# THE INTEGRATION OF PHYLOGENETIC AND FUNCTIONAL TRAIT INFORMATION INTO MONOCOT BIODIVERSITY CONSERVATION

Chris McOwen

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews

2012



Identifiers to use to cite or link to this thesis:

DOI: https://doi.org/10.17630/10023-3659
http://hdl.handle.net/10023/3659

# The Integration of Phylogenetic and Functional Trait Information into Monocot Biodiversity Conservation

**Chris Mcowen**

This thesis is submitted in partial fulfilment for the degree of Ph.D.

at the

University of St Andrews

**1. Candidate's declarations:**

I, Christopher James Mcowen, hereby certify that this thesis, which is approximately 28000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in September 2008 and as a candidate for the degree of PhD in September 2009; the higher study for which this is a record was carried out in the University of St Andrews between 2008 and 2011.

Date: 26/10/2011

Signature of candidate:

**2. Supervisor's declaration:**

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date: 26/10/2011

Signature of supervisor:

**3. Permission for electronic publication:**

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the electronic publication of this thesis:

Access to all of printed copy but embargo of all of electronic publication of thesis for a period of two years on the following ground:

Publication would preclude future publication

Date: 26/10/2011

Signature of candidate:

Signature of supervisor:

# Acknowledgements

The biggest thank you goes to Tom Meagher – not only was he a very supportive and enthusiastic supervisor with high standards, but, his dry sense of humour and wit lightened the beginning of many emails and meetings. I also thank my second supervisor, Neil Brummitt, for a number of very productive and helpful meetings during my trips to London and his dedication to the cause, spending many late evenings in the library searching through piles of books looking for trait and genetic information. I would also like to thank the team at Kew - Felix Forest for his invaluable phylogenetic advice, particularly in dating the tree and patiently helping me develop my final method. Eimear Nic Lughadha, for putting up with my minimal knowledge of plants, and in helping me view my results in the bigger picture and through a botanical lens. I am also grateful to Steve Bachman, and the people who gave their time and data to do the IUCN global assessments. I am also grateful to Malin Rivers for her critical eye and willingness to discuss my project over coffee or lunch. I would also like to thank Rich Grenyer, David Orme and Nick Isaac for code and R advice, a thank you that also goes out to the Silwood R mailing list for their troubleshooting tips. Finally, I would like to thank BBSRC for the studentship and the opportunity to undertake this research.

I would also to thank my family; in particular Jaya for not only desperately trying to improve my grammar, but also offering me her full support and patience, not only putting up with how stressed I was, but also marrying me in the midst of it! A huge thanks goes to my Mum for quite simple getting me where I am today. Also to Damion, Becky and Gran for instilling in me a love of nature and answering all my questions as a child. Thanks also goes out to my extended London family – Tamsin, Moo and John for their extreme hospitability in giving me a place to sleep when visiting and especially to Richard not only for his all round kindness but also his inexhaustible patience and enthusiasm in discussing my latest brainwave.

# Abstract

It is now widely acknowledged that a high proportion of the world's species are threatened with global extinction in the near future. Conserving all endangered species is impractical due to limited financial resources and manpower. Therefore, there is a real need to understand the mechanisms behind the extinction process in order to develop a *proactive* rather than *reactive* conservation strategy. Furthermore, given these limitations there is a need to be selective, prioritising species or areas not just on the basis of extinction risk or the number of species present, but with consideration for their role within the ecosystem. In this thesis, I investigate three aspects of current global extinction: What causes species to become threatened, how we should prioritise those that are, and how such information can be translated into area-based conservation priorities.

Firstly, I developed models of plant extinction risk in relation to environmental (anthropogenic, climatic, and physical) variables and used them to show that non-randomness in extinction risk is highly influenced by interactions between the plants traits and their environment. Incorporating such interactions, my models were able to explain ~30% of the variation in plant extinction risk. Results from this large-scale comparative study suggest the biological traits that increase species' susceptibility are dependent upon local environmental conditions.

Secondly, I looked at how much unique evolutionary history, a proxy for species value, will be lost under current extinction patterns, whilst investigating how effective current conservation methods are at prioritising species based on their evolutionary value. The results are not promising; species threatened with extinction had disproportionately high quantities of unique evolutionary history. Furthermore, the IUCN Red List makes no allowance for this, for example, species considered Critically Endangered, the highest threat status under the IUCN Red List, may not necessarily be the most evolutionary unique.

Finally, I looked at the relationships among species richness, versity (FD) and phylogenetic diversity (PD) as measures of biodiversity and therefore the conservation value of an area. It was found species richness is generally a good surrogate for FD and PD. However, agricultural development and climatic variation bring out discrepancies between species richness, FD and PD, challenging the claim of interchangeability of different diversity measures, with potential consequences for conservation planning.

# Table of contents

# List of figures

# List of Tables

# Supplementary material on accompanying CD

| | |
|---|---|
| **SUPPLEMENTARY MATERIAL 1.** | Genetic Sequences used |
| **SUPPLEMENTARY MATERIAL 2.** | Summary of data availability |
| **SUPPLEMENTARY MATERIAL 3**. | Phylogenetic tree with bootstrap support |
| **SUPPLEMENTARY MATERIAL 4**. | Traits used in analyses |
| **SUPPLEMENTARY MATERIAL 5.** | Sources of trait data |
| **SUPPLEMENTARY MATERIAL 6**. | Species composition of each eco-region |
| **SUPPLEMENTARY MATERIAL 7**. | Eco-region variables |

# Chapter 1: Introduction

## The 6th mass extinction crisis

There is now little doubt that many populations worldwide are declining or undergoing local extinction, causing many species to be threatened with global extinction. Indeed, it has been suggested we are entering the period of the 6th mass extinction (Wake and Vredenburg 2008, Dunn *et al.* 2009). Results from the latest IUCN Red List of Threatened Species corroborates this worrying observation, with 21% of mammal, 12% of bird and 32% of amphibian species currently listed as threatened with extinction, a total of 4268 threatened species just in these three taxa (IUCN 2008). However, the true extinction risk may be higher: 15% of mammals, 1% of birds and 25% of amphibians are classified as Data Deficient because insufficient data is available to evaluate their risk status (IUCN 2008). Furthermore, due to the global demands of a growing human population, the current extinction crisis looks set to worsen, likely resulting in more species becoming threatened and subsequent damaging alterations of ecosystems (Millennium Ecosystem Assessment 2005).

One group conspicuous by its absence from the above observations is plants: despite the survival of humanity and of many other species relying on plants to provide the structure of most terrestrial ecosystems (Wyse and Kennedy 2009). This absence does not mean that plants are not under threat; the latest global analysis of threatened plants indicates that 21.5% of plant species are threatened, making them more threatened than birds or mammals (Royal Botanic Gardens Kew 2010).

Furthermore, there are many ongoing environmental changes that are expected to further reduce plant biodiversity. For example, climate change models show that as

temperature increases, precipitation patterns change, desertification becomes more prevalent, sea-levels rise, growing seasons fluctuate, pollinators and seed dispersers decrease in number, and the frequency and intensity of extreme weather events, such as droughts, storms and floods, increase, plant species will be lost (Wyse and Kennedy 2009). The aim of conservation science is to inform policy and practice to counter extinction threats through the development of a scientifically rigorous methodology, firstly to understand what causes species to become extinct and secondly, with so many species facing extinction, to decide what criteria we should use to prioritise conservation efforts. My thesis will address several themes related to this central aim.

**Measuring extinction risk**

In order to protect species an understanding of how susceptible they are to extinction is a fundamental requirement. Throughout my thesis, I will use classifications from The Global IUCN Red List of Threatened Species (IUCN 2001) as estimates of species extinction risk. Assessments for the Red List are made using one or more of five quantitative criteria.

Species are assessed for population reduction (Criterion A); small geographic range size and its decline (Criterion B); small population size and its decline (Criterion C); very small or restricted population (Criterion D); predictions from a quantitative analysis regarding the time to extinction in the wild are taken into account (Criterion E).

Species are assigned to a Red List category based on an evaluation of range or population data against quantitative thresholds in the 5 Red List criteria. The criteria were devised so that listing species under any criterion for one of the threatened categories would be a probabilistic assessment of the time to extinction as follows:

Critically Endangered (CR) – at least 50% chance of extinction within 10 years or three generations (maximum of 100 years); Endangered (EN) – at least 20% chance of extinction within 20 years or five generations (maximum of 100 years); Vulnerable (VU) – at least10% chance of extinction within 100 years (IUCN 2001).However, these estimates are generalisations, and any detailed prediction should involve species-specific parameters (Mace 1995). In reality, timescales to extinction are likely to be longer than the estimates above (Redding and Mooers 2006). If species nearly, but do not quite, fulfill any of the quantitative criteria, they can be classified as Near Threatened (NT), whereas species not qualifying under any criterion are ranked as Least Concern (LC). Species with insufficient data to assign to a risk category a risk category, or with unresolved taxonomic status are designated as Data Deficient (DD).

Due to the quantitative basis of assigning a species to a category using the above criteria, and the easily comprehensible and flexible framework of assessments, the IUCN Red List can handle uncertainty and give consistent results across taxa (Akçakaya *et al.* 2000), and the criteria are designed so that a species should be assigned to the same category under different criteria, assuming sufficient data on range size and population size and its trend is known. Consequently, the IUCN Red List is the considered to the best available dataset of extinction risk estimates (Rodrigues *et al.* 2006).

High profile global Red List assessments have now been undertaken for a suite of vertebrate taxa (birds, amphibians and mammals) (IUCN 2008). However, there is a taxonomic bias to the assessments: highly diverse taxa are currently very poorly represented (IUCN 2011). One group that is poorly represented are plants. Fewer than one thousand species of plants are fully documented by the IUCN in their last report on threatened species (IUCN 2011); this is despite the fundamental role plants play in many ecosystem processes (Isbell *et al.* 2011). The limited representation of plants is primarily due to the relatively large number plant species; an estimated 352,000 for just the flowering plants, in relation to the number of botanical experts available to

assess the conservation status of each of the earth's species, and the lack of sufficient suitable, available information in order to apply the Red List Criteria for most plant species (Brummitt *et al.* 2008). The lack of information regarding the current risk of plant extinctions needs to be addressed.

The Sampled Red List Index (SRLI) was developed as an attempt to address the taxonomic bias in the coverage of IUCN assessments, based on a random subset of species for groups to be assessed (Butchart *et al.* 2005, Baillie *et al.* 2008).
In order to ensure a representative taxonomic and geographic distribution species were randomly selected, without re- placement, 1,500 species from each group, 50,000 times. On average, the proportion of species in each threat class (threatened, non-threatened, data deficient), Order and biogeographic realm, was not significantly different from the overall proportion of species in each class for the three groups with the 95% distributions less than 2.5% from the actual values (Baillie 2008, Meagher *et al.* pers comm.). In a number of taxonomic groups there will be strata with few representatives (e.g., one member of an Order), and these may not be represented in the sample, but this does not mean that the sample is not generally representative of the entire taxonomic group (Baillie 2008, Meagher *et al.* pers comm.). The SRLI species will be reassessed at regular intervals allowing the changes in conservation status for species and overall trends in more speciose groups to be quantified (Butchart *et al.* 2005; Baillie *et al.* 2008).The SRLI for Plants (Brummitt *et al.* 2008) assesses the threat status of a random sample of 1,500 species for each of four major taxa: bryophytes, pteridophytes, monocots, dicots; and all of the c.1000 gymnosperm species.

**Correlates of extinction risk**

Species are often the fundamental unit for conservation and the level at which many initiatives work (IUCN 2008), therefore understanding what drives species to extinction and why different taxa are not equally susceptible is of central importance, whilst the scale of the current extinction crisis is such that global approaches are needed both to understand the complex patterns of extinction risk, and to counter them (Rondinini *et al.* 2011).

Studying current extinction risk differs from the study of past extinctions, as it is likely the ultimate causes differ (Purvis 2008). Available data show that a range of factors associated with human impact, such as widespread habitat loss, contribute to current risk of extinction (Cardillo *et al.* 2008), whilst complex synergies exist between such factors and the species local environment to influence a species susceptibility (Freville *et al.* 2007). One method that has proved powerful in disentangling the extrinsic and intrinsic factors influencing a species extinction risk is a comparative analysis across species (Purvis *et al.* 2000a).

However, despite their importance in a number of ecosystems, global studies investigating correlates of extinction risk in plants are relatively scarce compared to those for other taxa (Cardillo *et al.* 2008, Cardillo *et al.* 2005, Cooper *et al.* 2008, Jones *et al.* 2003, Purvis 2008). This is likely to be a consequence of their large number of species, as described above.

Those studies investigating correlates of extinction risk in plants have found that a number of intrinsic traits can influence the risk of extinction of a species (Sodhi *et al.* 2008). These include breeding system; for example, dioecious species (with male and female reproductive functions separated into separate individuals) have been demonstrated to have an increased risk of extinction when compared with hermaphroditic species (Vamosi and Vamosi 2005). Mode of pollination also has an effect; species with biotic pollination syndromes have an increased risk of extinction over those with abiotic vectors (Pocock *et al.* 2006, Sodhi *et al.* 2008). One criticism of this trait-by-trait approach is that it treats traits in isolation from each other, neglecting the interactions between traits that may affect the extinction risk (Lawton *et al.* 1994). For example, most dioecious species require pollinators in order to set seed and have been demonstrated to be more susceptible to fluctuations in local pollinator abundance than are hermaphroditic species (Heilbuth *et al.* 2001).

However, using multivariate approaches, such as generalised linear models, complex synergistic effects and interactions can be taken into account.

In chapter 3 of this thesis, I use the comparative method to investigate the complex patterns of global species extinction risk for the sample of SRLI of monocot species. The monocotyledons are a large group of plants with an estimated 60,000 species, a large diversity of life forms including numerous commercially important species, with a worldwide distribution. They therefore make a suitable group to study global patterns of extinction. The majority of aforementioned studies have focused solely upon ecological traits (Sodhi *et al.* 2008, Bradshaw *et al.* 2008); however, anthropogenic (e.g. urbanisation) and climatic (e.g. temperature) factors have been demonstrated to play a significant role in shaping the observed variation in species extinction risk in a number of taxa (Cardillo *et al.* 2004, Cooper *et al.* 2008). I shall investigate how much influence these factors have in determining the susceptibility of plant species to extinction in order to provide a more holistic picture and allow direct comparisons with other taxa.

**Valuing species for conservation**

Given the number of species facing extinction, and the limited resources available to conservation, an understanding of the causal mechanisms leading to a species' extinction risk is required for the effective prioritisation of conservation effort. However, this is not achieved using species as the sole currency of conservation, which provides no method of filtering out those most deserving, i.e. it makes the assumption that all species are of equal value (Faith 1992). Consequently, conservation planners have long been forced into the unenviable situation of having to prioritise which species should receive the most protection (Vane-Wright *et al.* 1991) a dilemma that has been the focus of much discussion, described as the "agony of choice" or "Noah's Ark problem".

Attaching a weighting to a species raises the problem of defining value. A range of methods have been advocated, focusing on: restricted- range endemics (Roberts, *et al.* 2002), significant "umbrella" (Fleishman *et al.* 2000), "keystone" (Boswell *et al.* 1998) or "indicator" (Peterken 1974) species. A relatively recent approach has been to value species in economic terms; asking what services these species provide to humans and how much would it cost to replace them (Nunes and van den Bergh 2001; Millennium Ecosystem Assessment, 2005); however, the potential for human use has yet to be explored for most plant species. Furthermore, the role species play in sustaining the community in which they exist is not fully understood and so cannot usually be taken into account.

Due to the relative ease in determining the phylogenetic placement of a species, it has been suggested that species should be valued according to their evolutionary distinctiveness (Nee and May 1997). Species with greater evolutionary distinctiveness contain a greater quantity of unique evolutionary history (EH), as determined by the tempo and mode of divergence from its sister species, with the extinction of a species in an old, monotypic or species-poor clade resulting in a greater loss of EH than that of a more recently evolved species with many close relatives. Species with high EH values also have a greater influence on ecosystem structure and function, whilst, on average, having higher morphological variability and behavioral complexity (Erwin 2008). Conserving a species based on its EH may therefore be thought of as preserving future potential, increasing the chance of providing a species with the tools required to survive in an uncertain future (Forest *et al.* 2007). However, distinctiveness alone is not always a suitable method to prioritise species. For example, the most evolutionarily distinct species may not be at risk of extinction. Therefore, an understanding of the relationship between species' EH values and extinction is required to set priorities for plant conservation based not only on the likelihood that a species will be lost, but also on its irreplaceability (Isaac *et al.* 2007).

To date the results of studies relating EH to extinction risk have not been encouraging: mammal and bird species on longer branches (i.e. with high EH values) have higher extinction risk (Mace *et al.* 2003). The non-random extinction risk, demonstrated in many phylogenies (Chapter 3; Cooper *et al.* 2008, Cardillo *et al.* 2005, Davies *et al.* 2008) is thought to underlie this correlation and may have arisen due to either or both of the following scenarios: First, any phylogenetic clumping of factors that promote risk would increase the chance of all species in polytypic taxa, and hence those clades as a whole, being lost. Second, if such phylogenetically distributed traits have already mediated considerable extinction, then many monotypic genera or families might be the last survivors of once-larger clades (Purvis *et al.* 2000b).

Just as many of the studies investigating correlates of extinction risk have been of mammal, bird or amphibian taxa (Isaac *et al.* 2007, Purvis *et al.* 2000a), with only a minority focusing on plants, those looking at the relationship between EH and threat have similar taxonomic bias. Once again this is primarily due to a lack of detailed information, due to the large number of plant species and small number of specialists compared to those of other taxa. The few botanical studies of EH and threat have yielded mixed results. EH and threat have both been shown to be non-random with respect to phylogeny; however, unlike mammal and bird species, it has been suggested that threatened species tend to have low EH values, resulting in a decreased loss in EH compared to null model predictions (Schwartz and Simberloff 2001). On the other hand, a global scale analysis using a family level super-tree (Vamosi and Wilson 2008) provided some evidence that, as observed in other taxa (Isaac *et al.* 2007, Purvis *et al.* 2000a), plant extinction risk is highest in small clades, resulting in elevated EH loss compared with a null model of random extinction.

Using a species-level phylogeny of the species included in the SRLI sample of monocot species (Chapter 2), Chapter 4 of this thesis tests the relationship between EH and species Red List ratings, providing, to my knowledge, the first study of how effective IUCN criteria are not only at highlighting plant species for conservation but valuing them through an EH framework.

**Valuing areas for conservation**

Rather than focus on the species as the fundamental unit of conservation, a community level approach may be adopted, focusing upon habitats or ecosystems (Hodgson *et al.* 2011). Such an approach is generally used in the design of reserves or protected areas (Margules and Pressey 2000, Pressey *et al.* 2007), however the same problem as discussed above with regard to species also arise: How should we value an area and prioritise between areas? Vane-Wright *et al.* (1991) developed the complementarity approach to address this problem: an area or areas are compared against one another to find the set of areas that collectively maximise the total diversity captured. First, a reference set of species present in an area is selected, usually those present in the locality sampled first. Second, the species present in the potential additional areas are compared to the reference set. Those areas that have the largest number of unique species compared to the reference set are added, maximising the number of unique species protected across all localities. This approach focuses upon the diversity of species present, and has led to the development of biodiversity as the unit of measurement frequently used to prioritise areas.

In April 2002 the Convention on Biological Diversity (CBD, 2002) committed to achieve by 2010 a significant reduction of the current rate of biodiversity loss at the global, regional and national level – using the level of biodiversity within an area as a measure of its value.

However this raises a fundamental question: What is biodiversity? In recent decades, conservation prioritisation has focused on measures of species richness and endemism, the number of species in a given area and the number of those known only from that area, as a proxy for biodiversity. Areas with higher species richness or endemism that had already undergone significant loss of biodiversity were designated biodiversity hotspots and deemed more deserving of urgent conservation attention (Myers *et al.* 2001).

However, this approach assumes all species are equal, and makes no attempt to differentiate species based on the role they play in their ecosystem (function) or their evolutionary (phylogenetic) uniqueness. This is particularly relevant because a number of studies have demonstrated that under a range of scenarios species richness may not be a good measure of these other facets of diversity (Cavender-Bares *et al.* 2009, Meynard *et al.* 2011).

Using the species level conservation assessments prepared for the SRLI for monocots, in Chapter 5, I look at the relationship between species richness, phylogenetic and functional diversity to assess under what scenarios species richness may prove to be a poor surrogate, as a contribution to the debate as to how conservation resources may be allocated with maximum efficiency.

***Phylogenetic Diversity***

As discussed above and in Chapter 3, EH can be used as a method to attach value to species. This can be extrapolated from a species-based valuation to a broader scale, based on the complementarity approach of Vane-Wright *et al.* (1991). The value of an area is assigned according to the sum of the species' individual EH values within the area, known as the phylogenetic diversity (PD; Faith 1996). One advantage of this metric is the ease with which it can be calculated and modified as more species or

areas are added. A dated phylogenetic tree containing all species across all areas is used to sum EH values for individual species within an area for an overall PD, which can then be compared to the score for the other areas, with the best area the one that spans the greatest number of nodes on the phylogeny (Faith 1996).

Beyond preserving the species that are the most unique in evolutionary history (the benefits of this are discussed above), the quantity of PD within an area is an important factor in explaining community productivity patterns and stability, thereby benefiting the community as a whole (Cadotte *et al.* 2008).

