Rodriguez-Ezpeleta N., Albaina A. & Lehtiniemi M. (2016): Metabarcoding of marine zooplankton: prospects, progress and pitfalls. *Journal of Plankton Research*, 38: 393–400.

Buerki S., Devey D.S., Callmander M.W., Phillipson P.B. & Forest F. (2013): Spatio-temporal history of the endemic genera of Madagascar. *Botanical Journal of the Linnean Society*, 171: 304–329.

Christin P.-A. & Besnard G. (2009): Two independent C_4 origins in Aristidoideae (Poaceae) revealed by the recruitment of distinct phosphoenolpyruvate carboxylase genes. *American Journal of Botany*, 96: 2234–2239.

Christin P.-A., Besnard G., Samaritani E., Duvall M.R., Hodkinson T.R. et al. (2008): Oligocene CO_2 decline promoted C_4 photosynthesis in grasses. *Current Biology*, 18: 37–43.

Christin P.-A., Spriggs E., Osborne C.P., Strömberg C.A.E., Salamin N. & Edwards E.J. (2014): Molecular dating, evolutionary rates, and the age of the grasses. *Systematic Biology*, 63: 153–165.

Clay K. & Holah J. (1999): Fungal endophyte symbiosis and plant diversity in successional fields. *Science*, 285: 1742–1744.

Clay K. & Schardl C. (2002): Evolutionary origins and ecological consequences of endophyte symbiosis with grasses. *The American Naturalist*, 160: S99–S127.

Collins R.A. & Cruickshank R.H. (2013): The seven deadly sins of DNA barcoding. *Molecular Ecology Resources*, 13: 969–975.

Cook L.G. & Crisp M.D. (2005): Not so ancient: the extant crown group of *Nothofagus* represents a post-Gondwanan radiation. *Proceedings of the Royal Society B: Biological Sciences*, 272: 2535–2544.

Cordier T., Robin C., Capdevielle X., Fabreguettes O., Desprez-Loustau M.-L. & Vacher C. (2012): The composition of phyllosphere fungal assemblages of European beech (*Fagus sylvatica*) varies significantly along an elevation gradient. *New Phytologist*, 196: 510–519.

Cox C.B., Moore P.D. & Ladle R. (2016): *Biogeography: an ecological and evolutionary approach.* John Wiley & Sons, Chichester.

Craw R. (1989): New Zealand biogeography: a panbiogeographic approach. *New Zealand Journal of Zoology*, 16: 527–547.

Crisp M.D., Arroyo M.T.K., Cook L.G., Gandolfo M.A., Jordan G.J. et al. (2009): Phylogenetic biome conservatism on a global scale. *Nature*, 458: 754–756.

Croizat L. (1958): *Panbiogeography; or, An introductory synthesis of zoogeography, phytogeography, and geology, with notes on evolution, systematics, ecology, anthropology, etc.* Self-published, Caracas.

Delsuc F., Brinkmann H. & Philippe H. (2005): Phylogenomics and the reconstruction of the tree of life. *Nature Reviews Genetics*, 6: 361–375.

Donoghue M.J. (2008): A phylogenetic perspective on the distribution of plant diversity. *Proceedings of the National Academy of Sciences*, 105: 11549–11555.

Drummond A.J., Ho S.Y.W., Phillips M.J. & Rambaut A. (2006): Relaxed phylogenetics and dating with confidence. *PLOS Biology*, 4: e88.

Drummond A.J. & Suchard M.A. (2010): Bayesian random local clocks, or one rate to rule them all. *BMC Biology*, 8: 114.

Edwards E.J. & Donoghue M.J. (2013): Is it easy to move and easy to evolve? Evolutionary accessibility and adaptation. *Journal of Experimental Botany*, 64: 4047–4052.

Edwards E.J., Osborne C.P., Strömberg C.A.E. & Smith S.A. (2010): The origins of C_4 grasslands: integrating evolutionary and ecosystem science. *Science*, 328: 587–591.

Edwards E.J. & Smith S.A. (2010): Phylogenetic analyses reveal the shady history of C_4 grasses. *Proceedings of the National Academy of Sciences*, 107: 2532–2537.

Elton C.S. (1927): *Animal ecology.* University of Chicago Press, Chicago.

Engler A. (1879): *Versuch einer Entwicklungsgeschichte der Pflanzenwelt: insbesondere der Florengebiete seit der Tertiärperiode.* W. Engelmann, Leipzig.

Felsenstein J. (2004): *Inferring phylogenies.* Sinauer Associates, Sunderland, Mass.

Gibson D.J. (2009): *Grasses and grassland ecology.* Oxford University Press,

Gill S.R., Pop M., DeBoy R.T., Eckburg P.B., Turnbaugh P.J. et al. (2006): Metagenomic analysis of the human distal gut microbiome. *Science*, 312: 1355–1359.

Gillespie R.G., Baldwin B.G., Waters J.M., Fraser C.I., Nikula R. & Roderick G.K. (2012): Long-distance dispersal: a framework for hypothesis testing. *Trends in Ecology & Evolution*, 27: 47–56. GPWG – Grass Phylogeny Working Group (2001): Phylogeny and subfamilial classification of the grasses (Poaceae). *Annals of the Missouri Botanical Garden*, 88: 373–457.

GPWG II – Grass Phylogeny Working Group II (2012): New grass phylogeny resolves deep evolutionary relationships and discovers C_4 origins. *New Phytologist*, 193: 304–312.

Greiner S., Sobanski J. & Bock R. (2015): Why are most organelle genomes transmitted maternally? *BioEssays*, 37: 80–94.

Grinnell J. (1917): The niche-relationships of the California thrasher. *The Auk*, 34: 427–433.

Hardin G. (1960): The competitive exclusion principle. *Science*, 131: 1292–1297.

Hebert P.D.N., Cywinska A., Ball S.L. & deWaard J.R. (2003): Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B: Biological Sciences*, 270: 313–321.

Hey J. (2006): On the failure of modern species concepts. *Trends in Ecology & Evolution*, 21: 447–450.

Higgins K.L., Coley P.D., Kursar T.A. & Arnold A.E. (2011): Culturing and direct PCR suggest prevalent host generalism among diverse fungal endophytes of tropical forest grasses. *Mycologia*, 103: 247–260.

Hillebrand H. (2004): On the generality of the latitudinal diversity gradient. *The American Naturalist*, 163: 192–211.

Ho S.Y.W. & Phillips M.J. (2009): Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Systematic Biology*, 58: 367–380.

Hoffmann W.A., Geiger E.L., Gotsch S.G., Rossatto D.R., Silva L.C.R. et al. (2012): Ecological thresholds at the savanna-forest boundary: how plant traits, resources and fire govern the distribution of tropical biomes. *Ecology Letters*, 15: 759–768.

Hubbell S.P. (2001): *The unified neutral theory of biodiversity and biogeography*. Princeton University Press,

Humphreys A.M. & Linder H.P. (2013): Evidence for recent evolution of cold tolerance in grasses suggests current distribution is not limited by (low) temperature. *New Phytologist*, 198: 1261–1273.

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

Inda L.A., Segarra-Moragues J.G., Müller J., Peterson P.M. & Catalán P. (2008): Dated historical biogeography of the temperate Loliinae (Poaceae, Pooideae) grasses in the northern and southern hemispheres. *Molecular Phylogenetics and Evolution*, 46: 932–957.

Janssen T. & Bremer K. (2004): The age of major monocot groups inferred from 800+ *rbcL* sequences. *Botanical Journal of the Linnean Society*, 146: 385–398.

Jumpponen A. & Jones K.L. (2009): Massively parallel 454 sequencing indicates hyperdiverse fungal communities in temperate *Quercus macrocarpa* phyllosphere. *New Phytologist*, 184: 438–448.

Kellogg E.A. (2015): *Flowering Plants. Monocots: Poaceae*. Springer, Heidelberg.

Kidwell M.G. (1993): Lateral transfer in natural populations of eukaryotes. *Annual Review of Genetics*, 27: 235–256.

Kodandaramaiah U. (2010): Use of dispersal–vicariance analysis in biogeography – a critique. *Journal of Biogeography*, 37: 3–11.

Kõljalg U., Nilsson R.H., Abarenkov K., Tedersoo L., Taylor A.F.S. et al. (2013): Towards a unified paradigm for sequence-based identification of fungi. *Molecular Ecology*, 22: 5271–5277.

Lehmann C.E.R., Archibald S.A., Hoffmann W.A. & Bond W.J. (2011): Deciphering the distribution of the savanna biome. *New Phytologist*, 191: 197–209.

Leroy J.-F. (1978): Composition, origin, and affinities of the Madagascan vascular flora. *Annals of the Missouri Botanical Garden*, 65: 535–589.

Leuchtmann A., Bacon C.W., Schardl C.L., Jr J.F.W. & Tadych M. (2014): Nomenclatural realignment of *Neotyphodium* species with genus *Epichloë*. *Mycologia*, 106: 202–215.

Levin S.A. (1992): The problem of pattern and scale in ecology: the Robert H. MacArthur award lecture. *Ecology*, 73: 1943–1967.

Lindahl B.D., Nilsson R.H., Tedersoo L., Abarenkov K., Carlsen T. et al. (2013): Fungal community analysis by high-throughput sequencing of amplified markers – a user's guide. *New Phytologist*, 199: 288–299.

Linder H.P., Antonelli A., Humphreys A.M., Pirie M.D. & Wüest R.O. (2013): What determines biogeographical ranges? Historical wanderings and ecological constraints in the danthonioid grasses. *Journal of Biogeography*, 40: 821–834.

Lindström E.S. & Langenheder S. (2012): Local and regional factors influencing bacterial community assembly. *Environmental Microbiology Reports*, 4: 1–9.

Losos J.B. & Ricklefs R.E. (2009): Adaptation and diversification on islands. *Nature*, 457: 830.

MacArthur R. & Levins R. (1967): The limiting similarity, convergence, and divergence of

coexisting species. *The American Naturalist*, 101: 377–385.

MacArthur R.H. & Wilson E.O. (1967): *The theory of island biogeography*. Princeton University Press, Princeton, New Jersey.

Matthysen E. (2012): Multicausality of dispersal: a review. *Dispersal ecology and evolution* (ed. by J. Clobert, M. Baguette, T.G. Benton & J.M. Bullock): 3–18. Oxford University Press, Oxford, New York.

Matzke N.J. (2014): Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Systematic Biology*, 63: 951–970.

Maurin O., Davies T.J., Burrows J.E., Daru B.H., Yessoufou K. et al. (2014): Savanna fire and the origins of the "underground forests" of Africa. *New Phytologist*, 204: 201–214.

McGuinness K.A. (1984): Species–area curves. *Biological Reviews*, 59: 423–440.

McLoughlin S. (2001): The breakup history of Gondwana and its impact on pre-Cenozoic floristic provincialism. *Australian Journal of Botany*, 49: 271–300.

Meiser A., Bálint M. & Schmitt I. (2014): Meta-analysis of deep-sequenced fungal communities indicates limited taxon sharing between studies and the presence of biogeographic patterns. *New Phytologist*, 201: 623–635.

Merckx V.S.F.T., Hendriks K.P., Beentjes K.K., Mennes C.B., Becking L.E. et al. (2015): Evolution of endemism on a young tropical mountain. *Nature*, 524: 347–350.

Metzker M.L. (2010): Sequencing technologies – the next generation. *Nature Reviews Genetics*, 11: 31.

Nathan R. (2006): Long-distance dispersal of plants. *Science*, 313: 786–788.

Nathan R., Perry G., Cronin J.T., Strand A.E. & Cain M.L. (2003): Methods for estimating long-distance dispersal. *Oikos*, 103: 261–273.

Nelson G. (1978): From Candolle to Croizat: comments on the history of biogeography. *Journal of the History of Biology*, 11: 269–305.

Omacini M., Chaneton E.J., Ghersa C.M. & Müller C.B. (2001): Symbiotic fungal endophytes control insect host–parasite interaction webs. *Nature*, 409: 78–81.

Östman Ö., Drakare S., Kritzberg E.S., Langenheder S., Logue J.B. & Lindström E.S. (2010): Regional invariance among microbial communities. *Ecology Letters*, 13: 118–127.

Parr C.L., Lehmann C.E.R., Bond W.J., Hoffmann W.A. & Andersen A.N. (2014): Tropical grassy biomes: misunderstood, neglected, and under threat. *Trends in Ecology & Evolution*, 29: 205–213.

Pirie M.D., Humphreys A.M., Barker N.P. & Linder H.P. (2009): Reticulation, data combination, and inferring evolutionary history: an example from Danthonioideae (Poaceae). *Systematic Biology*, syp068.

Popp M., Mirré V. & Brochmann C. (2011): A single Mid-Pleistocene long-distance dispersal by a bird can explain the extreme bipolar disjunction in crowberries (*Empetrum*). *Proceedings of the National Academy of Sciences*, 108: 6520–6525.

Prasad V., Strömberg C. a. E., Leaché A.D., Samant B., Patnaik R. et al. (2011): Late Cretaceous origin of the rice tribe provides evidence for early diversification in Poaceae. *Nature Communications*, 2: 480.

Preston F.W. (1948): The commonness, and rarity, of species. *Ecology*, 29: 254–283.

Prinzing A. (2001): The niche of higher plants: evidence for phylogenetic conservatism. *Proceedings of the Royal Society of London B: Biological Sciences*, 268: 2383–2389.

Puillandre N., Lambert A., Brouillet S. & Achaz G. (2012): ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology*, 21: 1864–1877.

de Queiroz A. (2005): The resurrection of oceanic dispersal in historical biogeography. *Trends in Ecology & Evolution*, 20: 68–73.

de Queiroz K. (2007): Species concepts and species delimitation. *Systematic Biology*, 56: 879–886.

Quince C., Lanzén A., Curtis T.P., Davenport R.J., Hall N. et al. (2009): Accurate determination of microbial diversity from 454 pyrosequencing data. *Nature Methods*, 6: 639–641.

Ratnam J., Bond W.J., Fensham R.J., Hoffmann W.A., Archibald S. et al. (2011): When is a "forest" a savanna, and why does it matter? *Global Ecology and Biogeography*, 20: 653–660.

Redman R.S., Sheehan K.B., Stout R.G., Rodriguez R.J. & Henson J.M. (2002): Thermotolerance generated by plant/fungal symbiosis. *Science*, 298: 1581–1581.

Ree R.H., Moore B.R., Webb C.O. & Donoghue M.J. (2005): A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. *Evolution*, 59: 2299–2311.

Ree R.H. & Sanmartín I. (2009): Prospects and challenges for parametric models in historical biogeographical inference. *Journal of Biogeography*, 36: 1211–1220.

Ree R.H. & Smith S.A. (2008): Maximum Likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology*, 57: 4–14.

Renner S.S. (2005): Relaxed molecular clocks for dating historical plant dispersal events. *Trends in Plant Science*, 10: 550–558.

Ricklefs R.E. (1987): Community diversity: relative roles of local and regional processes. *Science*, 235: 167–171.

Ricklefs R.E. (2004): A comprehensive framework for global patterns in biodiversity. *Ecology Letters*, 7: 1–15.

Ricklefs R.E. & Jenkins D.G. (2011): Biogeography and ecology: towards the integration of two disciplines. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366: 2438–2448.

Rodriguez R.J., White Jr J.F., Arnold A.E. & Redman R.S. (2009): Fungal endophytes: diversity and functional roles. *New Phytologist*, 182: 314–330.

Ronquist F. (1997): Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Systematic Biology*, 46: 195–203.

Sage R.F., Sage T.L. & Kocacinar F. (2012): Photorespiration and the evolution of C_4 photosynthesis. *Annual Review of Plant Biology*, 63: 19–47.

Sánchez Márquez S., Bills G.F., Herrero N. & Zabalgogeazcoa Í. (2012): Non-systemic fungal endophytes of grasses. *Fungal Ecology*, 5: 289–297.

Sandve S.R., Rudi H., Asp T. & Rognli O.A. (2008): Tracking the evolution of a cold stress associated gene family in cold tolerant grasses. *BMC Evolutionary Biology*, 8: 245.

Sankaran M., Hanan N.P., Scholes R.J., Ratnam J., Augustine D.J. et al. (2005): Determinants of woody cover in African savannas. *Nature*, 438: 846–849.

Sanmartín I. (2016): Breaking the chains of parsimony: the development of parametric methods in historical biogeography. *Biogeography: an ecological and evolutionary approach* (ed. by C.B. Cox, P.D. Moore & R. Ladle): 241–245. John Wiley & Sons, Chichester.

Sanmartín I. & Ronquist F. (2004): Southern hemisphere biogeography inferred by event-based models: plant versus animal patterns. *Systematic Biology*, 53: 216–243.

Sanmartín I., Van Der Mark P. & Ronquist F. (2008): Inferring dispersal: a Bayesian approach to phylogeny-based island biogeography, with special reference to the Canary Islands. *Journal of Biogeography*, 35: 428–449.

Schardl C.L., Leuchtmann A. & Spiering M.J. (2004): Symbioses of grasses with seedborne fungal endophytes. *Annual Review of Plant Biology*, 55: 315–340.

Sclater P.L. (1858): On the general geographical distribution of the members of the class Aves. *Zoological Journal of the Linnean Society*, 2: 130–136.

Simon M.F., Grether R., Queiroz L.P. de, Skema C., Pennington R.T. & Hughes C.E. (2009): Recent assembly of the Cerrado, a neotropical plant diversity hotspot, by in situ evolution of adaptations to fire. *Proceedings of the National Academy of Sciences*, 106: 20359–20364.

Soreng R.J., Peterson P.M., Romaschenko K., Davidse G., Teisher J.K. et al. (2017): A worldwide phylogenetic classification of the Poaceae (Gramineae) II: An update and a comparison of two 2015 classifications. *Journal of Systematics and Evolution*, 55: 259–290.

Soreng R.J., Peterson P.M., Romaschenko K., Davidse G., Zuloaga F.O. et al. (2015): A worldwide phylogenetic classification of the Poaceae (Gramineae). *Journal of Systematics and Evolution*, 53: 117–137.

Spriggs E.L., Christin P.-A. & Edwards E.J. (2014): C_4 photosynthesis promoted species diversification during the Miocene grassland expansion. *PLoS ONE*, 9: e97722.

Strömberg C.A.E. (2005): Decoupled taxonomic radiation and ecological expansion of open-habitat grasses in the Cenozoic of North America. *Proceedings of the National Academy of Sciences of the United States of America*, 102: 11980–11984.

Strömberg C.A.E. (2011): Evolution of grasses and grassland ecosystems. *Annual Review of Earth and Planetary Sciences*, 39: 517–544.

Taberlet P., Coissac E., Hajibabaei M. & Rieseberg L.H. (2012a): Environmental DNA. *Molecular Ecology*, 21: 1789–1793.

Taberlet P., Coissac E., Pompanon F., Brochmann C. & Willerslev E. (2012b): Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, 21: 2045–2050.

Tansley A.G. (1935): The use and abuse of vegetational concepts and terms. *Ecology*, 16: 284–307.

Tedersoo L., Bahram M., Põlme S., Kõljalg U., Yorou N.S. et al. (2014): Global diversity and geography of soil fungi. *Science*, 346: 1256688.

Thorne J.L., Kishino H. & Painter I.S. (1998): Estimating the rate of evolution of the rate of molecular evolution. *Molecular Biology and Evolution*, 15: 1647–1657.

Tonti-Filippini J., Nevill P.G., Dixon K. & Small I. (2017): What can we do with 1000 plastid genomes? *The Plant Journal*, 90: 808–818.

Van Dam M.H. & Matzke N.J. (2016): Evaluating the influence of connectivity and distance on biogeographical patterns in the south-western deserts of North America. *Journal of Biogeography*, 43: 1514–1532.

Veldman J.W. (2016): Clarifying the confusion: old-growth savannahs and tropical ecosystem degradation. *Phil. Trans. R. Soc. B*, 371: 20150306.

Vicentini A., Barber J.C., Aliscioni S.S., Giussani L.M. & Kellogg E.A. (2008): The age of the grasses and clusters of origins of C_4 photosynthesis. *Global Change Biology*, 14: 2963–2977.

Waters C.N., Zalasiewicz J., Summerhayes C., Barnosky A.D., Poirier C. et al. (2016): The Anthropocene is functionally and stratigraphically distinct from the Holocene. *Science*, 351: aad2622.

Weiher E., Freund D., Bunton T., Stefanski A., Lee T. & Bentivenga S. (2011): Advances, challenges and a developing synthesis of ecological community assembly theory. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366: 2403–2413.

Wiens B.J.J. (2011): The causes of species richness patterns across space, time, and clades and the role of "ecological limits." *The Quarterly Review of Biology*, 86: 75–96.

Wiens J.J. & Donoghue M.J. (2004): Historical biogeography, ecology and species richness. *Trends in Ecology & Evolution*, 19: 639–644.

Wilson D. (1995): Endophyte: the evolution of a term, and clarification of its use and definition. *Oikos*, 73: 274–276.

Winkworth R.C., Wagstaff S.J., Glenny D. & Lockhart P.J. (2002): Plant dispersal N.E.W.S from New Zealand. *Trends in Ecology & Evolution*, 17: 514–520.

Wüest R.O., Antonelli A., Zimmermann N.E. & Linder H.P. (2015): Available climate regimes drive niche diversification during range expansion. *The American Naturalist*, 185: 640–652.

Yoder A.D. & Nowak M.D. (2006): Has vicariance or dispersal been the predominant biogeographic force in Madagascar? Only time will tell. *Annual Review of Ecology, Evolution, and Systematics*, 37: 405–431.

Zachos J.C., Dickens G.R. & Zeebe R.E. (2008): An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature*, 451: 279.

Zimmerman N.B. & Vitousek P.M. (2012): Fungal endophyte communities reflect environmental structuring across a Hawaiian landscape. *Proceedings of the National Academy of Sciences*, 109: 13022–13027.

2 Chapter I: Factors underlying the global dispersal of temperate Loliinae grasses[1]

2.1 Abstract

Dispersal of lineages depends on specific ecological requirements as well as on generalised factors such as distance and barriers. The temperate grass subtribe Loliinae (Poaceae: Poeae) has two main subgroups, broad-leaved (BL) and fine-leaved (FL) Loliinae, that differ in phenotypic syndrome and ecology and have both widespread, disjunct distributions in the northern and southern hemispheres. Previous studies have inferred phylogenetic relationships in this group and developed scenarios for its historical biogeography. Here, I used parametric models of range evolution combined with stochastic mapping to test the influence of distance, disjunction type, and phenotypic differences on dispersal rates. A model integrating founder events and scaling of dispersal by shortest distance between areas performed best amongst a set of alternative models and recovered a mean of 83 dispersal events. Overall dispersal rates were significantly higher in BL than in FL. Per-route dispersal rates showed a significant negative exponential relationship to shortest distance but were not affected by phenotypic syndrome or disjunction type. I conclude that Loliinae originated in the northern hemisphere, in agreement with previous studies, and evolved through recurrent range expansions and founder-event dispersals. Higher competitive ability, potentially related to the broad-leaved syndrome (i.e. tall, strongly rhizomatous plants, long-living individuals, occupancy of more stable habitats), may explain higher inferred dispersal rates in BL compared with FL Loliinae. However, the dominant factor impacting dispersal in both BL and FL Loliinae is the distance between suitable areas.

2.2 Introduction

Lineage dispersal across large distances is seen as a major process in the evolution of organismal diversity and distribution, especially in plants (Nathan 2006, Crisp et al. 2011). Studies using molecular phylogenetics have revealed its importance especially in the southern hemisphere, where dispersal occurring well after the split-up of the supercontinent has frequently been recovered (e.g. Sanmartín & Ronquist 2004). In plant groups with a global distribution, dispersal clearly played a key role, especially where plants are adapted to conditions found in widely disjunct regions, such as a temperate climate (e.g. Blattner 2006, Linder et al. 2013). Indeed, biome compatibility is thought to be a major determinant of dispersal (Donoghue 2008). However, factors independent of the species' ecology, such as the directionality of potential dispersal vectors (e.g. winds or ocean currents (e.g. Renner 2004, Sanmartín et al. 2007) or the distance between suitable areas (e.g. Linder et al. 2013, Van Dam & Matzke 2016) have also been shown to impact the frequency of dispersals.

1 The results of this chapter have contributed to the work led by P. Catalán and developed by M. Minaya (University of Zaragoza, doctoral dissertation 2015) and have been published in an article entitled "Contrasting dispersal histories of broad- and fine-leaved temperate Loliinae grasses: range expansion, founder events, and the roles of distance and barriers" in *Journal of Biogeography* (2017, vol. 44: pages 1980-1993, DOI: 10.1111/jbi.13012), authored by M. Minaya, J. Hackel (co-first author), M. Namaganda, C. Brochmann, M. S. Vorontsova, G. Besnard and P. Catalán. M.M., M.N. and P.C. originally devised the study, produced molecular data, performed phylogenetic and biogeographic analyses. C.B., G.B., M.N., M.S.V. and P.C. collected samples. My contribution was an analysis of range evolution with models integrating distance and founder-event speciation, stochastic mapping, and statistical analysis of dispersal rates (all presented here), and the writing of the corresponding parts of the article.

The almost cosmopolitan grass family (Poaceae) includes clades of grasses adapted to temperate climates and found in widely separated areas of the northern and southern hemisphere. The subtribe Loliinae (Pooideae: Poeae) is one of the largest among these, with over 600 species in the polyphyletic fescues (*Festuca*) and allied genera (Catalán 2006). These are found mainly in the northern hemisphere, but also in the temperate southern hemisphere (e.g. Patagonia and Australasia) and mountain ranges at tropical latitudes (e.g. the Andes and the East African Rift mountains) (Catalán 2006). Phylogenetic studies showed that the subtribe includes two main lineages encompassing the majority of species, the "broad-leaved" (BL) and the "fine-leaved" (FL) Loliinae, and a smaller "intermediate" lineage (Catalán et al. 2007, Inda et al. 2008). Their names reflect their phenotypes: BL Loliinae are mostly robust, perennial, and with a well developed rhizome, while FL Loliinae tend to be more slender, mostly annual, and without vigourous rhizome (Catalán et al. 2006). The "intermediate" lineage has a BL phenotypic syndrome, but is sister to the FL lineage, which together are sister to the BL Loliinae (Catalán et al. 2007, Inda et al. 2008). Phenotypic differences correspond to ecology, with BL tending to be more competitive and preferring mesic, stable habitats, while FL are rather colonist and found under harsher or more unstable conditions (Catalán et al. 2006). FL Loliinae also have higher rates of molecular evolution (Catalán et al. 2006).

Inda et al. (2008) performed the first historical biogeographic analysis of Loliinae. They estimated a crown age of around 13 million years (Ma) for the subtribe and a dispersal-dominated history with a Eurasian origin of both BL and FL Loliinae. The dissertation of Minaya (2015) extended this work, including a more comprehensive sample of species with many southern hemisphere taxa previously unsampled. Minaya recovered five new southern hemisphere lineages within Loliinae, three in BL and two in FL. He used a new set of macroscopic and phytolith fossils for calibration and estimated an age considerably older that that of Inda et al. (22.5 Ma). Implementing the dispersal–extinction–cladogenesis model (Ree & Smith 2008), he estimated additional dispersals: notably, long-distance dispersal events from South America to Tropical Africa, from Eurasia to Australasia, across the Antarctic between South America and Australasia, and between the Arctic and the Antarctic, were inferred. Minaya also suggested that BL were more successful at colonising adjacent areas, while FL more frequently crossed large distances.

Dispersal can be affected by the ecology of the species as well as unrelated, generalised factors such as distance. New parametric methods in biogeography allow to statistically test the importance of such factors (Van Dam & Matzke 2016). They can also include founder-event speciation, typical for island-like scenarios, and distinguish it from range expansion (Matzke 2014). In this chapter, my goal was to build on the work of Minaya (2015) by applying these recently developed models of range evolution. My main questions were the following: (i) Which role did distance and different modes of dispersal play in the biogeographic history of Loliinae? (ii) Which type of distance (centre vs. shortest) between areas of distribution would best explain lineage dispersal? (iii) Do dispersal rates differ between BL and FL Loliinae, and across oceanic vs. terrestrial barriers?

2.3 Materials and methods

Analysis of range evolution

I used the dated Loliinae phylogeny of Minaya (2015), which was estimated from molecular data both from the nuclear (internal transcribed spacer region) and the chloroplast genome (*trnT–L* and *trnL–F* spacers) and dated with fossil calibration points. It includes 214 accessions of 178 species representing the global distribution of the subtribe (Minaya 2015). Twelve operational areas (OAs) as defined by Minaya were used for range evolution analysis: South Africa, Madagascar & Mascarenes, Tropical Africa, Mediterranean, Irano−Turanian−Himalayan, North Eurasia, Australasia North & Central America, Northern South America, Southern South America, Macaronesia and Hawaii.

Maximum Likelihood analysis of range evolution in the *BioGeoBEARS* package (Matzke 2013) under *R* version 3.3.1 (R Core Team 2016), was used to implement the dispersal–extinction–cladogenesis (DEC) model of Ree & Smith (2008) and extensions thereof. I compared the basic DEC model, which assumes range expansion on branches as the only mode of dispersal, to models adding founder-event dispersal at speciation (DEC+*j*; Matzke 2014) or a distance-scaling parameter x (DEC+x; Van Dam & Matzke 2016), and models integrating both parameters (DEC+*j*+*x*). For the distance-dependent models (DEC+x, DEC+j+x), I tested both approximate pairwise shortest distances between OAs and pairwise distances between centres of each OA, calculated as Vincenty ellipsoid distances in the *R* package *geosphere* (Hijmans 2016) (Table A-I.1, annex). I assumed that the two types of distance roughly represent geographical distances through the time period studied (Late Oligocene to present).

A total of six models (DEC, DEC+*j*, DEC+*x* shortest distances, DEC+*x* centre distances, DEC+*j*+*x* shortest distances, DEC+*j*+*x* centre distances, see Table I.1) were thus optimised and compared, using the maximum clade credibility tree obtained by Minaya (2015) with outgroups removed. Only one accession was kept per species, except in the paraphyletic *Festuca abyssinica* that was represented by three accessions; exclusion of any accession here is unlikely to modify the ancestral areas estimated as all the species of the clade are distributed in the same OA (Tropical Africa, see Results). Two areas were assumed to be the maximum range size, based on current distribution patterns (Sanmartín 2003). I compared model performance with the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) and estimated relative likelihoods for ancestral areas at nodes under all models.

Dispersal counts and rates

To infer the number and rates of dispersals, I used stochastic mapping in the *BioGeoBEARS* package (Matzke 2016). I ran 1000 simulations for both the best-performing model (DEC+*j*+*x* with shortest distances, see Results) and the baseline DEC model and recovered the number of range expansions and founder-event dispersals for each branch and node, respectively. This allowed me to calculate and compare per-lineage dispersal rates for BL and FL Loliinae, by

dividing the number of dispersals by median stem age and the number of tips per lineage. For BL, I calculated dispersal rates for both the main BL clade alone and for the BL + "intermediate" clade, in the latter case dividing by the older BL stem age.

Dispersal rates (range expansion, founder-event, and total) were also calculated for each route, in each direction, and for BL (including the "intermediate" clade) and FL, yielding N = 264 data points (most of them zero). Quasi-Poisson regression in R, with a logarithmic link function, was used to test the relationship between dispersal rates and distance, using either shortest or centre distance as predictors, as well as lineage (BL or FL) and disjunction type (terrestrial or oceanic). For each dispersal rate and range evolution model, I fitted both a global regression model with distance, lineage, disjunction type and their interactions as predictors and a reduced model with distance as sole predictor.

2.4 Results

The DEC+*j*+*x* model of range evolution with shortest distances clearly outperformed the other five models in ML estimation, accounting for >0.99 of AIC and BIC weights (Table I.1). Figure 1 shows the relative likelihoods of ancestral areas at each node under this model. Estimations were similar for other models, but these assumed more widespread ancestral ranges when founder-event dispersal was excluded (i.e. under DEC and the DEC+*x* models, not shown).

Table I.1. Performance of six models of range evolution tested on the Loliinae phylogeny inferred by Minaya (2015), for 12 operational areas and a maximum range size of two. Estimated parameters: *d*, range expansion rate; *e*, extinction rate; *j*, relative weight for founder-event dispersal at node; *x*, distance-scaling parameter. P., number of parameters, logL, log(likelihood); AIC, Akaike Information Criterion; wAIC, Akaike weight; BIC, Bayesian Information Criterion; wBIC, Schwarz weight.

Range evolution model	***d***	***e***	***j***	***x***	**P.**	**logL**	**AIC**	**ΔAIC**	**wAIC**	**BIC**	**ΔBIC**	**wBIC**
DEC+*j*+*x* (shortest distances)	0.004	<0.001	0.037	-0.227	4	-366.24	740	0	>0.99	753	0	>0.99
DEC+*j*+*x* (centre distances)	0.009	<0.001	0.067	-0.896	4	-380	768	28	<0.001	780	28	<0.001
DEC+*j*	0.003	<0.001	0.024	0	3	-392.73	792	51	<0.001	801	48	<0.001
DEC+*x* (shortest distances)	0.013	0.015	0	-0.212	3	-421.09	848	108	<0.001	858	105	<0.001
DEC+*x* (centre distances)	0.035	0.014	0	-1.176	3	-430.7	867	127	<0.001	877	124	<0.001
DEC	0.01	0.012	0	0	2	-442.57	889	149	<0.001	895	142	<0.001

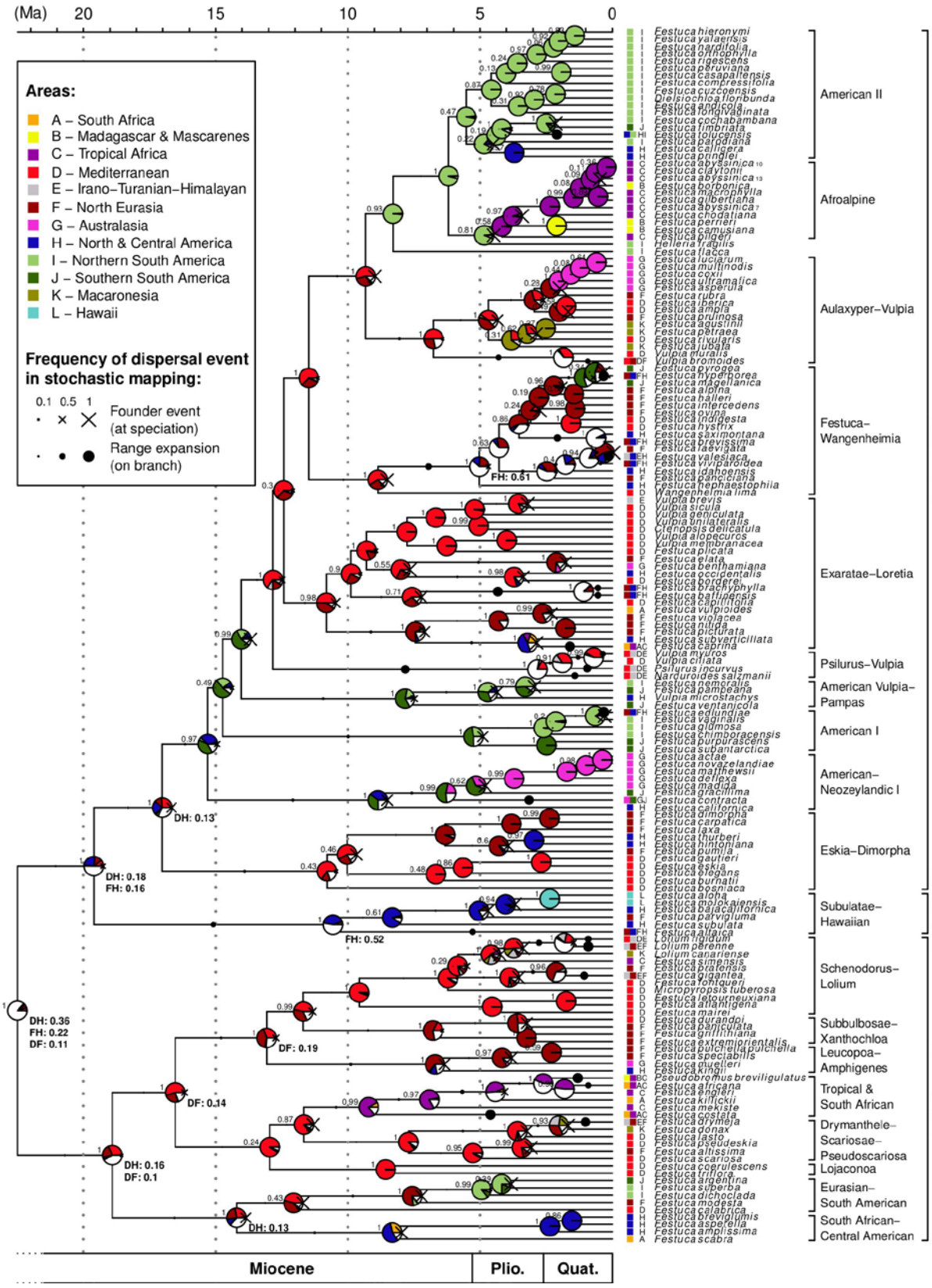
(Ma)
20
15
10
5
0
Areas:
A – South Africa
B – Madagascar & Mascarenes
C – Tropical Africa
D – Mediterranean
E – Irano–Turanian–Himalayan
F – North Eurasia
G – Australasia
H – North & Central America
I – Northern South America
J – Southern South America
K – Macaronesia
L – Hawaii
Frequency of dispersal event in stochastic mapping:
0.1 0.5 1
Founder event (at speciation)
Range expansion (on branch)
DH: 0.35
FH: 0.22
DF: 0.11
DH: 0.18
FH: 0.16
DH: 0.13
FH: 0.52
DF: 0.19
DF: 0.14
DH: 0.16
DF: 0.1
DH: 0.13
FH: 0.61
Miocene
Plio.
Quat.
American II
Afroalpine
Aulaxyper–Vulpia
Festuca–Wangenheimia
Exaratae–Loretia
Psilurus–Vulpia
American Vulpia–Pampas
American I
American–Neozeylandic I
Eskia–Dimorpha
Subulatae–Hawaiian
Schenodorus–Lolium
Subbulbosae–Xanthochloa
Leucopoa–Amphigenes
Tropical & South African
Drymanthele–Scariosae–Pseudoscariosa
Lojaconoa
Eurasian–South American
South African–Central American
Fine–leaved
Intermediate
Broad–leaved

Figure I.1 (previous page). Estimated ancestral ranges of Loliinae, obtained under the best-performing DEC+*j*+*x* model of range evolution for shortest distances between areas computed with BioGeoBEARS, mapped on the time-calibrated phylogenetic tree inferred by Minaya (2015). Posterior probability values are indicated at nodes. Pie charts at nodes represent Maximum Likelihood relative marginal probabilities of ancestral ranges, with their colour legend given at the insert chart (only single areas are coloured; white corresponds to two-area ranges). Letters at basal nodes indicate two-area ranges with the highest marginal probabilities. The upper time scale bar represents million years ago (Ma). Geological time scale is shown at the bottom. OAs assigned to each selected species (A to L) and the major lineages are indicated to the right of the tree. The inferred range expansions (black dots) and founder events (crosses) are indicated on branches and at nodes of the tree, respectively, with symbol size referring to the frequency of an event inferred in 1000 simulations.

A mean of 83 dispersal events (95% confidence interval [CI] 77–88) was recovered for Loliinae with stochastic mapping under the best model, DEC+*j*+*x* with shortest distances, compared to 72 (CI 67–77) under the baseline DEC model (Fig. I.2a). DEC+*j*+*x* allowed to distinguish between range expansion and founder-event dispersal, of which 28 (CI 24–32) and 55 (CI 48–62) events were estimated, respectively (Fig. I.2a). Range expansion, founder-event, and total dispersal rates were significantly lower for FL than for BL Loliinae, both under DEC and DEC+*j*+*x* and regardless of whether the intermediate clade was included in BL or not (Mann-Whitney U test, $P < 0.001$ in all cases; Fig I.2b).

Per-route dispersal rates showed a significant relationship to the shortest distance between OAs, and a negative exponential curve provided good fit under a quasi-Poisson regression model for range expansion, founder-event, and total dispersal rates under DEC and DEC+*j*+*x* ($P < 0.001$ for shortest distance in all cases; Fig. I.3; Table A-I.2). The estimated slope was more negative (log-scale factors $–5.3\times10^{–4} \pm$ SE $8\times10^{–5}$ vs. $–3.3\times10^{–4}$ SE $4.1\times10^{–5}$) for range expansion (Fig. I.3b) versus founder-event (Fig. I.3c) dispersal rates under DEC+*j*+*x*, indicating a stronger drop-off by distance. When shortest distance was taken into account, phenotypic syndrome (BL or FL), disjunction type (oceanic or terrestrial), and interactions had no significant effects on per-route dispersal rates ($P > 0.05$ in all cases; Table A-I.2). In regression on distance between OA centres, the results were less clear: no variable was significant, when tested together, for founder-event dispersal rates (DEC+*j*+*x*), whilst only disjunction type was significant for range expansion and total dispersal rates (Table AI.2, Fig. A-I.1). Centre distance was significant in all cases when tested alone.

2.5 Discussion

Historical biogeography of Loliinae

The biogeographical analyses of the Loliinae suggest recurrent colonizations through range expansions and founder-event dispersals that predominantly occurred within the northern hemisphere and from the northern to the southern hemisphere, but also within the southern hemisphere, and from S. America and S. Africa to N. and C. America (Fig. I.1). This is concordant with the complex scenario of neo- and re-colonisations suggested by Minaya (2015). Similar long-distance lineage dispersal patterns have also been suggested for other temperate pooids (*Hordeum*, Blattner 2006) and austral-temperate grasses (danthonioids, Linder et al. 2013). The strong support found here for a range evolution model including

founder-event speciation, typical of island clades (Matzke 2014), is in line with the picture of the cool-temperate Loliinae lineages moving about between suitable, often very distant areas. In agreement with Inda et al. (2008) and Minaya (2015), my biogeographical analysis lends support to an ancestral northern hemisphere distribution of Loliinae, but with uncertainty between Northern Eurasia, the Mediterranean region, North America, or a widespread distribution (Fig. I.1).

Dispersal rates: distance, barriers and differences between broad- and fine-leaved Loliinae

The analysis of range evolution selected a model including scaling by shortest distance between areas as the best one (Table I.1). Shortest distance thus seems to provide a reasonable approximation of the true dispersal corridors, which likely depend on complex combinations

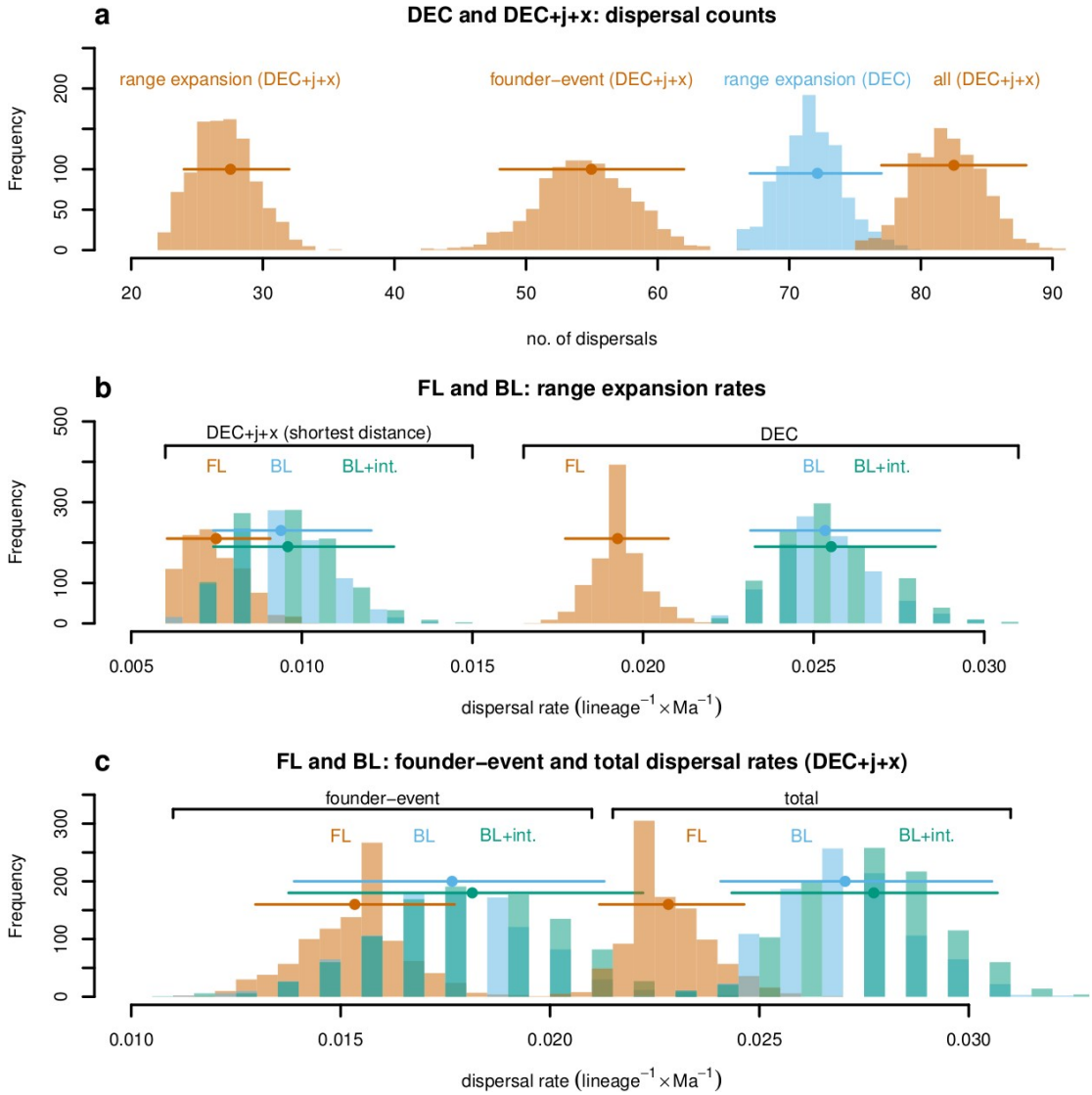

Figure I.2. Histograms of dispersal counts and rates for broad-leaved (BL, with and without Intermediate) and fine-leaved (FL) Loliinae, inferred through stochastic mapping. The DEC model, not including a distance effect, and a DEC+*j*+*x* model scaling dispersal rates by shortest distance were both used for 1000 simulations on tree inferred by Minaya (2015). (a) Global dispersal counts under DEC and DEC+*j*+*x*; (b) range expansion rates under DEC and DEC+*j*+*x*; and (c) founder-event and total dispersal rates under DEC+*j*+*x*.

of wind direction, ocean currents, animal migration routes, and suitable habitat patches. The regression of per-route dispersal rates, inferred through stochastic mapping, against shortest distance also shows a good fit with a negative exponential distribution. Only if one excludes the largest distances (>14,000 km) where dispersal is effectively zero, a fat-tailed dispersal kernel (Nathan 2006) may be distinguished. Overall, distance appears to be the major determinant of successful dispersal, while the type of disjunction (oceanic vs. terrestrial) was not supported as significant when shortest distance was included as predictor.

