EVOLUTIONARY HISTORY EXPLAINS DISTRIBUTION AND DIVERSITY OF MAMMALS AND THEIR PARASITES

by

SHAN HUANG

(Under the Direction of Sonia Altizer and John L. Gittleman)

ABSTRACT

Present biodiversity is shaped through species diversification and extinction in the evolutionary history. Therefore, phylogeny, as a representative of species evolutionary history, can lend insight towards examining current biodiversity patterns, including the distribution and diversity of mammals and of their parasites. Phylogenetic diversity (PD) has been proposed as an important measure of biodiversity because PD incorporates the number of species and the genetic and functional diversity of the species. Phylogeny of host species might also be associated with the distribution and diversity of parasites, many of which are relevant to wildlife conservation. My dissertation research focuses on evaluating phylogenetic information in predicting broad-scale patterns of distribution and diversity of mammals and their parasites. I combined multiple global comparative data sets compiled across large taxonomic and spatial scales, and conducted phylogenetic analyses to investigate two central topics in mammal conservation biology: (1) global distribution patterns of mammalian diversity, and (2) the parasite distribution and diversity in free-ranging mammals. I first compared phylogenetic diversity and species richness in predicting a third measure of biodiversity, the trait diversity of mammals. I

then searched for geographic regions where potential loss of phylogenetic diversity is higher than expected by the number of threatened mammal species and investigated mechanisms causing such extra loss. Next, I assessed the importance of host phylogeny, in comparison with other host ecological traits, in predicting the number of parasite species infecting a host species. I also compared the phylogenetic relatedness between host species, host ecological similarity and geographic range overlap, for predicting the parasite species assemblages between host species. Overall, my research has demonstrated that phylogenetic information is essential for quantifying mammal biodiversity and estimating potential loss in mammal biodiversity. Furthermore, knowledge of host phylogeny often provides the most important predictors, in comparison with host ecology, for the distribution and diversity of parasites among mammal hosts.

INDEX WORDS: Biodiversity, Distribution, Diversity, Global Analysis, Phylogeny, Mammal, Carnivora, Wild Host, Parasite Diversity, Parasite Sharing, Conservation.

EVOLUTIONARY HISTORY EXPLAINS DISTRIBUTION AND DIVERSITY OF MAMMALS AND THEIR PARASITES

by

SHAN HUANG

BSc, Sun Yat-sen (Zhongshan) University, China, 2005

MSc, Imperial College London, UK, 2006

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2012

EVOLUTIONARY HISTORY EXPLAINS DISTRIBUTION AND DIVERSITY OF MAMMALS AND THEIR PARASITES

by

SHAN HUANG

Major Professor:	Sonia Altizer John L. Gittleman
Committee:	John Drake Nate Nibbelink Charles Nunn

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2012

DEDICATION

For mom and dad.

ACKNOWLEDGEMENTS

I express most sincere thanks to Shai Meiri, David Orme, and other members from Ian Owens' research group at the Silwood Park Campus of Imperial College London for first exposing me to the field of macroecology and macroevolution. I sincerely thank my PhD committee, which includes John Drake, Nate Nibbelink, Charles Nunn, and my advisors John Gittleman and Sonia Altizer for the inspiration and guidance, and for providing just the right ratio of support and independence. I am extremely grateful to my colleagues, collaborators and fellow graduate students at UGA not only for their technical critiques of my work, but also to their contributions to my professional and personal development during my time. I particularly wish to thank the members of the Gittleman, Altizer, and Ezenwa lab groups for providing a constructive, friendly and stimulating environment. I am extremely thankful for the excellent technical support provided by Jeremy Sanderlin, Byron Ledbetter, Brian Perkins and Tyler Ingram at Odum School of Ecology. I am indebted to the undergraduate assistants, including Michael Cherry, Karen Christ, Matt Foretich, Ashton Griffin, Steven Howell, Tyler Halle, Cecilia Nix, and Jeff Shapiro, who worked tremendously on entering and cross-validating data for the global carnivore parasite database. Much of this work involved international collaboration with members of different academic institutions, and I particularly thank Patrick Stephens, Jonathan Davies and Olaf Bininda-Emonds for the inspiration and wonderful collaboration experience we have had.

This work was funded by Odum School of Ecology Small Grants; and research grants through my advisors John and Sonia. I am also grateful for the Graduate Lab Assistantships provided by UGA Graduate School.

Last but not least, I am extremely thankful for all the love, understanding and support from my family, friends, and J. Wang.

TABLE OF CONTENTS

Page

CHAPTER 1

INTRODUCTION

"Nothing in biology makes sense except in the light of evolution."

Theodosius Dobzhansky (1973)

Quantifying and understanding the determinants of biodiversity, the variety of life, is a central topic in macroecology (Gaston 2000, Brooks et al. 2006). Biodiversity can be seen as the raw material for providing "option value" – the ability to respond to unpredictable events or needs (McNeely et al. 1990), and it is not only important for maintaining the ecosystem functioning, but also for providing ecosystem services (Crozier 1997, Díaz and Cabido 2001, Loreau et al. 2001, Worm et al. 2006). Effective conservation action requires comprehensive, accurate, detailed data on where species occur and associated measures that indicate the likelihood of future species' survival. Phylogeny has been presented as a powerful tool for biodiversity research (Mace et al. 2003, Rodrigues et al. 2011), as it represents the evolutionary history of species and can be used to explain and predict many biodiversity patterns, including the origins (Murphy et al. 2001, Bininda-Emonds et al. 2007, Freckleton and Jetz 2009, Nyakatura and Bininda-Emonds 2012), geographic distribution (Davies and Pedersen 2008, Devictor et al. 2010), and potential loss of biodiversity (Purvis et al. 2000, Isaac et al. 2007, Fritz and

Purvis 2010b). The work presented here uses phylogenetic information for explaining and predicting global patterns of biodiversity.

Terrestrial mammals are an ideal group for this research. As one of the best studied groups of organisms, their geographic distribution, biological traits, and evolutionary history have been well described and comprehensive data sets have been published and made available to the public (Bininda-Emonds et al. 2007, Schipper et al. 2008, Fritz et al. 2009, Jones et al. 2009). The dramatic variation in their geographic distribution and biology (Schipper et al. 2008, Jones et al. 2009) makes them an interesting group for studying biodiversity patterns. Although mammals are often considered atypical, as they are mostly larger than other terrestrial organisms, many macroecological patterns identified in mammals apply to other groups (e.g. latitude gradient in geogrpahic range size: Rohde 1999, altitude gradient in geographic range size: McCain 2009, and latitude gradient in body size: Olson et al. 2009). The diversity of mammals is valuable to many terrestrial ecosystems, because mammals play key roles such as grazers, predators and seed dispersers (Wilson and Reeder 2005). However, 20% of the world's terrestrial mammals are currently at high risk of extinction (Schipper et al. 2008). Particularly, large, visible and charismatic mammal species at high risk of extinction have become conservation flagship species, such as the tiger (*Panthera tigris*), giant panda (*Ailuropoda melanoleuca*), and elephant (*Elepbus maximus*) (Caro and O'Doherty 1999, Williams et al. 2000). Effective and efficient conservation activity is urgently needed for preserving mammalian biodiversity. In my dissertation, the value of phylogenetic information is assessed with respect to two important conservation issues: quantifying mammalian biodiversity and predicting risk of infectious disease.

Phylogenetic information can be useful for conservation prioritization through quantifying biodiversity and estimating potential biodiversity loss in species assemblages. Ecologists have focused a great deal of effort on explaining the patterns of biodiversity in various spatial contexts and exploring their implications (Gaston 2000). The scope of studies on biodiversity ranges from genes to ecosystems, but most surveys tend to measure biodiversity as the number of species observed or estimated to occur within an assemblage (i.e. species richness) (Gaston 2000). However, it might be unrealistic to assume that all species are equal and that maximizing the number of species will maximize biodiversity measured in other ways (Soutullo et al. 2005). Recently, more attention has focused on understanding diversity beyond the number of species in an assemblage, such as the genetic and functional diversity among the species in question (Purvis and Hector 2000). While various measures have been proposed for quantifying either the phenotypic or genetic difference, phylogenetic diversity, or PD, has been advocated as a measure reflecting both (Purvis and Hector 2000, Rodrigues and Gaston 2002, Bininda-Emonds et al. 2007, Forest et al. 2007, Davies et al. 2008, Mace and Purvis 2008, Devictor et al. 2010, Safi et al. 2011).

In Faith's (1992) original definition of PD, the PD of a subset of x out of N species is defined as the sum of the lengths of all the branches in the minimum spanning path for x. The amount of PD in an assemblage depends not only on the number of species, but also on the relatedness of the species. In other words, phylogenetic diversity reflects not only the number of species in an assemblage, but also the accumulated evolutionary history through which diversity has evolved. Systematic conservation planning requires evaluation of the conservation value of potential target regions, and it is

thus essential to first determine the conservation currency – the unit of conservation value for planning and thus evaluation (Margules and Pressey 2000). PD has been proposed as a conservation currency, because it is expected to be a biodiversity surrogate that incorporates not only the number of species but also phenotypic and genetic diversity that have been accumulated through evolutionary history (Purvis and Hector 2000, Mace et al. 2003, Davies et al. 2008, Mace and Purvis 2008).

Chapter 2 presents a comprehensive assessment of whether PD is superior to the simple species count, also known as species richness (SR), for representing an important dimension of biodiversity, trait diversity (TD). I combined global data sets of mammal phylogeny (Bininda-Emonds et al. 2007, Fritz et al. 2009), geographic distribution (Schipper et al. 2008), and species-level biological traits (Jones et al. 2009) to examine the correlations among SR, PD and TD across phylogenetic clades and geographic space. The overall results support both PD and species richness as significant correlates of trait diversity. Across all the mammalian clades on the phylogenetic supertree (Bininda-Emonds et al. 2007, Fritz et al. 2009), PD tends to be a better representation of trait diversity than SR. However, when assessed across space, global geographic patterns of PD and SR seem to be similarly representative of patterns of trait diversity.

Chapter 3 describes how PD can provide extra information about potential biodiversity loss that cannot be conveyed by using SR as the measure of biodiversity. I calculated potential PD loss if all mammal species currently recognized as threatened (Schipper et al. 2008) were to go extinct, and compared potential PD loss with possible PD loss if random species of the same number in an area (100 * 100 km2 grid cell) were selected to go extinct. The impact of global species extinction on regional biodiversity

varies dramatically across regions. If all the threaten mammal species were extinct, several biodiversity hotspots in southern Asia and Amazonia would lose a larger proportion of currently existing PD than expected by random extinction. Extra PD losses might be due to phylogenetically clustered extinction risk and/or the tendency for highly distinctive species to suffer high risk of extinction, depending on the geographic locality. Overall results indicate that the impact of global extinction on regional biodiversity might be underestimated in global analyses because such impact might occur at fine scales.

Beyond the importance of PD for quantifying biodiversity and estimating biodiversity loss, mammalian evolutionary history is fundamentally linked to the distribution and diversity of parasites in wild mammals. Parasites can have a large influence on biodiversity from ecological, evolutionary and conservation perspectives (reviewed in: Hudson et al. 2002, Altizer and Pedersen 2008). On the one hand, parasites can induce serious threats to biodiversity by lowering host fitness and wiping out host species (Altizer et al. 2003, Smith et al. 2006, Altizer et al. 2007, Pedersen et al. 2007); on the other hand, parasites can contribute positively to biodiversity by maintaining host genetic diversity and potentially driving host diversification (Nunn et al. 2004). It is also important to note that parasite diversity might have conservation value unto itself because parasites are considered to be a significant part of biodiversity, constituting potentially half or more of species on earth (Poulin and Morand 2000). Therefore, concerns have been raised on the extra loss of biodiversity due to co-extinction of host and parasite species (Durden and Keirans 1996, Koh et al. 2004, Dunn et al. 2009). Identifying predictors of parasite distribution and diversity among host species can help to estimate parasitism in understudied host species and predict future parasite emergence, which

provides valuable information not only for wildlife host conservation, but also for preserving parasite diversity.

In recent decades, host-parasite interactions have attracted attention from conservation biologists, and great progress has been made in understanding the processes governing infection, especially using mathematical models. However, most of these studies focus on single-host-single-parasite interactions and ignore the effects of complex interactions in more realistic multi-host-multi-parasite systems. Only recently have global-scale databases (e.g. Nunn and Altizer 2005, Jones et al. 2008) become available to facilitate investigations of complex interactions across large numbers of host and parasites species. Recent advance in host parasite research has heavily emphasized a multi-host-multi-parasite framework (e.g. Esch et al. 1990, Poulin 1992, Poulin and Morand 2000, Holt et al. 2003, Poulin and Mouillot 2003, Pedersen and Fenton 2007, Pugliese 2010, Streicker et al. 2010). This part of my work is based on data from a global database of carnivore parasites that is part of the larger Global Mammal Parasite Database (Nunn and Altizer 2005; www.mammalparasites.org). The carnivore parasite database contains records of 930 parasite species reported in free-living populations of 159 carnivore species based on 1157 published studies searchable at major publication databases online. Data on host biology and ecology, examination methodology and parasite or pathogen prevalence or intensity were collected if disclosed in the reports.

Carnivores are an excellent representative group of mammals for studying the relationship between host evolutionary history and parasite infection risk. As one of the best studied mammal groups, carnivores exhibit large interspecific variation across a range of biological and ecological characteristics including body size (Gittleman 1985),

reproductive rate (Gittleman 1993), mating system, and group size (Gittleman 1989), and their geographic ranges reach nearly every type of habitat (Sunquist 2002). In addition to comprehensive biological and ecological data (Jones et al. 2009), newly updated geographic distribution data of the 275 land carnivore species are now available (Schipper et al. 2008) and 268 of them are included on a 90% resolved phylogenetic tree (Fritz et al.). Furthermore, carnivores are one of several mammalian orders that have relatively high percentages of globally threatened or extinct species (26.7%) (Schipper et al. 2008). Despite the fact that conservation efforts have focused on conserving threatened carnivore species, numerous examples have been reported where species experienced large declines in population size due to infectious diseases, including African wild dogs (Alexander and Appel 1994, Kat et al. 1995), Ethiopian wolves (Laurenson et al. 1998, Randall et al. 2004), Black-footed ferrets (Thorne and Williams 1988), Serengeti lions (Roelke-Parker et al. 1996) and Caspian seals (Kennedy et al. 2000).

The evolutionary history of host species is tightly linked to present patterns of parasite distribution and diversity among hosts through various mechanisms. Studies comparing host and parasite phylogenies have provided evidence of host parasite co-speciation (Brooks 1979, Hoberg et al. 1997), while several other mechanisms have also been suggested to underlie connections between host evolutionary history and parasite occurrence (Page 1994). For example, the diversification rate in a host lineage has been suggested to be associated with the diversity of parasites per host species, because host diversification might provide opportunities for a large diversity of parasites to persist, and conversely, the diversity of parasites might induce strong pressure on host survival and thus drive host diversification (Nunn et al. 2004). Furthermore, host species that diverged

recently tend to harbor co-inherent parasites from their common ancestor, and provide opportunities to other parasites infecting only one of them for host switching to the other (Davies and Pedersen 2008, Streicker et al. 2010). Therefore, knowledge of host evolutionary history is expected to provide important information on broad scale patterns of parasite occurrence. I combined parasite occurrence data in the global carnivore parasite database with phylogenetic data for carnivore species (Nyakatura and Bininda-Emonds 2012) to investigate whether host phylogeny, as a representation of carnivores' evolutionary history, can predict patterns of parasite diversity in different host species, and parasite sharing between host species.

Chapter 4 evaluates the importance of host evolutionary history in predicting the diversity of the parasites infecting the host species. Based on the data from the global carnivore parasite database, I used non-parametric methods to estimate parasite diversity in carnivore hosts and examined the association between parasite diversity and host evolutionary distinctiveness. I also assessed the importance of host evolutionary distinctiveness in predicting parasite diversity, in relation to a range of host ecological traits that have been suggested as important correlates of parasite diversity (Nunn et al. 2003a, Nunn et al. 2005, Ezenwa et al. 2006, Lindenfors et al. 2007). Results suggest that host evolutionary distinctiveness measured in the length of host terminal branch on the phylogenetic tree, is the most important predictor of parasite diversity in comparison with other ecological traits. The length of the terminal branch connecting a host species to the others is negatively correlated with parasite diversity. In other words, evolutionarily distinctive host species that diverged from others in the distant past harbor fewer parasite species than host species that have diverged from others recently. The association

between host species sexual dimorphism in body mass and parasite diversity supports the hypothesis that parasite diversity might often have driven host diversification through sexual selection (Nunn et al. 2004).

Chapter 5 directly addresses the relationship between the phylogenetic relatedness and parasite assemblage similarity between host species pairs. I used data in the global carnivore parasite database to quantify parasite assemblage similarity between host species, and assessed its association with three groups of host variables, phylogenetic relatedness, ecological similarity and geographic range overlap. Host phylogenetic relatedness is shown to be the most important predictor of parasite assemblage similarity, with closely phylogenetic related host species sharing higher proportions of their parasite species than distantly related host species. Host ecological similarity and host geographic overlap are also correlated with parasite assemblage similarity, but to a lesser extent. I also examined parasite species' host range on the carnivore phylogenetic tree and found that approximately 60% of the assessed parasite species, particularly helminths and viruses, have host ranges that are significantly constrained by host phylogeny. Overall, my results support the conclusion that host phylogeny is a powerful tool for predicting patterns of parasite distribution and diversity among host species. Particularly, host phylogenetic information allows us to estimate parasite diversity and parasite assemblage composition in understudied host species, as well as to predict future parasite emergence in novel host species.

CHAPTER 2

TRAITS, TREES AND TAXA: GLOBAL DIMENSIONS OF BIODIVERSITY IN MAMMALS[1]

[1] Huang, Shan, Patrick R. Stephens, and John L. Gittleman. 2012. *Proceedings of The Royal Society Biological Sciences.* Reprinted here with permission of publisher.

Abstract

Measures of biodiversity encompass variation along several dimensions such as species richness (SR), phylogenetic diversity (PD), and functional/ trait diversity (TD). At the global scale, it is widely recognized that SR and PD are strongly correlated, but the extent to which either tends to capture variation in TD is unclear. Here we assess relationships amongst PD, SR and TD for a number of traits both across clades and regional assemblages of mammals. We also contrast results using two different measures of TD, trait variance and a new measure we refer to as trait bin filling (the number of orders of magnitude of variation that contain at least one species). When TD is defined as trait variance, PD is a much stronger correlate of TD than SR across clades, consistent with hypotheses about the conservation value of PD. However, when TD is defined as bin filling, PD and SR show similar correlations with TD across clades and space. We also investigate potential losses of SR, PD, and TD if species that are currently threatened were to go extinct, and find that threatened PD is often a similar predictor of threatened TD as SR.

Key words

Biodiversity, species richness, phylogenetic diversity, trait diversity, mammals.

Introduction

Quantifying biodiversity is a central issue in ecology, biogeography, and conservation biology (Purvis and Hector 2000, Brooks et al. 2006, Pavoine and Bonsall 2011). Considered broadly, biodiversity encompasses variation in the composition and

characteristics of species assemblages (e.g., clades and communities). The diversity of assemblages can vary along several dimensions such as species richness (SR, taxonomic diversity sensu Devictor et al. 2010), genetic diversity, and trait diversity (TD) or functional diversity (Crozier 1997, Purvis and Hector 2000, Devictor et al. 2010, Safi et al. 2011). Of these, trait diversity has been the least studied even though it is potentially one of the most important aspects of biodiversity from a conservation perspective. For example, communities with high trait diversity and trait redundancy among species might be more resilient to perturbation (McNeely et al. 1990, Crozier 1997). Conversely, when species with distinctive ecological characteristics are lost from a local community, their extinction can have a disproportionate effect on the species that remain, and alter or reduce ecosystem function (Cardillo et al. 2005, Cardillo et al. 2008, Fritz et al. 2009, Huang et al. 2012a). Unfortunately, despite the availability of many sophisticated mathematical algorithms that have been developed for calculating TD of species (reviewed in Pavoine and Bonsall 2011), in most real-world cases it is not feasible to quantify TD for regional assemblages due to our limited knowledge of species' distributions and traits. Even when a complete species list is available, in many cases the traits of all but a few well-studied species are poorly known. For example, in mammals, perhaps the best-studied class of vertebrates, estimates of adult body mass are available in the literature for only approximately 60% species (Jones et al. 2009). Other characteristics such as litter size and population density have been quantified for even fewer species (Jones et al. 2009).

Recently there has been much interest in quantifying the phylogenetic diversity (PD) of species assemblages (Davies et al. 2008, Rodrigues et al. 2011), defined as the

summed branch lengths of a phylogeny of the species in an assemblage (Faith 1992). PD reflects both the number and the evolutionary distinctiveness of species in an assemblage, and thus can potentially act as a "silver bullet" ensconcing several dimensions of biodiversity (Faith 2002, Soutullo et al. 2005, Davies et al. 2008, Mace and Purvis 2008). We know from previous studies that PD is strongly correlated with SR in many systems (Polasky et al. 2001, Brooks et al. 2006, Schipper et al. 2008, Devictor et al. 2010, Morlon et al. 2011), due in large part to the way PD is calculated (i.e., when you add more species to a tree, the summed branch lengths of the tree are bound to increase). When PD is calculated from a tree with branch lengths that reflect sequence divergence, it can be presumed to also capture genetic diversity above the species level (Faith 1992, Bininda-Emonds et al. 2007). Whether PD generally captures variation in TD remains an open question. Several studies have investigated the relationship between PD and trait variance in experimental plant systems (Maherali and Klironomos 2007, Cadotte et al. 2008, Flynn et al. 2011), but with mixed results. A global study of mammals considered the relationships amongst SR, PD, and a composite measure of functional diversity (Safi et al. 2011), and showed all three variables to be strongly correlated; however, this analysis includes only four traits and the study did not assess variation in the strength of correlations for individual traits.

If PD is in general correlated with TD for a wide range of traits, it could serve as an important proxy for a dimension of biodiversity that is difficult to quantify otherwise. Knowledge of the phylogenetic structure of a community could also provide useful information such as which species might have unique ecological roles based on their phylogenetic distinctiveness. However, SR is easier to quantify than PD, the latter of

which requires not only knowledge of which species occur in an assemblage but also their evolutionary relationships. If PD is not a stronger correlate of TD than SR, then the extra effort required to quantify PD (i.e., constructing a phylogeny) would not be warranted when the primary goal is to estimate expected TD. To date this issue has rarely been investigated empirically. Fritz and Purvis (2010a) found that if threatened mammal species were to go extinct, the geographic patterns of PD losses across the world's ecoregions (Olson et al. 2001) would not be consistent with the losses of TD measured as adult body mass variance. However, the spatial scale they used (i.e., entire ecoregions) is arguably coarse for many conservation considerations, and they did not assess the relationship between total TD and total PD, examine traits other than mass, or consider measures of TD other than variance (Fritz and Purvis 2010a).

Here we investigate the relationship between TD, PD and SR for a variety of traits in terrestrial mammals at two different scales: across phylogenetic clades and across geographic regions. We first investigated the relationship between PD and TD across clades, since if no correlation occurs at the clade level it seems unlikely that such a relationship would hold between the PD and TD of communities (which can be thought of as somewhat random subsamples of full mammalian clades). We quantify TD in both individual level traits, such as litter size, as well as population and species level traits such as population density. We perform analyses to address three questions: (1) Is SR and PD correlated with TD across clades? (2) Is PD generally a stronger correlate of TD than SR? Finally, (3) does sample size or variation in phylogenetic signal explain differences in the strength of correlations between TD and PD or SR? Intuitively, the strength of phylogenetic signal in a trait could affect the degree to which PD is a better

predictor than SR, since traits with low signal exhibit variation that is independent of phylogeny. Additionally, we expect sample size for an individual trait might affect the observed patterns of TD. Because the amount of data available for mammals varies dramatically for different traits (Jones et al. 2009), low proportions of mammal species in a clade or a community that a trait has been sampled for could add "noise" to estimates of correlations between TD and PD (or SR). We further conduct global spatial analyses using two traits (body mass and geographic range size) that have particularly large sample sizes to see (4) whether PD (or SR) can be a good representative of TD across geographic regions; (5) whether the areas of highest PD and TD show geographic correspondence (i.e., are "hotspots" of PD and TD in mammals congruent; Myers et al. 2000, Sechrest et al. 2002), and (6) from a conservation perspective, whether the number of threatened species and amount of PD loss in a region predict the amount of TD that stands to be lost.

Methods

Defining TD

We here use the term "trait diversity" in preference to "functional diversity" in acknowledgement of the fact that the biological traits we considered may or may not directly reflect the ecosystem function of a species (e.g., mice and shrews have very different ecological roles, though they are both small, have high reproductive rates, and can reach high population densities). Among existing measures of TD (reviewed in Schleuter et al. 2010, Pavoine and Bonsall 2011), trait variance (e.g., Mason et al. 2003, Fritz and Purvis 2010a) and composite measures of functional diversity (e.g., Cornwell et

al. 2006, Mouchet et al. 2008, Devictor et al. 2010) are perhaps the most widely used in recent broad-scale studies. A previous global study of mammals (Safi et al. 2011) used a composite measure of trait variation based on dendrograms constructed using the unweighted pair group method with arithmetic averages (i.e., UPGMA). However, the summed branch lengths of a UPGMA tree will inevitably be correlated with the summed branch lengths of a phylogeny for the same number of species because both are sensitive to species richness, even when UPGMA trees are generated using completely random trait data (Fig. S1). Thus, their method is not well suited for testing for a general relationship between TD and PD. Other ways to examine composite TD include principal components and a multidimensional measure of functional richness such as FR_V or FR_{Im} (sensu Schleuter et al. 2010). However, the number of species that we could include using these methods would diminish rapidly as we included more traits since they all require complete case analysis (Fig. S2). Both to maximize sample sizes and to avoid using a measure of TD that is autocorrelated with PD and SR we focus on the variance of individual traits in this study.

While trait variance is a fairly standard measure of TD, it does not necessarily capture trait diversity in the sense of niche filling. A measure of niche filling that has been used in previous studies is functional range, the range of species trait values in a local community compared to the range of species trait values in a region or globally (Mason et al. 2005). While this does capture variation in the breadth of niche space occupied communities, it does not fully reflect "niche filling" in that a community with two species (e.g., a community consisting of a mouse species and an elephant species) could potentially could be considered as diverse as a community containing many species

with a greater variety of trait values. We therefore propose a new measure of TD that we refer to as trait bin filling, which more closely reflects trait diversity in the classical community ecological sense of the filling of semi-discrete niches (e.g., Macarthur's "broken stick" model of niche partitioning, MacArthur 1957) than either trait variance or other measures considering functional range (e.g. see Mason et al. 2005). Bin filling refers to the number of orders of magnitude of a given trait occupied by at least one species in a clade or community. For example, body mass spans roughly 8 orders of magnitude in mammals, and so an assemblage of mammal species can potentially fill up to 8 mass bins. If a community consisted of one large mammal of body mass $1.4x10^6$ grams and ten small mammals of between $1.0x10^2$ and $9.9x10^2$ grams, it would have two mass bins filled. In some cases where it was desirable to have larger numbers of bins, such as analyses of traits that span few orders of magnitude of variation in mammals and when producing global maps of bin filling, we used natural log bins instead of $\log^{10}$ bins.

We acknowledge that mammals similar in certain traits can fulfil very different ecological roles in a community (e.g., rodents versus insectivores). However, it at least stands to reason that species that differ by an order of magnitude in a trait such as mass or litter size have important ecological differences. Furthermore, the bin-filling definition of TD is potentially easier to measure accurately than variance or many other TD measures. Most species traits exhibit some degree of allometric scaling (West et al. 1997). In general, medium to large bodied species that occur in communities are known, whereas most unknown species will tend to be small-bodied. Due to the greater overall number of small bodied species in most communities and the potential for such species to reach extremely high abundance, at least a few small bodied species will generally have been

recorded for any given community (Brown 1995). By definition, as long as one species is reported falling into a given trait bin, the bin is considered filled. Thus, estimates of bin filling are potentially more robust to incomplete species sampling than other TD estimates.

Clade level analyses

Twelve traits extracted from PanTHERIA (Jones et al. 2009) were used for clade level analyses. Included traits span a wide variety of sample sizes from 535 to 4642 species (Table 1, Figure S2) and a wide range of phylogenetic dependence (Blomberg's K: Blomberg and Garland Jr 2002, Blomberg et al. 2003), from K = 0.079 to K = 1.855 (Table 1). Despite the wide variation in signal observed, all traits showed statistically significant signal (sensu Blomberg & Garland Jr 2002) when compared to randomized data ($P < 0.01$ in every case). Analyses were conducted for both trait variance and trait bins using a supertree of all extant mammals (Fritz et al. 2009). We analysed correlations using Spearman's rank-order test, and compared Spearman's ρ from models based on PD and SR to see which was more tightly correlated with TD (see electronic supplementary materials for more details). All the clade analyses were conducted in R v 2.12.2 (R Development Core Team 2011), phylogenetic analyses were performed using the package APE v 2.7 (Paradis et al. 2004).

Geographic Analyses

Geographic analyses focused on two traits for which we had particularly good species level sampling, adult body mass and geographic range area (Table 1, Fig. S3).

Both traits are considered important in ecology, evolution, and conservation biology (Brown 1995, Cardillo et al. 2008, Fritz and Purvis 2010a, Davies et al. 2011). Body mass displays strong phylogenetic signal, while range area shows weak phylogenetic signal. Geographic distribution data and threat status data of all terrestrial mammals were obtained from the IUCN database (www.iucnredlist.org) and were matched to the taxonomy (Wilson and Reeder 2005) used in the phylogenetic and trait data.

SR, PD and the two measures of TD using two traits, and the hypothetical reduction of them given extinction of threatened species were calculated for assemblages occurring in each 100 × 100 km^2 grid cell of an equal-area projection of the world's land surface. Only species with data available for at least one of the two traits, geographic distribution and phylogeny were included in the two series of analyses (see Table 1 for sample sizes). In order to investigate whether PD is a better indicator of TD than SR, we calculated residual PD from a global loess regression of PD and SR (Forest et al. 2007, Davies et al. 2008) in grid cells and assessed the correlations in regions where PD departs the most from the expectation of SR – the top and bottom five percentiles. Biodiversity loss, measured as SR, PD and TD loss, was defined as the difference between current biodiversity and the biodiversity left after removal of all the species that are currently recognized as threatened by IUCN (Huang et al. 2012a). Therefore, SR loss is the number of threatened species occurring in the region, while PD loss and TD loss are the differences between current estimated PD and TD in the region and the remnant PD and TD if all the threatened species in the region were extinct.

The primary purpose of our geographic analyses is to determine whether SR and PD are reliable predictors of variation in TD across the globe. Therefore, we analysed

spatial correlations using Spearman's rank without correcting for possible spatial autocorrelation (though we are aware of the problem of violating the assumption of data independency). In the case that PD and SR can be used to predict TD, it may well be due largely to spatial autocorrelation among all three variables. However, this would not diminish the utility of PD and SR in predicting TD. Further, a major goal of our analyses is to identify geographic localities with high biodiversity in multiple dimensions. The value of conserving a region with large biodiversity is not necessarily reduced because neighbouring regions also have large biodiversity. To visualize patterns of overlap between areas of high PD and TD, we constructed a map highlighting cells within the top five percentiles of PD, trait variance, and trait bins for both mass and range area, in comparisons with distributions of biodiversity hotspots recognized by Conservation International (CI, www.biodiversityhotspots.org). All the geographic data and results were processed in ArcGIS™ and all the spatial analyses were conducted in R (R Development Core Team 2011).

Results

Clade level analyses

Variances of eight out of 12 traits showed significant correlations with PD, while variances of only three showed significant correlations with SR (Table S1). The observed correlations (Spearman's ρ) were generally higher in models based on PD compared to models based on SR (7 of 8 traits that showed significant correlations). However, both PD and SR were similar predictors of bin filling. Five of 12 traits showed significant correlations for both PD and SR, while one additional trait showed a correlation with SR

alone and one with PD alone. Whether bin filling shows a stronger correlation with PD or SR varied among traits. Blomberg's K did not show a significant correlation with the strength of correlations observed across traits (n = 12; trait variance versus PD: $\rho = 0.164$, $p = 0.609$; trait variance versus SR: $\rho = 0.217$, $p = 0.499$; trait bins versus PD: $\rho = -0.245$, $p = 0.444$; trait bins versus SR: $\rho = -0.168$, $p = 0.604$) or the degree to which PD was a stronger correlate of SR than PD (n = 12; trait variance: $\rho = 0.147$, $p = 0.651$; trait bins: $\rho = 0.111$, $p = 0.733$). There was no correlation between sample size and the strength of correlations observed for trait variance (n = 12; PD: $r^2 = 0.021$, $p = 0.591$; SR: $r^2 = 0.193$, $p = 0.088$) or trait bins (n = 12; PD: $r^2 = 0.046$, $p = 0.427$; SR: $r^2 = 0.017$, $p = 0.589$). Simulations showed that trait variance shows no autocorrelation with PD or SR (Fig. S4). However, bin filling did show significant autocorrelation with both PD and SR, with either variable explaining roughly 45% of the variation in the number of "bins" filled with draws from a random lognormal distribution (Fig. S5).