***Functional diversity***

Above I discussed how species richness is not the only facet of biodiversity and therefore not the only way to value an area; similarly, phylogenetic diversity is not necessarily a complete measure either. It has long been suggested that the functioning of ecosystems depends on the diversity of ecological traits shown by the species within the ecosystems (Tilman *et al.* 1997, Loreau 1998, Ives *et al.* 1999) a diversity that can be quantified and also used to assign value to an area (Diaz and Cabido 1997).

Measures of trait (functional) diversity have existed in a number of forms for a number of years; however there are several stumbling blocks preventing the widespread use of trait diversity by conservation biologists, for example in how such functional diversity should be best measured (Petchey and Gaston 2002). Either all species are considered to be functionally unique, or some species are considered to be functionally identical, such that functional groups can be identified. However, in reality functional diversity is a spectrum, which can take any point between these two extremes, with some species being more similar than others at coarse functional scales and all species differing at finer functional scales (Petchey and Gaston 2006).

There have been a number of broad functional classification schemes. Naeem (1998) put organisms into one of four functional groups based on traits describing how they assimilate energy and carbon, whilst others differentiate functional groups based on resource use by their trophic position (Loreau *et al*. 2001); e.g. herbivores, primary consumers, secondary consumers and parasites. However, these classifications require that species are assigned to functional groups and, hence, arbitrary decisions about the scale at which differences between species are functionally insignificant are needed. Discontinuous measures have previously been used; i.e. the number of functional types or groups within the community (Stevens *et al.* 2003), as discussed above. Alternatively, continuous measures can be used and offer the advantage of providing a more informative and quantitative measure of the diversity present (Petchey and Gaston 2002).

The method I will use in this thesis is that of Petchey and Gaston (2002), termed functional diversity (FD): A species-by-trait matrix is converted into a distance matrix - a method similar to that of a principal component analysis, based on the spread of points (species) in *n* dimensional trait space, clustering algorithms are then used to convert the distance between the points into a functional dendrogram. A dendogram is similar to a phylogenetic tree in that it is a diagram used to represent relationships between species; however, rather than being based on the degree of genetic sequence similarity, in this case, it is based on the degree of functional similarity. Consequently the same method as used to calculate the PD of an area – the summating of branch lengths of the species present – can also be used to quantify the FD present in that area. One advantage of this method is that it was developed under the same methodological framework used to calculate PD; consequently it is a directly comparable metric Furthermore it is a continuous method and therefore avoids the artificial categorisation of species into classes (Petchey and Gaston 2002).

## Thesis aims

The overall aim of this thesis is to investigate why certain species face extinction, whilst others do not; and of those species facing extinction, how we should prioritise them in terms of conservation effort. This is achieved using a large, representative sample of monocots as the study group. Throughout, I develop and expand upon methods used in other taxa to understand patterns of plant extinction; this allows direct comparisons between the plight of plants and other groups to be made.

In Chapter 2, I outline the methods used to construct the phylogenetic tree used in all subsequent chapters.

In Chapter 3, I use a variety of statistical models to tease out the correlates of plant extinction and investigate how these interact with the plants' local environment, resulting in the global patterns of extinction risk observed today.

In Chapter 4, I ask, in the context of conservation, are all species equal or should we prioritise some over others and how well do current conservation methods achieve this?

Finally, in Chapter 5, I extend this to the community level asking what is biodiversity, how should we quantify it and how do environmental variables influence its distribution?

# Chapter 2: Developing a phylogenetic context

## Introduction

Genetic information has revolutionised evolutionary biology, providing a framework to test and develop a range of hypotheses. Subsequently there is a large demand for genetic sequences, which through advancements in computational methods and molecular techniques, is now relatively easy to produce in ever-increasing quantities. This has led to the development of large, online large databases such as GenBank, which act as repositories for sequence information (Benson *et al.* 2007). On December 15th 2010, GenBank contained an estimated 129 million sequences, and has doubled approximately every 18 months between 1982 and 2007 (Benson *et al.*2007).

Recently, there has been a rapidly increasing effort from a range of disciplines, such as ecology, to utilise this information, which has sought to put species into a phylogenetic context, bringing the evolutionary history and genealogical relationships of species to bear on questions of conservation, community assembly and patterns of diversity (Cavender-Bares *et al.* 2009). In the context of ecology, phylogenetic trees have proved to be most useful; providing a quantitative measure of species difference, therefore allowing the discrimination of species beyond their phenotypes (Faith 1996).

As sequence data for a wide variety of organisms has become easier to acquire, interest in inferring large-scale phylogenies has increased dramatically (Bininda- Emonds *et al.* 2002, Pisani *et al.* 2002, Bininda-Emonds 2004, McMahon and Sanderson 2006). However, the process of going from raw GenBank sequence data to a large, polished phylogeny involves several computing challenges.

A phylogeny of the 1500 monocot sampled red list index (SRLI) species is a cornerstone of my thesis, and a central requirement for the following chapters. However, the reconstruction of a phylogenetic tree for the monocot SRLI exemplifies this challenge: having a large sample size (1500 species), a disjointed taxon set (random sample from across the whole monocot phylogeny), and a number of families or genera that are infrequently studied, and are therefore less likely to have been sequenced.

## Developing the method

### *Marker selection*

As the phylogeny is large and spans multiple orders and families, both historical and contemporary processes need to be resolved in order to place species, a range of genetic markers were needed. Using a range of markers results in a phenomenon known as "signal enhancement" where the analysis of two or more combined phylogenetic data sets can yield a novel solution not supported by the analysis of either data set separately (Bininda-Emonds 2004). This arises because, when combined, the congruent sub-signals in the two data sets outweigh the incongruent primary signals in each. Generally, non-coding regions assign species to clades (shallow = recent common ancestors) and protein coding and nuclear ribosomal gene regions infer inter-clade (deep = less recent common ancestors) relationships (Pirie *et al.* 2008). In this study the non-coding regions - *trnL-F* and *ITS* - were selected because of their high variability, including length variation, whilst protein coding regions *matK* (including non-coding flanking *trnK* spacer regions), *rbcL*, and *ndhF* were sampled in order to resolve relationships between the clades (Pirie *et al.* 2008). The markers *matK* and *ndhF* have been shown to be sufficiently variable to be useful for phylogeny reconstruction at relatively low taxonomic levels in other plant groups

(Kim and Jansen 1995, Baum *et al.* 1998), while *rbcL* is less variable but has been used in large number of studies and therefore has a good representation on GenBank.

### *Gathering and standardising the data*

Sequences were downloaded from GenBank (Supplementary material 1) by entering the following Boolean expression into the Entrez nucleotide database: "Liliopsida [organism] AND (*matK* OR *trnL* OR *rbcL* OR *ndhF* OR *ITS*)", which searches the monocots for any of the keywords in any field. I further limited the records to include only those of 5kb length or less to exclude genomic sequences. Where species from the 1500 were not present surrogate species were selected based on taxonomic affiliation: i.e. subspecies or section (Supplementary material 1).

The next step was to standardise taxon names, this is typically a slow and error prone process but is essential to maintain the integrity and validity of the dataset. Typical problems that were corrected include; the numerous errors present in GenBank submissions (misspellings and incorrect suffixes) (Bidartondo, 2008), updating names where taxonomy has changed between submissions and dealing with name duplications where single species are represented by multiple names, using the International Plant Names Index (IPNI).

### *Data preparation*

Processing a large number of highly diverse sequences raises a range of methodological challenges, however a number of methods have been developed to handle such data, of which I shall discuss two of the most commonly used;

*Supertree or supermatrix?*

The supertree method compiles many source trees, which are generated by the independent analysis of the various loci; these trees are then combined either directly or indirectly to produce a single, joint estimate of phylogeny or a “supertree” (Fig.1-1).

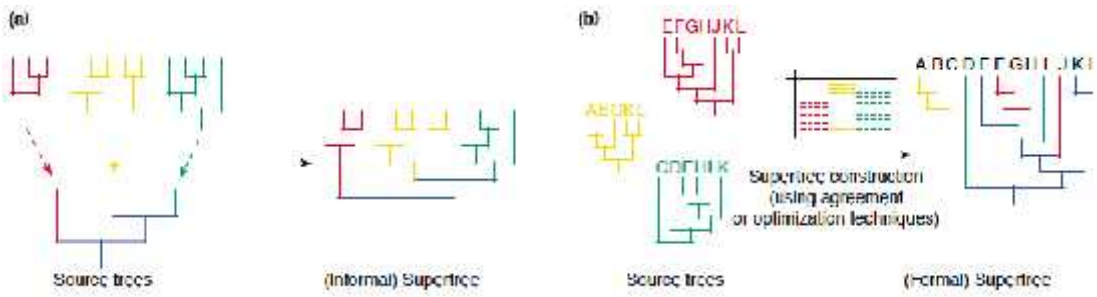


**Fig. 1-1.** The development of informal (a) and formal (b) supertrees, reproduced with permission (Bininda-Emonds 2004).

This approach has the advantage that it does not assume that all characters have experienced the same branching history and can work from incompatible data types (e.g. DNA sequences versus DNA–DNA hybridisation data) (Bininda-Emonds 2004). Although straightforward, supertree methods are not without their limitations, including problems related to data independence (the same data can contribute to more than one source tree) resulting in "artificial signal enhancement”, whilst novel relationships in each independent tree can contradict each other and perhaps most importantly, supertrees do not directly rely on the primary data for tree inference, making novel topologies suspect (Smith *et al.* 2009). Perhaps due to these limitations, and despite active development of methodologies, few large supertrees for diverse groups have been successfully constructed. (However, see Jones *et al.* 2002).

Alternatively, the sequences from the individual loci can be concatenated into one large "supermatrix" from which a tree is inferred directly. The gold standard for such an analysis remains a complete-matrix with the aim of scoring all characters for all taxa. However because the data is often not obtained by targeted projects the taxon sampling among loci tends to be very heterogeneous, leading to supermatrices that typically have high quantities of missing data. Despite this sparse-matrix studies have met with substantial success, even with up to 95% missing data (McMahon and Sanderson 2006). However there is always the risk that the heterogeneous nature of the data can mask the phylogenetic information contained in these datasets and give the misleading appearance of great uncertainty (Thomson and Shaffer 2010).

*Sequence alignment*

One of the biggest problems in using a range of different loci is the alignment of the sequence data, which is not really an issue in the supertree approach as each locus is aligned independently, however working under a supermatrix framework the matrix will contain both closely and distantly related taxa, which need to be aligned simultaneously. Whilst the majority of multiple sequence alignment algorithms generally assume that sequences differ only through processes of substitutions and small insertions or deletions, therefore large-scale differences in length and degree of overlap can lead to severe problems (McMahon 2006).

**Final method**

I experienced trouble aligning the supermatrix, the most difficult data sets originating from regions with high sequence variation, particularly the nuclear ribosomal *ITS* and the plastid *trnL-F* intron. However these are also among the most frequently sequenced regions and were therefore critical to the study. Consequently I had to modify the approach; I performed single gene alignments using MAFFT (Katoh *et al.* 2002), and then concatenated each of the gene sequences into a supermatrix using

CONCATENATE (Alexis Criscuolo, http://www.supertriplets.univ-montp2.fr/PhyloTools .php.). Taxa for which no sequences were gathered for a given partition were coded as missing values for the corresponding cells. This resulted in 1,305 species (Supplementary material 2), with 1,069 unambiguously aligned informative characters and 83% missing data, from which the phylogeny was inferred.

The general time-reversible model of sequence evolution with a gamma distribution, to account for among-site nucleotide substitution rate variation (GTR+gamma), was selected by jMODELTEST (Posada 2008) as the substitution model that best fit the data; and was subsequently used in all phylogenetic analyses. A single-partition ML analysis was performed using RAxML version 7.0.0 (Stamatakis 2006) with a 1000 rapid bootstrap analysis followed by the search of the best-scoring tree in one single run. This analysis was done using the facilities made available by the CIPRES portal in San Diego, USA (http://www.phylo.org/sub_sections/portal).

Due to missing data, and a lack of marker overlap, a number of species were misplaced; therefore, an order level backbone constraint was used. The constraint was built such that all taxa were present, but grouped in orders in accordance with APG III (The Angiosperm Phylogeny Group 2009). Species were arranged as a polytomy within orders in this constraint. This approach allowed the topology of the species within each order to be resolved with the sequence data from the supermatrix, while the ordinal backbone of the tree was defined *a priori*.

It is well established that molecular rates vary across lineages, and that this makes it difficult to estimate divergence dates accurately using DNA sequence data (Arbogast, *et al.* 2002). Therefore, I converted the molecular branch lengths from the ML analysis into units of time using nonparametric rate smoothing (NPRS; Sanderson, 1997) as implemented in the program *R8s* (version 1.7.1; Sanderson 2003), which

allows for different rates on each branch of the tree.

Non-parametric rate smoothing (NPRS) is a method for estimating divergence times when evolutionary rates are variable across lineages, using a least squares smoothing approach that penalizes rates that change too quickly from branch to neighboring branch (**Sanderson, 2002**). However NPRS makes a number of assumptions that I will outline below: Firstly, NPRS may overfit the data, leading to rapid fluctuations in rates in regions of a phylogeny where there are short branches (Bell and Donoghue 2005). Data sets with little information content (i.e., few inferred substitutions across the phylogeny) may tend to have zero-length branches in areas that are fairly unresolved, which would result in the appearance of rapid rate fluctuation (Bell and Donoghue 2005). Second, NPRS assumes that there is an autocorrelation of substitution rates and attempts to minimize rate changes between ancestral/descendant branches on a tree (i.e., at the nodes) (Bell and Donoghue 2005). However using a number of different genes with different patterns of molecular evolution would tend to compensate for unusual patterns in any single dataset **(**Rutschmann *et al.* 2004**)**. Alternatively, Bayesian based dating methods i.e. Mr Bayes and BEAST were trailed, however due to the large number of species and the variety of markers, the chain failed to converge.

In order to estimate the divergence times, the tree was calibrated using five well-characterised fossils (Table 2.1). The final tree was visually checked against published phylogenies and used as the basis for all phylogenetic analyses throughout the thesis (Supplementary material 3).

| Fossils | Reference | Node | Age (Mya) |
|---|---|---|---|
| *Spinizonocolpites* (Arecaceae; pollen) | (Harley 1996, Bremer 2000) | Stem node of Arecaceae | 89.3 (min) |
| *Spirematospermum* (Zingiberales; fruits) | (Knobloch and Mai 1986, Bremer 2000) | Stem node of Zingiberales | 83.5 (min) |
| *Monoporites* (Poaceae; pollen) | (Linder 1987, Bremer 2000) | Stem node of Poaceae | 65.5 (min) |
| Araceae | (Friis *et al.* 2004, Anderson and Janβen 2009) | Stem node of Alismatales | 120 (min) |
| Pandanales | (Gandolfo *et al.* 2002) | Stem node of Pandanales | 90 (min) |
| Secondary molecular estimate | (Bremer *et al.* 2003, Janssen and Bremer 2004) | Crown node of monocots | 134 (max) |

**Table 2-1.** Fossils and secondary molecular estimates used as minimum and maximum constraints used to obtain absolute time estimates for the phylogenetic tree of monocots.

# Chapter 3: Correlates of plant extinction risk: anthropogenic, climatic and life-history

## Introduction

Declines and extinctions in the world's flora have long been a cause for concern (Koopowitz 1983). However, due to the number of plant species, an estimated 352,000 flowering plants (Paton, Brummitt *et al.* 2008) and their great diversity (Ricklefs and Renner 1994, Kreft and Jetz 2007), very few studies compared to that of mammals (Cardillo, Mace *et al.* 2005, Cardillo, Mace *et al.* 2006, Isaac, Turvey *et al.* 2007) have sought to quantify or explain their decline. In fact fewer than one thousand species of plants were properly documented by the International Union for the Conservation of Nature (IUCN) in their last report (IUCN 2011). Given that we are in what many consider to be the 6$^{th}$ great mass extinction (Wake and Vredenburg 2008), together with the fundamental role plants play in many ecosystem processes, the lack of information regarding patterns and processes of plant extinctions needs to be addressed.

As an attempt to address the taxonomic bias in IUCN assessments, (IUCN 2011), and as a measure towards the World Summit on Sustainable Development 2010 target (to significantly reduce the current rate of biodiversity loss by 2010 (WHO 2002) the Sampled Red List Index (SRLI) was developed (Butchart, Stattersfield *et al.* 2005, Baillie, Collen *et al.* 2008). The SRLI, a modification of the Red List Index, reassesses the same random sample of species at regular intervals and therefore allows the relative changes in conservation status for more speciose groups to be quantified. For plants, the SRLI assesses the threat status of a random sample of 1500 species for each of 5 major taxa: bryophytes, pteriophytes, monocots, dicots and all of the <1,500 gymnosperms (Brummitt, Bachman *et al.* 2008).

The publication of the first report from the SRLI for plants project, "Plants under pressure – a global assessment" (Royal Botanic Gardens Kew 2010), brought the plight of plants and to the forefront of conservation biology. The study results, based on preliminary GIS based conservation assessments, found that of the 7000 plant species studied, 21.5% of species were threatened, i.e. were within the IUCN Red List categories of Vulnerable, Endangered or Critically Endangered (IUCN, 2001). This percentage is similar to that calculated for mammals (23% threatened) and much higher than birds (12% threatened) (Baillie, Hilton-Taylor *et al.* 2004). In addition, 33% of plant species (compared with 7.8% of mammals and 0.8% of birds; Baillie, Hilton-Taylor *et al.* 2004) are classed as data deficient (DD) (Royal Botanic Gardens Kew, 2010). The high proportion of threatened and DD plant species makes understanding the processes behind extinction risk in plants a priority in conservation biology research.

An informative method by which to assess mammal, amphibian and bird extinction risk has been the identification of correlates of extinction risk through statistical models (Bennett and Owens 1997, Purvis *et al.* 2000a, Jones, Purvis *et al.* 2003, Cardillo, Mace *et al.* 2008, Cooper, Bielby *et al.* 2008). The underlying assumption of such a method is that extinction risk is not random, but rather that there are intrinsic traits associated with species vulnerability to extinction (Bennett and Owens 1997). Traits associated with species ecology, reproduction, body size and range size have proved informative in modeling the variation in extinction risk across a number of animal taxa (Purvis *et al.* 2000a, Cardillo, Mace *et al.* 2005, Cooper, Bielby *et al.* 2008). Furthermore, the integration of extrinsic factors, i.e. anthropogenic (Cardillo, Purvis *et al.* 2004) and climatic (Purvis *et al.* 2000a, Cardillo, Mace *et al.* 2005, Cooper, Bielby *et al.* 2008), both as independent predictors and through interaction effects have increased the explanatory power of such models.

Studies have demonstrated that extinction risk in plants is also non-random (Vamosi and Wilson 2008); for example, species that are dioecious (separation of male and female function into separate individuals) have been demonstrated to have an increased risk of extinction when compared to hermaphroditic species (Vamosi and Vamosi 2005).

For example, in dioecious species only female plants disperse seed, potentially leading to a reduced dispersal in comparison to hermaphroditic species, with concomitant increased local resource competition (Heilbuth, Ilves *et al.* 2001). Furthermore, the mode of pollination has been demonstrated to influence risk of extinction (Sodhi, Koh *et al.* 2008). For example, species with biotic pollination syndromes are susceptible to fluctuations and declines in potential pollinators (Schweiger, Biesmeijer *et al.* 2010). With numerous potential pathways leading to a decline or local extinction of pollinators (Potts, Biesmeijer *et al.* 2010), an increased extinction risk for species with biotic pollination syndromes has been demonstrated (Pocock, Hartley *et al.* 2006, Sodhi, Koh *et al.* 2008). Such traits may not act independently from one another; rather complex synergistic effects and interactions may operate between traits to influence extinction risk (Freville, McConway *et al.* 2007, Brook, Sodhi *et al.* 2008). For example, most dioecious species require pollinators to set seed and therefore might be more susceptible to fluctuations in local pollinator abundance than hermaphroditic species (Heilbuth, Ilves *et al.* 2001).

Furthermore, synergies may occur between intrinsic traits and extrinsic factors to influence a species extinction risk. For example, it has been demonstrated that plant communities subjected to anthropogenic disturbance contain a non-random subset of species (Lôbo, Leão *et al.* 2011), with factors such as agricultural development and habitat fragmentation disproportionately increasing extinction in plant species based on life history traits; specifically, seed size (Laurance, Ferreira *et al.* 1998, de Melo, Dirzo *et al.* 2006), shade tolerance (Laurance, Ferreira *et al.* 1998), pollination syndrome (Laurance, Nascimento *et al.* 2006) and breeding system (Girao, Lopes *et al.* 2007).

The relative importance of intrinsic traits as drivers of extinction risk in plants has recently been called into question (Davies, Smith *et al.* 2011, Knapp 2011). Looking at the percentage of variance in threat explained by intrinsic traits in mammals (51%; Cardillo, Purvis *et al.* 2004) and frogs (59%; Cooper, Bielby *et al.* 2008) compared to plants (14-16%) (Bradshaw, Giam *et al.* 2008, Sodhi, Koh *et al.* 2008), this appears to be the case. Moreover, Davies *et al.* (2011) suggest

that extinction risk is in plants is largely associated with phylogenetic diversification rate, with younger lineages more at risk than older lineages, rather than specific traits. However, it could be argued that this apparent difference between animal and plant groups could be due in large part to the fact that different biological traits are relevant to the two different groups.