I recovered higher overall dispersal rates for BL than FL Loliinae under both the best DEC+*j*+*x* and the simplest DEC models. In contrast, per-route dispersal rates did not reveal a

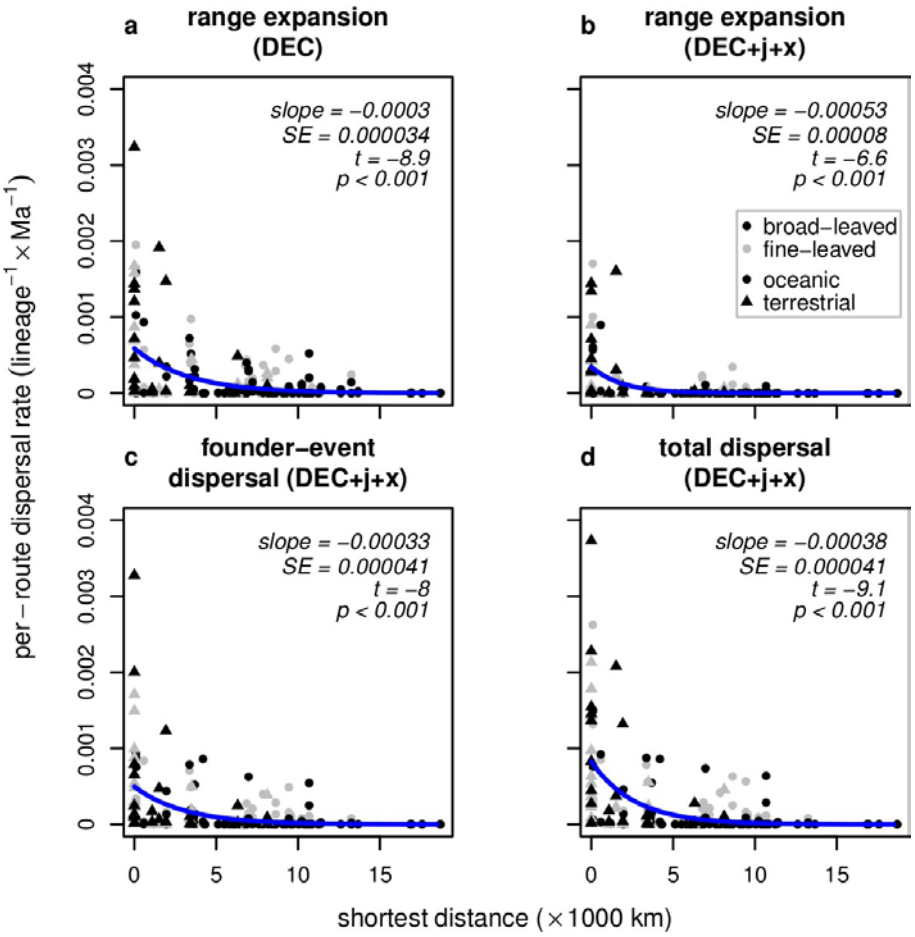

Figure I.3. Mean per-route dispersal rates in Loliinae, inferred with biogeographical stochastic mapping, plotted against the shortest distance between respective OAs. The DEC model, not including a distance effect, and a DEC+*j*+*x* model, scaling dispersal rates by shortest distance, were both used for 1000 simulations on the phylogenetic tree inferred by Minaya (2015). Dispersal rates were calculated per route, for each direction, for BL (including Intermediate; black) and FL (grey) Loliinae lineages, yielding 264 data points. Quasi-Poisson regression of dispersal rates on shortest distance showed a significant negative relationship for each range evolution model and type of dispersal (regression curve in blue, slope value is log-scale). X-axis shows geographical distance in km. Phenotypic syndrome (broad- vs. fine-leaved) and disjunction type (oceanic vs. terrestrial) had no significant effect on dispersal rates when distance was accounted for. (a) Range expansion rates (DEC); (b) range expansion rates (DEC+*j*+*x*); (c) founder-event dispersal rates (DEC+*j*+*x*); and (d) total dispersal rates (DEC+*j*+*x*). Regression on centre distances showed similar results (Fig. A-I.1, annex).

significant difference between BL and FL Loliinae once distance was taken into account. BL (including the intermediate lineage) are present in all 12 studied OAs (Fig. I.1, Minaya 2015), while the more speciose FL are only present in 11 OAs and are also proportionally less represented in some of them (e.g. A, Fig. I.1, Minaya 2015). BL may thus have successfully dispersed to more different OAs, while FL radiated more extensively in some OAs (Eurasia [D, E, F] and South America [I, J]).

The phenotypic syndromes of BL and FL Loliinae are associated with different evolutionary rates and adaptive capabilities (Catalán, 2006, Catalán et al., 2006). Perennial Loliinae show significantly lower molecular evolutionary rates than annual Loliinae (Catalán et al. 2006). Relative rate ratios support the minimum-generation-time hypothesis, implying the predominance of stabilizing selection in the mostly strong-rhizomatous perennial BL Loliinae versus more rapid adaptive evolution in the slender perennial and annual FL Loliinae (Catalán 2006, Catalán et al. 2006). This pattern is likely a consequence of the colonist syndrome of the FL Loliinae, which are, in general, better adapted to cool and dry ecological conditions than the BL Loliinae (Catalán 2006, Catalán et al. 2007). The colonist syndrome of the FL Loliineae, together with the existence of suitable habitats along the Andean corridor, Patagonia and the circum-Antarctic islands, might also explain why the FL lineages have spread more frequently in S. America and the South Pacific, but not in Africa, where BL ancestors have established more frequently in less abundant, but more stable habitats (Minaya 2015). This asymmetric dispersal pattern has resulted in the FL and BL Loliinae having, respectively, five and two lineages with most of their current species distributed in the southern hemisphere (Fig. I.1). Given relatively similar dispersal conditions for both lineages, BL Loliinae may be more competitive, both by being strongly rhizomatous, long-living plants, and by being adapted to more stable habitats than the more vagile FL Loliinae, which are less robust, short-living plants and usually adapted to less stable habitats. On the other hand, dispersal limitation by distance appears to be strong in both BL and FL Loliineae and would explain why no difference in dispersal rate per route was found between the two phenotypes once distance was taken into account.

Conclusion

This biogeographical analysis infers that the biogeographical history of the Loliinae began in the northern hemisphere and was shaped by frequent dispersals, both through founder events and through range expansion. Distance emerges as the overall most important factor in determining successful dispersal, whether across oceans or continents, with shortest distance between areas showing the best fit with the phylogenetic tree and the estimated dispersal rates. BL Loliinae appear more successful in the overall dispersal rates than FL Loliinae, presumably because of a more competitive life strategy. Nevertheless, dispersal per route is limited by distance in a similar fashion for both BL and FL Loliinae.

2.6 References

Blattner F.R. (2006): Multiple intercontinental dispersals shaped the distribution area of *Hordeum* (Poaceae). *New Phytologist*, 169: 603–614.

Catalán P. (2006): Phylogeny and evolution of *Festuca* L. and related genera of subtribe Loliinae (Poeae, Poaceae). *Plant genome: biodiversity and evolution*, 1: 255–303.

Catalán P., Torrecilla P., Lopez Rodriguez J. & Muller J. (2006): Molecular evolutionary rates shed new light on the relationships of *Festuca*, *Lolium*, *Vulpia* and related grasses (Loliinae, Pooideae, Poaceae). *Current taxonomic research on the British and European flora* (ed. by J. Bailey & R.G. Ellis): 45–70. Botanical Society of the British Isles, London, UK.

Catalán P., Torrecilla P., López-Rodríguez J., Müller J. & Stace C. (2007): A systematic approach to subtribe Loliinae (Poaceae: Pooideae) based on phylogenetic evidence. *Aliso: A Journal of Systematic and Evolutionary Botany*, 23: 380–405. Crisp M.D., Trewick S.A. & Cook L.G. (2011): Hypothesis testing in biogeography. *Trends in Ecology & Evolution*, 26: 66–72.

Donoghue M.J. (2008): A phylogenetic perspective on the distribution of plant diversity. *Proceedings of the National Academy of Sciences*, 105: 11549–11555.

Hijmans R.J. (2016): geosphere: Spherical Trigonometry. R package version 1.5-5.

Inda L.A., Segarra-Moragues J.G., Müller J., Peterson P.M. & Catalán P. (2008): Dated historical biogeography of the temperate Loliinae (Poaceae, Pooideae) grasses in the northern and southern hemispheres. *Molecular Phylogenetics and Evolution*, 46: 932–957.

Linder H.P., Antonelli A., Humphreys A.M., Pirie M.D. & Wüest R.O. (2013): What determines biogeographical ranges? Historical wanderings and ecological constraints in the danthonioid grasses. *Journal of Biogeography*, 40: 821–834.

Matzke N.J. (2013): *BioGeoBEARS: BioGeography with Bayesian (and Likelihood) Evolutionary Analysis in R Scripts.* University of California, Berkeley, Berkeley, CA.

Matzke N.J. (2014): Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Systematic Biology*, 63: 951–970.

Matzke N.J. (2016): Stochastic mapping under biogeographical models. Available at: http://phylo.wikidot.com/biogeobears#stochastic_mapping.

Minaya M. (2015): Phylogeny and biogeography of the temperate Loliinae grasses in the southern hemisphere. *A multigenomic approach to the phylogeny, evolution and biogeography of the grass subfamily Pooideae with an emphasis on the subtribe Loliinae*: 239–322. Doctoral dissertation, Universidad de Zaragoza, Spain.

Nathan R. (2006): Long-distance dispersal of plants. *Science*, 313: 786–788.

R Core Team (2016): *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ree R.H. & Smith S.A. (2008): Maximum Likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology*, 57: 4–14.

Renner S. (2004): Plant dispersal across the tropical Atlantic by wind and sea currents. *International Journal of Plant Sciences*, 165: S23–S33.

Sanmartín I. (2003): Dispersal vs. vicariance in the Mediterranean: historical biogeography of the Palearctic Pachydeminae (Coleoptera, Scarabaeoidea). *Journal of Biogeography*, 30: 1883–1897.

Sanmartín I. & Ronquist F. (2004): Southern hemisphere biogeography inferred by event-based models: plant versus animal patterns. *Systematic Biology*, 53: 216–243.

Sanmartín I., Wanntorp L. & Winkworth R.C. (2007): West Wind Drift revisited: testing for directional dispersal in the Southern Hemisphere using event-based tree fitting. *Journal of Biogeography*, 34: 398–416.

Van Dam M.H. & Matzke N.J. (2016): Evaluating the influence of connectivity and distance on biogeographical patterns in the south-western deserts of North America. *Journal of Biogeography*, 43: 1514–1532.

3 Chapter II: Grass diversification in Madagascar[1]

3.1 Abstract

Grasses (Poaceae) are found in all major habitats of Madagascar and have a particular importance in C_4 grasslands, whose origins are controversial. We aimed to estimate the number, age and origins of endemic grass lineages in the Madagascar region, and to compare the diversification of C_3 and C_4 taxa. We estimated 11 time-calibrated molecular phylogenies including 73% of Madagascar's known grass flora (65% of endemics), using two calibration scenarios. Integrating the available sequences from worldwide grass species, a total of 1928 accessions were analysed. We tested range evolution models, estimated ancestral ranges, and compared the patterns of lineage accumulation between endemic C_3 and C_4 grasses. We recovered 69 lineages endemic to or with an estimated origin in the Madagascar region, 25 of them C_3 and 44 C_4. Range evolution analysis suggests widespread distance-scaling of dispersal and strongest historical links to Africa. Extant grass diversity largely accumulated since the Miocene, with parallel increases in C_3 and C_4 taxa. Two large C_3 groups in the "Forest shade clade" (Paniceae: Boivinellinae) and the bamboos (subtribe Hickeliinae) have an estimated origin in the Madagascar region. Divergences and crown ages of endemic C_4 lineages largely coincide with the Miocene grassland expansion. We conclude that Madagascar's extant grass flora is the result of multiple overseas dispersals, predominantly from Africa, and diversified from the Miocene onwards. C_3 grasses are characterised by two large presumed *in situ* radiations of shade grasses in the Paniceae and bamboos. Endemic C_4 lineages result from twice as many immigration events, resulting in smaller clades. Ages of C_4 lineages are consistent with a Pliocene or Late Miocene origin of grasslands in Madagascar, but estimating the nature and expanse of such early grasslands will require further research.

3.2 Introduction

Madagascar's biodiversity is exceptional in its richness and endemism, and understanding the assembly of this island biota has been a major goal of biogeography. Molecular phylogenetics and time calibration have shown that despite Madagascar's long isolation (from mainland Africa since around 140 Ma and India since 88 Ma), the overwhelming majority of current-day endemics appear to be "neo-endemics" resulting from overseas immigration rather than ancient Gondwanan vicariance (Yoder & Nowak 2006, Buerki et al. 2013). This has refined ideas on the "evolution in isolation" of Malagasy lineages, yielding the picture of a biota more dynamic and, in many cases, younger than assumed before.

The flora of Madagascar, with over 80% endemism (Madagascar Catalogue 2013), is an assemblage with mixed affinity. It has the strongest taxonomic links with continental Africa, and with the surrounding West Indian Ocean islands of the Comoros, Seychelles, and Mascarenes, which may be considered part of a wider Madagascar region (Buerki et al. 2013). There are also evident links to the floras of India and Southeast Asia (Schatz 1996, Buerki et

1 This chapter was accepted for publication in *Journal of Biogeography*, pending final corrections and in slightly different format, in October 2017, under the title "Grass diversification in Madagascar: *in situ* radiation of two large C_3 shade clades and support for a Miocene to Pliocene origin of C_4 grassy biomes", co-authored by J. Hackel, M. S. Vorontsova, O. P. Nanjarisoa, R. C. Hall, J. Razanatsoa, P. Malakasi and G. Besnard. G.B., M.S.V., and myself devised the study. M.S.V., O.P.N., R.C.H. and G.B. collected samples. G.B. and P.M. produced molecular data. My contribution was the phylogenetic and biogeographic analysis and the writing of the manuscript.

al. 2013). Within Madagascar, a large diversity of habitats and vegetation types, ranging from high-mountain temperate vegetation to tropical humid forests and sub-arid spiny thickets (Humbert 1927, Moat & Smith 2007) may partly explain the exceptional levels of taxonomic richness.

Grasses (Poaceae) are one of the largest plant families in Madagascar, with roughly 550 recorded species (Vorontsova et al. 2016). They occur in all major terrestrial habitats. The proportion of endemic grass species is 40%, which is lower than that of most other plant families in Madagascar but higher than the average grass endemism for islands of similar size (Vorontsova et al. 2016). The age of the family is still a matter of debate, but different methods of time calibration and chloroplast vs. nuclear data place the diversification of its large crown group (the BOP–PACMAD clade) in the Late Cretaceous to Early Eocene, after Gondwana break-up (Christin et al. 2014). The cosmopolitan distribution of grasses must therefore be the result of recurrent dispersal between continents. Bio- and phylogeographic work has shown that the availability of suitable climatic conditions played a major role in allowing grasses to successfully colonise new continents (Linder et al. 2013, Visser et al. 2014, Lundgren et al. 2015). The multiple evolution of C_4 photosynthesis in grasses, first occurring in the Oligocene (Christin et al. 2008, Vicentini et al. 2008), and the expansion of grasslands in the Miocene were major drivers of grass diversification (Spriggs et al. 2014). This global Miocene grassland expansion occurred roughly between 3 and 8 Ma, but the exact timing differed between continents. It strongly modified biomes in the tropics, where C_4 grasses became dominant in open areas (Edwards et al. 2010). C_3 grasses, outcompeted under warm, high-light conditions, are mainly restricted to more humid and cooler microclimates in Madagascar and other tropical regions (Nanjarisoa et al. 2017, Edwards & Still 2008).

C_4 grasses, keystone species in open areas, are at the heart of the debate on the origin of the grasslands of Madagascar. These are now the dominant vegetation type on the island, in particular on the central plateaus (Moat & Smith 2007). Many authors interpreted these grasslands as secondary, created through cutting, grazing and fire after human colonization (Perrier de la Bâthie 1921, Humbert 1927, Koechlin 1972). In contrast, several lines of evidence suggest that at least some grasslands may be ancient, primary vegetation. These include grassland endemics (Bond et al. 2008, Vorontsova et al. 2016), pre-human pollen and charcoal deposits indicative of open habitats with natural fires (Burney 1987, 1997, Gasse & Van Campo 1998) and *Sporormiella* deposits suggesting presence of megaherbivores (Burney et al. 2003). Under this scenario, grasslands in Madagascar would be a regional outcome of the global Miocene grassland expansion (Bond et al. 2008). Around 358 C_4 grass species, with roughly 100 endemics (28%), are known in Madagascar, and many are regionally restricted (Vorontsova et al. 2016). Phylogenetic diversity of grassland communities was found to be negatively associated with grazing and trampling but independent of fire frequency, suggesting an endemic, fire-adapted grass flora (Vorontsova et al. 2016). Endemic clades of C_4 grasses would also support a natural origin of grasslands, but few of the endemic C_4 grass species have yet been included in phylogenetic analyses.

C_3 grasses in Madagascar are overall less diverse but have a significantly higher proportion of endemics than C_4 grasses (roughly 120 of 175, 68%; see checklist in Vorontsova et al. 2016). They include at least one phylogenetically isolated lineage (Panicoideae: *Lecomtella*; Besnard et al. 2013) as well as endemic and diverse genera of forest grasses and bamboos (Dransfield 2003, Vorontsova & Rakotoarisoa 2014). C_3 habitats in Madagascar and surrounding islands include the high mountains, the forest understorey, and various sheltered locations under rocks or other plants, including other grasses. The higher endemism in C_3 grasses suggests that they may belong to lineages which had more time for diversification or, alternatively, higher diversification rates than C_4 grasses.

In this study, we aimed to infer the diversification history of grasses in Madagascar, using molecular phylogenetics and dating combined with range evolution analysis. We first estimated the number and ages of endemic grass lineages in Madagascar. Second, by estimating ancestral areas, we sought to uncover the main biogeographical links of the Madagascan grass flora and to test if the region was only a "sink" for immigrant taxa or if *in situ* diversification and dispersal to other regions also occurred. Third, we specifically compared endemic C_3 and C_4 lineages, to see whether the ages of C_4 taxa were in line with a Miocene expansion of grasslands, and to test whether the C_3 and C_4 grass floras differed in age or lineage accumulation dynamics.

3.3 Materials and Methods

Taxon sampling, molecular biology, and data retrieval

We aimed to analyse a set of molecular data representative for the grass flora of Madagascar and the West Indian Ocean region and included available as well as new molecular data[1]. New sequence data were produced by PCR and Sanger sequencing or by high-throughput sequencing (see Annex A-II for details). We used three to five chloroplast markers (*ndhF*, *rbcL*, and *trnK–matK*, GPWG II 2012, Taylor et al. 2012; adding the *rpl16* and *rps16* introns for the slow-evolving Bambusoideae, see A-II). Sequences were organised in 11 datasets corresponding to major grass clades (GPWG II 2012; Kellogg 2015; Soreng et al. 2015): Oryzoideae, Bambusoideae, Brachypodieae, Poeae, Aristidoideae, Panicoideae (focusing on the "outlying Panicoideae" and *Lecomtella*, representing the large Andropogoneae–Arundinelleae, Paniceae and Paspaleae clades with two taxa each), Paniceae, Andropogoneae–Arundinelleae, Arundinoideae–Micrairoideae, Chloridoideae, and Danthonioideae. For each clade, we added sequence data available from *GenBank*. In total, an estimated 73% of the known species diversity in Madagascar, and 65% of the endemics are represented.

1 Sequence data were submitted to *GenBank*. The accession numbers of these new sequences, along with voucher information, as well as those of the accessions retrieved from *GenBank* have not been included in the annex of this dissertation as the table is very large (>1,900 rows). They will be available with the online article as supplementary information.

Phylogenetics, dating, and analysis of endemic clades

DNA sequences were aligned by marker for each clade using *MAFFT* v.7 (Katoh & Standley 2013). After manual inspection and trimming of poorly aligned ends, alignments were concatenated by clade and contained between 18 (Brachypodieae) and 516 (Paniceae) accessions, for a total of 1928. Maximum Likelihood (ML) and Bayesian phylogenetic analyses were performed with *RAxML* (Stamatakis 2014) and *BEAST2* (Drummond et al. 2006, Bouckaert et al. 2014, Rambaut et al. 2014), respectively, on the Cipres platform (Miller et al. 2012). We used two alternative, secondary calibrations: ages based on external angiosperm calibration and ages accounting for controversial phytolith fossils (Christin et al. 2014, see Table A-II.2). As estimates of the age of the Madagascan grass flora, we obtained the median ages and 95% HPD intervals for stem and crown of strictly endemic lineages (i.e. containing only endemic taxa) recovered in the MCC trees under both age calibrations. Lineages were scored as C_3 or C_4, based on the checklist of Osborne et al. (2014, see Table A-II.3). Details on alignment, phylogenetic models, and dating priors can be found in Annex A-II.

Analysis of range evolution

ML analysis of range evolution with the package *BioGeoBEARS* (Matzke 2013) in R was used to estimate the dispersal history between the Madagascar region and seven other areas: Tropical Africa, Southern Africa, Eurasia plus North Africa, the Indian subcontinent, Southeast Asia plus Australasia, North America, and the Neotropics. Species distribution was scored according to the World Checklist of Selected Plant Families (WCSP; Clayton et al. 2016). For each clade, we compared four range evolution models: (i) the dispersal–extinction–cladogenesis (DEC) model (Ree & Smith 2008); (ii) a DEC model expanded with *j*, a relative weight for founder-event speciation (DEC+*j*; Matzke 2014); (iii) an expanded DEC model scaling dispersal probabilities by distance between areas to power *x* (DEC+*x*; Van Dam & Matzke 2016); and (iv) a model integrating both *j* and *x* (DEC+*j*+*x*).

The best-fitting model was used to estimate ancestral areas and identify lineages with an estimated origin in the Madagascar region (i.e. with a marginal probability >0.5 for the region). We summed marginal probabilities for areas ancestral to clades estimated present in the Madagascar region, reflecting the origins of jump dispersals resulting in speciation in Madagascar. We correlated these figures with distance to Madagascar, calculated as Vincenty ellipsoid distance (Hijmans 2016) between approximate centres. This was compared to current-day occurrence counts in other areas of species native but not endemic to Madagascar, reflecting recent range expansions to or from Madagascar. We also traced the accumulation of C_3 and C_4 grass lineages in the Madagascar region through combined speciation and immigration with a lineage-through-time analysis: using MCC trees with median ages under external age calibration, we counted the number of C_3 and C_4 lineages in each 1 Myr time bin and weighted them by their marginal probability to occur in the region. Details on range evolution analysis can be found in Appendix A-II; Table A-II.4 contains area distances.

3.4 Results

Across the 11 phylogenies we estimated, we recovered 96 lineages that include only taxa strictly endemic to the Madagascar region (Table II.1). After accounting for common origins suggested by range evolution analysis, these strictly endemic lineages cluster in 69 lineages with an estimated origin in the region. Most of these contain only endemics, but some also include non-endemic or non-Madagascan taxa. C_4 lineages account for roughly twice as many Madagascar origins as C_3 lineages (44 vs. 25, Table II.1). About half of the Madagascar origin lineages fall in subfamily Panicoideae. The 11 time-calibrated phylogenies are summarised in Fig. II.1[1].

Table II.1. Summary of Poaceae lineages of the Madagascar region recovered in this study. Strictly endemic lineages contain only taxa endemic to the region. Madagascar origin lineages are those with an origin in the region estimated through range evolution analysis. These contain one or more strictly endemic lineages and can also include non-endemic and non-Madagascan taxa. Tables A-II.3 and A-II.6 in Annex A-II give details about individual strictly endemic and Madagascar origin lineages, respectively.

	Total	Photosynthetic pathway		Subfamily		Single accessions
		C_3	C_4	Panicoideae	others	
Strictly endemic	96	45	51	50	46	62
Madagascar origin	69	25	44	31	38	40

The two alternative age calibrations yielded different estimates, but these were close for more recent divergences, where the difference was mostly 1–3 Ma between the two median age estimates (Fig. II.2). Stem divergence times of most strictly endemic lineages estimation based on external calibration were placed in the Neogene, with the exception of *Lecomtella* which may already have diverged in the Oligocene (median stem age 22.3 Ma, HPD interval 14.8–30.3 Ma; Fig. II.2). Under phytolith calibration, *Lecomtella* may have diverged in the Eocene (upper bound 38.4 Ma), and upper bounds for the isolated C_4 species *Aristida ambongensis* and *Eragrostis chabouisii* marginally fall into the Oligocene (26.6 and 26.3 Ma, respectively). Roughly half of the single endemic species diverged from their non-endemic sister taxon only in the Quaternary according to median stem ages. Median stem ages of the majority of endemic C_4 lineages are estimated after the start of the global Miocene grassland expansion, around 8 Ma, under both calibrations (Fig. II.2). Estimated median crown ages of five well-supported endemic C_4 clades fall in the Late Miocene to Pliocene (lineages 21, 61, 62, 91, and 93 in Fig. II.2). Among endemic C_3 clades, the oldest estimated crown ages fall well in the Miocene and belong to the "Forest Shade Clade" (Panicoideae: Paniceae: Boivinellinae; lineages 41 and 43). Further endemic C_3 clades have crown ages in the Late Miocene to Quaternary.

1 The detailed plots of the individual phylogenies with support values and ancestral area probabilities have not been included in the annex for formatting issues (the largest phylogeny includes 516 accessions). They will be available with the online article as supplementary information. All alignments and phylogenies are available on *TreeBase* (study number S21602, reviewer access: http://purl.org/phylo/treebase/phylows/study/TB2:S21602?x-access-code=f8c9b299773662ae3f7625bb51f0520e&format=html)

The DEC+*j*+*x* model of range evolution ranked highest by AICc in 8 of 11 clades analysed (Table S2.6). In all but two clades (Brachypodieae and Danthonioideae), models incorporating distance (DEC+*j*+*x* and DEC+*x*) together aggregated >0.99 of AICc weight. The curve of lineage accumulation in the Madagascar region (Fig. II.3) suggests that grass diversity has increased exponentially, largely since around 20 Ma under external calibration. In many of the individual clades, most of the increase occurred very recently, i.e. in the last 5 to 10 Ma. Lineage accumulation curves were almost identical under the baseline DEC model (not shown).

Sums of marginal area probabilities for the ancestors of clades from the Madagascar region are highest for Tropical Africa, followed by Southern Africa, and lowest for India (Fig. II.4). They are not correlated with distance to Madagascar, whether tested on the estimates from

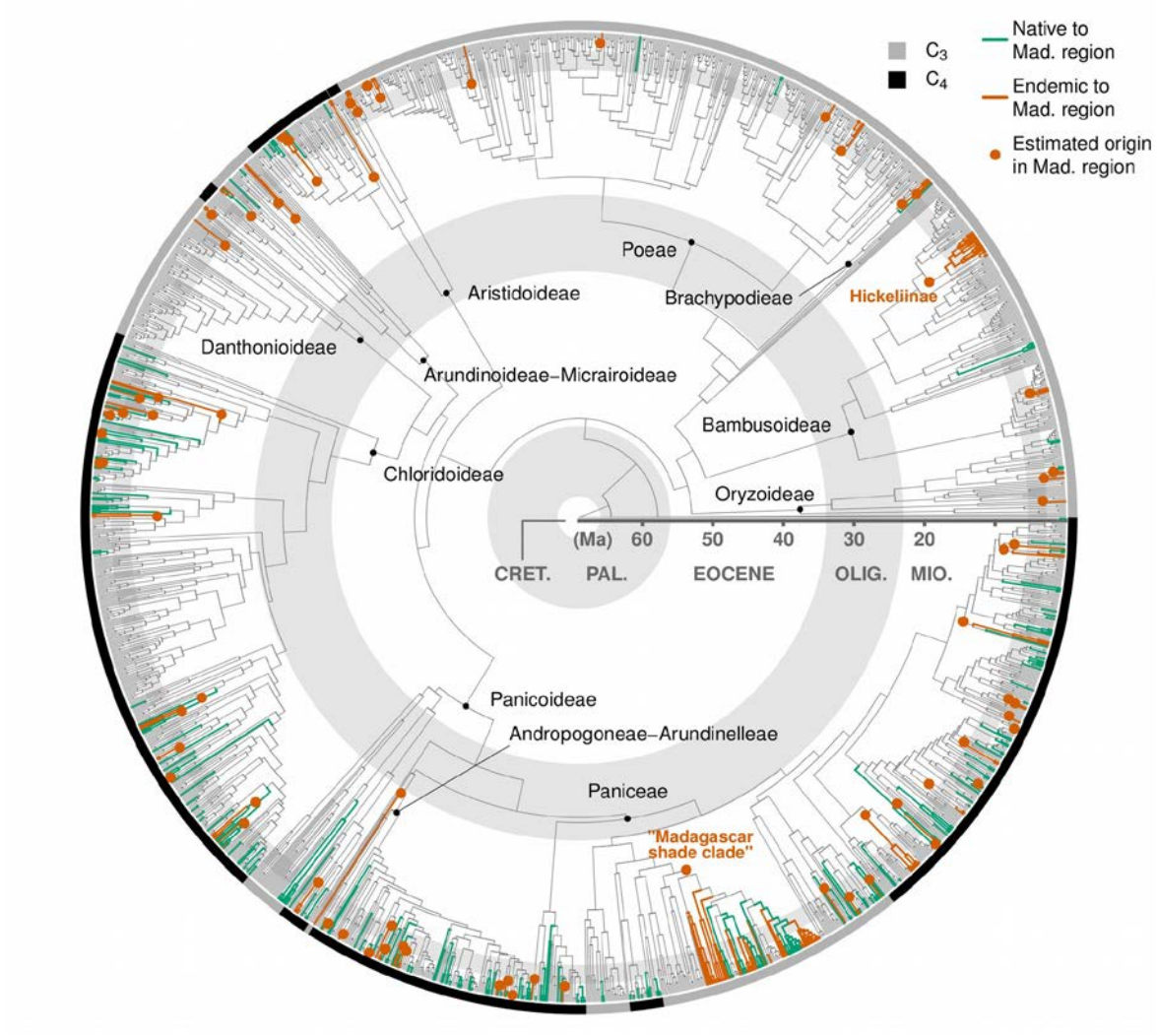

Figure II.1. Summary phylogenetic tree of the Poaceae, showing the 11 clades studied. Maximum Clade Credibility trees from *BEAST2* analyses, with median node heights under an external calibration without phytoliths, were grafted on a Poaceae backbone from Christin et al. (2014) for presentation. Lineages native to the Madagascar region in bluish green, endemic lineages in dark orange. The 69 lineages with an estimated origin in the Madagascar region are labelled with dark orange dots. Grey outer circle segments denote C_3, black segments C_4 taxa (*Alloteropsis* and Paspaleae are simplified as C_4 clades). Two large C_3 clades that diversified in the Madagascar region, the Hickeliinae and the "Madagascar shade clade", are labelled.

DEC or from the respective best model (one-sided Spearman test, $p = 0.48$ and $p = 0.13$, respectively). Tropical Africa, North America, and the Neotropics were mainly estimated as source regions for C_4 taxa (with C_4 taxa accounting for 56, 63, and 76%, respectively, of the summed ancestral area probabilities), while Eurasia and Southern Africa are predominantly C_3 sources (with 78 and 82%, respectively). Modern-day distribution ranges of non-endemics largely mirror the ancestral dispersal patterns, but India has a markedly higher share of occurrences compared to its estimated importance as ancestral area, while the relationship is inverse for the Neotropics (Fig. II.4). The proportion of C_4 species in each area is higher and less variable than for ancestral areas, ranging from 68 to 89%. The number of occurrences is negatively correlated with distance to Madagascar, whether tested for C_3 and C_4 species separately, or combined (all $p < 0.05$, one-sided Spearman test).

Eleven clades with nested non-Madagascan species had a relative ancestral area probability >0.5 for the Madagascar region estimated with both the respective best model and DEC, of which seven have strong clade support (Table A-II.6). The largest and oldest is a C_3 group in the "Forest shade clade" (Fig. II.5) with a relative probability of 0.86 for the Madagascar region and a median crown age of 15.6 Ma (18.7 Ma under phytolith calibration). This "Madagascar shade clade" alone includes 17 of the strictly endemic C_3 lineages. The second-largest is the Hickeliinae subtribe of bamboos with 24 accessions, a relative area probability of 0.92 and median crown age of 6 Ma (7.3 Ma under phytolith calibration). The divergence of the only continental species, *Hickelia africana*, happened less than 2 Ma; in fact, sequences of a *Hickelia* specimen from Madagascar conservatively excluded were identical to *H. africana*. Further Madagascan origins were estimated for smaller clades in Chloridoideae (*Acrachne–Sclerodactylon*, *Neostapfiella–Chloris humbertiana–Daknopholis*, all C_4), Micraioideae (*Isachne*, C_3), and Paniceae (*Brachiaria–Yvesia*, *Setaria–Stenotaphrum*, all C_4). In the endemic C_4 *Panicum cinctum–luridum–voeltzkowii* clade (Paniceae), a basal split is supported between taxa from Madagascar and from neighbouring Réunion, Europa, and Glorioso islands.

3.5 Discussion

C_3 grasses: forest panicoid and bamboo radiations, recent migration, and possible relicts

Our results suggest that within the "Forest shade clade" (Panicoideae: Boivinellinae, Giussani et al. 2001), a large C_3 lineage appears to have originated in Madagascar (Fig. II.5). This lineage comprises grasses found predominantly in shaded locations. Despite considerable intraspecific variation and often blurred species boundaries in genera such as *Poecilostachys*, it is clear that Madagascan taxa account for most of its diversity. Field experience (M.S. Vorontsova, pers. obs.) also suggests that a number of endemic species remain to be described. We specifically included additional non-Madagascan species in this group and believe we have a good representation of its worldwide diversity. The deep successive branching of Madagascan lineages strongly suggests the region as the clade's origin, which is supported by range evolution analysis. Besides regional endemics, this group also comprises

species with a pantropical distribution such as *Pseudechinolaena polystachya*, *Oplismenus burmanni*, and *O. compositus*. From its likely origin in Madagascar, the group appears to have dispersed as far as Asia in a clade of *Cyrtococcum*, and the Neotropics in the case of *Lasiacis*. Stem and crown of the "Madagascar shade clade" are here estimated to the Early Miocene, although phytolith calibration also allows for a Late Oligocene origin.

The tropical woody bamboos of subtribe Hickeliinae are the second-largest clade with a Madagascan origin estimated here. This group originated in the Late Miocene, presumably via

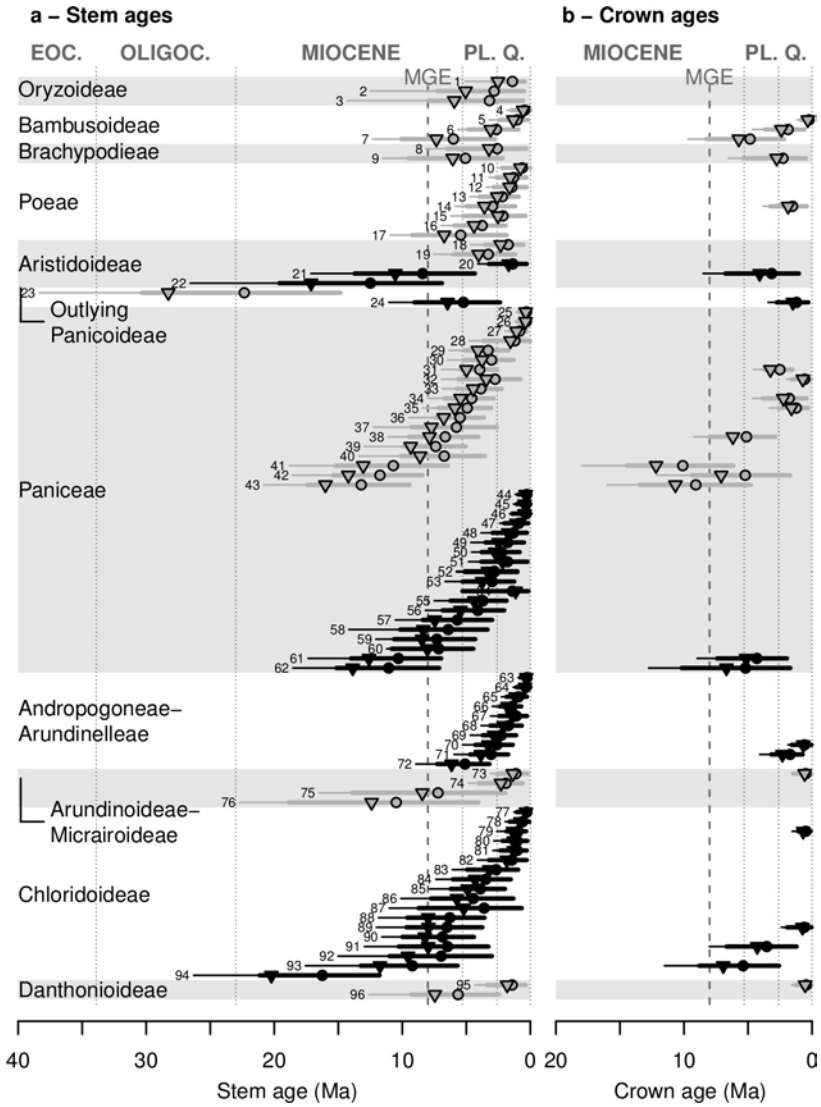

Figure II.2: Ages of 96 lineages strictly endemic to the Madagascar region, as inferred through divergence time estimation in *BEAST2*. C_3 lineages (N = 45) in grey, C_4 lineages (N = 51) in black. Note that these lineages cluster in 25 C_3 and 44 C_4 lineages when accounting for estimated origins in the Madagascar region (see text). Dots and thick lines are median ages and 95% High Posterior Density (HPD) intervals under external calibration; triangles and thinner lines are median ages and HPD intervals under calibration with phytolith ages. Dotted lines mark geological epoch boundaries; the dashed line at 8 Ma marks the approximate beginning of the global Miocene grassland expansion (MGE; Edwards et al., 2010). (a) Stem ages. Lineages are numbered left to age bars (see Table A-II.3 in Annex A-II). (b) Crown ages of lineages with more than one species.

immigration from Southeast Asia, and diversified in the Late Miocene to Pliocene. One species, *Hickelia africana*, recently dispersed to East Africa. Judging by sequence similarity, it may moreover occur in Madagascar, too, which would imply an even more recent range expansion to the continent. We could not include any of the two endemic species of *Cathariostachys*, but they very likely also fall in the Hickeliinae (Dransfield 1998). This bamboo subtribe and the "Madagascar shade clade" together probably account for more than half of the endemic C_3 grass species in Madagascar, and diversification in these two clades may thus largely explain the higher proportion of endemic species in C_3 compared to C_4 grasses in Madagascar.

Humid forests are believed to exist in Madagascar since the Oligocene (Wells 2003). The diversification of shade Paniceae and bamboos in the Madagascar region since the Miocene may have been favoured by factors such as the expansion of the Sambirano rainforests in the north (Wells 2003, Yoder & Nowak 2006) and the volcanic origin of the Comoros and Mascarenes (Warren et al. 2003). It is consistent with the estimated ages of other forest plant lineages in the region such as the palm tribe Areceae (Baker & Couvreur 2013), scaly tree ferns (Janssen et al. 2008), or *Canarium* (Federman et al. 2015). The extent of forests probably fluctuated with Pleistocene climatic cycles (Burney et al. 2004), which may have favoured recent allopatric speciation through fragmentation but also driven some species to peripheral humid refuges (e.g. *Chasechloa egregia* on Nosy Be island, Silva et al. 2017). The Hickeliinae bamboos have altitudinally patterned distributions (Dransfield 2003), suggesting parapatric speciation as potential mechanism of diversification. With more data on the distribution of forest grasses, it would be possible to test models of allopatric versus parapatric speciation proposed for Madagascar (Wilmé et al. 2006, Vences et al. 2009).

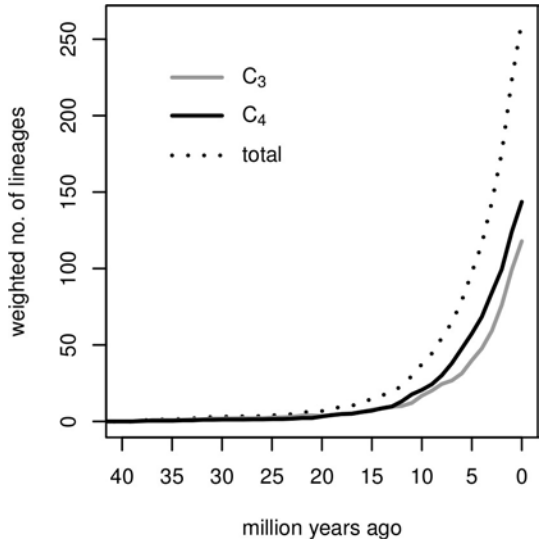

Figure II.3. Grass lineage accumulation in the Madagascar region, for all studied clades combined, as inferred through range evolution analysis. Lineages were counted per 1 Ma time bin (based on external calibration without phytoliths) and weighted by their marginal probability to occur in the Madagascar region.

In subfamily Oryzoideae, the forest endemics *Maltebrunia leersioides* (Oryzeae) and *Humbertochloa bambusiuscula* (Phyllorachideae) were both found to be sister to tropical African species, diverging from them in the Late Miocene to Quaternary. *Humbertochloa* fossils from the East African Miocene (Jacobs & Kabuye 1987), and the deep divergence of the Phyllorachideae dated here to the Eocene to Oligocene suggest this lineage of only three or four species (Soreng et al. 2015) may have originated in a hot and humid environment of continental Africa and dispersed to Madagascan forests only recently when forests expanded there. Another small, predominantly forest understorey clade was found in the temperate

bamboos (Arundinarieae: *Arundinaria* p.p.), grouping with the South African *Oldeania alpina* and diverging in the Pliocene to Quaternary; support for this group however is only moderate, as in most of this young bamboo tribe. Endemic C_3 lineages of high altitudes in the Pooideae (*Agrostis*, *Anthoxanthum*, *Festuca*, and *Poa*; less clear in *Brachypodium*) appear to have immigrated from Eurasia in the Pliocene or more recently. They clearly form part of a "pan-temperate element" in the African mountain flora (Gehrke & Linder 2009, Minaya et al. 2017, Tusiime et al. 2017). Further endemic mountain C_3 grasses in *Sartidia* (Aristidoideae), *Pentameris* and *Merxmuellera* (both Danthonioideae) appear to have their origin in Southern Africa, which is consistent with previous hypotheses (Besnard et al. 2014, for *Sartidia*, Linder et al. 2014, for Danthonioideae).