Geographic Analyses

All three dimensions of biodiversity showed complex patterns of variation across the globe (Fig. 1, Fig. S6, Fig. S7). When trait variance was used as our measure of TD, extremely weak but (generally) statistically significant global correlations between the TD and the PD or SR of regional assemblages were observed (Table S2). In contrast, strong correlations were observed at a continental scale (Table S3). PD loss appeared to be a good predictor of TD loss in terms of trait variance, especially for body mass, at both global and continental scales; the number of threatened species (i.e., threatened SR) was also correlated with TD loss though to a lower degree for body mass (Table S2 & S3; see

the electronic supplementary material for additional discussion of these results). We further found that TD was significantly correlated with PD and SR in cells where residual PD was either very high (top 5 percentile), such as those in the Eastern to Southern Africa and India, or very low (bottom 5 percentile), such as those in Western North America and the Amazon area (Table S4, Fig. S6). However, in these areas, PD did not appear to be a much stronger predictor of TD than SR.

When the number of trait bins filled by species in a grid cell was used as our measure of TD, PD and SR showed strong positive correlations with TD (Table S2 & S3). Based on comparison of the correlations, PD had a stronger predictive power than SR for the TD of geographic range area but not adult body mass. SR and PD losses also showed significant correlations with the number of bins that would be "emptied out" if all threatened species were to go extinct, though the correlations observed were much stronger for mass than range area (Table S2 & S3; see the electronic supplementary material for additional discussion of these results). At a global scale, there was relatively little overlap between the areas of highest diversity for PD and the two measures of TD. Areas of high PD and trait variance were generally outside CI hotspots. However, trait bin "hotspots" were generally well captured by the CI hotspots (Fig. S8).

Discussion

Whether or not PD is a "silver bullet" that captures TD in addition to SR depends on the definition of TD employed and the scale of analysis. At the clade level, when TD is defined as trait variance, models based on PD are generally a better fit than models based on SR (Table S1). However, when TD was defined as bin filling, which of the two

was a better fit varied from one trait to the next and both showed similar overall performance (i.e., when SR showed a significant correlation PD generally also did and vice versa). In general, PD is at worst a similar fit as SR to models of TD for clades, and it is often a better fit. When both PD and SR are available for mammal clades PD is generally to be preferred as a surrogate for expected TD, though the difference between the two measures is sometimes modest (i.e., $\Delta \rho < 0.2$ in some cases). Surprisingly neither phylogenetic signal nor sample size seemed to explain the strength of correlations observed between PD, SR, and TD across traits. Phylogenetic signal also did not explain the degree to which models based on PD were superior predictors of TD than models based on SR. The strength of observed correlations varied in a highly idiosyncratic way across traits in clade level analyses, and further study is needed to investigate the factors that cause the diversity of a trait to show strong or weak correlations with SR and PD. In particular, factors such as the model of trait evolution that best describes variation in different traits or the degree to which various traits seem to be under stabilizing selection may be informative.

Geographic analyses showed that neither SR nor PD is correlated with adult body mass variance of regional assemblages. In contrast, both SR and PD were significantly correlated with geographic range variance (Table 2). In addition to differences in their biological and ecological significance, differences in data availability and the range of variation between range area and mass may also have affected their correlations with SR and PD. Estimates of range area were available for almost all species of mammals, whereas estimates of adult body mass were available for only approximately 60% of all species. In addition, range area spans 11 orders of magnitude in terrestrial mammals

while body mass spans only seven. We also found that within continental regions, correlations of PD and trait variance do frequently occur, but the relationship varies among different continents (Table S2). We suspect that these patterns are shaped at least partly by the evolutionary history of mammal assemblages that occur in each region. When we focused our analyses on areas where PD departs from what would be expected from SR, both PD and SR showed statistically significant correlations with TD (Table S3). In high residual PD areas, mainly located in central Africa, PD is positively correlated with mass variance as expected from the clade-level analyses. Within these high residual PD areas, species diverged from each other in the relatively distant past and so in many cases are more morphologically distinctive from each other than species occurring in other regions of the world (Forest et al. 2007, Davies et al. 2008). Thus, as more species are sampled trait variance is likely to increase. In contrast, in areas such as Western North America, Amazon Basin, and South Asian islands where PD is much lower than expected from SR, there is a negative correlation between PD and TD. This is probably because these areas contain clades that arose from recent speciation events, resulting in high redundancy of traits among species and leading to a reduction in trait variance as more (relatively similar) species are sampled (Forest et al. 2007, Davies et al. 2008).

Our analyses of geographic range variance showed very different results from those of body mass variance. Not only were consistent correlations amongst PD, SR, and range area variance observed, but smaller scale analyses also showed that correlations between PD and geographic range variance were negative in areas of both high and low residual PD. This is likely related to the fact that areas of high and low residual PD are

often also areas of high species richness (Davies et al. 2008), and within these regions most species tend to have small geographic ranges (Figure S5, see also (Gaston 1996)). Thus, as more species are sampled species with similarly small geographic ranges will tend to be sampled, deflating variance. More generally, since geographic range area in mammals has very low phylogenetic signal (Table 1), the relationship between range variance and PD is likely mostly shaped by biogeographic mechanisms not directly considered by our study.

When we used a definition of TD that is more closely related to niche filling (i.e., trait bins) than trait variance, we found a strong relationship amongst PD, SR, and TD in regional assemblages. For body mass the correlation between trait bins and PD (and SR) in grid cells with high and low residual PD are similar to what we found for trait variance. However, for range area, grid cells with high residual PD actually showed a strong positive correlation between range area bins and PD (Table S3), and areas of low residual PD did not show any significant correlations. The latter pattern likely occurs due to the fact that areas with low residual PD are also areas where there are many species that are the product of recent speciation events (Davies et al. 2008). Such species will tend to have (similarly) small geographic ranges, and will tend to be ecologically similar (i.e., exhibit similar body mass).

Our results support bin filling as a useful measure of TD that captures different information about variation in TD among communities than trait variance. As bin filling is closer to the idea of niche filling than simple variance, it conceptually resembles how community ecologists and conservation biologists tend to think of TD than trait variance. For example, if two out of ten order of magnitude body mass bins have been emptied in a

community due to local extinction, it is safe to conclude that the ecological structure of the community has been profoundly altered; some ecological niches have been entirely vacated. In contrast, noting that the body mass variance in a community has been reduced by 20% due to local extinction conveys relatively little information on impacts to the overall structure of the community. Mass variance can even increase if species with intermediate traits go extinct from a community. Bin filling also has an analytical advantage over trait variance in that it can be much easier to quantify. In order to accurately quantify the trait variance of an assemblage, a complete list of species and their trait values are needed. In order to accurately assess bin filling, it is only necessary to know of at least one species that falls into each bin interval. An important caveat is that bin filling does show some degree of autocorrelation with SR, and thus with PD (Fig. S5). While the autocorrelation is not nearly as extreme as that shown for functional diversity (sensu [5], Figure S1), in cases where a measure of trait diversity that is completely independent of SR is desired bin filling should not be used. We also note that an aspect of TD not directly quantified by either trait variance or bin filling is the range of species trait values (e.g., functional range sensu (Mason et al. 2005)). An ideal measure of niche filling would combine information on both the number and range of niches that are filled.

Finally, preliminary analyses showed that there is relatively little overlap between the geographic regions that show the highest levels of TD and PD, and relatively little correspondence between the two measures of TD (Fig. S8). Current identified biodiversity hotspots also encompassed relatively few areas of high PD and trait variance. Somewhat surprisingly, however, they encompassed many areas where TD in terms of

bin filling is highest, particularly in Asia. We speculate that such areas represent relatively intact assemblages where species with very small geographic ranges and with large body mass still occur. This represents an exciting area for future research.

In summary, our results support the idea that PD is often a reliable measure of the overall biodiversity of an assemblage, including TD. However, two caveats apply. First, PD will generally only be a useful surrogate measure of spatial patterns of trait variance in areas with higher PD than would be expected from their SR, and the difference in the variation explained by the PD vs SR alone is generally modest (e.g., PD only generally outperforms SR by $\Delta \rho = 0.15 - 0.35$ in clade level analyses). Second, PD and SR show similar correlations with bin filling. In cases where researchers are primarily interested in bin filling or similar measures of TD (e.g., niche filling) the additional effort needed to quantify PD compared to SR (i.e., estimating the phylogenetic relationships of species) may not be warranted. Nevertheless, PD remains the only measure of biodiversity that can simultaneously capture variation in SR, TD and genetic diversity above the species level.

Acknowledgement

We thank Nate Nibbelink, Jonathan Davies, Susanne Fritz and the Altizer laboratory group and Ezenwa lab group at Odum School of Ecology for comments. This work was partially supported by a NSF grant (no. 0642619) and the Odum School of Ecology.

Table 1: Traits included in clade analyses.

traits	samples size	Blomberg's K
geographic range area (km^2)	4762	0.079
adult body mass (g)	3468	2.146
litter size (no. of offspring)	2478	0.504
adult head-body length (mm)	1910	1.554
gestation length (day)	1359	3.172
population density (number/km^2)	950	0.286
adult forearm length (mm)	892	1.671
litters per year (no. of litters)	889	0.358
home range (km^2)	703	0.419
teat number (no. of teats)	628	0.454
basal metabolic rate (ml o_2/hr)	571	0.834
mass specific metabolic rate (ml o_2/hr)	571	0.888

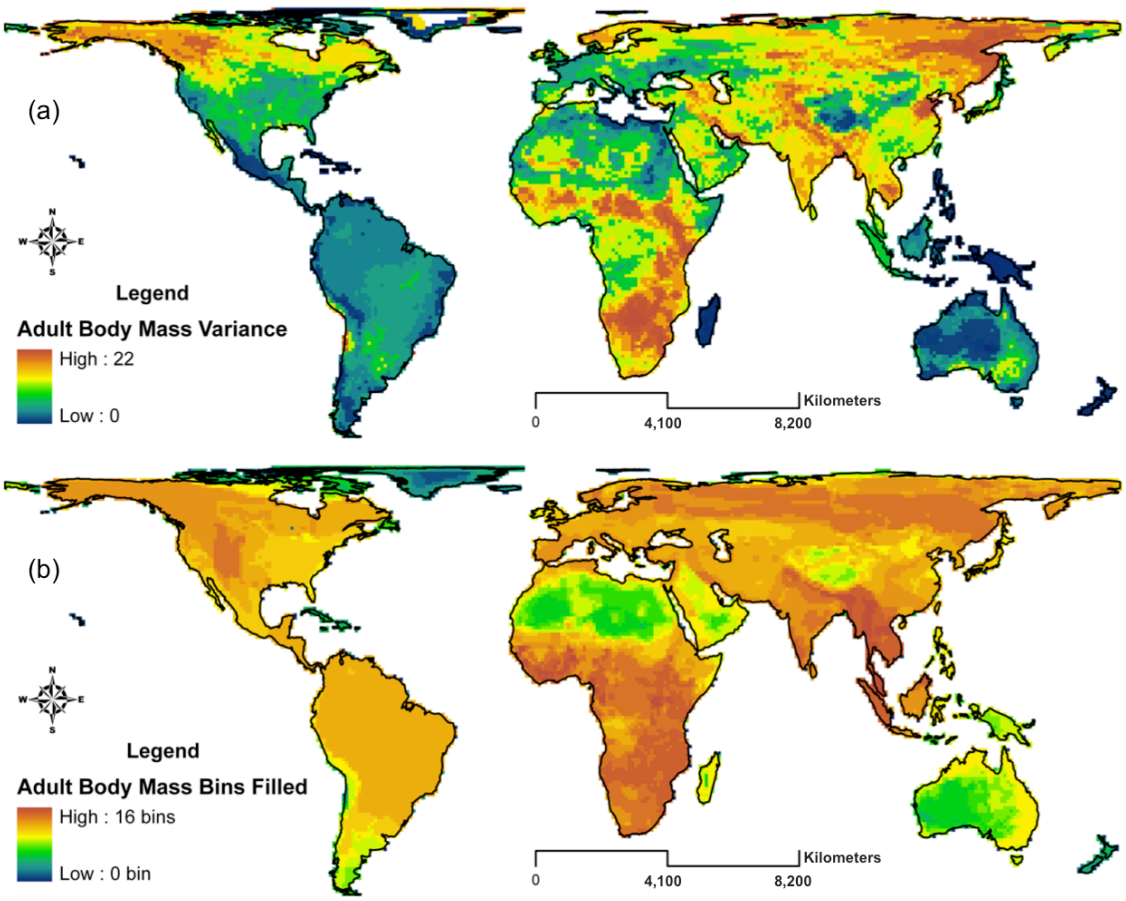


Figure 1: Global variations of mammalian trait diversity in adult body mass measured using (a) variance and (b) bin filling.

CHAPTER 3

HOW GLOBAL EXTINCTIONS IMPACT REGIONAL BIODIVERSITY IN MAMMALS[1]

[1] Huang, Shan, T. Jonathan Davies, John L. Gittleman. 2012. *Biology Letters*. Reprinted here with permission of publisher.

Summary

Phylogenetic diversity (PD) represents the evolutionary history of a species assemblage and is a valuable measure of biodiversity because it captures not only species richness but potentially also genetic and functional diversity. Preserving PD could be critical for maintaining the functional integrity of the world's ecosystems, and species extinction will have a large impact on ecosystems in areas where the ecosystem cost per species extinction is high. Here, we show that impacts from global extinctions are linked to spatial location. Using a phylogeny of all mammals, we compare regional losses of PD against a model of random extinction. At regional scales, losses differ dramatically: several biodiversity hotspots in southern Asia and Amazonia will lose an unexpectedly large proportion of PD. Global analyses may therefore underestimate the impacts of extinction on ecosystem processes and function because they occur at finer spatial scales within the context of natural biogeography.

Keywords

Phylogenetic diversity; biodiversity; threatened species; mammals; extinction.

1. Introduction

Phylogenetic diversity (PD (Faith 1992)) has been suggested as a valuable measure of biodiversity because it not only takes into account species richness but could also reflect genetic and functional diversity (Mace et al. 2003, Forest et al. 2007, Davies et al. 2008). Therefore, potential PD loss may provide an index of ecological vulnerability. Recently, Fritz and Purvis demonstrated dramatic geographic variations in

potential PD loss across terrestrial ecoregions (Fritz and Purvis 2010a). However, we still lack spatially explicit models describing the patterns of potential PD loss at the finer scales more relevant to practical conservation. In addition, the mechanisms explaining differences in PD loss remain poorly understood. Using a complete phylogeny (Fritz et al. 2009) and a comprehensive range database for terrestrial mammals (Schipper et al. 2008), we estimate impacts of global extinctions at a finer spatial scale than used conventionally in global studies (e.g. Grenyer et al. 2006, Davies et al. 2008) and provide novel insights into species' conservation values.

Knowledge of regional extinction patterns is important for developing strategies to conserve biodiversity. First, threatened mammal species richness varies dramatically across space (Grenyer et al. 2006, Schipper et al. 2008). Second, many drivers of extinction risk are shown to be geographically restricted, including environmental factors and anthropogenic disturbance (Cardillo et al. 2008, Fritz et al. 2009). Finally, the loss of particular species can impact different communities or ecosystems differently necessitating geographically specific assessment.

There are two non-exclusive mechanisms that explain high PD loss. First, species on longer terminal branches may be at higher risk of extinction than those subtending from shorter branches (Nee and May 1997). Species on longer branches are considered more evolutionarily distinct because they have no close relatives. At the global scale, such species have been termed Evolutionary Distinctive Globally Endangered (EDGE) (Isaac et al. 2007), and have attracted attention of conservationists in the past decade (Isaac et al. 2007, Mooers et al. 2008, Redding et al. 2010). Alternatively, species might have close relatives in other regions but be locally distinct. Second, if factors that

increase species extinction risk are phylogenetically clumped, PD loss may be high because we risk losing the internal branches connecting the species (Purvis et al. 2000, Sechrest et al. 2002). Phylogenetic structure in extinction risk has recently been demonstrated globally in mammals (Fritz and Purvis 2010b) but has never been assessed in a geographic context.

Here, we present the first global assessment of potential PD loss to focus on geographic patterns and mechanisms at regional scales. We 1) compare PD loss due to extinctions of threatened species with a global random extinction model; 2) identify localities with higher PD loss than expected from regionally random extinctions; and 3) assess the mechanisms explaining extra PD loss in these localities.

2. Materials and Methods

We combined two comprehensive databases on mammal distributions (Schipper et al. 2008) and phylogeny (Fritz et al. 2009). We considered 4796 terrestrial species for which we have both data on both phylogeny and geography. Excluded are recently described species with insufficient phylogenetic data; such species tend to capture little unique PD because they frequently represent recently elevated subspecies (Fritz et al. 2009).

Threat status was obtained from the IUCN (http://www.iucnredlist.org/ see Table S1). A total of 1004 mammal species have been recognized as threatened (listed as vulnerable, endangered or critically endangered). Species with deficient data for evaluating threat status were considered unthreatened, although we are aware of suggestions that such species are likely to be threatened (Purvis et al. 2000). We defined

PD loss as the difference between current PD and the remaining PD assuming the extinction of all threatened species (see Figure S1).

Distribution data were processed in ArcMap™ 9.2, using 100 × 100 km grid cells on a Behrman equal area projection of the world. All analyses were performed in R (http://www.r-project.org/) with the Ape package (Paradis et al. 2004).

For each level of extinction intensity (11%-60% extinct), we performed 1000 replicates, assigning extinction at random, to generate a null distribution of PD loss. To understand regional patterns, we additionally randomized extinction risk among species within each grid cell. The losses in 1) total PD, 2) terminal branch length and 3) internal branch length were calculated for each simulation respectively. If a cell's expected loss was higher than 95% of the losses predicted from simulations, this cell was considered at risk of significant higher loss than expected from random extinction.

3. Results

The proportion of the mammalian tree of life that is lost through random extinction increases approximately linearly with extinction intensity ($r^2 = 0.997$) (Figure S2). Of current mammalian PD, 14% would be lost if all currently threatened species become extinct. This result falls within the two-tailed 90% interval ([8151, 8961]) from our null model of random extinction (Table S1).

Threatened mammals occur in approximately 76% of global land areas and the intensity of threatened PD varies across space (Figure 1a). Large areas in Southeast Asia and western Madagascar are at risk of losing a large proportion of PD, while the proportional loss in PD in the rest of Africa appears relatively low. Among cells

harbouring threatened species, 22% will lose greater PD than predicted by random extinction (Figure 1b and S1).

The high PD loss we observe in the Amazon and Borneo in Indonesia can be largely attributed to the loss of locally distinctive species (Figure 1b, S3 and S4); whereas local mammalian diversity in South Asian and Mesoamerica, and the Ural-Caspian Steppes is threatened with the loss of many internal branches (Figure 1b, S3 and S5).

4. Discussion

In contrast to the optimistic hypothesis that much of the tree of life might survive through a large extinction event (Nee and May 1997), loss of mammal PD increases rapidly with extinction intensity. Over the complete mammal tree, the PD lost if currently threatened species become extinct is not significantly different from that predicted by chance. However, PD lost regionally varies strikingly across space. We used a null model of phylogenetically random extinction to search for areas where PD loss is higher than expected based simply on the number of threatened species. Estimated PD losses in most areas do not depart from random expectations; however, in some areas PD loss is much greater. For example, although the total amount of PD to be lost in the Amazon is not striking, PD is lost at a greater rate than that expected from random extinctions. Global extinctions that result in little loss of global PD may therefore still have a large impact on ecosystems in some areas because they can result in a disproportionate loss of regional diversity. Critically, estimating global PD losses will tend to underestimate the impact of extinctions on regional biodiversity and may thus misinform conservation prioritizing actions.

The loss of PD reflects the number of threatened species plus the phylogenetic structure of extinction risk; here we focus on explaining the latter. Large PD loss might result either because threatened species are evolutionarily distinct, or because they are clustered within clades so that close relatives share similar vulnerabilities. We show that the relative importance of these two mechanisms differs between regions. Higher-than-expected PD loss in the Amazon can be largely attributed to the loss of distinctive species. For example, in one Amazon cell (Figure S3), three of the 9 threatened species are in the 5% most evolutionary distinct quantile of the local assemblage: Amazonian manatee *Trichechus inunguis* (with a terminal branch as long as 98.9 myr), South American tapir *Tapirus terrestris* (85.3 myr) and Pacarana *Dinoimys branicki* (45.3) (Figure S4). The Pacarana is also evolutionarily distinct globally (1% quantile). However, the manatee and tapir might not be considered globally distinctive (time to most recent common extant ancestor globally: 19.5 myr and 8.1 myr for *Trichechus inunguis* and *Tapirus terrestris,* respectively), and therefore may be undervalued by global biodiversity metrics, such as EDGE (Isaac et al. 2007). Nonetheless, these species contribute substantially to regional biodiversity and should be considered as species of conservation importance. Other examples include the Asian wild ass *Equus hemionus* (endangered) in the Himalayas and the South American Tapir (vulnerable) in the Amazon, which have closely related species on the tree of life for mammals (< 3 myr apart), but are highly evolutionary distinct within their local communities (over 80 myr distant). Fine-scaled analyses are critical for identifying such species. In contrast, extinctions within South Asia and Mesoamerica would result in the loss of many internal branches, indicating that extinction is phylogenetically clumped. In one cell within the South Asian biodiversity

hotspot (Figure S3), extinctions would result in the loss of seven of the eight primate species removing several internal branches from the local ecosystem (Figure S5).

We have demonstrated that the impact of global species extinctions on regional biodiversity strongly depends on geographic locality and the regional species assemblage. Global analyses may underestimate impacts of extinction on ecosystem processes and function because they occur at finer spatial scales. Although extinction risk has been previously shown to demonstrate strong phylogenetic signal (Fritz and Purvis 2010b), we find that the phylogenetic distribution of risk varies considerably among regional mammal assemblages. It is critical, therefore, that phylogenetic approaches to conservation are considered within a spatially explicit framework. Furthermore, although national extinction risk of species in some countries mostly concurs with the global extinction risk assessment, species that are not recognized as globally threatened could potentially be lost at a regional scale (Brito et al. 2010). Future data collection allowing assessments of potential PD loss from regional extinctions would likely amplify our results and vastly improve our estimates of the true impacts of species extinctions.

Acknowledgements

We thank Patrick Stephens and Richard Grenyer for discussion and comments on this work. Funding was received from the National Science Foundation.

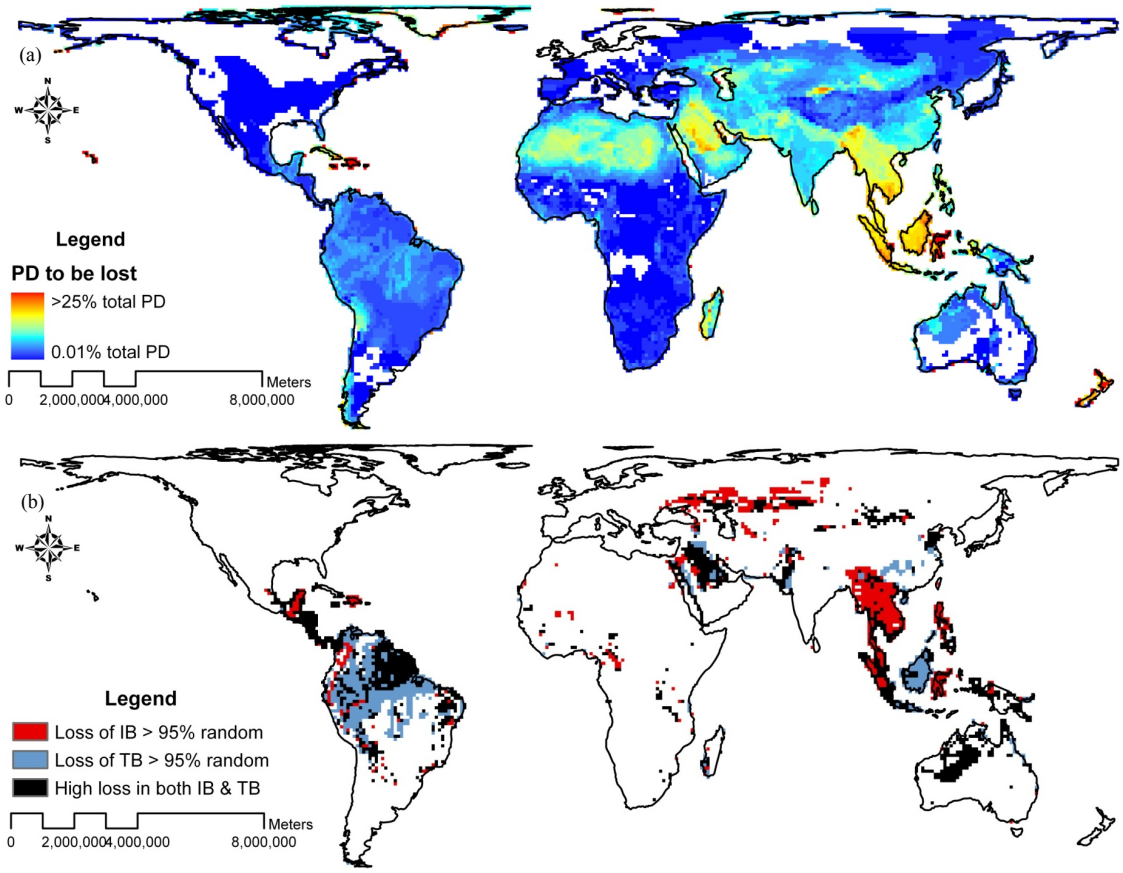


Figure 1. Predicted PD to be lost with extinction of currently threatened species within each 100 × 100 km grid cell. The colours in (a) represent the percentage of PD to be lost, from low (blue) to high (red). The coloured areas in (b) indicate cells with risk of higher loss in PD than 95% random: higher loss in internal branch length (IB, red), higher loss in terminal branch length (TB, blue), and a combination of both (black).

CHAPTER 4

PARASITE DIVERSITY DECLINES WITH HOST EVOLUTIONARY DISTINCTIVENESS: A GLOBAL ANALYSIS OF CARNIVORES[1]

[1] Huang, Shan, John M. Drake, John L. Gittleman, and Sonia Altizer. To be submitted to *Ecology Letters*.

Abstract

Older and more distinct host lineages could harbor fewer parasite species owing to fewer opportunities for parasite sharing with closely related hosts and parasite losses over evolutionary time. We examined this idea using data from over 900 species of parasites and pathogens reported to infect free-living carnivores at a global scale. We applied non-parametric richness estimators to calculate parasite diversity across a subset of well-sampled hosts and assessed how well host evolutionary distinctiveness, relative to other biological and environmental factors, explained variation in the diversity of parasites reported across hosts. We found that host evolutionary distinctiveness correlated negatively and most strongly with parasite diversity: carnivores sitting at the end of longer terminal branches of the phylogenetic tree had lower parasite diversity. Smaller-bodied carnivores and those with narrower geographic ranges also had fewer parasite species. We further tested whether fewer opportunities to share generalist parasites or less intense sexual selection (as might be mediated by parasites) could explain lower parasite diversity in more distinct carnivore lineages, but found limited support for these ideas. Our results indicate that hosts and parasites might have co-evolved through complex mechanisms. Nevertheless, host evolutionary history is an essential predictor of parasite diversity.

Key words

Parasite diversity, carnivore, evolutionary distinctiveness, phylogeny.

Introduction

Parasites account for a large fraction of global biodiversity and can both influence and be affected by the diversity of their hosts (reviewed in: Poulin and Morand 2000, Altizer and Pedersen 2008). In particular, recent work on cross-species transmission of parasites suggests that closely related host species are more likely to share generalist infectious agents and support evolutionary host shifts, and could thus collectively support higher parasite biodiversity (Nunn et al. 2004, Davies and Pedersen 2008, Streicker et al. 2010). At the same time, host species declines can lead to parasite extinction risk as emphasized by recent work on the conundrum of co-extinctions (Koh et al. 2004, Dunn et al. 2009), with the endangerment of evolutionarily unique hosts also placing their specialist parasites at risk. Thus, investigating factors that govern parasite diversity in wild populations, including the role of host ecology and evolutionary history, can help identify underlying drivers of biodiversity and predict the future loss of parasite species arising from host extinctions.

Parasite diversity in carnivores and other mammal groups has been shown to increase in host species with greater body mass, longevity, population density and geographic range size (Poulin 1995, Morand and Poulin 1998, Nunn et al. 2003a, Nunn et al. 2005, Ezenwa et al. 2006, Lindenfors et al. 2007). Other work showed that environmental variables such as proximity to the equator and temperature within the host range also predicted greater parasite diversity across host species (Poulin and Morand 2000, Krasnov et al. 2005, Nunn et al. 2005). Relatively less attention has focused on host evolutionary history as a predictor of parasite diversity, which is surprising because host evolutionary history affects processes such as host-parasite co-evolution, cross-

species transmission and opportunities for parasites to colonize (and be lost from) host lineages over evolutionary time (Page 2003, Nunn et al. 2004, Reed et al. 2007, Davies and Pedersen 2008). Here, we investigate the importance of host evolutionary distinctiveness in explaining parasite diversity in wild carnivore hosts, relative to other host biological/ecological traits previously shown to be important for explaining parasite diversity.

One hypothesis is that evolutionarily distinct host species will harbor fewer parasite species than host species that diverged more recently (Nunn et al. 2004). This pattern could arise if parasites affecting host fecundity and survivorship can induce strong selection pressure on their hosts to evolve defensive strategies that might ultimately drive host diversification (Dybdahl and Lively 1998, Johnson and Clayton 2003, Karvonen and Seehausen 2012, Marston et al. 2012). For example, work on interactions between bacteria and virulent phage shows that pathogens can drive allopatric divergence among host populations by selecting for anti-parasite defenses linked to different suites of host traits in different populations (Buckling and Rainey 2002). In mammals, sexually-selected traits that signal heritable resistance to parasite infection could facilitate divergent adaptation among populations owing to divergence in female mating preferences for male traits associated with resistance to infection (Maan and Seehausen 2011). Cross-species transmission processes could also cause negative associations between host evolutionary distinctiveness and parasite diversity. For example, past work on primate parasites showed that closely related host species are more likely to share generalist parasites than distantly related hosts (Davies and Pedersen 2008). Other work showed that opportunities for cross-species transmission and host shifts of pathogens are

more common among closely related hosts (Streicker et al. 2010). Both of these patterns could arise if divergent host species are more distinct biologically, immunologically or ecologically in ways that affect their susceptibility to parasite infection (Harvey and Pagel 1991, Harvey 1996, Freckleton et al. 2002). Thus, evolutionarily distinct hosts could harbor lower parasite diversity owing to fewer generalist parasites and fewer opportunities to acquire new parasites through cross-species transmission.

An alternative hypothesis is that evolutionarily distinct host species might be infected by a greater number of parasites if host species from more diversified clades experience parasite loss in one or more descendant lineages after host species divergence occurs (termed as the "missing the boat" process in Page 2003). Thus, host diversification events, including those due to pressures unrelated to parasitism, might result in reduced parasite diversity in descendant lineages. Co-evolutionary theories also suggest that bursts of host diversification might occur after the host lineage evolves effective defenses against parasites to escape from infection (Thompson 1994, Yoder and Nuismer 2010). Therefore, host species in clades that have experienced more diversification events might have lower number of parasite species on average than distinctive species. Furthermore, since clade diversification is a net outcome of speciation and extinction, distinctive species might be survivors of those once-large clades that have experienced low speciation rates and/or high extinction rates (Purvis et al. 2000). Such species might represent stable environments for parasites over evolutionary time, and if their parasite carrying capacity allows it, they could retain large numbers of parasite species. Previous studies of bacteria infected with virulent phage also found that infection could limit host

diversification when the host populations were distributed across a heterogeneous environment (Buckling and Rainey 2002, Brockhurst et al. 2004).