Thus, the relative importance of various drivers of extinction risk in plants requires clarification. The combination of intrinsic traits and extrinsic environmental factors has proven effective at providing guidance for conservation policy in several animal taxa, and the question is open as to whether modeling such effects would be useful for plant conservation policy. Moreover, few studies investigating correlates of extinction risks in plants have used multivariate methods (Bradshaw, Giam *et al.* 2008, Sodhi, Koh *et al.* 2008), therefore potentially neglecting the importance of the interaction between various effects in explaining extinction risk of plant species. Furthermore, few studies have examined the importance of extrinsic factors and the synergistic effects between these and intrinsic traits, despite the potential importance of these interactions (Freville, McConway *et al.* 2007). Therefore, despite the significance of anthropogenic and climatic factors as correlates of threat in other taxa (Cardillo, Purvis *et al.* 2004, Cooper, Bielby *et al.* 2008), the synergistic effect they have with life history traits and the demonstration that they significantly influence plant distribution patterns (O'Brien, 1993) and survival (Girao, Lopes *et al.* 2007), the explanatory power of models that comprehensively incorporate such effects and their interactions has yet to be explored.

The present study investigated the effectiveness of modeling intrinsic traits and extrinsic environmental factors and their interactions as drivers of plant extinction

risk. Using the species sampled for the monocot SRLI (Baillie *et al.* 2008), I aim to reveal the relative importance of intrinsic traits, anthropogenic environmental impacts and climatic environmental factors and their respective interactions in determining plant extinction risk. Thus, the overall objective is to determine whether such models can usefully inform plant biodiversity policy and also to identify specific factors that should be considered as such policy is developed.

## Method

### *Data collection*

IUCN categories (IUCN, 2001) were obtained from http://threatenedplants.myspecies.info, (accessed on 10/05/11) and used as a surrogate measure of extinction risk for the monocot species (Rodrigues, Pilgrim *et al.* 2006). I coded species extinction probability i.e. threatened vs. non-threatened as a binomial response variable (Bradshaw, Giam *et al.* 2008, Sodhi, Koh *et al.* 2008), re-classifying Vulnerable, Endangered or Critically Endangered species as threatened and Least Concern or Near Threatened species as non-threatened. Species with the IUCN designation DD and those without full assessments were not included in the analyses, resulting in threat codings for 871 species (Supplementary material 2).

Of 871 monocot species assessed as non-Data Deficient, 221 were threatened (Vulnerable, Endangered or Critically Endangered). These were predominantly assessed using Criteria B (62%) and D (31%) (Figure 3-1), although of course the same species can be assessed under multiple criteria (IUCN 2001).

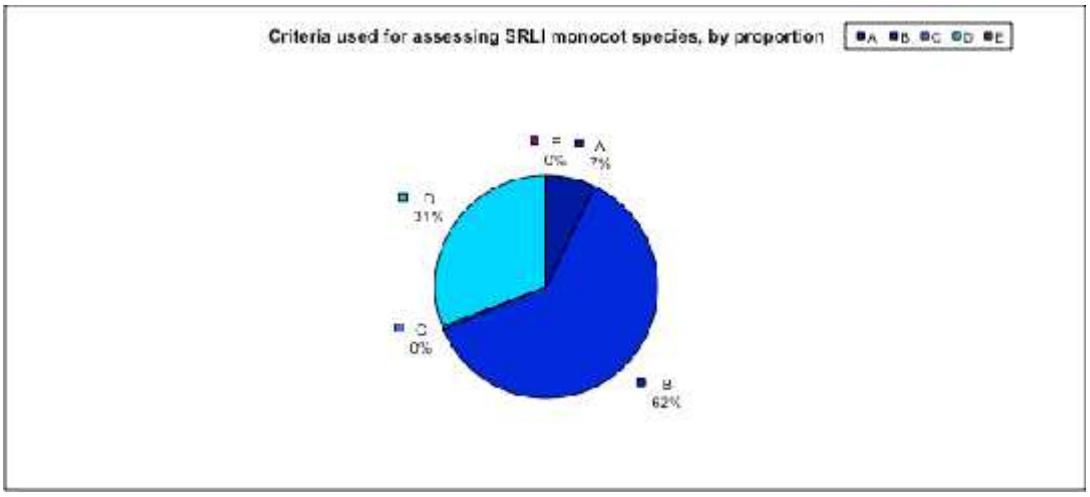


**Fig. 3-1.** The proportion of species assessed by each IUCN category.

Species intrinsic traits were collected from primary taxonomic literature and expert advice (Supplementary material 4, Supplementary material 5) and where species-specific information was not available, traits were inferred from the genus level. The spectrum of traits collected (Table 3-1), reflects those variables that have proved informative in previous similar studies (Vamosi and Vamosi 2005, Freville, McConway *et al.* 2007, Bradshaw, Giam *et al.* 2008, Sodhi, Koh *et al.* 2008);

Traits corresponding to pollination, breeding system and dispersal: primary pollinating mechanism (PD = abiotic or biotic), breeding system (BS = unisexual plant, unisexual flowers, or bisexual flowers), floral symmetry (FS = actinomorphic or zygomorphic), fruit type (FR = fleshy or non-fleshy) and the number of seeds in the fruit (SE = single, few (2- 40) or numerous (41+);

Traits corresponding to persistence: seasonality (SEA = annual or perennial), the presence of an underground storage organ (STO = present or absent), the presence or absence of an endosperm (END = present or absent) and their woodiness (WO = woody or not woody);

Traits corresponding to primary habitat (HAB = epiphytic, semi-aquatic or terrestrial) and altitudinal range (ALT = low: 0-1000m, mid: 1001-2000m, high: 2001 + and all: low – high).

Anthropogenic impact variables (Table 3-2) were based on gross domestic product (GDP) (CIESIN, 2004) and human population density (HPD) (CIESIN, 2005). In order to obtain species-specific measures for these variables, a range distribution shape file for each monocot species was generated using herbarium records, resulting in spatial data for 915 species (Supplementary material 2). The shape files were then overlaid with GDP and HPD data from which the median grid cell value for each variable across the geographical range of each species was recorded using the Spatial Analyst extension in the program ArcGIS (Cooper, Bielby *et al.* 2008).

Variables were measured across the ranges of individual species, but without reference to overall species range size. Although this means that variables show reduced variance for restricted-range species, it was not possible to also use range size as a functional trait since most species had already been assessed under IUCN Criterion B, which is based on measures of geographic range size, and it was felt important to apply a uniform set of functional traits across all species, in order to compare species responses. Restricted-range species may disproportionately show certain functional traits, and monocot species assessed for the SRLI project under Criterion B will have restricted ranges by definition. However, since the IUCN criteria measure threat rather than purely range size, many restricted-range species may be naturally rare but not threatened (e.g. in isolated locations within undisturbed habitats); therefore many species treated here as not threatened will also possess functional traits more likely to occur in restricted-range species.

| States | Level |
|---|---|
| Storage organs | Absent |
| | Present |
| Breeding system | Hermaphrodite |
| | Monoecious |
| | Dioecious |
| Pollen dispersal | Abiotic |
| | Biotic |
| Fruit | Dry |
| | Fleshy |
| Endosperm | Absent |
| | Present |
| Woodiness | Absent |
| | Present |
| Number of seeds | Few |
| | Numerous |
| | Single |
| Floral symmetry | Actinomorphic |
| | Zygomorphic |

**Table 3-1.** Life history traits and corresponding levels analysed.

| Group | Variable | Description |
|---|---|---|
| Climatic | STD PET | Standard deviation of annual potential evapotranspiration. |
| | Mean PET | Average annual potential evapotranspiration. |
| | Mean temp | The average annual temperature. |
| | STD temp | Standard deviation of temperature. |
| | Max temp | The maximum temperature. |
| | Min temp | The minimum temperature. |
| | Mean precipitation | The average annual precipitation. |
| | STD precipitation | Standard deviation of precipitation. |
| | Max precipitation | The maximum precipitation. |
| | Min precipitation | The minimum precipitation. |
| Anthropogenic | GDP | The gross domestic product. |
| | HPD | Population Density for 2000. |

**Table 3-2.** The climatic and anthropogenic variables used to explain variation in plant extinction risk.

The final analysis was conducted on 618 species, as that was the number of species with IUCN and anthropogenic data available (Supplementary material 2).

***Data preparation***

All variables were tested for significant collinearity using variance inflation factors, in order to identify those variables that could be dropped due to redundancy, however no significant redundancy was detected. Continuous variables were standardised to have a mean of 0 and a variance of 1, whilst categorical variables were coded as factors.

A principal component analysis was conducted on all the environmental variables with the first two components explaining 81% of the variance: PC1 = principal component-precipitation and PC2 = principal component –temperature.

Species are not evolutionarily independent from one another; and although extinction risk itself does not evolve along phylogenies, intrinsic biological traits do, making it necessary to use analyses that control for phylogeny to ensure statistical independence of data points (Cardillo, Mace *et al.* 2008). In order to test for statistical non-independence among species trait values I used Pagel's λ (1999), a measure that normally varies between 0 (phylogenetic independence) and (species' traits co-vary in direct proportion to their shared evolutionary history). A variance component analysis was performed to quantify how much of the variance in a trait and extinction risk can be explained by the position across taxonomic ranking (e.g. Class, Order, Family, etc.) (APE library v2.7-2; Paradis 2004) (R v2.13; R Development Team, http://www.r-project.org).

***Statistical analysis***

In order to account for phylogenetic non-independence, I used two approaches. Firstly, phylogenetic independent contrasts (Felsenstein, 1985) were generated and used in subsequent analyses (CAIC library; Purvis and Rambaut 1995) (R v2.13; R Development Team, http://www.r-project.org). Secondly, I coded the random-effects error structure of the GLMM as a hierarchical taxonomic effect i.e. the species order, family and genus was included as a random variable (Blackburn and Duncan 2001). However, as the analysis had an incomplete phylogeny, a large number of categorical variables and model selection was the statistical inference being used; the latter approach is more appropriate (Blackburn and Duncan 2001). Moreover, the final models selected from both methods were the same.

Therefore only the results from the random effects method are presented.

Generalised linear mixed effect models (GLMMs) were fitted to determine the influence of intrinsic traits, anthropogenic and climatic variables on the extinction risk of the sample of 618 monocot species. The models were fitted with a binomial response variable and each potential correlate as a linear predictor; each model had a binomial error distribution and a logit link function; and models were fitted using the lmer function (lme4 library v0.99; Bates 2007) in R (v2.13; R Development Team http://www.r-project.org).

Two trait:trait pairwise interactions (BS*FR and BS*HAB) were included in the analysis based on previous studies that have demonstrated their importance in patterns of plant extinction risk (Sodhi 2008), whilst four trait:climatic interactions (END*PC1, END*PC2, STO*PC1 and STO*PC2) were included to reflect the importance of underground storage organs and seed endosperm (Baroux 2002) in plant survival at extremes of temperature and water availability. Furthermore, four interactions between traits and anthropogenic factors (PD*GDP, PD*HPD, BS*GDP and BS*HPD) were also investigated to determine the impact of anthropogenic disturbance on pollination syndrome (Laurance, Nascimento *et al.* 2006) and the non-random effect it has on plant breeding systems (Girao, Lopes *et al.* 2007).

Because factors in this analysis were selected based on their effectiveness in other studies, and all the potential models are biologically possible, an extensive model building strategy was used. However, due to the potential number of candidate models to find the combination of variables resulting in the most parsimonious model that explains the variation in each metric, a supervised forward model selection procedure was adopted. Univariate models were tried first and the $AIC_c$ (Akaike's Information Criterion corrected for small sample sizes) (Burnham and Anderson 2002) calculated. The model with lowest $AIC_c$ was then retained and tried in combination with each of the remaining variables.

Then the best of the bivariate models were selected and tried with each of the remaining variables, and this process was repeated until introducing additional variables into the model failed to yield a lower $AIC_C$.

AIC based approaches can favour more complex models (Burnham and Anderson 2004), therefore, Bayesian Information Criterion (BIC), a criteria that penalises complexity more than AIC (Burnham and Anderson 2004) was trialled, however, the BIC approach made no difference to the most parsimonious models selected, therefore only the results from the AIC method are presented. Akaike weights ($w_i$) were calculated to estimate the relative support for each model; $w_i$ represents the probability that model *i* is the best model for the observed data, given the candidate set of models and sums to 1 across all models considered (Burnham and Anderson 2002).

Two methods were used to assess the relative importance of each model: $w_i$, as described above (Burnham and Anderson 2002) and the partial $R^2$ of the factors from the most parsimonious model. The area under the receiver operator curve (AUC) was used as a measure of model prediction accuracy (Bradley 1997) (ROCR library; Sing, 2005 in R v2.13; R Development Team, http://www.r-project.org). AUC values have a possible range of 0–1, an AUC score of 1 indicating a perfect threatened vs. non-threatened prediction. AUC values are calculated by plotting the sensitivity (ratio of correctly classified positives to the total number of positive cases) against 1-specificity (false positive rate) at all possible thresholds of presence–absence classification (Bradley 1997).

The above strategy was used to find the most parsimonious model for each grouping of variables: intrinsic traits only, intrinsic traits and anthropogenic variables, intrinsic traits and climatic variables, and full models of intrinsic traits, anthropogenic and

climatic variables, with the model selection process starting from the beginning for each grouping. The most parsimonious models from each grouping were then compared in terms of the variance explained, the $w_i$, value and the relative predictive power as indicated by the AUC score (Bradley 1997).

## Results

Including only intrinsic traits, the most parsimonious model ($w_i$ = 0.79) explained 11% of the variation in extinction risk, with pollination syndrome, altitude, habitat and the number of seeds all significant variables (Table 3-3). Specifically, species pollinated by biotic vectors had an increased extinction risk over those with abiotic vectors, extinction risk increased with altitude, epiphytic species were more at risk than terrestrial and those with several seeds also show increased risk of extinction (Fig. 3-2). Following Swets (1988), the overall predictive power of the optimal model, as measured by the AUC score (0.75), may be considered as moderate (Table 3-3).

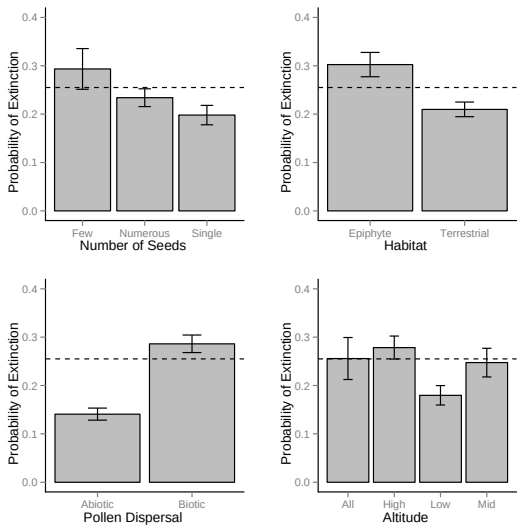


**Fig. 3-2.** Probability of a monocot species being threatened. The dotted line represents the threshold of being threatened (above the line) vs. non-threatened (below the line), with the error bars indicating the 95% confidence intervals based 10,000 bootstrap iterations. The dashed line is derived from a ROC curve and represents the cut-off value that best estimate the number of correct (i.e. maximizes sensitivity and specificity) threatened vs. non-threatened predictions across a range of potential cut-off values (Swets, 1988).

The addition of anthropogenic variables to the intrinsic traits resulted in the new most parsimonious model ($w_i$ = 0.69) explaining 20% of the variance having a predictive score of 0.76. The additional, significant variables were GDP, HPD, and the interaction terms BS*HPD, PD*GDP and PD*HPD (Table 3-3).

When climatic variables were added to intrinsic traits, the new most parsimonious model ($w_i$ = 0.67) explained 21% of the variance in threat with the predictive power increasing to 0.77. The most parsimonious model contained the same intrinsic traits as above in addition to the two climatic variables PC1 (precipitation) and PC2 (temperature) together with the interactions PC2*storage organ, PC1*endosperm and PC2*endosperm (Table 3-3).

Using all potential independent variables, the variance of the most parsimonious model ($w_i$ = 0.62) explained 29% of the variance in threat with a predictive power of 0.79 (Table 3-3). Thus, the final model was significantly better than any previous model, accounting for 100% of the information-theoretic weight ($w_i$ = 1: i.e. complete support).

The top 5 variables explained more than 50% of the observed variance, with the optimal model containing variables from each category; intrinsic trait, anthropogenic and climatic (Table 3-4). Rather than replacing variables from the previous models, the final model added extra variables; additional complexity increases the variation explained, suggesting that these extra variables are informative in a complimentary, rather than alternative, manner.

| Model | k | ΔAICc | $w_i$ | $R^2$ | AUC |
|---|---|---|---|---|---|
| *Traits only*<br>~PD+SE+ALT+HAB | 12 | 49.35 | 0 | 0.11 | 0.75 |
| *Traits and anthropogenic*<br>~PD + SE + ALT + HAB + GDP + HPD + PD*HPD + PD*GDP + BS*HPD | 20 | 20.3 | 0 | 0.2 | 0.76 |
| *Traits and climatic*<br>~PD+SE+ALT+HAB + WO + PC1 + PC2 + END*PC1 + END*PC2+ STO*PC2 | 22 | 18.53 | 0 | 0.21 | 0.77 |
| *Traits, anthropogenic and climatic*<br>~PD+SE+ALT+HAB+PC1+PC2+HPD+GDP+PC2*STO+ PC1*END+PC2*END + PD*GDP + PD*HPD + BS*HPD | 28 | 0 | 1 | 0.29 | 0.79 |

**Table 3-3.** The most parsimonious models from each category according to Akaike's Information Criterion corrected for small sample sizes (AICc). Terms shown are PD = pollen dispersal, SE = number of seeds, ALT = altitude, HAB = habitat, END = endospermous, STO = geophytic, BS = breeding system, GDP = gross domestic product, HPD = human population density, PC1 = principal component (precipitation), PC2 = principal component (temperature). Included are: the numbers of parameters (k), the difference in $AIC_c$ from the most parsimonious ($w_i$), the amount of variance the model explained ($R^2$) and the predictive power of the model (AUC).

Looking at the top 3 variables that explained more than 30% of the variation, the interaction between intrinsic traits and anthropogenic variables appear to be important, with PD*HPD explaining the greatest variance (12.5%) followed by the interaction BS*HPD (11.2) and pollination syndrome as a single factor (10.5) (Table 3-4).

| Trait | Deviance explained (%) |
|---|---|
| PD*HPD | 12.5 |
| BS*HPD | 11.2 |
| PD | 10.5 |
| HPD | 9.3 |
| PC2*STO | 9.1 |
| PC1*END | 9.0 |
| ALT | 7.0 |
| PC1 | 6.1 |
| PD*GDP | 5.4 |
| HAB | 5.1 |
| GDP | 4.5 |
| PC2*END | 3.6 |
| PC2 | 3.5 |
| SE | 3.2 |

**Table 3-4**. The deviance explained by the individual correlates of the most parsimonious model. Terms shown are PD = pollen dispersal, SE = number of seeds, ALT = altitude, HAB = habitat, WO = woodiness, END = endospermous, STO = geophytic, BS = breeding system, LC = life cycle, GDP = gross domestic product, HPD = human population density, PC1 = principal component 2 (precipitation), PC2 = principal component 3 (temperature).

## Discussion

As primary producers, plants play a fundamental role in a number of ecosystems and provide a range of services to the benefit of humans (Raskin, Ribnicky *et al.* 2002, Millennium Ecosystem Assessment 2005). Despite their importance, a high number of plant species are threatened with extinction (Royal Botanic Gardens Kew 2010); furthermore, the processes influencing their level of risk are relatively unknown compared to that of other taxa (Knapp 2011), making understanding the causal bases of extinction risk a priority.

It was found that plant models are inherently complex because amongst other factors discussed below, a species' interactions with its environment need to be taken into account in order to get reasonable explanatory power. When these interactions were added it was found that models performed comparably well with similar models of other taxa, providing greater predictive power for conservation initiatives.

***Correlates of plant extinction risk***

In this study the traits that best explained the observed variance; pollination syndrome, seed number, altitude and habitat, were consistent with those in other studies (Bradshaw, Giam *et al.* 2008, Sodhi, Koh *et al.* 2008), suggesting they influence a plant's susceptibility to extinction across a number of taxonomic groups and geographic locations.

However, rather than individual traits, the interaction between species intrinsic traits and the environment – both climate and anthropogenic - proved most informative. The interaction between human population density and pollen dispersal was the most significant, with increased human density significantly increasing the extinction risk of plants pollinated by biotic vectors, as has been found in other studies (Cunningham 2000). There are many mechanisms by which increased anthropogenic disturbance can alter plant-pollinator interactions, with habitat loss and increased fragmentation often a causal factor (Potts, Biesmeijer *et al.* 2010). The loss of suitable habitat can decrease the local abundance of pollinators (Grünewald 2010), increase the time taken for pollinators to move between plants (Hadley and Betts 2009) and through positive feedback, further reduce pollinator abundance as the quantity of pollen available for transport decreases (Memmott, Waser *et al.* 2004). The scale of habitat loss is well documented, and with the forecasted increase in human population density expected to further reduce and fragment natural landscapes (Sodhi, Kohi *et al.* 2004); coupled with the observed decline in important pollinators such as bees through viruses (Cameron, Lozier *et al.* 2011)

the dynamics of pollinator-plant and its influence on extinction risk is likely to remain an important factor.