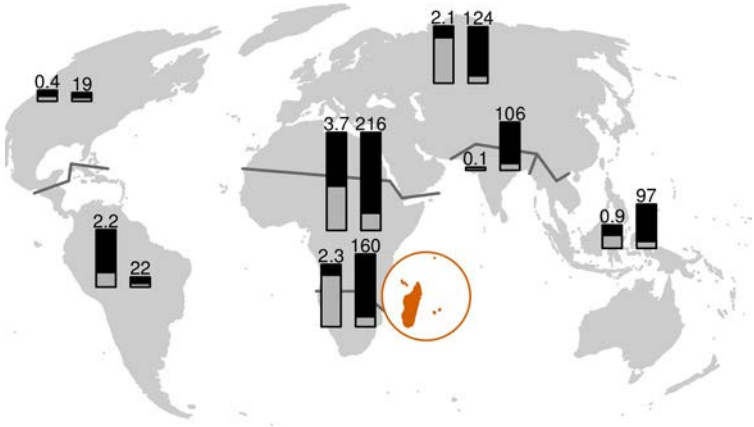

Figure II.4. Affinities of the Madagascan grass flora with other biogeographic regions. The map shows the areas defined for analysis of range evolution, including the Madagascar region (circle inset in orange). Bars show the sums of ancestral area probabilities for ancestors of Madagascar lineages inferred with the best-scoring range evolution model for each phylogeny (left bar for each area) and current-day occurrence counts of non-endemic Madagascan species (right bar), with grey and black segments denoting the proportion of C_3 and C_4 lineages or species, respectively. Both bars are scaled to the relative maximum value. Note that current-day occurrences and estimated historical distribution correlate well except for India and the Neotropics.

Older endemic C_3 lineages in Madagascar include *Styppeiochloa* (Arundinoideae), with further species awaiting description in addition to *S. hitchcockii* (J. Teisher, pers. comm.). The two accessions from Madagascar included here are highly divergent (~5–15 Ma), and the genus diverged in the Miocene to possibly Late Eocene. The lineage may have been more widespread in sun-exposed habitats before C_4 grasses overtook dominance. It now only persists on higher plateaus, where alternative adaptations to dry habitats such as strong desiccation tolerance may explain its frequent co-occurrence with C_4 grasses (G. Besnard, pers. obs.). *Lecomtella madagascariensis* (Panicoideae) is another possible relict of uncertain phylogenetic postion (Besnard et al. 2013). We estimated its divergence from Andropogoneae–Arundinelleae and the Asian *Jansenella griffithiana* at least to the Early Miocene. A recent phylogenomic study (Burke et al. 2016b) placed it at the base of the core Panicoideae, suggesting it may be even older. *Lecomtella* is very likely the oldest endemic grass lineage in Madagascar. It was possibly more widespread at lower altitudes during the cooler Oligocene and now persists only on the summit of the Andringitra massif (Besnard et al. 2013).

C_4 grasses: Miocene to Pliocene origins and two possible relicts

We show here that C_4 grasses count nearly twice as many lineage origins in Madagascar compared to C_3 grasses (44 versus 25 recovered) although they have less endemic species

(100 versus 120). This suggests that while *in situ* radiations produced larger endemic clades in C_3 grasses, there were more independent colonisations of C_4 grasses. Median stem ages of most endemic C_4 lineages, and crown ages of endemic clades in the Aristidoideae (*Aristida rufescens–similis*), Chloridoideae and Panicoideae, are younger than 8 Ma (some only marginally so under phytolith calibration) and thus compatible with the onset of the global Miocene grassland expansion (Edwards et al. 2010). Many colonisations appear to have occurred only in the Quaternary. In contrast, some lineages (including the *Aristida rufescens–similis* complex, a major component of present central plateau grasslands) may have been present before 8 Ma. Divergences of the isolated *Aristida ambongensis* and *Eragrostis chabouisii* were tentatively estimated to the Early to Mid Miocene. These potentially oldest endemic C_4 lineages in Madagascar are only known from the dry west and sub-arid southwest, respectively, and may be remnants from a once more widespread dry bush vegetation (Wells 2003).

Taking crown ages of endemic C_4 clades under external calibration as a minimum estimate, our results suggests that open-canopy habitats suitable for diversification of C_4 taxa existed in Madagascar at least by the Pliocene. They thus support the view that grasslands in Madagascar form part of the global expansion of C_4 grassy biomes (Bond et al. 2008). They could even be older if one considers the uncertainty about the age of Poaceae (see below). We stress, however, that this gives no indication of the area these may once have occupied. Primitive C_4 grassy biomes in Madagascar may have occurred in small patches, in a mosaic with other vegetation (Burney 1987, Lowry et al. 1997, Godfrey & Crowley 2016), or may have occupied more extensive areas. Moreover, their extent probably fluctuated with Quaternary climatic variations (Burney et al. 2004). Some may have resembled today's tapia (*Uapaca bojeri*) woodlands, where C_3 and C_4 grasses are frequently found in close proximity. As shown here, a number of endemic C_4 grasses colonised Madagascar only in the Quaternary, underlining the fact that species assemblages in open biomes continued evolving until very recently.

C_4 photosynthesis, although generally an adaptation to open and warm habitats, covers a large ecological spectrum (Christin & Osborne 2014), and C_4 grassland communities in South Africa have been shown to cluster phylogenetically depending on humidity, fire regime, and grazer disturbance (Visser et al. 2012). In Madagascar, Andropogoneae and Chloridoideae also tend to dominate in more mesic versus more dry habitats, respectively (Bond et al. 2008). The assembly histories of such different grassland types may not have been the same. The challenge therefore lies in characterising the different types of old-growth grasslands and savannahs in Madagascar and distinguishing them from secondary vegetation. More research on the distribution, regional endemism, community structure, and possibly phylogeographic patterning of grasses (e.g. McAllister & Miller 2016) as well as other plants endemic to open grassy biomes of Madagascar is needed to this end.

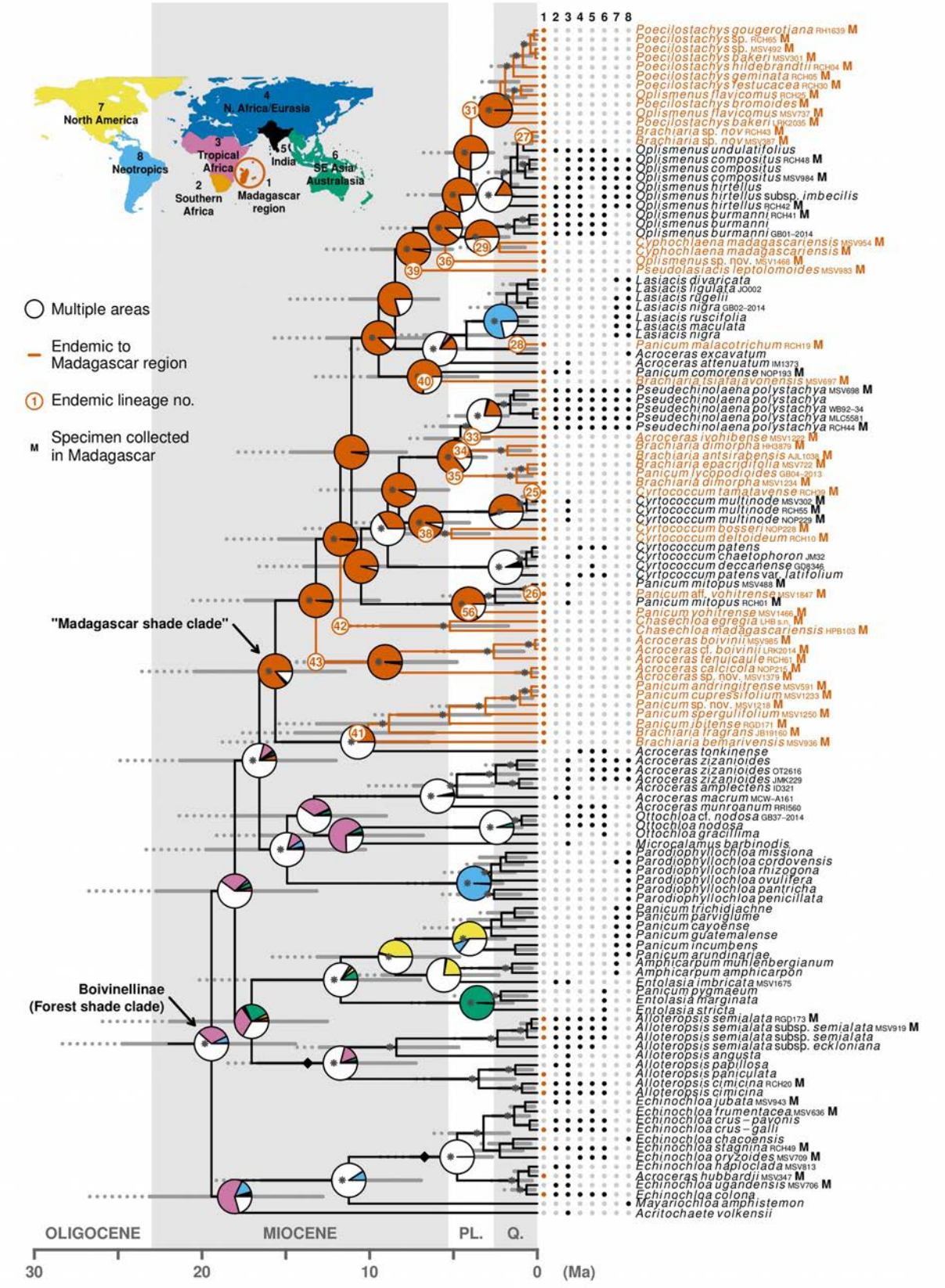
1 2 3 4 5 6 7 8
7 North America
4 N. Africa/Eurasia
3 Tropical Africa
5 India
6 SE Asia/ Australasia
8 Neotropics
2 Southern Africa
1 Madagascar region
Multiple areas
Endemic to Madagascar region
Endemic lineage no.
Specimen collected in Madagascar
"Madagascar shade clade"
Boivinellinae (Forest shade clade)
OLIGOCENE
MIOCENE
PL.
Q.
30
20
10
0
(Ma)

Figure II.5 (previous page). Time-calibrated phylogeny of the Forest shade clade (Panicoideae: Paniceae: Boivinellinae), with estimated Madagascan origin for a large C_3 "Madagascar shade clade". Current distribution (dots left to species labels) and ancestral areas inferred under a DEC+*j*+*x* model for selected nodes are shown; the map inset shows the areas defined. The tree was extracted from the Maximum Clade Credibility tree inferred with *BEAST2* for Paniceae, with median node heights under external calibration. Solid grey bars are 95% High Posterior Density (HPD) intervals; dotted bars are HPD intervals under a calibration with phytolith ages. Diamonds on branches denote C_4 clades (*Alloteropsis* is simplified as such), asterisks on nodes show a Bayesian posterior support ≥0.95. Strictly endemic lineages are numbered (see Table A-II.3 in Annex A-II).

Madagascar's extant grass flora: Neogene diversification and frequent dispersals

In this study, we aimed to understand the assembly of the grass flora in the Madagascar region, and to compare the diversification of C_3 and C_4 grasses. Studies like ours are subject to some limitations, in particular due to taxon sampling and the age calibration used.

We attempted to include as many taxa as possible, for Madagascar as well as other areas. The overall proportion of species sampled per clade ranged between roughly 10% (Andropogoneae–Arundinelleae) and 65% (Brachypodieae), while we have a better sampling of the Madagascan grass flora and its endemics (73% and 65%, respectively). We thus assume to have a good minimum estimate of the endemic lineages and lineage origins in Madagascar, although the true numbers may be higher. Ancestral range estimates are more likely to be affected by species sampling. They are consistent overall with current-day distribution patterns, with the notable exceptions of India and the Neotropics. It is possible that we underestimated India's historical links with Madagascar, given also the known affinities of the Indian and Madagascan floras in general (Schatz 1996, Buerki et al. 2013). In turn, the Neotropical grass flora is relatively well studied, which may have biased upwards our estimates for its historical contribution. Finally, the lineage-through-time curves (Fig. II.3) inferred for the Madagascar region are valid within the limits of our species sampling and give only minimum estimates of past diversity.

The age of Poaceae is subject of an ongoing debate, owing in particular to the uncertain value of phytolith fossils (Christin et al. 2014). The recent inclusion of an old macrofossil, presumably a grass spikelet (Poinar et al. 2015), added to the discussion and led to conspicuously old Poaceae age estimates (Burke et al. 2016a). Here, rather than subjectively choosing one set of fossils, we opted to use the two alternative hypotheses produced by Christin et al. (2014) as secondary calibrations, and test their implications for the Madagascan grass flora. We show that ages under both calibrations support the notion of a largely Neogene origin of Madagascan grasses. They approach each other for recent divergences, probably because they are effectively constrained where lineage sampling is denser. Much of the uncertainty in Poaceae concerns deeper divergences, which do not directly concern Madagascan lineages. However, we acknowledge that if the taxonomic placement of phytoliths and ancient macrofossils are confirmed and compatible with external angiosperm constraints, ages of grasses in Madagascar may lie nearer the upper end of our estimated confidence intervals. We expect that the precise branching order of the PACMAD subfamilies (see Cotton et al. 2015) has little effect on our age estimates, as the branches between them are very short (Cotton et al. 2015, GPWG II 2012).

The 69 lineages estimated here to have their origin in the Madagascar region suggest recurrent dispersals to Madagascar and its outlying islands. Dispersal appears to be constrained by distance, demonstrated by the very strong support for range evolution models incorporating a negative distance scaling parameter *x* in 9 of 11 phylogenies analysed. Linder et al. (2013) and Minaya et al. (2017) also found a negative correlation between dispersal rate and distance in the Danthonioideae and Loliinae (subfamily Pooideae) lineages of grasses, respectively. It appears that, despite often large ranges, dispersal limitation has shaped the distribution of the grass family. Few studies have so far included the *x* parameter, and we expect that it will provide a good fit for many plant phylogenies. The inference of Tropical and Southern Africa as areas with strongest historical and modern links with Madagascar is in line with patterns of taxonomic similarity in many plants groups (Buerki et al. 2013).

A final remark we would like to make concerns the role of extinction. This parameter is not easily tractable in historical biogeography, and our inferences are valid only for the lineage diversity we sampled, i.e. the branches leading to extant species. The island of Madagascar underwent significant environmental change since its split from Gondwana (Wells 2003). Species on long branches such as *Eragrostis chabousii*, only found in the south-western spiny bush, and *Lecomtella madagascariensis* from the high Andringitra, may be the only remnants from lineages that once were more diverse. We cannot exclude the possibility that an older, quite different assemblage of grasses was present on Madagascar before the extant grass flora emerged. If there was expansion of open habitats in Madagascar during the global Miocene grassland expansion, it is indeed likely that this was accompanied by significant lineage turnover.

To conclude, we infer that the extant grass flora of that Madagascan region diversified from the Miocene onwards, but there remains some possibility that lineages are older than we estimated. A few phylogenetically isolated lineages may be remnants from ancient biomes that retracted with environmental change, and extinction is likely to have occurred. Biogeographical links are strongest with Tropical and Southern Africa. C_3 grasses include two likely *in situ* radiations of bamboos and forest Paniceae that account largely for their higher proportion of endemics. There were twice as many immigration events in C_4 grasses than in C_3 grasses. C_4 grasses have been present and diversified in Madagascar since at least the Pliocene, supporting the idea that there are old-growth C_4 grassy biomes on the island.

3.6 References

Baker W.J. & Couvreur T.L.P. (2013): Global biogeography and diversification of palms sheds light on the evolution of tropical lineages. II. Diversification history and origin of regional assemblages. *Journal of Biogeography*, 40: 286–298.

Besnard G., Christin P.-A., Malé P.-J.G., Coissac E., Ralimanana H. & Vorontsova M.S. (2013): Phylogenomics and taxonomy of Lecomtelleae (Poaceae), an isolated panicoid lineage from Madagascar. *Annals of Botany*, 112: 1057–1066.

Besnard G., Christin P.-A., Malé P.-J.G., Lhuillier E., Lauzeral C. et al. (2014): From museums to genomics: old herbarium specimens shed light on a C_3 to C_4 transition. *Journal of Experimental Botany*, 65: 6711–6721.

Bond W.J., Silander Jr J.A., Ranaivonasy J. & Ratsirarson J. (2008): The antiquity of

Madagascar's grasslands and the rise of C_4 grassy biomes. *Journal of Biogeography*, 35: 1743–1758.

Bouckaert R., Heled J., Kühnert D., Vaughan T., Wu C.-H. et al. (2014): BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, 10: e1003537.

Buerki S., Devey D.S., Callmander M.W., Phillipson P.B. & Forest F. (2013): Spatio-temporal history of the endemic genera of Madagascar. *Botanical Journal of the Linnean Society*, 171: 304–329.

Burke S.V., Lin C.-S., Wysocki W.P., Clark L.G. & Duvall M.R. (2016a): Phylogenomics and plastome evolution of tropical forest grasses (*Leptaspis*, *Streptochaeta*: Poaceae). *Frontiers in Plant Science*, 7: 1993.

Burke S.V., Wysocki W.P., Zuloaga F.O., Craine J.M., Pires J.C. et al. (2016b): Evolutionary relationships in Panicoid grasses based on plastome phylogenomics (Panicoideae; Poaceae). *BMC Plant Biology*, 16: 140.

Burney D.A. (1987): Late Holocene vegetational change in central Madagascar. *Quaternary Research*, 28: 130–143.

Burney D.A. (1997): Theories and facts regarding Holocene environmental change before and after human colonization. *Natural change and human impact in Madagascar* (ed. by S.M. Goodman and B.D. Patterson): 75–89. Smithsonian Institution Press, Washington, D.C.

Burney D.A., Burney L.P., Godfrey L.R., Jungers W.L., Goodman S.M. et al. (2004): A chronology for late prehistoric Madagascar. *Journal of Human Evolution*, 47: 25–63.

Burney D.A., Robinson G.S. & Burney L.P. (2003): *Sporormiella* and the Late Holocene extinctions in Madagascar. *Proceedings of the National Academy of Sciences of the United States of America*, 100: 10800–10805.

Christin P.-A., Besnard G., Samaritani E., Duvall M.R., Hodkinson T.R. et al. (2008): Oligocene CO_2 decline promoted C_4 photosynthesis in grasses. *Current Biology*, 18: 37–43.

Christin P.-A. & Osborne C.P. (2014): The evolutionary ecology of C_4 plants. *New Phytologist*, 204: 765–781.

Christin P.-A., Spriggs E., Osborne C.P., Strömberg C.A.E., Salamin N. & Edwards E.J. (2014): Molecular dating, evolutionary rates, and the age of the grasses. *Systematic Biology*, 63: 153–165.

Clayton W.D., Govaerts R., Harman K.T., Williamson H. & Vorontsova M. (2016): *World Checklist of Poaceae, facilitated by the Royal Botanic Gardens, Kew.* Available at: http://apps.kew.org/wcsp/.

Cotton J.L., Wysocki W.P., Clark L.G., Kelchner S.A., Pires J.C. et al. (2015): Resolving deep relationships of PACMAD grasses: a phylogenomic approach. *BMC Plant Biology*, 15: 178.

Dransfield S. (1998): *Valiha* and *Cathariostachys*, two new bamboo genera (Gramineae-Bambusoideae) from Madagascar. *Kew Bulletin*, 53: 375–397.

Dransfield S. (2003): Poaceae, Bambuseae, bamboos. *The Natural History of Madagascar* (ed. by S.M. Goodman and J.P. Benstead): 467–471. University of Chicago Press, Chicago, London.

Drummond A.J., Ho S.Y.W., Phillips M.J. & Rambaut A. (2006): Relaxed phylogenetics and dating with confidence. *PLoS Biology*, 4: e88.

Edwards, E.J., Osborne, C.P., Strömberg, C.A.E. & Smith, S.A. (2010): The origins of C_4 grasslands: integrating evolutionary and ecosystem science. *Science*, 328: 587–591.

Edwards E.J. & Still C.J. (2008): Climate, phylogeny and the ecological distribution of C_4 grasses. *Ecology Letters*, 11: 266–276.

Federman S., Dornburg A., Downie A., Richard A.F., Daly D.C. & Donoghue M.J. (2015): The biogeographic origin of a radiation of trees in Madagascar: implications for the assembly of a tropical forest biome. *BMC Evolutionary Biology*, 15: 216.

Gasse F. & Van Campo E. (1998): A 40,000-yr pollen and diatom record from Lake Tritrivakely, Madagascar, in the southern tropics. *Quaternary Research*, 49, 299–311.

Gehrke B. & Linder H.P. (2009): The scramble for Africa: pan-temperate elements on the African high mountains. *Proceedings of the Royal Society of London B: Biological Sciences*, 276: 2657–2665.

Giussani L.M., Cota-Sánchez J.H., Zuloaga F.O. & Kellogg E.A. (2001): A molecular phylogeny of the grass subfamily Panicoideae (Poaceae) shows multiple origins of C_4 photosynthesis. *American Journal of Botany*, 88: 1993–2012.

Godfrey L.R. & Crowley B.E. (2016): Madagascar's ephemeral palaeo-grazer guild: who ate the ancient C_4 grasses? *Proceedings of the Royal Society B: Biological Sciences*, 283: 20160360.

GPWG II – Grass Phylogeny Working Group II (2012): New grass phylogeny resolves deep evolutionary relationships and discovers C_4 origins. *New Phytologist*, 193: 304–312.

Hijmans R.J. (2016): *geosphere: Spherical Trigonometry. R package version 1.5-5.* Available at: https://cran.r-project.org/web/packages/geosphere.

Humbert H. (1927): Destruction d'une flore insulaire par le feu : principaux aspects de la végétation à Madagascar. *Mémoires de l'Académie Malgache*, 5: 1–80.

Jacobs B.F. & Kabuye C.H.S. (1987): A Middle Miocene (12.2 my old) forest in the East African Rift Valley, Kenya. *Journal of Human Evolution*, 16: 147–155.

Janssen T., Bystriakova N., Rakotondrainibe F., Coomes D., Labat J.-N. & Schneider H. (2008):

Neoendemism in Madagascan scaly tree ferns results from recent, coincident diversification bursts. *Evolution*, 62: 1876–1889.

Katoh K. & Standley D.M. (2013): MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, 30: 772–780.

Kellogg E.A. (2015): *Flowering Plants. Monocots: Poaceae*. Springer, New York.

Koechlin J. (1972): Flora and vegetation of Madagascar. *Biogeography and ecology in Madagascar* (ed. by R. Battistini and R. Richard-Vindard): 145–190. J. Cramer, Vaduz.

Linder H.P., Antonelli A., Humphreys A.M., Pirie M.D. & Wüest R.O. (2013): What determines biogeographical ranges? Historical wanderings and ecological constraints in the danthonioid grasses. *Journal of Biogeography*, 40: 821–834.

Linder H.P., Rabosky D.L., Antonelli A., Wüest R.O. & Ohlemüller R. (2014): Disentangling the influence of climatic and geological changes on species radiations. *Journal of Biogeography*, 41: 1313–1325.

Lowry II, P.P., Schatz, G.E. & Phillipson P.B. (1997): The classification of natural and anthropogenic vegetation in Madagascar. *Natural change and human impact in Madagascar* (ed. by S.M. Goodman and B.D. Patterson): 93–123. Smithsonian Institution Press, Washington and London.

Lundgren M.R., Besnard G., Ripley B.S., Lehmann C.E.R., Chatelet D.S. et al. (2015): Photosynthetic innovation broadens the niche within a single species. *Ecology Letters*, 18: 1021–1029.

Madagascar Catalogue (2013): *Catalogue of the vascular plants of Madagascar.* Missouri Botanical Garden, St. Louis, Antananarivo. Available at: http://www.tropicos.org/Project/Madagascar.

Matzke N.J. (2013): *BioGeoBEARS: BioGeography with Bayesian (and Likelihood) Evolutionary Analysis in R Scripts.* University of California, Berkeley.

Matzke N.J. (2014): Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Systematic Biology*, 63: 951–970.

McAllister C.A. & Miller A.J. (2016): Single nucleotide polymorphism discovery via genotyping by sequencing to assess population genetic structure and recurrent polyploidization in *Andropogon gerardii*. *American Journal of Botany*, 103: 1314–1325.

Miller M.A., Pfeiffer W. & Schwartz T. (2012): The CIPRES science gateway: enabling high-impact science for phylogenetics researchers with limited resources. *Proceedings of the 1st Conference of the Extreme Science and Engineering Discovery Environment: Bridging from the eXtreme to the campus and beyond*, 39.

Minaya M., Hackel J., Namaganda M., Brochmann C., Vorontsova M.S., Besnard G. & Catalán P. (2017): Contrasting dispersal histories of broad- and fine-leaved temperate Loliinae grasses: range expansion, founder events, and the roles of distance and barriers. *Journal of Biogeography*, 44: 1980–1993.

Moat J. & Smith P. (2007): *Atlas of the Vegetation of Madagascar.* Royal Botanic Gardens, Richmond, Surrey, UK.

Nanjarisoa O.P., Besnard G., Ralimanana H. & Vorontsova M.S. (2017): Poaceae checklist of the Itremo Massif records endemic plateau grasses. *Madagascar Conservation & Development*, in press.

Osborne C.P., Salomaa A., Kluyver T.A., Visser V., Kellogg E.A. et al. (2014): A global database of C_4 photosynthesis in grasses. *New Phytologist*, 204: 441–446.

Perrier de la Bâthie H. (1921): La végétation malgache. *Annales du Musée Colonial de Marseille*, 9: 1–266.

Poinar G., Alderman S. & Wunderlich J. (2015): One hundred million year old ergot: psychotropic compounds in the Cretaceous? *Palaeodiversity*, 8: 13–19.

Rambaut A., Suchard M.A., Xie D. & Drummond, A. (2014): *Tracer v1.6.* Available at: http://beast.bio.ed.ac.uk/Tracer.

Ree R.H. & Smith S.A. (2008): Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology*, 57: 4–14.

Schatz G.E. (1996): Malagasy/Indo-Australo-Malesian phytogeographic connections. *Biogéographie de Madagascar* (ed. by W.R. Lourenço): 73–83. ORSTOM éditions, Paris.

Silva C., Besnard G., Piot A., Razanatsoa J., Oliveira R.P. & Vorontsova M.S. (2017): Museomics resolve the systematics of an endangered grass lineage endemic to north-western Madagascar. *Annals of Botany*, 119: 339–351.

Soreng R.J., Peterson P.M., Romaschenko K., Davidse G., Zuloaga F.O. et al. (2015): A worldwide phylogenetic classification of the Poaceae (Gramineae). *Journal of Systematics and Evolution*, 53: 117–137.

Spriggs E.L., Christin P.-A. & Edwards, E.J. (2014): C_4 photosynthesis promoted species diversification during the Miocene grassland expansion. *PLoS ONE*, 9: e97722.

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

Taylor S.H., Franks P.J., Hulme S.P., Spriggs E., Christin P.A. et al. (2012): Photosynthetic pathway and ecological adaptation explain stomatal trait diversity amongst grasses. *New Phytologist*, 193: 387–396.

Tusiime F.M., Gizaw A., Wondimu T., Masao C.A., Abdi A.A. et al. (2017): Sweet vernal grasses (*Anthoxanthum*) colonized African mountains along two fronts in the Late Pliocene, followed by secondary contact, polyploidization and local extinction in the Pleistocene. *Molecular Ecology*, 26: 3513–3532.

Van Dam M.H. & Matzke N.J. (2016): Evaluating the influence of connectivity and distance on biogeographical patterns in the south-western deserts of North America. *Journal of Biogeography*, 43: 1514–1532.

Vences M., Wollenberg K.C., Vieites D.R. & Lees D.C. (2009): Madagascar as a model region of species diversification. *Trends in Ecology & Evolution*, 24: 456–465.

Vicentini A., Barber J.C., Aliscioni S.S., Giussani L.M. & Kellogg E.A. (2008): The age of the grasses and clusters of origins of C_4 photosynthesis. *Global Change Biology*, 14: 2963–2977.

Visser V., Clayton W.D., Simpson D.A., Freckleton R.P. & Osborne C.P. (2014): Mechanisms driving an unusual latitudinal diversity gradient for grasses. *Global Ecology and Biogeography*, 23: 61–75.

Visser V., Woodward F.I., Freckleton R.P. & Osborne C.P. (2012): Environmental factors determining the phylogenetic structure of C_4 grass communities. *Journal of Biogeography*, 39: 232–246.

Vorontsova M.S., Besnard G., Forest F., Malakasi P., Moat J. et al. (2016): Madagascar's grasses and grasslands: anthropogenic or natural? *Proceedings of the Royal Society of London B: Biological Sciences*, 283: 20152262.

Vorontsova M.S. & Rakotoarisoa S.E. (2014): Endemic non-bambusoid genera of grasses (Poaceae) in Madagascar: review of current knowledge. *Malagasy Nature*, 8: 14–34.

Warren B.H., Bermingham E., Bowie R.C.K., Prys-Jones R.P. & Thébaud C. (2003): Molecular phylogeography reveals island colonization history and diversification of western Indian Ocean sunbirds (*Nectarinia*: Nectariniidae). *Molecular Phylogenetics and Evolution*, 29: 67–85.

Wells N.A. (2003): Some hypotheses on the Mesozoic and Cenozoic paleoenvironmental history of Madagascar. *The natural history of Madagascar* (ed. by S.M. Goodman and J.P. Benstead): 16–34. University of Chicago Press, Chicago, London.

Wilmé L., Goodman S.M. & Ganzhorn J.U. (2006): Biogeographic evolution of Madagascar's microendemic biota. *Science*, 312: 1063–1065.

Yoder A.D. & Nowak M.D. (2006): Has vicariance or dispersal been the predominant biogeographic force in Madagascar? Only time will tell. *Annual Review of Ecology, Evolution, and Systematics*, 37: 405–431.

4 Chapter III: Phylogenomics and biogeography of the core Panicoideae[1]

4.1 Abstract

Within the grass family (Poaceae), the subfamily Panicoideae includes diverse clades that dominate tropical C_4 grasslands, such as Andropogoneae, but also some C_3 species with uncertain positions. These species, notably in the two genera *Chandrasekharania* and *Jansenella*, are potentially important for our understanding of C_3– C_4 transitions and the origins of C_4 grasslands. We attempted to establish their positions in the grass phylogeny using phylogenomic methods. We assembled full chloroplast genomes, the nuclear genes *phyB*, the nuclear *nadpme* and *ppc* gene families and as the ITS region from shotgun sequencing data. Phylogenies were reconstructed and divergence times estimated from chloroplast genomes. We analysed specific C_4-related amino-acid sites in the *ppc* and *rbcL* genes. Distribution patterns were compared for the major core Panicoideae lineages. Results confirm *Gynerium* as sister to the remaining core Panicoideae, while the branching order of *Lecomtella* and Paniceae remains unclear. A polyploid origin was inferred for *Gynerium* and *Jansenella neglecta*, and potentially *Lecomtella*. The three species of *Chandrasekharania* and *Jansenella* are sister to the large C_4 clade Andropogoneae–Arundinelleae and have *ppc* and *rbcL* features typical of C_3 plants. The lineage, as Andropogoneae–Arundinelleae, has its centre of diversity and endemism in India. The position of the C_3 clade *Jansenella–Chandrasekharania* sister to Andropogoneae–Arundinelleae suggests that the transition to C_4 photosynthesis occurred on the Indian subcontinent, in the Early Miocene. The *Jansenella–Chandrasekharania* clade will be useful for comparative studies on the evolution of C_4 photosynthesis.

4.2 Introduction

Within the cosmopolitan grass familiy (Poaceae), Panicoideae is one of the largest subfamilies, with roughly 3,300 species (Soreng et al. 2015, Kellogg 2015). This lineage is mainly distributed in the tropics, where its members are among the dominant plants in open biomes (Osborne 2008). Tropical grassy biomes cover around 20% of the global land surface and have received increasing attention due to their importance for biodiversity and the global carbon cycle, particularly in a context of global change (Parr et al. 2014, Bond 2016, Lehmann & Parr 2016). The ecology of these biomes is tightly linked to the evolution of C_4 photosynthesis, which is an adaptation to the high photorespiration costs in warm and open habitats and plays a role in the productivity that fuels the regular fires (Osborne 2008). In the Miocene, between 3 and 10 million years ago (Ma), C_4 grasslands expanded globally, but timing and pace were not the same between different regions and continents (Edwards et al. 2010, Strömberg 2011). Multiple clades of grasses independently evolved the C_4 pathway, of which most are concentrated in subfamily Panicoideae, which alone counts 22–24 C_4 origins (Kellogg 2015).

1 The results of this chapter have not been published elsewhere when this dissertation was completed. The study was devised by G. Besnard and myself. G.B. assembled nuclear data and part of the plastome data. Alexandre Meunier assembled part of the plastome data and performed a first phylogenetic analysis as part of his Master thesis. P.-A. Christin and M. Bianconi shared valuable thoughts on divergence time estimation, on an earlier version of this manuscript, and intermediate results from a nuclear genome analysis (M.B., in preparation). My contribution was the assembly of part of the plastome data and the final analysis.

Molecular phylogenetic analyses of the Panicoideae have identified a "core" clade, including notably the large tribes Paniceae, Paspaleae, Andropogoneae, and Arundinelleae (GPWG II 2012, Burke et al. 2016). The latter two (also classified as a single tribe Andropogoneae, Kellogg 2015) are sister to each other and composed of C_4 species only, suggesting a single common C_4 origin. Paniceae and Paspaleae, on the contrary, each include several origins of C_4 photosynthesis (GPWG II 2012). These C_3-to-C_4 transitions led to some large radiations, albeit with different geographic distributions: Paniceae is largely pantropical, while most genera of Paspaleae are neotropical, and Andropogoneae–Arundinelleae are dominated by Asian members (Soreng et al. 2017). Andropogoneae, in particular, are the dominant grasses in all Asian savannah types (Ratnam *et al.* 2016). Resolving the origins and diversification of these C_4 groups, in their respective temporal and geographical settings, is necessary to understand both the common features and the particularities of modern tropical grassy biomes.

The C_4 species in Panicoideae have been intensively studied, particularly those of economic interest such as maize and sorghum, whose genomes have been sequenced. However, to understand the origin of C_4 clades, it is necessary to take into account their C_3 relatives, too (Christin et al. 2009a). Several of these are species-poor lineages and have often been overlooked in phylogenetic analyses. The monotypic genus *Gynerium* from tropical America is believed to be sister to the rest of the core panicoids based on analysis of the *rbcL*, *ndhF*, and *matK* plastid genes (Barker et al. 1995, GPWG II 2012). Another monotypic C_3 genus, *Lecomtella* from the Andringitra range in Madagascar, has been ambiguously placed so far. Besnard et al. (2013) showed that it is a distinct lineage in the core Panicoideae but with unclear relationship to the other tribes, depending on whether selected plastid markers, whole plastomes, or nuclear data were used. A recent whole-plastome analysis placed it as sister to a clade of Paniceae, Paspaleae and Andropogoneae–Arundinelleae (Burke et al. 2016).

The panicoid genera *Jansenella* and *Chandrasekharania*, both assumed to be C_3 (Osborne et al. 2014, Watson et al. 2017), have been enigmatic in terms of phylogenetic affinities. They have been included in Arundinelleae (Bor 1955, Sánchez-Ken & Clark 2010), referred tentatively to the otherwise C_4 tribe Tristachyideae (Kellogg 2015), considered *incertae sedis* in Panicoideae (Soreng et al. 2015), and again included in Arundinelleae (Soreng et al. 2017). Of the three known species, *Jansenella griffithiana* was described over 170 years ago (Müller 1856) and has the widest known distribution range, from Sri Lanka over India to Thailand (Bor 1955, Teerawatnanon & Hodkinson 2008). *Jansenella neglecta* and the single species *Chandrasekharania keralensis* were described more recently and are both known only from southern India (Nair et al. 1982, Yadav et al. 2010). *Jansenella griffithiana* is the only species so far included in a molecular phylogenetic study, which placed it sister to Andropogoneae–Arundinellae based on *ndhF*, *matK* and *rbcL* (see chapter II). This position is interesting for two reasons. First, a C_3 sister lineage of Andropogoneae–Arundinelleae would allow direct comparisons for studies on the origin of the photosynthetic pathway in one of the largest

known C_4 clades (Besnard & Christin 2010). Second, the high diversity of this C_4 clade in south and south-east Asia (Hartley 1958), combined with the known distribution of *Chandrasekharania* and *Jansenella*, would suggest an origin of Andropogoneae–Arundinelleae on the Indian subcontinent.

Here, we assembled complete chloroplast genomes (hereafter plastomes) and nuclear genes for all three species known in *Jansenella* and *Chandrasekharania*. We placed them within a phylogeny of the Panicoideae to test their sister relationship with Andropogoneae–Arundinelleae and checked whether two phyotosynthesis-related genes had a C_4 signature. Further, we re-investigate the positions of the two isolated, monotypic tropical C_3 genera *Lecomtella* and *Gynerium*. Finally, we estimate divergence times and discuss the biogeographical implications of the position of these tropical C_3 taxa.

4.3 Materials and Methods

Species sampling

We analysed one herbarium specimen for each of the three known species in *Chandrasekharania* and *Jansenella*, including isotypes of *Ch. keralensis* and *J. neglecta*. We also sequenced further specimens from major tribes of Panicoideae: Andropogoneae (*Glyphochloa forficulata* and *Lasiurus scindicus*), Arundinelleae (*Arundinella nepalensis*), Gynerieae (*Gynerium sagittatum*), Paspaleae (*Reynaudia filiformis*, presumably the earliest-diverging member of the tribe; GPWG II 2012), Tristachyideae (*Danthoniopsis stocksii*) and Paniceae (*Digitaria glauca*, *Melinis minutiflora*). Further sequences were obtained from *GenBank*: we generated a dataset of full chloroplast genomes (hereafter plastomes) focused on the core Panicoideae for which a good amount of sequences has been produced recently (Burke et al. 2016, Arthan et al. 2017, Piot et al. 2017), with one further representative plastome added for each of the other panicoid tribes (except Cyperochloeae and Steyermarkochloeae, both unavailable) and Poaceae subfamilies. Nuclear sequences were also assembled from shotgun data (see below), and we supplemented these with homologous Poaceae sequences which were retrieved directly and through *BLAST* searches from *GenBank*. All plastome sequences used are listed with voucher names and *GenBank* accessions in Table A-III.1 in annex A-III.

DNA extraction, sequencing, and sequence assembly

DNA was extracted from very small amounts of herbarium material, mostly leaves, or one single seed in the case of *Ch. keralensis*, using the BioSprint 15 DNA Plant Kit (Qiagen Inc.). DNA was then sequenced on a HiSeq 3000 machine at Genotoul, Castanet-Tolosan, France, or on a HiSeq 2000 machine at Genoscope, Evry, France, applying a low-coverage, genome-skimming strategy.

The 100 or 150 bp paired-end reads obtained were first used to assemble full plastomes. We used the *Org.Asm* organelle assembler v. 1.0 (http://pythonhosted.org/ORG.asm) with coding sequences of *Axonopus fissifolius* (KU291491.1, Paspaleae) as seeds for initiating

contig assembly. Final assemblies were checked by mapping reads on the obtained circular sequence in *Geneious* v. 9.0.5 (Biomatters Ltd., Auckland). Low-copy nuclear genes were assembled manually for *A. nepalensis*, *Ch. keralensis*, *Gl. forficulata*, *Gy. sagittatum*, *J. griffithiana*, *J. neglecta*, *La. scindicus*, *Le. madagascariensis* and *R. filiformis*, following the method described by Besnard et al. (2014). We focused on genes frequently employed in grass phylogenetics and with a large species sampling available: *phyB* (phytochrome B) and the gene families *nadpme* (coding for NADP-malic enzyme) and *ppc* (coding for phosphoenolpyruvate carboxylase). The latter two include genes that are implicated in the C_4 photosynthetic pathway. While for some accessions assembly was straightforward, only incomplete sequences could be assembled for others. Some accessions were polymorphic at the studied loci (see Results). We were able to separate the different copies in some cases by phasing the reads, but in other cases we could only separate short stretches (in general exons) and were not able to match fragments of each copy throughout the sequence. The complete nuclear ribosomal cluster, including the internal transcribed spacer (ITS) region, was also assembled for *Ch. keralensis*, *J. grifftihiana*, and *J. neglecta*.

Phylogenetics and divergence time estimation

The 88 full plastomes were aligned with *MAFFT* v. 7.2 (Katoh & Standley 2013) with default parameters, after removing one of the inverted repeat regions. A visual check of coding sequence (CDS) translations revealed no obvious alignment errors. We then estimated phylogenies on (i) the whole alignment, (ii) only alignment columns without any gaps (as done by Burke et al. 2016), and (iii) only CDS regions. *RAxML* v. 8.2.4 (Stamatakis 2014) was used for Maximum Likelihood (ML) estimation, applying a GTR substitution model and the CAT model of site heterogeneity with 25 rate categories and specifying *Anomochloa marantoidea* as outgroup. One thousand rapid bootstrap searches were run to assess node support, followed by a thorough ML search (-f a option of *RAxML*). Sequences of *phyB* (91, part of exon 1), ITS (210) and the *nadpme* and *ppc* families (116 and 145, exons 2–17 and 2–9, respectively) were each aligned using *MAFFT*, using the *E-INS-i* algorithm for the indel-rich ITS. *RAxML* was then used to estimate ML phylogenies, with parameters as stated above but using a Gamma model of rate heterogeneity.

Divergence time estimation using *BEAST2* was performed on the CDS-only plastome dataset. We specified a normal prior with a mean of 54.9 Ma and a standard deviation of 5.8 Ma for the BOP–PACMAD split, based on a previous Poaceae-wide study using external angiosperm calibration (Christin et al. 2014). We applied a GTR substitution model with four rate categories, a Yule tree prior and a log-normal relaxed clock. Two Markov Chain Monte Carlo runs with 20 million generations were performed, sampling parameter estimates and trees every 10,000 generations. The log files were examined in *Tracer* v. 1.6 (http://tree.bio.ed.ac.uk/software/tracer/) to ensure adequate mixing and Effective Sample Sizes of >200 for all parameters (except rate variation among branches which is difficult to estimate, P.-A. Christin, pers. comm.). A Maximum Clade Credibility Tree was computed from the combined chains, discarding a 10% burn-in each, using *TreeAnnotator* from the

BEAST2 package. Both *RAxML* and *BEAST2* analyses were performed on the computing cluster of the *Évolution & Diversité Biologique* laboratory, Toulouse.

Codon changes in ppc-B2 and rbcL

We compared amino-acid sequences of the nuclear gene *ppc-B2* (coding for phosphoenolpyruvate carboxylase) and the chloroplast gene *rbcL* (coding for the large subunit of Rubisco) between accessions in our alignments. Both genes have been found to be under C_4-related positive selection in grasses, with 12 particular codons found in both genes that underwent convergent adaptive changes (Christin et al. 2007, Piot et al. 2017). We were interested in the amino-acid residues of *Chandrasekharania* and *Jansenella* at these sites, particularly with respect to their putative C_4 sister clade Andropogoneae–Arundinelleae.

Distribution and endemism or core Panicoideae lineages

We generated maps of global distribution and endemism for the lineages of the core Panicoideae. Species-level distribution data for TDWG level-3 botanical regions, which broadly correspond to countries or states, were obtained from the World Catalogue of Selected Plant Families (Clayton et al. 2016). Grass genera were assigned to tribes according to the most recent classification (Soreng et al. 2017). We retrieved the total number of species and the number of endemics per TDWG level 3 botanical region for the large tribes Andropogoneae, Arundinelleae, Paniceae, and Paspaleae. Location records marked as doubtful were not considered. Note that the regions have different area, but correcting for this is not simple because very small regions would distort the figures. Botanical region shapefiles were obtained from the Royal Botanic Gardens, Kew (https://www.kew.org/gis/tdwg/index.html, accessed on 10 August 2017). The *R* package *rgdal* (Bivand et al. 2017) was used to read shapefiles and plot distribution and endemism data. For *Gy. sagittatum* and *J. griffithiana*, we reported the level-3 regions they occur in. For the range-restricted *Ch. keralensis*, *J. neglecta* and *Le. madagascariensis*, we referred to the relevant literature (Bor 1955, Nair et al. 1982, Teerawatnanon and Hodkinson 2008, Yadav et al. 2010, Besnard et al. 2013, Ramachandran and Betty 2015).

4.4 Results

Sequence assembly

We generated new complete plastome sequences for 11 species, including *Chandrasekharania keralensis* and the two known species of *Jansenella*, using genome-skimming data from herbarium material. We also assembled low-copy nuclear genes, *phyB* and the *nadpme* and *ppc* gene families, for nine of these taxa, with varying success, and complete nuclear ribosomal clusters for *Ch. keralensis*, *J. griffithiana* and *J. neglecta*. *Jansenella griffithiana* and *Ch. keralensis* appeared to be homozygous at all sites, and sequences complete with introns were assembled for all genes. Mostly complete sequences were also obtained for *Lasurus scindicus*, *Lecomtella madagascariensis* and *Reynaudia filiformis*. Assembly was more challenging in other cases due to low read coverage and/or

polymorphic sites. The accessions of *Jansenella neglecta* and *Gynerium sagittatum* consistently revealed two copies of each gene. In *J. neglecta*, one copy was always very similar to *J. griffithiana*, which allowed to consistently separate them. In contrast, while copies could be separated by phasing the reads for each exon in *Gy. sagittatum*, it was impossible to track them through introns. In *Le. madagascariensis*, we found two copies for *nadpme-II* and *ppc-aL1a* but not for the other genes. The *ppc-B1* gene appeared to be pseudogenized in *Le. madagascariensis* and *Gy. sagittatum*, with long deletions and/or frameshift mutations. In *R. filiformis*, *ppc-B2* was potentially duplicated, with two variants of equal proportion in the first portion of the gene, and we assembled a consensus sequence. The *ppc-aL1b* gene could not be detected in *J. griffithiana*, despite good sequence coverage. Finally, three copies of *nadpme-IV* were detected in *A. nepalensis*.