Here we use data on parasites and pathogens from free-living carnivores to directly examine the relationship between parasite biodiversity and host phylogenetic distinctiveness, relative to other ecological and environmental variables could affect parasite diversity. Carnivores are one of the best-studied mammal groups and exhibit tremendous interspecific variation across a range of biological and ecological traits (Gittleman 1985, 1989, 1993, Sunquist 2002). Carnivore species occupy geographic ranges from the tropics to the poles, capture extreme differences in body size across species, and exhibit further variation in population density, diet and life history (Gittleman 2001a). Importantly, carnivores have also been relatively well studied for parasites and infectious diseases, in part due to their close relationship to domesticated dogs and cats (Smith et al. 1993, Appel and Summers 1995, Butler et al. 2004, De Castro and Bolker 2005, Lembo et al. 2008), and also because infectious diseases have caused devastating losses and conservation concerns for some carnivore species in the recent past (Alexander and Appel 1994, Kat et al. 1995, Roelke-Parker et al. 1996, Sillero-Zubiri et al. 1996, Murray et al. 1999, Randall et al. 2004).

In this study, we combine multiple global-scale datasets on carnivores and their parasites, here defined broadly to include viruses, bacteria, fungi, protozoa, helminths and arthropods, to test the main hypothesis that host species phylogenetic distinctiveness is an important predictor of parasite biodiversity. Specifically, we used data on parasite occurrence from wild populations based on studies published between 1986 – 2010, and we estimated the number of parasite species, as a measure of parasite diversity, using

modern mathematical estimators rarely applied to global-scale parasite data sets. We then combined parasite data with phylogenetic information from a recently-published carnivore supertree (Nyakatura and Bininda-Emonds 2012) and carnivore biological and ecological trait data (Jones et al. 2009) to examine the relationship between host phylogenetic distinctiveness and parasite diversity, and to compare the relative importance of phylogenetic distinctiveness with other ecological traits of carnivores. Finally, we examined potential explanations for associations between parasite diversity and host evolutionary distinctiveness, including a possible link between parasites, sexual selection and evolutionary diversification, and the idea that more distinct host species experience fewer opportunities to share generalist parasites with close relatives.

Materials & Methods

Global carnivore parasite database

We collated records of parasite occurrence in free-living terrestrial (fissiped) and marine (pinniped) carnivore populations published in the primary literature between 2002 and 2010, and combined them with previously compiled datasets of carnivore parasites from publications spanning the years 1986-2002 (Lindenfors et al. 2007). The data collection procedure was similar to that of the Lindenfors et al. (2007) study, in that systematic searches of the online data base Web of Science (http://isi3.isiknowledge.com) were conducted using carnivore species Latin binomials. We included only primary references of a particular infectious agent and followed a more up-to-date version of mammal taxonomy (Wilson and Reeder 2005), which is consistent with the most recently published carnivore phylogeny (Nyakatura and Bininda-Emonds 2012). A total of 1157

references were included in our updated data set, providing records of 930 parasite species found in free-living populations of 159 terrestrial carnivore species. Because we included all records where host species were sampled for parasites, even if infections were not found, a total of 67 parasite species and 7 host species are retained in our data set even although their records of host-parasite occurrence are negative.

For each host–parasite combination, we recorded parasite type (helminth, protozoan, virus, arthropod, bacteria and fungus), sampling locality, dates of sampling and information on the number of animals sampled. For all host-parasite records, we corrected host species names that deviated from Wilson and Reeder (2005) and updated parasite species names using the most current taxonomic information available. For viruses, we followed names provided by the International Committee on Taxonomy of Viruses (ictvonline.org, last accessed on February 20, 2012). Fungi taxonomy was standardized according to Index Fungorum (http://www.indexfungorum.org, last accessed on February 20, 2012). The taxonomy of other all parasites was standardized according to Global Biodiversity Information Facility (GBIF, http://www.gbif.org, last accessed on February 20, 2012), or, if not available on GBIF, the National Center for Biotechnology Information (NCBI) taxonomy database (http://www.ncbi.nlm.nih.gov/taxonomy, last accessed on February 17, 2012). Most parasite records (when viewed as the number of parasite species, or unique host-parasite combinations) were from helminthes (n = 471 species), bacteria (n = 228 species), arthropods (n = 181 species), protozoa (n = 88 species) and viruses (n = 77 species), with fewer records of fungi (n = 16 species) and rotifers (n = 2 species). The best represented host groups in the data set were mustelidae (n = 30 species), felidae (n = 28 species) and canidae (n = 28 species), with a total of 13

carnivore families included in the data set, covering 56% of the currently recognized carnivore species.

Comparisons of parasite diversity estimators

Observed parasite diversity for each carnivore species was quantified as the number of parasite species reported for which prevalence was > 0 for at least one sampled population of the host species. Parasite species that were identified to genus but not to the species level were counted only when no other species from the same genus has been detected in the same host species. Parasite species that were not identified to the genus level were excluded in this study. Carnivores not identified to the species level were also excluded from analysis, and we did not treat sub-species of hosts or parasites as separate taxa in computing observed parasite diversity.

Past studies showed that observed parasite diversity correlates strongly with the degree to which host species have been sampled (Poulin 1995, Walther et al. 1995, Poulin and Morand 2000, Nunn et al. 2003a, Ezenwa et al. 2006, Lindenfors et al. 2007). One popular approach to control for uneven sampling effort in comparative analyses is to include measures of sampling effort as a covariate in multivariate analyses, which is analogous to examining residual parasite diversity from a regression between observed parasite diversity and sampling effort (e.g. Nunn et al. 2003a, Nunn et al. 2005, Lindenfors et al. 2007). While straightforward to implement, this approach provides no information on true parasite diversity for each host and assumes a similar (often linear) relationship between parasite diversity and sampling effort across host species. Nonparametric diversity estimators that control for sampling effort have been widely

used in a variety of biodiversity surveys of free-living taxa, and these approaches have also been applied to comparative analyses of parasite species richness where similar sampling protocols are employed for each host species (e.g. Poulin 1998, Chiarucci et al. 2003, Blakeslee and Byers 2008, Gihring et al. 2012). However, to our knowledge, non-parametric estimators have not been used in broad scale comparative studies of parasite diversity like ours where heterogeneous data were assembled from the primary literature. Nevertheless, we explored different non-parametric estimators for parasite diversity using our host-parasite dataset and compared their performance on a subset of data for well-sampled hosts.

We used three non-parametric estimators that showed low bias and high accuracy in previous parasite diversity studies, including Chao2, Jackknife, and Bootstrap (Colwell and Coddington 1994, Poulin 1998, Walther and Morand 1998, Walther and Moore 2005, Hortal et al. 2006). We also used the more recently developed abundance-based coverage estimators (ACE1 and ACE2) with $\kappa = 2$ (Chao et al. 2006). These estimators augment observed richness with the number of species that were likely to have been missed, using information on sampling intensity and the number of rare species (or in our case, parasites for which very few hosts were infected). Detailed methods for computing each estimate are provided in the Supplemental Materials. To compare the performance of estimators, parasite diversity was examined for the 6 carnivore species best documented for parasites: *Vulpes vulpes* (227 reports), *Procyon lotor* (151 reports), *Meles meles* (84 reports), *Canis latrans* (80 reports), *Canis lupus* (71 reports), and *Phoca vitulina* (63 reports). We treated each paper as a sampling event, much like a random pool of samples from the species of interest. To visualize the comparison between the different estimators

and the reported parasite diversity, we generated accumulative curves using each method for each species. From the N published reports of a host species, we randomly drew sub-pools of reports from 5 to N, 1000 draws of each size, and we counted and estimated parasite diversity based on the sup-pools of reports. To compare the performance of estimators at low sample sizes, we searched for the minimum number of reports required to have at least a 5% chance covering the final estimate (based on all the included reports) for each method. We also used this minimum number as the cut-off sampling effort for including host species in further analyses described below.

Host evolutionary history & ecological traits

We quantified host phylogenetic distinctiveness as the terminal branch length, and two recently developed distinctiveness measures based on the equal-split (Redding and Mooers 2006) and the fair-proportion (Isaac et al. 2007) algorithms, using the most recently published supertree by Nyakatura and Bininda-Emonds (2012). These different measures made different assumptions about how internal branch length should be assigned to descendant species. Terminal branch length only accounts for the evolutionary distance of a species to all the other species in the clade and ignores the effect of past evolutionary history before the most recent diversification event. In comparison, the other two incorporate the length of internal branches into the calculations of distinctiveness differently (Cadotte and Davies 2010).

We obtained host ecological data from the PanTHERIA database (data available online: http://esapubs.org/archive/ecol/E090/184/default.htm; Jones et al. 2009), which compiled species level data from the published literature for 5416 mammal species,

including information on life history, ecology and geography. We included 9 trait variables with data for at least 90% of our host species. These included ecological traits that have been suggested or shown to be correlated with parasite diversity: adult body mass (g), population density (number of individuals per km^2), geographic range area (km^2) and centroid latitude degree of the species geographic range (absolute degree) (Nunn et al. 2005, Lindenfors et al. 2007). Additionally we included four environmental variables: average temperature (degree C), average precipitation (mm), average potential evapo-transpiration (PET, mm) and average annual evapo-transpiration (AET, mm) within each species range, which might be associated with the distribution of the host species and their associated parasites (Guernier et al. 2004, Antonovics 2009, Fuller et al. 2012).

Because initial analyses showed that parasite diversity declined with carnivore evolutionary distinctiveness, we examined two potential mechanisms that could cause this association. First, we examined whether parasite diversity and carnivore evolutionary distinctiveness were linked to sexual selection by focusing on sexual dimorphism in host body mass using the Lovich-Gibbons revised two-step ratio (See supplementary materials for the formula; Gibbons and Lovich 1990, Lovich and Gibbons 1992). We obtained sex specific body mass data for the host species from a carnivore species trait database (Bininda-Emonds and Gittleman 2000) to ask whether host species harboring a large diversity of parasite display strong host sexual dimorphism (Nunn et al. 2004). Second, we asked whether generalist parasites (that infect > 1 carnivore species) were more common in host species that had lower measures of evolutionary distinctiveness, which might cause carnivores with more extant close relatives to have

higher parasite diversity through sharing of generalist parasites (Davies and Pedersen 2008). Here, we examined the proportion of parasites that infect more than one carnivore host species as the response variable, and we calculated this using the reported (rather than estimated) richness measures as our analysis focused on a ratio rather than actual parasite numbers.

Statistical analysis

All analyses were conducted in R (R Development Core Team 2011) using the R packages: ape (Paradis et al. 2004), CAIC (Orme et al. 2009) and picante (Kembel et al. 2010) for phylogenetic analyses, randomForest (Liaw and Wiener 2002) for the random forest analyses, and MASS (Enhances et al. 2011), AICcmodavg (Mazerolle and Mazerolle 2012), and arm (Gelman and Hill 2007) for the multivariate analyses. We first examined the influence of sampling effort on both observed and estimated (Chao2, Jackknife, Bootstrap, ACE1 and ACE2) parasite diversity using the Spearman rank-order test to determine whether sampling effort should be included as a covariate in later analyses. Here, sampling effort was estimated based on the number of reports included in the database. For each analysis described below, we report results for both estimated and reported parasite diversity as response variables in separate models. We tested the reported and estimated richness for phylogenetic signals using significant tests of Blomberg's K (Blomberg and Garland Jr 2002, Blomberg et al. 2003) to determine whether further analyses needed to incorporate phylogenetically independent contrasts ($p \leq 0.05$ for 1000 randomization).

We next investigated the importance of host evolutionary distinctiveness in predicting parasite diversity. Because initial analyses showed that sampling effort correlates strongly with the response variables, we used a multivariate generalized linear model (GLM) with sampling effort included as a covariate. We then used random forest analyses (Breiman 2001) to examine the importance of evolutionary distinctiveness relative to host ecological traits, including adult body mass, population density, geographic range area, centroid latitude of the species range, and environmental conditions (temperature, precipitation, PET, and AET) within the host species' range. Random forest analysis is a machine learning algorithm that assigns the importance of a variable based on the increase in residual sum of squares following variable inclusion. We generated confidence intervals for the importance of each variable through 1000 random forest analyses of bootstrapped data. As an effort to provide information for predicting parasite diversity in understudied host species, we searched for the best GLM models to predict parasite diversity based on AICc scores using the four most important variables identified in the random forest analyses.

Because host evolutionary distinctiveness was negatively correlated with parasite diversity measures, we conducted a final set of analyses to test potential underlying mechanisms. Specifically, we used Spearman rank-order tests to examine how parasite diversity was associated with host sexual dimorphism in body mass (kg) and host longevity (years) to test the prediction that parasite diversity drives host diversification through sexual selection (Nunn et al. 2004). We also used Spearman rank-order tests to compare host evolutionary distinctiveness to the proportion of generalist parasites (i.e. parasites found in more than one host species) to test the prediction that host species from

more recently diverged lineages harbor a higher diversity of generalist parasites through sharing with close relative species (Nunn et al. 2004, Davies and Pedersen 2008).

Results

Reported and estimated PSR

For the six carnivore species best sampled for parasites in the primary literature, accumulative curves of reported parasite diversity showed no sign of saturation (Figure S1). Comparisons of nonparametric estimators showed that Chao 2 and ACE consistently required lower number of reports to cover the final estimates in all cases than Jackknife and Bootstrap methods (Table S2; Figure S1). Therefore, we only report results of Chao 2 and ACE estimates below. Moreover, because ACE1 and ACE2 estimates are identical for all species, we refer to these collectively as ACE estimates below.

Parasite diversity was estimated for 45 host species (Table S1) with at least 9 published studies of parasitism, based on our estimated minimum number of reports required for a > 5% chance of covering the final richness estimate. Although Chao2 and ACE estimates were positively correlated with sampling effort based on the number of reports in our database (Figure S2; Chao 2: $R^2 = 0.232$ with $p < 0.001$; ACE: $R^2 = 0.244$ with $p < 0.001$), the observed parasite diversity was more strongly associated with sampling effort than either of the nonparametric estimators ($R^2 = 0.730$; $p < 0.001$). Sampling effort was retained as a covariate in all further analyses. We did not detect significant phylogenetic signals in observed ($K = 0.163$, $p = 0.227$) or estimated richness (Chao 2: $K = 0.177$, $p = 0.171$; ACE: $K = 0.192$, $p = 0.109$). The total parasite diversity

across all carnivore hosts was estimated from 1722 (Chao 2 estimate) to 2251 (ACE estimate), relative to 930 reported parasite species.

Parasite diversity, host evolutionary distinctiveness and ecological traits

A total of 30 host species were included in our analyses of parasite diversity in relation to host evolutionary distinctiveness and ecology (with complete data for all ecological traits as a requirement for the random forest analysis). All three measures of evolutionary distinctiveness (terminal branch length, equal-split, and fair-proportion) were highly correlated with each other (Figure S3), but only terminal branch length was significantly associated with estimated parasite diversity. In this case, host species on longer terminal branches harbored lower estimated parasite diversity (for both ACE and Chao 2) (Table S3) although observed parasite diversity did not show this same association. Sampling effort was retained as a significant covariate in models for all three measures of parasite diversity. Bivariate analyses of other ecological variables (with sampling effort as a covariate) showed strong associations between estimated parasite diversity and adult body mass, and between observed parasite diversity and mean temperature, mean PET and the absolute latitude degree of the centroid point in the host's geographic range (Table S3).

Random forest analyses showed that variable rankings were different for reported and estimated parasite diversity, but similar for the two non-parametric parasite diversity estimates (Figure 1). For both of the non-parametric richness estimates, terminal branch length, adult body mass, and geographic range area appear to be the most important predictors (in rank order but with overlapping confidence intervals, Figure 1). Partial

dependence analyses of the random forest models show that estimated parasite diversity is negatively associated with host terminal branch length, and positively associatied with host body mass and geographic range area (Figure 2; final psudo-$R^2 = 0.178$ for both Chao 2 and ACE estimates, and see figure S4). In comparison, for observed parasite diversity, sampling effort is the most important predictor, followed by centroid point latitude degree, average AET, population density and the average precipitation in the host geographic range. We repeated the analysis for reported parasite diversity to include all host species regardless the number of reports for parasitism ($n = 63$) and found similar patterns, with the number of reports as a highly important predictor relative to other variables (Figure S5).

The four most important variables identified by each random forest of regression trees for the estimated parasite diversity are also the variables with best data availability. Using these variables, we were able to include 44 host species to identify the best GLM models for predicting estimated parasite diversity. Both measures of estimated parasite diversity were best predicted by terminal branch length, together with sampling effort (Table 1).

Because significant phylogenetic structure was detected in the adult body mass ($K = 0.530$, $p = 0.001$) and terminal branch length ($K = 0.282$, $p = 0.007$), we repeated several analyses using independent contrasts of the variables, even though we detected no phylogenetic signal for parasite diversity (noted above), geographic range area ($K = 0.092$, $p = 0.880$) or sampling effort ($K = 0.145$, $p = 0.459$). According to the best GLM models of independent contrasts, Chao2 estimates of parasite diversity was best predicted by terminal branch length while ACE estimates of parasite diversity was best predicted

by terminal branch length and geographic range area, together with sampling effort (Table S4).

Mechanisms underlying the evolutionary distinctiveness – parasite diversity relationship

We examined the association between estimated parasite diversity and host sexual dimorphism using 41 carnivore species for which data on all variables were available. 35 of these species displayed up to 230% larger body mass in male adults than in females, 1 species did not show dimorphism, and the rest showed up to 25% larger mass in females than in males. Because we detected significant phylogenetic structures in carnivore sexual dimorphism ($K = 0.325$, $p = 0.031$), although not in parasite diversity (Chao 2: $K = 0.181$, $p = 0.268$; ACE: $K = 0.190$, $p = 0.209$), we repeated analyses with and without phylogenetic independent contrasts. To account for sampling effort, we included contrasts of the number of reports ($K = 0.184$, $p = 0.372$) in the analyses. The sexual dimorphism in body mass was significantly correlated with estimated parasite diversity, although no significant associations were found in non-phylogenetic analysis (Table S5; Figure 4). We did not detect correlations between sexual dimorphism and terminal branch length.

Finally, we examined the relationship between host evolutionary distinctiveness and the proportion of generalist parasites (occurring in > 1 carnivore host) to ask whether less distinct host species have higher proportion of generalist parasites due to sharing of parasites with other related host species. We constrained this analysis in relatively well-studied host species (sampled in at least 9 reports, $n = 45$). All host species harbor above

50% generalist parasite species. We found no significant association between the proportion of generalist parasites and the host terminal branch length (Spearman's ρ = 0.041, p = 0.791).

Discussion

Identifying factors associated with high parasite diversity in well-studied host groups such as carnivores is vital for understanding the ecological and evolutionary principles that govern parasite biodiversity. Because most host species on Earth have not been well sampled for parasites, understanding key determinants of parasite diversity can also help predict parasite diversity in understudied hosts for which biological trait data are available in the absence of parasite sampling.

Our analysis of carnivores showed that overall parasite diversity declined with terminal branch length as a summary measure of host evolutionary distinctiveness, and this was the most consistent and robust predictor of parasite diversity of the variables examined here. Carnivore species on longer terminal branches that diverged in the more distant past harbored fewer parasite species than hosts with many recent phylogenetic relatives. This finding agrees with a past study of primates that showed greater parasite diversity in host species from more rapidly diversifying clades (Nunn et al. 2004), and suggests that evolutionarily distinct mammal species, which have fewer living relatives and are often unusual in their physical or behavioral traits, support less diverse parasite communities.

Two other host ecological traits, body mass and geographic range area, were positively associated with parasite diversity in carnivores, a finding that is consistent with

previous comparative analyses of parasite diversity in carnivores (Lindenfors et al. 2007), primates (Nunn et al. 2003a), and ungulates (Ezenwa et al. 2006). Partial dependence analyses using the random forest models suggested that large-bodied carnivores tend to support the most diverse parasite communities. Further, host species distributed in large geographic range areas also tend to harbor many parasite species. In our data set, host species with these characteristics included brown bears (*Ursus arctos*), the leopards (*Panthera pardus*), lions (*Panthera leo*) and grey wolves (*Canis lupus*). However, only terminal branch length was included in our final GLM models as an essential factor for predicting parasite diversity. Surprisingly, when we examined two other measures of evolutionary distinctiveness, we found no significant associations with parasite diversity. In addition to the terminal branch length, these two other measures also account for deeper branches (internal branches), which represent the evolutionary history that occurred before the most recent diversification between an existing species and the rest of the clade represented. Thus, it is possible that the more recent evolutionary history captured by the terminal branch length measure is more important for contemporary host-parasite associations than other measures of host distinctiveness.

Our analyses to explore mechanisms that might account for the association between host evolutionary distinctiveness and parasite diversity focused initially on the idea that parasites might drive host diversification, and that this could be mediated by sexual selection. Specifically, hosts with greater parasite diversity could experience more intense sexual selection on traits that signal parasite resistance or tolerance to prospective mates (Hamilton and Zuk 1982; Andersson 1994; Moore and Wilson 2002). At the same time, sexual selection has been identified as a key driver of speciation, such that species

with more intense sexual selection might also be from clades that recently diversified (Lande 1981; Owens et al. 1999; Maan and Seehausen 2011). Our phylogenetically controlled analyses supported a significant correlation between parasite diversity and carnivore body mass dimporphism, such that host species harboring large parasite diversity tend to display stronger male-biased sexual dimorphism than host species harboring few parasite species. This is similar to the result of an earlier primate study (Nunn et al. 2004), indicating that parasite diversity might play a role in host sexual selection. However, analyses conducted without correcting for phylogenetic structures did not detect a significant association between parasite diversity and sexual size dimorphism. This indicates that the phylogenetic structure of the sexual dimorphism data might have obscured the relationship between sexual dimorphism and parasite diversity (Gittleman 2001b). Meanwhile, we did not detect a significant correlation between terminal branch length and sexual dimorphism in carnivore hosts. We note that the terminal branch length data might not represent the real age of the species as extinct species and lineages are not included in the phylogenetic tree (Nyakatura and Bininda-Emonds 2012). Therefore, further phylogenetic data on real species ages are likely to help to clarify this relationship. We also suggest further work focus on examining host parasite associations through evolution at a fine scale, for example, by estimating the ancestral host status of sexual dimorphism and parasite infection, and identifying evolutionary association between parasitism and sexual dimorphism on the host phylogenetic tree.

We also followed Nunn et al. (2004) by examining the idea that more distinct carnivore species might have lower parasite diversity owing to less frequent opportunities

to share generalist parasites with close relatives. In other words, carnivores from clades with greater numbers of related species might provide greater opportunities for generalist parasites to be shared among hosts, an idea supported by recent work showing that host evolutionary relatedness is one of the most important predictors of parasite sharing (Davies and Pedersen 2008) and host-shifts (Streicker et al. 2010) among free-living animal hosts. However, our analysis of the proportion of generalist parasites in carnivore hosts did not support this idea, in that the proportion of generalist parasites does not vary according to the terminal branch length of the host species. This result is consistent with previous findings in primates, which suggest no association between host diversification and opportunities for parasite sharing between sympatric host species (Nunn et al. 2004). The lower frequency with which evolutionarily distinctive hosts sharing parasites with other host species suggested by a different study (Davies and Pedersen 2008) might be the reason why host species on longer terminal branches have fewer generalist parasites contributing to lower overall parasite diversity. However, the low overall parasite diversity in distinctive host species is more likely due to other mechanisms, such as the impact of parasite diversity on host lineage diversification.

Importantly, this is the first study where non-parametric diversity estimators have been used for predicting parasite diversity in a comparative analysis that uses data built from the primary literature covering broad taxonomic and geographic scales. According to our interpretation of estimated parasite diversity, current records of parasite diversity in carnivores are far from complete. In particular, the total number of parasite species harbored by all carnivore hosts is estimated to be at least double that of reports in the published literature up to 2010. Even in the best-studied carnivore species, the rarefaction

curves for parasite diversity do not asymptote, and all of our four non-parametric estimates of parasite diversity greatly exceed reported parasite numbers. Thus, despite the effort invested in this well-studied group for which infectious diseases have caused significant conservation problems, at least half of the parasites and pathogens from wild carnivores remain to be discovered. With this in mind, our results can help guide further research on parasite diversity in several ways. In particular, our statistical models can identify host species that are currently not well sampled for parasites, but for which we might expect parasite diversity to be relatively high. This might include species such as tigers (*Panthera tigris*), jaguars (*Panthera onca*) and side-striped jackals (*Canis adustus*), that are large-bodied, have large geographic ranges, and/or are from more rapidly diversifying clades, but that are not currently well represented in the parasite literature.

Our analysis also points to problems with conventional approaches to controlling for uneven sampling effort among hosts in comparative analyses of parasite diversity. In particular, previous approaches that include sampling effort measures as a covariate in statistical models and that assume linear relationships between observed parasite diversity and sampling effort (Nunn et al. 2003a, Nunn et al. 2005, Lindenfors et al. 2007) are not ideal for investigation at broad taxonomic scales. This is in part because heterogeneity often exists in sampling strategies, sample materials, sensitivity of detection techniques and sample sizes for different host populations and parasite species based on studies in the published literature. For example, the bovine tuberculosis pathogen *Mycobacterium bovis* (common in Meles meles, Murphy et al. 2010) is frequently sampled mainly because of the significant economic impacts of the disease. Thus, some wildlife hosts are studied over and over again for the same parasite species because they are considered

pathogen sources for livestock. Host species for which these studies take place might be assigned to a low residual measure of parasite diversity based on the high sampling effort, but this by no means indicates that further searches for more parasite species will not result in a large addition to current records of parasite diversity. In contrast, the nonparametric diversity estimators that we used in this study weight the multiple reports focused on the same parasite species differently. Because true parasite diversity is not known for wild carnivore species, we cannot test the accuracy of the estimators used in our study (but see: Poulin 1998). However, our further analyses showed stronger correlations between estimated richness and host evolutionary distinctiveness and ecological factors than with sampling effort, whereas sampling effort remained the strongest predictor of variation in observed richness. Thus, our analyses provide support for the use of diversity estimators for broad-scale studies of parasite diversity, for which parasite diversity is almost always under-sampled. The primary drawback to our use of the nonparametric estimators in this study is the loss of host species from the analysis for which sampling did not meet our minimum threshold of 9 published studies, thus reducing the data set from 159 carnivore hosts with parasite data to just 45 species with estimated parasite diversity.

Carnivores are one of several mammalian orders that have high percentages of globally threatened species, with 26.7% of wild carnivores at risk of extinction or already extinct in the wild (Schipper et al. 2008). Despite ongoing conservation efforts, numerous examples have been reported where species experienced large declines in population size due to infectious diseases, with examples including black-footed ferrets, lions, African wild dogs and Ethiopian wolves (e.g. Thorne and Williams 1988, Alexander and Appel

1994, Roelke-Parker et al. 1996). In many of these cases, spillover of multi-host pathogens such as rabies and canine distemper viruses from domesticated dogs to wild carnivores have caused the greatest concerns. Indeed, past work indicates that the close evolutionary relationships between wild canids, felids and domesticated dogs and cats could enhance the vulnerability of these groups to pathogens circulating in domesticated species (Smith et al. 1993, Appel and Summers 1995, Butler et al. 2004, De Castro and Bolker 2005, Pedersen et al. 2005, Lembo et al. 2008). It would be interesting in future studies to examine key predictors of outbreaks of these multi-host pathogens in wild carnivore hosts, including asking whether the presence of diverse parasite communities can buffer wildlife against novel virulent pathogens. It is also worth noting that recent studies have focused attention on the problem of co-extinctions, whereby specialist pathogens and parasites could be lost following the extinction of their hosts (Koh et al. 2004, Dunn et al. 2009). Thus, analyses such as the one presented here could be extended to predict the numbers of parasite species that might be lost as a result of future carnivore extinctions.

Prediction of future outbreaks and emergent infectious diseases in wild populations requires comprehensive information of host-parasite associations (Harris and Dunn 2010). Although it is still impossible to achieve a complete record of parasite species infecting wild carnivore populations, our finding that evolutionary distinctiveness, body mass and geographic range area are strong predictors of parasite diversity sheds light into the mechanisms shaping parasite distribution patterns. We also developed a simple model for predicting parasite diversity in unevenly and incompletely-sampled carnivore species, which are often species of conservation concern due to low

population densities and restricted geographic distributions (Cardillo et al. 2005, Cardillo et al. 2008, Fritz et al. 2009). Our study also highlights the importance of host phylogenetic information in predicting parasite diversity, and in particular, parasite diversity might have contributed to driving host diversification and promoting host diversity. Parasites have often been considered a threat to biodiversity because of their negative impact on host population persistence (Murray et al. 1999, De Castro and Bolker 2005, Smith et al. 2006), but considered from an evolutionary perspective, parasites might actually play important roles in maintaining and promoting biodiversity.

Acknowledgements

We thank Patrick Stephens, Charles Nunn, Nate Nibbelink and the Altizer and Ezenwa lab groups for discussion and comments on earlier drafts of this manuscript. We thank Michael Cherry, Karen Christ, Matt Foretich, Ashton Griffin, Steven Howell, Tyler Halle, Cecilia Nix, and Jeff Shapiro for assistance with building the database, and Charles Nunn and Patrick Lindenfors for access to the earlier (pre-2002) parasite data. This project was supported by Conservation International, the Odum School of Ecology, and National Science Foundation (Grant #0642619).

Table 1 Generalized linear models (GLM) of estimated parasite diversity. Sampling effort measured as the number of reports was included as a covariate in all models. Note that complete data were available for 44 host species to be included in these models. ΔAICc for each model was calculated as the difference from the lowest AICc.

	Chao 2 estimator		*ACE estimator*	
Models	**AICc**	**ΔAICc**	**AICc**	**ΔAICc**
Terminal branch length (TBL)	74.86	0	77.36	0
Adult body mass (ABM)	88.16	13.30	84.86	7.50
Geographic range area (GRA)	89.47	14.61	84.41	7.05
TBL + ABM	75.37	0.51	78.58	1.22
TBL + GRA	76.48	1.62	79.05	1.69
ABM + GRA	90.08	15.22	86.44	9.08
All	76.67	1.81	80.73	3.37

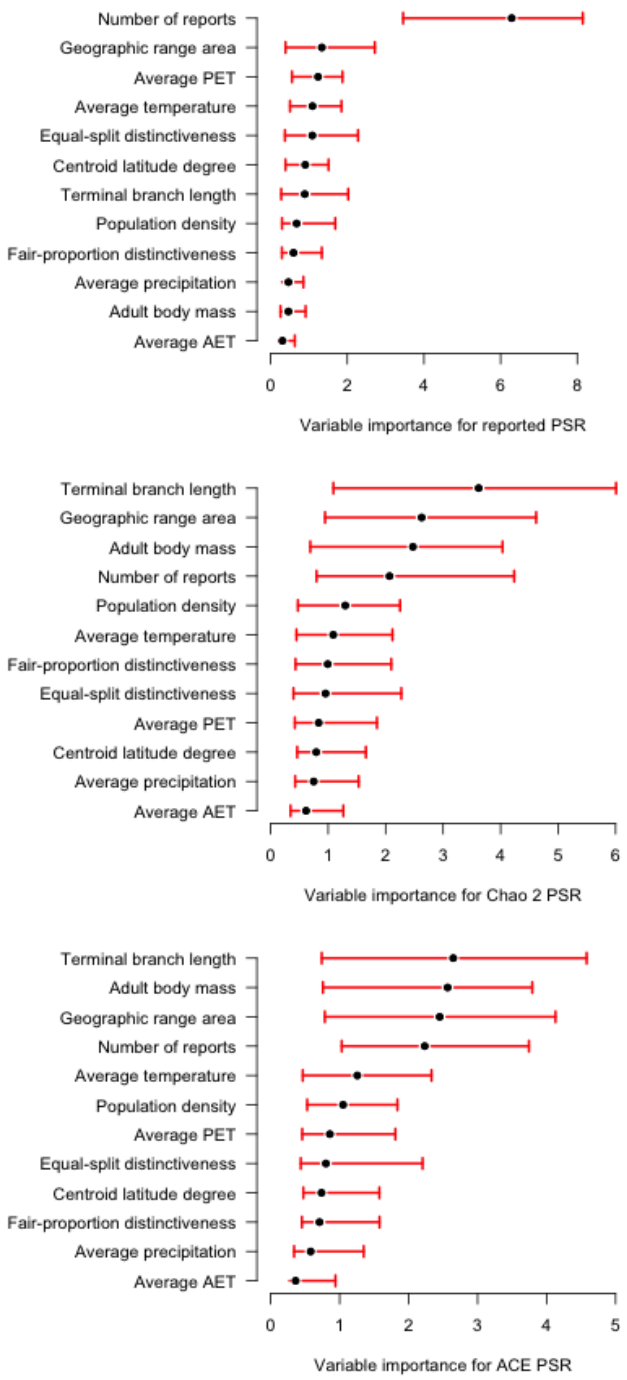


Figure 1 Variable importance determined by the random forest analysis for observed parasite species richness (PSR) (a) and PSR estimated using Chao 2 (b) and ACE (c) methods. The importance of each variable is quantified as the increase of node impurity (residual sum of squares) if the variable is excluded, where a high increase indicates high importance of the variable. The 90% confidence intervals (indicated by error bars) are generated from bootstrapped data (5% and 95% quantiles).