Human population density has also been shown to interact with a plant's breeding system to influence its susceptibility to extinction (Girao, Lopes *et al.* 2007). This interaction proved to be the second most important factor in the model, accounting for 11.2% of the explained variance. It was found that species with unisexual breeding systems face a higher extinction risk than bisexual species as human population density increases. Once again, it is probable the link between increasing human population density and increased habitat loss and fragmentation influences this observation. With decreasing population sizes or increased fragmentation, species that require another individual to reproduce i.e. are unisexual, are likely to face a decrease in fitness and increased extinction risk in small populations (Heilbuth 2000).

The other important factor was the interaction between the climate and species' intrinsic traits, in particular the relationship between precipitation and possession of storage organ, and the interaction between temperature and the presence or absence of a seed endosperm. Both of these factors represent a species' ability to persist under unfavourable conditions; the ability to persist through desiccation caused by high temperature or low rainfall increases with the possession of a storage organ or a seed endosperm (Diaz and Cabido 1997). Given the expected increase in climatic stochasticity, particularly in relation to temperatures (New, Liverman *et al.* 2011) and the quantity and frequency of rainfall (Allan and Soden 2008), the ability to persist through unfavourable conditions is likely to become increasingly important. Furthermore, as a consequence of non-random extinction due to climatic filtering (Meynard, Devictor *et al.* 2011), the retention of only the species that can persist in a particular environment, the species present in communities may become less functionally diverse (Lôbo, Leão *et al.* 2011), resulting in the loss of not only groups of species but also disproportionately high quantities of functional and genetic diversity (Devictor, Mouillot *et al.* 2010).

### *Modeling plant extinction risk*

Studies have indicated that methods that effectively explain extinction risk in other groups, such as the development of explanatory models based on intrinsic traits are not suitable for plants (Davies, Smith *et al.* 2011, Knapp 2011). The relatively poor explanatory power of models in explaining the variance in plant extinction risk (Bradshaw, Giam *et al.* 2008, Sodhi, Koh *et al.* 2008) leads to the question - are plants different to other taxa in regard to the underlying causality of extinction risk? Or have previous models omitted important explanatory variables?

#### *Lineage age*

A recent study by Davies *et al.* (2011) showed threatened species are more common in young lineages that are diversifying quickly, suggesting the mode of plant speciation contributes to the observed non-random distribution of threat across taxonomic ranks. However, there was no indication of this in the results; both taxon age and the number of species per genera proved poor explanatory variables of extinction risk. Nevertheless, looking at the results from the variance component analysis, the variance in extinction risk was highest at the species level (0.98), and was relatively conserved within deep clades, supporting the findings of Davies *et al.* (2011) that plants undergo a “late-burst model” of evolution. However, the sample of monocot species used was relatively broad and widely dispersed across the monocot phylogeny. As a consequence, closely related species are rarely immediate phylogenetic “siblings”, therefore “recent events” in this analysis are relatively old, adding uncertainty to the findings in regards to the importance of taxon age.

#### *Traits*

Compared to other taxa the use of ecological and life history traits have not previously proved informative in explaining variation in plant extinction risk

(Bradshaw, Giam *et al.* 2008, Sodhi, Koh *et al.* 2008). When considered on their own, a finding confirmed in the present study; intrinsic traits explain only 11% of the observed variance. However this comparison is misleading for a variety of reasons as outlined below.

*Range size*

The results to which models of plant extinction risk have been compared often use range size as a variable, which has frequently been shown to be the most important variable across a number of taxa (Cardillo, Mace *et al.* 2005, Cardillo, Mace *et al.* 2008, Cooper, Bielby *et al.* 2008). However, range size is rarely used in plant-based models, as it is a primary method used to quantify plant risk (under Criteria B and D2) (Brummitt, Bachman *et al.* 2008); therefore its inclusion as a variable would be circular (Purvis *et al.* 2000a). Consequently, the potentially most informative predictor of plant extinction risk cannot be included, likely contributing to poor explanatory power of models of plant extinction risk. This is demonstrated by a study restricted to legumes not classified by Criterion B or D2, where the optimal models included range size and explained between 57% and 61% of the observed deviance (Bradshaw, Giam *et al.* 2008), percentages similar to the levels observed in models of mammal (Cardillo, Mace *et al.* 2005) and amphibian (Cooper, Bielby *et al.* 2008) extinction risk. This suggests that the technical difference behind the method of risk assessment, which results in an important variable being excluded, rather than a fundamental difference in the importance of traits contributes to the previously observed poor performance of plant models.

*Omission of important variables from plant based models*

As outlined above anthropogenic and climatic variables significantly contribute to a species risk of extinction, and when these variables were included the models explanatory power increased. These variables have been demonstrated to play

significant roles in explaining a significant proportion of variance in other taxa (Cardillo, Mace *et al.* 2008, Cooper, Bielby *et al.* 2008); however, they had previously not been incorporated into large-scale studies of plant extinctions. Interestingly, it was not the addition of the individual variables that contributed most to the increased explanatory power, rather it was the interaction between the traits and the climatic and anthropogenic factors that proved most significant.

Whilst the variance explained by the final model (29%) may not be as high as for other taxa, it is similar to the variance explained by mammal models independent of range size (27%). Demonstrating that when similar variables are used, the explanatory powers of plant models are not significantly different to those observed for other taxa.

*Habit*

It has been demonstrated that optimal models explaining extinction risk in mammals are region specific (Cardillo, Mace *et al.* 2008); however, as a consequence of the sedentary nature of plants and the significance of local variation, it is likely that due to variations in micro-climatic and soil nutrients, locality is *more* important in explaining patterns of plant extinction risk than for other groups. This was demonstrated by a study investigating patterns of plant extinctions in park grass communities in the United Kingdom. Where it was found the optimal models included complex three way interactions between local scale variation in extrinsic factors (pH, nitrogen) and life history traits (phenology, height, diaspore mass) (Freville, McConway *et al.* 2007). This was supported by the findings outlined above; the optimal model being relatively complex compared to other taxa; requiring interactions and traits from all categories (biological, climatic and anthropogenic) to explain similar levels of variance observed in other taxa. As a result it is likely plant models require more detailed local information to effectively describe variation in plant extinction risk than has been previously investigated.

Of course it is possible that the increased variance explained by increased models may be a consequence of adding more variables to the model resulting in a better model fit ($R^2$ value). However, the optimal models were selected based upon AIC and BIC scores, which penalise model complexity (Burnham and Anderson 2002).

## Conclusion

In conclusion the pattern of plant extinction *is* different to other taxa; however, the reason for the difference is complicated. It is likely the method of speciation and the age of the lineage are important variables; however, this is not at the exclusion of the importance of ecological, climatic and anthropogenic information, as demonstrated by three findings: Once climatic and anthropogenic factors, were included the models explained similar levels of variance as observed for other taxa. However, in the absence of range size as a predictor variable, the optimal models needed to be more complex compared to that of other taxa. Finally, given the sedentary nature of plants, greater significance is given to interactions between intrinsic biological traits and the local environment.

Of note to plant conservation is the most significant factors are interactions that include climate and human population density, both of which are likely to increase in the future. Therefore, given the current high percentage of plant species already at risk of extinction, further studies aimed at unraveling complex local scale causation of extinction are of great importance.

# Chapter 4: The relationship between IUCN ratings and evolutionary history

## Introduction

Plants provide a large number of services to the benefit of humans (Isbell *et al.* 2011), whilst influencing the stability and function of communities across a wide range of ecosystems (Tilman *et al*. 2006). Despite their relative importance, a large number of plant species are facing extinction (Royal Botanic Gardens Kew 2010). Consequently, conservation biologists are faced with "The agony of choice", a phrase coined by Vane-Wright *et al.* (1991), which reflects the challenge of prioritising which species should receive the most protection, a decision that is especially relevant today given many consider we have entered the 6$^{th}$ great mass extinction (Wake and Vredenburg 2008) with limited financial resources (McDonald-Madden *et al.* 2008).

One method used to prioritise species for conservation is the use of threatened species lists, such as the IUCN Red List (IUCN 2011), which provides the best available comparable estimates of species' extinction risk (Rodrigues *et al.* 2006). However, such lists do not take into account the fact that the extinction of a species not only removes it from the global species pool; it also removes a unique branch from the tree of life (Faith 1992). Due to different evolutionary pathways the loss of some species removes a greater proportion of unique evolutionary history (EH) than the loss of others (Faith 1992), potentially leading to a reduction in ecosystem productivity and community function to a greater extent than the loss of a species with a lower EH (Cadotte *et al.* 2008).

Combining species' conservation status with a measure of their evolutionary distinctiveness is therefore desirable. One method that has aimed to reconcile these two aspects is the EDGE (Evolutionarily Distinct and Globally Endangered) program, an index of urgency, that weights a species' evolutionary distinctiveness with its IUCN rating to generate a list of species that are both evolutionarily distinct and globally endangered (Isaac *et al.* 2007). Species can then be ranked according to their EDGE scores, a method that has been applied to a number of groups including mammals, coral reefs, amphibians and birds (The Zoological Society of London 2011). However, no such study exists for plants, due to the size of the group (an estimated 352,000 species (Paton *et al.* 2008), and the consequent absence of (a) a near-complete phylogeny, which is required in order to assign EDGE scores (Isaac *et al.* 2007) and (b) conservation status assessments for most species. Therefore, the most evolutionarily distinct plant species are not necessarily being identified as high priorities in existing conservation frameworks.

Although a complete phylogeny would be needed for a full EDGE assessment, using a partial phylogeny (for a sample of species) can still provide insight into the relationship between EH and IUCN ratings within a taxonomic group, such as the monocots. Patterns of plant extinction risk (Sodhi *et al.* 2008) and EH (Vamosi and Wilson 2008) have both been shown to be non-random with respect to phylogeny. Importantly, unlike mammal and bird species where species on long branches (i.e. those with high EH values) are particularly likely to be at risk (Mace *et al.* 2003), it has been suggested that the opposite pattern may hold for plants, due to differentiation in speciation and extinction patterns, resulting in a decrease in the loss of EH compared with a null model of random extinction (Schwartz and Simberloff 2001). However, a global-scale analysis using a family level supertree (Vamosi and Wilson 2008), provided some evidence that plant extinction risk is highest in small clades, as observed in other taxa (Purvis *et al.* 2000b, Isaac *et al.* 2007), resulting in elevated loss of EH compared with a null model of random extinction.

Studies of the distribution of EH at the level of plant families have thus come to different conclusions, and to date no study with plants has worked at the species level. A consequence of working at the family-level is a less complete picture of variation in EH values *between* families (Vamosi and Wilson 2008) and no indication of the variation of EH *within* families. Furthermore, in the absence of species-level EH values and conservation assessments for most species, no study has investigated the relationship between EH values and IUCN ratings; therefore it is unknown if the pattern observed in other taxa i.e. the uneven distribution of EH across Red List categories applies to plants, potentially resulting in many species with high EH values not benefiting from existing conservation projects (Isaac *et al.* 2007).

This gap in species-level knowledge was addressed using a randomly selected sample of 1500 monocot species (Brummitt *et al.* 2008), a group with a large diversity of life forms and a worldwide distribution, thereby allowing an understanding of how effective IUCN ratings are not only in highlighting extinction risk but also prioritising the conservation of EH.

The analysis is focused on two primary questions: (1) Are EH and extinction risk evenly distributed across the phylogeny of monocots, (2) What is the relationship between EH and extinction risk?

## Methods

### *Data collection*

The IUCN categories (IUCN, 2001) obtained from http://threatenedplants.myspecies.info, (accessed on 10/05/11) were used as a measure of extinction risk for the sample of monocot species. Species with the IUCN

designation DD and those without full assessments were not included in the analyses, resulting in threat codings for 871 species (Supplementary material 2).
After combining filtering out those species with genetic and IUCN data the analysis was conducted with 769 species (Supplementary material 2).

### *Analysis*

*Is evolutionary history and threat evenly distributed across the phylogeny?*

Due to the non-normal distribution of EH across families, generalised linear models with a gamma distribution were used, in which EH, the response variable, was coded as continuous whilst families were coded as factors. In order to determine which families differed significantly from others a post hoc test, "general linear hypothesis" (Hothorn *et al.* 2008), was conducted on those variables with significant P-values, using the packages multcomp and multcompView (Hothorn *et al.* 2008) in R (R, v.2.13.0; R Development Core Team, 2011). P-values were corrected using the "Holm method for multiple comparisons" (Aickin and Gensler 1996).

The same method was employed to test for the non-random distribution of threat, where a species' extinction probability was coded as a binomial response variable (Bradshaw *et al.* 2008, Sodhi *et al.* 2008) with Vulnerable (VU), Endangered (EN) or Critically Endangered (CR) species classified as threatened and Least Concern (LC) or Near Threatened (NT) species as non-threatened.

*What is the relationship between evolutionary history and threat?*

Two independent methods were used:

**1.** In order to investigate the relationship between EH and threatened or non-threatened species, extinction risk was coded as a binomial response variable, as described above. To investigate the relationship between EH and IUCN rating, the IUCN categories were coded on an ordinal scale from Least Concern (0) to Critically Endangered (4). Generalised linear models were then used to regress EH scores against each response variable (R, v.2.13.0; R Development Core Team, 2011) with a post hoc general linear hypothesis test (as described above) used on those with significant P-values.

2. A simulation method; iteratively, 100 species from each category (threatened or non-threatened) were removed at random and the EH remaining after each iteration calculated, the process repeated 1000 times. A chi-squared test was then used to test for a significant difference in the percentage of EH remaining between the two groups. This procedure was repeated, removing 100 species from each IUCN category, calculating the percentage of EH remaining in each IUCN category.

Results

***The distribution of evolutionary history and threat across the phylogeny***

The total amount of EH contained within the phylogeny was 15,126 My, with individual species scores ranging from 2.31 My (*Chondrorhyncha viridisepala*) to 68.14 My (*Cleistes ionoglossa*). Scores were approximately log-normally distributed, with a median of 13.83 My and geometric mean of 18.46 My.

However, EH was not evenly distributed across the phylogeny; seven families had significantly higher EH values than the remaining families (Table 4-1), whilst extinction risk was also clustered; three families had a significantly higher proportion of threatened species than other families (Table 4-1).

| Family | EH | Threat |
|---|---|---|
| Arecaceae | 0.003 | 0.050 |
| Bromeliaceae | 0.002 | 0.021 |
| Colchicaceae | 0.006 | NS |
| Convallariaceae | 0.004 | NS |
| Costaceae | 0.012 | NS |
| Liliaceae | 0.048 | NS |
| Melanthiaceae | 0.021 | NS |
| Zingiberaceae | NS | 0.059 |

**Table 4-1.** Families with significantly (P-values) more EH per species than the average and those with a significantly more threatened species than average. Non-significant values – NS.

***Evolutionary history and threat status***

EH was not evenly distributed among the Red List categories; instead it significantly increased with elevated threat ($F_1$, 783 = 15.36, p = 0.0001). Combined, threatened species (CR, EN and VU) had significantly higher ($F_1$, 783 = 10.43, p = 0.001) EH values than those non-threatened (NT and LC) (Table 4-2), a difference also observed in the simulation analysis with significantly more EH lost through the removal of

threatened species compared to non-threatened ($X^2$ = 4472, P <0.001, d.f. = 4084, chi-square test; Fig. 4-1).

Looking at individual Red List categories no significant difference in the quantity of EH among subcategories within threatened (VU, EN, CR) or non-threatened (NT and LC) was found. However, there was significantly more EH within CR than LC ($F_1$, 381 = 2.29, p = 0.0307) and NT ($F_1$, 96 = 1.05, p = 0.0471) categories (Table 4-2), a finding also observed in the simulation analysis with significantly more EH lost when CR species were removed than NT ($F_1$, 2166 = 600, p = < 2.2e-16) or LC ($F_1$, 3001 = 712, p = < 2.2e-16) (Fig. 4-2).

| | **IUCN CATEGORY** | | |
|---|---|---|---|
| | NT | LC | Non-threatened |
| EN | 0.05 | 0.04 | |
| CR | 0.04 | 0.03 | |
| Threatened | | | 0.001 |

**Table 4-2.** Significant (P-Values) differences in the quantity of EH between each IUCN category, and grouped.

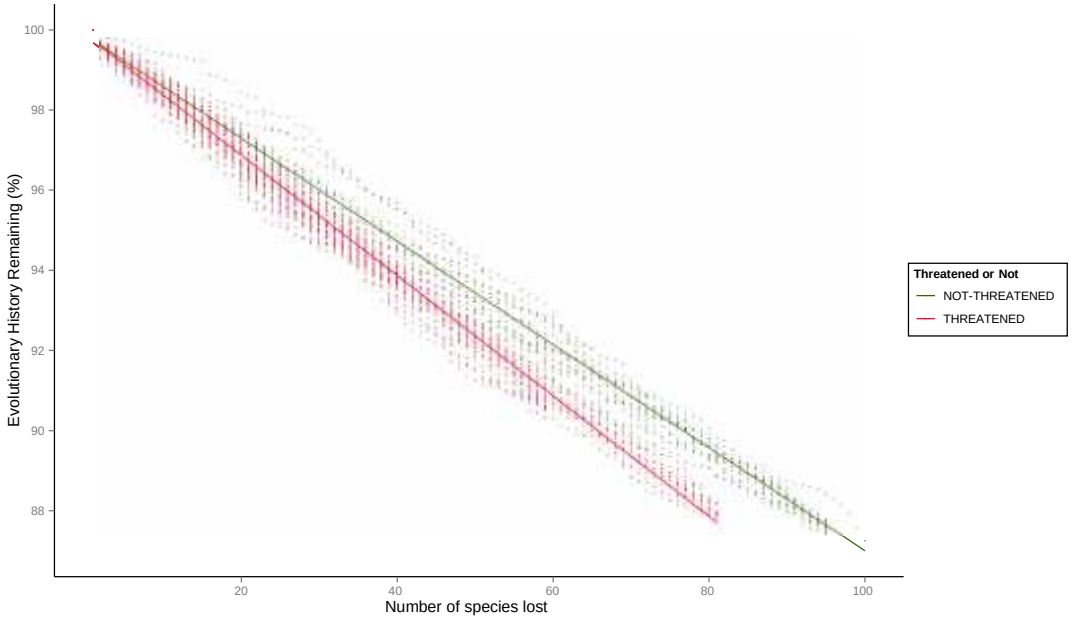


**Fig. 4-1.** The percentage of evolutionary history remaining in non-threatened and threatened species after the simulated loss of 100 species, 1000 times.

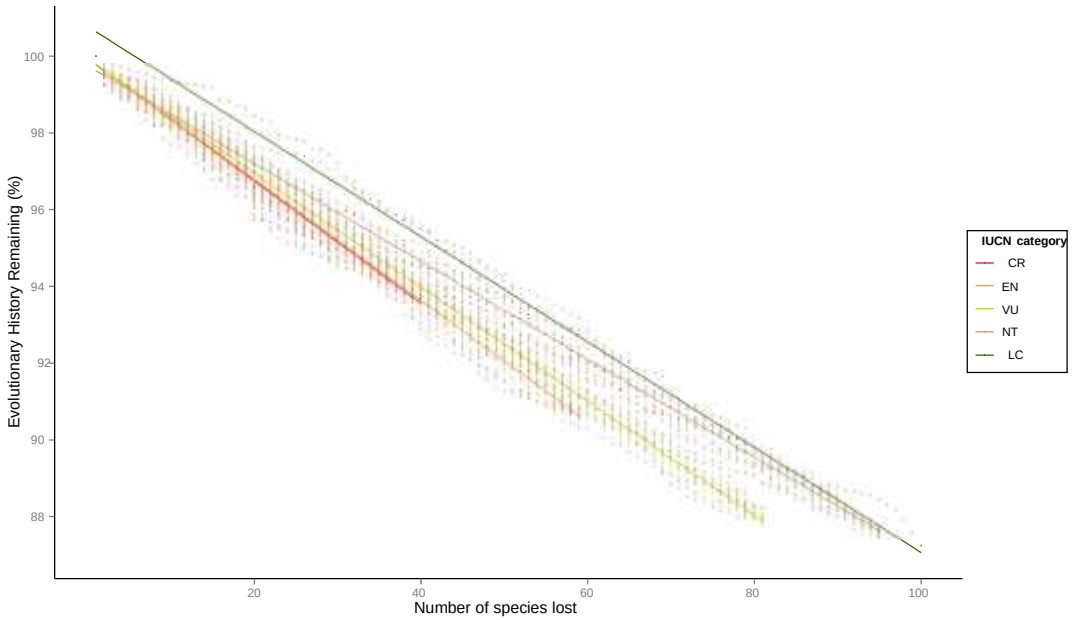


**Fig. 4-2.** The percentage of evolutionary history remaining in species from each IUCN category after the simulated loss of 100 species, 1000 times.