Plastome phylogeny and divergence times

Table III.1. Estimated ages, in Ma, of major nodes in Panicoideae, based on a Bayesian analysis of plastome coding sequences in *BEAST2*. Given are the estimated median ages and the 95% high posterior density (HPD) interval.

		95% HPD interval	
Node	**median**	**lower**	**upper**
Jansenella-Chandrasekharania split	13.65	7.19	20.75
Jansenella-Chandrasekharania stem	22.55	16.18	29.25
Arundinelleae crown	17.63	11.86	24.24
Andropogoneae crown	13.48	9.6	17.86
Andropogoneae-Arundinelleae split	19.46	13.79	26.04
Paspaleae crown	24.2	17.5	31.02
Paniceae crown	23.74	17.07	31.22
Lecomtella stem	30.23	22.61	38.27
Gynerium stem	33.19	24.81	41.8
Panicoideae crown	36.74	27.54	45.99

We estimated phylogenetic relationships through ML on three different subsets of a complete plastome alignment. Relationships were mostly congruent between the three topologies, which all had high bootstrap support for most nodes (i.e. >90; summarised in Fig. III.1, full trees can be found in Fig. A-III.1). In Andropogoneae, several nodes separated by very short branches had weak (i.e. <70) bootstrap support, but *Lasiurus scindicus* was always found to be sister to the rest. Arundinelleae was always sister to Andropogoneae but its monophyly was weakly to moderately supported. *Chandrasekharania* and *Jansenella* always formed a well supported clade sister to Andropogoneae–Arundinelleae. *Reynaudia filiformis* was confirmed as sister to the remaining Paspaleae, and *Gynerium* as sister to the remaining core Panicoideae. The position of *Lecomtella varied among* the alignment subsets: when only CDS were used, *Lecomtella* diverged just after *Gynerium*, with high support, but when the full alignment or an alignment without gaps were used, its divergence was placed after that of Paniceae, with weak support. When *Gynerium* was excluded, as in Burke et al. (2016), *Lecomtella* was always sister to the rest of the core Panicoideae, but with weak support except for the CDS-only subset (Fig. A-III.1, D–F). Note that the branching order among the

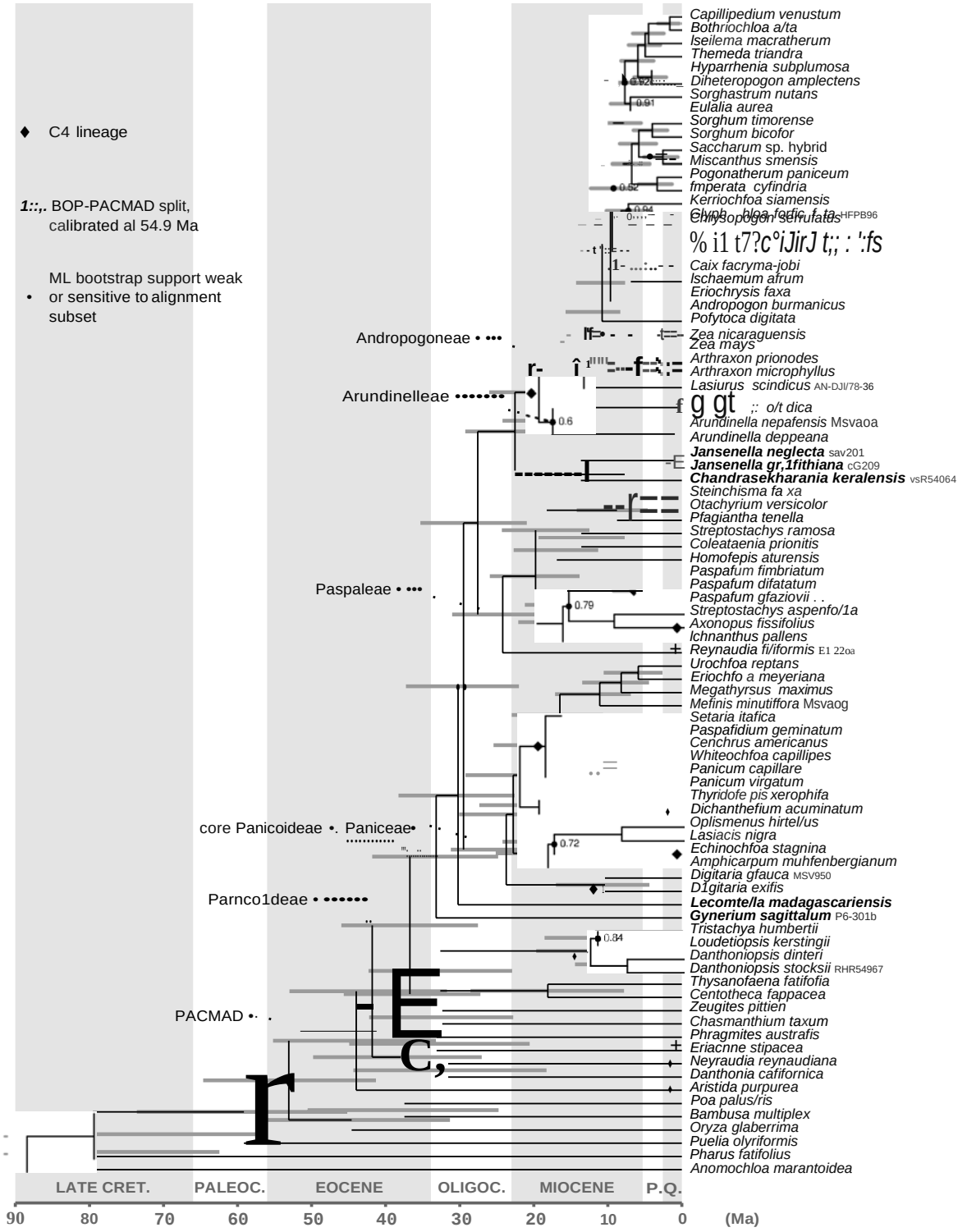

Figure III.1. Time-calibrated phylogeny of the core Panicoideae, based on complete plastome coding sequences. Shown is the maximum clade credibility tree from a BEAST2 analysis. Node heights are common ancestor ages; grey bars represent 95% high posterior density intervals. Values at nodes are Bayesian posterior probabilities smaller than 1. Nodes that received weak support with other alignment subsets are indicated. Focal species of this study are highlighted in bold.

PACMAD subfamilies also depends on subset choice, with either Aristidoideae or Panicoideae placed as sister to the others, with varying support.

The relationships found with Bayesian divergence time estimation on the CDS subset were congruent with the ML analysis (Fig. III.1). Table III.1 lists age estimates for selected nodes with 95% high posterior density (HPD) intervals for the two calibrations, and Fig. 1 shows median ages and HPD intervals for all nodes.

Nuclear gene phylogenies

The topologies estimated with nuclear sequences are partially congruent with those estimated with plastomes. The *phyB* tree confirmed *Chandrasekharania–Jansenella* as sister clade of Andropogoneae–Arundinelleae (Fig. III.2a, detailed version in Fig. A-III.2). One copy of the *J. neglecta phyB* gene was more closely related to the *J. griffithiana phyB* sequence than to the other copy from the same genome. *Lecomtella madagascariensis* diverged after Paniceae but with weak bootstrap support. The two *phyB* copies of *Gy. sagittatum* successively branched at the base of the core Panicoideae, one clustering with a *Gynerium phyB* sequence from *GenBank* (EU272433.1). Panicoideae was not monophyletic, and relationships within Panicoideae tribes and among PACMAD subfamilies differed strongly from the plastome topology. The ITS tree recovered the main lineages of the Panicoideae, but placed *Gy. sagittatum* with outlying Panicoideae and *Le. madagascariensis* with Paniceae and Paspaleae (Figs III.2b, A-III.3). *Chandrasekharania* and *Jansenella* form a strongly supported clade with Andropogoneae, *Arundinella* and *Garnotia* but are paraphyletic at its base, with weak support. *Arundinella* and *Garnotia* successively diverge next.

The topologies found for the *nadpme* and *ppc* gene families (Figs III.3, A-III.4–5) confirmed the relationships among genes found in previous studies (Christin et al. 2007, Christin et al. 2009b). Within the clade of almost every individual gene, *Jansenella–Chandrasekharania* was supported as sister group of Andropogoneae–Arundinelleae. Sequences of *ppc-B2* placed the two genera in a paraphyletic grade of C_3 species (plus *Stipagrostis*, which recruited *ppc-aL1b* instead for the C_4 function, Christin & Besnard 2009) basal to a C_4 cluster. In the clade of *ppc-B1*, absent from Andropogoneae–Arundinelleae, they were sister to a poorly supported clade of Paniceae, Paspaleae, and *Gy. sagittatum*. The *nadpme* and *ppc* copies of *Lecomtella* were mostly poorly supported as diverging before or after Paniceae, or as sister to either Paniceae or Paspaleae. The two copies of *ppc-aL1a* in *Lecomtella* were paraphyletic at the base of the core Panicoideae (*Gy. sagittatum* not being included). For *Gy. sagittatum*, only *nadpme-IV* could be included of the *nadpme* sequences, the others yielding only short exons. The two *nadpme-IV* copies were paraphyletic at the base of the core Panicoideae, one clustering with *Orthoclada laxa* (tribe Zeugiteae). Sequences of *ppc-aL1b*, *ppc-B1* and *ppc-aR* placed *Gy. sagittatum* sister to Paniceae, with poor support, while *ppc-aL2* supported it as sister to the rest of core Panicoideae.

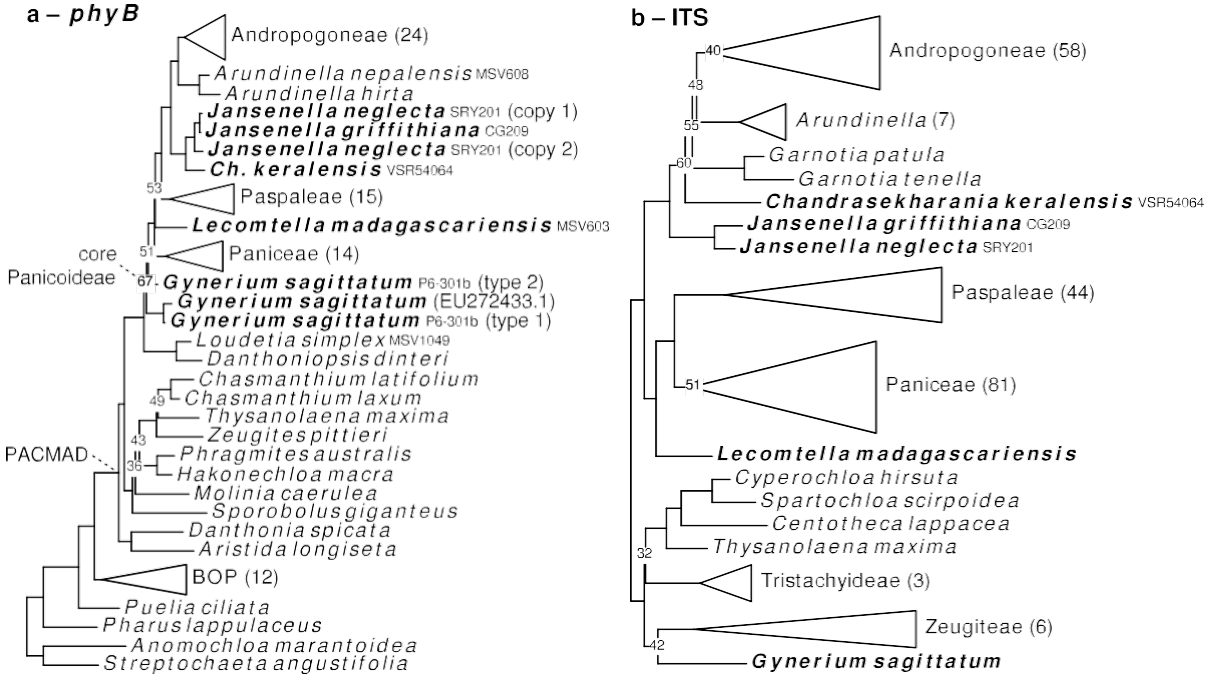

Figure III.2: Maximum Likelihood phylogeny of Poaceae *phyB* sequences. Values at nodes are bootstrap values smaller than 70 from 1000 replicates in RAxML. Larger clades are collapsed, with number of tips indicated. Focal species of this study are highlighted in bold.

Codon changes in ppc-B2 and rbcL

Amino-acid translations at sites previously shown to be under C_4-related selection in *ppc-B2* and *rbcL* are more similar to C_3 grasses in *Chandrasekharania* and *Jansenella* (Table III.2). Most importantly, site 780 of *ppc-B2* which is a serine (S) in most C_4 plants is an alanine (A) in all three species, as in all C_3 grasses and *Stipagrostis*. The same site also translates to alanine in all other *ppc* paralogs for the three species (not shown). Residues at other selected sites of *ppc-B2* also place them with C_3 groups and distinguish them from Andropogoneae–Arundinelleae. In *rbcL*, *Chandrasekharania* and *Jansenella* are likewise more similar to C_3 than to C_4 groups, with the exception of a lysine (K) at site 471, which is also the most frequent amino acid at this site in Andropogoneae but an alanine (A) in Arundinelleae. Note that in *La. scindicus*, *rbcL* encodes a protein with six additional amino acids at the C-terminal part, which were absent in all other Andropogoneae.

Distribution and endemism

Maps of distribution and endemism show quite different biogeographic patterns for the lineages of the core Panicoideae, all most diverse in tropical areas (Fig. III.4). While *G. sagittatum* is widespread in the Neotropics, *Le. madagascariensis* is confined to a single mountain range in southern Madagascar. The largest tribe, Paniceae, is pantropical but has its centres of diversity and endemism in tropical Africa and especially Madagascar. This pattern remains when the polyphyletic *Panicum* – accounting for almost one third of Paniceae, with some probably belonging to Paspaleae (Kellog 2015) – is excluded (not shown). Paspaleae is largely neotropical, with diversity and endemism peaking in west–central Brazil. In *Jansenella–Chandrasekharania*, *J. griffithiana* is known from India, Sri Lanka, Myanmar and

Thailand, while the other two species are restricted to the Western Ghats of southwest India, with only one record known of *Ch. keralensis*. Andropogoneae and Arundinelleae are both widespread but have their centres of diversity and endemism in India.

Table III.2. Amino acid residues at sites in the chloroplast gene *rbcL* and the nuclear gene *ppc-B2* previously shown to be under C_4-related positive selection (Christin et al. 2007, Piot et al. 2017). Shown are the dominant residues for selected grass clades (*Jansenella–Chandrasekharania* highlighted), with minor residues in brackets. Asterisks indicate that values could be determined for only one of the two *ppc-B2* copies in *Jansenella neglecta*. Note that C_3 and C_4 clades in Paniceae and Paspaleae were each grouped together. Codon positions refer to NC_001666.2 for *rbcL* and CAA33317.1 for *ppc-B2* (both *Zea mays*).

Clade/group												
	rbcL											
codon position	101	142	145	281	282	309	328	468	471	472	473	476
Andropogoneae (C_4)	V (I)	P	A (V/S)	S	H	M (I)	S	D	K (Q/E/V)	A (T)	M (Q/V)	L (I/V)
Arundinelleae (C_4)	V	P	A/S	A/S	H	I/M	A/S	D/E	A	P	V	I
***Jansenella-Chandrasekharania* (C_3)**	V	P	S	A	H	M	A	E	K	P	V	I
Paspaleae (C_3)	V	P (T)	S	A (S)	H	M	A	E (D)	A (K)	P	V	I
Paspaleae (C_4)	I (V)	P	S	S	H	I (M)	S	D (E)	D/K (E)	P	V	I
Paniceae (C_3)	V	P/T	S	A	H	M	A	E	A	P	V	I
Paniceae (C_4)	I (V)	P (V/I/T)	A/S	A (S)	H (Y)	M (I)	S (A)	D (E)	E (A/K/Q)	P (A)	V (M)	I (V/L)
Lecomtella (C_3)	V	P	S	A	H	M	A	E	A	P	V	I
Gynerium (C_3)	V	P	S	A	H	M	A	E	A	P	V	I
	ppc-B2											
codon position	466	517	531	560	577	579	625	637	761	780	794	807
Andropogoneae	I	A	P	P	S	E	A	L	A	S	V	K (R)
Arundinelleae	I	A	P	R	S	E	A	F	A	S	F	K
Jansenella-Chandrasekharania	L	T*	A	R*	A	A	V	M	S*	A*	F*	R*
Paspaleae (C_3)	V	T	A	R	A	A	V	M	S	A	F	R
Paspaleae (C_4)	I	A	P	P	A	E	V	M	S	S	F	K
Paniceae (C_3)	L	T	A	R	A	A	V	M	S	A	F	R
Paniceae (C_4)	I	C	P	P	S	E	A	M	A	S	V (F)	K
Lecomtella (C_3)	L	T	A	R	A	A	V	M	?	A	?	R
Chasmanthieae (C_3)	I	T	A	R	A	A	V	M	S	A	F	R
Chloridoideae (C_4)	V	C	P	P	S	E	A	F	A	S	V	K
Arundinoideae (C_3)	L	T	A	R	A	A	V	M	S	A	F	R
Aristidoideae: *Aristida* (C_4)	I	T	P	P	S	E	Q	F	S	S	I	K
Aristidoideae: *Stipagrostis* (C_4)	L	T	A	R	A	A	V	M	S	A	F	R
Pooideae (C_3)	L	T	A	R	A	A	V	M	S	A	F	R

4.5 Discussion

Phylogeny of the core Panicoideae and the roles of polyploidy and duplication

We here report a phylogenomic analysis of the core Panicoideae including for the first time full plastomes of the phylogenetically isolated *Gynerium sagittatum*, the enigmatic genera *Chandrasekharania* and *Jansenella* (see next section), the earliest-branching Paspaleae species *Reynaudia filiformis*, and *Lasiurus scindicus* which we found be the earliest-branching Andropogoneae species. We supplemented the plastome analysis with data from

nuclear genes. Overall, the results confirm previously found relationships in the core Panicoideae (GPWG II 2012, Burke et al. 2016) but still leave some uncertainty on the branching order or *Lecomtella* and Paniceae and the monophyly of Arundinelleae.

Gynerium sagittatum is confirmed as earliest-branching core panicoid by the plastome analysis, diverging in the Early Oligocene (Fig. III.1). Sequence assembly from shotgun data consistently revealed two divergent nuclear gene copies. While we could not include sufficiently long sequences of all *nadpme* and *ppc* paralogs in our phylogenies, the topologies of *phyB* (Fig. III.2) and *nadpme-IV* (Fig. III.3) revealed paraphyly of the two copies at the base of the core Panicoideae, suggesting allopolyploidy, though poorly supported in the case of *phyB*. Indeed, 44 chromosomes were determined for this species (Pohl & Davidse 1971) and interpreted as a basic number of $x = 11$ (Hilu 2004). *Gynerium sagittatum* is widespread in the Neotropics (Fig. III.4), but it is unknown whether all populations are polyploid. One of the two variants of *nadpme-IV* in *Gy. sagittatum* was placed with *Orthoclada laxa* (tribe Zeugiteae). That species is equally widespread in the Neotropics (Clayton et al. 2016), so a horizontal transfer of that gene could have occurred in an ancestor of the two species.

Lecomtella madagascariensis was previously placed in various positions in the core Panicoideae depending on the data analysed (Besnard et al. 2013, Burke et al. 2016). Here, we were still not able to fully resolve its position, showing that it depends on the choice of characters in the plastome alignment and the inclusion of *Gy. sagittatum* (Fig. A-III.1). Its divergence occurred either before or after Paniceae. The branches between these two lineages are extremely short, which explains the difficulties of resolving them. In any case, the stem of *Le. madagascariensis* would fall in the Early Oligocene (Fig. III.1, Table III.1), confirming it as the likely oldest endemic grass lineage in Madagascar (see chapter II). We also observed two paraphyletic copies of *ppc-aL1a* in *Le. madagascariensis* (Fig. III.3), suggesting an ancient duplication event or an allopolyploid origin of the lineage which has since been obscured by chromosomal rearrangements and homogenisation. The chromosome number of this species is unknown (Besnard et al. 2013).

Jansenella–Chandrasekharania, a C_3 sister group of Andropogoneae–Arundinelleae

Full plastomes and nuclear sequences show that *Chandrasekharania* and *Jansenella* are distinct from all other accepted tribes in Panicoideae and form a well-supported sister group of the large C_4 clade Andropogoneae–Arundinelleae. *Jansenella* had first been placed in Arundinelleae (which then also included what is now tribe Tristachyideae), but differed from most of its members by a combination of morphological characters: a punctiform hilum, membranaceous ligule, lemma of the upper floret with two tufts of hairs at the base of the two lobes, and especially the two-keeled, two-toothed palea of the upper floret with unusual one-celled hairs. These characters were observed in *J. griffithiana* and led to its segregation from *Danthoniopsis* by Bor (1955). The later described *J. neglecta* is very similar and was presumably confounded with *J. griffithiana* before (Yadav et al. 2010). *Chandrasekharania* was not assigned to any known group at its description but shared the morphology of the upper floret's lemma and the punctiform hilum (Nair et al. 1982).

Basic chromosome numbers are synapomorphic for many grass clades (Kellogg 2015), and in the core Panicoideae, Paspaleae was separated from the $x = 9$ Paniceae based on phylogeny and its basic chromosome number of $x = 10$ (Morrone et al. 2012). This ancestral number indeed appeared to be a synapomorphy for the Paspaleae–Andropogoneae–Arundinelleae clade, classified as supertribe Andropogonodae by Soreng et al. (2017). In *Jansenella–Chandrasekharania*, chromosome numbers have only been determined for *J. griffithiana* to our knowledge: Christopher & Abraham (1970) found a diploid number of 20, and Phipps & Mahon (1970) $2n = 40$, suggesting two ploidy levels. While data for the other two species would be desirable, this tentatively suggests an ancestral number of $x = 10$ for the *Jansenella–Chandrasekharania* clade and is in agreement with its position sister to Andropogoneae–Arundinelleae. We observed two copies of all nuclear genes studied in *J. neglecta*, of which one was always closer to *J. griffithiana* (Figs III.2 and III.3). The divergence of the two species' plastomes was estimated to the Quaternary (Fig. III.1). This strongly suggests that *J. neglecta* is an allopolyploid, with *J. griffthiana* or a recent ancestor as one of the parent species. It is possible that the $2n = 40$ specimen of Phipps and Mahon (1970) was in fact *J. neglecta*. Its source locality, Mahabaleshwar, falls into the known range of that species in the Western Ghats, but both species appear to be widely sympatric (Yadav et al. 2010).

Chandrasekharania and *Jansenella* were thought to have a C_3 photosynthetic pathway based on leaf anatomy: in an early investigation in Arundinelleae s.l. (including Tristachyideae), Conert (1957) observed a radiate, Kranz-like chlorenchyma and single bundle sheaths, as in other members of the tribe, for *J. griffithiana*, and this was cited without further analyses by Metcalfe (1960). Türpe (1970) performed a detailed leaf anatomical study in *J. griffithiana*. She observed double bundle sheaths, with few chloroplasts, but did not discuss photosynthetic pathway. Türpe's drawings were cited as evidence for a C_3 pathway by Clayton & Renvoize (1986), who also reported that the leaf anatomy of *Ch. keralensis* was "almost identical" to *J. griffithiana*. Renvoize (1982) confirmed the (partially) radiate chlorenchyma for *J. griffithiana* but identified it as C_3 because of an adaxial palisade chlorenchyma and more than four cells between vascular bundles. Watson et al. (2017) reported *Jansenella* as "supposedly C_3" based on the drawings of Türpe (1970) but added that further study was needed.

Here, we found that the *ppc-B2* gene recruited for primary carbon fixation in most C_4 species shows a C_3 signature in *Ch. keralensis* and the two species of *Jansenella*. This also explains the position of their *ppc-B2* sequences in a basal grade of C_3 grasses in the *ppc-B2* phylogeny, distinct from the C_4 cluster (Fig. III.3; Christin et al. 2007). Most importantly, no serine, typically present in C_4 grasses (Christin et al. 2007), was found at position 780 in either *ppc-B2* or any of the other *ppc* paralogs. Sites in the chloroplast gene *rbcL* under selection in C_4 grasses (Piot et al. 2017) also showed a signature closer to C_3 species. This genetic evidence thus supports both genera as C_3 grasses, but further data such as carbon isotope ratios would be useful. Some of the presumably intermediate anatomic traits could be explained by inheritance from the common ancestor with Andropogoneae–Arundinelleae that

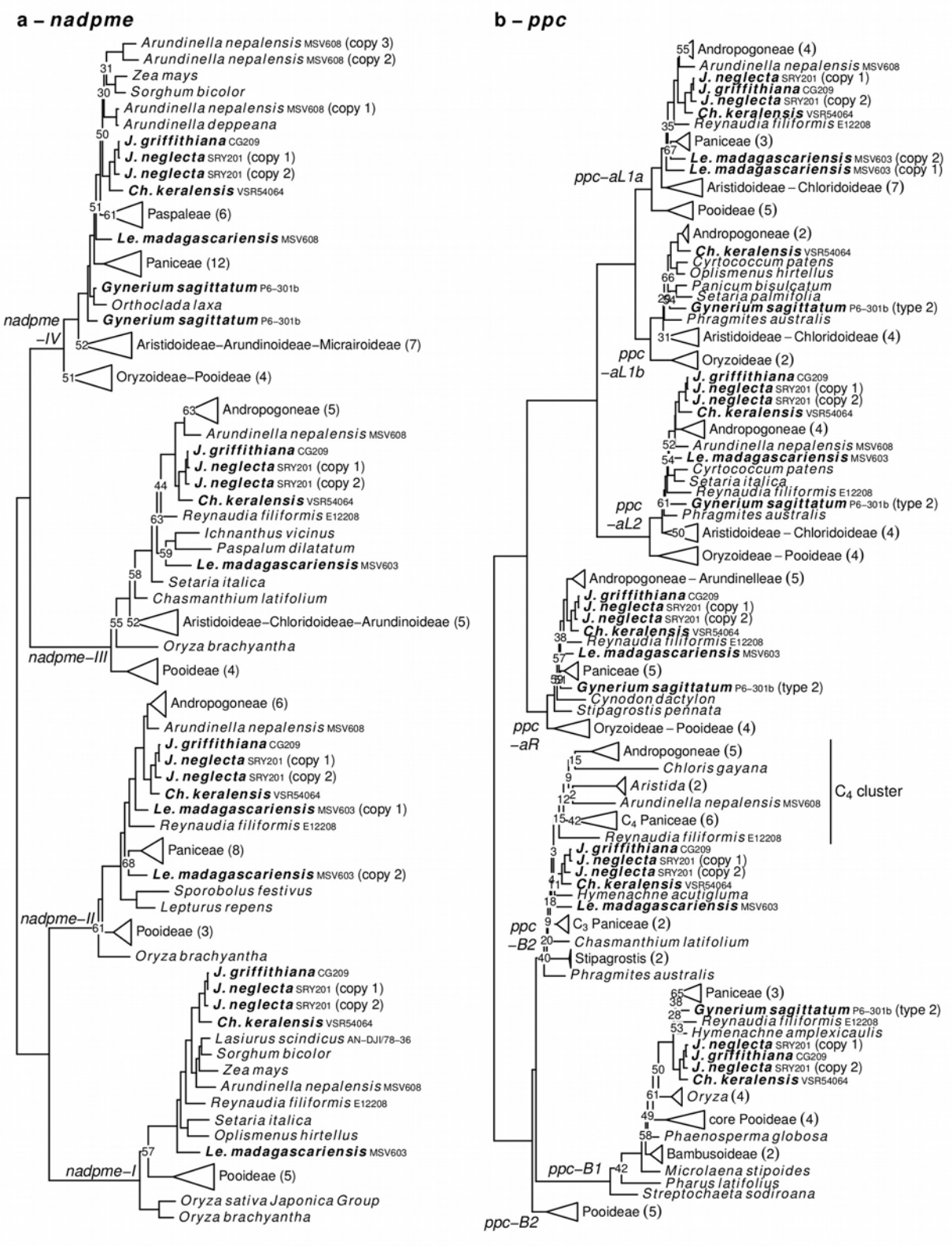

Figure III.3: Maximum Likelihood phylogenies of the *nadpme* (A) and *ppc* (B) gene families in Poaceae. Values at nodes are bootstrap values smaller than 70 from 1000 replicates in RAxML. Larger clades are collapsed, with number of tips indicated. Focal species of this study are highlighted in bold.

may have already featured some C_4 predispositions. Such "anatomical enablers", present in C_4 taxa and their close relatives, are common in the PACMAD clade (Christin et al. 2013).

Taken together, the phylogenetic position of *Chandrasekharania* and *Jansenella* sister to Andropogoneae–Arundinelleae, in combination with their morphology, karyology and photosynthetic pathway would justify recognising them as separate tribe within Panicoideae. This newly discovered sister group will be of interest for comparative studies on the genetic basis of C_4 photosynthesis in Andropogoneae–Arundinelleae (Besnard & Christin 2010). Given the rarity especially of *Ch. keralensis*, the phylogenetic uniqueness of this clade also underlines its conservation value and calls for better data on its distribution.

Biogeography of Andropogoneae–Arundinelleae and the origin of Indian grasslands

The C_4 sister tribes Andropogoneae and Arundinelleae both have their highest diversity and endemism on the Indian subcontinent (Fig. III.4). Their sister group *Jansenella–Chandrasekharania* has two species endemic to India, and one found from India to South-east Asia. These patterns, although not in themselves sufficient evidence, suggest that the divergence of these clades, and the transition to C_4 photosynthesis in Andropogoneae–Arundinelleae, occurred on the subcontinent. Hartley (1958) already noted the high diversity of Andropogoneae in India. Further evidence supports an Indian origin: the species *Lasiurus scindicus*, here supported as the earliest-diverging Andropogoneae, has two cytotypes, of which the $2n = 18$ one is found in India and Pakistan and the $2n = 56$ one in North Africa (Faruqi et al. 1987), suggesting westward dispersal from Asia. A South Asian origin was inferred for both the core Andropogoneae genus *Themeda* and the species *Th. triandra*, one of the dominant savannah grasses from Africa through Australasia (Dunning et al. submitted).

India has been connected to the rest of Eurasia since around 40–50 Ma (Ali & Aitchison 2008, Bouilhol et al. 2013) but is isolated through the Himalayas to the north, which restrict dispersal of tropical taxa and favoured radiations on the subcontinent (Karanth 2015). The monsoon climate which affects most of the Indian Ocean region originated in the Late Eocene, creating seasonally alternating high rainfall and aridity (Licht et al. 2014). This seasonality increased with intensified Himalayan orogeny in the Mid-Miocene, around 15 Ma (Ganjoo & Shaker 2007). The first stable-isotope evidence for C_4 plants in the Siwaliks of India and Pakistan date to 9–6 Ma (Quade & Cerling 1995, Nelson 2007, Sanyal et al. 2010, Singh et al. 2011) and are concordant with the global rise of C_4 grasslands in the Late Miocene (Edwards et al. 2010). We here estimated the divergence of *Jansenella–Chandrasekharania* from Andropogoneae–Arundinelleae to the Early Miocene, around 22.6 Ma (Fig. III.1, Table III.1). The split of the two C_4 tribes occurred roughly 3 Myr later, with a median of 19.5 Ma and a lower 95% HPD bound of 13.8 Ma. The transition to C_4 photosynthesis would thus have occurred several million years before the Late Miocene grassland spread and the first isotopic C_4 evidence on the subcontinent, but at a time of increased seasonality. The later C_4 diversification would however have been concomitant with grassland expansion. Estep et al. (2014) estimated that many Andropogoneae lineages had an allopolyploid origin, most of them in the Late Miocene or after. Evidence from other taxa for Indian grasslands dating to

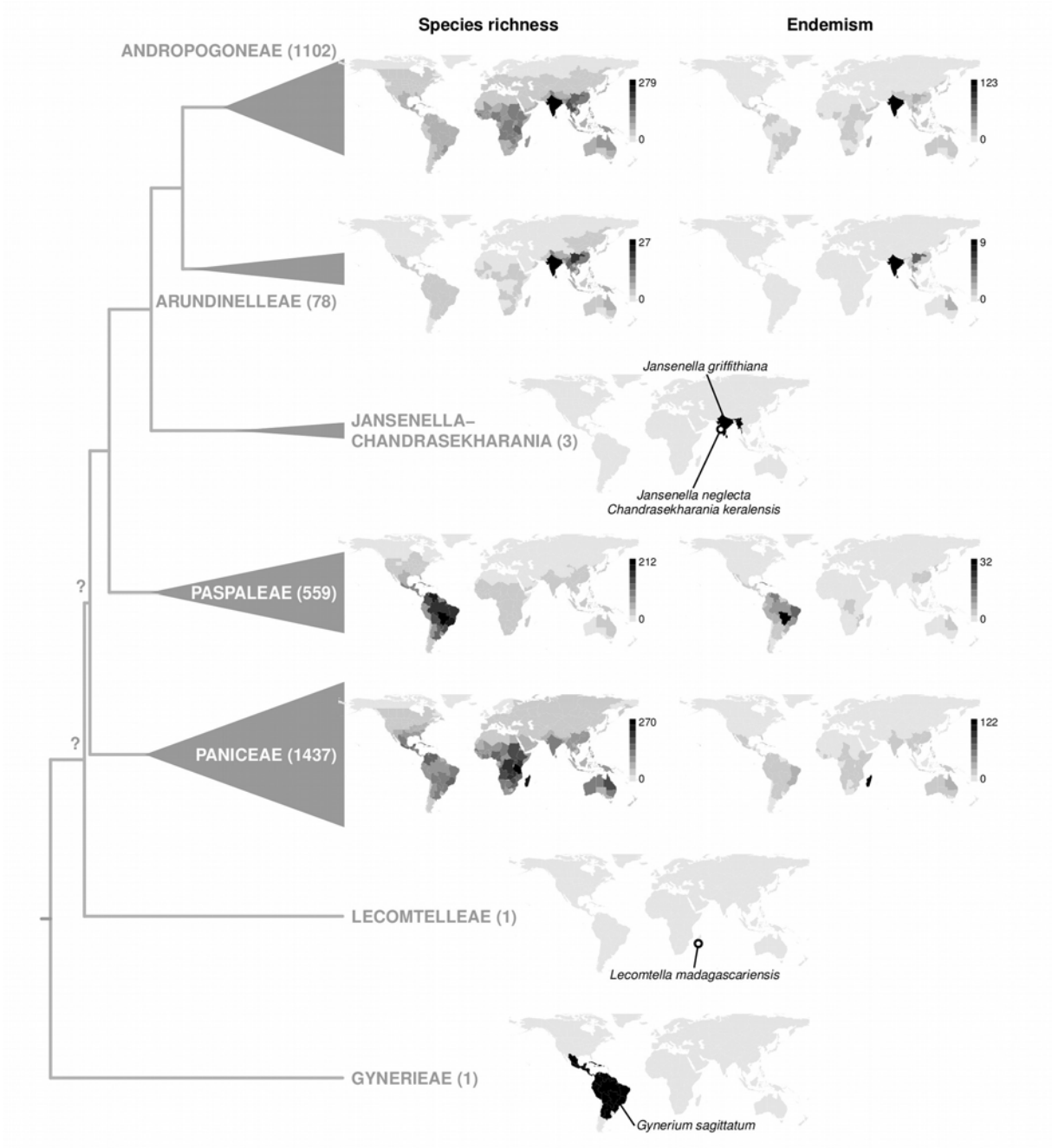

Fig. III.4: Distribution and endemism of major lineages in the core Panicoideae. Shown are species numbers and numbers of endemics per TDWG level 3 botanical region, extracted from the World Catalogue of Selected Plant Families (Clayton et al. 2016). Value ranges are given in numbers of species. The phylogeny as suggested by plastome sequences is shown to the left, with question marks indicating unclear branching order. Species number per lineage is given in brackets.

this period is the Late Miocene increase of diversification in open-habitat *Ophisops* lizards in India (Agarwal & Ramakrishnan 2017).

A monsoon climate is thought to have been favourable to the evolution of C_4 photosynthesis (Sage et al. 2012) and was in place when Andropogoneae–Arundinelleae diverged. Andropogoneae in particular are mostly species of mesic habitats, adapted to fast accumulation of biomass during the wet season which then promotes fire in the dry season (Osborne 2008). Within India, the Western Ghats mountain range captures much of the monsoon precipitation and features old-growth and diverse grasslands (Vasanthy 1988, Sukumar et al. 1995, Sankaran 2009). We could not further resolve patterns of diversity within India here because most of the country constitutes one single TDGW level-3 region (Fig. III.4). However, much of the diversity of Andropogoneae also appears to be concentrated in the west of the country (Hartley 1958). The Western Ghats, a global hotspot of biodiversity, seem to have been the centre of radiations in Andropogoneae, e.g. in the genus *Glyphochloa* (Gosavi et al. 2016). Three presumably endemic *Arundinella* species were also recently described from the region (Sunil & Kumar 2014, Sunil et al. 2014, 2017).

In the newly found C_3 sister group, *J. neglecta* and *Ch. keralensis* are both restricted to the Western Ghats (Nair et al. 1982, Yadav et al. 2010). Only *J. griffithiana* has a wider distribution, but always occurs at higher altitudes, and is widespread also in the Western Ghats (Yadav et al. 2010). All three are annual, open-habitat species. In the Khasi hills of northern India (Megalaya), *J. griffithiana* was recorded in diverse grasslands, where it appeared to be slightly negatively affected by fire and grazing (Shilla & Tiwari 2015). The *Jansenella–Chandrasekharania* lineage may have emerged in the first open grasslands before they became dominated by C_4 grasses. Possibly, the transition from C_3 to C_4 photosynthesis in Andropogoneae–Arundinelleae may have occurred in montane, seasonally humid grasslands resembling those found today in the Western Ghats. The C_4 clade would then have diversified into lower areas during the Late Miocene. Alternatively, the C_3 clade may have been present at lower altitudes but retreated to montane habitats after being outcompeted by the now more diverse C_4 grasses.

To conclude, the sister relationship of *Jansenella–Chandrasekharania* and Andropogoneae–Arundinelleae suggests that the transition to C_4 and the initial diversification of the large C_4 clade occurred in India, potentially in the seasonally humid environment of the Western Ghats. From there, it would have spread into and dominated savannahs in Asia and other continents. Phylogenies with a sufficient species sampling for this large clade would allow to formally test this hypothesis.

4.6 References

Agarwal I. & Ramakrishnan U. (2017): A phylogeny of open-habitat lizards (Squamata: Lacertidae: *Ophisops*) supports the antiquity of Indian grassy biomes. *Journal of Biogeography*, 44: 2021–2032.

Ali J.R. & Aitchison J.C. (2008): Gondwana to Asia: Plate tectonics, paleogeography and the biological connectivity of the Indian sub-continent from the Middle Jurassic through latest Eocene (166–35 Ma). *Earth-Science Reviews*, 88: 145–166.

Arthan W., McKain M.R., Traiperm P., Welker C.A.D., Teisher J.K. & Kellogg E.A. (2017): Phylogenomics of Andropogoneae (Panicoideae: Poaceae) of mainland Southeast Asia. *Systematic Botany*, 42: 418–431.

Barker N.P., Linder H.P. & Harley E.H. (1995): Polyphyly of Arundinoideae (Poaceae): Evidence from *rbcL* sequence data. *Systematic Botany*, 20: 423–435.

Besnard G. & Christin P.-A. (2010): Evolutionary genomics of C_4 photosynthesis in grasses requires a large species sampling. *Comptes Rendus Biologies*, 333: 577–581.

Besnard G., Christin P.-A., Malé P.-J.G., Coissac E., Ralimanana H. & Vorontsova M.S. (2013): Phylogenomics and taxonomy of Lecomtelleae (Poaceae), an isolated panicoid lineage from Madagascar. *Annals of Botany*, 112: 1057–1066.

Bivand R., Keitt T. & Rowlingson B. (2017): *rgdal: Bindings for the Geospatial Data Abstraction Library*.

Bläsing O.E., Westhoff P. & Svensson P. (2000): Evolution of C_4 phosphoenolpyruvate carboxylase in *Flaveria*, a conserved serine residue in the carboxyl-terminal part of the enzyme is a major determinant for C_4-specific characteristics. *Journal of Biological Chemistry*, 275: 27917–27923.

Bond W.J. (2016): Ancient grasslands at risk. *Science*, 351: 120–122.

Bor N.L. (1955): Notes on Asiatic grasses: XXIII. *Jansenella* Bor, a new genus of Indian grasses. *Kew Bulletin*, 10: 93–99.

Bouilhol P., Jagoutz O., Hanchar J.M. & Dudas F.O. (2013): Dating the India–Eurasia collision through arc magmatic records. *Earth and Planetary Science Letters*, 366: 163–175.

Burke S.V., Wysocki W.P., Zuloaga F.O., Craine J.M., Pires J.C. et al. (2016): Evolutionary relationships in Panicoid grasses based on plastome phylogenomics (Panicoideae; Poaceae). *BMC Plant Biology*, 16: 140.

Christin P.-A. & Besnard G. (2009): Two independent C_4 origins in Aristidoideae (Poaceae) revealed by the recruitment of distinct phosphoenolpyruvate carboxylase genes. *American Journal of Botany*, 96: 2234–2239.

Christin P.-A., Osborne C.P., Chatelet D.S., Columbus J.T., Besnard G. et al. (2013): Anatomical enablers and the evolution of C_4 photosynthesis in grasses. *Proceedings of the National Academy of Sciences*, 110: 1381–1386.

Christin P.-A., Salamin N., Kellogg E.A., Vicentini A. & Besnard G. (2009a): Integrating phylogeny into studies of C_4 variation in the grasses. *Plant Physiology*, 149: 82–87.

Christin P.-A., Salamin N., Savolainen V., Duvall M.R. & Besnard G. (2007): C_4 photosynthesis evolved in grasses via parallel adaptive genetic changes. *Current Biology*, 17: 1241–1247.

Christin P.-A., Samaritani E., Petitpierre B., Salamin N. & Besnard G. (2009b): Evolutionary insights on C_4 photosynthetic subtypes in grasses from genomics and phylogenetics. *Genome Biology and Evolution*, 1: 221–230.

Christin P.-A., Spriggs E., Osborne C.P., Strömberg C.A.E., Salamin N. & Edwards E.J. (2014): Molecular dating, evolutionary rates, and the age of the grasses. *Systematic Biology*, 63: 153–165.

Christopher J. & Abraham A. (1970): Studies on the cytology and phylogeny of South Indian grasses. I. Subfamilies Bambusoideae, Oryzoideae, Arundinoideae and Festucoideae. *Cytologia*, 36: 579–592.

Clayton W.D., Govaerts R., Harman K.T., Williamson H. & Vorontsova M. (2016): *World Checklist of Poaceae, facilitated by the Royal Botanic Gardens, Kew.* Available at: http://apps.kew.org/wcsp/.

Clayton W.D. & Renvoize S.A. (1986): Genera graminum. Grasses of the World. Royal Botanic Gardes, Kew, Richmond, Surrey.

Conert H.J. (1957): Beiträge zur Monographie der Arundinelleae. *Botanische Jahrbücher*, 77: 226–353.

Dunning L.T., Liabot A.-L., Olofsson J.K., Smith E.K., Vorontsova M.S. et al. (submitted): Paleotropical grassland dominance out of Asian savannas.

Edwards E.J., Osborne C.P., Strömberg C.A.E. & Smith S.A. (2010): The origins of C_4 grasslands: integrating evolutionary and ecosystem science. *Science*, 328: 587–591.

Estep M.C., McKain M.R., Diaz D.V., Zhong J., Hodge J.G. et al. (2014): Allopolyploidy, diversification, and the Miocene grassland expansion. *Proceedings of the National Academy of Sciences*, 111: 15149–15154.

Faruqi S., Quraish H. & Inamuddin M. (1987): Studies in Libyan grasses X: Chromosome number and some interesting features. *Annali di Botanica*, 45: 75–102.

Ganjoo R.K. & Shaker S. (2007): Middle Miocene pedological record of monsoonal climate from NW Himalaya (Jammu & Kashmir State), India. *Journal of Asian Earth Sciences*, 29: 704–714.

Gosavi K.V.C., Yadav S.R., Praveen Karanth K. & Surveswaran S. (2016): Molecular phylogeny of *Glyphochloa* (Poaceae, Panicoideae), an endemic grass genus from the Western Ghats, India. *Journal of Systematics and Evolution*, 54: 162–174.

GPWG II – Grass Phylogeny Working Group II (2012): New grass phylogeny resolves deep evolutionary relationships and discovers C_4 origins. *New Phytologist*, 193: 304–312.

Hartley W. (1958): Studies on the origin, evolution and distribution of the Gramineae. I. The tribe

Andropogoneae. *Australian Journal of Botany*, 6: 116–128.