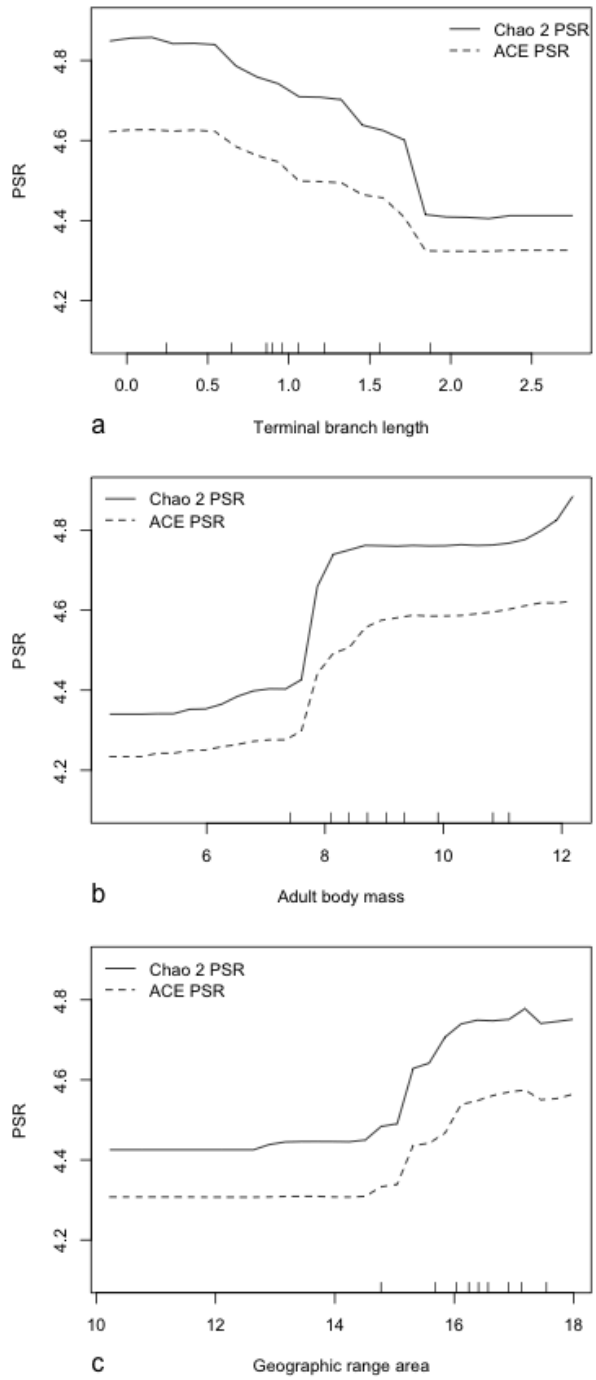


Figure 2 Partial dependency of Chao 2 (solid) and ACE (dashed) estimated parasite diversity (PSR) on the three most important variables from random forest analyses: (a) host terminal branch length, (b) adult body mass, and (c) geographic range area. In each of the three cases, for each value of the variable on the horizontal axes, average PSR estimates were calculated based on the data of all the other variables not included using the random forest model for Chao 2 and ACE estimates of parasite diversity. All axes were natural-log transformed.

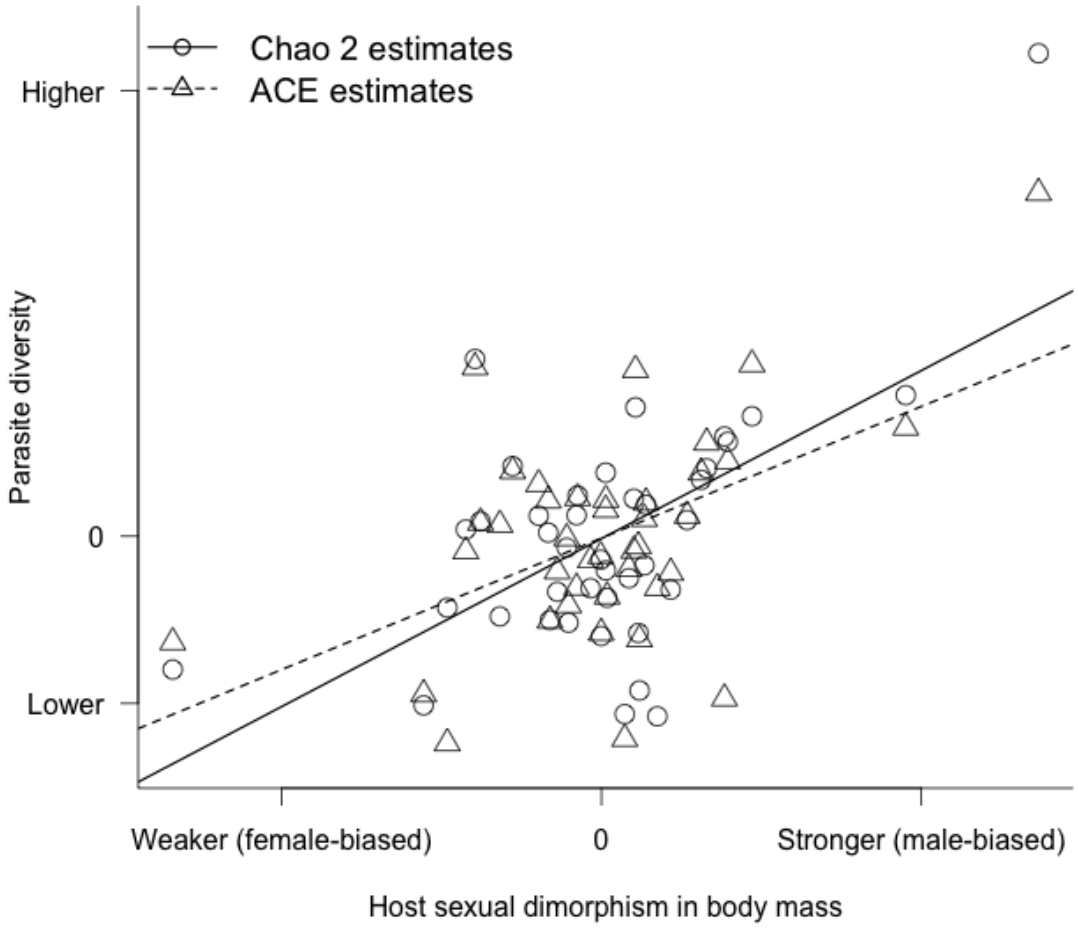


Figure 3 The relationship between parasite diversity and host sexual dimorphism in body mass based on phylogenetic independent contrasts. Parasite diversity was based on the residual estimated parasite diversity (Chao 2 and ACE estimators) from linear regressions with the number of reports. The illustrated trend lines were based on the parameters from bivariate GLM models. All variables have been transformed into phylogenetic independent contrasts.

CHAPTER 5

PHYLOGENETICALLY RELATED AND ECOLOGICALLY SIMILAR CARNIVORES HARBOR SIMILAR PARASITE ASSEMBLAGES[1]

[1] Huang, Shan, Olaf R. P. Bininda-Emonds, Patrick R. Stephens, John L. Gittleman, and Sonia Altizer. To be submitted to *Journal of Animal Ecology*.

Abstract

Many parasites can infect more than one host species. Understanding the broad scale determinants of parasite sharing across host lineages is important for predicting pathogen emergence in new hosts and for estimating pathogen diversity in understudied host species. In this study, we focused on a large dataset of 793 parasite species infecting 64 carnivore host species to show that parasites are more commonly shared between phylogenetically related host pairs. Overall ecological similarity based on multiple life history and behavioral traits also predicted parasite sharing between host species, particularly for viruses. Geographic range overlap between host species was retained as a significant variable in multi-variate models, but with relatively less importance. When viewing the host range for individual parasite species across the carnivore phylogenetic tree, viruses and helminths showed more restricted host ranges than expected by chance, whereas protozoa, bacteria and arthropods were less constrained by host phylogeny. Overall, our results underscore the importance of host evolutionary history as a key determinant of parasite host range, even when simultaneously considering other factors such as host ecology and geographic distribution. Findings here further suggest that knowledge of parasitism in well-studied host species can provide expectations for parasite assemblage composition in their understudied phylogenetically close relatives, especially for helminths and viruses.

Key words

Parasite assemblage similarity, phylogenetic relatedness, ecological similarity, geographic range overlap, Jaccard index, carnivore.

Introduction

One of the most important questions in host parasite ecology and evolution is what limits the host range that a parasite can infect (Anderson and May 1992, Agrawal 2000, Hudson et al. 2002, Weaver and Barrett 2004). Recent advance in host parasite research has heavily emphasized a multi-host-multi-parasite framework (e.g. Esch et al. 1990, Poulin 1992, Poulin and Morand 2000, Holt et al. 2003, Poulin and Mouillot 2003, Pedersen and Fenton 2007, Pugliese 2010, Streicker et al. 2010). Parasites that infect multiple host species can cause severe problems for wildlife host species that may already suffer from high extinction risk due to low population density, because maintenance of such parasites in abundant host species allow the parasites to persist in the threatened host species resulting in parasite mediated extinction (Hudson et al. 2002, Altizer and Pedersen 2008). Additionally, reports of emerging infectious diseases (EID) in human (Jones et al. 2008) and wildlife populations (Drexler et al. 2012, Fuller et al. 2012) have further emphasized the importance of a broad-scale multi-species framework for understanding mechanisms underlying parasite emergence and occurrence in novel host species (Fenton and Pedersen 2005, Keesing et al. 2010).

Two factors have been proposed as important predictors for the likelihood of a parasite being shared between host species: host geographic range overlap and phylogenetic relatedness (Pfennig 2000, Antonovics et al. 2002, Ricklefs and Fallon 2002, Perlman and Jaenike 2003, Poulin 2003, Streicker et al. 2010). Geographically, parasites can only exist where their hosts exist (Poulin 1997, Harris and Dunn 2010). Overlap between host species' geographic ranges provides greater opportunities for hosts to share parasite species compared to hosts that do not occur in sympatry (Antonovics et

al. 2002, Poulin 2003, Streicker et al. 2010). Phylogenetically closely related host species might have inherited similar parasite assemblages from their most recent common ancestor (Page 2003). They also tend to be biologically similar to each other, with low genetic and ecological barriers to cross-species transmission (Pfennig 2000, Ricklefs and Fallon 2002, Gilbert and Webb 2007, De Vienne et al. 2009, Streicker et al. 2010, Longdon et al. 2011). Despite this conventional wisdom about key determinants of host range, most studies to date focused on very narrow scales of natural host parasite associations (e.g. Ricklefs and Fallon 2002, Streicker et al. 2010). We are aware of only one study (Davies and Pedersen 2008) that has considered parasite sharing in a multi-host-multi-parasite comparative framework. Specifically, Davies and Pedersen (2008) identified predictors of parasite assemblage similarity in free-ranging primates populations. Notably, they found whether the host species' geographic ranges overlap or not to be an important predictor of how similar their parasite assemblages are, but the degree of the geographic range overlap seemed less important (Davies and Pedersen 2008). Host phylogenetic relatedness, was shown to be a stronger predictor for parasite assemblage similarity than geographic range overlap, such that closely related primate host species support more similar parasite assemblages than distantly related primates (Davies and Pedersen 2008).

Host species phylogeny represents the evolutionary history through which host lineages diverged and then accumulated genetic and functional differences. Each time a host lineage diverged, both descendants probably are likely to maintain only a proportion of the parasites infecting the ancestral host (i.e. parasites might be "missing the boat", Page 2003). Thus, if each of the two host species or lineages has gone through additional

divergence events after they diverged from each other, this increases the chance for them to lose different species of parasites they inherited from their most recent common ancestor. As a result, relative to host species separated by many divergence events, closely related host species that have gone through few or no additional divergence events should be more likely to share a large proportion of their parasite species assemblages. Additionally, compared with closely related species, species that are connected through a long phylogenetic distance are expected to be more dissimilar genetically, biologically and ecologically (Harvey and Pagel 1991, Harvey 1996, Freckleton et al. 2002). The genetic distance accumulated through evolutionary processes might underlie key differences in host defenses (e.g. immune responses) against parasite infection as well as in other interactions with parasites (Nunn et al. 2000, 2003b, Litman et al. 2005). Further, the phenotypic dissimilarity in host ecological and physiological traits could influence the probability of parasite establishment following cross-species transmission and hence parasite sharing (Nunn et al. 2000, Nunn et al. 2003a, Nunn et al. 2003b, Nunn and Altizer 2006, Lindenfors et al. 2007).

Beyond phylogeny and geography, the ecological similarity between two host species might actually be a third important predictor of parasite assemblage similarity. Ecological similarity between phylogenetically related host species may contribute to why phylogenetic relatedness is often a strong predictor for parasite assemblage overlap in primate systems (Antonovics et al. 2002, Davies and Pedersen 2008). However, no study has specifically tested whether this factor can directly predict parasite assemblage similarity at broad taxonomic scales. Previous studies of the relationship between host ecological traits (e.g. body mass, geographic range area and population density) and

parasite diversity in mammals (Nunn et al. 2003a, Ezenwa et al. 2006, Lindenfors et al. 2007) have implied that host ecology plays an essential role in shaping parasite occurrence patterns. Compared with infecting two host species highly similar in ecology, it could be very challenging for a parasite species to infect two host species with very different ecological traits, such as physical conditions, behaviors, and diets, because the parasite's adaptation to the traits in one host species might result in a maladaptation to the other. Furthermore, while phylogenetic structure is commonly found in many mammal ecological traits, the evolutionary models that best fits each trait and the strength of phylogenetic signal among traits could varies dramatically (Nunn 2011). Therefore, the ecological dissimilarity between two host species might not be fully captured by the phylogenetic related of the host species, and its predictive power needs to be assessed directly.

Carnivore species play key roles in many ecosystems and are known to display dramatic variation in many of their ecological traits and play key roles in many ecosystems (Gittleman 1985, 1989, 1993). Fore example, the average adult body mass of a carnivore species ranges from around 78 g (*Mustela nivalis*) to over one ton (*Mirounga angustirostris*), the average length of gestation period ranges from around 25 days (*Mustela frenata*) to one year (*Phocarctos hookeri*), and different carnivore species can take from 93 days (*Mustela ermine*) to 7 years (*Arctocephalus forsteri*) to reach sexual maturity (Jones et al. 2009). Importantly, carnivores are one of several mammalian orders that have high percentages of globally threatened or extinct species (26.7%) (Schipper et al. 2008). Despite the efforts to conserve threatened carnivore species, there are numerous recent examples of generalist parasites negatively impacting wild carnivore

populations, including rabies virus and canine distemper virus in African wild dogs (Alexander and Appel 1994, Kat et al. 1995), Ethiopian wolves (Laurenson et al. 1998, Randall et al. 2004), Black-footed ferrets (Thorne and Williams 1988), Serengeti lions (Roelke-Parker et al. 1996) and Caspian seals (Kennedy et al. 2000). Thus, understanding the mechanisms underlying parasites sharing among host species has crucial implications for wildlife management and conservation.

Here, we examine the relative importance of host phylogenetic distance, geographic range overlap, and ecological similarity in predicting parasite assemblage similarity among carnivores. Similar to observed patterns in primates (Davies and Pedersen 2008), we expect that phylogenetic relatedness and geographic range overlap between carnivore species will correlate positively with parasite assemblage similarity. We also tested whether the importance of host phylogeny in predicting host ranges differed among different types of parasites. Additionally, we expect that ecologically similar host species that overlap in traits such as body mass, trophic level and life history traits might share more parasites in common either because such similarity could expose them to the same parasites, or particular traits determine whether parasites can establish within individual hosts.

Materials & Methods

Host parasite data

We compiled a global data set of parasite species occurrence in free-ranging carnivore populations published between 1986 and 2010. Briefly, we conducted systematic searches of the online data base Web of Science (http://isi3.isiknowledge.com)

using carnivore species Latin binomials. We included only primary references of a particular infectious agent and followed a more up-to-date version of mammal taxonomy (Wilson and Reeder 2005), which is consistent with the most recently published carnivore phylogeny (Nyakatura and Bininda-Emonds 2012). For each host–parasite combination, we recorded parasite type (helminth, protozoan, virus, arthropod, bacteria and fungus), sampling locality, dates of sampling and information on the number of animals sampled (for more details, see Huang et al. in press). Our data set included 1157 references providing records of 930 parasite species found in free-living populations of 159 carnivore species. Using these data, we summarized all the host and parasite species-specific associations and quantified the similarity between the parasite assemblages in any pair of host species with infection records. Low numbers of parasite species reported from any particular host typically indicates the host species has been poorly sampled for parasites (Huang et al. In prep)[more ref]. Moreover, in order to detect parasite sharing between host species, the hosts in question must each have at least a modest number of parasite species reported from them. Therefore, we only included host species reported with at least 5 parasite species in further analyses (therefore excluding about 60% of host species that were poorly sampled for parasites).

Parasite assemblage similarity

We used two similarity measures in our analyses: the Jaccard index (*J*) and a modified version of Jaccard similarity coefficient (i.e. corrected Jaccard index). The Jaccard index is a commonly used measure for quantifying differences between two (or more) assemblages are (Koleff et al. 2003):

$$J = \frac{|A \cap B|}{|A \cup B|}$$

In our case, A represents the parasite assemblage infecting host species X, and B represents the parasites infecting host species Y. Thus, *J* between the parasite assemblages in X and Y is the ratio between the size of the intersection of A and B (parasite shared by X and Y) and the size of the union of A and B (all the parasite infecting either X or Y). When assemblages A and B differ in size, *J* is limited by the size of the smaller parasite assemblage (potential maximum *J* is equal to A divided by B, if A is a subset of B). In such cases, the difference between parasite assemblages in the two host species indicated by *J* is largely due to the difference in the parasite species richness infecting the two host species instead of the difference in the parasite species identity. Therefore, we developed a corrected Jaccard index (*CJ*) by dividing *J* using the potential maximum *J*:

$$\text{Corrected } J = \frac{|A \cap B|}{|A \cup B|} \Big/ \frac{\min(|A|, |B|)}{\max(|A|, |B|)}$$

We acknowledge that the difference in assemblage size is an important component of the assemblage similarity; in particular, the difference between two host species in their capacities to support parasite diversity has valuable implications for wildlife disease management. However, we also note that the number of parasite species associated with a host species in existing records is strongly influenced by the effort that has been put on sampling the host species for parasites (Nunn et al. 2003a, Lindenfors et al. 2007).

Different types of parasites might require different conditions to infect multiple host species, and some parasite taxa might be more commonly shared among host

species. For example, helminths are typically restricted to specific host ranges as compared to viruses and protozoa (Cleaveland et al. 2001, Pedersen et al. 2005, Szymanski and Lovette 2005). To investigate whether parasite type can be used to help predict parasite assemblage similarity between host species, we grouped parasite species into five major taxa: arthropods, bacteria, helminthes, protozoa, and viruses. To assess how effectively host phylogenetic relatedness, geographic range overlap, and ecological similarity predict parasite assemblage similarity of different taxa, we calculated J and CJ based on the parasite assemblages of each taxon between host pairs for each of the five main parasite taxa.

Host phylogenetic distance, geographic range overlap and ecological dissimilarity

We quantified host phylogenetic distance for each pair of hosts using unique measures for three different measures to capture three different aspects of evolutionary history. We extracted time since divergence from a dated species level phylogenetic supertree (chronogram) of existing carnivore species (Nyakatura and Bininda-Emonds 2012). Because host genetic difference is likely to cause fundamentally different associations with parasites, we extracted the genetic distance between host species from a molecular phylogenetic supertree reconstructed using DNA sequences published in GenBank. Additionally, to assess the effect of host divergence on parasite assemblage similarity (Poulin and Morand 2000), we counted the number of existing divergence events between pairs of host species included in the analyses using the carnivore supertree. We acknowledge that this is an underestimate of the actual number of

divergence events in the evolutionary history, particularly for host pairs that diverged a long time ago, as extinct species are not included in our phylogenetic data. All three phylogenetic variables were correlated with each other, but time since divergence and genetic distance were more highly correlated than either was with number of divergence events (Figure S1). Because genetic distance and time since divergence generally produced congruent results, we only report results of time since divergence in the main result text, as well as results of number of divergence events.

We quantified ecological trait similarity between host pairs in two ways. First, we used adult body mass as a single indicator of overall biological similarity. Adult body mass is correlated with a wide range of other ecological and physiological traits in plants and animals (Western 1979, Brown 1995), as well as parasite diversity in primates and carnivores (Nunn et al. 2003a, Lindenfors et al. 2007). Further, adult body mass is one of the best-studied ecological traits in mammals with excellent data availability (over 87% of carnivores). Using data of adult body mass (g) from the species-level mammalian trait database PanTHERIA (Jones et al. 2009), we calculated the absolute difference in adult body mass for host pairs as a surrogate of their overall ecological similarity. Second, we used a composite measure that included adult body mass and 8 additional biological traits that have been quantified for at least 90% of our host species from the PanTHERIA database (Jones et al. 2009). Additional traits included adult head body length (mm), gestation length (days), inter birth interval (days), litter size (number of offspring), maximum longevity (months), sexual maturity age (days), trophic level (1 = herbivore, 2 = omnivore, and 3 = carnivore), and weaning age (days). For any trait of a species with missing data, we used the genus average. We used the resulting log-transformed data to

construct a Euclidean distance matrix of pair-wise host ecological distances. The body mass difference and overall ecological dissimilarity of carnivore host species pairs were significantly correlated with each other (Figure S2).

Finally, we obtained host geographic range maps from the IUCN Global Mammal Assessment (Schipper et al. 2008), and matched the data up to the taxonomy used in the phylogenetic trees (Wilson and Reeder 2005). Using these data, we calculated three continuous variables that describe host geographic range. First, we calculated the areas of all the host species ranges and all the pair-wise range intersections. Next, we calculated the proportional overlap area as the ratio of the intersection area to the area of the union of the two ranges, which is analogous to the Jaccard index of parasite assemblages. To test the association with the corrected Jaccard index for parasite assemblages, we divided the proportional geographic overlap by the maximum expected value (the ratio of the small range to the large range) to generate a variable we call corrected proportional geographic overlap. Additionally, we included a binary variable indicating whether the two hosts' ranges overlap or not. All the geographic data were processed in ArcGIS™ 10. All three continuous variables of geographic range overlap are highly correlated (Figure S3) and produced congruent results; thus, we only report results for geographic range overlap area in the main result text.

Statistical analyses

We first conducted non-parametric correlation tests (Spearman's rank-order test) between parasite assemblage similarity (*J* and *CJ*) and each of the seven host variables: three measures of phylogenetic relatedness, three measures of geographic range overlap,

and two measures of ecological similarity. Next, we constructed a generalized linear model (GLM) to predict parasite assemblage similarity focusing on host variables significant in pair-wise correlation tests. Our full model includes all host variables; we selected the final model through model simplification based on a low relative AIC score. Because one of our goals was to compare the patterns of parasite assemblage similarity in carnivores with those previously identified in primates (Davies and Pedersen 2008), if either (or both) of the range overlap variables was important, we re-analyzed the data of only host pairs with overlapped geographic ranges. We also analyzed the relationship between host variables (phylogenetic relatedness, geographic range overlap and ecological similarity) and parasite assemblage similarity of each of the five major parasite taxonomic groups (i.e. arthropods, bacteria, helminthes, protozoa, and viruses).

Finally, because initial results showed that phylogenetic relatedness was an important factor for predicting parasite sharing, we used a different approach to examine phylogenetic constraint on host range. Specifically, we examined the distribution of individual parasite species across the host phylogenetic tree, and calculated a measure of phylogenetic species variability (PSV; Helmus et al. 2007a, Helmus et al. 2007b) of all the hosts infected by the parasite species. PSV is a measure of the overall phylogenetic relatedness of species in an assemblage, and it is independent of the number of species in the assemblage (Helmus et al. 2007a). Based on the same phylogenetic tree, species in an assemblage with a low PSV value are more related to each other than species with a high PSV, regardless the number of species in the assemblage. To test whether the host range of a parasite that infects multiple host species is constrained by host phylogenetic relatedness, we simulated random associations with host species and estimate the lower

5% quantile of the PSV calculated using randomly selected hosts. If a parasite species has a host PSV below the 5% quantile of random simulation with the same number of host species, the parasite's host range is considered highly constrained by host phylogeny. All the analyses were conducted in R 2.12.2 (R Development Core Team 2011) with packages ape (Paradis et al. 2004) and picante (Kembel et al. 2010) for phylogenetic analyses, and MASS (Enhances et al. 2011) for other statistical analyses.

Results

We included 64 carnivore host species that were reported to be infected with at least 5 parasite species in the analyses to reduce the influence of sampling effort. These gave rise to 2016 pairwise host combinations, among which 1522 unique host pairs shared at least one parasite species. A total of 793 parasite species were reported in these host species. The Jaccard index (*J*) and corrected Jaccard index (*CJ*) calculated from pairwise host combination were correlated (Spearman's $\rho = 0.530$, $p < 0.001$), but their values differ from each other for most (74.6%) host pairs (Figure S4). In general, *J* and *CJ* showed similar associations with all the factors we assessed, so in the remainder of the paper, we present results for *CJ*, as we expect *CJ* to be more independent of sampling effort for parasitism in host species than *J*. Between carnivore species, parasite assemblage *CJ* ranges from 0 to 1, with a mean about 0.148 and median about 0.093.

All of the individual host variables were significantly correlated with the overall *CJ* for all parasite species, with the strongest correlate being the number of divergence events ($\rho = -0.347$, $p < 0.001$) (Figure 1, Table S1). Overall ecological similarity was the second strongest correlate ($\rho = -0.312$, $p < 0.001$). When we assessed the associations

within major taxonomic groups of parasites, we found different results across parasite groups (Figure 1, Table S2). Specifically, the strongest correlations existed between helminth *CJ* and number of divergence events between host pairs ($\rho = -441$, $p < 0.001$) as well as time since divergence ($\rho = -0.322$, $p < 0.001$). Strong correlations also existed between virus *CJ* and host ecological dissimilarity ($\rho = -326$, $p < 0.001$), as well as time since host divergence ($\rho = -292$, $p < 0.001$).

We included all the significant correlates in the full GLM models and conducted model simplification based on AIC scores to search for the best model for predicting parasite assemblage *CJ* between host species. Number of divergence events, overall ecological similarity and geographic overlap area were the three essential factors for predicting overall parasite assemblages *CJ*, and the predictive power was stronger when only hosts with overlapped geographic ranges were considered (Figure 2, Table 1). When we repeated the analyses for assemblage *CJ* of different parasite groups, we again found different factors were important in the final models, but in general, models for predicting parasite assemblage *CJ* in host pairs with overlapping geographic ranges perform better than those for *CJ* in all host pairs (Table S2).

Finally, because host phylogenetic distance was the strongest predictor for overall parasite assemblage similarity and for assemblage similarity of four out of the five parasite groups, we calculated a measure of host phylogenetic relatedness (host PSV) for each of the parasite species that have been reported to infect at least five carnivore host species. Comparison of average PSV across all parasites showed that there is no relationship between the number of host species and host PSV (Figure S5). Among all the parasite species infecting at least five host species, approximately 50-60% infected host

species with PSV values equal to or lower than the 5% quantile of the PSV calculated from randomly selected host species based on the three phylogenetic trees (chronogram: 59.2% and cladogram: 49.6%; Figure S5). Among the five major groups of parasites, helminthes and viruses appeared to be more constrained among host species that were phylogenetically related than parasites in other groups (Figure 3, Figure S6).

Discussion

Our broad scale comparative analyses showed that host phylogenetic relatedness is the most important factor for predicting parasite assemblage similarity, in that closely related carnivore species tend to share higher proportions of their parasite species than distantly related host species on the carnivore phylogeny. Closely related host species might share more co-inherent parasite species from their common ancestor than distantly related hosts because of a process termed as "missing the boat", through which descendant host species or lineages only inherent part of the ancestor host's parasite assemblage (Page 2003). Compared with distantly related hosts, closely related hosts also tend to show higher genetic and ecological similarity (Harvey and Pagel 1991, Harvey 1996, Freckleton et al. 2002) leading to more similar probabilities of acquiring the same parasite species and higher chance of cross-species transmission for their parasites (Nunn et al. 2000, 2003b, Page 2003, Streicker et al. 2010). Among the different measures of phylogenetic relatedness that we considered, the number of divergence events that occurred after two host species diverged is the strongest correlate of overall parasite assemblage similarity. The more divergence events two host species have gone through,

the lower similarity their parasite assemblages show. This result is consistent with the "missing the boat" hypothesis (Page 2003).

Our final model for predicting overall parasite assemblage similarity includes not only the number of divergent events, but also overall ecological similarity and the area of overlapping geographic ranges. Carnivore host ecological similarity is the second strongest correlate of overall parasite assemblage similarity, in that ecological similar carnivore species tend to share a large proportion of their parasite species. Carnivores could be dramatically different from each other in their ecological traits such as behavior, life pace and metabolism (Gittleman 1985, 1989, 1993), and thus ecologically dissimilar carnivore hosts are likely to encounter and acquire different parasite species. Their overall ecological dissimilarity might also reduce the chance of cross-species transmission of parasites. Host geographic range overlap is also positively correlated with parasite assemblage similarity. This is likely because extensive overlap of two hosts species' geographic ranges provides many opportunities for cross-species transmission, which often requires co-existing of the host species in the same region (Streicker et al. 2010). The predicting power of the final model increases when we constrained the analysis on carnivore species pairs with overlapped geographic ranges only.

The general association of parasite sharing with host phylogenetic relatedness and geographic range overlap we identified in carnivores are congruent with findings in other groups of mammals from the few previous studies on this topic. Host phylogenetic relatedness and geographic range overlap have been found to be two important predictors for parasite assemblage similarity between primate species (Davies and Pedersen 2008), for lentivirus host switching in primates (Charleston and Robertson 2002) and for cross-

species transmission of rabies between bat species (Streicker et al. 2010). Additionally, many of generalist parasites in primates have taxonomic constrained host ranges (Pedersen et al. 2005). We note that the correlations we found in carnivores are not as strong as those found in primates (Davies and Pedersen 2008), possibly due to the difference between these two groups of animals. For example, primates are mostly concentrated in low latitude regions, while carnivores are widely distributed throughout low to high latitudes, experiencing a large variety of habitats and climatic conditions (Sunquist 2002, Schipper et al. 2008). Patterns of parasite occurrence in carnivores are likely to be affected by environmental conditions. Geographic distance has also been suggested a strong predictor for parasite assemblage similarity in fish (Poulin 2003, Timi et al. 2010) and marine invertebrate (Thieltges et al. 2009). Therefore, despite all the biological difference and evolutionary distance, it is reasonable to expect that these patterns can be generalized to most animal hosts. However, the association between host ecological similarity and parasite assemblage similarity has not been studied in many systems (but see Streicker et al. 2010). We suggest future work include host ecological similarity in consideration, as well as expand on the taxonomic scale to explore the patterns of parasite sharing between host species from different orders or classes.

We further investigated the sharing patterns of major parasite taxa, including the arthropods, bacteria, helminths, protozoa and viruses and found that patterns vary among different parasite taxa. In particular, helminth assemblage similarity is most correlated with host phylogenetic relatedness measured in the number of existing divergence events that occurred after the two host species diverged, while virus assemblage similarity is most correlated with overall ecological similarity. Similarities in other groups of parasites

are less correlated with these host variables than those for helminthes and viruses, and in particular, protozoa assemblage similarity is only weakly correlated with phylogenetic relatedness. When we looked at this in a different way, by asking for each parasite species, how closely related are the hosts it infects, we found that for helminthes and viruses, their hosts tend to be more closely related than expected by chance. In other words, the host ranges of helminthes and viruses are more constrained by host phylogeny relative to other parasite types, such as protozoa, which are widely spread across host phylogeny. It is important to note that all parasites included in our analyses infected multiple host species, and our analyses quantify host phylogenetic range based on *which* host species were infected, not how many.

The finding that helminth distribution is constrained by host phylogeny is consistent with previous findings that many helminth species tend to have taxonomically restricted host ranges (Poulin and Mouillot 2003, Pedersen et al. 2005, Rosas-Valdez and de León 2010). This is possibly because helminths interact with hosts through complex processes that are often associated with specific host behavioral and physiological traits (Anderson and May 1992, Nunn and Altizer 2006), and phylogenetically distantly related carnivore species tend to be dissimilar in behaviors and physiology (Gittleman 1985, Harvey and Pagel 1991, Bininda-Emonds and Gittleman 2000, Gittleman 2001b, Freckleton et al. 2002). Therefore, expanding the phylogenetic range of hosts might require dramatic changes in helminths' strategies for transmission and within-host persistence, and such changes are likely to be enough for the helminths to be classified as different species. As a result, most helminthes species are highly specialist for small host lineages.