## Discussion

Nee and May (1997) studied the relationship between extinction and the loss of evolutionary history (EH) in a theoretical framework and showed that with the loss of 50% of species from a phylogenetic tree, 70-80% of the total EH still remained. However, this conclusion has since been shown to be optimistic (Purvis *et al.* 2000b), in part because the non-random relationship between extinction risk and phylogeny was not considered within the theoretical study. Phylogenies are non-random in shape and extinction risk is rarely random, but rather shows a strong phylogenetic component. Therefore, the degree of correlation between speciation rate and extinction risk determines how much EH will remain.

Using IUCN ratings as a proxy for extinction risk and a species level super-tree of monocotyledons, one of the major conclusions from this study is that the relationship previously found at family level for plants (Vamosi and Wilson 2008) and in other taxa (Purvis *et al.* 2000b, Isaac *et al.* 200) that EH is not randomly distributed across the IUCN categories but rather is significantly clustered within threatened groups – is supported at the species level for the monocots. This finding suggests that the loss of EH via extinction may be greater than the relatively optimistic predictions made by Nee and May (1997); however, it also suggests that prioritisation based solely on threat status may effectively capture EH (Purvis *et al.* 2000b).

The clustering of EH within threatened groups of monocots is likely due to the extreme tree imbalance seen in the angiosperm phylogeny (Davies *et al.* 2004). For example, within the phylogeny, species from the Bromeliaceae and Arecaceae families were found to have significantly more EH and threatened species than other families, a combination that likely contributed to the observed correlation between high EH values and extinction risk. The Bromeliaceae is a large neo-tropical family (58 genera, 3140 species) with eight subfamilies, and is extremely ecologically diverse (Givnish *et al.* 2011). The historical biogeography of the family shows a

relatively long time period (ca. 81My) between the origin of the family and the divergence into modern lineages, possibly due to high extinction rates within the family (Givnish *et al.* 2011), rendering the phylogenetic placement of the group a challenge due to its relatively isolated position, with no clear out-group (Givnish *et al.* 2011). A large number of Bromeliaceae species are threatened due to their habitat – specialised epiphytes growing upon isolated outcrops in Brazil often containing minerals such as limestone or iron – resulting in increased extinction risk from mining and habitat fragmentation (Jacobi *et al.* 2007, Versieux and Wendt 2007, Barbara *et al.* 2008, Leme *et al.* 2010).

Arecaceae, the palms, have an estimated 183 genera and 2400 species (Baker *et al.* 2011). As for the Bromeliaceae, there is difficulty in selecting appropriate out-groups, leading to difficulties in phylogenetic reconstruction (Stewart 1994). Part of this problem arises due to Arecaceae being a relatively old lineage of monocots, the first modern monocot family clearly represented in the fossil record (Stewart 1994). Furthermore, like the Bromeliaceae, Arecaceae underwent a rapid burst of evolution, but this time early in its history, and by ca. 60 (My) many modern-day palms had evolved (Stewart 1994). The fact that both of these families are relatively old lineages with rapid bursts of evolution likely contributes to their relatively high EH values compared to those of other families.

For the dataset as a whole EH increases moving from the LC category to higher levels of threat; whilst the EH of threatened species (taken collectively) was greater than that of non-threatened species (taken collectively). However, differences in EH between threatened species in different categories were not statistically significant; nor were differences in EH between LC and NT.

This lack of complete congruence between high EH and threat rating is replicated in all other taxa from the EDGE program, with 30% of the top 20 EDGE mammals classified as EN (a lower priority than CR), and 15% of the top 20 EDGE amphibians classified as EN or VU. Coral species are, to date, the taxonomic group that could potentially lose the most EH if prioritisation were based solely on IUCN categories, with 85% of the 13 species studied classified as EN and VU, and the top 4 EDGE species classified as VU (The Zoological Society of London 2011). Consequently, based on the findings of this study and on trends from other groups as outlined above, there is a high probability that the most evolutionarily distinct monocot species, such as members of the Bromeliaceae, are not being ranked highest by current IUCN criteria, potentially leading to a significant loss of EH.

In a botanical context, I advocate the development of larger, species-level supertrees with greater taxonomic scope, and increased research exploring the relationship between EH and IUCN criteria to further test and expand upon these findings. Furthermore, given the expected increase in species extinction, together with financial constraints to conservation, there needs to be continued development and practical implementation of methods for valuing species or areas for conservation, such as preferentially conserving evolutionarily distinct species, to ensure that species-level conservation efforts are prioritised most effectively.

# Chapter 5: The global distribution of phylogenetic and functional diversity in relation to environmental variables

## Introduction

Biodiversity is decreasing and shifting on a global scale due to anthropogenic activity and climate stochasticity; therefore understanding the processes underlying and regulating its distribution is of interest to conservation practitioners (Pimm, Russell *et al.* 1995). However, biodiversity is complex, incorporating many facets of variation, including phylogenetic (taxonomic) or functional (trait) diversity among species (Magurran 2004)**.** Species richness is commonly used as a surrogate for these other facets and as such used as a simple metric of biodiversity (Kerr 1997, Persha, Fischer *et al.* 2010). In recent years a number of studies have demonstrated that this surrogacy can break down, resulting in areas with disproportionate quantities of one facet relative to another (Vamosi, Heard *et al.* 2009).

However, what does the consideration of phylogenetic diversity or functional diversity offer to conservation that species richness does not? Phylogenetic diversity represents the phylogenetic relationships and evolutionary history between species and within communities, whilst effectively capturing the range of features present (Forest, Grenyer *et al.* 2007). As a consequence, phylogenetic diversity (hereafter PD) is potentially a useful conservation metric that enables the identification and prioritisation of groups of species or geographical areas that are phylogenetically diverse (Faith, Reid *et al.* 2004), therefore enhancing evolutionary potential for adaptation and hence long-term survival in an uncertain future (Forest, Grenyer *et al.* 2007).

Similarly, the quantity of functional diversity (hereafter FD) present in a community has been demonstrated to influence its stability in the face of environmental or anthropogenic stochasticity (Hooper, Chapin *et al.* 2005). Furthermore, FD influences the productivity of a community and is related to the provision of ecosystem services (Hooper, Chapin *et al.* 2005) resulting in communities with higher FD being both more productive and more resistant to change.

A key reason why PD and FD have not been more widely used is the difficulty of quantifying them. Unlike species richness, which is intuitive and relatively simple to quantify, measuring the phylogenetic relationships and the distribution of functional traits within a group of species occupying a particular area has proved challenging (Petchey and Gaston 2006). However, the rapid development of molecular tools has made the production of phylogenetic trees easier and faster, providing a framework from which phylogenetic relationships and PD can be quantified, of which Vane-Wright *et al.* (1991) and Faith (1992) were the earliest proponents. Functional diversity has previously been calculated after *a priori* classification of species into functional groups based on relatively arbitrary categories, such trophic level or food source, resulting in a discontinuous measure of functional relationships (Petchey and Gaston 2002). However by fitting species into multidimensional space based on a range of ecological characteristics the FD of a community can be quantified in a continuous and therefore more powerful measure (Petchey and Gaston 2002).

The integration of PD and FD into community ecology has provided a framework from which the processes regulating community assembly can be identified (Cavender-Bares, Kozak *et al.* 2009). The spatial distribution of PD and FD has been found to be non-random in several cases; some areas showing a disproportionate amount of PD or FD given the number of species present (Vamosi, Heard *et al.* 2009). Such deficits or excesses of PD or FD have been attributed to historical niche related-processes, suggesting the dominance of niche over neutral processes in the

assembly of natural communities (Wiens and Graham 2005). Species in communities have been found to be more similar functionally than expected given the number of species present, a finding often attributed to environmental filtering; a process that clusters functionally similar species due to the abiotic conditions of the area in which they are found (Fine and Kembel 2010). If these functional traits are phylogenetically conserved and therefore occur in closely related species, filtering will tend to cause closely related species to co-occur, resulting in species being more phylogenetically similar in a given area than expected by chance alone (Donoghue 2008).Alternatively, species in communities have been found to be less functionally or phylogenetically similar than expected given the number of species present (Cavender-Bares, Kozak *et al.* 2009). This pattern is often attributed to competitive exclusion, which limits the coexistence of closely related species resulting in species being less phylogenetically or functionally similar in a given area than expected by chance alone (Donoghue, 2008). However, these scenarios may be somewhat simplistic, as the spatial/taxonomic scale of the analysis and the model of trait evolution (conserved vs. convergent) has a large influence on the likelihood of detecting non-random patterns (Cavender-Bares, Keen *et al.* 2006).

Estimation of PD and FD has also allowed larger scale biodiversity patterns to be investigated, such as the PD patterns across the Cape of South Africa (Forest, Grenyer *et al.* 2007) or at national/supranational level (Devictor, Mouillot *et al.* 2010). At larger spatial scales, the effects of biotic interactions such as competition decreases; instead a combination of climatic and environmental filtering and large-scale biogeographic processes determine the species found (Cavender- Bares, Kozak *et al.* 2009). Variability in climatic factors such as temperature and precipitation along spatial gradients influence the spatial distribution of biodiversity (Gaston 2000), so much so that models have been developed that can effectively predict levels of species richness based on these climatic variables (Field, O'Brien *et al.* 2005). However, studies separating biodiversity into its constituent parts, have demonstrated that these metrics may be distributed differently from species richness, and to each other, along environmental gradients (Meynard, Devictor *et al.* 2011).

Physical gradients such as elevation have been shown to effect species richness, PD and FD in complex ways. For example, species richness is generally highest at mid-elevations before decreasing at high elevations, possibly due to a time-for- speciation' effect (TSE) that suggests that mid-elevations have been occupied the longest and will therefore have more species (Wiens, Ackerly *et al.* 2010). However Tallents *et al.* (2005) demonstrated that PD increased from mid-to-high elevation in the Usambara Mountains of Tanzania, whilst species richness decreased. This pattern was attributed to the independent occurrence in distantly related lineages of traits allowing persistence at high elevation (Tallents, Lovett *et al.* 2005). Furthermore, anthropogenic gradients, measured using proxies such as percentage of urbanisation have been shown to affect PD and FD independent of their effect on species richness (Barragan, Moreno *et al.* 2011).

Despite the number of studies investigating their relationships, the precise distribution of PD and FD along climatic, physical and anthropogenic gradients and the key variables affecting their quantity and spatial distribution are yet to be fully understood on a global scale. In this study the global distribution of phylogenetic and functional diversity in relation to climatic, physical and anthropogenic factors independently of species richness for a randomly-selected sample of 1500 monocot species that are the basis for the monocot component of the Sampled Red List Index (hereafter, SRLI) for plants (Brummitt, Bachman *et al.* 2008). The monocots are a good group to study such patterns due to their immense diversity, comprising some 60,000 species in 97 families, over a quarter of all flowering plants and their worldwide distribution, dominating significant parts of the world's ecosystems.

The analysis focused on three primary questions: (1) what are the climatic, physical or anthropogenic determinants of PD and FD and how do they differ from determinants of species richness? (2) How are PD and FD distributed along environmental gradients and how does this differ from species richness? (3) How do PD and FD relate to each other?

## Method

### *Data collection*

The spatial units used in the analysis were the 867 WWF ecoregions; subdivisions of the globe that are geographically and ecologically distinct (Olson, Dinerstein *et al.* 2001) and therefore considered to represent natural distributions of particular species assemblages more accurately than grid cell based approaches (Fritz and Purvis 2010). The ecoregions are at a relatively coarse scale, thus eliminating the problem of artificially forcing species into finer spatial scales. Despite their coarse scale they are, on the other hand, at a sufficiently small enough scale to allow the inference of global patterns (Fritz and Purvis 2010).

Range distribution for each of the monocotyledon species included in the SRLI were generated using herbarium records, resulting in range maps for 915 species, the range of the remaining 585 species were not inferred due to insufficient spatial data (Supplementary material 2). The monocot range maps were overlaid with the ecoregion shapefiles (http://www.worldwildlife.org/science/data/item1875.html, accessed January 2011) and the ecoregion occurrences for each species extracted. In total 637 occupied ecoregions were identified, of which 502 had more than one species (Supplementary material 6) and were subsequently used in the analysis (at least two species are required to calculate PD and FD) On average the ecoregions used contained 15 monocot species from the SRLI sample, ranging in number from 2 -78 (Supplementary material 6).

In order to model underlying causality of species distributions, a series of environmental variables were obtained for each ecoregion from the Terrestrial Ecoregions Base Global Dataset (http://www.worldwildlife.org/science/data/item1874.html, accessed January 2011) (Supplementary material 7) falling into three broad covariate groups; climatic, physical and anthropogenic (Table 3-2). All covariates were tested for significant co-linearity using variance inflation factors (VIF) to identify variables that could be dropped before starting the analyses due to redundancy. The remaining covariates were standardised to have a mean of 0 and a variance of 1.

The aim of the analysis on FD was to investigate potential environmental factors underlying its global distribution. The traits selected for this analysis were selected under 3 criteria – first, traits were selected from broad spectrum of characteristics related to species persistence, dispersal potential, species-species interactions and fitness. Secondly I used traits that have proved informative in previous studies of plant distributions and extinctions (see Chapter 1 and 3 and references therein)**.** Finally to ensure taxonomic and geographic representation, it was important the selection procedure did not exclude those rare species or which little is known about, therefore only traits for which I could gain information for the entire sample were included (Table 3-1).

Traits were compiled from taxonomic literature and expert advice (Supplementary material 4; Supplementary material 5); where species information was not available, species were assigned traits that were characteristic of the genus to which they belong.

After all available spatial, functional and phylogenetic data was obtained, 617 species remained that could be included in subsequent analyses (Supplementary material 2).

***Phylogenetic signal analysis***

In order to test for significant phylogenetic signal, λ, a measure of phylogenetic correlation (Pagel 1999) was used. The metric λ normally varies between (phylogenetic independence) and (species' traits covary in direct proportion to their shared evolutionary history). A variance component analysis (VCA) was also performed, in order to quantify how much of the variance in a trait can be explained by the distribution of a trait across taxonomic levels. These analyses were performed in R (v.2.13.0; R Development Core Team 2011) using functions from the package "APE" (Paradis 2004).

***PD and FD calculation***

Phylogenetic diversity is a continuous measure corresponding to the sum of branch lengths connecting a given set of species in a phylogenetic tree to assess species relatedness (Faith 1992). In order to calculate the PD for each ecoregion, the length of the branch leading to the root node of the minimum-spanning tree for species occurring within that ecoregion was calculated. The calculation of PD was performed in R (v.2.13.0; R Development Core Team 2011) using the method of Isaac *et al.* (2007).

Functional diversity was calculated using the method of Petchey and Gaston (2006) as follows: i) construction of a species-trait matrix; ii) conversion of species-trait matrix into a distance matrix; iii) clustering of the distance matrix into a dendrogram; and iv) calculating FD by summing dendrogram branch lengths of community species (Petchey and Gaston 2002). This method is particularly appropriate to this study as it accounts for covariance between traits (Petchey and Gaston 2002). As the trait matrix contains categorical variables, Gower's dissimilarity coefficient (Gower and Krzanowski 1999) was used to calculate the distance matrix, a method that is better suited for categorical variables (Gower and Krzanowski 1999). To produce the dendrogram the GFD script of Mouchet *et al.* (2008) was used, which selects the clustering algorithm that best fits the species distribution in the functional trait space by minimising the dissimilarity between the distance matrix and the ultrametric matrix of the functional tree. The unweighted pair group method using arithmetic

averages (UPGMA) was selected as the best cluster method for the subsequent analysis as it most often led to the highest cophenetic correlation.

***Statistical analysis***

*Modeling strategy*

Significant spatial autocorrelation was detected in the dataset through visual inspection of correlograms (Moran's I, p= 4.44E-16<0.001) validating the assumption of independently distributed errors, therefore a spatial simultaneous autoregressive modeling technique (SAR; spdep library in R v2.8.1), (Kissling and Carl 2008) was used. A weighted spatial neighbourhood matrix representing the spatial relationship between ecoregions was used as an error term in the SAR. The matrix was constructed based on Euclidean distance and coded by row-standardisation (the covariance based on the number of neighbours of each ecoregion in each row of the spatial weights matrix).

The accuracy of SAR models depend on the model type and the neighbour distance used. Therefore, they are not always more accurate than "simple ordinary least square models" (Kissling and Carl 2008). There are two forms of spatial autocorrelation; "inherent" where the autocorrelation is present in the response variable, and "induced", where autocorrelation is present in the predictor variables (Kissling and Carl 2008). The type of autocorrelation present determines the best model type to use. Spatial error, lagged and mixed SAR models were trialled on the data; the spatial error model was selected and used in all subsequent analyses, based on the lowest value of the Akaike information criterion (AIC). This model also performed better than simple and smoothed ordinary least squares OLS models. Several distances were tested to construct the spatial neighbourhood matrix and the distance of 750km was selected after visualisation of the spatial correlograms.

In order to test for significant non-linear effects, four weight functions were tested ($x^2$, $x^3$/, 1/x and 1/2x) for each response variable; 1/x was selected as it produced the AIC value in a spatial regression with no other predictors. Liner functions were selected as they produced the lowest AIC value in a spatial regression with no other predictors (Kissling and Carl 2008).

In order to ascertain which covariate group(s) best explains the variance in each metric; each covariate was concatenated to their representative group (climatic, anthropogenic and physical) and used as the predictor variables. Akaike weights ($w_i$) were calculated to estimate the relative support for each model; $w_i$ represents the probability that model *i* is the best model for the observed data, given the candidate set of models, and sums to 1 across all models considered (Burnham and Anderson 2002). The evidence ratio for the models was calculated, which provides an estimate of the evidence for one model over the other; ratios over 0.90 are considered strong for the support of one model over another (Burnham and Anderson 2002). In order to assess the relative importance of each covariate group, the AIC weights ($w_i$) for each model in which the covariate group appeared was summed (Burnham and Anderson 2002).

To find the combination of covariates resulting in the most likely model that explains the variation in each metric, a supervised forward selection procedure was adopted; single covariate models were tried first and the AICc calculated. The model with the lowest AIC value was retained and tried in combination with each of the remaining covariates, both as main effect terms and interaction terms. The better of the two-covariate models were then selected and tried with each of the remaining covariates, and this process was repeated until introducing another term into the model failed to yield a model with lower AICc.

Models with a difference in AICc of less than 5 (Burnham and Anderson 2002) were retained as candidate models from which Akaike weights ($w_i$) and evidence ratios were calculated. Once again, in order to assess the relative importance of each covariate group and individual covariate, the AIC weights ($w_i$) for each model in which the variable appeared was summed.

***The climatic, physical and anthropogenic drivers of species richness, PD and FD.***

The addition of a species to a region results in branches added to the resulting minimum spanning tree/dendrogram and therefore leads to an increase in the PD/FD measured in that area. This results in significant correlation between the species richness of an area and levels of PD ($R^2$, 0.95, p, 2.00E-16) and FD ($R^2$, 0.93, p, 2.00E-16). Therefore, to accurately understand the spatial distribution of PD and FD and to determine the climatic and anthropogenic correlates of this pattern, the effect of species richness must be removed, achieved by using species richness as a covariate in the SAR error models

***How do PD and FD relate to each other?***

In order to test the relationship between PD and FD, both indices were standardised to have a mean of 0 and a variance of 1 and the difference between them used as the response variable in the SAR models.

**Results**

Both methods used to account for the monotonic relationship between PD/FD and species richness resulted in the same predictors selected as significant; therefore, only the results from the covariate method are presented.

***The climatic, anthropogenic and physical drivers of species richness, PD and FD.***

The model of covariate groups that best explains the variation in species richness was composed of climatic and physical factors ($w_i$ = 0.52, evidence ratio to next best model = 1.3); climatic factors being the most important (0.92), followed by physical (0.64), with little support for anthropogenic factors (0.07) (Table 5-1). Similarly, the

model of covariate groups that best explains the variation in PD comprised of climatic and physical factors ($w_i$ = 0.66, evidence ratio to next best model = 1.7); climatic factors being the most important (0.94), followed by physical factors (0.71) with little support for anthropogenic factors (0.03) (Table 5-1). In contrast, FD was best explained by the model containing all covariate groups; climatic, anthropogenic and physical ($w_i$ = 0.29, evidence ratio to next best model = 1.4); climatic factors being the most important (0.49), followed by anthropogenic factors (0.41) and then physical factors (0.31) (Table 5-1).

The covariates that best modeled variation in species richness were:

Mean Elevation + STD Temp + STD Precipitation + Max Precipitation + Max Temp + Min PET ($w_i$ = 0.32, evidence ratio to next best model = 1.19) (Table 5-2)

The covariates that best modeled variation in PD were:

Mean Elevation + Min Temp + Max Temp + STD Temp + Min Precipitation ($w_i$ = 0.41, evidence ratio to next best model = 1.42) (Table 5-2).