Hilu K.W. (2004): Phylogenetics and chromosomal evolution in the Poaceae (grasses). *Australian Journal of Botany*, 52: 13–22.

Karanth P.K. (2015): An island called India: phylogenetic patterns across multiple taxonomic groups reveal endemic radiations. *Current Science*, 108: 1847–1851.

Katoh K. & Standley D.M. (2013): MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, 30: 772–780.

Kellogg E.A. (2015): *Flowering Plants. Monocots: Poaceae*. Springer, Heidelberg.

Lehmann C.E.R. & Parr C.L. (2016): Tropical grassy biomes: linking ecology, human use and conservation. *Philosophical Transaction of the Royal Society B: Biological Sciences*, 371: 20160329.

Licht A., Van Cappelle M., Abels H., Ladant J., Trabucho-Alexandre J. et al. (2014): Asian monsoons in a late Eocene greenhouse world. *Nature*, 513: 501.

Metcalfe C.R. (1960): *Anatomy of the monocotyledons. I. Gramineae*. Clarendon press, Oxford.

Morrone O., Aagesen L., Scataglini M.A., Salariato D.L., Denham S.S. et al. (2012): Phylogeny of the Paniceae (Poaceae: Panicoideae): integrating plastid DNA sequences and morphology into a new classification. *Cladistics*, 28: 333–356.

Müller C. (1856): Manipulus Graminearum novarum. *Botanische Zeitung (Berlin)*, 14: 345–350.

Nair V.J., Ramachandran V.S. & Sreekumar P.V. (1982): *Chandrasekharania*: A new genus of Poaceae from Kerala, India. *Proceedings of the Indian Academy of Science (Plant Science)*, 91: 79–82.

Nelson S.V. (2007): Isotopic reconstructions of habitat change surrounding the extinction of *Sivapithecus*, a Miocene hominoid, in the Siwalik Group of Pakistan. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 243: 204–222.

Osborne C.P. (2008): Atmosphere, ecology and evolution: what drove the Miocene expansion of C_4 grasslands? *Journal of Ecology*, 96: 35–45.

Osborne C.P., Salomaa A., Kluyver T.A., Visser V., Kellogg E.A. et al. (2014): A global database of C_4 photosynthesis in grasses. *New Phytologist*, 204: 441–446.

Parr C.L., Lehmann C.E.R., Bond W.J., Hoffmann W.A. & Andersen A.N. (2014): Tropical grassy biomes: misunderstood, neglected, and under threat. *Trends in Ecology & Evolution*, 29: 205–213.

Phipps J.B. & Mahon J.B. (1970): Studies in the Arundinelleae (Gramineae): IX. Chromosome numbers of 21 species. *Canadian Journal of Botany*, 48: 1419–1423.

Piot A., Hackel J., Christin P.-A. & Besnard G. (2017): One third of the plastid genes evolved under positive selection in PACMAD grasses. *Planta*: in press.

Pohl R.W. & Davidse G. (1971): Chromosome numbers of Costa Rican grasses. *Brittonia*, 23: 293–324.

Quade J. & Cerling T.E. (1995): Expansion of C_4 grasses in the Late Miocene of Northern Pakistan: evidence from stable isotopes in paleosols. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 115: 91–116.

Ramachandran V.S. & Betty T. (2015): On the occurrence of *Jansenella neglecta* Yadhav et al., and *Polypogon nilgiricus* Kabeeer & V.J.Nair (Poaceae) from the Nilgiris, Tamil Nadu. *International Letters of Natural Sciences*, 31: 57–59.

Ratnam J., Tomlinson K.W., Rasquinha D.N. & Sankaran M. (2016): Savannahs of Asia: antiquity, biogeography, and an uncertain future. *Phil. Trans. R. Soc. B*, 371: 20150305.

Renvoize S.A. (1982): A survey of leaf-blade anatomy in grasses II. Arundinelleae. *Kew Bulletin*, 37: 489–495.

Sage R.F., Sage T.L. & Kocacinar F. (2012): Photorespiration and the evolution of C_4 photosynthesis. *Annual Review of Plant Biology*, 63: 19–47.

Sánchez-Ken J.G. & Clark L.G. (2010): Phylogeny and a new tribal classification of the Panicoideae s.l. (Poaceae) based on plastid and nuclear sequence data and structural data. *American Journal of Botany*, 97: 1732–1748.

Sankaran M. (2009): Diversity patterns in savanna grassland communities: implications for conservation strategies in a biodiversity hotspot. *Biodiversity and Conservation*, 18: 1099–1115.

Sanyal P., Sarkar A., Bhattacharya S.K., Kumar R., Ghosh S.K. & Agrawal S. (2010): Intensification of monsoon, microclimate and asynchronous C_4 appearance: Isotopic evidence from the Indian Siwalik sediments. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 296: 165–173.

Shilla U. & Tiwari B.K. (2015): Impact of fire and grazing on plant diversity of a grassland ecosystem of Cherrapunjee. *Keanean Journal of Science*, 4: 67–78.

Singh S., Parkash B., Awasthi A.K. & Kumar S. (2011): Late Miocene record of palaeovegetation from Siwalik palaeosols of the Ramnagar sub-basin, India. *Current Science*, 100: 213–222.

Soreng R.J., Peterson P.M., Romaschenko K., Davidse G., Teisher J.K. et al. (2017): A worldwide phylogenetic classification of the Poaceae (Gramineae) II: An update and a comparison of

two 2015 classifications. *Journal of Systematics and Evolution*, 55: 259–290.

Soreng R.J., Peterson P.M., Romaschenko K., Davidse G., Zuloaga F.O. et al. (2015): A worldwide phylogenetic classification of the Poaceae (Gramineae). *Journal of Systematics and Evolution*, 53: 117–137.

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

Strömberg C.A.E. (2011): Evolution of grasses and grassland ecosystems. *Annual Review of Earth and Planetary Sciences*, 39: 517–544.

Sukumar R., Suresh H.S. & Ramesh R. (1995): Climate change and its impact on tropical montane ecosystems in Southern India. *Journal of Biogeography*, 22: 533–536.

Sunil C.N. & Kumar V.V.N. (2014): A new species of *Arundinella* (Poaceae: Panicoideae: Arundinelleae) from Kerala, India. *Webbia*, 69: 249–252.

Sunil C.N., Narayanan M.K.R., Parameswaran P., Sivadasan M. & Alfarhan A.H. (2014): A new species of *Arundinella* Raddi (Poaceae) from the Western Ghats, India. *Bangladesh Journal of Plant Taxonomy*, 21: 153–157.

Sunil C.N., Prabhukumar K.M., Kumar V.V.N., Thomas V.P., Jose J. & Balachandran I. (2017): *Arundinella muthikulamensis* (Poaceae), a new species from the Western Ghats of Kerala, India. *Webbia*, 72: 101–104.

Teerawatnanon A. & Hodkinson T.R. (2008): *Jansenella griffithiana* (Müll. Hal.) Bor (Gramineae/Poaceae): a new record for Thailand, and notes on its typification. *Thai Forest Bulletin (Botany)*, 36: 109–113.

Türpe A.M. (1970): Sobre la anatomía foliar de *Jansenella griffithiana*. *Senckenbergiana biologica*, 51: 277–285.

Vasanthy G. (1988): Pollen analysis of late quaternary sediments: Evolution of upland savanna in Sandynallah (Nilgiris, South India). *Review of Palaeobotany and Palynology*, 55: 175–192.

Watson L., Macfarlane T.D. & Dallwitz M.J. (2017): Grass genera of the world. Available at: delta-intkey.com/grass.

Yadav S.R., Chivalkar S.A. & Kosavi K.V.C. (2010): On the identity of *Jansenella griffithiana* (Poaceae) with a new species from Western Ghats, India. *Rheedea*, 20: 38–43.

5 Chapter IV: A metabarcoding method for the analysis of grass endophyte communities in Madagascar[1]

5.1 Abstract

Fungal endophytes, living in plant tissues without causing disease, are widespread. In grasses, there are well-described examples of positive effects of leaf endophytes on the host plant that can even impact community and food web composition. Diversity of grass endophytes is much higher than the well known model species, particularly in the tropics, but community composition and structuring factors are unknown in most natural settings. Here, we developed a metabarcoding approach to study grass leaf endophyte communities in three regions of highland Madagascar. A total of 768 samples were collected from 58 grass species in three habitat types: open grassland, tapia woodland and forest. We amplified and sequenced the ITS2 region to describe fungal communities. Results show preferential amplification of grass DNA and recovered only few fungal OTUs. Grass ITS2 sequences recovered allowed testing the original species identification and highlighted some possible errors. Grass communities both based on original identification and ITS2 clustering clearly distinguished forest from grassland and tapia woodland. Among the few fungal OTUs recovered are species of *Epichloë* associating with temperate Pooideae grasses and recorded for the first time in Madagascar. While analysis of fungal communities was limited, the results suggest that our preservation method yielded sufficient DNA for metabarcoding, that the structure of mock communities was repeatable across PCRs, and that technical biases such as tag switching were manageable. Our dataset has rich associated ecological data and would benefit from re-analysis with more specific primers.

5.2 Introduction

Plants, like all higher organisms, associate with microbes (Zilber-Rosenberg & Rosenberg 2008). Symbioses with bacteria or fungi play a central role in the interaction of plants with their environment, through beneficial functions like improved biotic or abiotic resistance, or as pathogens (Saikkonen et al. 1998, Reinhold-Hurek & Hurek 2011). Microbes growing within living tissues of plants without causing apparent symptoms, termed endophytes, are believed to cover a wide array of functions (Wilson 1995). Fungal endophytes in particular include well-characterised cases of mutualistic relationships, for example protection against stress through heat (Redman et al. 2002), salinity (Rodriguez et al. 2008), pathogens (Arnold et al. 2003), or herbivory (Clay et al. 1985). However, mutualism may shift to antagonism under certain environmental conditions (Saikkonen et al. 1998), and many fungal endophytes are believed to be latent pathogens (Wilson 1995). Whatever their precise function, bioactive compounds produced by endophytes are seen as a promising reservoir for natural product research (Strobel 2003, Guo et al. 2008).

1 Results of this chapter have not been published elsewhere when this dissertation was completed. The study was devised by G. Besnard and myself. M. S. Vorontsova and H. Ralimanana helped with the selection of sample sites and the identification of grass species (M.S.V.). H. Andrianirina and J. Razanatsoa helped with sampling in the field. G.B. collected samples in the field and amplified and sequenced *matK* grass barcodes. S. Manzi prepared libraries for sequencing. My contribution was field sampling, DNA extraction, PCR with fungal primers and data analysis. Grass community data from nine plots sampled here have been included in a study entitled "Grass traits and species composition differentiate tapia woodlands from forest in Madagascar", authored by C.L. Solofondranohatra, M.S. Vorontsova, J. Hackel, G. Besnard, S. Cable, V. Jeannoda and C.E.R. Lehmann and submitted to *Frontiers in Ecology and Evolution* in October 2017.

Fungal leaf endophytes associated with grasses (Poaceae), with the well-studied case of the ascomycete genus *Epichloë* (syn. *Neotyphodium*), have been a focus of research (Saikkonen et al. 2004). Several species of *Epichloë* have demonstrated positive effects on their host, increasing its fitness under various forms of abiotic stress and protecting it against herbivores and pathogens through alkaloid production (Clay & Schardl 2002). The symbiosis is characterised by co-evolution with vertical transmission through seeds, partial co-divergence of lineages and loss of sexual reproduction in the fungi (Schardl et al. 1997, Clay & Schardl 2002). Moreover, effects of the endophytes beyond the host level have been demonstrated in experiments: their presence was shown to impact the composition of plant communities (Clay et al. 1993, Clay & Holah 1999) and food webs (Omacini et al. 2001). They have also been of considerable interest to agronomists (Malinowski & Belesky 2006). However, most of this research was restricted to one model system, *E. coenophiala* associating with *Lolium* species, and its application to most natural systems remains to be tested (Saikkonen et al. 2006). *Epichloë* species are restricted to the temperate grass subfamily Pooideae (Clay & Schardl 2002), but grasses are believed to associate with a much larger array of fungi over their cosmopolitan distribution (Sánchez Márquez et al. 2012). A genus close to *Epichloë*, *Parepichloë*, was described from tropical grass species in Africa and Asia, but its function in the host is unknown (White & Reddy 1998).

Endophytes in the tropics are still poorly investigated but thought to much more diverse than at temperate latitudes (Arnold et al. 2000). Time-intensive culturing techniques have mostly limited research to the isolation and description of particular species, mostly from trees (e.g. Arnold et al. 2003, Costa et al. 2012). Among the few examples of community-wide studies in the tropics, work in Panama based on isolation and culturing showed that leaf endophyte communities associated with forest understorey grasses were dominated by host generalists and structured by spatial distance (Higgins et al. 2011, 2014). The rise of environmental metabarcoding has allowed to study communities at a larger scale and depth. For example, Zimmerman & Vitousek (2012) found that fungal communities of a single tree species varied strongly with environmental gradients in Hawaii. However, while advances have been made on the global biogeography and community ecology of soil fungi (Tedersoo et al. 2014) or bacteria (Martiny et al. 2006), the factors structuring endophyte communities at different scales remain largely unknown in most natural habitats and especially so in the tropics (Rodriguez et al. 2009).

Grasses cover a large part of the terrestrial surface, with tropical grasslands alone accounting for roughly 20% (Parr et al. 2014). Tropical grasslands are of global importance, sequestering carbon, feeding large stocks of herbivores, and sustaining human livelihoods. Progress has been made on their definition, ecology and biogeography, underlining the importance of herbivory and fire (Ratnam et al. 2011, Lehmann & Parr 2016). Microbial symbionts of grasses have not been integrated in frameworks of tropical grassland ecology. Yet, they may potentially impact productivity, species composition, and herbivore–plant interactions, whether as mutualistic or pathogenic symbionts. Basic information on fungal endophytes in the tropics and in tropical grasslands in particular is lacking, including on

taxonomic composition, host selectivity, and the importance of environmental versus spatial determinants of community assembly.

Here, we attempted to develop a metabarcoding method to study foliar fungal endophytes communities in grasses of highland Madagascar. Grasslands, in a mosaic with forest and woodland, dominate plateau landscapes of Madagascar, and are only beginning to receive more attention from ecologists and biogeographers (see Vorontsova et al. 2016 and chapter II). We sampled a large and phylogenetically diverse set of grasses from open grassland, woodland, and closed forest. Our goal was to establish whether the fungal communities present in the leaves could be analysed with a metabarcoding approach. This could then allow to study the importance of various structuring factors, such as host identity, host evolutionary history, habitat type, and distance.

5.3 Materials and Methods

Field sampling

Fig. IV.1: The three sampling regions in central Madagascar. The scale is in kilometres.

We collected leaves of grass species in 92 plots distributed over three regions of central Madagascar (Fig. IV.1) in April and May 2016. The highlands of central Madagascar are dominated by grasslands, in a mosaic with bushland, woodlands, and some remaining forests which are typically gallery forests along streams. One particular type of open woodland is tapia woodland, dominated by *Uapaca bojeri* (tapia) and other tree species, all endemic (Moat & Smith 2007). Forests and tapia woodland are primary vegetation, rich in diversity and endemics, while the status of the open grasslands is controversial (see chapter II). The three protected areas we collected in are: (i) Ankafobe (34 plots), a small reserve covering a remnant gallery forest and surrounding grassland next to a national road; (ii) Ibity (12 plots), a mountain massif featuring abundant tapia woodland; (iii) Itremo (46 plots), a large and relatively isolated quartzite plateau featuring grassland, tapia woodland, and larger remnants of humid gallery forest. All sampling sites ranged from approximately 1300 to 1650 m elevation a.s.l. Necessary permits for collection and sample export were obtained before fieldwork.

Plots were established during random walks in each region, ensuring at least 100 m distance between them. They were placed in three types of vegetation: (i) open grassland, (ii) tapia woodland and (iii) forest. Plots at the edge of forests, often disturbed, were also classified as forest. In every plot, we collected eight grass specimens within a 5-m radius. We

collected 1–4 specimens for 3–8 species per plot, depending on local diversity, with at least 1 m distance between specimens. We collected healthy-looking individuals, if possible with inflorescences, and stored them in individual plastic bags. Location, elevation, slope, estimated ground and canopy cover, the total number of grass species, and, if present, signs of disturbance and recent fires, were recorded for each plot.

Grass specimens were treated in the field camp maximum five hours after collection. For some, we kept voucher material (herbarium M. S. Vorontsova, K/TAN). We selected 3–5 intact leaves from each specimen. These were surface-sterilised by sequential washes of roughly 30 s each in 70% ethanol, 2% sodium hypochlorite and 90% ethanol, and finally rinsed with sterile water (following U'Ren et al. 2014 but with shorter immersion times). A segment of 2–5 mm was then cut from the middle of each leaf. Tweezers and scissors were sterilised by flaming over an alcohol lamp between specimens. Leaf segments were pooled per specimen and preserved in 2× CTAB buffer solution in a 200-µl well of a 96-well plate. Sample plates were stored in a styrofoam box buffered against heat with bottles of cold water. They were transferred to a refrigerator at ~8 °C upon the return from each field trip; the maximum time before refrigeration was 11 days. During air transit to France, samples were again kept in a styrofoam box and then stored at 4 °C.

DNA extraction, amplification and sequencing

DNA was extracted using the Wizard® Genomic DNA Purification Kit (Promega, Fitchburg, WI). For each sample, leaf segments were transferred with flame-sterilised tweezers to a tube with 300 µl of Nuclear Lysis Solution (NLS) and a 3-mm sterilised steel grinding ball. Samples were ground in a Tissue Lyser (Quiagen, Hilden) for 2×2 min, flipping them half-time through, at a speed of 30 s^{-1}. After 20 min waiting time at room temperature to reduce foam, the grinding balls were removed and another 200 µl NLS added. Tubes were then incubated at 65 °C for 30 min. After cooling down for 5 min, 50 µl of Protein Precipitation Solution were added, tubes were vortexed and then centrifuged for 15 min at 13,000 rpm. One hundred µl of the supernatant was transferred to a new tube with 600 µl isopropanol, avoiding foam from the top of the liquid. Tubes were inversed ten times and then centrifuged for 15 min at 13,000 rpm to precipitate DNA. The supernatant was discarded and 100 µl ethanol (95%) added to the pellet before centrifuging again for 10 min at 13,000 rpm. After discarding the supernatant, tubes were left open to dry for 3 h and DNA was then suspended in 100 µl nuclease-free water. Samples were handled under a laminar flow hood using filter-tip pipettes.

DNA was quantified with a Qubit fluorometer (Thermo Fisher Scientific, Waltham, MA) and samples were brought to equal concentration prior to amplification. Polymerase chain reaction (PCR) was performed using the primers ITS3 and ITS4 (White et al. 1990) which target the nuclear ribosomal internal transcribed spacer 2 region (ITS2) and have been employed in fungal metabarcoding (e.g. Blaalid et al. 2013, Tedersoo et al. 2015). Primers were tagged to allow for multiplexing, using different 8-bp tags for the forward and the reverse primer. The PCR mix per sample (total volume 15 µl) contained 7.5×10^{-3} µmol of each of the primers, 150 µg bovine serum albumin, 7.5 µl AmpliTaq Gold® 360 Master Mix

(Thermo Fisher Scientific, Waltham, MA) containing a hot-start polymerase, and 0.25 ng template DNA. We prepared nine 96-well PCR plates containing randomly shuffled samples as well as one positive and one negative control. The positive control was a mock community prepared with DNA extracted from pure cultures of eight fungal species unlikely to be found in plant leaves (Annex 1). A diagonal of eight wells was left empty, thus not using eight possible tag combinations per plate, which would allow estimating the frequency of tag switches. The PCR program included 10 min of initial denaturation and polymerase activation at 95 °C, followed by 30 cycles of 30 s denaturation at 95°C, 30 s annealing at 51.5 °C, 1 min elongation at 72 °C, and 7 min final elongation at 72 °C. PCR success was checked with gel electrophoresis.

The sequencing library was prepared and the sequencing run performed by the GeT-PlaGe core facilities of Genotoul (Toulouse, France). High-throughput parallel sequencing was done in a single, paired-end 2×250 bp run on an MiSeq machine (Illumina, San Diego, CA).

Sequence data processing

Reads obtained from the sequencing run were analysed using the *obitools* suite (Boyer et al. 2016). Forward and reverse reads were first quality-checked by calculating mean phred scores for the entire reads and for the first and last 10 nucleotides. Paired reads were merged and then matched to samples by their tag combinations, accepting two mismatches for primers and removing primer and tag sequences in the same step. In a subsequent quality filtering step, reads shorter than 150 bp, containing any ambiguities or with a paired-read alignment score <50 were discarded. Reads were dereplicated and any singletons removed. Reads were then clustered into Operational Taxonomic Units (OTUs) with *sumaclust* (Mercier et al. 2013), applying a threshold of 97% pairwise identity commonly used in fungal metabarcoding (Lindahl et al. 2013). To identify fungal reads and match them to known taxa, we used the curated *UNITE* collection (Kõljalg et al. 2013), in version 7.2 of the general release with 49,407 fungal ITS sequences, as reference database. The *ecotag* program of the *obitools* was used to assign taxonomic ranks and names to the most abundant sequence of each cluster, applying a minimum identity of 95%.

Grass barcoding and community structure

In some cases, the original identification of grass species in the field was doubtful, particularly in the genera *Oplismenus*, *Hyparrhenia* and *Digitaria*, and in some specimens collected without inflorescence. In *Hyparrhenia* in particular, we were unsure about the distinction between *H. rufa* and *H. schimperi*. Thus, we amplified a short, ~660 bp barcode segment of the *matK* chloroplast gene for 52 samples, using previously described primers and protocols (GPWG II 2012). The barcode sequences were checked via alignments against available sequences of the species in question (see chapter II), using single-nucleotide polymorphisms and in some cases indels to differentiate groups.

Grass community structure was visualised with non-metric multidimensional scaling (NMDS), using the *metaMDS* function of the R package *vegan* (Oksanen et al. 2017) with

Bray–Curtis distances based on presence–absence data. We also looked at phylogenetic community structure, based on a phylogeny extracted from the grass supertree of chapter II. Three species not included in that phylogeny were replaced by others: *Andropogon trichozygus* was represented by *A. eucomus* ssp. *huillensis* which was identical on the markers analysed (see chapter II); *Arundinaria itremoensis* was the only bamboo in our samples and represented by the closely related *A. humbertii*; *Hyparrhenia variabilis* was represented by *H. cymbaria* to which it was most similar based on *matK*. Using the R package *picante* (Kembel et al. 2010), we calculated phylogenetic diversity (PD, Faith 1992) and two measures of phylogenetic structure (clustering vs. overdispersion, Webb 2000), the net relatedness index (NRI) and the nearest taxon index (NTI), both inferred by comparisons against 999 randomisations of the community matrix where species richness per plot was maintained. We visualised community structure by habitat to see if grassland, forest, and tapia woodland could be distinguished.

As the DNA primers unintentionally amplified plant rather than fungal DNA (see Results), we made use of these data by checking the congruence of ITS2 genetic clusters with grass identifications by morphology and *matK* sequences. For this, we first extracted the most common read per sample. The extracted reads were then clustered with *sumaclust* as described above, but using a threshold of 98% identity. This matched the original species identification quite well while over-splitting species slightly. The most abundant read per cluster was checked via *BLAST* searches (Altschul et al. 1990) against *GenBank* as implemented in Geneious v. 9.0.5 (Biomatters, Auckland). The clusters were then used to repeat NMDS of community structure as described above.

Fungal diversity

The original goal of analysing fungal endophyte community structure was compromised by the preferential amplification of plant DNA (see Results). We thus focused our analysis on the sequenced mock community and the taxonomic identity of the few fungal OTUs recovered. First, we checked the abundance of the mock community OTUs over the nine PCR plates to see whether community structure and abundance were repeatable. Second, we quantified the occurrence of mock OTUs in other samples, which would indicate either tag switches or cross-contaminations. Finally, we excluded all controls and removed OTUs with a count of <25 reads per sample, based on the estimations of tag switches and mock OTUs occurring in other samples. Of the remaining OTUs, we submitted the most abundant read for those occurring in more than one sample to *BLAST* searches against *GenBank* to check the taxonomic assignment by *ecotag*. We did not further analyse fungal community structure because of the low number of OTUs; test NMDS runs did not converge, suggesting insufficient information in the data.

Finally, we extracted the most abundant reads for two OTUs identified as *Epichloë.* We aligned them, using the *E-ins-i* option of MAFFT (Katoh & Standley 2013), with complete ITS sequences from *GenBank*, including all identified *Epichloë* species available, the respective closest *BLAST* matches, and four *Balansia* species as outgroup. A Maximum

Likelihood phylogeny was inferred with *RAxML* (Stamatakis 2014), applying a GTR+G substitution model and performing 1000 bootstrap pseudoreplicates.

5.4 Results

Data quantity and quality

From 92 plots, we collected a total of 768 samples belonging to 58 grass species; 512 were from grassland, 104 from tapia woodland and 152 from forest. Sequencing yielded a total of 12,444,774 raw paired-end reads. Mean phred quality was >25 in the large majority of the samples, but reverse reads overall had lower quality, especially at the 3' ends (Fig. IV.2a). After tag matching, quality filtering and removal of singletons (87.5% of unique reads), 7,347,559 (59%) of the reads, corresponding to 302,150 unique reads, remained (Fig. IV.2b). Only 1.4% of these were identified as fungal OTUs through *ecotag* (Fig. IV.2b). Per sample, there was a mean of 9,441 reads (without controls), while positive controls (i.e. mock communities) had a mean of 11,751 reads, and the number of reads in negative controls and with unused tag combinations was always below 20 and 35, respectively (Fig. IV.2c).

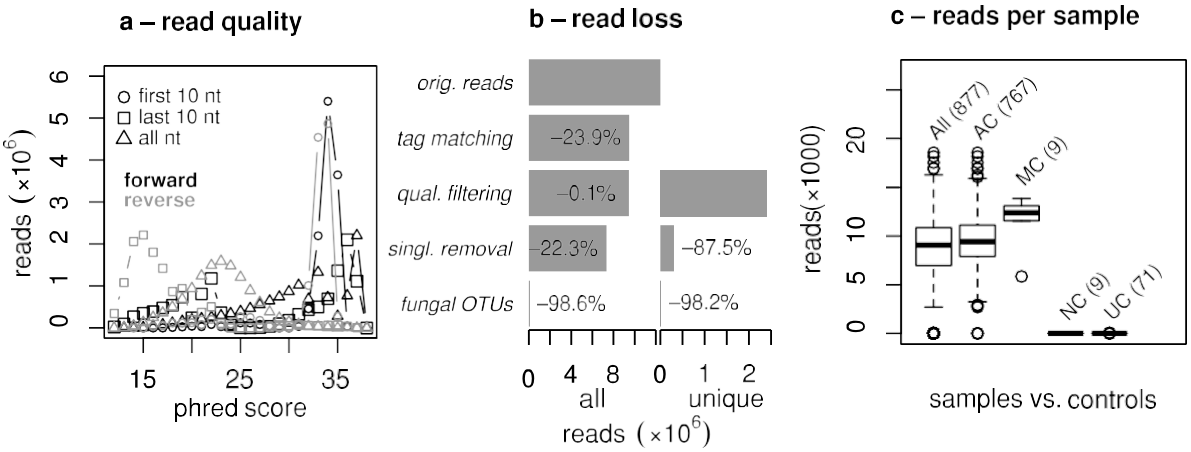

Figure IV.2. Quality and quantity of reads obtained through MiSeq sequencing. (a) refers to the original reads, (b) shows read loss through the filtering steps, and (c) shows the distribution of quality-filtered reads among samples. In (c), AC stands for all samples without controls, MC means mock community, NC negative controls, and UC unused tag combinations.

Grass identification and community structure

The 58 grass species sampled were phylogenetically diverse, but the majority belongs to subfamily Panicoideae (Fig. IV.3). The use of *matK* barcodes confirmed the distinction of three *Digitaria* species but suggested that *Hyparrhenia rufa* and *H. schimperi* were very close and could not be reliably distinguished; these were thus considered as one species, *H.* cf. *rufa*. Two C_4 species dominating grassland and tapia woodland plots, *Loudetia simplex* and *Schizachyrium sanguineum*, account for roughly one third of the samples. In forest plots, a clade of closely related C_3 species with many endemics accounts for most samples (Fig. IV.3). Plots from all three habitats types had a mean species richness of six, but grassland and tapia had a higher mean phylogenetic diversity (Fig. IV.4a–b). Forest grass communities had clearly positive values of NRI and NTI, indicating phylogenetic clustering (Fig. IV.4c–d). NRI values

indicate random assembly or slight phylogenetic overdispersion and NTI values random assembly to slight phylogenetic clustering for tapia woodland and grassland communities. NMDS clearly distinguished forest grass communities from grassland and tapia woodland based on original species identification coupled with *matK* barcodes (Fig. IV.5a).

Plant ITS2 sequences were the dominant reads in all but 11 samples. Clustering of ITS2 sequences overall confirmed species identifications and *matK* barcoding. However, 23 samples were assigned to clearly aberrant clusters, which were mostly from the same habitats. *Schizachyrium sanguineum* split in two ITS2 clusters of 45 samples each, while the marker

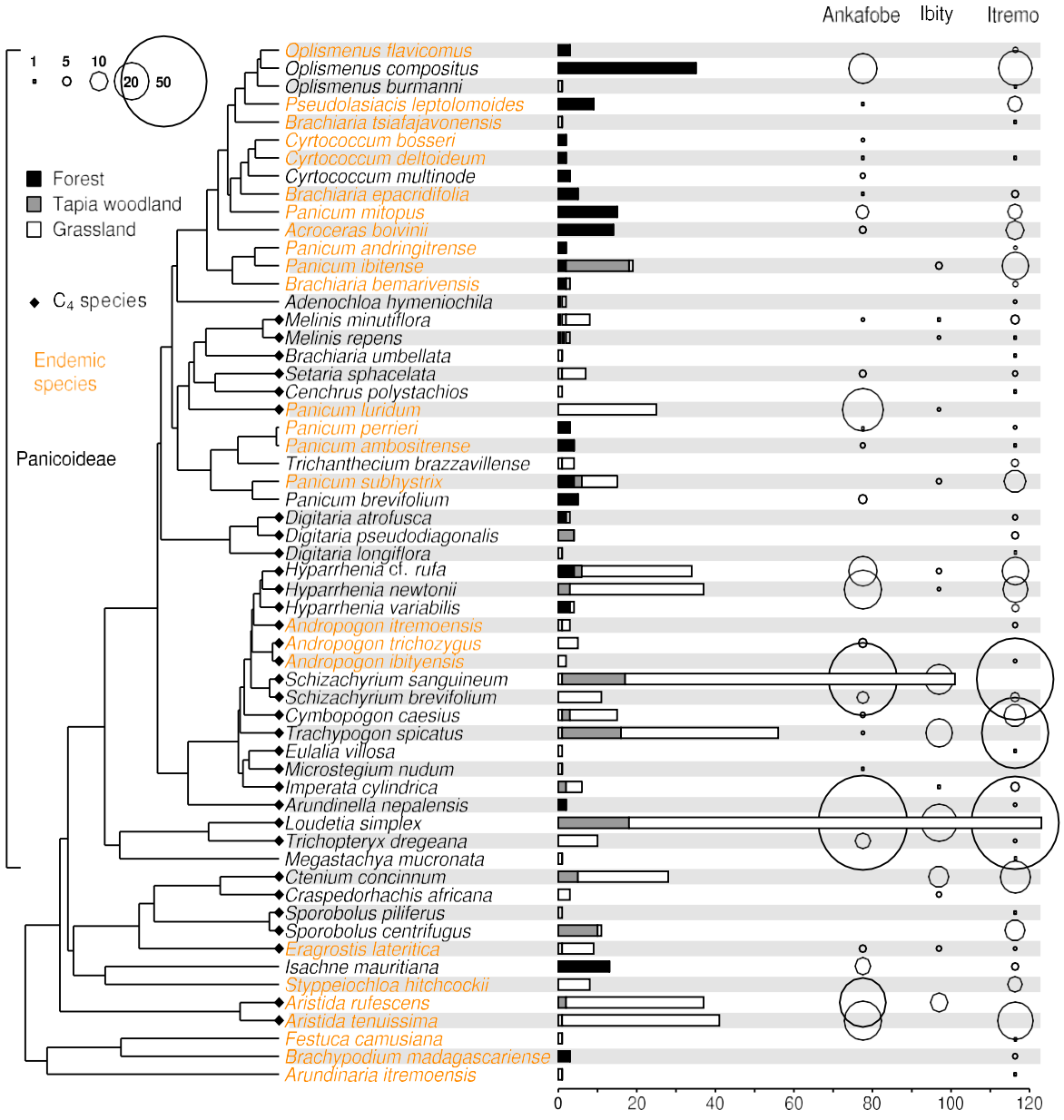

Figure IV.3. Distribution of the 768 samples over grass species and habitats (bars) and regions (circles). Plotted left to the species labels is the phylogeny inferred from chloroplast markers (see chapter II), with branch lengths relative to time.

lumped *Panicum ambositrense* and *P. perreri* together. *Oplismenus* split in three clusters which suggested that several samples of *O. burmanni* and *O. flavicomus* were misidentified as

O. *compositus.* These discrepancies did not affect the distinction of forest grass communities from tapia woodland and grassland through NMDS (Fig. IV.5b).

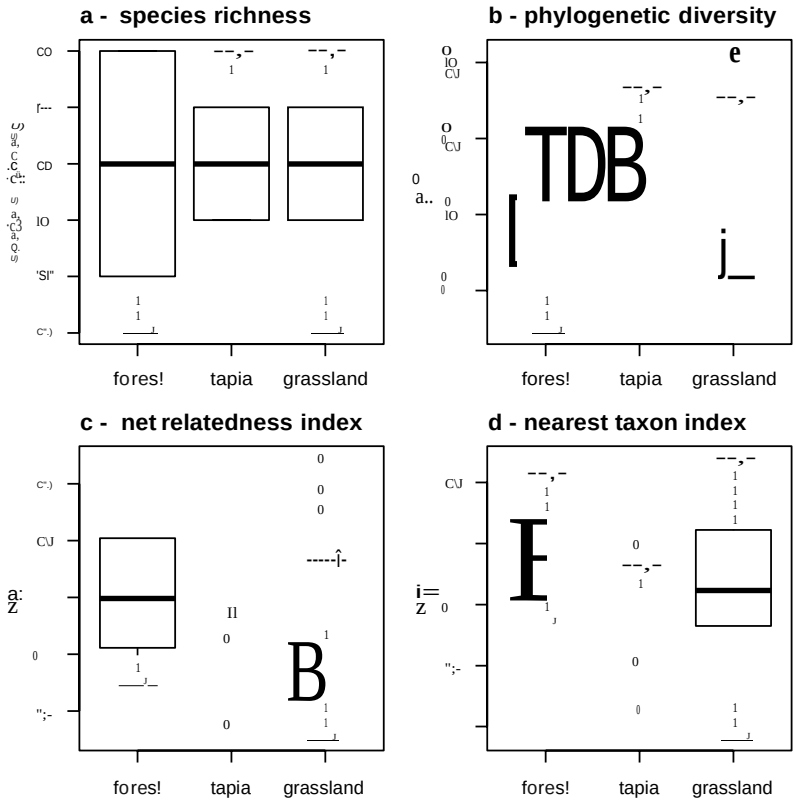

Figure IV.4. Grass species richness and phylogenetic diversity per plot, compared between habitats.

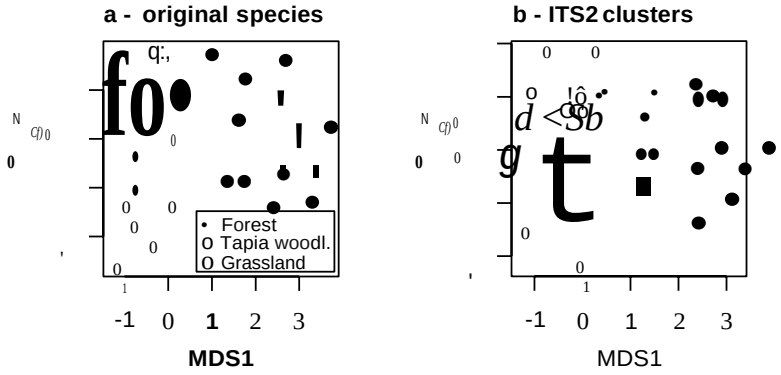

Figure IV.5. Differentiation of grass communities by habitat, visualised through non-metric multidimensional scaling (NMDS). NMDS was performed based on both original identification supported through *matK* barcodes as well as 98% identity ITS2 clusters, based on Bray–Curtis distances of presence–absence data.

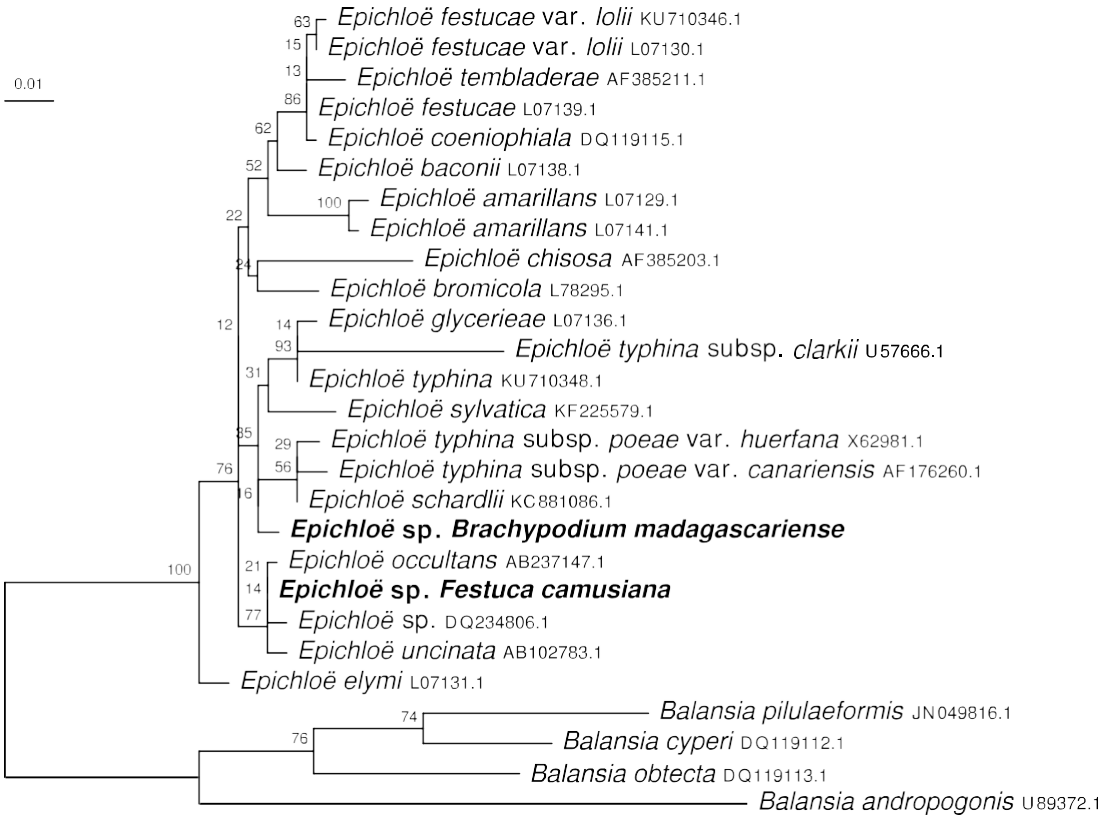

Figure IV.7. Maximum Likelihood ITS phylogeny of *Epichloë*, including ITS2 sequences of two OTUs recovered in our dataset (in bold). Values at nodes are bootstrap support values. The scale is in substitutions per site.

Endophyte diversity

The 104,563 quality-filtered reads identified as fungi clustered in 398 OTUs. In the mock communities, six out of eight species were always recovered, while two, *Paxillus adelphus* (Basidiomycota) and *Glonium stellatum* (Ascomycota), were not found and apparently failed to amplify. Read numbers and abundance ranks of mock OTUs were repeatable, with an outlying lower yield only in plate 4 (Fig. IV.6a). Yield overall increased with the amount of mock OTU template DNA but revealed strong taxon-specific variability (Fig. IV.6b). Among the six amplified mock OTUs, no general bias towards ascomycetes or basidiomycetes was observed. Reads of the most abundant mock OTUs were also found in other samples, but with a maximum number of 23. Excluding mock OTUs and all occurrences with less than 25 reads, only 38 OTUs were left of which 22 were only found in one sample (Table A-IV.1 in annex A-IV). Most were only assigned to the fungal kingdom by *ecotag* but several could be assigned to genera or higher-level ranks by *BLAST* searches against *GenBank*. The most frequent OTU, found in 12 samples, had 100% identity with several *Phoma* and *Epicoccum* species, including the sorghum grain mold fungus *Epicoccum sorghinum*. It was associated with various genera, including *Aristida*, *Hyparrhenia*, and *Schizachyrium*.

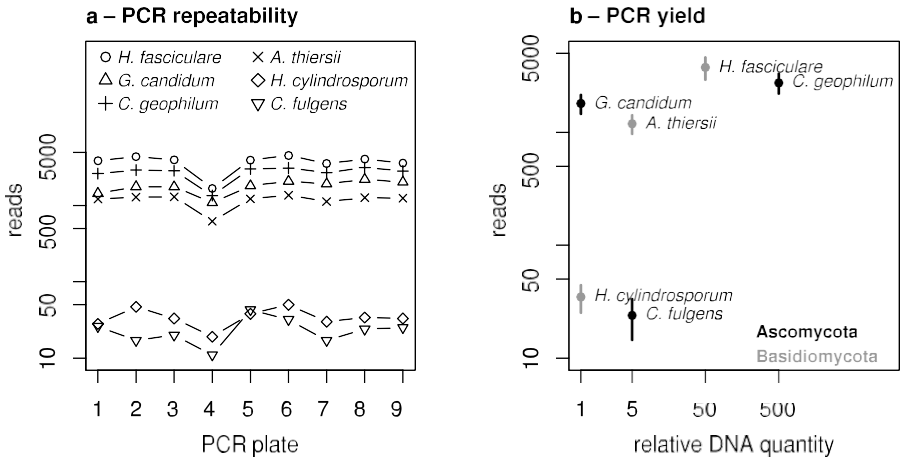

Figure IV.6. Read number for fungal mock community OTUs. Note that the y axis in (a) and both axes in (b) have a logarithmic scale. Vertical lines in (b) are standard deviations.

Two OTUs were identified as Clavicipitaceae by *ecotag* and could be assigned to *Epichloë* through BLAST searches. One was found in the single sample of *Festuca camusiana*, collected in the shade in a tapia woodland plot, the other in the three samples of *Brachypodium madagascariense*, all from forest plots. This means that all four Pooideae samples contained *Epichloë*. An ITS phylogeny of the genus was overall not well resolved but placed the *F.-camusiana*-associated OTU with two other Loliinae-associated species (*E. occultans* and *E. uncinata*) while the OTU associated with *B. madagascariense* was distinct but without supported close relatives (Fig. IV.7). One OTU from the *F. camusiana* sample classified as fungal by *ecotag* was identified as *Festuca* through BLAST searches (Table A-IV.1).

5.5 Discussion

The analysis we present here is preliminary and was hampered by the unintended amplification of plant DNA. Nevertheless, some conclusions can be drawn that will help direct future re-analysis of the dataset. First, we found that grassland and forest had not only distinct grass species communities but also differed in phylogenetic structure. Forest communities are phylogenetically clustered, which can be explained by the dominance of one particular panicoid clade, the "Madagascar shade clade" (see chapter II). Grass communities in tapia woodland are more similar to grassland but also feature some particular species (Fig. 3). This is in line with results from a larger sample of grass communities in central Madagascar (C. Solofondranohatra et al. submitted). In any case, this means that host grass phylogeny should be included as variable in a future analysis of endophyte community structure, given its interaction with habitat.

Primers for amplifying the ITS2 region were chosen because there is thought to be a potential sequencing bias in ITS1 against ascomycetes due to an upstream SSU intron (Lindahl et al. 2013), and ascomycetes are the dominant group among endophytes (Rodriguez et al. 2009). The primers ITS3 and ITS4 have been employed since more than 25 years in fungal ITS sequencing (White et al. 1990) and were also used in metabarcoding studies of soil

and root-associated fungi (e.g. Blaalid et al. 2013, Tedersoo et al. 2015). A comparison of different primer sets found ITS3/ITS4 to have fungal coverage comparable to other ITS2 primers and at least one mismatch in each primer to embryophytes (Toju et al. 2012). It is possible that for analysis of the leaf microbiome where plant DNA is dominant, stronger discrimination is necessary. Also, preferential amplification specific to Poaceae cannot be excluded. Metabarcoding studies of leaf endophytes that did use the ITS1 region recovered a majority of ascomycetes (e.g. Cordier et al. 2012, Bálint et al. 2015), suggesting the potential bias is weak. Sequencing the whole ITS region (500–600 bp), the accepted fungal barcode (Schoch et al. 2012), would be desirable but is not feasible with current Illumina technologies which have a maximum sequencing length of 300 bp per direction. A future re-analysis of this dataset should explore the use of the ITS1 region but specifically test the primers against grasses first using *in silico* PCR (Ficetola et al. 2010).