Owing to their high mutation rate, short generation time, and potential for rapid evolution, viruses were previously reported to be less constrained by host phylogeny than other parasite types and are more likely to emerge in novel host species that are not closely related to existing host species (Cleaveland et al. 2001, Pedersen et al. 2005). However, in these studies, virus host ranges were not compared to random selections of the same total number of host species. In fact, some of the viruses in our database infect a wide range of carnivore species, such as the canine distemper virus (53 host species) and rabies virus (34 host species), but these two viruses and most of other viruses in our database have host ranges that are significantly more constrained by host phylogeny based on divergence time than random selections of host species. This is possibly because cross-species transmission is more likely between ecologically similar hosts than between dissimilar hosts, and ecological similarity is often seen between carnivore species that diverged in recent past (Gittleman 1985, Harvey and Pagel 1991, Bininda-Emonds and Gittleman 2000, Gittleman 2001b, Freckleton et al. 2002). Additionally, because viruses, especially RNA viruses, are fast-evolving parasites (Nunn and Altizer 2006), viral lineages that infect a wide variety of host species will often accumulate sufficient sequence diversity through mutations to be considered multiple species.

Our study emphasizes the importance of host evolutionary and ecological information for predicting parasite occurrence patterns, particularly parasite occurrence in understudied host species and future parasite emergence in novel host species. To our knowledge, about 44% carnivore species have not been sampled for parasitism by 2010 and are not included in our database. Similarly, despite the availability of global data sets of parasite infection in primates and hoofed mammals, parasite data for large proportions

of these species have not been available (Nunn et al. 2003a, Ezenwa et al. 2006). Current knowledge of mammal phylogeny, ecology and distribution could be used for inferring the parasite assemblage composition in these understudied host species. It is highly likely that these understudied species share parasites with their phylogenetically close relatives that are ecologically similar to them, and co-existing in the same regions increases this likelihood. In other words, with information on a host species' position on the phylogeny, its geographic range, and some ecological traits, the coefficients of models, such as those presented here could predict the probability of particular parasite species to be present in understudied hosts.

The severe impacts of recent emerging infectious diseases on human, domestic animals and wildlife have raised the critical issue of predicting future disease risk for effective prevention (Cleaveland et al. 2001, Cleaveland et al. 2002, Jones et al. 2008, Fuller et al. 2012). Nowhere more apparent, carnivore species have been severely impacted by disease outbreaks, including diseases shared by domestic cats and dogs such as canine distemper virus (Alexander and Appel 1994, Roelke-Parker et al. 1996, Laurenson et al. 1998), rabies (Smith et al. 1993, Kat et al. 1995, Sillero-Zubiri et al. 1996, Laurenson et al. 1998, Butler et al. 2004, Randall et al. 2004), parvavirus (Laurenson et al. 1998) and heartworms (Pappas and Lunzman 1985). Global databases of host phylogeny, ecology and geographic distribution could be powerful tools for identifying future emergent pathogen by inferring a candidate pool of parasites based on the phylogenetic distance, ecological similarity and geographic range overlap between current and potential hosts (Davies and Pedersen 2008, Pedersen and Davies 2009). For example, we might expect carnivores that are phylogenetically related and ecologically

similar to domesticated dogs to be vulnerable to pathogen spillover from high-density populations of the reservoir host in the same region.

In summary, our results suggest that host phylogenetic relatedness is the most important predictor for parasite sharing patterns, and strongly indicate that host evolutionary history, represented by phylogeny, underlies the distribution of parasites among animal hosts. Specifically, host species could inherent the same parasite species from their common ancestor but might lose these parasites through later evolution (Page 2003). Furthermore, the evolutionary distance between host species affects their genetic and functional difference (Harvey and Pagel 1991, Harvey 1996, Freckleton et al. 2002) and thus affects the likelihood of cross-species transmission (Charleston and Robertson 2002, Streicker et al. 2010). Additionally, host ecological similarity and geographic range overlap provide additional predictive power for parasite sharing. Assemblage similarities of helminths and viruses are more predictable using our host variables than similarity of other parasite groups. Based on the weak association between parasite sharing pattern in arthropods, bacteria and protozoa found in our study, we suggest future investigation focus on other factors such as the environmental conditions within host ranges.

Acknowledgement

We thank Patrick Lindenfors and Charles Nunn for initial access to the database and all the undergraduate students helping with developing the global carnivore parasite database: Michael Cherry, Karen Christ, Matt Foretich, Ashton Griffin, Steven Howell, Tyler Halle, Cecilia Nix, and Jeff Shapiro. We thank Byron Ledbetter and Brian Perkins for technical assistance. We also thank the Altizer lab group and Ezenwa lab group at Odum School of Ecology for comments. This study is supported by Conservation

International, Odum School of Ecology, and National Science Foundation (Grant #0642619).

Table 1 Generalized linear models for predicting parasite assemblage *CJ* (log transformed) between all carnivore host pairs and pairs with overlapped geographic ranges. All the coefficients have $p < 0.001$ in both models. Pseudo-R2 is calculated as 1-(residual deviance / null deviance). To avoid zeros in the data, we added 1 to all the variables before log transformation.

	All hosts		***Overlapped hosts***	
Variables	**Coefficient (*p*)**	**Pseudo-R^2**	**Coefficient (*p*)**	**Pseudo-R^2**
Number of divergence events	-0.006 ± 0.0005 ($p < 0.001$)	0.195	-0.005 ± 0.001 ($p < 0.001$)	0.278
log Ecological similarity	0.078 ± 0.007 ($p < 0.001$)		0.086 ± 0.015 ($p < 0.001$)	
log Geographic overlap	0.002 ± 0.0002 ($p < 0.001$)		0.005 ± 0.001 ($p < 0.001$)	

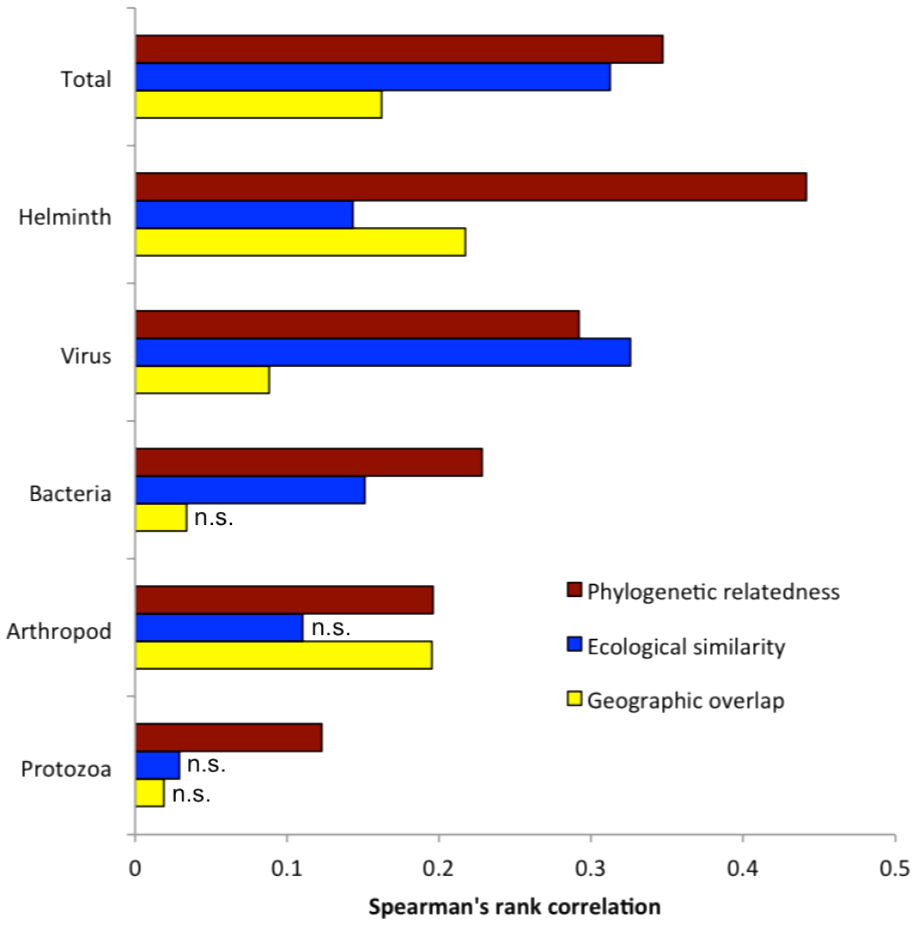


Figure 1 Strength of Spearman's rank correlations between parasite assemblage similarity (CJ) and pair-wise host factors from three categories. For each factor category, only the highest correlation is shown. Statistically insignificant correlations are indicated (n.s.). See Table S1 and S2 for more details.

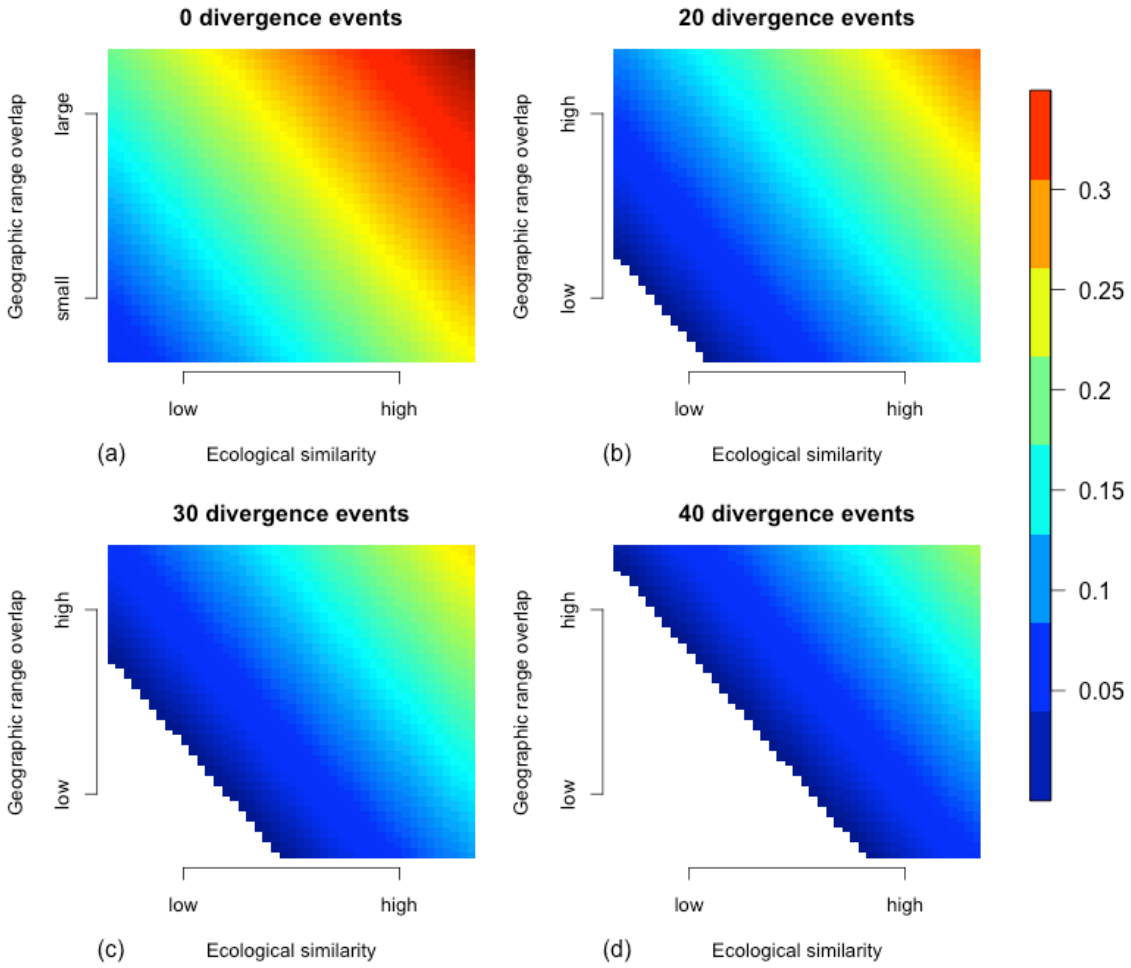


Figure 2 The parasite assemblage similarity (from low to high in blue to red) predicted by the final GLM model for pairs of host species with overlapped geographic ranges (see Table 1 for specific parameters). Where there is no colour, CJ is predicted to be 0 or negative.

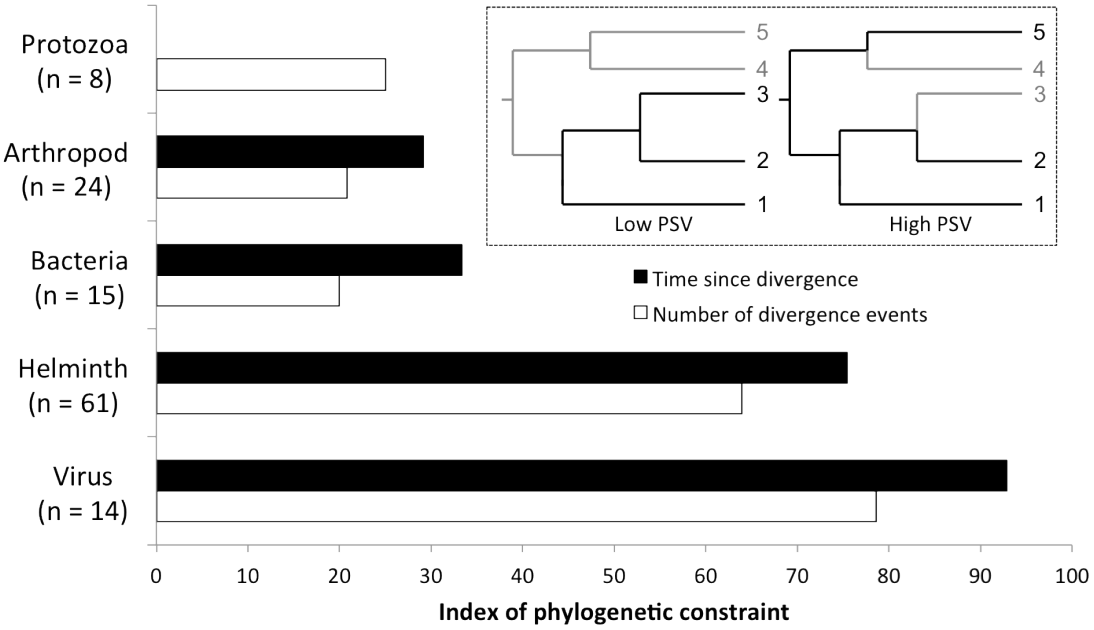


Figure 3 Degree to which parasites from major groups are constrained by host phylogeny, expressed as the percentage of parasite species in each major group that have observed host phylogenetic species variability (PSV) equal to or lower than the lower 5% quantile of the PSV values calculated from randomly selected host species. As illustrated in the conceptual graph in the upper right corner, host species showing a low PSV are more constrained by phylogeny than host species with a high PSV value. PSV was calculated based on two different host phylogenetic trees: a chronogram of host evolutionary history (black), and a cladogram of host evolutionary history (white). Note that because none of the protozoan species is significantly constrained by host phylogenetic relatedness measured in time since divergence, there is no black bar for protozoa.

CHAPTER 6

CONCLUDING REMARKS AND FUTURE DIRECTIONS

Throughout my PhD research, I aim to explain broad scale patterns of biodiversity. My research has used terrestrial mammals as a model system to investigate (1) the implication of mammalian phylogenetic diversity (PD) distribution for global mammal diversity conservation, and (2) the power of host phylogeny in predicting the risk of natural parasite infection in free-ranging mammals (particularly in the order of Carnivora). Overall, my findings emphasize that knowledge of mammal evolutionary history is highly valuable for explaining the global distribution of mammal biodiversity and estimating potential diversity loss if we fail to conserve species currently recognized as at high risk of extinction. Furthermore, the results from my studies of natural parasites in mammals have suggested that host evolutionary history is closely linked to the diversity and distribution patterns of parasites among host species, and therefore is an important predictor of parasite occurrence in wildlife populations.

Results presented in Chapter 2 (Huang et al. 2012b) confirm that PD is a valuable measure of biodiversity, because it is informative of multiple dimensions of biodiversity, including the number of species and the variation in these species' biological traits. This finding is particularly of interest for conservation practice. The diversity of species biological traits is an essential element of ecosystem stability, and therefore, is a dimension of biodiversity of high conservation value (McNeely et al. 1990, Crozier 1997). However, even in mammals, one of the most-studied and best-described groups,

the biological traits of a vast majority of them are under-described (Jones et al. 2009). Knowing that PD reliably indicates species variation in several biological traits, we can use PD as a conservation currency for quantifying biodiversity and determining the priority of conserving different mammal clades and geographic regions (Forest et al. 2007, Davies et al. 2008, Safi et al. 2011).

Using the evolutionary history represented by PD as a measure of biodiversity, I presented in chapter 3 the regional variation in potential biodiversity loss scenarios across the globe (Huang et al. 2012a). In particular, I found that some severe impacts of species extinction on biodiversity in specific regions are only visible in region specific assessment. Moreover, my comparison of the biodiversity loss estimations based on PD versus species richness conforms that species evolutionary history provides important insights of current status of biodiversity. Despite the intensive effort on studying and conserving mammals, approximately one out of four mammal species on earth is at high risk of extinction (Schipper et al. 2008), and in some geographic regions, high extinction risk is distributed among species in such ways that potential PD loss would be significantly higher than expected based on the number of threatened species. Considering the high rate of losing biodiversity per species, conservation effort is urgently needed in such regions.

According to recent assessment of mammal conservation status (Schipper et al. 2008), infectious disease has been recognized as one of the major threats driving many populations of threatened mammals close to extinction. Yet, our understanding of broad-scale parasite occurrence is still very limited. Through my work presented in chapter 4, I identified an important predictor of the diversity of parasites infecting a host species – the

evolutionary distinctiveness of the host species. My finding of evolutionary distinctive host species tend to host fewer parasites agrees with previous findings of parasite diversity in primates (Nunn et al. 2004), and results from my further investigation strongly suggested that parasite diversity is associated with host species diversification. Although the causal relationship between the two is not possible to determine with my data, it is reasonable to speculate that large diversity of parasite might have driven host speciation and thus promoted host diversity (Page 2003). It has also been suggested that parasites compose a large portion of biodiversity, and possibly have important conservation value (Durden and Keirans 1996, Poulin and Morand 2000, Koh et al. 2004). Nevertheless, according to my assessment on the records of parasites infecting free-ranging carnivores, one of the best-studied mammal host orders, current records of parasite diversity are still incomplete and some host species have been better-studied than others for parasite infection. My results supported that knowledge of host evolutionary history can be a powerful tool for predicting parasite diversity in understudied host species.

Host evolutionary history is also informative of the parasite assemblage similarity between two host species, as suggested by my results in chapter 5. This is consistent with previous findings in primates (Davies and Pedersen 2008). Furthermore, my simulation analyses suggest that parasites might not be distributed among host species randomly, but rather, even some widely spreading viruses are infecting host species that are more closely related in evolutionary history than randomly selected host species. Therefore, knowledge of evolutionary relationship among host species can not only help estimate the composition of current parasite assemblages in understudied host species, but also help

predict future risk of emerging infectious disease by inferring a candidate parasite pool (Pedersen and Davies 2009). Predicting future disease risk for effective prevention is critical, as emerging infectious diseases can have devastating impacts on wildlife, on domestic animals, and on human (Cleaveland et al. 2001, Cleaveland et al. 2002, Jones et al. 2008, Fuller et al. 2012).

In summary, my work presented in this dissertation has strongly support that knowledge of evolutionary history is essential for explaining and predicting biodiversity patterns, and therefore, has important values in informing biodiversity conservation practice. Identification of the strong association of species evolutionary history to other dimensions of biodiversity and parasite diversity patterns not only provides powerful tools for conservation practice, but also sheds light onto mechanisms shaping current biodiversity. For further understanding global biodiversity distribution, an important future direction is to explain the heterogeneity in biodiversity composition across space by quantifying another dimension of global biodiversity, the geographic variation of biodiversity in a phylogenetic framework (Devictor et al. 2010, Morlon et al. 2011). As my work has clearly indicated that evolutionary processes underlie current parasite occurrence patterns, future investigations comparing host phylogeny to parasite phylogeny should provide valuable insights of the evolutionary relationship between host and parasite species (Brooks 1979).

For my PhD dissertation, I took the approach of integrating computer programing with global-scale databases to explore the value of the knowledge of evolutionary history in explaining different aspects of global biodiversity. I am very grateful that recent computational developments are allowing synthesizing complicated broad scale

biodiversity patterns in multiple dimensions. Practically, compared with the traditional fieldwork approach, advanced analytic tools enabled by powerful modern computational techniques have made it possible to collate and analyze larger amount of data across broader spatial and temporal scales within a reasonable time frame for a PhD program. However, as pointed out above, current records of parasites in even the best-studied host groups are incomplete, and further effort on sampling wildlife populations, particularly populations of threatened species is much needed. For my PhD work on parasite diversity, I employed non-parametric methods to estimate true parasite diversity in a host species based on the frequency of the parasite species being reported and the number of reports on the host species. The actual accuracy of these estimates are impossible to assess without complete records of parasite diversity, but improvement of these methods could probably be made by considering the abundance and transmission modes of the parasites and distinguishing different types of parasite reports. It is also worth pointing out that the most recent phylogenetic supertree of mammals does not cover all the identified terrestrial mammal species in the world, and still includes a large proportion of artificial branches based on simplified macroevolutionary models (Fritz et al. 2009). Improving current evolutionary history data is undoubtedly a timely task. Perhaps more urgently, investigation is needed on how the quality of phylogenetic supertree might affect analyses of biodiversity patterns at the global scale and how we can control for potential bias (Nee 2006).

REFERENCES

Agrawal, A. A. 2000. Host-range evolution: adaptation and trade-offs in fitness of mites on alternative hosts. Ecology **81**:500-508.

Alexander, K. A. and M. J. Appel. 1994. African wild dogs (Lycaon pictus) endangered by a canine distemper epizootic among domestic dogs near the Masai Mara National Reserve, Kenya. J Wildl Dis **30**:481-485.

Altizer, S., D. Harvell, and E. Friedle. 2003. Rapid evolutionary dynamics and disease threats to biodiversity. Trends in Ecology & Evolution **18**:589-596.

Altizer, S., C. L. Nunn, and P. Lindenfors. 2007. Do threatened hosts have fewer parasites? A comparative study in primates. Journal of Animal Ecology **76**:304-314.

Altizer, S. and A. B. Pedersen. 2008. Host-parathogen evolution, biodiversity, and disease risks for natural populations. Pages 259-277 *in* S. P. C. C. W. Fox, editor. Conservation Biology: Evolution in Action. Oxford University Press Inc., New York.

Anderson, R. M. and R. M. May. 1992. Infectious diseases of humans: dynamics and control. Oxford University Press, USA.

Antonovics, J. 2009. The effect of sterilizing diseases on host abundance and distribution along environmental gradients. Proceedings of the Royal Society B: Biological Sciences **276**:1443-1448.

Antonovics, J., M. Hood, and J. Partain. 2002. The ecology and genetics of a host shift: Microbotryum as a model system. The American Naturalist **160**:S40-S53.

Appel, M. J. G. and B. A. Summers. 1995. Pathogenicity of morbilliviruses for terrestrial carnivores. Veterinary Microbiology **44**:187-191.

Bininda-Emonds, O. R. P., M. Cardillo, K. E. Jones, R. D. E. MacPhee, R. M. D. Beck, R. Grenyer, S. A. Price, R. A. Vos, J. L. Gittleman, and A. Purvis. 2007. The delayed rise of present-day mammals. Nature **446**:507-512.

Bininda-Emonds, O. R. P. and J. L. Gittleman. 2000. ARE PINNIPEDS FUNCTIONALLY DIFFERENT FROM FISSIPED CARNIVORES? THE IMPORTANCE OF PHYLOGENETIC COMPARATIVE ANALYSES. Evolution **54**:1011-1023.

Blakeslee, A. M. H. and J. E. Byers. 2008. USING PARASITES TO INFORM ECOLOGICAL HISTORY: COMPARISONS AMONG THREE CONGENERIC MARINE SNAILS. Ecology **89**:1068-1078.

Blomberg, S. and T. Garland Jr. 2002. Tempo and mode in evolution: phylogenetic inertia, adaptation and comparative methods. Journal of Evolutionary Biology **15**:899.

Blomberg, S. P., T. G. Jr., and A. R. Ives. 2003. Testing For Phylogenetic Signal In Comparative Data: Behavioral Traits Are More Labile. Evolution **57**:717-745.

Breiman, L. 2001. Random Forests. Machine Learning **45**:5-32.

Brito, D., R. G. Ambal, T. Brooks, N. D. Silva, M. Foster, W. Hao, C. Hilton-Taylor, A. Paglia, J. P. RodrÌguez, and J. V. RodrÌguez. 2010. How similar are national red lists and the IUCN Red List? Biological Conservation **143**:1154-1158.

Brockhurst, M. A., P. B. Rainey, and A. Buckling. 2004. The effect of spatial heterogeneity and parasites on the evolution of host diversity. Proceedings of the Royal Society of London. Series B: Biological Sciences **271**:107-111.

Brooks, D. R. 1979. Testing the Context and Extent of Host-Parasite Coevolution. Systematic Biology **28**:299-307.

Brooks, T. M., R. A. Mittermeier, G. A. B. da Fonseca, J. Gerlach, M. Hoffmann, J. F. Lamoreux, C. G. Mittermeier, J. D. Pilgrim, and A. S. L. Rodrigues. 2006. Global Biodiversity Conservation Priorities. Science **313**:58-61.

Brown, J. 1995. Macroecology. University of Chicago Press.

Buckling, A. and P. B. Rainey. 2002. The role of parasites in sympatric and allopatric host diversification. Nature **420**:496-499.

Butler, J. R. A., J. T. du Toit, and J. Bingham. 2004. Free-ranging domestic dogs (Canis familiaris) as predators and prey in rural Zimbabwe: threats of competition and disease to large wild carnivores. Biological Conservation **115**:369-378.

Cadotte, M. W., B. J. Cardinale, and T. H. Oakley. 2008. Evolutionary history and the effect of biodiversity on plant productivity. Proceedings of the National Academy of Sciences **105**:17012-17017.

Cadotte, M. W. and T. J. Davies. 2010. Rarest of the rare: advances in combining evolutionary distinctiveness and scarcity to inform conservation at biogeographical scales. Diversity and Distributions **16**:376-385.

Cardillo, M., G. M. Mace, J. L. Gittleman, K. E. Jones, J. Bielby, and A. Purvis. 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., G. M. Mace, K. E. Jones, J. Bielby, O. R. P. Bininda-Emonds, W. Sechrest, C. D. L. Orme, and A. Purvis. 2005. Multiple Causes of High Extinction Risk in Large Mammal Species. Science **309**:1239-1241.

Caro, T. M. and G. O'Doherty. 1999. On the Use of Surrogate Species in Conservation Biology

Uso de Especies Sustituatas en la Conservación Biológica. Conservation Biology **13**:805-814.

Chao, A., P. C. Li, S. Agatha, and W. Foissner. 2006. A statistical approach to estimate soil ciliate diversity and distribution based on data from five continents. Oikos **114**:479-493.

Charleston, M. A. and D. L. Robertson. 2002. Preferential Host Switching by Primate Lentiviruses Can Account for Phylogenetic Similarity with the Primate Phylogeny. Systematic Biology **51**:528-535.

Chiarucci, A., N. J. Enright, G. L. W. Perry, B. P. Miller, and B. B. Lamont. 2003. Performance of nonparametric species richness estimators in a high diversity plant community. Diversity and Distributions **9**:283-295.

Cleaveland, S., G. R. Hess, A. P. Dobson, M. K. Laurenson, H. I. McCallum, M. G. Roberts, and R. Woodroffe. 2002. The role of pathogens in biological conservation. The ecology of wildlife diseases:139?150.

Cleaveland, S., M. K. Laurenson, and L. H. Taylor. 2001. Diseases of humans and their domestic mammals: pathogen characteristics, host range and the risk of emergence. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences **356**:991-999.

Colwell, R. K. and J. A. Coddington. 1994. Estimating Terrestrial Biodiversity through Extrapolation. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences **345**:101-118.

Cornwell, W. K., D. W. Schwilk, and D. D. Ackerly. 2006. A trait-based test for habitat filtering: convex hull volume. Ecology **87**:1465-1471.

Crozier, R. H. 1997. Preserving The Information Content Of Species:Genetic Diversity, Phylogeny, and Conservation Worth. Annual Review of Ecology and Systematics **28**:243-268.

Davies, T. J., L. B. Buckley, R. Grenyer, and J. L. Gittleman. 2011. The influence of past and present climate on the biogeography of modern mammal diversity. Philosophical Transactions of the Royal Society B: Biological Sciences **366**:2526-2535.

Davies, T. J., S. A. Fritz, R. Grenyer, C. D. L. Orme, J. Bielby, O. R. P. Bininda-Emonds, M. Cardillo, K. E. Jones, J. L. Gittleman, G. M. Mace, and A. Purvis. 2008. Phylogenetic trees and the future of mammalian biodiversity. Proceedings of the National Academy of Sciences **105**:11556-11563.

Davies, T. J. and A. B. Pedersen. 2008. Phylogeny and geography predict pathogen community similarity in wild primates and humans. Proceedings of the Royal Society B: Biological Sciences.

De Castro, F. and B. Bolker. 2005. Mechanisms of disease-induced extinction. Ecology Letters **8**:117-126.

De Vienne, D. M., M. E. Hood, and T. Giraud. 2009. Phylogenetic determinants of potential host shifts in fungal pathogens. Journal of Evolutionary Biology **22**:2532-2541.

Devictor, V., D. Mouillot, C. Meynard, F. Jiguet, W. Thuiller, and N. Mouquet. 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.

Díaz, S. and M. Cabido. 2001. Vive la différence: plant functional diversity matters to ecosystem processes. Trends in Ecology & Evolution **16**:646-655.

Drexler, J. F., V. M. Corman, M. A. Muller, G. D. Maganga, P. Vallo, T. Binger, F. Gloza-Rausch, A. Rasche, S. Yordanov, A. Seebens, S. Oppong, Y. A. Sarkodie, C. Pongombo, A. N. Lukashev, J. Schmidt-Chanasit, A. Stocker, A. J. B. Carneiro, S. Erbar, A. Maisner, F. Fronhoffs, R. Buettner, E. K. V. Kalko, T. Kruppa, C. R. Franke, R. Kallies, E. R. N. Yandoko, G. Herrler, C. Reusken, A. Hassanin, D. H. Kruger, S. Matthee, R. G. Ulrich, E. M. Leroy, and C. Drosten. 2012. Bats host major mammalian paramyxoviruses. Nat Commun **3**:796.

Dunn, R. R., N. C. Harris, R. K. Colwell, L. P. Koh, and N. S. Sodhi. 2009. The sixth mass coextinction: are most endangered species parasites and mutualists? Proceedings of the Royal Society B: Biological Sciences **276**:3037-3045.

Durden, L. A. and J. E. Keirans. 1996. HostParasite Coextinction and the Plight of Tick Conservation. American Entomologist **42**:87-91.