Whereas the covariates that best modeled variation in FD were:

Mean Elevation + STD Temp + Min Precipitation + STD Precipitation - Urban + PET ($w_i$ = 0.21, evidence ratio to next best model = 1) (Table 5-2).

| **Model** | **Species Richness** | | | | **Phylogenetic Diversity** | | | | **Functional Diversity** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ΔAIC | $W_i$ | Importance | Rank | ΔAIC | $w_i$ | Importance | Rank | ΔAIC | $w_i$ | Importance | Rank |
| Anthropogenic | 9.7 | 0 | 0.07 | 7 | 10.7 | 0 | 0.03 | 7 | 1.9 | 0.11 | 0.41 | 4 |
| Climatic | 3.4 | 0.31 | 0.92 | 2 | 2.8 | 0.27 | 0.94 | 2 | 1.7 | 0.2 | 0.49 | 3 |
| Physical | 5.9 | 0.02 | 0.64 | 3 | 5.4 | 0.04 | 0.71 | 3 | 5 | 0.02 | 0.31 | 7 |
| Anthropogenic/Climatic | 6.6 | 0.02 | - | 4 | 7.8 | 0.01 | - | 4 | 1.6 | 0.21 | - | 2 |
| Anthropogenic/Physical | 7.9 | 0.01 | - | 5 | 8.2 | 0.01 | - | 5 | 2.6 | 0.08 | - | 5 |
| Full | 10.4 | 0 | - | 6 | 12.1 | 0 | - | 6 | 0 | 0.29 | - | **1** |
| Physical/Climatic | 0 | 0.52 | - | **1** | 0 | 0.66 | - | **1** | 2.7 | 0.08 | - | 6 |

**Table 5-1.** The models of covariate groups investigated to explain the variation in species richness, PD and FD together with the associated model summary statistics. ΔAIC - the difference in AIC from the best model, $w_i$ - Akaike weight (the relative support) and the importance of each covariate group.

| Covariate | Species Richness | | | | Phylogenetic Diversity | | | | Functional diversity | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Estimate | P-Value | Importance | Rank | Estimate | P-Value | Importance | Rank | Estimate | P-Value | Importance | Rank |
| Urban | - | - | - | - | - | - | - | - | -0.11 | <0.001 | 0.85 | 5 |
| Max Precipitation | 0.24 | <0.001 | 0.83 | 2 | - | - | - | - | - | - | - | - |
| Max Temp Mean | 0.08 | <0.001 | 0.7 | 5 | 0.08 | <0.001 | 0.68 | 6 | - | - | - | - |
| Elevation Min | -0.08 | <0.001 | 0.69 | 6 | 0.09 | <0.001 | 0.73 | 5 | -0.12 | <0.001 | 0.87 | 4 |
| Precipitation Min | - | - | - | - | - | - | - | - | 0.09 | <0.001 | 0.83 | 6 |
| PET | 0.37 | <0.001 | 0.87 | 1 | 0.3 | <0.001 | 0.89 | 1 | - | - | - | - |
| Min Temp | - | - | - | - | 0.18 | <0.001 | 0.79 | 4 | - | - | - | - |
| STD PET | - | - | - | - | - | - | - | - | 0.18 | <0.001 | 0.91 | 3 |
| STD Precipitation | 0.18 | <0.001 | 0.77 | 3 | 0.22 | <0.001 | 0.82 | 3 | 0.19 | <0.001 | 0.93 | 2 |
| STD Temp | 0.1 | <0.001 | 0.71 | 4 | 0.28 | <0.001 | 0.87 | 2 | 0.23 | <0.001 | 0.97 | 1 |

**Table 5-2.** The covariates included in the optimal model explaining the variation in species richness, PD and FD together with the associated summary statistic

Variation in species richness was best explained by variables associated with energy and water availability, all of which showed linear relationships. The minimum level of potential evapotranspiration was the most important ($w_i$ = 0.87), having a positive relationship to species richness whilst the next two variables were related to water; with species richness increasing with maximum precipitation of the wettest month ($w_i$ = 0.83) and with increasing variability in precipitation ($w_i$ = 0.77). Similarly, an increase in species richness resulted from increasing variability in temperature ($w_i$ = 0.71) and an increase in the maximum temperature of the maximum month ($w_i$ = 0.70). Finally, there was a negative relationship with elevation ($w_i$ = 0.69); however, this was not a linear relationship, in fact a quadratic (humped) relationship explained the variation better, with increasing species richness up to mid-elevation (1000-1200m), and decreasing species richness at higher elevations (Fig. 5-1).

As with species richness, variation in the quantity of PD was best explained by energy- and water-related variables; however, energy-related variables were more significant for PD than was observed for species richness. The minimum level of potential evapotranspiration was again the most important ($w_i$ = 0.89), having a positive, linear relationship with PD. The seasonal variation in temperature, the variability in precipitation, and the minimum temperature of the warmest month all had a positive, linear relationship to the observed variation in PD ($w_i$ = 0.87, $w_i$ = 0.82, and $w_i$ = 0.79, respectively). Overall PD had a positive relationship with elevation ($w_i$= 0.73); however, the relationship was not linear, instead a logistic curve best described the variation. PD increased with elevation, leveling off slightly at mid-elevation (1000-1200m); however, PD continued to increase beyond 1500m, beyond the mid-elevation maximum observed for species richness (Fig. 5-1). Finally PD showed a positive, linear relationship with the maximum temperature of the warmest month ($w_i$ = 0.68).

Variation in FD was best explained by variables associated with energy and water availability, more specifically with their variation. FD had a positive and linear relationship with variability in temperature ($w_i$ = 0.97), precipitation ($w_i$ = 0.93) and PET ($w_i$ = 0.91). Like species richness, FD showed a quadratic (humped) relationship with elevation ($w_i$ = 0.87), increasing up to mid-elevation (1000-1200m), and then decreasing at higher elevations (Fig. 5-1). A similar relationship was observed with urbanisation ($w_i$ = 0.85), with FD decreasing in ecoregions where urbanisation development exceeds 1500 km$^2$ (Fig. 5-2). The minimum level of precipitation of the driest month precipitation was also significant ($w_i$ = 0.83) in explaining the quantity of FD, showing a positive, linear relationship.

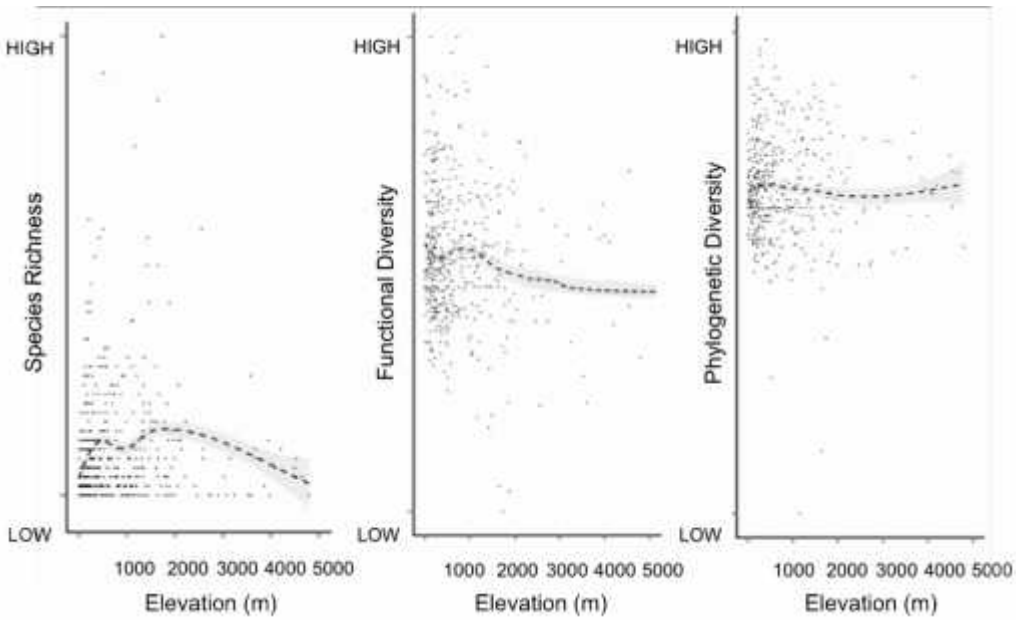


**Fig. 5-1.** The relationship between species richness, PD and FD with elevation.

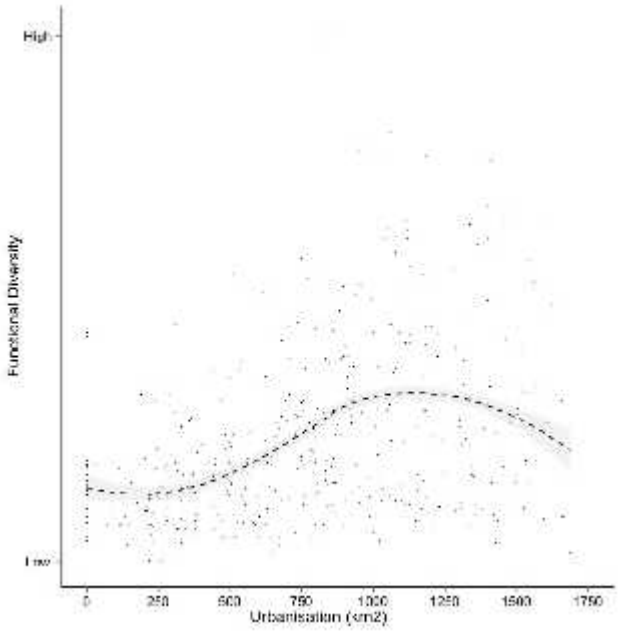


**Fig. 5-2.** The relationship between functional diversity and urbanisation.

There was no support to suggest that spatial factors having an influence in the distribution of FD or PD; biome, realm and latitude all showed no significant relationship with the quantity of either metric.

***How do PD and FD relate to each other?***

Independently of species richness, PD and FD were still significantly correlated with each other (p= 0.001); however, 69% of the variance in the observed difference between the two was still unexplained by this relationship.

The variation in PD and FD was best explained by the model containing all covariate groups, climatic, anthropogenic and physical ($w_i$ = 0.58, evidence ratio to next best model = 2.45); climatic factors being the most important ($w_i$ = 0.35) followed by anthropogenic factors ($w_i$ = 0.27) and then physical ($w_i$ = 0.22) (Table 5-3).

| Model | ΔAIC | $w_i$ | Rank | Support |
|---|---|---|---|---|
| Anthropogenic | 6.8 | 0 | 6 | 0.27 |
| Climatic | 6.2 | 0.31 | 5 | 0.35 |
| Physical | 9.1 | 0.02 | 7 | 0.22 |
| Anthropogenic/Climatic | 3.1 | 0.02 | 2 | - |
| Anthropogenic/Physical | 4.5 | 0.01 | 4 | - |
| Full | 0 | 0.58 | **1** | - |
| Physical/Climatic | 3.7 | 0.52 | 3 | - |

**Table 5-3.** The models of covariate groups investigated to explain the difference between PD and FD in relation to the covariate groups, together with summary statistics. Δ AIC - the difference in AIC from the best model, $w_i$ - Akaike weight (the relative support) and the importance of each covariate group.

Standardised differences in expected-observed values between PD and FD showed the model of covariates that most likely explaining the relationship was:

Urban + Mean elevation + Precipitation range + Min AET + Max Temp ($w_i$ = 0.58, evidence ratio to next best model = 2.45).

The most important variable explaining the difference between the metrics was maximum temperature ($w_i$ = 0.71), followed by elevation ($w_i$ = 0.69) and urbanisation ($w_i$ = 0.64). The variation in precipitation ($w_i$ = 0.37) and the minimum level of potential evapotranspiration ($w_i$ = 0.28) were also significant in explaining the variation between the PD and FD metrics (Table 5-4).

| Covariate | Estimate | P-Value | Importance | Rank |
|---|---|---|---|---|
| Urbanisation | 0.14 | <0.001 | 0.69 | 2 |
| Max. Temp | -0.11 | <0.001 | 0.64 | 3 |
| Mean Elevation | 0.21 | <0.001 | 0.71 | **1** |
| Min. PET | 0.02 | <0.001 | 0.28 | 5 |
| STD. Precipitation | 0.08 | <0.001 | 0.37 | 4 |

**Table 5-4.** The covariates included in the optimal model explaining the difference between PD and FD together with the associated model summary statistics and the AIC derived importance of each covariate**.**

Once again there was no support for spatial factors having an influence, with biome, realm and latitude showing no significant relationship with the observed difference between PD and FD values.

***Phylogenetic signal***

All traits used in the analysis have significant phylogenetic signal compared to Brownian motion, whilst the majority of the variation in the traits occurs at species/genus level, with the traits showing relatively little variation at the order/family level **(**Table 5-5).

| | | Variance | | | |
|---|---|---|---|---|---|
| **Trait** | **Signal (λ)** | **Order** | **Family** | **Genus** | **Species** |
| Breeding system | 0.99 | 0.20 | 0.70 | 0.96 | 0.98 |
| Endosperm | 0.96 | 0.32 | 0.45 | 0.89 | 0.96 |
| Floral symmetry | 0.93 | 0.13 | 0.45 | 0.96 | 0.98 |
| Fruit | 0.97 | 0.18 | 0.66 | 0.98 | 0.98 |
| Pollen dispersal | 0.99 | 0.48 | 0.56 | 0.97 | 0.98 |
| Storage organ | 0.85 | 0.00 | 0.12 | 0.87 | 0.96 |
| Woodiness | 0.61 | 0.26 | 0.31 | 0.94 | 0.98 |

**Table 5-5**. The phylogenetic signal (λ) for each trait and variance explained by the traits position across taxonomic levels.

## Discussion

The relationship between PD, FD and species richness has mostly been studied at the community level, where deviations in one metric from another can be used to infer processes regulating the assembly of communities (Webb 2000). The results have indicated that the scale of the analysis influences the type of community structuring observed, with smaller scale studies showing that biotic processes influence communities, whilst communities at larger (regional) scales are more influenced by environmental factors (Cavender-Bares, Keen *et al.* 2006). However, very few studies have investigated the global distribution of PD and FD and investigated how they relate to each other and species richness, or how these patterns are influenced by environmental factors (Meynard, Devictor *et al.* 2011).

Based on an analysis of representative sample of monocots, and consistent with the findings of other studies (Meynard, Devictor *et al.* 2011), it was found that the patterns used to describe the relationship between the environment and global patterns of species richness (O'Brien 1993) can generally be extended to PD and FD. Using spatial simultaneous autoregressive models, energy- and water-related variables proved to be significantly important variable categories in explaining the variation in the three facets of biodiversity. This is related to the relationship between water and energy, which are key in regulating photosynthesis and patterns of plant species richness (O'Brien 1993). Furthermore, species richness has been shown to vary with climate across biomes in a similar way, suggesting the relationships are not the result of pooling across biomes (Francis and Currie 2003). It is likely that the significance of these variables to general plant species patterns is contributing to the correlation between the three metrics.

Moreover, the relationship between species richness and PD and FD observed in this study was more complex, differing in the importance of the individual climatic variables. Whilst variation in species richness was described by variables linked to photosynthesis that have been previously demonstrated and effectively used in niche modeling, the other two metrics behaved different. Rather than responding to individual measures of PET, temperature and precipitation (e.g. the maximum or minimum), FD was primarily related to within year variation of energy and water. As variation in the variables across ecoregions increases the quantity of FD, the diversity of life history traits present, also increases, a finding attributed to the effect of environmental heterogeneity on FD. As the environment becomes increasingly heterogeneous, the number of niches and habitats available to plant species also increases (Dufour, Gadallah *et al.* 2006), therefore the quantity of FD present will be expected to be higher than in relatively stable and homogeneous environmental conditions.

It was found that the level of species richness to be significantly correlated with the maximum precipitation and temperature across ecoregions, agreeing with previous studies in alpine vascular plants (Moser, Dullinger *et al.* 2005) and the woody flora of South African (O'Brien 1993). However, FD was correlated with the *minimum* precipitation and temperature of the month rather than the *maximum*, a difference attributed to the environmental stress hypothesis (Bertness and Callaway 1994), which assumes that the pool of species available to colonise an area declines with increasing harshness of the environmental conditions. In this case species not adapted to low temperatures and precipitation levels are "filtered" out from the region, leaving species with traits adapted to the environment and therefore reducing the FD present (Zobel 1997). The ecoregions in which FD was observed to be low support this explanation, predominantly occurring in desert/xeric shrubland and montane grassland/shrubland biomes (low precipitation) and in the tundra and boreal forests / taiga (low temperature).

The theory that the species environment filters non-random selections of traits can be carried forward to explain the effect of anthropogenic variables on the metrics. No significant effect was found between anthropogenic variables and the quantity of species richness or PD, however beyond a threshold of 1500 $km^2$ of urbanised land within an ecoregion, FD significantly decreased. This finding is supported by a number of studies demonstrating that plant (Lôbo, Leão *et al.* 2011), bird (Devictor, Julliard *et al.* 2007), and fish (Clavero and Hermoso 2011) communities subjected to anthropogenic disturbance contain a non-random subset of species bearing particular life history traits. Urbanisation and habitat fragmentation have been shown to negatively affect a large range of plant functional traits disproportionally, namely seed size (Laurance, Ferreira *et al.* 1998, de Melo, Dirzo *et al.* 2006), shade tolerance (Laurance, Ferreira *et al.* 1998), pollination syndrome (Laurance, Nascimento *et al.* 2006) and breeding system (Girao, Lopes *et al.* 2007).

The mechanism behind this functional homogenisation of plant communities probably reflects a complex combination of biotic and abiotic processes; however the result is the filtering of species into “loser” and “winner” species based on life history traits (Smart, Thompson *et al.* 2006). Looking at the effect of urbanisation development on specific functional traits, a significant decrease in species with biotic pollinators and species with unisexual breeding systems was detected in ecoregions encompassing more than 1500 $km^2$ of urban development. As urbanisation increases the suitable habitat for plant species decreases and becomes increasingly fragmented. Therefore, the shifts in pollination syndrome and breeding systems likely reflects a decrease in biotic pollinators in edge-dominated small forest fragments and a shift to unisexuality as a breeding strategy as the size of the population decreases (Girao, Lopes *et al.* 2007). The threshold of 1500 $km^2$ can be explained when looking at the ecoregions most significantly affected by a decrease in FD in relation to urban development. They are similar in that they all belong to the tropical and subtropical moist broadleaf forest biome; a biome that has high levels of anthropogenic disturbance (Douglas, Wood *et al.* 2007). Furthermore, the majority are relatively small ecoregions, with a history of urban development and deforestation such as the Carolines tropical moist forests (WWF 2001); the relatively small size means that 1500 $km^2$ of development represents a significant area of the region.

In addition another significant difference between the metrics was the effect of elevation, with PD increasing while FD and species richness decreased. A characteristic mid-elevation hump in species richness (Kessler, Kluge *et al.* 2011) was found, a phenomenon that has been hypothesised to occur due to the time-for-speciation’ effect (TSE) that suggests that mid-elevations have been occupied the longest and will therefore have more species (Wiens, Ackerly *et al.* 2010). Beyond mid-elevations, environmental conditions shift, resulting in high elevation regions becoming increasingly extreme environments (Kessler, Kluge *et al.* 2011). As a consequence, species richness decreases as the number of species that can survive at high elevations decreases, due to physiological demands (Grau, Grytnes *et al.* 2007).

Through this mechanism of environmental filtering, species occurring in extreme environments possess a lower diversity of functional traits, which then allows them to persist at high elevations, resulting in the decrease in FD with elevation. Looking at the shift in functional traits in relation to increased elevation, there was a significant decline in species with biotic pollination syndromes and species with unisexual breeding systems above 1500m. It has been demonstrated that the pollination syndrome, or specifically the local pollinator community, can act as a habitat filter (Sargent and Ackerly 2008) with changes in the abiotic environment influencing the relationship between plants and their pollinators (Sargent and Ackerly 2008), favouring one syndrome over the other, particularly in montane environments (Kay, Reeves *et al.* 2005). In this study there appeared to be a shift from a biotic pollination syndrome to an abiotic pollination syndrome with increasing elevation due to the decrease in biotic pollinators at higher elevations, a pattern observed in other studies (Kay, Reeves *et al.* 2005). Further to the shift in pollination syndrome, and related to it, there was a shift in breeding system above 1500m with significantly more bisexual plant species occurring at higher elevations, which is likely to be a consequence of decreased pollinators and potential mates.

I have demonstrated how the distribution of FD and species richness can differ, resulting in areas with disproportional quantities of one over another in relation to environmental stress, habitat filtering and anthropogenic factors. However how does this impact upon PD? Specifically in relation to elevation, why does PD increase at high elevations when species richness and FD decreases? An explanation lies in the distribution of traits across the phylogenetic tree; specifically the phylogenetic signal and the results from the variance component analysis (VCA).

The traits that appear to most influence the spatial mismatch between species richness and FD are breeding system and pollination syndrome. These traits have relatively high phylogenetic signal ($\lambda$ =.  ), meaning they are not randomly distributed across

the phylogenetic tree. Instead, certain clades of the tree, due to shared evolutionary history, have significantly more of one trait state than the other – in other words, the traits are conserved. The results from the VCA indicate that trait conservation occurs at deep nodes in the phylogeny, the variance in breeding system at the order level being 0.12, compared to the variance at the species level of 0.98 (Table 7). Thus traits required to survive at high elevations are conserved at the order level while there is more variation at the species level, meaning the species which "pass" through the filter have come from disparate phylogenetic lineages, a phenomenon common in plant communities (Pennington and Dick 2004), which results in increased PD within the high elevation ecoregions (see also Tallents, Lovett *et al.* 2005).