The preferential recovery of plant ITS2 sequences, while unintended, allowed to highlight some potential cases of confusion or misidentification of grass species. These are possible since especially the open grass communities are very dense and individuals can be easily confused. Exclusion of samples with conflicting identification would probably be most conservative with regard to downstream analyses and would affect less than 5% of the samples. Close species that were sometimes confused, such as *Oplismenus* species, or *Panicum ambositrense* and *P. perrieri*, should probably be lumped together. This would increase sample size for each species and mask processes such as hybridisation that may take place between close species but are not of interest at the scale of this study. Unrelated to the objective of endophyte community analysis, the plant ITS2 sequences generated may also allow to look at the genetic structure of grass populations from three distinct regions, which is relevant to the history of grasslands in Madagascar (see chapter II). Considering possible intragenomic ITS variability and polyploidy, a strategy to extract variants per individual and distinguish them from errors would have to be found.

Overall, the ITS2 analysis suggests that metabarcoding is a suitable tool to study endophyte communities even in remote locations. In the absence of better options for preservation and cooling in the field, storage in CTAB buffer appears to yield sufficient quantity and quality of DNA. U'Ren et al. (2014) also suggested that CTAB preservation is preferable over silica drying when direct freezing is not feasible. Our analysis also suggests that while tag switches do occur (Carlsen et al. 2012), they do so at very low levels relative to the total number of reads. They can thus be controlled for by removing reads below an estimated abundance threshold. In our study, this removed most of the presumed fungal OTUs, some of which may be true occurrences. However, this is rather a problem of very low coverage relative to reads of plant origin, which should be resolved with more appropriate primers (see above). The use of a mock community demonstrated that PCR, while showing bias among taxa, yields repeatable community composition as long as there is sufficient sequencing depth (Smith & Peay 2014). The assignment of OTUs to fungal taxa could be improved; here, our first goal was to recover those OTUs that are fungal so we chose a rather conservative limit of 95% identity. This means that the most common ancestor of all reference

sequences similar to an OTU, including many unidentified environmental samples, could not be resolved further than to the kingdom Fungi by *ecotag*. The use of all fungal sequences available on *GenBank*, whose identification is less certain but which provide a higher number of references, should be explored. The assignment of a *Festuca* sequence as fungal by *ecotag* moreover highlights potential errors, even in expert-curated databases such as *UNITE* (Kõljalg et al. 2013).

The few fungal OTUs recovered support the idea that many endophytes are latent pathogens (Wilson 1995): an OTU identified as *Epicoccum* and close to the common pathogen *E. sorghinum* was the most widespread. That species was recently recovered through metabarcoding from sorghum seeds in Burkina Faso (Stokholm et al. 2016), suggesting vertical transmission and making an endophytic life stage plausible. However, the complete identity of the OTU's ITS2 sequence with several *Epicoccum* accessions on *GenBank* suggests that ITS may not be sufficient to distinguish species in this group, a conclusion reached by Gazis et al. (2011) for other groups of endophytes. A significant result of our study is the confirmation of *Epichloë* endophytes in all four Pooideae samples analysed. This is, to our knowledge, the first record of *Epichloë* in Madagascar. The ITS phylogeny (Fig. IV.7) tentatively suggests that the two OTUs are distinct species and do not have a common origin. *Epichloë* species have been found in various temperate regions of the world, always associated with Pooideae, and have a demonstrated co-divergence history of some subgroups with the grass subfamily (Schardl et al. 1997, Clay & Schardl 2002). It is likely that they co-dispersed to Madagascar with the ancestors of the endemic *B. madagascariense* and *F. camusiana*, potentially with seeds given the possibility of vertical transmission.

The original objective of our study was an analysis of fungal endophyte community structure across several habitats, regions, and grass species. The goal was notably to distinguish habitat and host filters from neutral factors such as dispersal limitation by distance. Our dataset includes a large number of samples and covers a wide range of potential ecological factors, with 58 grass species from contrasting habitats, a spatially explicit sampling, and a full phylogeny of the host grasses. Re-analysis with more appropriate primers would give valuable insights into endophyte community structure in a region completely unknown in terms of fungal endophytes.

5.6 References

Altschul S.F., Gish W., Miller W., Myers E.W. & Lipman D.J. (1990): Basic local alignment search tool. *Journal of Molecular Biology*, 215: 403–410. Arnold A.E., Maynard Z., Gilbert G.S., Coley P.D. & Kursar T.A. (2000): Are tropical fungal endophytes hyperdiverse? *Ecology Letters*, 3: 267–274.

Arnold A.E., Mejía L.C., Kyllo D., Rojas E.I., Maynard Z. et al. (2003): Fungal endophytes limit pathogen damage in a tropical tree. *Proceedings of the National Academy of Sciences*, 100: 15649–15654.

Bálint M., Bartha L., O'Hara R.B., Olson M.S., Otte J. et al. (2015): Relocation, high-latitude warming and host genetic identity shape the foliar fungal microbiome of poplars. *Molecular Ecology*, 24: 235–248.

Blaalid R., Kumar S., Nilsson R.H., Abarenkov K., Kirk P.M. & Kauserud H. (2013): ITS1 versus

ITS2 as DNA metabarcodes for fungi. *Molecular Ecology Resources*, 13: 218–224.

Boyer F., Mercier C., Bonin A., Le Bras Y., Taberlet P. & Coissac E. (2016): obitools: a unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources*, 16: 176–182.

Carlsen T., Aas A.B., Lindner D., Vrålstad T., Schumacher T. & Kauserud H. (2012): Don't make a mista(g)ke: is tag switching an overlooked source of error in amplicon pyrosequencing studies? *Fungal Ecology*, 5: 747–749.

Clay K., Hardy T.N. & Hammond A.M. (1985): Fungal endophytes of grasses and their effects on an insect herbivore. *Oecologia*, 66: 1–5.

Clay K. & Holah J. (1999): Fungal endophyte symbiosis and plant diversity in successional fields. *Science*, 285: 1742–1744.

Clay K., Marks S. & Cheplick G.P. (1993): Effects of insect herbivory and fungal endophyte infection on competitive interactions among grasses. *Ecology*, 74: 1767–1777.

Clay K. & Schardl C. (2002): Evolutionary origins and ecological consequences of endophyte symbiosis with grasses. *The American Naturalist*, 160: S99–S127.

Cordier T., Robin C., Capdevielle X., Fabreguettes O., Desprez-Loustau M.-L. & Vacher C. (2012): The composition of phyllosphere fungal assemblages of European beech (*Fagus sylvatica*) varies significantly along an elevation gradient. *New Phytologist*, 196: 510–519.

Costa I.P.M.W., Maia L.C. & Cavalcanti M.A. (2012): Diversity of leaf endophytic fungi in mangrove plants of Northeast Brazil. *Brazilian Journal of Microbiology*, 43: 1165–1173.

Faith D.P. (1992): Conservation evaluation and phylogenetic diversity. *Biological Conservation*, 61: 1–10.

Ficetola G.F., Coissac E., Zundel S., Riaz T., Shehzad W. et al. (2010): An *In silico* approach for the evaluation of DNA barcodes. *BMC Genomics*, 11: 434.

Gazis R., Rehner S. & Chaverri P. (2011): Species delimitation in fungal endophyte diversity studies and its implications in ecological and biogeographic inferences. *Molecular Ecology*, 20: 3001–3013.

GPWG II (2012): New grass phylogeny resolves deep evolutionary relationships and discovers C_4 origins. *New Phytologist*, 193: 304–312.

Guo B., Wang Y., Sun X. & Tang K. (2008): Bioactive natural products from endophytes: a review. *Applied Biochemistry and Microbiology*, 44: 136–142.

Higgins K.L., Arnold A.E., Coley P.D. & Kursar T.A. (2014): Communities of fungal endophytes in tropical forest grasses: highly diverse host- and habitat generalists characterized by strong spatial structure. *Fungal Ecology*, 8: 1–11.

Higgins K.L., Coley P.D., Kursar T.A. & Arnold A.E. (2011): Culturing and direct PCR suggest prevalent host generalism among diverse fungal endophytes of tropical forest grasses. *Mycologia*, 103: 247–260.

Katoh K. & Standley D.M. (2013): MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, 30: 772–780.

Kembel S.W., Cowan P.D., Helmus M.R., Cornwell W.K., Morlon H. et al. (2010): Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, 26: 1463–1464.

Kõljalg U., Nilsson R.H., Abarenkov K., Tedersoo L., Taylor A.F.S. et al. (2013): Towards a unified paradigm for sequence-based identification of fungi. *Molecular Ecology*, 22: 5271–5277.

Lehmann C.E.R. & Parr C.L. (2016): Tropical grassy biomes: linking ecology, human use and conservation. *Philosophical Transaction of the Royal Society B: Biological Sciences*, 371: 20160329.

Lindahl B.D., Nilsson R.H., Tedersoo L., Abarenkov K., Carlsen T. et al. (2013): Fungal community analysis by high-throughput sequencing of amplified markers – a user's guide. *New Phytologist*, 199: 288–299.

Malinowski D.P. & Belesky D.P. (2006): Ecological importance of *Neotyphodium* spp. grass endophytes in agroecosystems. *Grassland Science*, 52: 1–14.

Martiny J.B.H., Bohannan B.J.M., Brown J.H., Colwell R.K., Fuhrman J.A. et al. (2006): Microbial biogeography: putting microorganisms on the map. *Nature Reviews Microbiology*, 4: 102–112.

Mercier C., Boyer F., Bonin A. & Coissac E. (2013): SUMATRA and SUMACLUST: fast and exact comparison and clustering of sequences. *Programs and abstracts of the SeqBio 2013 workshop*, 27–29.

Moat J. & Smith P. (2007): *Atlas of the vegetation of Madagascar*. Royal Botanic Gardens, Richmond, Surrey, UK.

Oksanen J., Blanchet F.G., Friendly M., Kindt R., Legendre P. et al. (2017): *vegan: community ecology package version 2.4-3.*

Omacini M., Chaneton E.J., Ghersa C.M. & Müller C.B. (2001): Symbiotic fungal endophytes control insect host–parasite interaction webs. *Nature*, 409: 78–81.

Parr C.L., Lehmann C.E.R., Bond W.J., Hoffmann W.A. & Andersen A.N. (2014): Tropical grassy biomes: misunderstood, neglected, and under threat. *Trends in Ecology & Evolution*, 29: 205–213.

Ratnam J., Bond W.J., Fensham R.J., Hoffmann W.A., Archibald S. et al. (2011): When is a "forest" a

savanna, and why does it matter? *Global Ecology and Biogeography*, 20: 653–660.

Redman R.S., Sheehan K.B., Stout R.G., Rodriguez R.J. & Henson J.M. (2002): Thermotolerance generated by plant/fungal symbiosis. *Science*, 298: 1581–1581.

Reinhold-Hurek B. & Hurek T. (2011): Living inside plants: bacterial endophytes. *Current Opinion in Plant Biology*, 14: 435–443.

Rodriguez R.J., Henson J., Van Volkenburgh E., Hoy M., Wright L. et al. (2008): Stress tolerance in plants via habitat-adapted symbiosis. *The ISME Journal*, 2: 404–416.

Rodriguez R.J., White Jr J.F., Arnold A.E. & Redman R.S. (2009): Fungal endophytes: diversity and functional roles. *New Phytologist*, 182: 314–330.

Saikkonen K., Faeth S.H., Helander M. & Sullivan T.J. (1998): Fungal endophytes: a continuum of interactions with host plants. *Annual Review of Ecology and Systematics*, 29: 319–343.

Saikkonen K., Lehtonen P., Helander M., Koricheva J. & Faeth S.H. (2006): Model systems in ecology: dissecting the endophyte–grass literature. *Trends in Plant Science*, 11: 428–433.

Saikkonen K., Wäli P., Helander M. & Faeth S.H. (2004): Evolution of endophyte–plant symbioses. *Trends in Plant Science*, 9: 275–280.

Sánchez Márquez S., Bills G.F., Herrero N. & Zabalgogeazcoa Í. (2012): Non-systemic fungal endophytes of grasses. *Fungal Ecology*, 5: 289–297.

Schardl C.L., Leuchtmann A., Chung K.-R., Penny D. & Siegel M.R. (1997): Coevolution by common descent of fungal symbionts (*Epichloë* spp.) and grass hosts. *Molecular Biology and Evolution*, 14: 133–143.

Schoch C.L., Seifert K.A., Huhndorf S., Robert V., Spouge J.L. et al. (2012): Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences*, 109: 6241–6246.

Smith D.P. & Peay K.G. (2014): Sequence depth, not PCR replication, improves ecological inference from next generation DNA sequencing. *PLoS ONE*, 9: e90234.

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

Stokholm M.S., Wulff E.G., Zida E.P., Thio I.G., Néya J.B. et al. (2016): DNA barcoding and isolation of vertically transmitted ascomycetes in sorghum from Burkina Faso: *Epicoccum sorghinum* is dominant in seedlings and appears as a common root pathogen. *Microbiological Research*, 191: 38–50.

Strobel G.A. (2003): Endophytes as sources of bioactive products. *Microbes and Infection*, 5: 535–544.

Tedersoo L., Anslan S., Bahram M., Põlme S., Riit T. et al. (2015): Shotgun metagenomes and multiple primer pair–barcode combinations of amplicons reveal biases in metabarcoding analyses of fungi. *MycoKeys*, 10: 1.

Tedersoo L., Bahram M., Põlme S., Kõljalg U., Yorou N.S. et al. (2014): Global diversity and geography of soil fungi. *Science*, 346: 1256688.

Toju H., Tanabe A.S., Yamamoto S. & Sato H. (2012): High-coverage ITS primers for the DNA-based identification of ascomycetes and basidiomycetes in environmental samples. *PLoS ONE*, 7: e40863.

U'Ren J.M., Riddle J.M., Monacell J.T., Carbone I., Miadlikowska J. & Arnold A.E. (2014): Tissue storage and primer selection influence pyrosequencing-based inferences of diversity and community composition of endolichenic and endophytic fungi. *Molecular Ecology Resources*, 14: 1032–1048.

Vorontsova M.S., Besnard G., Forest F., Malakasi P., Moat J. et al. (2016): Madagascar's grasses and grasslands: anthropogenic or natural? *Proceedings of the Royal Society B*, 283: 20152262.

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

White J.F. & Reddy P.V. (1998): Examination of structure and molecular phylogenetic relationships of some graminicolous symbionts in genera *Epichloë* and *Parepichloë*. *Mycologia*, 90: 226–234.

White T.J., Bruns T.D., Lee S.B. & Taylor J.W. (1990): Analysis of phylogenetic relationships by amplification and direct sequencing of ribosomal RNA genes. *PCR protocols: A guide to methods and applications* (ed. by M.A. Innis, D.H. Gelfand, J.N. Sninsky & T.J. White). Academic Press, London.

Wilson D. (1995): Endophyte: the evolution of a term, and clarification of its use and definition. *Oikos*, 73: 274–276.

Zilber-Rosenberg I. & Rosenberg E. (2008): Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiology Reviews*, 32: 723–735.

Zimmerman N.B. & Vitousek P.M. (2012): Fungal endophyte communities reflect environmental structuring across a Hawaiian landscape. *Proceedings of the National Academy of Sciences*, 109: 13022–13027.

6 Conclusion and perspectives

6.1 Lineage dispersal in grasses

In this dissertation, I studied the dispersal history of a globally distributed plant family, the grasses. Chapter I explicitly looked at factors underlying dispersal in one temperate grass lineage. Chapter II analyses the origins and timing of arrival of the various lineages that constitute the extant grass flora of Madagascar. Chapter III placed a small C_3 lineage in the grass phylogeny, confirming a position that suggests its sister lineage, of the largest known C_4 clades, originated in and dispersed out of India. Chapter IV, despite technical limitations, demonstrated the presence of vertically transmitted endophytes in Madagascar, suggesting potential co-dispersal with their host grass lineages.

Lineage dispersal has only become accessible to statistical inference with the arrival of molecular phylogenetics and especially divergence time estimation (see Introduction). Rather than rare and elusive, it appears to be a central process in the biogeography of at least plants (Renner 2004, Sanmartín & Ronquist 2004). Higher plants, with their embryos well protected in seed coats, have an excellent vehicle allowing transport even over large distances and under rough conditions. Large seeds with hard shells such as the sea-travelling coconuts are the best examples for global dispersal, but it appears that long-distance dispersal is common even in rather small-seeded groups such as grasses (chapters I and II, Blattner 2006, Inda et al. 2008, Linder et al. 2013). Grass seeds have a well developed, starchy endosperm that provides a nutrient reservoir, while the double wall structure of the caryopsis, where the fruit wall is fused with the seed coat, additionally protects the seed (Kellogg 2015). Dispersal units sometimes not only include the caryopsis, but whole spikelets or clusters of spikelets, which may have hooks or barbs that facilitate dispersal (Kellogg 2015). Many bird species feed on grass seeds (e.g. Columbidae, Frith et al. 1976) and represent a likely dispersal vector. Testing whether reconstructed dispersal routes are statistically associated with dominant migration routes of birds (Gillespie et al. 2012) would be worthwhile; range evolution analysis combined with stochastic mapping as employed in chapter I could be used for this. Large phylogenies such as now increasingly available for the grasses (Spriggs et al. 2014, for example, compiled a phylogeny for 3,595 species, i.e. 30% of the family) would provide the necessary sample size and statistical power.

A factor naturally expected to affect dispersal success is distance. Studies at the margin between ecology and biogeography have demonstrated “dispersal kernels” where dispersal probability drops off with distance (Nathan et al. 2003). Here, I inferred through statistical model comparison that distance is an important factor in one specific plant lineage, the Loliinae (chapter I), and in several lineages that contributed to the assembly of Madagascar’s grass flora (chapter II). Surprisingly few studies have explicitly tested the role of distance in range evolution (e.g. Linder et al. 2013, Van Dam & Matzke 2016). Models such as DEC+*x* including this factor would certainly provide good fit to many phylogenies and probably

improve ancestral area inference. It will be interesting to see, too, how the strength of dispersal limitation as measured by *x* differs between lineages, and how it is related to other factors such as dispersal vectors, barriers etc. However, it is not quite clear whether *x* can be directly compared when inferred on separate phylogenies where sample size, area definition, and distance measures differ; this will have to be tested using simulations. The parameter *x* alone also does not allow to estimate the shape of the dispersal kernel, which requires stochastic mapping. Unlike a study on danthonioid grasses (Linder et al. 2013), I could not confirm a "fat-tailed" dispersal kernel for the Loliinae (chapter I) where dispersal probability would essentially be independent from distance at very large distances (Nathan et al. 2003). This may be partly due to methodological choices, such whether "zero" dispersals between areas are counted as I did for the Loliinae example. Also, the shape of the dispersal kernel may not be the same for ecological dispersal and the "effective dispersal" (including establishment) usually inferred in biogeography. Again, the increasing number and size of phylogenies will allow to properly test hypotheses for different taxonomic groups.

Beyond factors that act on the dispersal process itself, successful establishment will determine dispersal rates and probabilities inferred from phylogenies. Plants are advantaged over most animals in that often a single individual may suffice to found a new population, as observed for example in an orchid (Willems 1982). Asexual reproduction, through vegetative growth or apomixis, is common in plants, including grasses (Brown & Emery 1958, Calzada et al. 1996, Kellogg 2015). Moreover, the high frequency of polyploidisation, including between distinct species (Kellogg 2015), may improve local persistence and facilitate the recruitment of genes adaptive in a new environment. The invasive grass *Spartina anglica*, an anthropogenic polyploid hybrid, is a good example for this (Gray et al. 1991). In the grass subfamily Danthonioideae, long-distance dispersal events were found to be more likely in polyploid lineages (Linder & Barker 2014) and several lineages resulted from new combinations of chloroplast and nuclear lineages through hybridisation (Pirie et al. 2009). Other mechanisms of horizontal transfer probably underlie the recruitment of C_4 genes from co-occurring but phylogenetically distant lineages observed in the model grass *Alloteropsis semialata* (Christin et al. 2012). Increasingly available data from various genomes (nuclear, mitochondrial, and chloroplastic) will probably reveal that such events of adaptation through recombination or horizontal transfers are more frequent than currently known (Soucy et al. 2015). Other traits conserved phylogenetically and already present at the time of dispersal may facilitate establishment. In chapter I, I inferred only weak support for dispersal rates differing between two clades with contrasting phenotypic syndromes (roughly, competitive vs. vagile). It may be useful to decompose such syndromes into individual traits and test if there are any key traits that affect dispersal and establishment. Rhizome strength, lateral spread, overall growth rate, seed production, and ploidy level (all southern hemisphere Loliinae appear to be polyploid, Catalán 2006), are candidate traits. Another factor affecting local establishment is that of microbial symbionts: frequently, plants will have to adapt to locally occurring microbial communities, but conservatism and co-dispersal of adapted symbionts is another possibility (see last section, below).

6.2 The assembly of tropical grasslands

The importance of tropical grasslands has been increasingly acknowledged over the last years (Parr et al. 2014, Bond 2016, Lehmann & Parr 2016). That grasslands expanded globally since the Miocene is supported by several lines of evidence (Edwards et al. 2010, Strömberg 2011). However, the status of many grasslands as natural or anthropogenic is debated (see e.g. Bond et al. 2008 for Madagascar, and Ratnam et al. 2016 for Asian grasslands). Phylogenetic analyses of grass diversity (e.g. Vorontsova et al. 2016) represent one way to test whether grasslands are natural and to estimate their age. Here, I analysed grass diversification in Madagascar and found endemic C_4 clades diverging in the Miocene to Pleistocene, which supports the existence of C_4 habitats before human colonisation (chapter II). I also recovered a small lineage of Indian C_3 grasses that are found in grasslands and phylogenetically sister to one of the major tropical C_4 clades, supporting the ancient nature of Indian grasslands (chapter III). Such phylogenetic analyses are important to complement studies looking at the macroecological determinants of grasslands (e.g. Sankaran et al. 2005, Lehmann et al. 2011, Ratnam et al. 2011): they reveal patterns that may be specific and contingent on the regional context. For example, while both Madagascar and India were presumably affected by the Miocene grassland expansion, Madagascar saw repeated immigration of C_4 grasses, mainly from Africa (chapter I), while India was potentially the cradle of a major *in situ* C_4 clade diversification (chapter III). With more data available, future studies could compare the phylogenetic structure of grasslands in various regions of the world, identify bursts of diversification and evaluate the importance of local lineage recruitment versus immigration over larger distances (Donoghue 2008). The next step would be to couple phylogenetics with trait evolution and estimate how key traits related to fast growth, resprouting after fire, herbivore resistance, and drought resilience, accumulated over time, how trait diversification relates to niche space (Walker & Valentine 1984, Freckleton & Harvey 2006), and whether trait composition and phylogenetic diversity of grasses differ between types of grassland (e.g. herbivore- vs. fire-controlled, or mesic versus xeric) or between different regional contexts.

Phylogenetic analyses can only be large in scale and are hampered by the uncertainties of divergence time calibration (see general introduction and discussion in chapter II). Even if they may support the existence of C_4 grassy biomes before widespread human intervention, they provide no clues on how important and expansive these ancient grasslands were. Stable isotope analyses looking at the stability of forest–savannah boundaries through time (e.g. Desjardins et al. 1996, Wiedemeier et al. 2012) as well as additional stratigraphic records of pollen, charcoal, and dung fungi (e.g. Burney 1987, Gasse & Van Campo 1998, Burney et al. 2003) could increase spatial and temporal resolution of grassland dynamics. Population genetic analyses may provide further evidence through the inference of past population contractions and expansions. Grasses, as the keystone species of grassy biomes, would be a natural subject for population genetics (such as done by McAllister & Miller 2016 for a North American grassland species); however, their often large populations mean that signatures of geographical structure may be difficult to detect. More sparsely distributed savannah plants

with presumably short (ecological) dispersal distance, such as the "underground trees" (geoxylic suffrutices, Maurin et al. 2014) of Africa, or tapia (*Uapaca bojeri*) in Madagascar may be better suited to such analyses. Also, precisely those species not adapted to open habitats may give clues on the past extent of grasslands: in Madagascar, there is high local endemism in forest-dwelling species such as lemurs, and allopatric speciation through past fragmentation has been proposed as a mode of diversification on the island (Wilmé et al. 2006, Vences et al. 2009). The striking *in situ* diversification of forest grasses and bamboos (chapter II) may represent an equivalent pattern in plants. If forests were indeed fragmented and separated by the expansion of grassland, this could have restricted gene flow and favoured speciation. Thus, population genetics of open-habitat species would ideally be complemented with analyses of forest specialists in the same region.

6.3 Fungal endophytes of grasses

Looking beyond the grass phylogeny, the original objective of the last chapter (IV) was to study the community structure of symbionts associated with grasses. Whereas the first three chapters used phylogenies inferred from DNA to look back in time, here molecular methods were used to study diversity in space. The unintended amplification of plant rather than fungal DNA underlined the methodological challenges of such barcoding studies (Taberlet et al. 2012). Nevertheless, some results suggest perspectives for future research.

First, grass communities from forest and open habitats (grassland and tapia woodland) had contrasting phylogenetic structure. This suggests that host phylogeny should be integrated in endophyte community analysis since it may, at least partly, explain potential differences between habitats. It may also allow to detect any host specificity that operates at a higher level than that of species. I expect however that forest and open habitats, independently from host phylogeny, harbour different fungal communities, as they represent environments with very different conditions for growth and dispersal. First, plant leaves in open grasslands are subject to high irradiation, a factor which is thought to require specific adaptations of the leaf microbiome and thus represent a strong selective force (Vorholt 2012). Second, higher airflow in open habitats may facilitate the contagious spread of fungal spores and homogenise communities over larger distances; this would create weaker spatial structure in grassland than in forest communities. Furthermore, the frequent fires which characterise tropical grasslands regularly consume most of the grass biomass (Bond & Keeley 2005). I thus expect that latent saprotrophs have less abundant resources and may be under-represented among grassland endophytes compared to forest grasses. The most frequent OTU recovered in chapter IV, an *Epicoccum* species, is likely a latent pathogen, but this is only anecdotal evidence for any dominant lifestyle. Finally, grasslands in Madagascar – and elsewhere – have a grazing megafauna; these are mainly zebu cattle in modern Malagasy grasslands (including the ones sampled in chapter IV), but there was presumably a much more diverse grazing fauna less than a few thousand years ago (Burney 2003). Selective pressure through grazers may favour mutualistic fungi protecting their host through toxins, as do *Epichloë* in temperate Pooideae

grasses. Less studied relatives of *Epichloë* found in tropical grasses, such as *Balansia* (Reddy et al. 1998) and *Parepichloë* (White & Reddy 1998), may have similar functions, given the widespread occurrence of diverse alkaloids in the family Clavicipitaceae (Schardl et al. 2013).

The discovery of Pooideae-specific *Epichloë* endophytes in Madagascar demonstrates the conservatism of this symbiosis even at tropical latitudes. Co-divergence of mutualistic *Epichloë* with their host plant lineages has been demonstrated (Clay & Schardl 2002), but no study has yet investigated the combined biogeography of symbionts and host plants to my knowledge. The low probability of randomly dispersed spores to find a host plant in the tropics, where Pooideae only occur in "mountain islands", and the frequent vertical transmission with seeds suggest that both dispersed together. If this is the case, phylogenies of *Epichloë* should show a north-to-south dispersal pattern as in the Loliinae (chapter I) and the Pooideae in general (Bouchenak-Khelladi et al. 2010). The frequent hybrid origin of mutualistic *Epichloë* species (Clay & Schardl 2002) complicates phylogenetic analyses, but it also suggests some interesting hypotheses on the relationship between hybridisation and dispersal. On one hand, hybrid taxa may be advantaged in new environments (see first section above) and thus more frequently associated with grass lineages derived from long-distance dispersal. Then again, the disjunct distribution of Pooideae in the Tropics may provide less opportunities for hybridisation. Such hypotheses could be tested with well-sampled phylogenies of *Epichloë*. Unfortunately, molecular data is mainly available from species that were isolated from readily accessible live plants, with a bias towards the northern hemisphere (e.g. Craven et al. 2001, Li et al. 2006, Charlton et al. 2012, but see Moon et al. 2002, Gentile et al. 2005). Modern sequencing technologies have allowed to recover not only highly represented chloroplast DNA, but also low-coverage nuclear loci from herbarium specimens (see e.g. chapter III). It may thus be possible to sequence the symbiont along with the plant, especially if the nuclear ribosomal complex, present in repeated copies and commonly used for fungal phylogenetics and barcoding, is targeted. *Epichloë* infections are typically systemic, i.e. the hyphae are present in most aboveground plant tissues (Clay & Schardl 2002). This is supported by the fact that all four Pooideae samples from Madagascar yielded *Epichloë* DNA despite the overall bias towards plant DNA (chapter IV). If fungal DNA could be recovered from herbarium material or silica-dried samples, this would open exciting perspectives for building well-sampled phylogenies, estimating dispersal histories, and confronting them with the evolutionary history of the host plants.

6.4 References

Blattner F.R. (2006): Multiple intercontinental dispersals shaped the distribution area of *Hordeum* (Poaceae). *New Phytologist*, 169: 603–614.

Bond W.J. (2016): Ancient grasslands at risk. *Science*, 351: 120–122.

Bond W.J. & Keeley J.E. (2005): Fire as a global "herbivore": the ecology and evolution of flammable ecosystems. *Trends in Ecology & Evolution*, 20: 387–394.

Bond W.J., Silander Jr J.A., Ranaivonasy J. & Ratsirarson J. (2008): The antiquity of Madagascar's grasslands and the rise of C_4 grassy biomes. *Journal of Biogeography*, 35: 1743–1758.

Bouchenak-Khelladi Y., Verboom G.A., Savolainen V. & Hodkinson T.R. (2010): Biogeography of the grasses (Poaceae): a phylogenetic approach to reveal evolutionary history in geographical space

and geological time. *Botanical Journal of the Linnean Society*, 162: 543–557.

Brown W.V. & Emery W.H.P. (1958): Apomixis in the Gramineae: Panicoideae. *American Journal of Botany*, 45: 253–263.

Burney D.A. (1987): Late Holocene vegetational change in central Madagascar. *Quaternary Research*, 28: 130–143.

Burney D.A. (2003): Madagascar's prehistoric ecosystems. *The natural history of Madagascar* (ed. by S.M. Goodman & J.P. Benstead): 47–51. University of Chicago Press, Chicago, London.

Burney D.A., Robinson G.S. & Burney L.P. (2003): *Sporormiella* and the late Holocene extinctions in Madagascar. *Proceedings of the National Academy of Sciences*, 100: 10800–10805.

Calzada J.-P.V., Crane C.F. & Stelly D.M. (1996): Apomixis: the asexual revolution. *Science*, 274: 1322–1324.

Catalán P. (2006): Phylogeny and evolution of *Festuca* L. and related genera of subtribe Loliinae (Poeae, Poaceae). *Plant genome: biodiversity and evolution*, 1: 255–303.

Charlton N.D., Craven K.D., Mittal S., Hopkins A.A. & Young C.A. (2012): *Epichloë canadensis*, a new interspecific epichloid hybrid symbiotic with Canada wildrye (*Elymus canadensis*). *Mycologia*, 104: 1187–1199.

Christin P.-A., Edwards E.J., Besnard G., Boxall S.F., Gregory R. et al. (2012): Adaptive evolution of C_4 photosynthesis through recurrent lateral gene transfer. *Current Biology*, 22: 445–449.

Clay K. & Schardl C. (2002): Evolutionary origins and ecological consequences of endophyte symbiosis with grasses. *The American Naturalist*, 160: S99–S127.

Craven K.D., Hsiau P.T.W., Leuchtmann A., Hollin W. & Schardl C.L. (2001): Multigene phylogeny of *Epichloë* species, fungal symbionts of grasses. *Annals of the Missouri Botanical Garden*, 88: 14–34.

Desjardins T., Filho A.C., Mariotti A., Girardin C. & Chauvel A. (1996): Changes of the forest-savanna boundary in Brazilian Amazonia during the Holocene revealed by stable isotope ratios of soil organic carbon. *Oecologia*, 108: 749–756.

Donoghue M.J. (2008): A phylogenetic perspective on the distribution of plant diversity. *Proceedings of the National Academy of Sciences*, 105: 11549–11555.

Edwards E.J., Osborne C.P., Strömberg C.A.E. & Smith S.A. (2010): The origins of C_4 grasslands: integrating evolutionary and ecosystem science. *Science*, 328: 587–591.

Freckleton R.P. & Harvey P.H. (2006): Detecting non-Brownian trait evolution in adaptive radiations. *PLOS Biology*, 4: e373.

Frith H.J., Wolfe T.O. & Barker R.D. (1976): Food of eight species of Columbidae, in the genera *Geopelia*, *Phaps*, *Geophaps* and *Petrophassa*. *Wildlife Research*, 3: 159–171.

Gasse F. & Van Campo E. (1998): A 40,000-yr pollen and diatom record from Lake Tritrivakely, Madagascar, in the southern tropics. *Quaternary Research*, 49: 299–311.

Gentile A., Rossi M.S., Cabral D., Craven K.D. & Schardl C.L. (2005): Origin, divergence, and phylogeny of *Epichloë* endophytes of native Argentine grasses. *Molecular Phylogenetics and Evolution*, 35: 196–208.

Gillespie R.G., Baldwin B.G., Waters J.M., Fraser C.I., Nikula R. & Roderick G.K. (2012): Long-distance dispersal: a framework for hypothesis testing. *Trends in Ecology & Evolution*, 27: 47–56. Gray A.J., Marshall D.F. & Raybould A.F. (1991): A century of evolution in *Spartina anglica*. *Advances in Ecological Research* (ed. by M. Begon, A.H. Fitter & A. Macfadyen): 1–62. Academic Press, Cambridge, MA.

Inda L.A., Segarra-Moragues J.G., Müller J., Peterson P.M. & Catalán P. (2008): Dated historical biogeography of the temperate Loliinae (Poaceae, Pooideae) grasses in the northern and southern hemispheres. *Molecular Phylogenetics and Evolution*, 46: 932–957.

Kellogg E.A. (2015): *Flowering Plants. Monocots: Poaceae.* Springer, Heidelberg.

Lehmann C.E.R., Archibald S.A., Hoffmann W.A. & Bond W.J. (2011): Deciphering the distribution of the savanna biome. *New Phytologist*, 191: 197–209.

Lehmann C.E.R. & Parr C.L. (2016): Tropical grassy biomes: linking ecology, human use and conservation. *Philosophical Transaction of the Royal Society B: Biological Sciences*, 371: 20160329.

Li W., Ji Y., Yu H. & Wang Z. (2006): A new species of *Epichloë* symbiotic with Chinese grasses. *Mycologia*, 98: 560–570.

Linder H.P., Antonelli A., Humphreys A.M., Pirie M.D. & Wüest R.O. (2013): What determines biogeographical ranges? Historical wanderings and ecological constraints in the danthonioid grasses. *Journal of Biogeography*, 40: 821–834.

Linder H.P. & Barker N.P. (2014): Does polyploidy facilitate long-distance dispersal? *Annals of Botany*, 113: 1175–1183.

Maurin O., Davies T.J., Burrows J.E., Daru B.H., Yessoufou K. et al. (2014): Savanna fire and the origins of the "underground forests" of Africa. *New Phytologist*, 204: 201–214.

McAllister C.A. & Miller A.J. (2016): Single nucleotide polymorphism discovery via genotyping by sequencing to assess population genetic structure and recurrent polyploidization in *Andropogon gerardii*. *American Journal of Botany*, 103: 1314–1325.

Moon C.D., Miles C.O., Järlfors U. & Schardl C.L. (2002): The evolutionary origins of three new *Neotyphodium* endophyte species from grasses indigenous to the southern hemisphere. *Mycologia*, 94: 694–711.

Nathan R., Perry G., Cronin J.T., Strand A.E. & Cain M.L. (2003): Methods for estimating long-distance dispersal. *Oikos*, 103: 261–273.

Parr C.L., Lehmann C.E.R., Bond W.J., Hoffmann W.A. & Andersen A.N. (2014): Tropical grassy biomes: misunderstood, neglected, and under threat. *Trends in Ecology & Evolution*, 29: 205–213.

Pirie M.D., Humphreys A.M., Barker N.P. & Linder H.P. (2009): Reticulation, data combination, and inferring evolutionary history: an example from Danthonioideae (Poaceae). *Systematic Biology*, syp068.

Ratnam J., Bond W.J., Fensham R.J., Hoffmann W.A., Archibald S. et al. (2011): When is a "forest" a savanna, and why does it matter? *Global Ecology and Biogeography*, 20: 653–660.

Ratnam J., Tomlinson K.W., Rasquinha D.N. & Sankaran M. (2016): Savannahs of Asia: antiquity, biogeography, and an uncertain future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371: 20150305.

Reddy P.V., Bergen M.S., Patel R. & White J.F. (1998): An examination of molecular phylogeny and morphology of the grass endophyte *Balansia claviceps* and similar species. *Mycologia*, 90: 108–117.

Renner S. (2004): Plant dispersal across the tropical Atlantic by wind and sea currents. *International Journal of Plant Sciences*, 165: S23–S33.

Sankaran M., Hanan N.P., Scholes R.J., Ratnam J., Augustine D.J. et al. (2005): Determinants of woody cover in African savannas. *Nature*, 438: 846–849.

Sanmartín I. & Ronquist F. (2004): Southern hemisphere biogeography inferred by event-based models: plant versus animal patterns. *Systematic Biology*, 53: 216–243.

Schardl C.L., Young C.A., Hesse U., Amyotte S.G., Andreeva K. et al. (2013): Plant-symbiotic fungi as chemical engineers: multi-genome analysis of the Clavicipitaceae reveals dynamics of alkaloid loci. *PLOS Genetics*, 9: e1003323.

Soucy S.M., Huang J. & Gogarten J.P. (2015): Horizontal gene transfer: building the web of life. *Nature Reviews Genetics*, 16: 472–482.

Spriggs E.L., Christin P.-A. & Edwards E.J. (2014): C_4 photosynthesis promoted species diversification during the Miocene grassland expansion. *PLoS ONE*, 9: e97722.

Strömberg C.A.E. (2011): Evolution of grasses and grassland ecosystems. *Annual Review of Earth and Planetary Sciences*, 39: 517–544.

Taberlet P., Coissac E., Pompanon F., Brochmann C. & Willerslev E. (2012): Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, 21: 2045–2050.

Van Dam M.H. & Matzke N.J. (2016): Evaluating the influence of connectivity and distance on biogeographical patterns in the south-western deserts of North America. *Journal of Biogeography*, 43: 1514–1532.

Vences M., Wollenberg K.C., Vieites D.R. & Lees D.C. (2009): Madagascar as a model region of species diversification. *Trends in Ecology & Evolution*, 24: 456–465.

Vorholt J.A. (2012): Microbial life in the phyllosphere. *Nature Reviews Microbiology*, 10: 828–840.

Vorontsova M.S., Besnard G., Forest F., Malakasi P., Moat J. et al. (2016): Madagascar's grasses and grasslands: anthropogenic or natural? *Proceedings of the Royal Society of London B: Biological Sciences*, 283: 20152262.

Walker T.D. & Valentine J.W. (1984): Equilibrium models of evolutionary species diversity and the number of empty niches. *The American Naturalist*, 124: 887–899.

White J.F. & Reddy P.V. (1998): Examination of structure and molecular phylogenetic relationships of some graminicolous symbionts in genera *Epichloë* and *Parepichloë*. *Mycologia*, 90: 226–234.

Wiedemeier D.B., Bloesch U. & Hagedorn F. (2012): Stable forest–savanna mosaic in north-western Tanzania: local-scale evidence from $\delta^{13}C$ signatures and ^{14}C ages of soil fractions. *Journal of Biogeography*, 39: 247–257.

Willems J.H. (1982): Establishment and development of a population of *Orchis simia* Lamk. in the Netherlands, 1972 to 1981. *New Phytologist*, 91: 757–765.

Wilmé L., Goodman S.M. & Ganzhorn J.U. (2006): Biogeographic evolution of Madagascar's microendemic biota. *Science*, 312: 1063–1065.

7 Annex

7.1 A-I (chapter I: Factors underlying the global dispersal of temperate Loliinae grasses)

Table A-I.1. Distances used for DEC+x and DEC+j+x models in BioGeoBEARS. They were calculated by locating approximately the centre points of or shortest-distance points between Operational Areas on a world map and then calculating Vincenty ellipsoid distances from them. Distances are here given in km but were normalised to the respective smallest value for the analysis to avoid parameter scaling problems in *BioGeoBEARS* optimisation. Note that for shortest distance, the distance between contiguous regions was set to a very small value (0.01 of the smallest distance) as the distance scaling parameter *x* requires non-zero values.

Shortest distances

	A	B	C	D	E	F	G	H	I	J	K	L
A	0											
B	1177	0										
C	1515	576	0									
D	6314	5792	1927	0								
E	6108	4308	1083	0	0							
F	8118	6664	3392	0	0	0						
G	10404	9254	10406	13251	6038	3367	0					
H	10677	13669	9563	3458	6779	94	10227	0				
I	7620	10843	7868	6858	10046	6974	10582	0	0			
J	6389	9088	6857	7222	9669	9442	8634	3494	0	0		
K	6139	8195	3626	109	3691	1959	17581	1950	5128	5506	0	
L	18692	16865	16933	12613	13254	11394	6060	4191	8434	10717	11247	0

Centre distances

	A	B	C	D	E	F	G	H	I	J	K	L
A	0											
B	2514	0										
C	3134	3257	0									
D	7633	7609	4621	0								
E	7840	6120	5186	4356	0							
F	10939	9103	8165	5600	3100	0						
G	11836	10499	13703	16282	11929	11142	0					
H	14787	16231	12974	8801	11632	9507	13245	0				
I	10390	12895	11071	10155	14509	14473	13608	5891	0			
J	8206	10675	9883	10951	14732	16391	12489	8696	2808	0		
K	7835	9040	5810	3209	7550	8488	19444	7230	6961	7920	0	
L	19262	17549	17603	13111	12763	9673	7625	5682	9572	11536	12776	0

Table A-I.2. Statistics of quasi-Poisson regression of per-route dispersal rates, estimated through stochastic mapping under two range evolution models (DEC and DEC+j+x with shortest distances), on distance, disjunction type, and phenotypic syndrome, and their interactions. Regression models 1–4 use shortest distances between Operational Areas (OAs) as predictor, models 5–8 instead use distances between approximate centres of OAs. Disjunction type is terrestrial (T) or oceanic, Phenotypic syndrome is fine-leaved (FL) or broad-leaved (including intermediate). Significant variables ($p < 0.05$) are highlighted in bold. Coefficient estimates are log-scale.