Dybdahl, M. F. and C. M. Lively. 1998. Host-Parasite Coevolution: Evidence for Rare Advantage and Time-Lagged Selection in a Natural Population. Evolution **52**:1057-1066.
Enhances, A., G. P. A. FrF, and M. B. Ripley. 2011. Package , MASS.
Esch, G. W., A. O. Bush, and J. M. Aho. 1990. Parasite communities: Patterns and processes.
Ezenwa, V. O., S. A. Price, S. Altizer, N. D. Vitone, and K. C. Cook. 2006. Host traits and parasite species richness in even and odd-toed hoofed mammals, Artiodactyla and Perissodactyla. Oikos **115**:526-536.
Faith, D. P. 1992. Conservation evaluation and phylogenetic diversity. Biological Conservation **61**:1-10.
Faith, D. P. 2002. Quantifying Biodiversity: a Phylogenetic Perspective. Conservation Biology **16**:248-252.
Fenton, A. and A. B. Pedersen. 2005. Community epidemiology framework for classifying disease threats. Emerging Infectious Diseases **11**:1815-1821.
Flynn, D. F. B., N. Mirotchnick, M. Jain, M. I. Palmer, and S. Naeem. 2011. Functional and phylogenetic diversity as predictors of biodiversity‚Äìecosystem-function relationships. Ecology **92**:1573-1581.
Forest, F., R. Grenyer, M. Rouget, T. J. Davies, R. M. Cowling, D. P. Faith, A. Balmford, J. C. Manning, S. Proches, M. van der Bank, G. Reeves, T. A. J. Hedderson, and V. Savolainen. 2007. Preserving the evolutionary potential of floras in biodiversity hotspots. Nature **445**:757-760.
Freckleton, R. P., P. H. Harvey, and M. Pagel. 2002. Phylogenetic Analysis and Comparative Data: A Test and Review of Evidence. American Naturalist **160**:712.
Freckleton, R. P. and W. Jetz. 2009. Space versus phylogeny: disentangling phylogenetic and spatial signals in comparative data. Proceedings of the Royal Society B: Biological Sciences **276**:21-30.
Fritz, S. A., O. R. P. Bininda-Emonds, and A. Purvis. 2009. Geographical variation in predictors of mammalian extinction risk: big is bad, but only in the tropics. Ecology Letters **12**:538-549.
Fritz, S. A. and A. Purvis. 2010a. 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.
Fritz, S. A. and A. Purvis. 2010b. Selectivity in Mammalian Extinction Risk and Threat Types: a New Measure of Phylogenetic Signal Strength in Binary Traits. Conservation Biology **24**:1042-1051.
Fuller, T., S. Bensch, I. Müller, J. Novembre, J. Pérez-Tris, R. Ricklefs, T. Smith, and J. Waldenström. 2012. The Ecology of Emerging Infectious Diseases in Migratory Birds: An Assessment of the Role of Climate Change and Priorities for Future Research. EcoHealth **9**:80-88.
Gaston, K. J. 1996. Species-range-size distributions: patterns, mechanisms and implications. Trends in Ecology & Evolution **11**:197-201.
Gaston, K. J. 2000. Global patterns in biodiversity. Nature **405**:220-227.
Gelman, A. and J. Hill. 2007. Data analysis using regression and multilevel/hierarchical models. Cambridge University Press New York.

Gibbons, J. W. and J. E. Lovich. 1990. Sexual dimorphism in turtles with emphasis on the slider turtle (Trachemys scripta). Herpetological monographs **4**:1-29.
Gihring, T. M., S. J. Green, and C. W. Schadt. 2012. Massively parallel rRNA gene sequencing exacerbates the potential for biased community diversity comparisons due to variable library sizes. Environmental Microbiology **14**:285-290.
Gilbert, G. S. and C. O. Webb. 2007. Phylogenetic signal in plant pathogen–host range. Proceedings of the National Academy of Sciences **104**:4979-4983.
Gittleman, J. L. 1985. Carnivore Body Size: Ecological and Taxonomic Correlates. Oecologia **67**:540-554.
Gittleman, J. L. 1989. Carnivore group living: comparative trends. Pages 183-207 *in* J. L. Gittleman, editor. Carnivore behavior, ecology, and evolution. Cornell University Press, Ithaca, NY.
Gittleman, J. L. 1993. Carnivore life histories: a re-analysis in the light of new models. Pages 65-86 *in* N. Dunstone and M. L. Gorman, editors. Mammals as predators. Oxford University Press, Oxford.
Gittleman, J. L. 2001a. Carnivore conservation. Cambridge Univ Pr.
Gittleman, J. L. 2001b. Comparative Methods in the Study of Behavior. Pages 337-350 Encyclopedia of the life sciences: Biology of Behavior. Academic Press, New York.
Grenyer, R., C. D. Orme, S. F. Jackson, G. H. Thomas, R. G. Davies, T. J. Davies, K. E. Jones, V. A. Olson, R. S. Ridgely, and P. C. Rasmussen. 2006. Global distribution and conservation of rare and threatened vertebrates. Nature **444**:93-96.
Guernier, V., M. E. Hochberg, and J.-F. Guégan. 2004. Ecology Drives the Worldwide Distribution of Human Diseases. PLoS Biol **2**:e141.
Harris, N. C. and R. R. Dunn. 2010. Using host associations to predict spatial patterns in the species richness of the parasites of North American carnivores. Ecology Letters **13**:1411-1418.
Harvey, P. and M. Pagel. 1991. The comparative method in evolutionary biology. Oxford University Press Oxford.
Harvey, P. H. 1996. Phylogenies for Ecologists. Journal of Animal Ecology **65**:255-263.
Helmus, M. R., T. J. Bland, C. K. Williams, and A. R. Ives. 2007a. Phylogenetic Measures of Biodiversity. The American Naturalist **169**:E68-E83.
Helmus, M. R., K. Savage, M. W. Diebel, J. T. Maxted, and A. R. Ives. 2007b. Separating the determinants of phylogenetic community structure. Ecology Letters **10**:917-925.
Hoberg, E., D. Brooks, D. Siegel-Causey, D. Clayton, and J. Moore. 1997. Host-parasite co-speciation: history, principles, and prospects. Host-parasite evolution: general principles and avian models.:212-235.
Holt, R. D., A. P. Dobson, M. Begon, R. G. Bowers, and E. M. Schauber. 2003. Parasite establishment in host communities. Ecology Letters **6**:837-842.
Hortal, J., P. A. V. Borges, and C. Gaspar. 2006. Evaluating the performance of species richness estimators: sensitivity to sample grain size. Journal of Animal Ecology **75**:274-287.
Huang, S., T. J. Davies, and J. L. Gittleman. 2012a. How global extinctions impact regional biodiversity in mammals. Biology Letters.

Huang, S., P. R. Stephens, and J. L. Gittleman. 2012b. Traits, trees and taxa: global dimensions of biodiversity in mammals. Proceedings of the Royal Society B: Biological Sciences **279**:4997-5003.

Hudson, P. J., A. Rizzoli, B. T. Grenfell, H. Heesterbeek, and A. P. Dobson. 2002. The ecology of wildlife diseases. Oxford University Press Oxford, UK:.

Isaac, N. J. B., S. T. Turvey, B. Collen, C. Waterman, and J. E. M. Baillie. 2007. Mammals on the EDGE: Conservation Priorities Based on Threat and Phylogeny. PLoS ONE **2**.

Johnson, K. P. and D. H. Clayton. 2003. Coevolutionary history of ecological replicates: comparing phylogenies of wing and body lice to Columbiform hosts. Tangled Trees: phylogeny, cospeciation and coevolution:262-286.

Jones, K. E., J. Bielby, M. Cardillo, S. A. Fritz, J. O'Dell, C. D. L. Orme, K. Safi, W. Sechrest, E. H. Boakes, C. Carbone, C. Connolly, M. J. Cutts, J. K. Foster, R. Grenyer, M. Habib, C. A. Plaster, S. A. Price, E. A. Rigby, J. Rist, A. Teacher, O. R. P. Bininda-Emonds, J. L. Gittleman, G. M. Mace, A. Purvis, and W. K. Michener. 2009. PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. Ecology **90**:2648-2648.

Jones, K. E., N. G. Patel, M. A. Levy, A. Storeygard, D. Balk, J. L. Gittleman, and P. Daszak. 2008. Global trends in emerging infectious diseases. Nature **451**:990-993.

Karvonen, A. and O. Seehausen. 2012. The Role of Parasitism in Adaptive Radiations&# 8212; When Might Parasites Promote and When Might They Constrain Ecological Speciation? International Journal of Ecology **2012**.

Kat, P. W., K. A. Alexander, J. S. Smith, and L. Munson. 1995. Rabies and African Wild Dogs in Kenya. Proceedings: Biological Sciences **262**:229-233.

Keesing, F., L. K. Belden, P. Daszak, A. Dobson, C. D. Harvell, R. D. Holt, P. Hudson, A. Jolles, K. E. Jones, C. E. Mitchell, S. S. Myers, T. Bogich, and R. S. Ostfeld. 2010. Impacts of biodiversity on the emergence and transmission of infectious diseases. Nature **468**:647-652.

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

Kennedy, S., T. Kuiken, P. D. Jepson, R. Deaville, M. Forsyth, T. Barrett, M. W. van de Bildt, A. D. Osterhaus, T. Eybatov, C. Duck, A. Kydyrmanov, I. Mitrofanov, and S. Wilson. 2000. Mass die-Off of Caspian seals caused by canine distemper virus. Emerging Infectious Diseases **6**:637-639.

Koh, L. P., R. R. Dunn, N. S. Sodhi, R. K. Colwell, H. C. Proctor, and V. S. Smith. 2004. Species Coextinctions and the Biodiversity Crisis. Science **305**:1632-1634.

Koleff, P., K. J. Gaston, and J. J. Lennon. 2003. Measuring beta diversity for presence–absence data. Journal of Animal Ecology **72**:367-382.

Krasnov, B. R., G. I. Shenbrot, D. Mouillot, I. S. Khokhlova, and R. Poulin. 2005. Spatial variation in species diversity and composition of flea assemblages in small mammalian hosts: geographical distance or faunal similarity? Journal of Biogeography **32**:633-644.

Laurenson, K., C. Sillero-Zubiri, H. Thompson, F. Shiferaw, S. Thirgood, and J. Malcolm. 1998. Disease as a threat to endangered species: Ethiopian wolves, domestic dogs and canine pathogens. Animal Conservation **1**:273-280.

Lembo, T., K. Hampson, D. T. Haydon, M. Craft, A. Dobson, J. Dushoff, E. Ernest, R. Hoare, M. Kaare, T. Mlengeya, C. Mentzel, and S. Cleaveland. 2008. Exploring reservoir dynamics: a case study of rabies in the Serengeti ecosystem. Journal of Applied Ecology **45**:1246-1257.

Liaw, A. and M. Wiener. 2002. Classification and Regression by randomForest. R news **2**:18-22.

Lindenfors, P., C. L. Nunn, K. E. Jones, A. A. Cunningham, W. Sechrest, and J. L. Gittleman. 2007. Parasite species richness in carnivores: effects of host body mass, latitude, geographical range and population density. Global Ecology and Biogeography.

Litman, G. W., J. P. Cannon, and L. J. Dishaw. 2005. Reconstructing immune phylogeny: new perspectives. Nat Rev Immunol **5**:866-879.

Longdon, B., J. D. Hadfield, C. L. Webster, D. J. Obbard, and F. M. Jiggins. 2011. Host Phylogeny Determines Viral Persistence and Replication in Novel Hosts. PLoS Pathog **7**:e1002260.

Loreau, M., S. Naeem, P. Inchausti, J. Bengtsson, J. P. Grime, A. Hector, D. U. Hooper, M. A. Huston, D. Raffaelli, B. Schmid, D. Tilman, and D. A. Wardle. 2001. Biodiversity and Ecosystem Functioning: Current Knowledge and Future Challenges. Science **294**:804-808.

Lovich, J. and J. W. Gibbons. 1992. A review of techniques for quantifying sexual size dimorphism. Growth, development, and aging: GDA **56**:269.

Maan, M. E. and O. Seehausen. 2011. Ecology, sexual selection and speciation. Ecology Letters **14**:591-602.

MacArthur, R. H. 1957. On the relative abundance of bird species. Proceedings of the National Academy of Sciences of the United States of America **43**:293.

Mace, G. M., J. L. Gittleman, and A. Purvis. 2003. Preserving the Tree of Life. Science **300**:1707-1709.

Mace, G. M. and A. Purvis. 2008. Evolutionary biology and practical conservation: bridging a widening gap. Molecular Ecology **17**:9-19.

Maherali, H. and J. N. Klironomos. 2007. Influence of Phylogeny on Fungal Community Assembly and Ecosystem Functioning. Science **316**:1746-1748.

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

Marston, M. F., F. J. Pierciey, A. Shepard, G. Gearin, J. Qi, C. Yandava, S. C. Schuster, M. R. Henn, and J. B. H. Martiny. 2012. Rapid diversification of coevolving marine Synechococcus and a virus. Proceedings of the National Academy of Sciences **109**:4544-4549.

Mason, N. W. H., K. MacGillivray, J. B. Steel, and J. B. Wilson. 2003. An index of functional diversity. Journal of Vegetation Science **14**:571-578.

Mason, N. W. H., D. Mouillot, W. G. Lee, and J. B. Wilson. 2005. Functional richness, functional evenness and functional divergence: the primary components of functional diversity. Oikos **111**:112-118.

Mazerolle, M. J. and M. M. J. Mazerolle. 2012. Package ‚ÄòAICcmodavg‚Äô.

McCain, C. M. 2009. Vertebrate range sizes indicate that mountains may be 'higher' in the tropics. Ecology Letters **12**:550-560.

McNeely, J. A., K. R. Miller, W. V. Reid, R. A. Mittermeier, and T. B. Werner. 1990. Conserving the world's biological diversity. International Union for Conservation of Nature and Natural Resources, Washington, DC.

Mooers, A., D. Faith, and W. Maddison. 2008. Converting endangered species categories to probabilities of extinction for phylogenetic conservation prioritization. PLoS ONE **3**.

Morand, S. and R. Poulin. 1998. Density, body mass and parasite species richness of terrestrial mammals. Evolutionary Ecology **12**:717-727.

Morlon, H., D. W. Schwilk, J. A. Bryant, P. A. Marquet, A. G. Rebelo, C. Tauss, B. J. M. Bohannan, and J. L. Green. 2011. Spatial patterns of phylogenetic diversity. Ecology Letters **14**:141-149.

Mouchet, M., F. Guilhaumon, S. Villéger, N. W. H. Mason, J. A. Tomasini, and D. Mouillot. 2008. Towards a consensus for calculating dendrogram−based functional diversity indices. Oikos **117**:794-800.

Murphy, D., E. Gormley, E. Costello, D. O'Meara, and L. Corner. 2010. The prevalence and distribution of< i> Mycobacterium bovis</i> infection in European badgers (< i> Meles meles</i>) as determined by enhanced post mortem examination and bacteriological culture. Research in veterinary science **88**:1-5.

Murphy, W. J., E. Eizirik, W. E. Johnson, Y. P. Zhang, O. A. Ryder, and S. J. O'Brien. 2001. Molecular phylogenetics and the origins of placental mammals. Nature **409**:614-618.

Murray, D. L., C. A. Kapke, J. F. Evermann, and T. K. Fuller. 1999. Infectious disease and the conservation of free-ranging large carnivores. Animal Conservation **2**:241-254.

Myers, N., R. A. Mittermeier, C. G. Mittermeier, G. A. B. da Fonseca, and J. Kent. 2000. Biodiversity hotspots for conservation priorities. Nature **403**:853-858.

Nee, S. 2006. Birth-Death Models in Macroevolution. Annual Review of Ecology, Evolution, and Systematics **37**:1-17.

Nee, S. and R. M. May. 1997. Extinction and the Loss of Evolutionary History. Science **278**:692-694.

Nunn, C. L. 2011. The comparative approach in evolutionary anthropology and biology. University of Chicago Press.

Nunn, C. L. and S. Altizer. 2006. Infectious Diseases in Primates: Behavior, Ecology and Evolution.

Nunn, C. L., S. Altizer, K. E. Jones, and W. Sechrest. 2003a. Comparative Tests of Parasite Species Richness in Primates. The American Naturalist **162**:597-614.

Nunn, C. L., S. Altizer, W. Sechrest, K. E. Jones, R. A. Barton, and J. L. Gittleman. 2004. Parasites and the Evolutionary Diversification of Primate Clades. The American Naturalist **164**:90-103.

Nunn, C. L. and S. M. Altizer. 2005. The global mammal parasite database: An online resource for infectious disease records in wild primates. Evolutionary Anthropology: Issues, News, and Reviews **14**:1-2.

Nunn, C. L., S. M. Altizer, W. Sechrest, and A. A. Cunningham. 2005. Latitudinal Gradients of Parasite Species Richness in Primates. Diversity and Distributions **11**:249-256.

Nunn, C. L., J. L. Gittleman, and J. Antonovics. 2000. Promiscuity and the Primate Immune System. Science **290**:1168-1170.
Nunn, C. L., J. L. Gittleman, and J. Antonovics. 2003b. A Comparative Study of White Blood Cell Counts and Disease Risk in Carnivores. Proceedings: Biological Sciences **270**:347-356.
Nyakatura, K. and O. Bininda-Emonds. 2012. Updating the evolutionary history of Carnivora (Mammalia): a new species-level supertree complete with divergence time estimates. BMC Biology **10**:12.
Olson, D. M., E. Dinerstein, E. D. Wikramanayake, N. D. Burgess, G. V. N. Powell, E. C. Underwood, J. A. D'Amico, I. Itoua, H. E. Strand, and J. C. Morrison. 2001. Terrestrial Ecoregions of the World: A New Map of Life on Earth. BioScience **51**:933-938.
Olson, V. A., R. G. Davies, C. D. L. Orme, G. H. Thomas, S. Meiri, T. M. Blackburn, K. J. Gaston, I. P. F. Owens, and P. M. Bennett. 2009. Global biogeography and ecology of body size in birds. Ecology Letters **12**:249-259.
Orme, D., R. Freckleton, G. Thomas, T. Petzoldt, and S. Fritz. 2009. CAIC: Comparative Analyses using Independent Contrasts.
Page, R. D. M. 1994. PARALLEL PHYLOGENIES: RECONSTRUCTING THE HISTORY OF HOST-PARASITE ASSEMBLAGES. Cladistics **10**:155-173.
Page, R. D. M. 2003. Tangled trees: phylogeny, cospeciation, and coevolution. University of Chicago Press.
Pappas, L. G. and A. T. Lunzman. 1985. Canine heartworm in the domestic and wild canids of southeastern Nebraska. The Journal of Parasitology:828-830.
Paradis, E., J. Claude, and K. Strimmer. 2004. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics **20**:289-290.
Pavoine, S. and M. B. Bonsall. 2011. Measuring biodiversity to explain community assembly: a unified approach. Biological Reviews **86**:792-812.
Pedersen, A. and T. Davies. 2009. Cross-Species Pathogen Transmission and Disease Emergence in Primates. EcoHealth **6**:496-508.
Pedersen, A. B., S. Altizer, M. Poss, A. A. Cunningham, and C. L. Nunn. 2005. Patterns of host specificity and transmission among parasites of wild primates. International Journal for Parasitology **35**:647-657.
Pedersen, A. B. and A. Fenton. 2007. Emphasizing the ecology in parasite community ecology. Trends in Ecology & Evolution **22**:133-139.
Pedersen, A. B., K. E. Jones, C. L. Nunn, and S. Altizer. 2007. Infectious Diseases and Extinction Risk in Wild Mammals. Conservation Biology **21**:1269-1279.
Perlman, S. J. and J. Jaenike. 2003. EVOLUTION OF MULTIPLE COMPONENTS OF VIRULENCE IN DROSOPHILA-NEMATODE ASSOCIATIONS. Evolution **57**:1543-1551.
Pfennig, D. W. 2000. Effect of Predator-Prey Phylogenetic Similarity on the Fitness Consequences of Predation: A Trade-off between Nutrition and Disease? The American Naturalist **155**:335-345.
Polasky, S., B. Csuti, C. A. Vossler, Meyers, and S. Mark. 2001. A comparison of taxonomic distinctness versus richness as criteria for setting conservation priorities for North American birds. Biological Conservation **97**:99-105.

Poulin, R. 1992. Determinants of host-specificity in parasites of freshwater fishes. International Journal for Parasitology **22**:753-758.
Poulin, R. 1995. Phylogeny, Ecology, and the Richness of Parasite Communities in Vertebrates. Ecological Monographs **65**:283-302.
Poulin, R. 1997. Species Richness of Parasite Assemblages: Evolution and Patterns. Annual Review of Ecology and Systematics **28**:341-358.
Poulin, R. 1998. Comparison of Three Estimators of Species Richness in Parasite Component Communities. The Journal of Parasitology **84**:485-490.
Poulin, R. 2003. The decay of similarity with geographical distance in parasite communities of vertebrate hosts. Journal of Biogeography **30**:1609-1615.
Poulin, R. and S. Morand. 2000. The Diversity of Parasites. The Quarterly Review of Biology **75**:277-293.
Poulin, R. and D. Mouillot. 2003. Parasite specialization from a phylogenetic perspective: a new index of host specificity. Parasitology **126**:473-480.
Pugliese, A. 2010. Modelling Multi-Species Parasite Transmission
Modelling Parasite Transmission and Control. Pages 32-50 *in* E. Michael and R. C. Spear, editors. Springer New York.
Purvis, A., P. M. Agapow, J. L. Gittleman, and G. M. Mace. 2000. Nonrandom Extinction and the Loss of Evolutionary History. Science **288**:328.
Purvis, A. and A. Hector. 2000. Getting the measure of biodiversity. Nature **405**:212-219.
R Development Core Team. 2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
Randall, D. A., S. D. Williams, I. V. Kuzmin, C. E. Rupprecht, L. A. Tallents, Z. Tefera, K. Argaw, F. Shiferaw, D. L. Knobel, C. Sillero-Zubiri, and M. K. Laurenson. 2004. Rabies in endangered Ethiopian wolves. Emerging Infectious Diseases **10**:2214-2217.
Redding, D. W., C. V. DeWolff, and A. Ø. Mooers. 2010. Evolutionary Distinctiveness, Threat Status, and Ecological Oddity in Primates. Conservation Biology **24**:1052-1058.
Redding, D. W. and A. Ø. Mooers. 2006. Incorporating Evolutionary Measures into Conservation Prioritization. Conservation Biology **20**:1670-1678.
Reed, D. L., J. E. Light, J. M. Allen, and J. J. Kirchman. 2007. Pair of lice lost or parasites regained: the evolutionary history of anthropoid primate lice. BMC Biology **5**:7.
Ricklefs, R. E. and S. M. Fallon. 2002. Diversification and host switching in avian malaria parasites. Proceedings of the Royal Society of London. Series B: Biological Sciences **269**:885-892.
Rodrigues, A. S. L. and K. J. Gaston. 2002. Maximising phylogenetic diversity in the selection of networks of conservation areas. Biological Conservation **105**:103-111.
Rodrigues, A. S. L., R. Grenyer, J. E. M. Baillie, O. R. P. Bininda-Emonds, J. L. Gittlemann, M. Hoffmann, K. Safi, J. Schipper, S. N. Stuart, and T. Brooks. 2011. Complete, accurate, mammalian phylogenies aid conservation planning, but not much. Philosophical Transactions of the Royal Society B: Biological Sciences **366**:2652-2660.

Roelke-Parker, M. E., L. Munson, C. Packer, R. Kock, S. Cleaveland, M. Carpenter, S. J. O'Brien, A. Pospischil, R. Hofmann-Lehmann, H. Lutz, G. L. M. Mwamengele, M. N. Mgasa, G. A. Machange, B. A. Summers, and M. J. G. Appel. 1996. A canine distemper virus epidemic in Serengeti lions (Panthera leo). Nature **379**:441-445.

Rohde, K. 1999. Latitudinal Gradients in Species Diversity and Rapoport's Rule Revisited: A Review of Recent Work and What Can Parasites Teach Us about the Causes of the Gradients? Ecography **22**:593-613.

Rosas-Valdez, R. and G. P.-P. de León. 2010. Patterns of Host Specificity Among the Helminth Parasite Fauna of Freshwater Siluriforms: Testing the Biogeographical Core Parasite Fauna Hypothesis. Journal of Parasitology **97**:361-363.

Safi, K., M. V. Cianciaruso, R. D. Loyola, D. Brito, K. Armour-Marshall, and J. A. F. Diniz-Filho. 2011. Understanding global patterns of mammalian functional and phylogenetic diversity. Philosophical Transactions of the Royal Society B: Biological Sciences **366**:2536-2544.

Schipper, J. and J. S. Chanson and F. Chiozza and N. A. Cox and M. Hoffmann and V. Katariya and J. Lamoreux and A. S. L. Rodrigues and S. N. Stuart and H. J. Temple and J. Baillie and L. Boitani and T. E. Lacher, Jr. and R. A. Mittermeier and A. T. Smith and D. Absolon and J. M. Aguiar and G. Amori and N. Bakkour and R. Baldi and R. J. Berridge and J. Bielby and P. A. Black and J. J. Blanc and T. M. Brooks and J. A. Burton and T. M. Butynski and G. Catullo and R. Chapman and Z. Cokeliss and B. Collen and J. Conroy and J. G. Cooke and G. A. B. da Fonseca and A. E. Derocher and H. T. Dublin and J. W. Duckworth and L. Emmons and R. H. Emslie and M. Festa-Bianchet and M. Foster and S. Foster and D. L. Garshelis and C. Gates and M. Gimenez-Dixon and S. Gonzalez and J. F. Gonzalez-Maya and T. C. Good and G. Hammerson and P. S. Hammond and D. Happold and M. Happold and J. Hare and R. B. Harris and C. E. Hawkins and M. Haywood and L. R. Heaney and S. Hedges and K. M. Helgen and C. Hilton-Taylor and S. A. Hussain and N. Ishii and T. A. Jefferson and R. K. B. Jenkins and C. H. Johnston and M. Keith and J. Kingdon and D. H. Knox and K. M. Kovacs and P. Langhammer and K. Leus and R. Lewison and G. Lichtenstein and L. F. Lowry and Z. Macavoy and G. M. Mace and D. P. Mallon and M. Masi and M. W. McKnight and R. A. Medellin and P. Medici and G. Mills and P. D. Moehlman and S. Molur and A. Mora and K. Nowell and J. F. Oates and W. Olech and W. R. L. Oliver and M. Oprea and B. D. Patterson and W. F. Perrin and B. A. Polidoro and C. Pollock and A. Powel and Y. Protas and P. Racey and J. Ragle and P. Ramani and G. Rathbun and R. R. Reeves and S. B. Reilly and J. E. Reynolds, III and C. Rondinini and R. G. Rosell-Ambal and M. Rulli and A. B. Rylands and S. Savini and C. J. Schank and W. Sechrest and C. Self-Sullivan and A. Shoemaker and C. Sillero-Zubiri and N. De Silva and D. E. Smith and C. Srinivasulu and P. J. Stephenson and N. van Strien and B. K. Talukdar and B. L. Taylor and R. Timmins and D. G. Tirira and M. F. Tognelli and K. Tsytsulina and L. M. Veiga and J.-C. Vie and E. A. Williamson and S. A. Wyatt and Y. Xie and B. E. Young. 2008. The Status of the World's Land and Marine Mammals: Diversity, Threat, and Knowledge. Science **322**:225-230.

Schleuter, D., M. Daufresne, F. Massol, and C. Argillier. 2010. A user's guide to functional diversity indices. Ecological Monographs **80**:469-484.
Sechrest, W., T. M. Brooks, G. A. B. da Fonseca, W. R. Konstant, R. A. Mittermeier, A. Purvis, A. B. Rylands, and J. L. Gittleman. 2002. Hotspots and the conservation of evolutionary history. Proceedings of the National Academy of Sciences **99**:2067.
Sillero-Zubiri, C., A. King, and D. Macdonald. 1996. Rabies and mortality in Ethiopian wolves (Canis simensis). Journal of Wildlife Diseases **32**:80-86.
Smith, J. S., P. A. Yager, and L. A. Orciari. 1993. Rabies in wild and domestic carnivores of Africa: epidemiological and historical associations determined by limited sequence analysis. The Onderstepoort journal of veterinary research **60**:307-314.
Smith, K. F., D. F. Sax, and K. D. Lafferty. 2006. Evidence for the Role of Infectious Disease in Species Extinction and Endangerment
Evidencia del Papel de Enfermedades Infecciosas en la Extinción y Puesta en Peligro de Especies. Conservation Biology **20**:1349-1357.
Soutullo, A., S. Dodsworth, S. B. Heard, and A. Mooers. 2005. Distribution and correlates of carnivore phylogenetic diversity across the Americas. Animal Conservation **8**:249-258.
Streicker, D. G., A. S. Turmelle, M. J. Vonhof, I. V. Kuzmin, G. F. McCracken, and C. E. Rupprecht. 2010. Host Phylogeny Constrains Cross-Species Emergence and Establishment of Rabies Virus in Bats. Science **329**:676-679.
Sunquist, F. 2002. Wild cats of the world. University of Chicago Press.
Szymanski, M. M. and I. J. Lovette. 2005. HIGH LINEAGE DIVERSITY AND HOST SHARING OF MALARIAL PARASITES IN A LOCAL AVIAN ASSEMBLAGE. Journal of Parasitology **91**:768-774.
Thieltges, D., M. Ferguson, C. Jones, M. Krakau, X. de Montaudouin, L. Noble, K. Reise, and R. Poulin. 2009. Distance decay of similarity among parasite communities of three marine invertebrate hosts. Oecologia **160**:163-173.
Thompson, J. N. 1994. The coevolutionary process. University Of Chicago Press.
Thorne, E. T. and E. S. Williams. 1988. Disease and Endangered Species: The Black-Footed Ferret as a Recent Example. Conservation Biology **2**:66-74.
Timi, J. T., J. L. Luque, and R. Poulin. 2010. Host ontogeny and the temporal decay of similarity in parasite communities of marine fish. International Journal for Parasitology **40**:963-968.
Walther, B. A., P. Cotgreave, R. D. Price, R. D. Gregory, and D. H. Clayton. 1995. Sampling Effort and Parasite Species Richness. Parasitology Today **11**:306-310.
Walther, B. A. and J. L. Moore. 2005. The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. Ecography **28**:815-829.
Walther, B. A. and S. Morand. 1998. Comparative performance of species richness estimation methods. Parasitology **116**:395-405.
Weaver, S. C. and A. D. T. Barrett. 2004. Transmission cycles, host range, evolution and emergence of arboviral disease. Nat Rev Micro **2**:789-801.
West, G. B., J. H. Brown, and B. J. Enquist. 1997. A General Model for the Origin of Allometric Scaling Laws in Biology. Science **276**:122-126.

Western, D. 1979. Size, life history and ecology in mammals. African Journal of Ecology **17**:185-204.
Williams, P. H., N. D. Burgess, and C. Rahbek. 2000. Flagship species, ecological complementarity and conserving the diversity of mammals and birds in sub-Saharan Africa. Animal Conservation **3**:249-260.
Wilson, D. and D. Reeder. 2005. Mammal Species Of The World: A Taxonomic And Geographic Reference. Johns Hopkins University Press.
Worm, B., E. B. Barbier, N. Beaumont, J. E. Duffy, C. Folke, B. S. Halpern, J. B. C. Jackson, H. K. Lotze, F. Micheli, S. R. Palumbi, E. Sala, K. A. Selkoe, J. J. Stachowicz, and R. Watson. 2006. Impacts of Biodiversity Loss on Ocean Ecosystem Services. Science **314**:787-790.
Yoder, Jeremy B. and Scott L. Nuismer. 2010. When Does Coevolution Promote Diversification? The American Naturalist **176**:802-817.

APPENDIX A

SUPPORTING INFORMATION FOR CHAPTER 2

Additional Methods

Clade level analyses

The primary purpose of our clade level analyses was to determine whether PD or SR is a better predictor of TD, and to see if the strength of the correlation varies in a predictable way across traits. We analysed correlations using Spearman's rank-order test, and compared Spearman's ρ values to see which of SR and PD was more tightly correlated with TD. Because phylogenetic distance between two species is associated with their difference in traits that show strong phylogenetic signals [1, 2], we investigated whether the degree of phylogenetic signal in a trait determines the strength of the correlation between PD and trait variance. We also investigated whether sample size affected the outcome of any of our analyses, and performed a series of simulations to assess whether our TD measures, trait variance and bin filling, are auto-correlated with SR and PD (see supplementary materials).

Analyses were conducted using a subsampling of unnested whole clades from the supertree of all mammals. In the subsampling scheme, clades of various sizes were chosen so that the sample sizes added up to as close as possible to the total number of species for which trait data were available. For example, if an analysis included 3000 species, one unnested clade of each of sizes ten through 77 (the sum of 10-77 = 2958) would have been sampled if possible. In practice, since unnested clades of every possible sample size rarely occur in any given tree, between 70% and 90% of all possible species were included in analyses of unnested clades.

Regardless of whether there is a true functional relationship between the variables we examined or one driven by phylogeny, we aim to address the empirical question of whether PD and SR are reliable predictors of TD in an assemblage or community. Therefore, we did not perform phylogenetically informed analyses such as independent contrasts or phylogenetic GLM.

Geographic Analyses

Only species with data available for at least one of the two traits, geographic distribution and phylogeny were included in the geographical analyses (see Table 1 for sample sizes). Body mass displays strong phylogenetic signal, while range area shows weak phylogenetic signal (see Table 1).