High variation of trait states at the species level could be seen as inconsistent with the phylogenetic niche conservatism (PNC) hypothesis which predicts that there is a tendency for species to retain most of their ancestral characters, and as a consequence it is easier for species to track the environment than succeed in a new one (Donoghue 2008). However, the apparent inconsistency may simply be due to sampling intensity, for two reasons: Firstly, the sample of monocot species was relatively small and broadly dispersed across the monocot phylogeny, as a consequence, the most closely related species in the phylogeny are rarely immediate siblings and therefore —recent events in this analysis are relatively old compared to sibling-species events, meaning the two explanations maybe compatible but operating in different time scales (Vamosi, Heard *et al.* 2009). Secondly, and directly related to the first explanation, as the species-level splits seen here relate to older biogeographical events, if a complete species-level monocot phylogeny could be used we may see increased species level trait conservatism in line with other findings (Prinzing, Durka *et al.* 2001, Wiens, Ackerly *et al.* 2010), resulting in a decrease of observed levels of PD.

## Conclusion

In line with previous studies (Meynard, Devictor *et al.* 2011) it was found species richness, PD and FD vary in similar ways in response to environmental gradients. A finding explained by the significant effect of energy and water gradients on photosynthesis, and therefore on plant biodiversity. However, the observed correlation between the three metrics breaks down in situations of stress, either environmental or anthropogenic, with increasing stress resulting in the filtering of specific traits and a consequent homogenisation of trait states and as a result a decrease in FD. The relationship between PD and FD was a little more complex. In this study, the ecological traits displayed an ecological convergent model of evolution, resulting in a spatial mismatch between PD and FD; however in line with other studies if the traits showed high levels of conservatism, then PD and FD would be expected to respond in a similar way i.e. both diverge from species richness in stressful environments.

It is demonstrated that these three facets of biodiversity are likely to respond differently in relation to environmental and anthropogenically induced stress and believe this should be further investigated. This is especially pertinent given the expected increase in anthropogenic disturbance and increased climatic perturbation. Further investigation should be done using a full species level phylogeny where possible, using an increased scope of traits e.g. physiological. This approach will help improve the understanding of the spatial distribution of different facets of biodiversity in response to human activities and environmental extremes, and allow a more holistic approach to biodiversity management.

# Chapter 6: Concluding Remarks

## Summary of results and implications

In this thesis I have investigated why certain plant species are threatened with extinction whilst others are not; and, given limited resources, how we should value those that are. I also looked at this from a biogeographical perspective, investigating how areas can be valued and how well current measures of biodiversity, e.g. species richness, perform as a measure of the functional and phylogenetic diversity present.

The study group that formed the basis of this analysis was derived from Sampled Red List Index (SRLI); an index developed for a number of speciose groups in order to address the taxonomic bias in the coverage of IUCN assessments. Specifically I worked on the SRLI of monocots, a random sample of 1500 monocots from the estimated 60,000 known species. The monocots were selected as a study group due to their large diversity, global distribution and data availability compared to other groups. These questions have been investigated using a broad spectrum of factors – anthropogenic, climatic and physical – all demonstrated to influence the distribution of extinction risk, thereby increasing the understanding of underlying extinction processes and how conservation efforts should be prioritised.

Consistent with previous studies, a non-random pattern of extinction risk was found (Chapter 3) confirming that the current extinction crisis is targeting species with certain characteristics (Bennett and Owens 1997, Cardillo *et al.* 2005, Bradshaw *et al.* 2008, Brook *et al.* 2008, Sodhi *et al.* 2008). Four variables emerged as consistent predictors of extinction risk across taxa and threat types, with previous studies

reporting similar correlations, namely:

**1.** The mode of pollen dispersal, with biotically pollinated species facing higher extinction risk.

**2.** The number of seeds in the fruit, with species whose fruit contains fewer seed facing higher extinction risk than those with numerous seeds.

**3.** The species altitude, with species at higher altitudes facing higher extinction risk than those at low altitudes.

**4.** The habitat of the species, with epiphytes facing higher extinction risk than terrestrial species.

In modeling the variance in extinction risk, it was found although these factors had a significant effect, the variance they explained for the monocots was relatively low (~ 11%), compared to that of non-plant taxa (~30%), a finding that has been replicated in other studies (Bradshaw *et al.* 2008, Sodhi *et al.* 2008) and recently has led people to question the ability to model patterns of plant extinction risk (Davies *et al.* 2011). However, once environmental factors i.e. anthropogenic and climatic variables were included the explanatory power of the models increased. Interestingly, adding these variables as independent factors resulted in a marginal increase in explanatory power; but when interaction terms between the traits and the environment were included the explanatory power (~29%) became comparable with values observed in other taxa (Cooper *et al.* 2008). Through the findings outlined here it appears rather than plants being difficult to model, it was more a matter that the right questions were not being asked, resulting in important variables and interactions being omitted from previous studies.

Likely as a consequence of plants sedentary habit, their environment more immediately influences them than mobile taxa, and once this has been taken into account, the power of the plant models increases. Furthermore, although not directly tested here, based on my findings and those of others (Freville *et al.* 2007), I would hypothesise that plant extinction risk is driven more by localised heterogeneity in environmental factors, i.e. fine scale patterns of nutrient availability, than other taxa. Consequently, to develop these models further I believe we need to complement global scale analyses with localised studies in order take into account regional and biological heterogeneity and variation. Despite this, the global approaches I have taken in this thesis have proven powerful in expanding our knowledge of patterns in plant extinction risk whilst providing potentially fruitful avenues for future research.

Chapter 4 of my thesis highlights that under current patterns of extinction risk significantly more evolutionary history (EH) will be lost than would be expected if extinction were random i.e. there appears to be significantly more EH within threatened species than those non-threatened. This is due to the non-random distribution of threat and EH across the monocot phylogeny, with species on long branches more prone to extinction than those on shorter branches, a pattern found across the whole phylogeny. This finding was particularly apparent in members of the Bromeliaceae and Arecaceae, a consequence of both families having disproportionately high EH values due to both families splitting from the major trunk of the monocots relatively early and having high “speciation bursts”, coupled with both families having a relatively high percentage of threatened species, perhaps due to a number of endemic species from both groups occurring in close proximity to areas of high human disturbance.

The most frequently used index to prioritise species for conservation is the IUCN Red List, however this approach makes no allowances for phylogenetic distinctiveness of

species. For example, beyond saying a species is critically endangered or vulnerable it does not take into account how unique the species is or what would be lost if one critical species became extinct rather than another.

Given significantly more EH is likely to be lost with current extinction patterns than under a random scenario, it would be pertinent to combine conservation status (i.e. IUCN classification) with a measure of EH, thereby weighting species for conservation. However, no significant difference was observed in EH values *within* the IUCN threatened categories, resulting in threatened species with high EH values not being discriminated from those threatened species with low values. This has been found to be the case in all the taxa studied to date (The Zoological Society of London 2011), with coral species being the group that would lose the most EH if prioritisation were based solely on IUCN categories with 85% of the 13 species studied classified as EN and VU, i.e. not given the highest IUCN ranking (The Zoological Society of London 2011).

However, in the absence a full phylogeny it is not possible to explicitly test how many, or which species are, under a EH framework, not receiving sufficient conservation priority. Nevertheless, the study is the first to date that has looked at the relationship between a plant species EH value and its IUCN rating, and what it achieves is highlighting the fact that the findings with plants are consistent with other taxa, making this a priority area for future research.

Another frequently used method by conservation practitioners is the prioritisation of one area over another, the quantity of biodiversity present is often the currency used for discrimination. This thesis explored the use of species richness as a surrogate for biodiversity and its different facets, namely functional diversity (FD) and phylogenetic diversity (PD). In accordance with other studies it was found the

surrogacy generally corresponds, with the different diversity measures responding in similar ways to the environment (Devictor *et al.* 2010). This was due to the overriding importance of energy-and-water related variables in determining the global distribution of plant species (Kreft and Jetz 2007), resulting in high correlations between the different facets.

Moreover, discrepancies between species richness, FD and PD were identified, challenging the claim of interchangeability of different biodiversity measures. The results suggest that environmental filtering limits species occupying a particular habitat to those that are pre-adapted, and thus functionally similar, which would result in disproportionately low values of FD in a number of regions compared to the number of species present. This filtering is attributed to two underlying causes, each of which has been demonstrated in previous studies to cause functional homogenisation, but never over a global scale;

1. *Climatic stress*. In areas where the average monthly temperature or levels of precipitation were below a critical level, FD was significantly lower than the level of species richness. Species were filtered from these regions of climatic stress due to a number of traits, for example absence of an underground storage organ or seed endosperm.

2. *Anthropogenic stress*. When the area of agriculturally developed land within an ecoregion passed a threshold, FD significantly decreased, with the species present in these disturbed habitats generally having bisexual breeding systems and abiotic pollination vectors. On the other hand, there was no detectable effect on the level of species richness or PD in relation to the degree of agricultural development.

A pattern of phylogenetic over-dispersion (i.e. higher PD than expected given the number of species present) was detected at high elevations can be attributed to filtering of traits that have evolved in a number of orders through a convergent evolution, resulting in communities at high elevations being composed of a number of evolutionary distinct lineages.

The significance of these results is two-fold: Firstly, it is the first global analysis of plant biodiversity aimed at unraveling the mechanisms regulating the observed patterns of biodiversity taking into account more than the diversity of species present. Second, the mechanisms that have caused the interchangeability of species richness, FD and PD as measures of biodiversity to breakdown, i.e. climatic extremes and anthropogenic disturbance, are predicted to increase in terms of frequency and severity in the future.

Clearly, studying areas where discrepancies in different diversity measures occur can provide us with the means to understand what are the central causes underlying variation in biological diversity.

**Potential bias**

One shortcoming intrinsic to all analyses of extinction risk as identified using the IUCN Red List (Chapter 3) is that many species are categorised as Data Deficient (DD). I had to exclude such species (289) from my analyses, which potentially removed a biased subset of biodiversity; as species are often categorised as DD because of insufficient data, due to there rareness, and thus they are more likely to be threatened. If a DD species turn out to be mostly threatened, the detected correlations of risk may underestimate or overestimate the true importance of these factors in determining extinction risk.

Furthermore, if the distribution of DD species across the phylogeny is not taxonomically random then a further bias will be introduced. However looking at the percentage of families of which there are IUCN ratings and those with DD ratings I found no difference in the taxonomic "make up" of the two lists: Looking at the top 5 families (which make up over 80% of the species) the order is the same and the percentages comparable – Orchidaceae(26.6% – IUCN / 35.3% – DD); Poaceae (19.3% – IUCN / 11.6% – DD); Bromeliaceae (13.3% – IUCN / 10.8% – DD) Cyperaceae (13.1% – IUCN / 7.5% – DD) and Araceae (6.1% – IUCN / 4.7% – DD)

A possible solution to the problem of DD species for future studies could be multiple imputation (substitution) of the missing values for extinction risk. In practice, this is likely to be difficult as the imputation for risk needs to be phylogenetically explicit, considering the strong phylogenetic signal in extinction risk and threat types. Additionally, many DD species may be so poorly known that their biological trait values will be missing as from record as well.

The problem of missing data can be extended to the others chapters (4 and 5) in particular in the calculation of EH, where due to the methods used to calculate EH, missing species from the phylogeny could potentially inflate the EH scores of the close relatives present (Isaac *et al.* 2007). However the phylogeny used contains taxa that are sampled randomly with respect to the complete phylogeny, a factor that will minimise this bias. If the sampling method was non-random, e.g. one representative of each subgenus was selected, this would lead to over-dispersed sampling and would raise the type I error (Pybus and Harvey 2000). On the other hand, if taxa are excluded because they exhibit a trait (such as rarity or geographical location) that is correlated with phylogeny, the sampled set will be under-dispersed raising the type II error (Pybus and Harvey 2000). Crucially, where there is bias due to sampling it will be consistent throughout the phylogeny, therefore any inflation will have no taxonomic, geographic or IUCN rating bias. This is due to the large sample size used, where the sample analysed has been shown, through simulation, to be representative of the complete taxonomic group in terms of species richness per order and biogeographic realm (Baillie *et al.* 2008).

As is typical for complex, global analyses, there are many other possible sources of error. However, only large-scale, complex analyses as presented here can give insight into general patterns and processes. Furthermore, a trade-off exists; if too much time is spent collecting data before making decisions, we may stand to miss opportunities for conservation of biodiversity, particularly if there is on-going loss of habitat.

Conversely, not enough data may lead to poor decisions. However, in the context of dynamic and ever-changing world, decisions are required, and waiting for the "complete" dataset before taking any action would, in my opinion, be naïve.

**Conclusion**

Plants have colonised much of the land's surface, resulting in high numbers of species with amazing diversity. However, despite their importance to a large number of the earth's biological processes, there are an ever-increasing number of documented mechanisms leading to their extinction. In spite of their importance, processes resulting in the extinction risk for plant species and ways to counter them are not as other well studied as smaller, more charismatic groups, such as mammals.

Given the number of species studied and the global scope of the analysis, there are a number of potential conservation implications resulting from this thesis. Firstly, the key drivers of plant extinction risk are not that different from other taxa. Secondly, given the wealth of studies suggesting so, anthropogenic disturbance is plays a key role in plant extinction risk. However future studies trying to predict future risk levels for terrestrial plants should be more locally-focused than for other groups, given the sedentary nature of plants and the importance of local scale environment:plant interactions.

Conservation efforts need to focus on preserving ecosystem function. The lack of equivalence between different measures of biodiversity under a range of scenarios

suggests that conserving areas with high numbers of threatened species is not

necessarily the most effective for global biodiversity conservation. My results strongly suggest that human actions have selectively caused declines in a number of species, pruning unique evolutionary history from the tree of life whilst homogenising the trait states of the species remaining. However, with a combination of global and local scale analysis, and the increasing recognition of how vital plants are not only to the environment but also to the future welfare of humankind, there is still hope, partly influenced by future academic research, that it may not be too late to save significant parts of plant global biodiversity.

# References

Akçakaya, H. R. *et al.* 2000. Making consistent IUCN classifications under uncertainty. *Conservation Biology* **14**:1001-3.

Allan, R. P. and B. J. Soden. 2008. Atmospheric warming and the amplification of precipitation extremes. *Science* **321**:1481-1484.

Anderson, C. J. and T. Janβen. 2009. Monocots. Pages 3-212 *in* S. B. Hedge and S. Kumar, editors. The timetree of life. Oxford University Press.

Arbogast, B. S. *et al.* 2002. Estimating divergence times from molecular data on phylogenetic and population genetic timescales. *Annual Review of Ecology and Systematics* **33**:707-740.

Baillie, J. E. M. *et al.* 2004. IUCN Red List of threatened species. A global species assessment. IUCN, Gland, Switzerland and Cambridge.

Baillie, J. E. M. *et al.* 2008. Toward monitoring global biodiversity. *Conservation Letters* **1**:18-26.

Baker, W. J. *et al.* 2011. Phylogenetic relationships among Arecoid palms (Arecaceae: Arecoideae). *Annals of Botany* **33**:707-740.

Barbara, T. *et al.* 2008. Within-population spatial genetic structure in four naturally fragmented species of a neotropical inselberg radiation, *Alcantarea imperialis*, *A. geniculata*, *A. glaziouana* and *A. regina* (bromeliaceae). *Heredity* **101**:285-296.

Baroux, C. *et al.* 2002. Evolutionary origins of the endosperm in flowering plants. *Genome Biology* **3**:1-5.

Bates, D. and D. Sarkar. 2007. Lme4: Linear mixed-effects models using s4 classes.

Baum, D. A. *et al.* 1998. Biogeography and floral evolution of baobabs (*Adansonia*, *Bombacaceae*) as inferred from multiple data sets. *Systematic Biology* **47**:181-207.

Bell, C. D. and M. J. Donoghue. 2005. Dating the Dipsacales: comparing models, genes, and evolutionary implications *American Journal of Botany* **92**:284-296.

Bennett, P. M. and I. P. F. Owens. 1997. Variation in extinction risk among birds: Chance or evolutionary predisposition? *Proceedings of the Royal Society of London. Series B: Biological Sciences* **264**:401-408.

Benson, D. A. *et al.* 2007. GenBank. *Nucleic Acids Research* **35**:D21-D25.

Bertness, M. D. and R. Callaway. 1994. Positive interactions in communities. *Trends in Ecology and Evolution* **9**:191-193.

Bidartondo, M. I. 2008. Preserving accuracy in GenBank. *Science* **319**:1616.

Bininda-Emonds, O. R. P. 2004. The evolution of supertrees. *Trends in Ecology and Evolution* **19**:315-322.

Bininda-Emonds, O. R. P. *et al.* 2002. The (super)tree of life: Procedures, problems, and prospects. *Annual Review of Ecology and Systematics* **33**:265-289.

Blackburn, T. M. and R. P. Duncan. 2001. Establishment patterns of exotic birds are constrained by non-random patterns in introduction. *Journal of Biogeography* **28**:927-939.

Boswell, G. P. *et al.* 1998. Habitat fragmentation, percolation theory and the conservation of a keystone species. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **265**:1921-1925.

Bradley, A. P. 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition* **30**:1-1159.

Bradshaw, C. *et al.* 2008. Threat or invasive status in legumes is related to opposite extremes of the same ecological and life-history attributes. *Journal of Ecology* **96**:869-883.

Bremer, B. *et al.* 2003. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG II. *Botanical Journal of the Linnean Society* **141**:399-436.

Bremer, K. 2000. Early cretaceous lineages of monocot flowering plants. *Proceedings of the National Academy of Sciences* **97**:4707-4711.

Brook, B. W. *et al.* 2008. Synergies among extinction drivers under global change. *Trends in Ecology and Evolution* **23**:453-460.

Brummitt, N. *et al.* 2008. Applications of the IUCN Red List: Towards a global barometer for plant diversity. *Endangered Species Research* **6**:127-135.

Burnham, K. P. and D. R. Anderson. 2002. Model selection and multi-model inference. 3rd edition. Springer.

Burnham, K. P. and D. R. Anderson. 2004. Multimodel inference. *Sociological Methods and Research* **33**:261-304.

Butchart, S. H. M. *et al.* 2005. Using Red List indices to measure progress towards the 2010 target and beyond. *Philosophical Transactions of the Royal Society B: Biological Sciences* **360**:255-268.

Cadotte, M. W. *et al.* 2008. Evolutionary history and the effect of biodiversity on plant productivity. *Proceedings of the National Academy of Sciences* **105**:17012-17017.

Cameron, S. A. *et al.* 2011. Patterns of widespread decline in North American bumble bees. *Proceedings of the National Academy of Sciences* **108**:662-667.

Cardillo, M. *et al.* 2008. The predictability of extinction: Biological and external correlates of decline in mammals. *Proceedings of the Royal Society B: Biological Sciences* **275**:1441-1448.

Cardillo, M. *et al.* 2004. Human population density and extinction risk in the world's carnivores. *PLoS ONE* **2**:0909-0914.

Cardillo, M. *et al.* 2005. Multiple causes of high extinction risk in large mammal species. *Science* **309**:9-1241.

Cardillo, M. *et al.* 2006. Latent extinction risk and the future battlegrounds of mammal conservation. *Proceedings of the National Academy of Sciences of the United States of America* **103**:4157-4161.

Cavender-Bares, J. *et al.* 2006. Phylogenetic structure of Floridian plant communities depends on taxonomic and spatial scale. *Ecology* **87**:109-.

Cavender-Bares, J. *et al.* 2009. The merging of community ecology and phylogenetic biology. *Ecology Letters* **12**:693-715.

CBD (2002). Strategic plan for the conventionon biological diversity. URL http://www.cbd.int/2010-target/about.shtml

CIESIN. 2004. Global grids of gross domestic product (GDP) density, [online database]. Center for International Earth Science Information Network (CIESIN), Socioeconomic Data and Applications Center (SEDAC), Columbia University, Palisades, USA.

CIESIN and CIAT. 2005. Gridded Population of the World Version 3 (GPWV3), [online database]. Center for International Earth Science Information Network (CIESIN), Centro Internacional de Agricultura Tropical (CIAT), Socioeconomic Data and Applications Center (SEDAC). Columbia University, Palisades, USA.

Clavero, M. and V. Hermoso. 2011. Reservoirs promote the taxonomic homogenization of fish communities within river basins. *Biodiversity and Conservation* **20**:41-57.

Cooper, N. *et al.* 2008. Macroecology and extinction risk correlates of frogs. *Global Ecology and Biogeography* **17**:211-221.

Cunningham, S. A. 2000. Depressed pollination in habitat fragments causes low fruit set. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **267**:1-1.

Currie, D. J. and A. P. Francis. 1999. Some general propositions about the study of spatial patterns of species richness. *Ecoscience* **6**:392-399.

Davies, T. J. *et al.* 2004. Darwin's abominable mystery: Insights from a supertree of the angiosperms. *Proceedings of the National Academy of Sciences of the United States of America* **101**:1904-1909.

Davies, T. J. *et al.* 2011. Extinction risk and diversification are linked in a plant biodiversity hotspot. *PLoS Biol* **9**:e1000620.

De Melo, F. P. L. *et al.* 2006. Biased seed rain in forest edges: Evidence from the Brazilian Atlantic forest. *Biological Conservation* **132**:50-60.

Devictor, V. *et al.* 2007. Functional homogenization effect of urbanization on bird communities. *Conservation Biology* **21**:741-751.

Devictor, V. *et al.* 2010. Spatial mismatch and congruence between taxonomic, phylogenetic and functional diversity: The need for integrative conservation strategies in a changing world. *Ecology Letters* **13**:1030-1040.