Coefficient	Estimate	SE	t	p value
1: Range expansion rate – null deviance 0.13 (263 df), residual deviance 0.08 (256 df), dispersion 0.0004				
Intercept	**-7.74227**	**0.31327**	**-24.71**	**<0.001**
Phenotypic syndrome (FL)	0.03174	0.4348	0.07	0.942
Disjunction type (T)	0.72555	0.3686	1.97	0.05
Shortest distance	**-0.00027**	**0.00006**	**-4.48**	**<0.001**
Phenotypic syndrome (FL) : Disjunction type (T)	-0.63194	0.54152	-1.17	0.244
Phenotypic syndrome (FL) : Shortest distance	0.00002	0.00008	0.26	0.798
Disjunction type (T) : Shortest distance	-0.00009	0.00013	-0.69	0.491
Phenotypic syndrome (FL) : Disjunction type (T) : Shortest distance	0.00006	0.00019	0.31	0.755
1b: Range expansion rate – null deviance 0.13 (263 df), residual deviance 0.08 (262 df), dispersion 0.00042				
Intercept	**-7.43967**	**0.12953**	**-57.44**	**<0.001**
Shortest distance	**-0.0003**	**0.00003**	**-8.88**	**<0.001**
2: Range expansion rate – null deviance 0.08 (263 df), residual deviance 0.04 (256 df), dispersion 0.00034				
Intercept	**-8.18359**	**0.45841**	**-17.85**	**<0.001**
Phenotypic syndrome (FL)	0.10299	0.60581	0.17	0.865
Disjunction type (T)	0.57749	0.52245	1.11	0.27
Shortest distance	**-0.00061**	**0.00018**	**-3.29**	**0.001**
Phenotypic syndrome (FL) : Disjunction type (T)	-0.78545	0.74403	-1.06	0.292
Phenotypic syndrome (FL) : Shortest distance	0.00017	0.00021	0.8	0.425
Disjunction type (T) : Shortest distance	0.00006	0.00028	0.21	0.834
Phenotypic syndrome (FL) : Disjunction type (T) : Shortest distance	-0.00015	0.00041	-0.38	0.707
2b: Range expansion rate – null deviance 0.08 (263 df), residual deviance 0.04 (262 df), dispersion 0.00037				
Intercept	**-7.97621**	**0.1746**	**-45.68**	**<0.001**
Shortest distance	**-0.00053**	**0.00008**	**-6.55**	**<0.001**
3: Founder-event dispersal rate – null deviance 0.11 (263 df), residual deviance 0.06 (256 df), dispersion 0.00041				
Intercept	**-7.80806**	**0.33987**	**-22.97**	**<0.001**
Phenotypic syndrome (FL)	-0.32361	0.50485	-0.64	0.522
Disjunction type (T)	0.53648	0.40922	1.31	0.191
Shortest distance	**-0.0003**	**0.00007**	**-4.32**	**<0.001**
Phenotypic syndrome (FL) : Disjunction type (T)	0.0656	0.61027	0.11	0.914
Phenotypic syndrome (FL) : Shortest distance	0.00007	0.0001	0.68	0.497
Disjunction type (T) : Shortest distance	-0.00015	0.00017	-0.88	0.382
Phenotypic syndrome (FL) : Disjunction type (T) : Shortest distance	0.00001	0.00024	0.02	0.982
3b: Founder-event dispersal rate – null deviance 0.11 (263 df), residual deviance 0.07 (262 df), dispersion 0.00043				
Intercept	**-7.60978**	**0.14563**	**-52.26**	**<0.001**
Shortest distance	**-0.00033**	**0.00004**	**-8.03**	**<0.001**
4: Total dispersal rate – null deviance 0.16 (263 df), residual deviance 0.08 (256 df), dispersion 0.00052				
Intercept	**-7.34525**	**0.31835**	**-23.07**	**<0.001**
Phenotypic syndrome (FL)	-0.11645	0.45151	-0.26	0.797
Disjunction type (T)	0.61262	0.3746	1.64	0.103
Shortest distance	**-0.00036**	**0.00007**	**-4.81**	**<0.001**
Phenotypic syndrome (FL) : Disjunction type (T)	-0.29872	0.54828	-0.54	0.586
Phenotypic syndrome (FL) : Shortest distance	0.00006	0.0001	0.61	0.541
Disjunction type (T) : Shortest distance	-0.00013	0.00016	-0.81	0.417
Phenotypic syndrome (FL) : Disjunction type (T) : Shortest distance	0.00001	0.00024	0.03	0.973
4b: Total dispersal rate – null deviance 0.16 (263 df), residual deviance 0.08 (262 df), dispersion 0.00054				

Intercept	**-7.09961**	**0.12936**	**-54.88**	**<0.001**
Shortest distance	**-0.00038**	**0.00004**	**-9.15**	**<0.001**
5: Range expansion rate – null deviance 0.13 (263 df), residual deviance 0.1 (256 df), dispersion 0.0007				
Intercept	**-8.58045**	**0.7563**	**-11.35**	**<0.001**
Phenotypic syndrome (FL)	-0.25412	1.05953	-0.24	0.811
Disjunction type (T)	**2.85393**	**0.97333**	**2.93**	**0.004**
Centre-distance	-0.00007	0.00007	-0.99	0.321
Phenotypic syndrome (FL) : Disjunction type (T)	-1.5627	1.44818	-1.08	0.282
Phenotypic syndrome (FL) : Centre-distance	0.00004	0.0001	0.37	0.709
Disjunction type (T) : Centre-distance	-0.00029	0.00015	-1.94	0.054
Phenotypic syndrome (FL) : Disjunction type (T) : Centre-distance	0.00023	0.00021	1.09	0.276
5b: Range expansion rate – null deviance 0.13 (263 df), residual deviance 0.11 (262 df), dispersion 0.00071				
Intercept	**-7.29591**	**0.29479**	**-24.75**	**<0.001**
Centre-distance	**-0.00017**	**0.00004**	**-4.7**	**<0.001**
6: Range expansion rate – null deviance 0.08 (263 df), residual deviance 0.05 (256 df), dispersion 0.0006				
Intercept	**-8.55017**	**1.14698**	**-7.45**	**<0.001**
Phenotypic syndrome (FL)	-1.14907	1.59128	-0.72	0.471
Disjunction type (T)	**3.57341**	**1.51857**	**2.35**	**0.019**
Centre-distance	-0.00022	0.00013	-1.66	0.099
Phenotypic syndrome (FL) : Disjunction type (T)	-1.65652	2.17344	-0.76	0.447
Phenotypic syndrome (FL) : Centre-distance	0.00017	0.00017	1.05	0.294
Disjunction type (T) : Centre-distance	-0.0005	0.00029	-1.73	0.085
Phenotypic syndrome (FL) : Disjunction type (T) : Centre-distance	0.00034	0.00037	0.91	0.365
6b: Range expansion rate – null deviance 0.08 (263 df), residual deviance 0.06 (262 df), dispersion 0.00061				
Intercept	**-7.49776**	**0.4141**	**-18.11**	**<0.001**
Centre-distance	**-0.00027**	**0.00006**	**-4.59**	**<0.001**
7: Founder-event dispersal rate – null deviance 0.11 (263 df), residual deviance 0.09 (256 df), dispersion 0.00057				
Intercept	**-8.38297**	**0.72033**	**-11.64**	**<0.001**
Phenotypic syndrome (FL)	-0.38464	1.05468	-0.36	0.716
Disjunction type (T)	1.55804	0.95615	1.63	0.104
Centre-distance	-0.00011	0.00007	-1.56	0.12
Phenotypic syndrome (FL) : Disjunction type (T)	-0.39653	1.41484	-0.28	0.779
Phenotypic syndrome (FL) : Centre-distance	0.00003	0.0001	0.33	0.743
Disjunction type (T) : Centre-distance	-0.00008	0.00014	-0.59	0.553
Phenotypic syndrome (FL) : Disjunction type (T) : Centre-distance	0.00008	0.00019	0.39	0.698
7b: Founder-event dispersal rate – null deviance 0.11 (263 df), residual deviance 0.09 (262 df), dispersion 0.00066				
Intercept	**-7.43678**	**0.31744**	**-23.43**	**<0.001**
Centre-distance	**-0.00018**	**0.00004**	**-4.61**	**<0.001**
8: Total dispersal rate – null deviance 0.16 (263 df), residual deviance 0.12 (256 df), dispersion 0.00086				
Intercept	**-7.85486**	**0.74581**	**-10.53**	**<0.001**
Phenotypic syndrome (FL)	-0.59154	1.07212	-0.55	0.582
Disjunction type (T)	**2.24682**	**0.96606**	**2.33**	**0.021**
Centre-distance	-0.00014	0.00008	-1.79	0.074
Phenotypic syndrome (FL) : Disjunction type (T)	-0.85614	1.42963	-0.6	0.55
Phenotypic syndrome (FL) : Centre-distance	0.00007	0.00011	0.68	0.497
Disjunction type (T) : Centre-distance	-0.0002	0.00015	-1.36	0.175
Phenotypic syndrome (FL) : Disjunction type (T) : Centre-distance	0.00015	0.00021	0.72	0.471
8b: Total dispersal rate – null deviance 0.16 (263 df), residual deviance 0.13 (262 df), dispersion 0.00088				
Intercept	**-6.82333**	**0.29575**	**-23.07**	**<0.001**
Centre-distance	**-0.00021**	**0.00004**	**-5.47**	**<0.001**

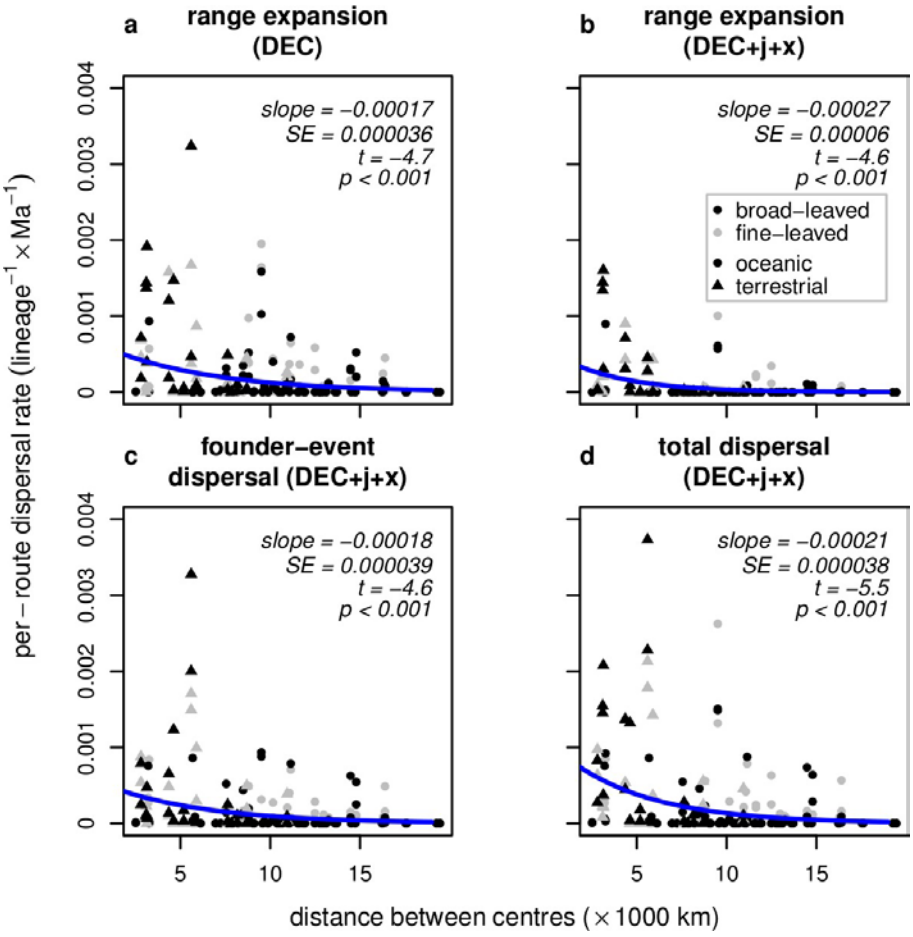

Figure A-I.1. Mean per-route dispersal rates in Loliinae, inferred with biogeographical stochastic mapping, plotted against the centre distance between respective OAs. Compare with Fig. I.4, which shows the regression on shortest distances.

7.2 A-II (chapter II: Grass diversification in Madagascar)

Taxon sampling and datasets

We aimed to analyse a representative set of molecular data for the grass flora of Madagascar and the wider Madagascar region, focusing especially on endemic species. To this end, we sampled specimens collected in the field since 2011 as well as herbarium specimens. We also included additional species putatively close to Malagasy taxa, especially for groups particularly diverse in Madagascar such as the Hickeliinae (Bambusoideae) and the Boivinellinae (Panicoideae). For some species with complex species-level taxonomy or a large distribution range, we included two or more specimens. Altogether, we produced new data for 238 specimens from Madagascar or nearby islands and 133 specimens from other areas. Including sequence data from *GenBank* (see below), an estimated 396 species occurring in Madagascar are represented in the phylogenies (73% of the known species diversity, including some putative new taxa; 77% of the known C_3 species and 71% of the C_4 species). Of these, 144 are endemic (65% of known endemics, 71% of the endemic C_3 species and 59% of the endemic C_4 species), and another four are endemic to Réunion, Glorioso or Europa Island.

Molecular data were organised in 11 datasets corresponding to major grass clades recovered in previous work, mostly subfamilies or large tribes (GPWG II 2012, Kellogg, 2015, Soreng et al. 2015), following the classification of Soreng et al. (2015): Oryzoideae, Bambusoideae, Brachypodieae, Poeae, Aristidoideae, Panicoideae (focusing on the "outlying Panicoideae" and *Lecomtella*, representing the large Andropogoneae–Arundinelleae, Paniceae and Paspaleae clades with two taxa each), Paniceae, Andropogoneae–Arundinelleae, Arundinoideae–Micraioideae, Chloridoideae, and Danthonioideae. Total species sampling in these clades ranged from roughly 10% (Andropogoneae–Arundinelleae) to 65% (Brachypodieae). The clades Paspaleae and Pharoideae, present with two and one native species in Madagascar, respectively (Vorontsova et al. 2016), have no endemics and were not analysed here. *Bromus* was not analysed either as we were unable to collect the two endemic species.

DNA extraction, PCR, and sequencing

We selected three chloroplast genes, *ndhF*, *rbcL*, and *trnK–matK*, as molecular markers. This combination has provided good phylogenetic resolution at both shallow and deep levels in previous work on grasses (GPWG II 2012). DNA was extracted from ground herbarium or silica-dried material using the BioSprint 15 DNA Plant Kit (Qiagen Inc.). Markers were PCR-amplified, using primers and procedures as described before (GPWG II 2012, Taylor et al. 2012), and sequenced with Sanger technology. For some samples, extracted DNA was instead processed with a genome-skimming sequencing procedure on a HiSeq platform that has also been described previously (Besnard et al. 2013). Bases from Sanger-derived chromatographs were assigned in *Geneious* 8 (Biomatters Ltd., Auckland, New Zealand) and checked

manually. Genome-skimming sequence reads were assembled into full chloroplast genomes or contigs containing the three marker genes, depending on sequence quality; see Besnard et al. (2013) for the method followed. All newly produced sequences, listed in Appendix S1, are accessible through *GenBank*.

Sequence mining from GenBank and alignment

To represent as much of the phylogenetic grass diversity as possible, we retrieved all *ndhF*, *trnK–matK* and *rbcL* sequences available on *GenBank* (as of August 2015) for each clade and selected the longest sequence with fewest ambiguities for each marker per species or subspecies recognised in the NCBI taxonomy. In the Bambusoideae, where low rates of molecular evolution produced insufficient phylogenetic information, we added sequences for the *rps16* and *rpl16* introns from *GenBank*. The thus assembled data were combined with the newly produced sequences, and outgroup sequences were added to each dataset.

Using *MAFFT* v.7 (http://mafft.cbrc.jp/alignment/server/; Katoh & Standley 2013) with default settings, we first roughly aligned each marker per dataset. Rough Maximum Likelihood (ML) trees produced with *RAxML* v.8 (Stamatakis 2014) were used to inspect each alignment for sequences in strong conflict between marker sets, or on very long branches, suggesting errors; these were removed or replaced. We only kept accessions represented by at least *ndhF* or *trnK–matK*, whichever was more common in each dataset. Thus, all accessions in an alignment had at least one marker in common. We removed several *GenBank* accessions in the large genera *Chusquea*, *Muhlenbergia*, *Poa*, and *Rytidosperma* that showed little or no divergence and were mostly represented by a single marker.

Sequences for some specimens collected in Madagascar were removed because they showed no genetic divergence from others on the markers studied although they were sometimes identified as different species morphologically. They are nevertheless available on *GenBank*. In three cases, potential endemics (*Andropogon trichozygus, Hickelia madagascariensis* MSV986 and *Perotis arenacea*) were excluded in favour of a non-endemic relative. This was done to conservatively give less weight to dubious endemics in the analysis of range evolution.

The datasets were then re-aligned, trimmed and manually adjusted. Some ambiguously aligned regions were removed from the *trnK–matK* alignments (see Table A-II.1). For nomenclature we followed the World Checklist of Selected Plant Families (WCSP, Clayton et al. 2016); in cases where WCSP suggested synonymy but taxa appeared divergent or non-monophyletic in the tree, we retained the NCBI nomenclature. Final datasets contained between 18 (Brachypodieae) and 516 (Paniceae) accessions, for a total of 1928.

Table A-II.1. Ambiguously aligned regions that were removed from the *trnK–matK* alignments.

Clade	Position in alignment	Length (bp)	Type
Andropogoneae–Arundinelleae	155	32	poly-T
	245	9	poly-A
	426	11	poly-T
Aristidoideae	76	14	poly-T
Arundinoideae–Micrairoideae	83	16	poly-T
	298	8	poly-T
Bambusoideae	409	19	poly-T
Chloridoideae	320	13	poly-T
	520	22	poly-T
	2600	14	poly-AAAT
	2628	15	poly-A
Danthonioideae	353	23	poly-T
	530	10	poly-T
Oryzoideae	348	13	poly-T
	2362	78	longer repeated motif
Panicoideae	76	19	poly-T
Paniceae	104	33	poly-T
	320	14	poly-T
	2638	63	ambiguous region

Phylogenetics and dating

We first estimated ML phylogenies for each concatenated dataset per clade in *RAxML* on the *CIPRES* web platform (Miller et al. 2012), applying a GTR substitution model and accounting for rate heterogeneity with 25 Gamma rate categories per marker. Clade support was assessed with 1000 bootstrap pseudoreplicates. Joint Bayesian tree and divergence time estimation was then performed in *BEAST2* (Bouckaert et al. 2014), using the ML phylogeny as starting tree. Partitions and substitution models were as described for *RAxML* but using only four Gamma rate categories; the diversification process was approximated with a Yule model; molecular rate heterogeneity was modelled with a lognormal relaxed clock (Drummond et al., 2006). The choice of the parameter-rich GTR+G substitution model was validated *a posteriori* by ensuring that substitution rates were well estimated and distinct. First tests estimating also the proportion of invariable sites (+I) led to poor estimates for that additional parameter and thus we did not include it.

Dating priors for a first estimation in *BEAST2* were taken from a previous study that had used external angiosperm calibration to estimate divergence times in the Poaceae (Christin et al. 2014). Widely used macrofossils are younger than estimates of Christin et al. (2014) for the respective nodes and thus not very informative as minimum constraints, e.g. *Cleistochloa* and *Distichlis* with ca. 14 Ma (Dugas & Retallack 1993) or the oldest known grass spikelet from the Paleocene/Eocene boundary (Crepet & Feldman 1991). Phytolith fossils suggest older ages but are controversial, see below. Median ages for clades recovered by Christin et al. (2014) were used as mean (μ) for normal age priors on nodes, and the 95% high posterior

density (HPD) intervals were used as normal standard deviation (*SD*) by dividing the difference between upper and lower bound by 4 (see Table A-II.2). For Andropogoneae–Arundinelleae, not represented as a clade in Christin et al. (2014), we used the estimates for the crown of Andropogoneae s.str. and the split from Paspaleae.

We ran three to five MCMC chains with 10 to 50 million iterations, depending on the dataset. In *Tracer* v.1.6 (Rambaut et al. 2014), we checked that chains had converged and ESS values for all relevant parameters were >200, and we combined independent runs, discarding a 10% burn-in each, in *LogCombiner*. Age estimates and posterior clade probabilities from the combined estimates were summarised on the Maximum Clade Credibility (MCC) tree for each dataset in *TreeAnnotator*.

Christin et al. (2014) found that calibrating divergences in the Poaceae with controversial phytolith fossils yielded significantly older estimates that appear to contradict some of the assumed major events in grass evolution; they concluded that external and phytolith calibration are best treated as alternative hypotheses until more conclusive evidence becomes available. Consequently, we used phytolith-derived ages from that study as calibration priors in a second *BEAST2* analysis, using the MCC trees estimated under the first calibration as starting trees with fixed topology.

Table A-II.2. Calibrated nodes and age priors (in Ma) used (taken from Christin et al. 2015).

Node	Calibration 1 (external, without phytoliths)		Calibration 2 (with phytoliths)	
	μ	***SD***	***μ***	***SD***
Andropogoneae crown	11.792	2.9555	14.45	3.383
Andropogoneae–Paspaleae split	28.502	3.335	34.296	3.7335
Aristidoideae crown	32.59	6.034	41.532	7.7445
Arundinoideae–Micrairoideae split	38.518	5.1265	47.506	6.9835
Bambusoideae crown	28.376	7.0585	34.208	9.0975
Brachypodieae divergence	34.218	3.33	43.712	4.2565
Chloridoideae crown	38.532	3.6305	49.116	4.359
Danthonioideae crown	29.242	4.6135	38.732	6.4875
Oryzoideae crown	38.548	6.126	67.99	0.941
Panicoideae crown	38.188	3.8605	48.096	4.9415
Paniceae crown	25.46	2.767	30.74	2.971
Poeae crown	26.588	2.8435	32.296	3.636

As estimates of the age of the Madagascan grass flora, we obtained the median ages and 95% HPD intervals for stem and crown of endemic lineages recovered in the MCC trees under both age calibrations. C_4 clades were identified based on known C_4 genera or the photosynthetic pathway of close species in the case of the polyphyletic genera *Brachiaria* and *Panicum* (Osborne et al. 2014). Some taxa have not been analysed for their photosynthetic pathway, but their position within known C_3 or C_4 clades was well supported.

Table A-II.3. Lineages strictly endemic to the Madagascar region, recovered through phylogenetic analysis. Photosynthetic pathway, median stem age under external calibration (cal. 1) and calibration with phytoliths (cal. 2), tip number, and an estimate of true species richness are given.

No.	Clade	Tips	Photos. pathway	Median stem age (Ma)		No. of tips	Estimated species richness
				Cal. 1	Cal. 2		
1	Oryzoideae	*Leersia perrieri*, *Leersia perrieri* HPB17323	C_3	1.4	2.52	2	1
2	Oryzoideae	*Maltebrunia leersioides* GES3926	C_3	2.82	5.06	1	2
3	Oryzoideae	*Humbertochloa bambusiuscula* MSV956	C_3	3.19	5.94	1	1
4	Bambusoideae	*Hickelia madagascariensis*	C_3	0.45	0.55	1	1
5	Bambusoideae	*Hickelia madagascariensis* MSV724, *Hickelia madagascariensis* MSV731	C_3	1.03	1.32	2	3
6	Bambusoideae	*Arundinaria* sp. MSV723, *Arundinaria marojejyensis* MSV487, *Arundinaria humbertii* MSV1223	C_3	2.65	3.11	3	7
7	Bambusoideae	*Hitchcockella baronii* DR and TA 3430, *Sokinochloa chapelieri* MSV319, *Sokinochloa australis* LRK780, *Sokinochloa brachyclada* SD1352, *Cephalostachyum chapelieri* MSV499, *Perrierbambus* sp. MSV1321, *Perrierbambus madagascariensis* TRS2665, *Valiha diffusa* MSV1487, *Sirochloa parvifolia* NOP224, *Decaryochloa diadelpha* SD1116, *Nastus borbonicus*, *Nastus* cf. *perrieri* MSV495, *Nastus aristatus* LG2683, *Nastus elongatus* MSV309	C_3	6.01	7.34	14	27
8	Brachypodieae	*Brachypodium perrieri* MSV1242	C_3	2.58	3.22	1	1
9	Brachypodieae	*Brachypodium madagascariense* MSV735, *Brachypodium perrieri* MSV616, *Brachypodium perrieri* LRK2010	C_3	5.07	6.06	3	3
10	Poeae	*Festuca borbonica* GB03-2013	C_3	0.64	0.77	1	1
11	Poeae	*Agrostis elliotii* LRK2033, *Agrostis elliotii* MSV1037	C_3	1.31	1.57	2	1
12	Poeae	*Poa perrieri* MSV1232	C_3	1.44	1.62	1	2
13	Poeae	*Agrostis* sp. MSV1846	C_3	2.18	2.58	1	5
14	Poeae	*Festuca camusiana* LRK2032, *Festuca camusiana* MSV617, *Festuca perrieri* NOP85	C_3	2.96	3.58	3	2
15	Poeae	*Anthoxanthum madagascariense* MSV613	C_3	2.14	2.53	1	1
16	Poeae	*Agrostis emirnensis* MSV600	C_3	3.77	4.40	1	1
17	Poeae	*Pseudobromus breviligulatus* MSV725	C_3	5.44	6.73	1	3
18	Aristidoideae	*Sartidia perrieri* HPB10751	C_3	1.73	2.32	1	1
19	Aristidoideae	*Sartidia isaloensis* PM3609, *Sartidia isaloensis* MSV1325	C_3	3.29	4.04	2	1
20	Aristidoideae	*Aristida tenuissima* MSV1053	C_4	1.37	1.69	1	1
21	Aristidoideae	*Aristida similis* P02260258, *Aristida similis* LRK2006, *Aristida rufescens* MSV586, *Aristida rufescens* MSV337, *Aristida rufescens* MSV333, *Aristida* cf. *similis* MSV346	C_4	8.41	10.54	6	2
22	Aristidoideae	*Aristida ambongensis* MPG17367, *Aristida ambongensis* RNF2144	C_4	12.49	17.13	2	1
23	Panicoideae	*Lecomtella madagascariensis*	C_3	22.34	28.27	1	1
24	Panicoideae	*Tristachya humbertii* MSV1369, *Tristachya betsileensis* MSV1003, *Danthoniopsis isalensis* MSV1370	C_4	5.23	6.47	3	3
25	Paniceae	*Cyrtococcum tamatavense* RCH39	C_3	0.3	0.36	1	1
26	Paniceae	*Panicum* aff. *vohitrense* MSV1847	C_3	0.31	0.37	1	3
27	Paniceae	*Brachiaria* sp. nov. MSV387, *Brachiaria* sp. nov. RCH43	C_3	0.83	1.08	2	1
28	Paniceae	*Panicum malacotrichum* RCH19	C_3	1.21	1.55	1	1
29	Paniceae	*Cyphochlaena madagascariensis*, *Cyphochlaena madagascariensis* MSV954	C_3	3.3	4.09	2	1

No.	Clade	Tips	Photos. pathway	Median stem age (Ma)		No. of tips	Estimated species richness
				Cal. 1	Cal. 2		
30	Paniceae	*Sacciolepis viguieri* RGD190	C_3	3.03	3.76	1	2
31	Paniceae	*Poecilostachys bakeri* MSV301, *Poecilostachys* sp. MSV492, *Poecilostachys* sp. RCH65, *Poecilostachys gougerotiana* RH1639, *Poecilostachys hildebrandtii* RCH04, *Poecilostachys geminata* RCH05, *Poecilostachys bromoides*, *Oplismenus flavicomus* RCH25, *Poecilostachys festucacea* RCH30, *Oplismenus flavicomus* MSV737, *Poecilostachys bakeri* LRK2035	C_3	3.95	4.99	11	16
32	Paniceae	*Panicum ambositrense* MSV736, *Panicum perrieri* MSV749	C_3	2.75	3.44	2	2
33	Paniceae	*Acroceras ivohibense* MSV1222	C_3	3.87	4.46	1	1
34	Paniceae	*Brachiaria antsirabensis* AJL1038, *Brachiaria dimorpha* HH3879	C_3	4.58	5.45	2	2
35	Paniceae	*Panicum lycopodioides* GB04-2013, *Brachiaria epacridifolia* MSV722, *Brachiaria dimorpha* MSV1234	C_3	4.93	5.90	3	3
36	Paniceae	*Oplismenus* sp. MSV1468	C_3	5.49	6.76	1	1
37	Paniceae	*Panicum subhystrix* MSV748	C_3	5.77	7.72	1	3
38	Paniceae	*Cyrtococcum deltoideum* RCH10, *Cyrtococcum bosseri* NOP228	C_3	6.64	7.87	2	4
39	Paniceae	*Pseudolasiacis leptolomoides* MSV983	C_3	7.39	9.34	1	3
40	Paniceae	*Brachiaria tsiafajavonensis* MSV697	C_3	6.73	8.62	1	1
41	Paniceae	*Brachiaria fragrans* P06769493, *Panicum ibitense* RGD171, *Panicum spergulifolium* MSV1250, *Panicum* sp. nov. MSV1218, *Panicum cupressifolium* MSV1233, *Panicum andringitrense* MSV591, *Brachiaria bemarivensis* MSV936	C_3	10.7	13.05	7	7
42	Paniceae	*Chasechloa madagascariensis* HPB103, *Chasechloa egregia* LHB s.n.	C_3	11.74	14.21	2	2
43	Paniceae	*Acroceras* cf. *boivinii* LRK2014, *Acroceras boivinii* MSV985, *Acroceras tenuicaule* RCH61, *Acroceras* sp. nov. MSV1379, *Acroceras calcicola* NOP215	C_3	13.21	16.00	5	7
44	Paniceae	*Thuarea perrieri* MBV and AMT 1534	C_4	0.28	0.35	1	1
45	Paniceae	*Eriochloa subulifera* MSV351	C_4	0.34	0.45	1	1
46	Paniceae	*Brachiaria pseudodichotoma* MSV942	C_4	0.33	0.45	1	2
47	Paniceae	*Panicum mahafalense* GMP14929	C_4	0.9	1.08	1	2
48	Paniceae	*Brachiaria subrostrata* MSV716	C_4	1.32	1.73	1	2
49	Paniceae	*Pennisetum pseudotriticoides* MSV718	C_4	1.77	2.47	1	1
50	Paniceae	*Digitaria glauca* MSV950	C_4	2.24	2.74	1	3
51	Paniceae	*Setaria bathiei* RGD196	C_4	1.79	2.17	1	1
52	Paniceae	*Cenchrus cafer*	C_4	2.81	3.31	1	1
53	Paniceae	*Digitaria ankaratrensis* MSV1044	C_4	2.98	3.76	1	3
54	Paniceae	*Brachiaria pseudodichotoma* PBP3180	C_4	1.4	1.14	1	3
55	Paniceae	*Stenotaphrum unilaterale* MSV1043	C_4	3.71	4.37	1	1
56	Paniceae	*Panicum vohitrense* MSV1466	C_4	4.1	5.46	1	3
57	Paniceae	*Brachiaria humbertiana* MSV1378, *Brachiaria* cf. *humbertiana* MSV646	C_4	5.7	7.46	2	1
58	Paniceae	*Yvesia madagascariensis* MSV957	C_4	6.41	8.36	1	1
59	Paniceae	*Setaria humbertiana* MSV1414	C_4	7.3	8.45	1	1
60	Paniceae	*Stenotaphrum oostachyum* MSV1042	C_4	7.16	8.04	1	1
61	Paniceae	*Panicum* cf. *voeltzkowii* CF, *Panicum voeltzkowii* EUR27, *Panicum* sp. GLO19, *Panicum cinctum* MSV1027, *Panicum cinctum* MSV1038, *Panicum* cf. *voeltzkowii* MSV937, *Panicum luridum* MSV741, *Panicum luridum* MSV909	C_4	10.3	12.60	8	3

No.	Clade	Tips	Photos. pathway	Median stem age (Ma)		No. of tips	Estimated species richness
				Cal. 1	Cal. 2		
62	Paniceae	*Setaria scottii* MSV1041, *Setaria madecassa* MSV979, *Setaria vatkeana* MSV1813	C_4	11.06	13.87	3	8
63	Andropogoneae-Arundinelleae	*Chrysopogon humbertianus* MSV1413	C_4	0.27	0.32	1	1
64	Andropogoneae-Arundinelleae	*Andropogon andringitrensis* MSV1260	C_4	0.34	0.42	1	1
65	Andropogoneae-Arundinelleae	*Elionurus tristis* MSV589	C_4	0.93	1.16	1	1
66	Andropogoneae-Arundinelleae	*Andropogon ibityensis* MSV754	C_4	1.43	1.77	1	5
67	Andropogoneae-Arundinelleae	*Arthraxon antsirabensis* NOP217	C_4	1.1	1.39	1	1
68	Andropogoneae-Arundinelleae	*Ischaemum koleostachys* MSV1299, *Ischaemum koleostachys*	C_4	1.72	2.09	2	1
69	Andropogoneae-Arundinelleae	*Andropogon itremoensis* MSV753	C_4	2.27	2.83	1	1
70	Andropogoneae-Arundinelleae	*Saccharum hildebrandtii* MSV739, *Saccharum perrieri* NOP140	C_4	2.62	3.28	2	2
71	Andropogoneae-Arundinelleae	*Saccharum viguieri* MSV1248, *Saccharum hildebrandtii* LRK2008	C_4	3.07	3.86	2	2
72	Andropogoneae-Arundinelleae	*Dimeria manongarivensis* MSV343	C_4	5.1	6.15	1	3
73	Arundinoideae-Micrairoideae	*Isachne muscicola* MSV597, *Isachne humbertiana* MSV496	C_3	1.13	1.36	2	4
74	Arundinoideae-Micrairoideae	*Isachne humicola* MSV563	C_3	1.9	2.29	1	1
75	Arundinoideae-Micrairoideae	*Styppeiochloa hitchcockii* MSV993	C_3	7.21	8.44	1	3
76	Arundinoideae-Micrairoideae	*Styppeiochloa* sp. MSV1306	C_3	10.48	12.40	1	3
77	Chloridoideae	*Eragrostis capuronii* NOP186	C_4	0.28	0.34	1	1
78	Chloridoideae	*Eragrostis lateritica* MSV638	C_4	0.61	0.77	1	1
79	Chloridoideae	*Eragrostis* cf. *bemarivensis* MSV923, *Eragrostis hildebrandtii* MSV590	C_4	0.97	1.35	2	2
80	Chloridoideae	*Lepturus* cf. *humbertianus* MSV1403	C_4	1.03	1.30	1	1
81	Chloridoideae	*Eragrostis betsileensis* MSV1002	C_4	1.04	1.29	1	1
82	Chloridoideae	*Neostapfiella perrieri* JB20177	C_4	1.43	1.84	1	3
83	Chloridoideae	*Eragrostis boinensis* MSV935	C_4	2.67	3.20	1	2
84	Chloridoideae	*Chloris humbertiana* NOP206	C_4	3.48	4.32	1	1
85	Chloridoideae	*Lepturus anadabolavensis* NOP203	C_4	3.9	4.89	1	5
86	Chloridoideae	*Eragrostis humbertii* MSV1417	C_4	4.46	5.73	1	1
87	Chloridoideae	*Decaryella madagascariensis* TBC31391, *Decaryella madagascariensis* MSV1398	C_4	3.61	5.20	2	1
88	Chloridoideae	*Dactyloctenium capitatum* MSV1393	C_4	6.29	7.98	1	1
89	Chloridoideae	*Neostapfiella* sp. MSV1486, *Neostapfiella* sp. MSV1772	C_4	6.51	7.96	2	2
90	Chloridoideae	*Eragrostis stolonifera* MSV1046	C_4	6.86	8.24	1	1
91	Chloridoideae	*Acrachne perrieri* AJML13867, *Acrachne* sp. MSV1767	C_4	6.47	7.98	2	2
92	Chloridoideae	*Viguierella madagascariensis* MSV966	C_4	6.96	9.55	1	1
93	Chloridoideae	*Sporobolus perrieri* MSV1415, *Sporobolus* sp. AA156, *Sporobolus halophilus* MSV585	C_4	9.21	11.74	3	3
94	Chloridoideae	*Eragrostis chabouisii* NOP185	C_4	16.25	20.22	1	1
95	Danthonioideae	*Merxmuellera ambalavaoensis* MSV1235, *Merxmuellera tsaratananensis* MSV486	C_3	1.42	1.81	2	3
96	Danthonioideae	*Pentameris andringitrensis* MSV1229	C_3	5.64	7.46	1	3

Analysis of range evolution

ML analysis of range evolution, implemented in the R package *BioGeoBEARS* (Matzke 2013), was used to estimate the history of lineage dispersal to and from Madagascar. The large possible state space resulting from the wide distribution ranges of many grasses limited the analyses to eight areas, defined as follows: the Madagascar region, Tropical Africa, Southern Africa, Eurasia plus North Africa, the Indian subcontinent, Southeast Asia plus Australasia, North America, and the Neotropics (see also map inset in Fig. II.5). We chose to combine Madagascar with the smaller West Indian Ocean islands as only four of the taxa included were native to one of the islands but not Madagascar. Presence in each area (ignoring introductions) was scored according to WCSP as of April 2016, and occurrences in Madagascar were cross-checked against the list of Vorontsova et al. (2016). We calculated Vincenty ellipsoid distances between approximate area centres with the R package *geosphere* (Hijmans 2016) and rescaled them by the smallest distance (see Table S2.5).

Table A-II.4. Areas defined for range evolution analysis, approximate geographical centres, and Vincenty ellipsoid distances between area centres, scaled by the smallest distance.

Area		Appr. centre		Vincenty ellipsoid distance to area centre (relative)							
Code	**Description**	**Lat.**	**Long.**	**E**	**W**	**I**	**S**	**U**	**M**	**O**	**F**
E	Eurasia and North Africa	49.8	80.5	0							
W	West Indian Ocean (incl. Madagascar)	-18.4	47.1	3.22	0						
I	Indian Subcontinent	21.6	78.8	1.23	2.20	0					
S	Southern Africa	-25.5	23.5	3.93	1	3.10	0				
U	Southeast Asia and Australasia	-1.81	126	2.81	3.43	2.25	4.37	0			
M	North America	43.6	-102.1	3.78	6.26	5.00	5.77	5.24	0		
O	Neotropics	-13.4	-61.1	5.74	4.46	6.14	3.46	7.10	2.96	0	
F	Tropical Africa	3.5	24.8	2.86	1.35	2.41	1.26	4.41	4.92	3.78	0

For each dataset, we used the MCC tree obtained under the external calibration without phytoliths, outgroups removed, to compare four range evolution models: (i) the dispersal–extinction–cladogenesis (DEC) model (Ree & Smith, 2008); (ii) a DEC model expanded with *j*, a relative weight for founder-event speciation (DEC+*j*; Matzke 2014); (iii) an expanded DEC model scaling dispersal probabilities by distance between areas to power *x* (DEC+*x*; Van Dam & Matzke 2016); (iv) a model integrating both *j* and *x* (DEC+*j*+*x*). As maximum range size, we set the present-day maximum in each lineage (from 3 in Bambusoideae to 8 in several clades; see Table A-II.6). Likelihood optimisation was performed for each model, and convergence was verified with the output of the *optimx* function from the package of the same name (Nash 2014). Model performance was assessed using the Akaike Information Criterion corrected for sample size (AICc), penalising parameter-rich models more strongly for smaller trees.

Table A-II.5. Range evolution model statistics for 11 Poaceae phylogenies. Estimated parameters: *d* – probability of anagenetic dispersal (range expansion); *e* – probability of local extinction; *j* – relative weight for founder-event speciation (dispersal to new area at cladogenesis); *x* – scaling factor for distance (*d* and *j* are multiplied by distance to power *x*). K – number of estimated parameters.

Model	Maximum range size	Number of tips	*d*	*e*	*j*	*x*	K	logL	AICc	ΔAICc	AICc weight
Oryzoideae											
DEC+*j*+*x*	8	54	0.152	0.023	0.295	-1.977	4	-228.96	466.74	0	0.78
DEC+*x*			0.197	0.044	-	-2.023	3	-231.4	469.27	2.53	0.22
DEC+*j*			0.016	<0.001	0.033	-	3	-249.85	506.19	39.45	<0.001
DEC			0.021	0.029	-	-	2	-253.67	511.57	44.83	<0.001
Bambusoideae											
DEC+*j*+*x*	3	167	0.07	<0.001	0.021	-1.96	4	-231.88	472.02	0	0.61
DEC+*x*			0.08	<0.001	-	-2	3	-233.39	472.93	0.91	0.39
DEC+*j*			0.007	<0.001	0.003	-	3	-256.65	519.44	47.43	<0.001
DEC			0.008	<0.001	-	-	2	-258.89	521.84	49.83	<0.001
Brachypodieae											
DEC+*x*	4	13	0.064	<0.001	-	-1.472	3	-43.62	95.9	0	0.74
DEC			0.017	0.011	-	-	2	-47.11	99.43	3.53	0.13
DEC+*j*+*x*			0.062	<0.001	0.026	-1.45	4	-43.49	99.98	4.08	0.1
DEC+*j*			0.015	<0.001	0.012	0	3	-46.73	102.13	6.23	0.03
Poeae											
DEC+*j*+*x*	6	389	0.074	<0.001	0.019	-1.333	4	-1116.35	2240.81	0	1
DEC+*x*			0.08	<0.001	-	-1.35	3	-1123.19	2252.44	11.62	0
DEC+*j*			0.018	<0.001	0.005	-	3	-1168.91	2343.88	103.07	<0.001
DEC			0.019	<0.001	-	-	2	-1176.69	2357.42	116.61	<0.001
Aristidoideae											
DEC+*j*+*x*	8	74	0.077	<0.001	0.028	-1.41	4	-241	490.58	0	0.52
DEC+*x*			0.085	0.007	-	-1.445	3	-242.21	490.76	0.18	0.48
DEC+*j*			0.019	<0.001	0.006	-	3	-258.22	522.78	32.2	<0.001
DEC			0.02	0.005	-	-	2	-259.33	522.83	32.25	<0.001
Panicoideae											
DEC+*j*+*x*	6	57	0.096	<0.001	0.697	-2.333	4	-208.92	426.62	0	0.93
DEC+*x*			0.128	0.01	-	-2.353	3	-212.64	431.73	5.12	0.07
DEC+*j*			0.009	<0.001	0.025	-	3	-258.06	522.57	95.95	<0.001
DEC			0.01	0.006	-	-	2	-260.24	524.71	98.09	<0.001
Paniceae											
DEC+*j*+*x*	8	512	0.134	0.004	0.025	-1.484	4	-2075.45	4158.99	0	0.75
DEC+*x*			0.142	0.012	-	-1.495	3	-2077.58	4161.2	2.22	0.25
DEC+*j*			0.03	0.002	0.006	-	3	-2249.19	4504.44	345.45	<0.001
DEC			0.032	0.013	-	-	2	-2251.46	4506.95	347.96	<0.001
Andropogoneae–Arundinelleae											
DEC+*x*	8	145	0.469	0.027	-	-2.16	3	-614.1	1234.37	0	0.74
DEC+*j*+*x*			0.464	0.025	0.016	-2.156	4	-614.08	1236.45	2.08	0.26
DEC			0.053	0.028	-	-	2	-719.3	1442.69	208.32	<0.001
DEC+*j*			0.053	0.028	<0.001	-	3	-719.3	1444.78	210.41	<0.001
Arundinoideae–Micraroideae											
DEC+*x*	7	43	0.045	<0.001	-	-1.675	3	-118.74	244.1	0	0.77
DEC+*j*+*x*			0.047	<0.001	<0.001	-1.712	4	-118.74	246.53	2.43	0.23
DEC			0.008	<0.001	-	-	2	-129.71	263.71	19.61	<0.001
DEC+*j*			0.008	<0.001	<0.001	-	3	-129.71	266.03	21.93	<0.001
Chloridoideae											
DEC+*j*+*x*	8	371	0.141	<0.001	0.177	-1.884	4	-1298.86	2605.84	0	1

Model	Maximum range size	Number of tips	*d*	*e*	*j*	*x*	K	logL	AICc	ΔAICc	AICc weight
DEC+*x*			0.165	0.009	-	-1.93	3	-1317	2648.07	42.23	<0.001
DEC+*j*			0.02	<0.001	0.014	-	3	-1503.93	3013.92	408.08	<0.001
DEC			0.021	0.005	-	-	2	-1519.55	3043.13	437.3	<0.001
Danthonioideae											
DEC+*j*+*x*	4	99	0.005	<0.001	0.022	-0.614	4	-129.62	267.66	0	0.52
DEC+*j*			0.003	<0.001	0.011	-	3	-130.79	267.84	0.18	0.48
DEC			0.005	<0.001	-	-	2	-142.71	289.54	21.88	<0.001
DEC+*x*			0.009	<0.001	-	-0.525	3	-141.77	289.79	22.13	<0.001

We obtained the ages of clades estimated as occurring in the Madagascar region with marginal probability >0.5 under the best-performing model for each phylogeny. As an estimate of the relative historical importance of each defined area as source area for dispersal to Madagascar, we summed the relative likelihoods for nodes that were direct ancestors to a node estimated present in the Madagascar region. This sum thus incorporates both the number of clades for which dispersal is likely, and the likelihood of each area being the ancestral area of those clades. We then tested for correlation of these estimates with distance to Madagascar. To compare estimated historical dispersal with current distribution patterns, we also correlated the occurence counts of species native but not endemic to Madagascar (excluding other West Indian Ocean islands, based on checklist from Vorontsova et al., 2016) in each of the defined areas, as per WCSP, with distance to Madagascar. Finally, we traced the accumulation of C_3 and C_4 grass lineages in the Madagascar region with a lineage-through-time analysis: using MCC trees with median ages under external age calibration, we counted the number of C_3 and C_4 lineages in each 1 Myr time bin and weighted them by their marginal probability (under the best range evolution model) to occur in the Madagascar region. Multiple accessions per species were counted as a singe lineage when monophyletic. In the case of multiple accessions that did not form a single clade, we only counted the accession collected in the Madagascar region.

Table A-II.6. Lineages with an estimated origin in the Madagascar region, recovered through phylogenetic analysis and range evolution analysis. The median stem age given is under external calibration. The numbers of included strictly endemic lineages refer to Table A-II.3.