In order to investigate whether PD is a better indicator of TD than SR, we calculated residual PD from a global loess regression of PD and SR [3, 4] in grid cells and assessed the correlations in regions where PD departs the most from the expectation of SR – the top and bottom five percentiles. Biodiversity loss, measured as SR, PD and TD loss, was

defined as the difference between current biodiversity and the biodiversity left after removal of all the species that are currently recognized as threatened by IUCN [5]. Therefore, SR loss is the number of threatened species occurring in the region, while PD loss and TD loss are the differences between current estimated PD and TD in the region and the remnant PD and TD if all the threatened species in the region were extinct.
The primary purpose of our geographic analyses is to determine whether SR and PD are reliable predictors of variation in TD across the globe. Therefore, we analysed spatial correlations using Spearman's rank without correcting for possible spatial autocorrelation (though we are aware of the problem of violating the assumption of data independency). In the case that PD and SR can be used to predict TD, it may well be due largely to spatial autocorrelation among all three variables. However, this would not diminish the utility of PD and SR in predicting TD. Further, a major goal of our analyses is to identify geographic localities with high biodiversity in multiple dimensions. The value of conserving a region with large biodiversity is not necessarily reduced because neighbouring regions also have large biodiversity. To visualize patterns of overlap between areas of high PD and TD, we constructed a map highlighting cells within the top five percentiles of PD, trait variance, and trait bins for both mass and range area, in comparisons with distributions of biodiversity hotspots recognized by Conservation International (CI, www.biodiversityhotspots.org).

Discussion of Correlations Between Threatened PD loss and TD loss

Despite the weak correlation between total PD and mass variance, PD loss appears a good indicator of variance loss for both adult body mass and geographic range area (Table S2 and S3). Our results contrast somewhat with those of Fritz and Purvis [6], who found no correspondence between areas where unusually high amounts of PD and mass variance were at risk. However, Fritz and Purvis did not examine the overall global relationship between threatened PD and mass variance. This could explain the difference between their results and ours. Similar to their study, we showed that there is relatively little geographic overlap between areas that happen to have the highest PD and mass variance (Fig. S6a). However, when all areas are considered we show that there is a strong correlation between threatened PD and mass variance both globally (Table S2) and within continental regions (Table S3). The overall correlation between PD loss and loss in adult body mass variance may be explained by the observation that high extinction risk is associated with large body mass in most regional assemblages [6, 7]. As large bodied species are generally outliers in a mass distribution of a local assemblage (i.e., there tend to be relatively few of them) they contribute disproportionately to the mass variance of the assemblage. Conversely, as species with extremely restricted geographic ranges often have high extinction risk [8, 9], extinctions of threatened species might reduce the number of species on the small end of the range area spectrum. However, unlike the case for the large body mass, because most species in an assemblage have relatively small range areas [10], removing threatened species has a relatively small effect on the variance in range area resulting in a weaker correlation between PD loss and TD loss in terms of geographic range variance.
We also found that the number of threatened species as well as potential PD loss was a good predictor of the number of bins that would be emptied by extinction for both traits. In the case of body mass this is most likely because large body mass bins tend to only

include a few species, and because large species are often threatened, especially in the tropics where threats to wildlife are most severe [7, 8, 11]. Thus, extinction of threatened species will often tend to eliminate all large bodied species from local assemblages. In comparison to mass, geographic range area shows a weaker correlation with losses in bin filling, PD, and SR. We speculate that this is because species with small geographic ranges are much more likely to be threatened than species with large ranges [8, 9]. Species with small ranges also tend to outnumber species with large geographic ranges [10], and thus even after loss of many species with restricted ranges at least a few species with small geographic ranges are likely to remain.

Additional References

1. Blomberg S., Garland Jr T. 2002 Tempo and mode in evolution: phylogenetic inertia, adaptation and comparative methods. *Journal of Evolutionary Biology* 15(6), 899.
2. Blomberg S.P., Jr. T.G., Ives A.R. 2003 Testing For Phylogenetic Signal In Comparative Data: Behavioral Traits Are More Labile. *Evolution* 57(4), 717-745.
3. Davies T.J., Fritz S.A., Grenyer R., Orme C.D.L., Bielby J., Bininda-Emonds O.R.P., Cardillo M., Jones K.E., Gittleman J.L., Mace G.M., et al. 2008 Phylogenetic trees and the future of mammalian biodiversity. *Proceedings of the National Academy of Sciences* 105(Supplement 1), 11556-11563. (doi:10.1073/pnas.0801917105).
4. Forest F., Grenyer R., Rouget M., Davies T.J., Cowling R.M., Faith D.P., Balmford A., Manning J.C., Proches S., van der Bank M., et al. 2007 Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature* 445(7129), 757-760.
5. Huang S., Davies T.J., Gittleman J.L. 2012 How global extinctions impact regional biodiversity in mammals. *Biology Letters*. (doi:10.1098/rsbl.2011.0752).
6. Fritz S.A., Purvis A. 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(1693), 2435-2441. (doi:10.1098/rspb.2010.0030).
7. Fritz S.A., Bininda-Emonds O.R.P., Purvis A. 2009 Geographical variation in predictors of mammalian extinction risk: big is bad, but only in the tropics. *Ecology Letters* 12(6), 538-549.
8. Cardillo M., Mace G.M., Gittleman J.L., Jones K.E., Bielby J., Purvis A. 2008 The predictability of extinction: biological and external correlates of decline in mammals. *Proceedings of the Royal Society B: Biological Sciences* 275(1641), 1441-1448.
9. Cardillo M., Mace G.M., Gittleman J.L., Purvis A. 2006 Latent extinction risk and the future battlegrounds of mammal conservation. *Proceedings of the National Academy of Sciences* 103(11), 4157-4161.
10. Gaston K.J. 1996 Species-range-size distributions: patterns, mechanisms and implications. *Trends in Ecology & Evolution* 11(5), 197-201.
11. Cardillo M., Mace G.M., Jones K.E., Bielby J., Bininda-Emonds O.R.P., Sechrest W., Orme C.D.L., Purvis A. 2005 Multiple Causes of High Extinction Risk in Large Mammal Species. *Science* 309(5738), 1239-1241. (doi:10.1126/science.1116030).

Additional Tables & Figures

Table S1. Spearman's rank correlations (ρ) of TD with SR and PD across clades. TD is measured as trait variance and trait bins, for 12 traits in mammals.

	Trait Variance		Trait Bins	
Traits	**PD vs TD**	**SR vs TD**	**PD vs TD**	**SR vs TD**
Geographic range area (km^2)	**0.389 (<0.001)**	**0.292 (0.013)**	**0.649 (<0.001)**	**0.690 (<0.001)**
Adult body mass (g)	**0.350 (0.004)**	0.124 (0.327)	**0.438 (<0.001)**	**0.273 (0.029)**
Litter size (no. of offspring)	**0.310 (0.027)**	0.171 (0.230)	0.256 (0.070)	0.218 (0.124)
Adult head-body length (mm)	**0.491 (<0.001)**	**0.342 (0.025)**	0.250 (0.107)	**0.328 (0.032)**
Gestation length (day)	**0.701 (<0.001)**	0.267 (0.121)	**0.599 (<0.001)**	0.261 (0.129)
Population density (number/km^2)	0.198 (0.321)	0.145 (0.468)	**0.491 (0.009)**	**0.546 (0.003)**
Litters per year (no. of litters)	**0.414 (0.033)**	0.044 (0.828)	0.096 (0.632)	0.068 (0.736)
Home range (km^2)	-0.016 (0.943)	-0.048 (0.825)	0.397 (0.057)	**0.521 (0.009)**
Adult forearm length (mm)	0.383 (0.060)	**0.442 (0.028)**	0.336 (0.100)	**0.498 (0.011)**
Teat number (no. of teats)	**0.491 (0.020)**	0.221 (0.323)	**0.522 (0.013)**	0.210 (0.346)
Basal metabolic rate (mL O_2/hr)	-0.136 (0.554)	-0.300 (0.186)	0.197 (0.393)	-0.077 (0.739)
Mass specific metabolic rate (mL O_2h^{-1}/ g)	0.390 (0.082)	0.0.068 (0.771)	0.371 (0.096)	0.265 (0.244)

Table S2: Spearman's rank correlations of TD with PD and SR across geographic regions

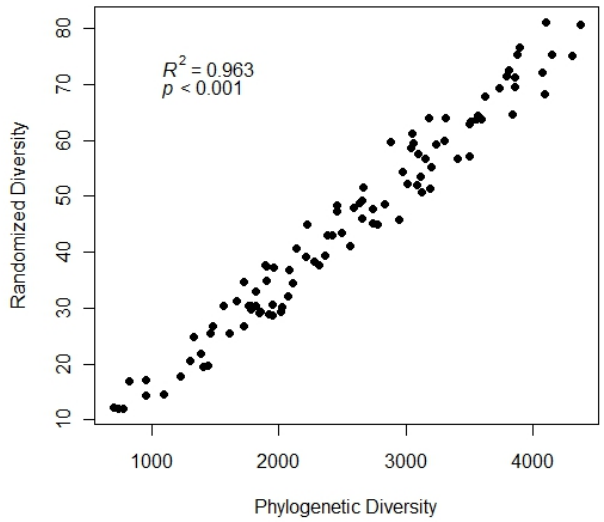
TD measures	Traits	Total		Loss	
		SR	PD	SR	PD
Trait variance	Adult body mass	-0.01 (p=0.087)	-0.02 (p=0.001)	0.675 (p<0.001)	0.711 (p<0.001)
	Geographic range area	0.231 (p<0.001)	0.246 (p<0.001)	0.366 (p<0.001)	0.323 (p<0.001)
Trait bins	Adult body mass	0.675 (p<0.001)	0.601 (p<0.001)	0.626 (p<0.001)	0.696 (p<0.001)
	Geographic range area	0.640 (p<0.001)	0.650 (p<0.001)	0.388 (p<0.001)	0.388 (p<0.001)

Table S3. Spearman's rank correlations of TD with SR and PD in five regions.

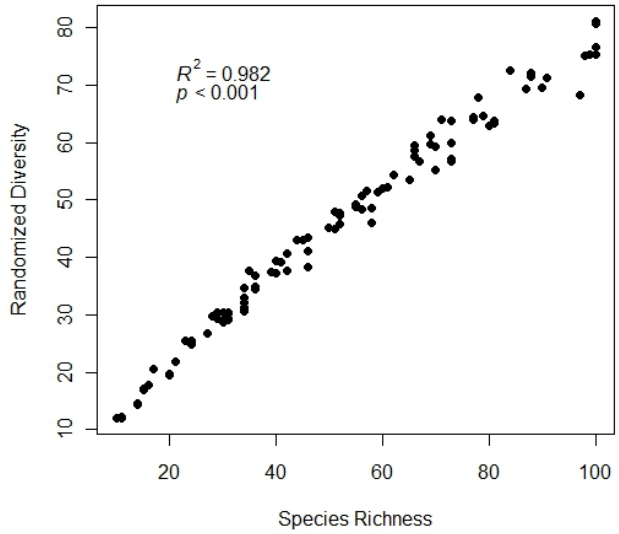
Continent	TD measures	Traits	Total		Loss	
			SR	PD	SR	PD
Africa	Trait variance	Adult body mass	0.582 (p<0.001)	0.597 (p<0.001)	0.684 (p<0.001)	0.820 (p<0.001)
		Geographic range area	0.112 (p<0.001)	0.107 (p<0.001)	0.367 (p<0.001)	0.460 (p<0.001)
	Trait bins	Adult body mass	0.897 (p<0.001)	0.903 (p<0.001)	0.676 (p<0.001)	0.780 (p<0.001)
		Geographic range area	0.595 (p<0.001)	0.591 (p<0.001)	0.472 (p<0.001)	0.540 (p<0.001)
Australia	Trait variance	Adult body mass	-0.121 (p<0.001)	-0.129 (p<0.001)	-0.590 (p<0.001)	-0.502 (p<0.001)
		Geographic range area	0.322 (p<0.001)	0.276 (p<0.001)	0.444 (p<0.001)	0.271 (p<0.001)
	Trait bins	Adult body mass	0.843 (p<0.001)	0.853 (p<0.001)	0.263 (p<0.001)	0.322 (p<0.001)
		Geographic range area	0.544 (p<0.001)	0.495 (p<0.001)	0.509 (p<0.001)	0.337 (p<0.001)
Euro-Asia	Trait variance	Adult body mass	0.481 (p<0.001)	0.490 (p<0.001)	0.530 (p<0.001)	0.706 (p<0.001)
		Geographic range area	0.208 (p<0.001)	0.232 (p<0.001)	0.351 (p<0.001)	0.200 (p<0.001)
	Trait bins	Adult body mass	0.858 (p<0.001)	0.849 (p<0.001)	0.328 (p<0.001)	0.451 (p<0.001)
		Geographic range area	0.625 (p<0.001)	0.642 (p<0.001)	0.347 (p<0.001)	0.357 (p<0.001)
North America	Trait variance	Adult body mass	-0.151 (p<0.001)	-0.194 (p<0.001)	0.128 (p<0.001)	0.337 (p<0.001)
		Geographic range area	0.383 (p<0.001)	0.373 (p<0.001)	0.211 (p<0.001)	-0.020 (p=0.268)
	Trait bins	Adult body mass	0.602 (p<0.001)	0.523 (p<0.001)	0.429 (p<0.001)	0.546 (p<0.001)
		Geographic range area	0.807 (p<0.001)	0.794 (p<0.001)	0.511 (p<0.001)	0.322 (p<0.001)
South America	Trait variance	Adult body mass	-0.329 (p<0.001)	-0.333 (p=0.001)	0.562 (p<0.001)	0.465 (p<0.001)
		Geographic range area	-0.232 (p<0.001)	-0.240 (p<0.001)	0.541 (p<0.001)	0.514 (p<0.001)
	Trait bins	Adult body mass	0.723 (p<0.001)	0.726 (p<0.001)	0.585 (p<0.001)	0.568 (p<0.001)
		Geographic range area	0.057 (p=0.013)	0.037 (p=0.107)	0.200 (p<0.001)	0.178 (p<0.001)

Table S4. Spearman's rank correlations of TD with SR and PD in regions with residual PD in the highest and lowest 5% quantiles.

Traits	TD measures	High residual PD		Low residual PD	
		SR	PD	SR	PD
Adult body mass	Trait variance	0.480 (p=0.087)	0.481 (p=0.001)	-0.464 (p<0.001)	-0.433 (p<0.001)
	Trait bins	0.733 (p<0.001)	0.736 (p<0.001)	-0.146 (p<0.001)	-0.124 (p<0.001)
Geographic range area	Trait variance	-0.220 (p<0.001)	-0.223 (p<0.001)	-0.152 (p<0.001)	-0.167 (p<0.001)
	Trait bins	0.344 (p<0.001)	0.344 (p<0.001)	0.045 (p=0.215)	0.028 (p=0.433)



(a)



(b)

Figure S1: Randomized trait diversity versus the (a) phylogenetic diversity and (b) species richness of clades of between 10 and 100 random species drawn from the mammal supertree of Fritz et al. (2009). Randomized diversity was calculated from the summed branch lengths of a UPGMA dendrogram generated from four random normal trait variables (e.g., if a clade of ten random species was drawn from the mammal supertree, a UPGMA tree would be constructed from four "traits" for the ten species with values drawn from a random normal distribution).

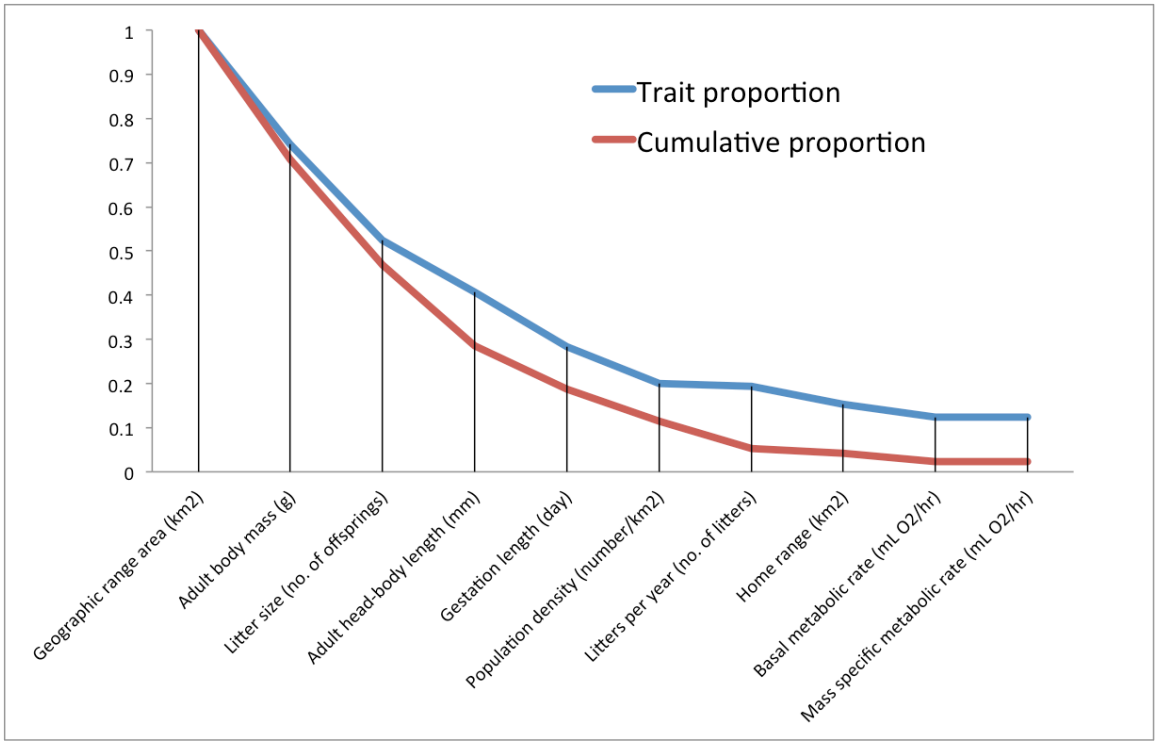


Figure S2: Proportion of all terrestrial mammal species sampled for various traits data used for this study. Cumulative proportion is the proportion included when complete case analysis is applied to successive additional traits (e.g., when the best sampled four traits are used, complete case analysis includes less than 30% of all species). Illustrated are the ten traits that yield the maximum possible cumulative ten trait proportion.

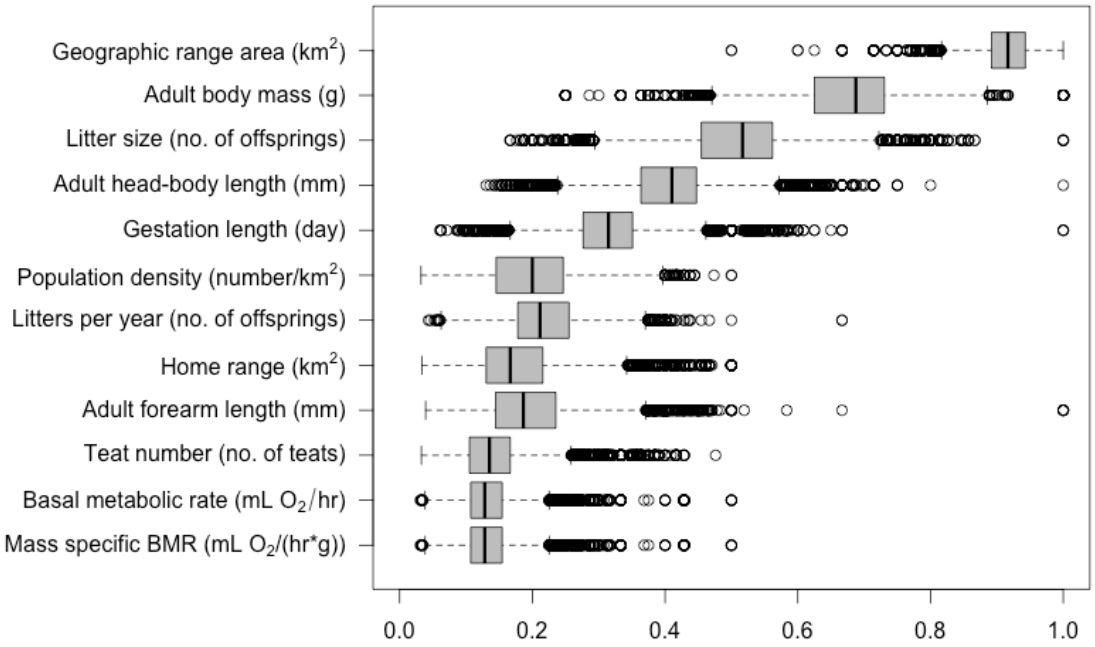
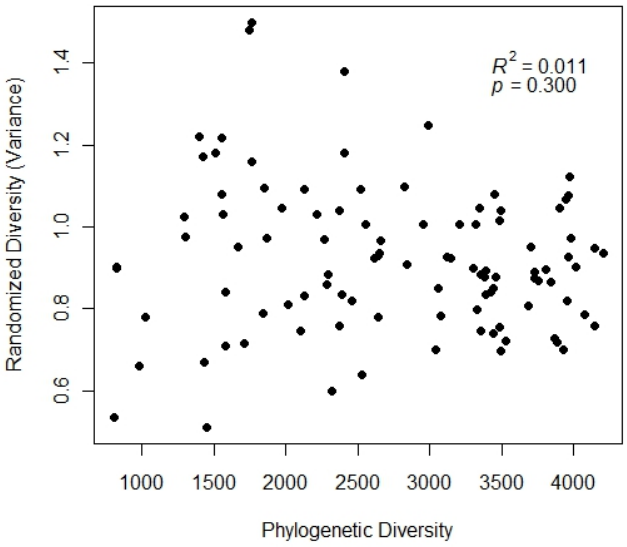
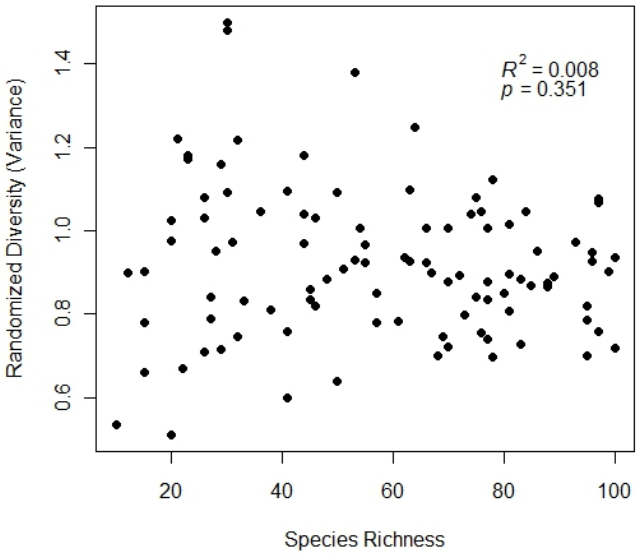


Figure S3: Proportion of the species in each 100 * 100 km^2 grid used in analysis of global biodiversity patterns cell with data available for each species trait.

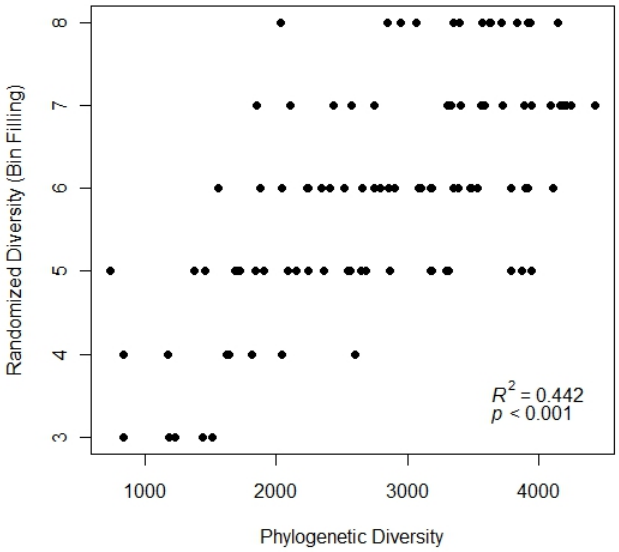


(a)

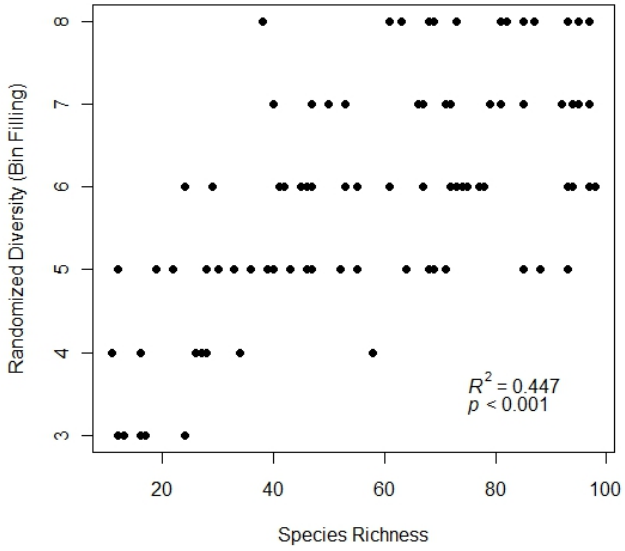


(b)

Figure S4: Randomized trait variance versus the (a) phylogenetic diversity and (b) species richness of clades of between 10 and 100 random species drawn from the mammal supertree of Frits et al. (2009). Randomized variance was calculated using draws from a random normal distribution.



(a)



(b)

Figure S5: Randomized bin filling versus the (a) phylogenetic diversity and (b) species richness of clades of between 10 and 100 random species drawn from the mammal supertree of Frits et al. (2009). Randomized bin filling was calculated using draws from a random normal lognormal distribution spanning eight orders of magnitude of variation.

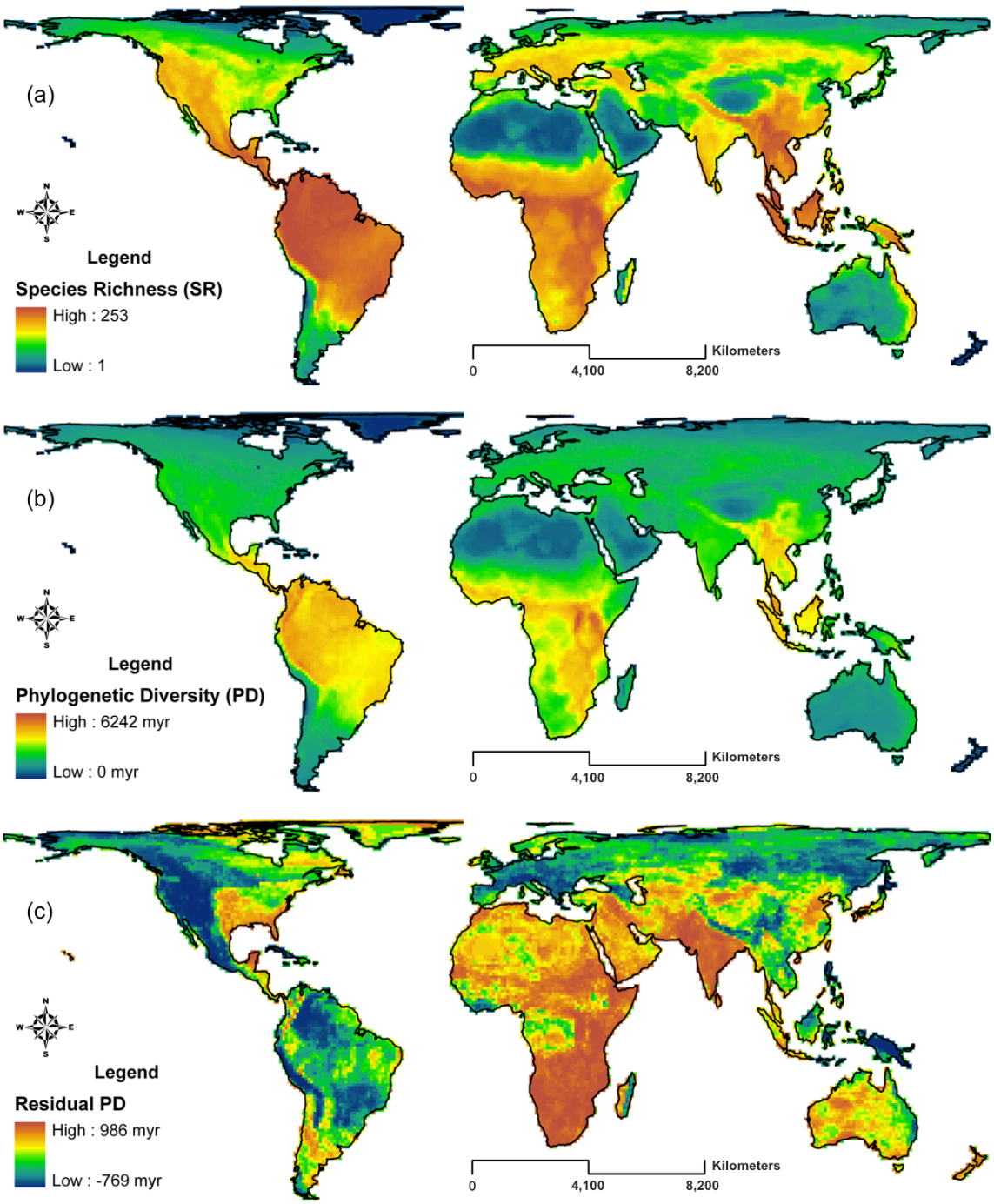


Figure S6: SR (a), PD (b), and residual PD (c) of terrestrial mammals globally.

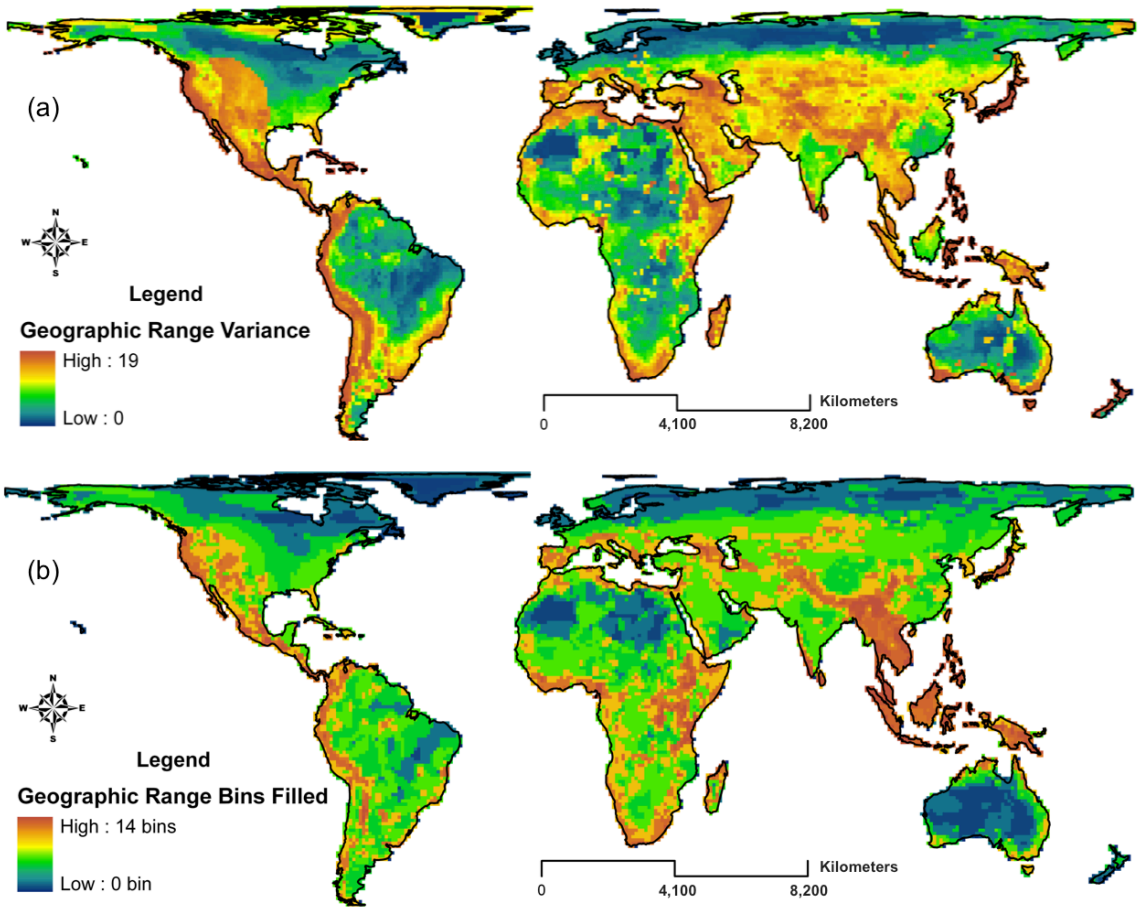


Figure S7: Variance (a) and bins (b) of geographic range area of terrestrial mammals globally.