Diaz, S. and M. Cabido. 1997. Plant functional types and ecosystem function in relation to global change. *Journal of Vegetation Science* **8**:463-474.

Diaz, S. and M. Cabido. 2001. Vive la differrence: Plant functional diversity matters to ecosystem processes. *Trends in Ecology and Evolution* **16**:646-655.

Donoghue, M. J. 2008. A phylogenetic perspective on the distribution of plant diversity. *Proceedings of the National Academy of Sciences* **105**:19-11555.

Douglas, E. *et al.* 2007. Policy implications of a pan-tropic assessment of the simultaneous hydrological and biodiversity impacts of deforestation. *Water Resources Management* **21**:211-232.

Dufour, A. *et al.* 2006. Plant species richness and environmental heterogeneity in a mountain landscape: Effects of variability and spatial configuration. *Ecography* **29**:573-584.

Dunn, R. R. *et al.* 2009. The sixth mass coextinction: Are most endangered species parasites and mutualists? *Proceedings of the Royal Society B: Biological Sciences* **276**:3037-3045.

Erwin, D. H. 2008. Extinction as the loss of evolutionary history. *Proceedings of the National Academy of Sciences* **105**:10-17.

Faith, D. 1992. Systematics and conservation: On prediciting the feature diversity of subsets of taxa. *Cladistics* **8**:361-373.

Faith, D. 1996. Conservation priorities and phylogenetic pattern. *Conservation Biology* **10**:1286-1289.

Faith, D. *et al.* 2004. Integrating phylogenetic diversity, complementarity, and endemism for conservation assessment. *Conservation Biology* **18**:255-261.

Felsenstein, J. 1985. Phylogenies and the comparative method. *The American Naturalist* **125**:1-15.

Field, R. *et al.* 2005. Global models for predicting woody plant richness from climate: Development and evaluation. *Ecology* **86**:2263-2277.

Fine, P. V. A. and S. W. Kembel. 2010. Phylogenetic community structure and phylogenetic turnover across space and edaphic gradients in western amazonian tree communities. *Ecography* **34**: 552–565.

Fleishman, E. *et al.* 2000. A new method for selection of umbrella species for conservation planning. *Ecological Applications* **10**:569-579.

Forest, F. *et al.* 2007. Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature* **445**:757-760.

Francis, A. P. and D. J. Currie. 2003. A globally consistent richness-climate relationship for angiosperms. *The American Naturalist* **161**:523-536.

Freville, H. *et al.* 2007. Prediction of extinction in plants: Interaction of extrinsic threats and life history traits. *Ecology* **88**:2662-2672.

Friis, E. M. *et al.* 2004. Araceae from the early Cretaceous of Portugal: Evidence on the emergence of monocotyledons. *Proceedings of the National Academy of Sciences of the United States of America* **101**:16565-16570.

Fritz, S. A. and A. Purvis. 2010. Phylogenetic diversity does not capture body size variation at risk in the world's mammals. *Proceedings of the Royal Society B: Biological Sciences* **277**:2435-2441.

Gandolfo, M. A. *et al.* 2002. *Triuridaceae* fossil flowers from the upper Cretaceous of New Jersey. *American Journal of Botany* **89**:1940-1957.

Gaston, K. J. 2000. Global patterns in biodiversity. *Nature* **405**:220-227.

Girao, L. C. *et al.* 2007. Changes in tree reproductive traits reduce functional diversity in a fragmented Atlantic forest landscape. *PLoS ONE* **2**:e908.

Givnish, T. J. *et al.* 2011. Phylogeny, adaptive radiation, and historical biogeography in *Bromeliaceae*: Insights from an eight-locus plastid phylogeny. *American Journal of Botany* **98**:872-895.

Gower, J. C. and W. J. Krzanowski. 1999. Analysis of distance for structured multivariate data and extensions to multivariate analysis of variance. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **48**:505-519.

Grau, O. *et al.* 2007. A comparison of altitudinal species richness patterns of *Bryophytes* with other plant groups in Nepal, Central Himalayas. *Journal of Biogeography* **34**:1907-1915.

Grünewald, B. 2010. Is pollination at risk? Current threats to and conservation of bees. *GAIA - Ecological Perspectives for Science and Society* **19**:61-67.

Hadley, A. S. and M. G. Betts. 2009. Tropical deforestation alters hummingbird movement patterns. *Biology Letters* **5**:207-210.

Harley, M. M. 1996. Palm pollen and the fossil record. PhD Dissertation. Royal Botanic Gardens, Kew, UK.

Heilbuth, J. C. 2000. Lower species richness in dioecious clades. *The American Naturalist* **156**:221-241.

Heilbuth, J. *et al.* 2001. The consequences of dioecy for seed dispersal: Modeling the seed-shadow handicap. *Evolution* **55**:880-888.

Hodgson, J. A. *et al.* 2011. Habitat area, quality and connectivity: Striking the balance for efficient conservation. *Journal of applied ecology* **48**:148-152.

Hooper, D. U. *et al.* 2005. Effects of biodiversity onecosystem functioning: A consensus of current knowledge. *Ecological Monographs* **75**:3-35.

Isaac, N. J. B. *et al.* 2007. Mammals on the EDGE: Conservation priorities based on threat and phylogeny. *PLoS ONE* **2**:e296.

Isbell, F. *et al.* 2011. High plant diversity is needed to maintain ecosystem services. *Nature* **477**:199-202.

IUCN. 2001. IUCN Red List categories: Version 3.1. IUCN Species Survival Commission, IUCN. Gland, Switzerland and Cambridge, UK.

IUCN. 2008. IUCN Red List of threatened species. URL http://www.iucnredlist.org

IUCN. 2011. Red List overview. URL http://www.iucnredlist.org/about/red-list-overview

Ives, A. R. *et al.* 1999. Stability and variability in competitive communities. *Science* **286**:542-544.

Jacobi, C. *et al.* 2007. Plant communities on ironstone outcrops: A diverse and endangered Brazilian ecosystem. *Biodiversity and Conservation* **16**:2185-2200.

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

Jones, K. *et al.* 2002. A phylogenetic supertree of the bats (mammalia: Chiroptera). *Biological Reviews* **77**:223 - 259.

Jones, K. E. *et al.* 2003. Biological correlates of extinction risk in bats. *The American Naturalist* **161**:601-614.

Katoh, K. *et al.* 2002. MAFFT: A novel method for rapid multiple sequence alignment based on fast fourier transform. *Nucleic Acids Research* **30**:3059-3066.

Kay, K. M. *et al.* 2005. Rapid speciation and the evolution of hummingbird pollination in neotropical costus subgenus costus (costaceae): Evidence from nrdna its and ets sequences. *American Journal of Botany* **92**:1899-1910.

Kerr, J. T. 1997. Species richness, endemism, and the choice of areas for conservation. *Conservation Biology* **11**:1094-1100.

Kessler, M. *et al.* 2011. A global comparative analysis of elevational species richness patterns of ferns. *Global Ecology and Biogeography* **20**: 868–880.

Kim, K. J. and R. K. Jansen. 1995. *ndhF* sequence evolution and the major clades in the sunflower family. *Proceedings of the National Academy of Sciences* **92**:10379-10383.

Kissling, W. D. and G. Carl. 2008. Spatial autocorrelation and the selection of simultaneous autoregressive models. *Global Ecology and Biogeography* **17**:59-71.

Knapp, S. 2011. Rarity, species richness, and the threat of extinction,are plants the same as animals? *PLoS Biol* **9**:e1001067.

Knobloch, E. and D. H. Mai. 1986. Monographie der früchter und samen in der kreide von mitteleuropa. Rozpr. *Ustavu. Geol* **47**:1–219.

Koopowitz, H. 1983. Plant extinction a global crisis. Stone Wael Press, Washington.

Kreft, H. and W. Jetz. 2007. Global patterns and determinants of vascular plant diversity. *Proceedings of the National Academy of Sciences* **104**:5925-5930.

Laurance, W. F. *et al.* 1998. Effects of forest fragmentation on recruitment patterns in Amazonian tree communities. *Conservation Biology* **12**:460-464.

Laurance, W. F. *et al.* 2006. Rapid decay of tree-community composition in Amazonian forest fragments. *Proceedings of the National Academy of Sciences of the United States of America* **103**:19010-19014.

Lawton, J. H. *et al.* 1994. Population dynamic principles [and discussion]. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **344**:61-68.

Leme, E. M. C. *et al.* 2010. Three new Pitcairnia species (*Bromeliaceae*) from the inselbergs of Espírito Santo, Brazil. *Systematic Botany* **35**:487-496.

Linder, H. P. 1987. The evolutionary history of the Poales/Restionales: A hypothesis. *Kew Bulletin* **42**:297-318.

Linder, H. P. 1991. Environmental correlates of patterns of species richness in the South-Western Cape Province of South Africa. *Journal of Biogeography* **18**:509-518.

Lôbo, D. *et al.* 2011. Forest fragmentation drives Atlantic forest of Northeastern Brazil to biotic homogenization. *Diversity and Distributions* **17**:287-296.

Loreau, M. 1998. Biodiversity and ecosystem functioning: A mechanistic model. *Proceedings of the National Academy of Sciences* **95**:5632-5636.

Loreau, M. *et al.* 2001. Biodiversity and ecosystem functioning: Current knowledge and future challenges. *Science* **294**:804-808.

Mace, G. M. 1995. Classification of threatened species and its role in conservation planning. Pages 197-213 *in* L. J. H and R. M. May, editors. Extinction rates. Oxford Univeristy Press, Oxford, UK.

Mace, G. M. *et al.* 2003. Preserving the tree of life. *Science* **300**:1707-1709.

Magurran, A. E. 2004. Measuring biological diversity. Blackwell Publishing, Oxford.

Margules, C. R. and R. L. Pressey. 2000. Systematic conservation planning. *Nature* **405**:243-253.

McDonald-Madden, E. *et al.* 2008. Making robust decisions for conservation with restricted money and knowledge. *Journal of applied ecology* **45**:1630-1638.

McMahon, M. M. and M. J. Sanderson. 2006. Phylogenetic supermatrix analysis of GenBank sequences from 2228 Papilionoid Legumes. *Systematic Biology* **55**:818-836.

Memmott, J. *et al.* 2004. Tolerance of pollination networks to species extinctions. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **271**:2605-2611.

Meynard, C. N. *et al.* 2011. Beyond taxonomic diversity patterns: How do α, β and γ components of bird functional and phylogenetic diversity respond to environmental gradients across France? *Global Ecology and Biogeography* **20**: 893–903.

Millennium Ecosystem Assessment. 2005. Ecosystems and human well-being: Biodiversity synthesis., Washington DC.

Moser, D. *et al.* 2005. Environmental determinants of vascular plant species richness in the Austrian Alps. *Journal of Biogeography* **32**:1117-1127.

Mouchet, M. *et al.* 2008. Towards a consensus for calculating dendrogram-based functional diversity indices. *Oikos* **117**:794-800.

Myers, N. *et al.* 2000. Biodiversity hotspots for conservation priorities. *Nature* **403**:853-858.

Naeem, S. 1998. Species redundancy and ecosystem reliability. *Conservation Biology* **12**:39-45.

Nee, S. and R. M. May. 1997. Extinction and the loss of evolutionary history. *Science* **278**:692-694.

New, M. *et al.* 2011. Four degrees and beyond: The potential for a global temperature increase of four degrees and its implications. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **369**:6-19.

Nunes, P. A. L. D. and J. C. J. M. van den Bergh. 2001. Economic valuation of biodiversity: Sense or nonsense? *Ecological Economics* **39**:203-222.

O'Brien, E. M. 1993. Climatic gradients in woody plant species richness: Towards an explanation based on an analysis of Southern Africa's woody flora. *Journal of Biogeography* **20**:181-198.

Olson, D. M. *et al.* 2001. Terrestrial ecoregions of the world: A new map of life on earth. *BioScience* **51**:933-938.

Owens, I. P. F. and P. M. Bennett. 2000. Ecological basis of extinction risk in birds: Habitat loss versus human persecution and introduced predators. *Proceedings of the National Academy of Sciences* **97**:44-48.

Pagel, M. 1999. Inferring the historical patterns of biological evolution. *Nature* **401**:877-884.

Paradis, E. *et al.* 2004. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**:289-290.

Paton, A. J. *et al.* 2008. Towards target 1 of the Global Strategy for Plant Conservation: A working list of all known plant speciesprogress and prospects. *Taxon* **57**:602-611.

Pennington, R. T. and C. W. Dick. 2004. The role of immigrants in the assembly of the South American rainforest tree flora. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **359**:1611-1622.

Persha, L. *et al.* 2010. Biodiversity conservation and livelihoods in human-dominated landscapes: Forest commons in South Asia. *Biological Conservation* **143**:2918-2925.

Petchey, O. L. and K. J. Gaston. 2002. Extinction and the loss of functional diversity. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **269**:1721-1727.

Petchey, O. L. and K. J. Gaston. 2006. Functional diversity: Back to basics and looking forward. *Ecology Letters* **9**:741-758.

Peterken, G. F. 1974. A method for assessing woodland flora for conservation using indicator species. *Biological Conservation* **6**:239-245.

Pimm, S. L. *et al.* 1995. The future of biodiversity. *Science* **269**:347-350.

Pirie, M. D. *et al.* 2008. A novel supermatrix approach improves resolution of phylogenetic relationships in a comprehensive sample of Danthonioid grasses. *Molecular Phylogenetics and Evolution* **48**:1106-1119.

Pisani, D. *et al.* 2002. A genus-level supertree of the Dinosauria. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **269**:915-921.

Pocock, M. J. O. *et al.* 2006. Ecological correlates of range structure in rare and scarce British plants. *Journal of Ecology* **94**:581-596.

Posada, D. 2008. jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution* **25**:1253-1256.

Potts, S. G. *et al.* 2010. Global pollinator declines: Trends, impacts and driversity. *Trends in Ecology and Evolution* **25**:345-353.

Pressey, R. L. *et al.* 2007. Conservation planning in a changing world. *Trends in Ecology and Evolution* **22**:583-592.

Prinzing, A. *et al.* 2001. The niche of higher plants: Evidence for phylogenetic conservatism. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **268**:2389-2389.

Purvis, A. 2008. Phylogenetic approaches to the study of extinction. *Annual review of ecology, evolution, and systematics* **39**:301-319.

Purvis, A. and A. Rambaut. 1995. Comparative Analysis by Independent Contrasts (CAIC): An Apple Macintosh application for analysing comparative data. *Computer applications in the biosciences : CABIOS* **11**:247-251.

Purvis, A. *et al.* 2000a. Predicting extinction risk in declining species. *Proceedings of the Royal Society B: Biological Sciences* **267**:1947-1952.

Purvis, A. *et al.* 2000b. Non-random extinction and the loss of evolutionary history. *Science* **288**:328-330.

Pybus, O. G. and P. H. Harvey. 2000. Testing macro-evolutionary models using incomplete molecular phylogenies. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **267**:2267-2272.

R Development Core Team (2011). R: A language and environment for statistical computing. URL http://www.R-project.org/

Raskin, I. *et al.* 2002. Plants and human health in the Twenty-First Century. *Trends in Biotechnology* **20**:522-531.

Redding, D. W. and A. Ø. Mooers. 2006. Incorporating evolutionary measures into conservation prioritization. *Conservation Biology* **20**:1670-1678.

Ricklefs, R. E. and S. S. Renner. 1994. Species richness within families of flowering plants. *Evolution* **48**:1619-1636.

Roberts, C. M. *et al.* 2002. Marine biodiversity hotspots and conservation priorities for tropical reefs. *Science* **295**:1280-1284.

Rodrigues, A. S. L. *et al.* 2006. The value of the IUCN Red List for conservation. *Trends in Ecology and Evolution* **21**:71-76.

Rondinini, C. *et al.* 2011. The key elements of a comprehensive global mammal conservation strategy. *Philosophical Transactions of the Royal Society B: Biological Sciences* **366**:2591-2597.

Royal Botanic Gardens Kew. 2010. Plants Under Pressure a Global Assessment. The first report of the IUCN Sampled Red List Index for plants., Royal Botanic Gardens, Kew, UK, London.

Rutschmann, F. *et al.* 2004. Did Crypteroniaceae really disperse out of India? Molecular dating evidence from *rbcL*, *ndhF* and *rpl16* intron sequences. *Int. J. Plant Sci* **165:(**4 suppl.).

Sanderson, M. 1997. A non-parametric approach to estimating divergence times in the absence of rate constancy. *Molecular Biology and Evolution* **14**:8-1.

Sanderson, M. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Molecular Biology and Evolution* **19**:101-109.

Sanderson, M. J. 2003. R8s: Inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**:301-302.

Sargent, R. D. and D. D. Ackerly. 2008. Plant-pollinator interactions and the assembly of plant communities. *Trends in Ecology and Evolution* **23**:-130.

Schwartz, M. W. and Simberloff. 2001. Taxon size predicts rates of rarity in vascular plants. *Ecology Letters* **4**:464-469.

Schweiger, O. *et al.* 2010. Multiple stressors on biotic interactions: How climate change and alien species interact to affect pollination. *Biological Reviews* **85**:777-795.

Sing, T. *et al.* 2005. ROCR: Visualizing classifier performance in R. *Bioinformatics* **21**:3940-3941.

Smart, S. M. *et al.* 2006. Biotic homogenization and changes in species diversity across human-modified ecosystems. *Proceedings of the Royal Society B: Biological Sciences* **273**:2659-2665.

Smith, S. *et al.* 2009. Mega-phylogeny approach for comparative biology: An alternative to supertree and supermatrix approaches. *Bmc Evolutionary Biology* **9**:37.

Sodhi, N. S. *et al.* 2004. Southeast Asian biodiversity: An impending disaster. *Trends in Ecology and Evolution* **19**:654-660.

Sodhi, N. S. *et al.* 2008. Correlates of extinction proneness in tropical angiosperms. *Diversity and Distributions* **14**:1-10.

Stamatakis, A. 2006. Raxml-vi-hpc: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**:2688 - 2690.

Stevens, R. D. *et al.* 2003. Patterns of functional diversity across an extensive environmental gradient: Vertebrate consumers, hidden treatments and latitudinal trends. *Ecology Letters* **6**:1099-1108.

Stewart, L. 1994. A guide to Palms and Cycads of the world. Angus and Robertson, Sydney.

Swets, J. 1988. Measuring the accuracy of diagnostic systems. *Science* **240**:1285-1293.

Tallents, L. A. *et al.* 2005. Phylogenetic diversity of forest trees in the Usambara mountains of Tanzania: Correlations with altitude. *Botanical Journal of the Linnean Society* **149**:217-228.

The Angiosperm Phylogeny Group. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* **161**:105-.

The Zoological Society of London (2011). EDGE of existence. URL http://www.edgeofexistence.org/

Thomson, R. C. and H. B. Shaffer. 2010. Sparse supermatrices for phylogenetic inference: Taxonomy, alignment, rogue taxa, and the phylogeny of living turtles. *Systematic Biology* **59**:42-58.

Tilman, D. *et al.* 1997. Plant diversity and ecosystem,productivity: Theoretical considerations. *Proceedings of the National Academy of Sciences* **94**:1857-1861.

Tilman, D. *et al.* 2006. Biodiversity and ecosystem stability in a decade-long grassland experiment. *Nature* **441**:629-632.

Vamosi, J. C. and S. M. Vamosi. 2005. Present day risk of extinction may exacerbate the lower species richness of dioecious clades. *Diversity and Distributions* **11**:25-32.

Vamosi, J. C. and J. R. U. Wilson. 2008. Non-random extinction leads to elevated loss of angiosperm evolutionary history. *Ecology Letters* **11**:1047-1053.

Vamosi, S. M. *et al.* 2009. Emerging patterns in the comparative analysis of phylogenetic community structure. *Molecular Ecology* **18**:572-592.

Vane-Wright, R. I. *et al.* 1991. What to protect? Systematics and the agony of choice. *Biological Conservation* **55**:235-254.

Versieux, L. and T. Wendt. 2007. Bromeliaceae diversity and conservation in Minas Gerais State, Brazil. *Biodiversity and Conservation* **16**:2989-3009.

Wake, D. B. and V. T. Vredenburg. 2008. Are we in the midst of the sixth mass extinction? A view from the world of amphibians. *Proceedings of the National Academy of Sciences* **105**:16-13.

Webb, C. O. 2000. Exploring the phylogenetic structure of ecological communities: An example for rain forest trees. *The American Naturalist* **156**:-156.

WHO (2002). World Summit on Sustainable Development (WSSD). URL http://www.who.int/trade/glossary/story097/en/index.html

Wiens, J. J. *et al.* 2010. Niche conservatism as an emerging principle in ecology and conservation biology. *Ecology Letters* **13**:1310-1324.

Wiens, J. J. and C. H. Graham. 2005. Niche conservatism: Integrating evolution, ecology, and conservation biology. *Annual review of ecology, evolution, and systematics* **36**:519-539.

WWF (2001). Carolines Tropical Moist Forests (oc0). URL http://www.worldwildlife.org/wildworld/profiles/terrestrial/oc/oc0_full.html

Wyse, P. and K. Kennedy. 2009. The Global Strategy for Plant Conservation: A challenge and opportunity for the international community. *TRENDS in Plant Science* **14**:578-580.

Zobel, M. 1997. The relative of species pools in determining plant species richness: An alternative explanation of species coexistence? *Trends in Ecology and Evolution* **12**:266-269.