Clade	Genus/Genera	No. of tips	No. of non-endemic tips	Photosynthetic pathway	Median stem age (Ma)		Endemic lineage numbers
					Cal. 1	Cal. 2	
Oryzoideae	*Humbertochloa*	1	0	C_3	3.19	5.94	3
Oryzoideae	*Leersia*	2	0	C_3	1.40	2.52	1
Oryzoideae	*Maltebrunia*	1	0	C_3	2.82	5.06	2
Bambusoideae	*Arundinaria*	3	0	C_3	2.65	3.11	5
Bambusoideae	*Decaryochloa, Hickelia, Hitchcockella, Nastus, Perrierbambus, Sirochloa, Sokinochloa, Valiha*	18	1	C_3	9.08	10.73	4, 6, 7
Brachypodieae	*Brachypodium*	1	0	C_3	2.58	3.22	8
Brachypodieae	*Brachypodium*	3	0	C_3	5.07	6.06	9

Clade	Genus/Genera	No. of tips	No. of non-endemic tips	Photosynthetic pathway	Median stem age (Ma)		Endemic lineage numbers
					Cal. 1	Cal. 2	
Poeae	*Agrostis*	5	1	C_3	4.91	5.99	11, 13, 16
Poeae	*Anthoxanthum*	1	0	C_3	2.14	2.53	15
Poeae	*Festuca*	1	0	C_3	0.64	0.77	10
Poeae	*Festuca*	3	0	C_3	2.96	3.58	14
Poeae	*Poa*	1	0	C_3	1.44	1.62	12
Poeae	*Pseudobromus*	1	0	C_3	5.44	6.73	17
Aristidoideae	*Aristida*	1	0	C_4	1.37	1.69	20
Aristidoideae	*Aristida*	2	0	C_4	12.49	17.13	22
Aristidoideae	*Aristida*	6	0	C_4	8.41	10.54	21
Aristidoideae	*Sartidia*	1	0	C_3	1.73	2.32	18
Aristidoideae	*Sartidia*	2	0	C_3	3.29	4.04	19
Panicoideae	*Lecomtella*	1	0	C_3	22.34	28.27	23
Panicoideae	*Tristachya*	3	0	C_4	5.23	6.47	24
Paniceae	*Acroceras, Brachiaria, Chasechloa, Cyphochlaena, Cyrtococcum, Lasiacis, Oplismenus, Panicum, Poecilostachys, Pseudechinolaena, Pseudolasiacis*	79	35	C_3	16.58	19.84	26, 27, 28, 29, 31, 33, 34, 35, 36, 39, 40, 41, 42, 43, 56
Paniceae	*Brachiaria*	2	0	C_4	5.70	7.46	57
Paniceae	*Brachiaria*	3	1	C_4	6.55	9.05	46, 54
Paniceae	*Brachiaria, Yvesia*	3	1	C_4	8.20	10.38	48, 58
Paniceae	*Cenchrus*	1	0	C_4	2.81	3.31	52
Paniceae	*Cenchrus*	1	0	C_4	1.77	2.47	49
Paniceae	*Digitaria*	1	0	C_4	2.98	3.76	53
Paniceae	*Digitaria*	1	0	C_4	2.24	2.74	50
Paniceae	*Eriochloa*	1	0	C_4	0.34	0.45	45
Paniceae	*Panicum*	1	0	C_4	0.90	1.08	47
Paniceae	*Panicum*	1	0	C_3	5.77	7.72	37
Paniceae	*Panicum*	2	0	C_3	2.75	3.44	32
Paniceae	*Panicum*	8	0	C_4	10.30	12.60	61
Paniceae	*Panicum, Paratheria, Setaria*	5	2	C_4	12.52	15.28	62
Paniceae	*Sacciolepis*	1	0	C_3	3.03	3.76	30
Paniceae	*Setaria*	1	0	C_4	1.79	2.17	51
Paniceae	*Setaria, Stenotaphrum*	5	3	C_4	8.58	10.31	55, 59
Paniceae	*Stenotaphrum*	1	0	C_4	7.16	8.04	60
Paniceae	*Thuarea*	1	0	C_4	0.28	0.35	44
Andropogoneae –Arundinelleae	*Andropogon*	1	0	C_4	1.43	0.42	66
Andropogoneae –Arundinelleae	*Andropogon*	1	0	C_4	2.27	2.83	69
Andropogoneae –Arundinelleae	*Andropogon*	1	0	C_4	0.34	1.77	64
Andropogoneae –Arundinelleae	*Arthraxon*	1	0	C_4	1.10	1.39	67
Andropogoneae	*Chrysopogon*	1	0	C_4	0.27	0.32	63

Clade	Genus/Genera	No. of tips	No. of non-endemic tips	Photosynthetic pathway	Median stem age (Ma)		Endemic lineage numbers
					Cal. 1	Cal. 2	
–Arundinelleae							
Andropogoneae –Arundinelleae	*Dimeria*	1	0	C_4	5.10	6.15	72
Andropogoneae –Arundinelleae	*Elionurus*	1	0	C_4	0.93	1.16	65
Andropogoneae –Arundinelleae	*Ischaemum*	2	0	C_4	1.72	2.09	68
Andropogoneae –Arundinelleae	*Lasiorrhachis*	2	0	C_4	3.07	3.86	71
Andropogoneae –Arundinelleae	*Lasiorrhachis*	2	0	C_4	2.62	3.28	70
Arundinoideae–Micrairoideae	*Isachne*	4	1	C_3	5.77	6.95	73, 74
Arundinoideae–Micrairoideae	*Styppeiochloa*	1	0	C_3	7.21	8.44	75
Arundinoideae–Micrairoideae	*Styppeiochloa*	1	0	C_3	10.48	12.40	76
Chloridoideae	*Acrachne, Sclerodactylon*	3	1	C_4	9.68	12.20	91
Chloridoideae	*Chloris, Daknopholis, Neostapfiella*	5	1	C_4	7.82	9.57	82, 84, 89
Chloridoideae	*Dactyloctenium*	1	0	C_4	6.29	7.98	88
Chloridoideae	*Decaryella*	2	0	C_4	3.61	5.20	87
Chloridoideae	*Eragrostis*	1	0	C_4	4.46	5.73	86
Chloridoideae	*Eragrostis*	1	0	C_4	2.67	3.20	83
Chloridoideae	*Eragrostis*	1	0	C_4	6.86	8.24	90
Chloridoideae	*Eragrostis*	1	0	C_4	16.25	20.22	94
Chloridoideae	*Eragrostis*	1	0	C_4	0.28	0.34	77
Chloridoideae	*Eragrostis*	1	0	C_4	1.04	1.29	81
Chloridoideae	*Eragrostis*	1	0	C_4	0.61	0.77	78
Chloridoideae	*Eragrostis*	2	0	C_4	0.97	1.35	79
Chloridoideae	*Lepturus*	6	5	C_4	4.61	6.23	85
Chloridoideae	*Sporobolus*	3	0	C_4	9.21	11.74	93
Chloridoideae	*Viguierella*	1	0	C_4	6.96	9.55	92
Danthonioideae	*Merxmuellera*	2	0	C_3	1.42	1.81	95
Danthonioideae	*Pentameris*	1	0	C_3	5.64	7.46	96

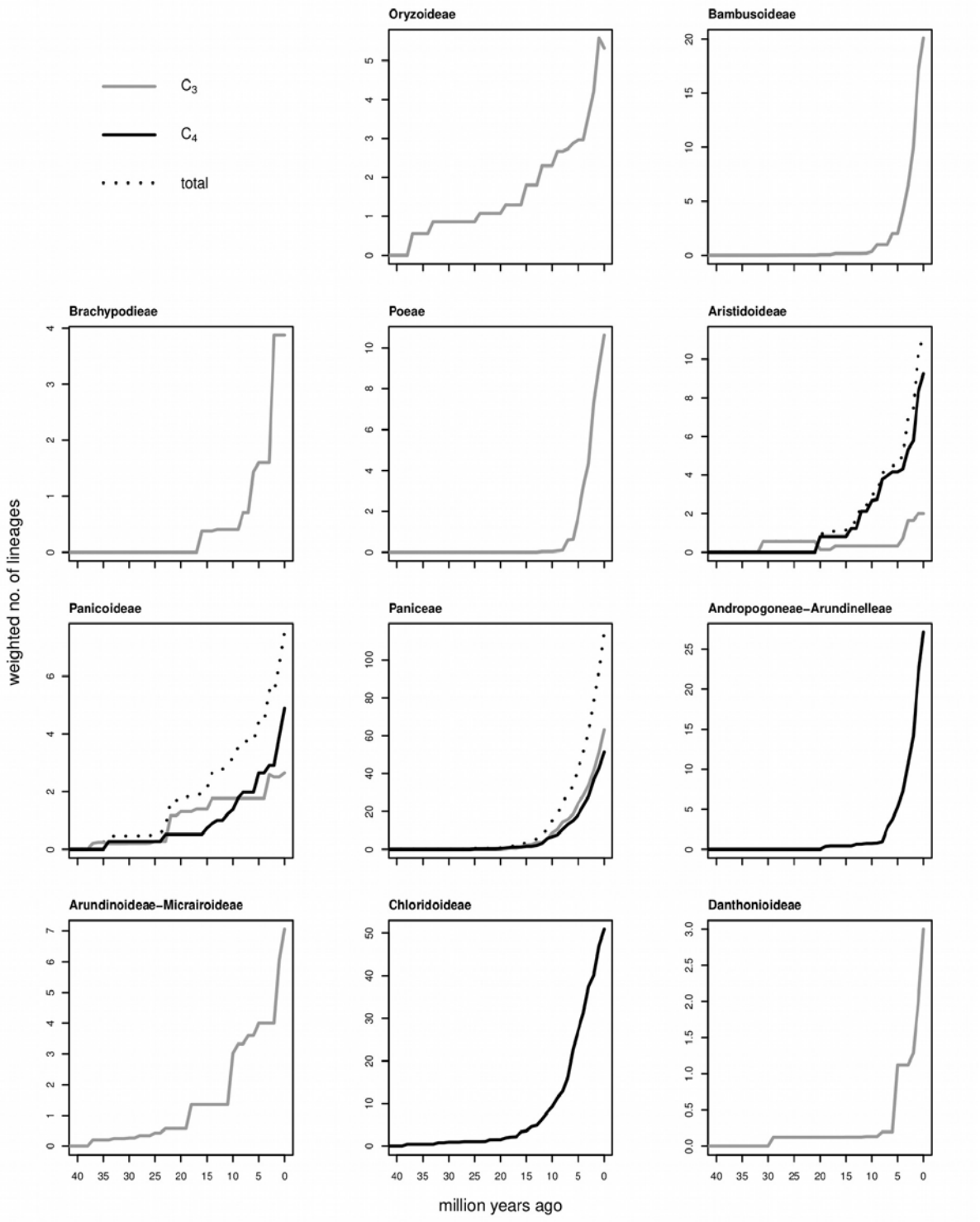

Fig. A-II.1. Grass lineage accumulation in the Madagascar region per clade, as inferred through ancestral range estimation. Lineages were counted per 1 Myr time bin (based on external calibration without phytoliths) and weighted by their marginal probability to occur in the Madagascar region. Total counts are only shown when the clade has both C_3 and C_4 lineages and their relative probabilities to occur in the region are non-negligible (i.e., >0.1 over the whole time span). Note that as relative probability weights of lineages vary through time, the accumulation curve is not necessarily monotonous.

References

Besnard G., Christin P.-A., Malé P.-J.G., Coissac E., Ralimanana H. & Vorontsova M.S. (2013): Phylogenomics and taxonomy of Lecomtelleae (Poaceae), an isolated panicoid lineage from Madagascar. *Annals of Botany*, 112: 1057–1066.

Bouckaert R., Heled J., Kühnert D., Vaughan T., Wu C.-H. et al. (2014): BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, 10: e1003537.

Christin P.-A., Spriggs E., Osborne C.P., Strömberg C.A.E., Salamin N. & Edwards E.J. (2014): Molecular dating, evolutionary rates, and the age of the grasses. *Systematic Biology*, 63: 153–165.

Clayton W.D., Govaerts R., Harman K.T., Williamson H. & Vorontsova M. (2016): *World Checklist of Poaceae, facilitated by the Royal Botanic Gardens, Kew*. Available at: http://apps.kew.org/wcsp/.

Crepet W.L. & Feldman G.D. (1991): The earliest remains of grasses in the fossil record. *American Journal of Botany*, 78: 1010–1014.

Drummond A.J., Ho S.Y.W., Phillips M.J. & Rambaut A. (2006): Relaxed phylogenetics and dating with confidence. *PLOS Biology*, 4: e88.

Dugas D.P. & Retallack G.J. (1993): Middle Miocene fossil grasses from Fort Ternan, Kenya. *Journal of Paleontology*, 67: 113–128.

GPWG II – Grass Phylogeny Working Group II (2012): New grass phylogeny resolves deep evolutionary relationships and discovers C_4 origins. *New Phytologist*, 193: 304–312.

Hijmans R.J. (2016): *geosphere: Spherical Trigonometry. R package version 1.5-5*. Available at: https://cran.r-project.org/web/packages/geosphere.

Katoh K. & Standley D.M. (2013): MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, 30: 772–780.

Kellogg E.A. (2015): *Flowering Plants. Monocots: Poaceae.* Springer, New York.

Matzke N.J. (2013): *BioGeoBEARS: BioGeography with Bayesian (and Likelihood) Evolutionary Analysis in R Scripts.* University of California, Berkeley.

Matzke N.J. (2014): Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Systematic Biology*, 63: 951–970.

Nash J.C. (2014): On best practice optimization methods in R. Journal of Statistical Software, 60, 1–14.

Miller M.A., Pfeiffer W. & Schwartz T. (2012): The CIPRES science gateway: enabling high-impact science for phylogenetics researchers with limited resources. *Proceedings of the 1st Conference of the Extreme Science and Engineering Discovery Environment: Bridging from the eXtreme to the campus and beyond*, 39.

Osborne C.P., Salomaa A., Kluyver T.A., Visser V., Kellogg E.A. et al. (2014): A global database of C_4 photosynthesis in grasses. *New Phytologist*, 204: 441–446.

Rambaut A., Suchard M.A., Xie D. & Drummond A. (2014): *Tracer v1.6.* Available at: http://beast.bio.ed.ac.uk/Tracer.

Ree R.H., & Smith S.A. (2008) Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology*, 57: 4–14.

Soreng R.J., Peterson P.M., Romaschenko K., Davidse G., Zuloaga F.O. et al. (2015): A worldwide phylogenetic classification of the Poaceae (Gramineae). *Journal of Systematics and Evolution*, 53: 117–137.

Stamatakis, A. (2014): RAxML Version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30: 1312-1313.

Taylor S.H., Franks P.J., Hulme S.P., Spriggs E., Christin P.A. et al. (2012): Photosynthetic pathway and ecological adaptation explain stomatal trait diversity amongst grasses. *New Phytologist*, 193: 387–396.

Van Dam M.H. & Matzke N.J. (2016): Evaluating the influence of connectivity and distance on biogeographical patterns in the south-western deserts of North America. *Journal of Biogeography*, 43: 1514–1532.

Vorontsova M.S., Besnard G., Forest F., Malakasi P., Moat J. et al. (2016) Madagascar's grasses and grasslands: anthropogenic or natural? *Proceedings of the Royal Society B*, 283, 20152262.

7.3 A-III (chapter III: Phylogenomics and biogeography of the core Panicoideae)

Table A-III.1. Voucher names and *GenBank* accession numbers for plastome sequences used in this study, including new data.

Species	Voucher/isolate	Accession no.	Pathway	New data
Amphicarpum muhlenbergianum	ISC:Clark et al. 1695	KU291489.1	C_3	
Andropogon burmanicus	BKF:Traiperm 572	KY596167.1	C_4	
Anomochloa marantoidea	ISC:L. Clark 1299	NC_014062.1	C_3	
Aristida purpurea	N/A	NC_025228.1	C_4	
Arthraxon microphyllus	BKF:Traiperm 537	KY596183.1	C_4	
Arthraxon prionodes	MO:Kellogg PI 659331	KY596138.1	C_4	
Arundinella deppeana	XAL:Clark et al. 1680	KU291490.1	C_4	
Arundinella nepalensis	K:MSV608		C_4	•
Axonopus fissifolius	ISC:Clark et al. 1703	KU291491.1	C_4	
Bambusa multiplex	N/A	NC_024668.1	C_3	
Bothriochloa alta	DEK:Duvall s.n.	NC_030621.1	C_4	
Capillipedium venustum	PI:11713	NC_030622.1	C_4	
Cenchrus americanus	N/A	NC_024171.1	C_4	
Centotheca lappacea	N/A	NC_025229.1	C_3	
Chandrasekharania keralensis	K:VSR54064 (isotype)		C_3	•
Chasmanthium laxum subsp. *sessiliflorum*	ISC:Sanchez-Ken s.n.	KU291494.1	C_3	
Chrysopogon serrulatus	MO:219580	NC_029884.1	C_4	
Coix lacryma-jobi	DEK:Duvall s.n.	NC_013273.1	C_4	
Coleataenia prionitis	N/A	NC_025231.1	C_3	
Danthonia californica	N/A	NC_025232.1	C_3	
Danthoniopsis dinteri	SRR2163566	MF035977.1	C_4	
Danthoniopsis stocksii	M:RHR54967		C_4	•
Dichanthelium acuminatum	CAN:Saarela 666	KU291496.1	C_4	
Digitaria exilis	N/A	NC_024176.1	C_4	
Digitaria glauca	K:MSV950		C_4	•
Diheteropogon amplectens var. *catangensis*	PI:12585	KU291497.1	C_4	
Echinochloa stagnina	K:RCH49	MF563381.1	C_4	
Eriachne stipacea	N/A	NC_025234.1	C_4	
Eriochloa meyeriana	DEK:Duvall s.n.	KU291498.1	C_4	
Eriochrysis laxa	ICN:Welker 489	NC_029883	C_4	
Eulalia aurea	PI:12153	NC_030503.1	C_4	
Garnotia tenella	BKF:Traiperm 552	KY596184.1	C_4	
Garnotia thailandica	BKF:Traiperm 535	KY596171.1	C_4	
Glyphochloa forficulata	K:HFP896		C_4	•
Gynerium sagittatum	K:P6-301b		C_3	•
Homolepis aturensis	P:GB06-2012	MF563378.1	C_3	
Hyparrhenia subplumosa	PI:12665	NC_030625.1	C_4	
Ichnanthus pallens	GB06-2014	MF563377.1	C_3	
Imperata cylindria	DEK:Burke 21	NC_030487.1	C_4	
Ischaemum afrum	PI:364924	KU291467.1	C_4	
Iseilema macratherum	PI:257760	NC_030611.1	C_4	
Jansenella griffithiana	K:CG209		C_3	•
Jansenella neglecta	K:SRY201 (isotype)		C_3	•
Kerriochloa siamensis	BKF:Traiperm 580	KY596120.1	C_4	
Lasiacis nigra	GB02-2014	MF563376.1	C_3	
Lasiurus scindicus	K:A Naegelé DJI/78-36		C_4	•

Lecomtella madagascariensis	K:MSV603	NC_024106.1	C_3	
Loudetiopsis kerstingii	PI:12679	KU291469.1	C_4	
Megathyrsus maximus	PI 12181	KU291470.1	C_4	
Melinis minutiflora	K:MSV609		C_4	•
Miscanthus sinensis	N/A	KR822688.1	C_4	
Neyraudia reynaudiana	RSA:Columbus 5302	NC_024262.1	C_4	
Oplismenus hirtellus	ISC:Clark & Lewis 1644	KU291473.1	C_3	
Oryza glaberrima	SAMN02415034	KJ513090.1	C_3	
Otachyrium versicolor	SI:Zuloaga 7027	KU291474.1	C_3	
Panicum capillare	CAN:Saarela 769	KU291475.1	C_4	
Panicum virgatum cv. *Summer*	N/A	HQ822121.1	C_4	
Paspalidium geminatum	SI:Giussani 313	KU291476.1	C_4	
Paspalum dilatatum	CAN:Peterson 19673	KU291477.1	C_4	
Paspalum fimbriatum	SI:Morrone 3651	NC_030495.1	C_4	
Paspalum glaziovii	SI:Filgueiras 3482	KU291479.1	C_4	
Pharus latifolius	MO:J. Triplett 421	NC_021372.1	C_3	
Phragmites australis	N/A	KF730315.1	C_3	
Plagiantha tenella	SI:Zuloaga 6953	KU291480.1	C_3	
Poa palustris	CAN:Saarela 1080	NC_027484.1	C_3	
Pogonatherum paniceum	MO:Clark s.n.	NC_029881	C_4	
Polytoca digitata	BKF:Arthan 054	KY596178.1	C_4	
Puelia olyriformis	MO:Clayton 1060	NC_023449.1	C_3	
Reynaudia filiformis	K:E12208		C_4	•
Rottboellia cochinchinensis	ISC:Clark et al. 1698	KU291481.1	C_4	
Saccharum arundinaceum	JW630	LC160130.1	C_4	
Saccharum sp. hybrid cv. RB72454	N/A	LN849914.1	C_4	
Setaria italica	N/A	KJ001642.1	C_4	
Sorghastrum nutans	DEK:Wysocki s.n.	NC_030498.1	C_4	
Sorghum bicolor	N/A	NC_023800.1	C_4	
Sorghum timorense	N/A	NC_023800.1	C_4	
Steinchisma laxa	SI:Zuloaga 7416	KU291483.1	C_3	
Streptostachys asperifolia	P:GB01-2012	MF563369.1	C_3	
Streptostachys ramosa	SI:Zuloaga 6960	KU291472.1	C_3	
Themeda triandra	AL94	KY707772.1	C_4	
Thyridolepis xerophila	CAN:Saarela 1643	KU291485.1	C_3	
Thysanolaena latifolia	N/A	NC_025229	C_3	
Tristachya humbertii	K:MSV1369	MF563368.1	C_4	
Urochloa reptans	HAW:Morden 1221	KU291486.1	C_4	
Whiteochloa capillipes	DEK:Duvall s.n.	KU291487.1	C_4	
Zea mays	N/A	X86563.2	C_4	
Zea nicaraguensis	N/A	KU291447.1	C_4	
Zeugites pittieri	ISC:Clark et al. 1171	KU291488.1	C_3	

Fig A-III.1 (continued on next page). Whole plastome Maximum Likelihood phylogenies inferred with *RAxML* from different alignment subsets. The respective scale is in substitutions per site; values at nodes are bootstrap support values <100. A–C, all taxa; D–F (next page), without *Gynerium sagittatum*.

Fig A-III.1 (continued).

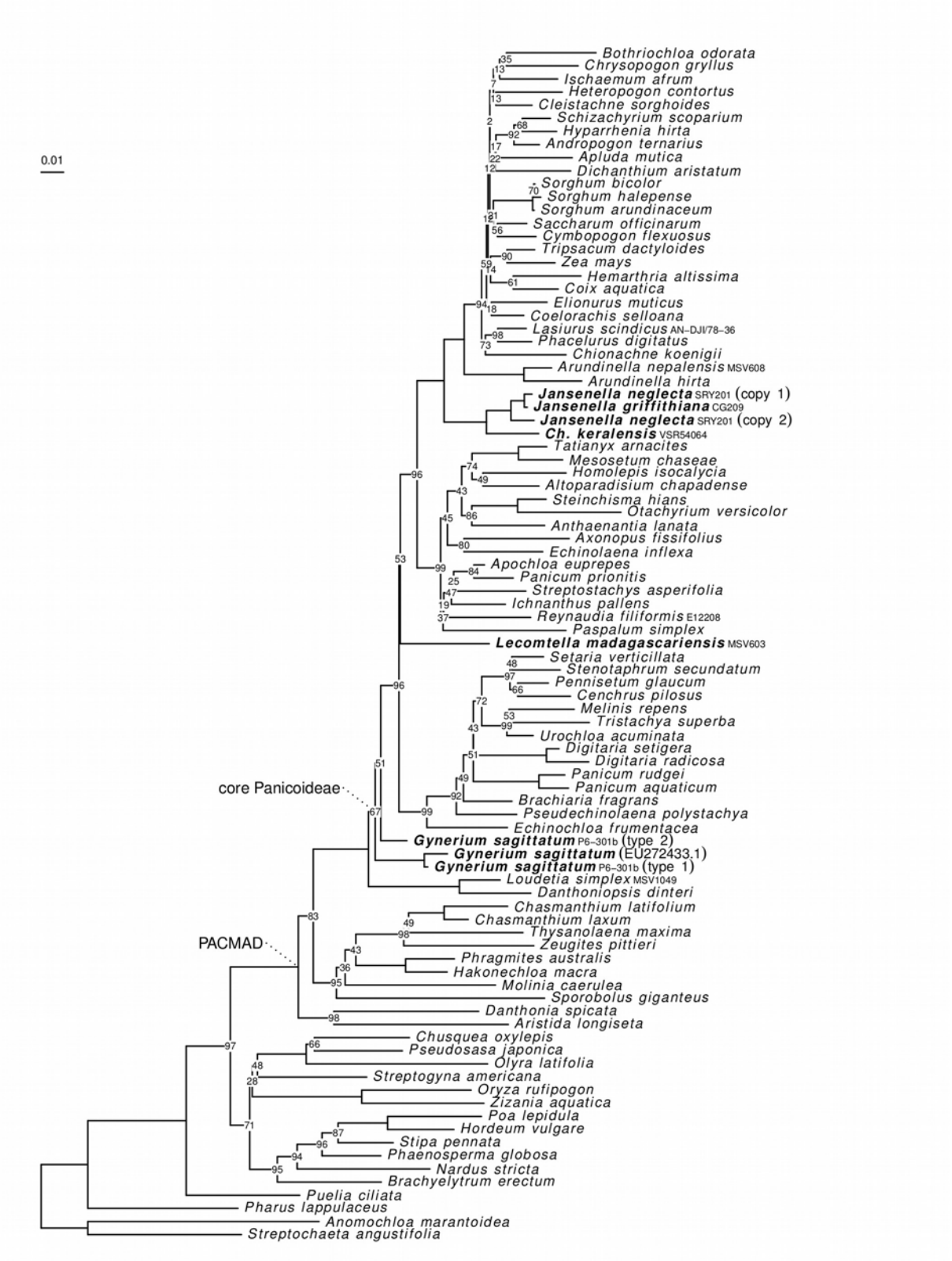

Fig A-III.2. *phyB* Maximum Likelihood phylogeny inferred with *RAxML*. The scale is in substitutions per site; values at nodes are bootstrap support values <100.

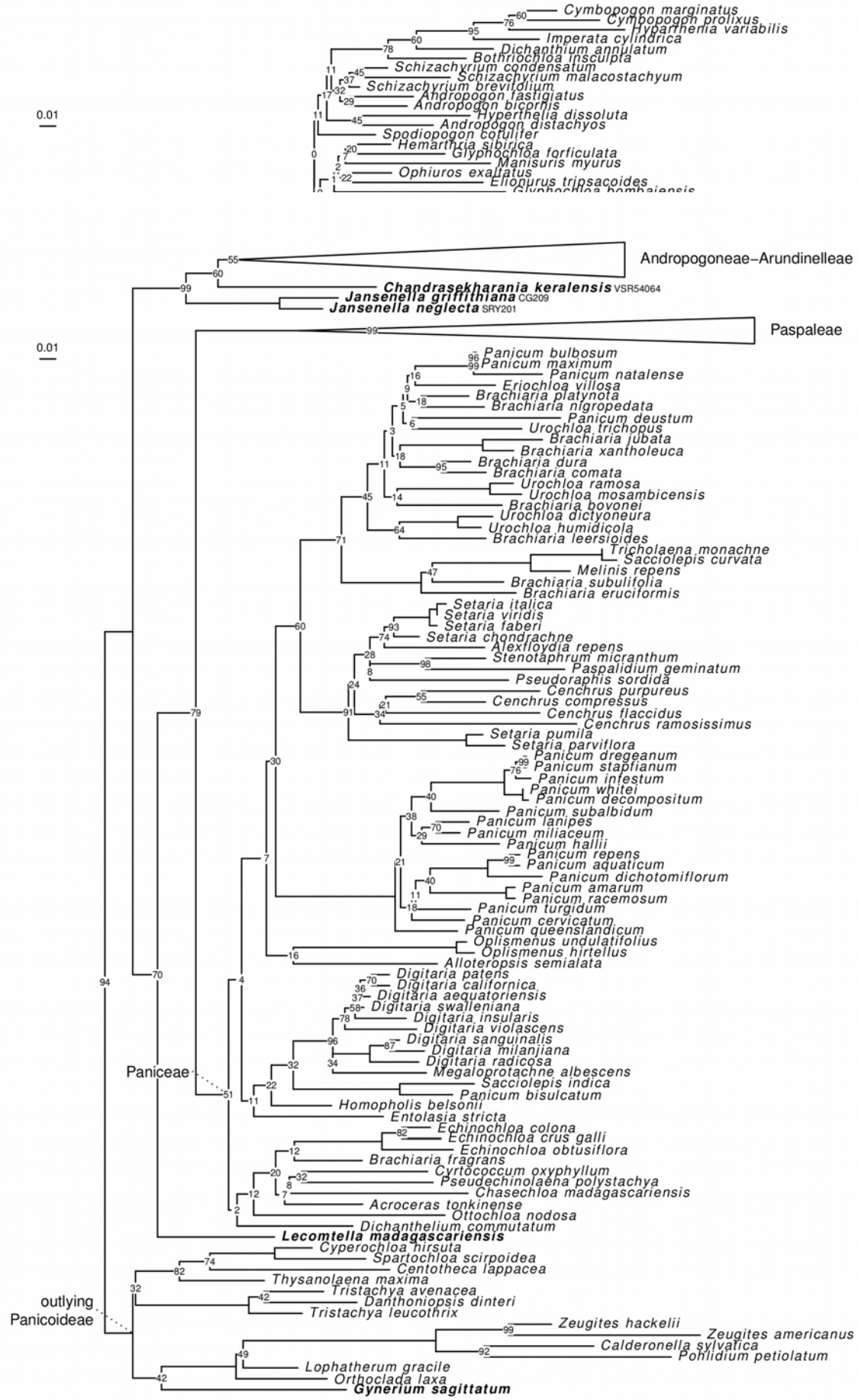

Fig A-III.3 (continued).

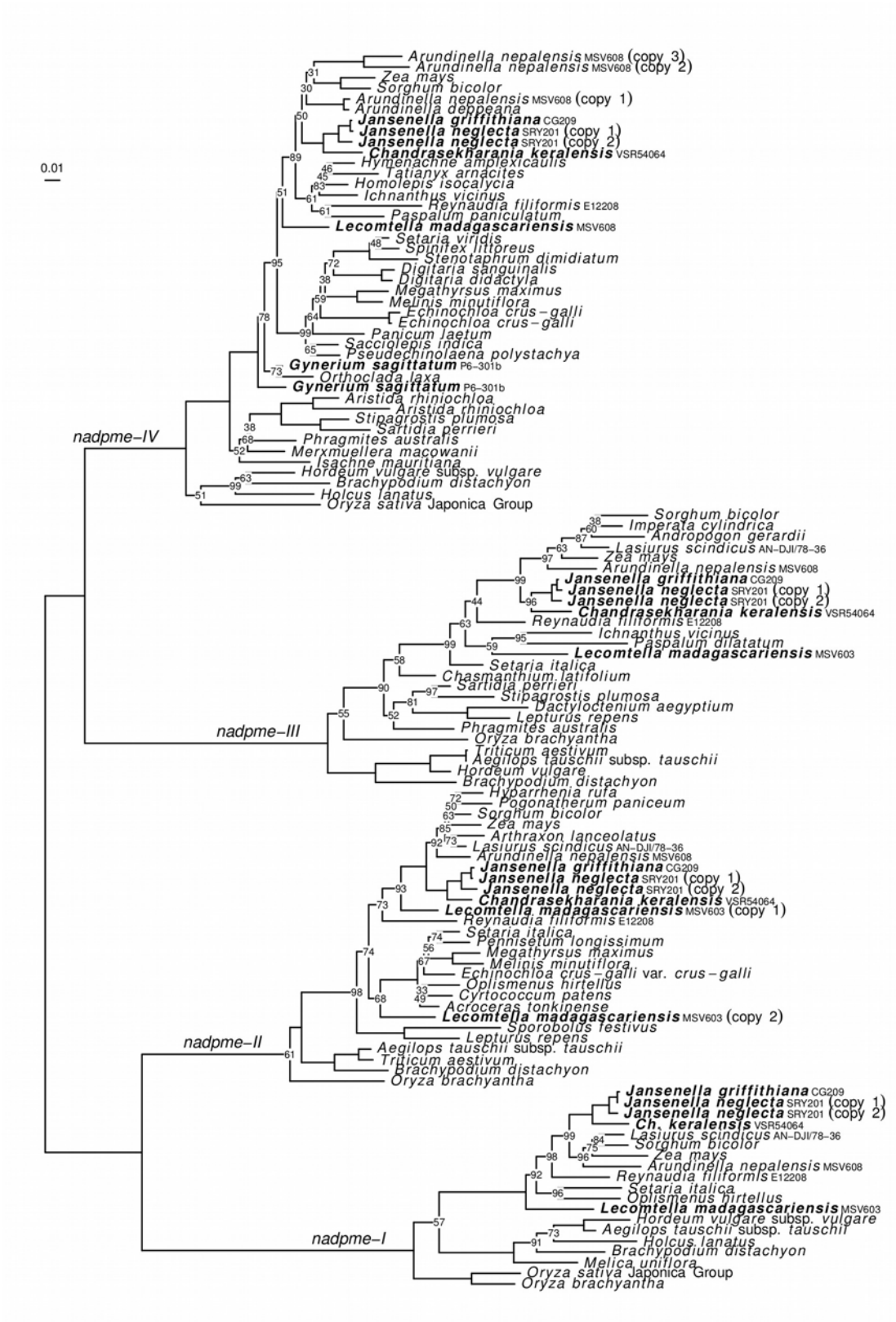

Fig A-III.4. *nadpme* Maximum Likelihood phylogeny inferred with *RAxML*. The scale is in substitutions per site; values at nodes are bootstrap support values <100.

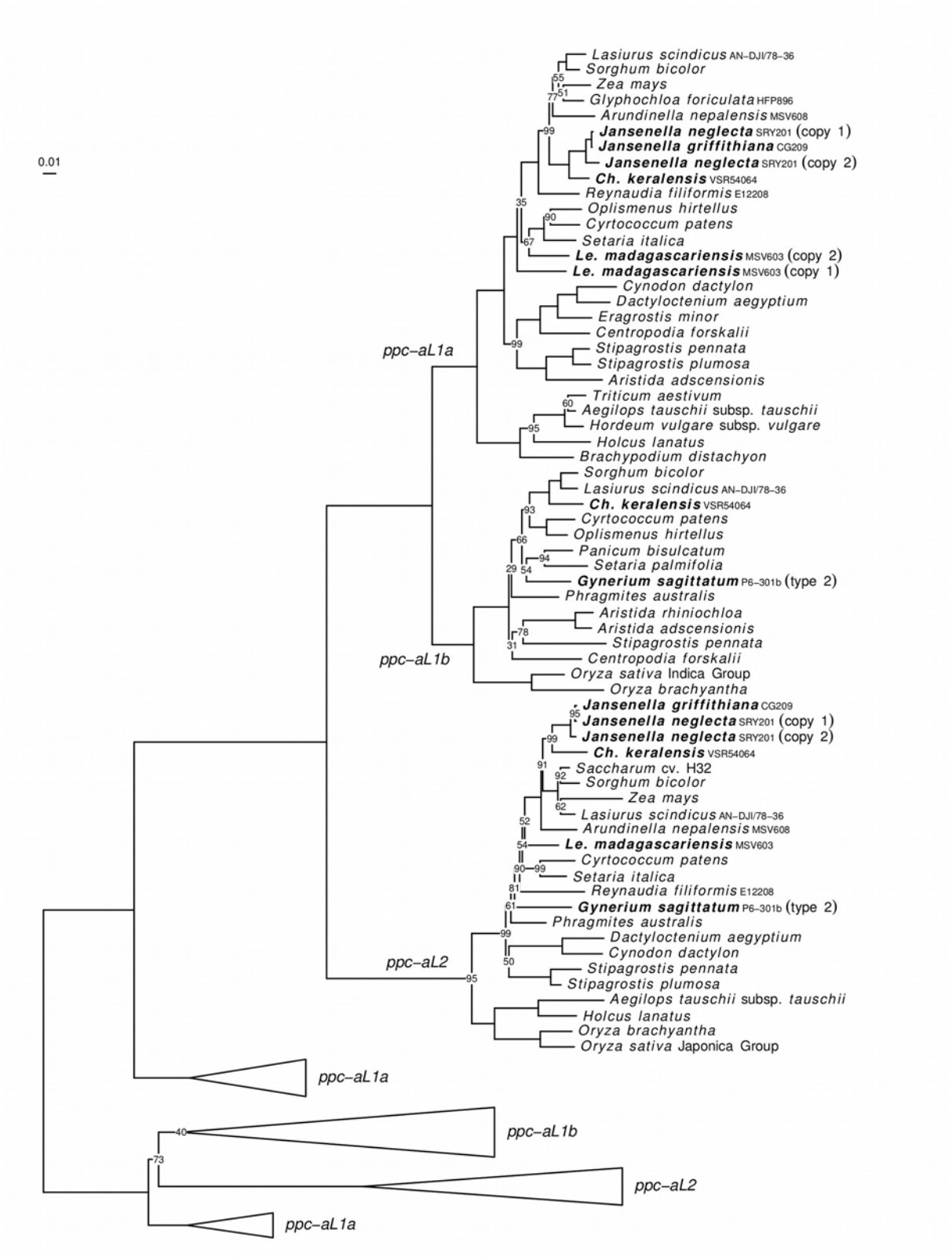

Fig A-III.5 (continued on next page). *ppc* Maximum Likelihood phylogeny inferred with *RAxML*. The scale is in substitutions per site; values at nodes are bootstrap support values <100.

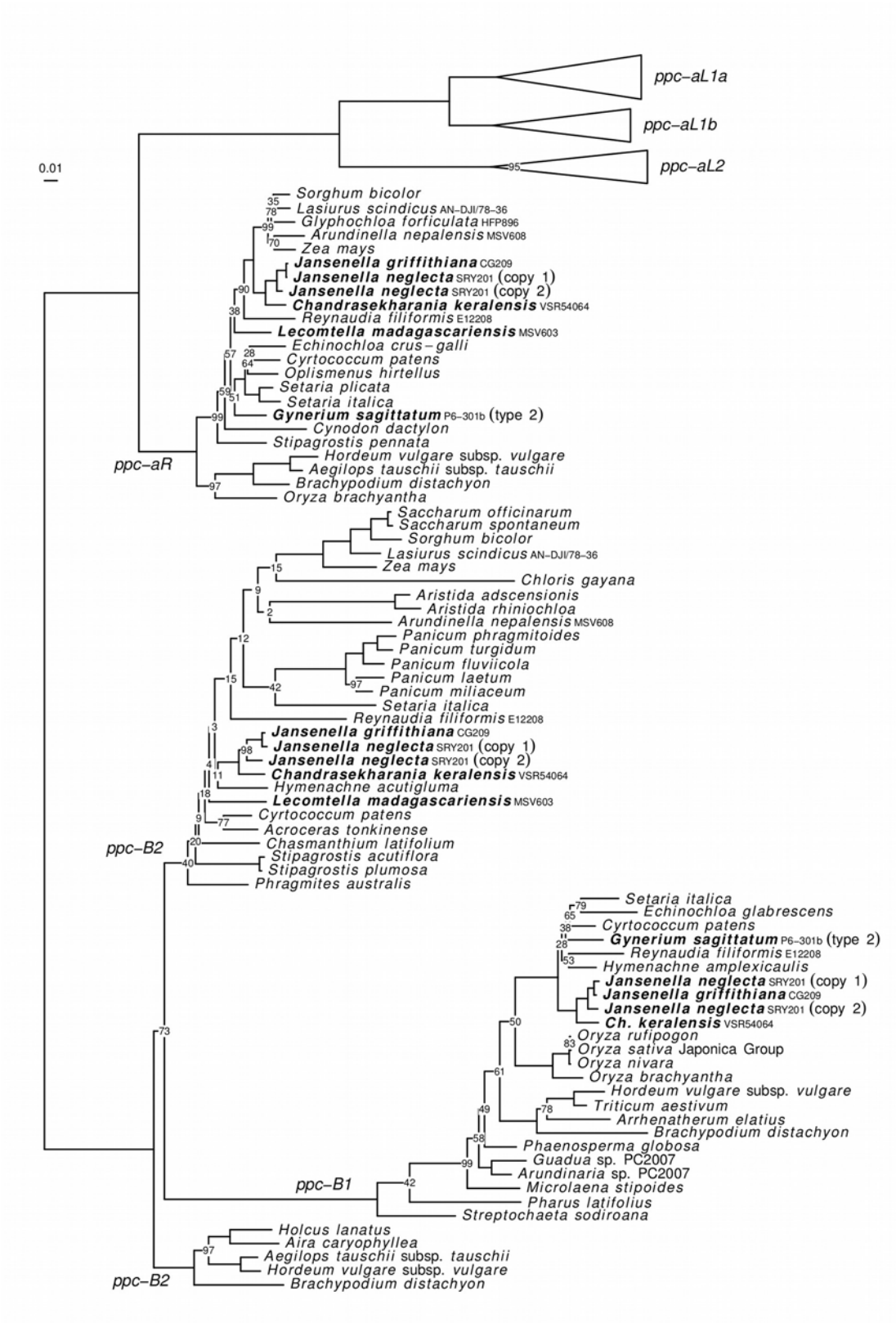

Fig A-III.5 (continued).

7.4 A-IV (chapter IV: A metabarcoding method for the analysis of grass endophyte communities in Madagascar)

Table A-IV.1. OTUs based on 97% ITS2 identity and with >25 reads per sample, identified as fungal with *ecotag* and the *UNITE* database. Only OTUs present in more than one sample (and a conspicuously abundant OTU, no. 17) were additionally submitted to *BLAST* searches against *GenBank*.

OTU	No. of samples	*UNITE* taxon	No. of reads	*GenBank* taxon	Potential ecology	Best match	Identity
1	12		1379	*Phoma/Epicoccum*	pathogen/endophyte	KT310093	100
2	6	Capnodiales	486	*Toxicocladosporium*	saprophyte/endophyte (tropical)	KF777190	99.7
3	5		552	*Cladosporium* sp.	saprophyte/pathogen	KU986780	100
4	5		476	?		MF569901	97.1
5	5		593	?		MF568876	95.3
6	4		334	*Ramichloridium* sp.	saprophyte/pathogen	KM357331	98.7
7	3	Capnodiales	241	Capnodiales		KY228722	97.3
8	3	Hypocreales: Clavicipitaceae	181	*Epichloë* sp.	endophyte	KC881085	97.3
9	3		361	*Septoria* sp.	pathogen	MF569709	97.6
10	3	Capnodiales	143	?		KF435160	98
11	2		359	?		KJ869155	96.7
12	2		169	Sordariomycetes		EU002898	97.7
13	2		229	?		JF951143	93
14	2		358	?		MF569334	94
15	2		76	*Arthrobotrys/Microdochium*	pathogen	KP689198	100
16	2		115	?		KF436353	95.7
17	1		4766	*Festuca* (grass)			
18	1	Hypocreales: Clavicipitaceae	367				
19	1		836				
20	1		228				
21	1		207				
22	1		182				
23	1		187				
24	1	Pleosporales: *Paraphaeosphaeria michotii*	162				
25	1		97				
26	1		70				
27	1		205				
28	1		70				
29	1		70				
30	1		60				
31	1		72				
32	1		57				
33	1		85				
34	1		108				
35	1		236				
36	1		29				
37	1		47				
38	1		27				

AUTHOR: Jan HACKEL

TITLE: Molecular biogeography of grasses and tropical grasslands

ABSTRACT:

Grasses (Poaceae) are a large, cosmopolitan plant family. In this dissertation, I used molecular methods to study their biogeographic history. The first chapter focuses on determinants of lineage dispersal in the temperate subtribe Loliinae, with distance found to be the dominant factor. The second chapter analyses the origins of Madagascar's grass flora. Two large *in situ* radiations of C_3 grasses were found while C_4 grasses immigrated more frequently and support the pre-human presence of grasslands in Madagascar. The third chapter resolves relationships of an Asian C_3 lineage using phylogenomic methods, with implications for C_4 photosynthesis evolution and the assembly of tropical grasslands. The fourth and final chapter developed a metabarcoding method for the analysis of fungal endophyte communities associated to grasses in Madagascar, with results highlighting methodological limitations.

KEYWORDS :

Biogeography, C_4 photosynthesis, dispersal, grasslands, phylogeny, Poaceae

AUTEUR : Jan HACKEL

TITRE : Molecular biogeography of grasses and tropical grasslands

DIRECTEUR DE THESE : Guillaume BESNARD

LIEU ET DATE DE SOUTENANCE : Toulouse, 13/12/2017

RÉSUMÉ :

Les graminées (Poaceae) sont une très grande famille cosmopolite de plantes. Dans la thèse présentée ici, j'ai utilisé des méthodes moléculaires pour analyser leur histoire biogéographique. Le premier chapitre s'intéresse aux déterminants de la dispersion dans la sous-tribu tempérée Loliinae. La distance apparaît comme le facteur dominant. Le deuxième chapitre présente une analyse des origines des graminées de Madagascar. Les résultats montrent deux grandes radiations *in situ* de graminées C_3 et une immigration répétée de graminées C_4, soutenant l'existence de savanes malgaches avant la présence humaine. Le troisième chapitre résout la position d'une lignée C_3 d'Asie à l'aide de la phylogénomique, avec des implications pour l'évolution de la photosynthèse C_4 et des savanes tropicales. Le quatrième chapitre présente une méthode de méta-barcoding pour l'analyse des communautés d'endophytes fongiques associées aux graminées à Madagascar, démontrant des limitations méthodologiques.

MOTS-CLES :

Biogéographie, dispersion, phylogénie, photosynthèse C_4, savanes, Poaceae

DISCIPLINE ADMINISTRATIVE : Écologie – Évolution

INTITULÉ ET ADRESSE DU LABORATOIRE :

Laboratoire Évolution & Diversité Biologique
UMR 5174
Université Toulouse III – Paul Sabatier
118 route de Narbonne
31062 Toulouse cedex 9