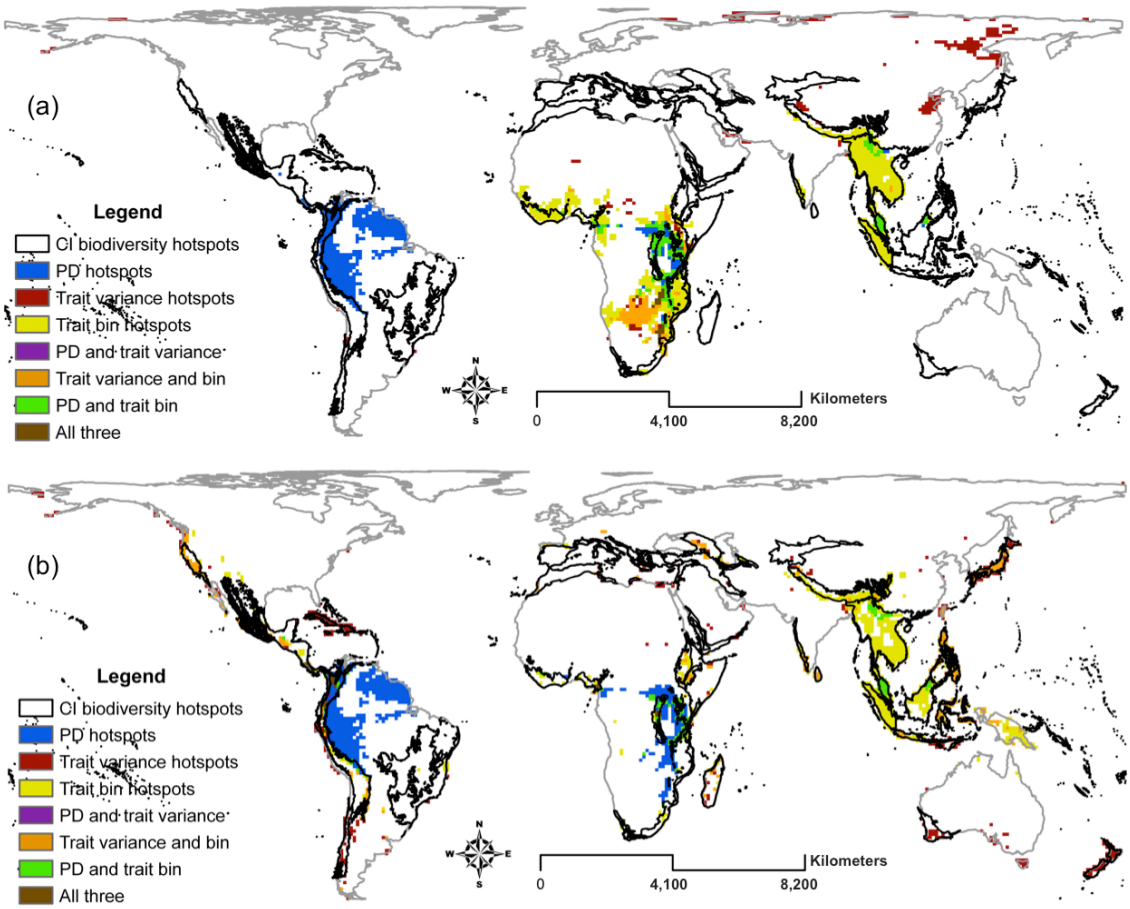


Figure S8: Hotspots (top 5% quantile) of PD and TD using two different measures, for two traits: (a) adult body mass, and (b) geographic range area. Dark outlines are Conservation International biodiversity hotspots.

APPENDIX B

SUPPORTING INFORMATION FOR CHAPTER 3

Table 1 PD losses if species in currently recognized threatened categories are extinct. IUCN categories: VU – Vulnerable; EN – Endangered; and CR – Critically Endangered.

IUCN Threat status	**VU + EN + CR**	**EN + CR**	**CR**
Number of species	1004 (21%)	552 (12%)	151 (3%)
PD loss if the species are extinct (million year)	8470 (14%)	4312 (7%)	1106 (2%)
Two-tailed 90% quantile of PD loss from random extinction simulation (million year)	[8044, 8803]	[4249, 4839]	[1066, 1380]
Terminal branch length loss if the species are extinct (million year)	5700	2783*	740
Two-tailed 90% quantile of terminal branch loss from random extinction simulation (million year)	[5458, 6387]	[2847, 3567]	[661, 1068]
Internal branch length loss if the species are extinct (million year)	2771	1529	365
Two-tailed 90% quantile of internal branch loss from random extinction simulation (million year)	[2232, 2760]	[1136, 1544]	[253, 480]

*At the global scale, endangered species tend to be on shorter terminal branches, resulting in underestimates of the impact of global extinction on regional communities and ecosystems.

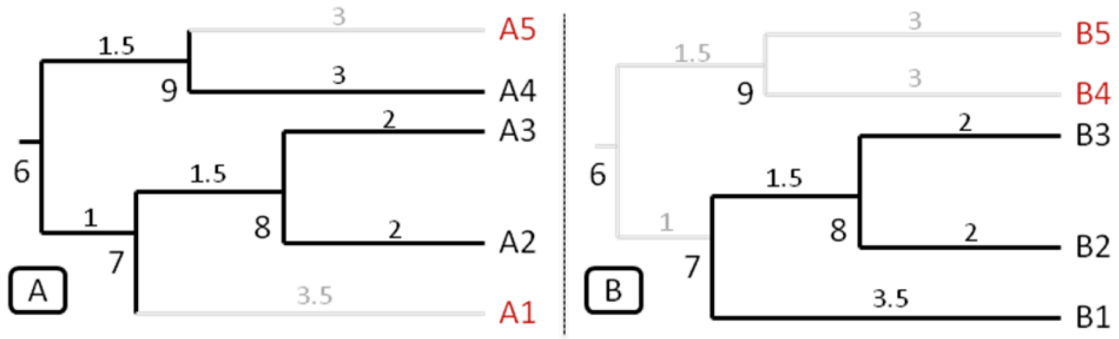


Fig. S1. Hypothetical phylogeny (branch lengths indicated as the numbers above the branches) of species assemblages in two different areas illustrating how losing the same number of species can lead to low (A) or high (B) loss of local phylogenetic diversity. The PD of the original species assemblages in both area A and B are the same: $PD_{originalA} = PD_{originalB} = 17.5$ myr. If threatened species (in red) become extinct, the remaining species would only have $PD_{remainingA} = 11$ myr in (A) and $PD_{remainingB} = 9$ myr in (B). By our definition, these two assemblages (areas) would have lost different amounts of PD due to loss of the same total numbers of species: $PD_{lostA} = 6.5$ myr but $PD_{lostB} = 8.5$ myr.

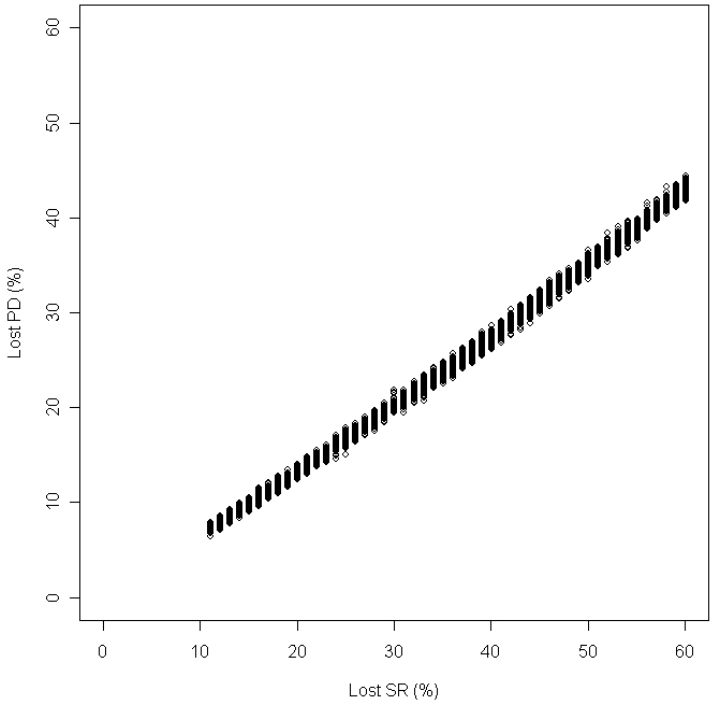


Fig. S2. The relationship between percent loss in phylogenetic diversity (PD) and species richness (SR) through random extinction (terrestrial mammals, n = 4796). For each level of SR loss, random extinction was simulated for 1000 times.

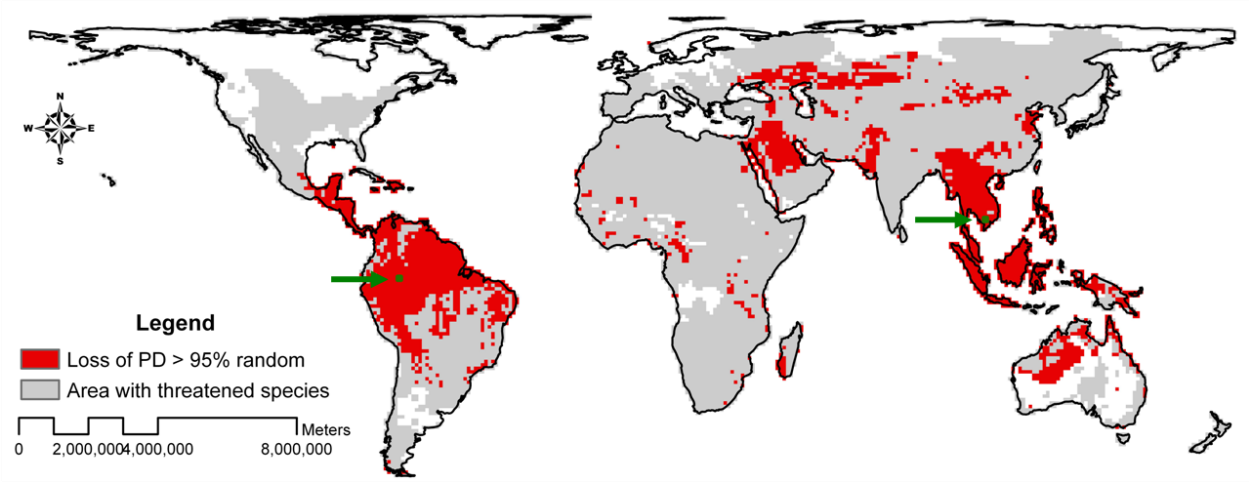


Fig. S3. Results of the comparison between estimated and simulated PD loss in 1000 simulations for each 10,000 km^2 grid cell. Areas shown in red will lose greater PD following the loss of currently threatened species than at least 95% of all random extinction events (assuming the loss of the same number of randomly chosen species). Phylogenetic trees of species in two example grid cells (green cells, indicated by arrows), one in the Amazon and the other in South Asia, are illustrated in Fig. S2-3.

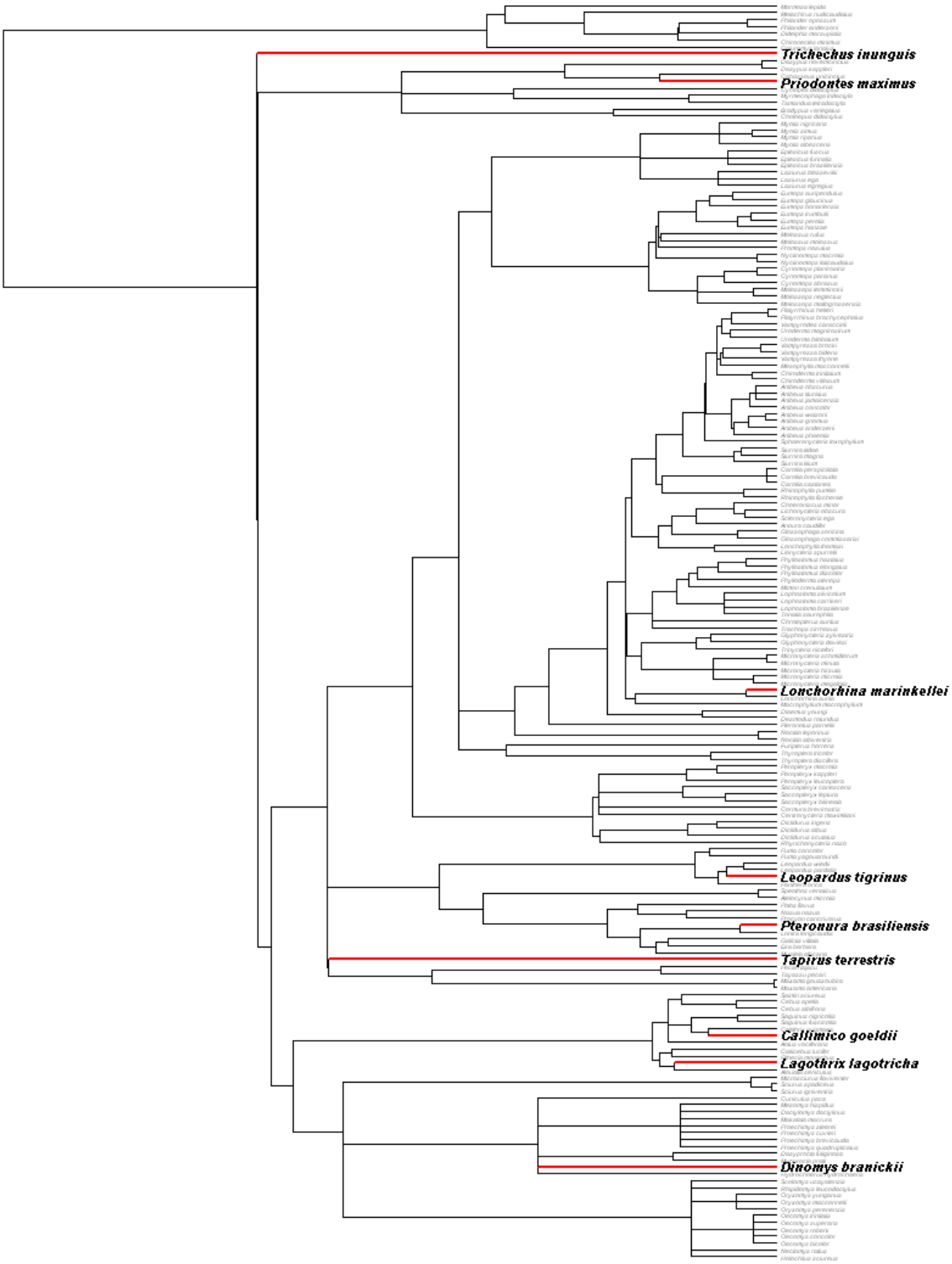


Figure S4. The phylogeny of species in a grid cell in the Amazon, as indicated on figure S1. Branches to be lost are colored in red. Only threatened tips (species) are visibly labeled with species Latin binomials.

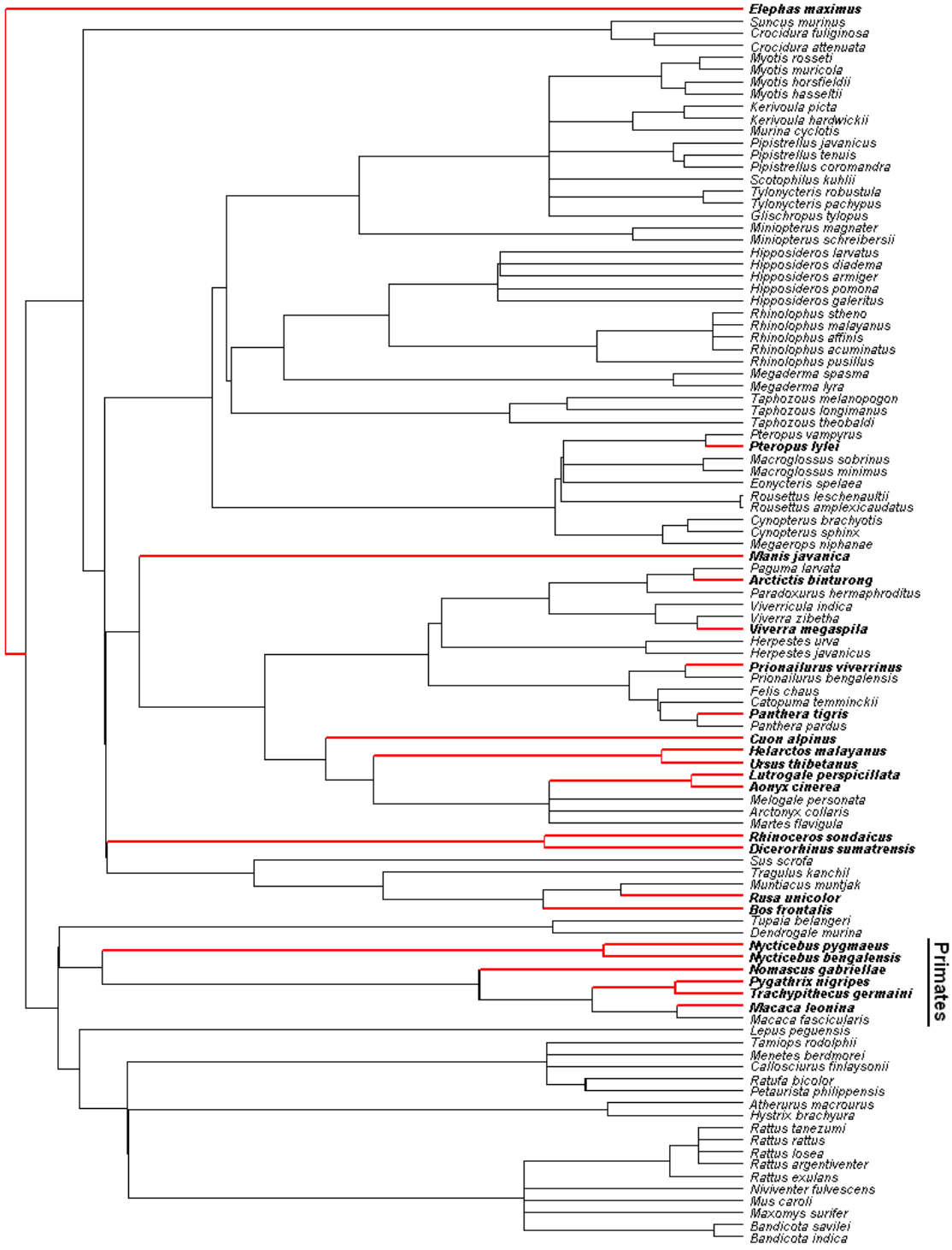


Figure S5. The phylogeny of species in a grid cell in the South Asian biodiversity hotspot, as indicated on figure S1. Threatened species are labeled in bold and branches to be lost are colored in red. Note that 6 out of 7 species of primates (labeled) from this location are threatened with extinction.

APPENDIX C

SUPPORTING INFORMATION FOR CHAPTER 4

Non-parametric estimators of species diversity

Using four non-parametric estimators, we estimated parasite species richness (PSR) based on the reported PSR (PSR_R) and information on how the number of reported parasites increased with greater sampling effort. Specifically, these methods estimate how many more species would be found if extra sampling effort is made based on the number of rare species (found in only one and/or two samples) that have been found. Specifically, we used four non-parametric estimators, described as below.

Chao 2 estimator (PSR_C):

$$PSR_C = PSR_R + (\frac{a^2}{2b})$$

where *a* is the number of parasite species reported in only 1 study and *b* is the number of parasite species reported in 2 studies (Poulin 1998; Chao 2005).

Jackknife estimator (PSR_J):

$$PSR_J = PSR_R + \frac{a(H-1)}{H}$$

where *H* is the number of studies (Poulin 1998).

Bootstrap estimator (PSR_B):

$$PSR_B = PSR_R + \sum_{j=1}^{PSR_R} \left[1 - \left(\frac{h_j}{H}\right)\right]^H$$

where h_j is the number of studies that reported parasite species j (Poulin 1998).

ACE (ACE 1 & ACE 2) estimators (PSR_{A1} & PSR_{A2}):

$$PSR_{A1} = PSR_R - (a+b) + \frac{a+b}{1-\frac{a}{a+b}} + \frac{a}{1-\frac{a}{a+b}} \cdot \gamma$$

where $\gamma = max\left\{\frac{a+b}{1-\frac{a}{a+b}} \cdot \frac{b}{(a+b)(a+b-1)} - 1, 0\right\}$, and for highly heterogeneous case,

$$PSR_{A2} = PSR_R - (a+b) + \frac{a+b}{1-\frac{a}{a+b}} + \frac{a}{1-\frac{a}{a+b}} \cdot \gamma'$$

where $\gamma' = max\left\{\gamma \cdot \left(1 + \frac{\frac{ab}{a+b}}{(ab-1)\left(1-\frac{a}{a+b}\right)}\right), 0\right\}$ (Chao 2005; Chao *et al.* 2006).

The Lovich-Gibbons revised two-step ratio of size sexual dimorphism

We calculated host sexual dimorphism (D) of body mass (kg) :

$$D = \begin{cases} \frac{M_f}{M_m} - 1, if\ M_f \geq M_m \\ 1 - \frac{M_m}{M_f}, if\ M_m \geq M_f \end{cases}$$

where M_m is the average adult body mass of male individuals of the host species, and female (M_f) is the average adult body mass of females (Gibbons & Lovich 1990; Lovich & Gibbons 1992).

References:

Chao A. (2005). Species richness estimation. In: *Encyclopedia of statistical sciences* (ed. Samuel Kotz CBR, N. Balakrishnan, Brani Vidakovic). Wiley-Interscience New York, pp. 7907-7916.

Chao A., C. Li P., Agatha S. & Foissner W. (2006). A statistical approach to estimate soil ciliate diversity and distribution based on data from five continents. *Oikos*, 114, 479-493.

Gibbons J.W. & Lovich J.E. (1990). Sexual dimorphism in turtles with emphasis on the slider turtle (Trachemys scripta). *Herpetological monographs*, 4, 1-29.

Lovich J. & Gibbons J.W. (1992). A review of techniques for quantifying sexual size dimorphism. *Growth, development, and aging: GDA*, 56, 269.

Poulin R. (1998). Comparison of Three Estimators of Species Richness in Parasite Component Communities. *The Journal of Parasitology*, 84, 485-490.

Additional tables and figures

Table S2 The 45 host species that have been sampled in at least 9 reports for parasites. The total number of parasite species (reported PSR) and number of generalist parasite species (reported generalist PSR) reported from literature are summarized. Parasite diversity was also estimated using four non-parametric estimators: Chao 2, ACE, Jackknife, and Bootstrap estimators. Host species included in the random forest analyses are in bold.

Host species	Host family	Number of reports	Reported PSR	Reported generalist PSR	Chao 2 estimator	ACE estimator	Jackknife estimator	Bootstrap estimator
Acinonyx jubatus	Felidae	13	14	13	24.13	25.45	22.25	17.65
Canis aureus	Canidae	15	31	29	93.50	90.62	54.08	40.43
Canis latrans	Canidae	80	92	74	152.63	135.69	139.35	112.43
Canis lupus	Canidae	71	80	77	160.67	117.19	123.31	98.00
Cerdocyon thous	Canidae	24	24	18	124.00	60.92	42.95	31.42
Crocuta crocuta	Hyaenidae	14	20	17	92.25	57.78	35.79	26.26
Cystophora cristata	Phocidae	13	14	14	26.50	31.50	22.89	17.88
Enhydra lutris	Mustelidae	26	27	18	107.67	118.67	48.08	35.29
Erignathus	Phocidae	13	10	10	22.25	20.00	16.22	12.66

Host species	Host family	Number of reports	Reported PSR	Reported generalist PSR	Chao 2 estimator	ACE estimator	Jackknife estimator	Bootstrap estimator
barbatus								
Felis silvestris	Felidae	28	38	34	129.13	71.30	63.83	48.36
Genetta genetta	Viverridae	22	21	17	53.00	49.00	36.11	27.21
Halichoerus grypus	Phocidae	29	21	21	119.00	35.74	34.42	26.31
Herpestes ichneumon	Phocidae	15	20	15	68.17	76.67	35.30	26.25
Leopardus pardalis	Felidae	12	33	33	273.25	288.75	60.56	43.95
Lontra canadensis	Mustelidae	14	22	17	112.25	81.71	39.42	28.94
Lutra lutra	Mustelidae	30	60	38	253.14	243.53	109.64	79.59
Lycaon pictus	Canidae	16	12	11	44.00	23.00	19.43	15.02
Lynx lynx	Felidae	20	35	33	87.07	90.59	60.31	45.48
Lynx pardinus	Felidae	25	47	45	66.53	71.46	70.96	58.11
Lynx rufus	Felidae	39	65	54	132.50	130.45	108.59	83.31
Martes foina	Mustelidae	26	39	34	91.56	88.17	66.79	50.49
Meles meles	Musteliae	84	54	42	76.53	73.14	79.61	65.84
Mephitis mephitis	Mephitidae	42	48	41	184.90	118.92	84.05	62.19
Mirounga angustirostris	Phocidae	13	37	32	109.90	71.45	61.75	47.14
Mustela erminea	Mustelidae	23	34	32	38.24	42.16	45.29	40.46
Mustela nivalis	Mustelidae	14	24	23	24.66	27.16	28.29	27.50
Mustela putorius	Mustelidae	32	46	45	68.56	57.06	64.14	54.44
Neovison vison	Mustelidae	37	64	55	98.23	101.00	99.68	80.10
Nyctereutes procyonoides	Canidae	40	66	52	80.00	86.92	93.10	79.94
Odobenus rosmarus	Odobenidae	10	13	12	43.25	48.75	22.63	16.98
Pagophilus groenlandicus	Phocidae	15	15	15	51.00	40.71	26.00	19.48
Panthera leo	Felidae	40	47	38	169.50	99.50	81.08	60.44
Panthera pardus	Felidae	13	19	19	181.00	190.00	35.20	25.38
Phoca vitulina	Phocidae	63	67	57	101.03	92.90	99.35	81.47
Procyon lotor	Procyonidae	151	182	92	358.73	303.32	289.16	226.81
Puma concolor	Felidae	46	69	63	95.27	96.20	102.23	84.61
Pusa hispida	Phocidae	24	23	15	77.00	57.50	39.94	29.85
Urocyon cinereoargenteus	Canidae	34	56	51	144.89	115.46	94.57	71.80
Ursus americanus	Ursidae	44	76	65	170.53	170.65	129.63	98.13
Ursus arctos	Ursidae	22	43	37	423.25	210.70	79.71	57.23
Ursus maritimus	Ursidae	14	21	11	102.00	75.00	37.00	27.47
Vulpes lagopus	Canidae	19	32	27	110.13	76.44	55.33	41.43
Vulpes velox	Canidae	10	31	27	58.00	46.88	47.20	38.02
Vulpes vulpes	Canidae	227	157	120	238.67	214.44	226.66	187.75
Zalophus californianus	Otariidae	30	66	46	274.29	210.00	118.07	86.48

[1] Adult body mass data was not available for *Neovison vison*, thus this species was not included in analyses constructing GLM models.

Table S2 Minimum sampling effort (MinSE) to cover (with 95% quantiles) the final estimates of PSR using different methods. The lowest MinSE for each host species are in bold.

Host species	Number of reports*	Chao 2		ACE		Bootstrap		Jackknife	
		PSR	**MinSE**	**PSR**	**MinSE**	**PSR**	**MinSE**	**PSR**	**MinSE**
Vulpes vulpes	207	239	9	214	**8**	188	192	227	178
Procyon lotor	127	359	**5**	303	**5**	227	121	289	114
Meles meles	66	77	**5**	73	**5**	64	58	80	49
Canis latrans	73	153	**5**	136	**5**	112	68	139	64
Canis lupus	63	161	**5**	117	**5**	98	51	123	53
Phoca vitulina	50	101	**5**	93	**5**	81	45	99	41

*The number of reports that have observed a parasite in the host population, which is sometimes lower than the overall sampling effort for which we count all the reports that have searched for parasites but might or might not have observed any parasite infection.

Table S3 Summary statistics of GLM models showing the relationship between parasite diversity (observed and estimated) and individual variables. All models have the number of reports as a covariable. The significant associations (except those with sampling effort) and the lowest AICc score for each response variable are in bold.

	Observed PSR		**Chao 2 estimate**		**ACE estimate**	
Model/*Variable*	**Coefficients (*p*)**	**AICc**	**Coefficients (*p*)**	**AICc**	**Coefficients (*p*)**	**AICc**
Number of reports	0.781 ± 0.074 ($p < 0.001$)	17	0.469 ± 0.160 ($p = 0.007$)	63	0.473 ± 0.152 ($p = 0.004$)	60
+ Terminal branch length		15		58		56
Number of reports	0.776 ± 0.071 ($p < 0.001$)		0.455 ± 0.143 ($p = 0.004$)		0.460 ± 0.140 ($p = 0.003$)	
Terminal branch length	-0.163 ± 0.084 ($p = 0.062$)		**-0.480 ± 0.169 ($p = 0.008$)**		**-0.405 ± 0.166 ($p = 0.021$)**	
+ Adult body mass		19		58		58
Number of reports	0.781 ± 0.076 ($p < 0.001$)		0.450 ± 0.143 ($p = 0.004$)		0.458 ± 0.143 ($p = 0.004$)	
Adult body mass	-0.006 ± 0.032 ($p = 0.857$)		**0.169 ± 0.060 ($p = 0.009$)**		**0.131 ± 0.060 ($p = 0.039$)**	
+ Centroid latitude degree		15		65		62
Number of reports	0.744 ± 0.072 ($p < 0.001$)		0.485 ± 0.167 ($p = 0.007$)		0.476 ± 0.159 ($p = 0.006$)	
Centroid latitude degree	**-0.006 ± 0.003 ($p = 0.041$)**		-0.003 ± 0.006 ($p = 0.680$)		-0.001 ± 0.006 ($p = 0.923$)	
Mean temperature		15		65		62
Number of reports	0.746 ± 0.072 ($p < 0.001$)		0.468 ± 0.167 ($p = 0.009$)		0.460 ± 0.159 ($p = 0.008$)	
Centroid latitude degree	**-0.001 ± 0.001 ($p = 0.050$)**		-0.00002 ± 0.001 ($p = 0.988$)		-0.0004 ± 0.001 ($p = 0.736$)	
Mean PET		**14**		65		62
Number of reports	0.739 ± 0.072 ($p < 0.001$)		0.475 ± 0.169 ($p = 0.009$)		0.463 ± 0.164 ($p = 0.008$)	
Mean PET	**-0.0003 ± 0.0001 ($p = 0.033$)**		0.00004 ± 0.00003 ($p = 0.886$)		-0.00007 ± 0.0003 ($p = 0.814$)	

Table S4 Generalized linear models (GLM) of phylogenetic independent contrasts of estimated parasite diversity. Sampling effort measured as the number of reports was included as a covariate in all models. Complete data were available for 44 host species to be included in these models. ΔAICc for each model was calculated as the difference from the lowest AICc.

	Chao 2 estimator		*ACE estimator*	
Models	**AICc**	**ΔAICc**	**AICc**	**ΔAICc**
Terminal branch length (TBL)	-20.25	0.30	-22.20	3.53
Adult body mass (ABM)	-11.60	8.95	-19.24	6.49
Geographic range area (GRA)	-13.74	6.81	-23.54	2.19
TBL + ABM	-17.72	2.83	-20.41	5.32
TBL + GRA	-20.55	0	-25.73	0
ABM + GRA	-11.18	9.37	-21.63	4.10
All	-18.00	2.55	-23.46	2.27

Table S4 Generalized linear models (GLM) of estimated parasite diversity and host sexual dimorphism. Analyses were repeated using the raw data and using phylogenetic independent contrasts of the variables. Complete data were available for 40 host species to be included in these models. Note that negative values in sexual dimorphism indicate a male bias, and 85% of included host species displayed male biased dimorphism in body mass.

	Chao 2 estimate	**ACE estimate**
Variable	***Coefficients (p)***	***Coefficients (p)***
	Phylogenetic independent contrasts	
Intercept	0.014 ± 0.028 (p = 0.628)	0.009 ± 0.026 (p = 0.736)
Number of reports	**0.008 ± 0.002 (p < 0.001)**	**0.008 ± 0.002 (p < 0.001)**
Sexual dimorphism in body mass	**-1.005 ± 0.224 (p < 0.001)**	**-0.787 ± 0.212 (p < 0.001)**
	Original data	
Intercept	**4.166 ± 0.167 (p < 0.001)**	**4.000 ± 0.156 (p < 0.001)**
Number of reports	**0.008 ± 0.003 (p = 0.003)**	**0.008 ± 0.002 (p = 0.002)**
Sexual dimorphism in body mass	-0.353 ± 0.204 (p = 0.091)	-0.314 ± 0.191 (p = 0.109)

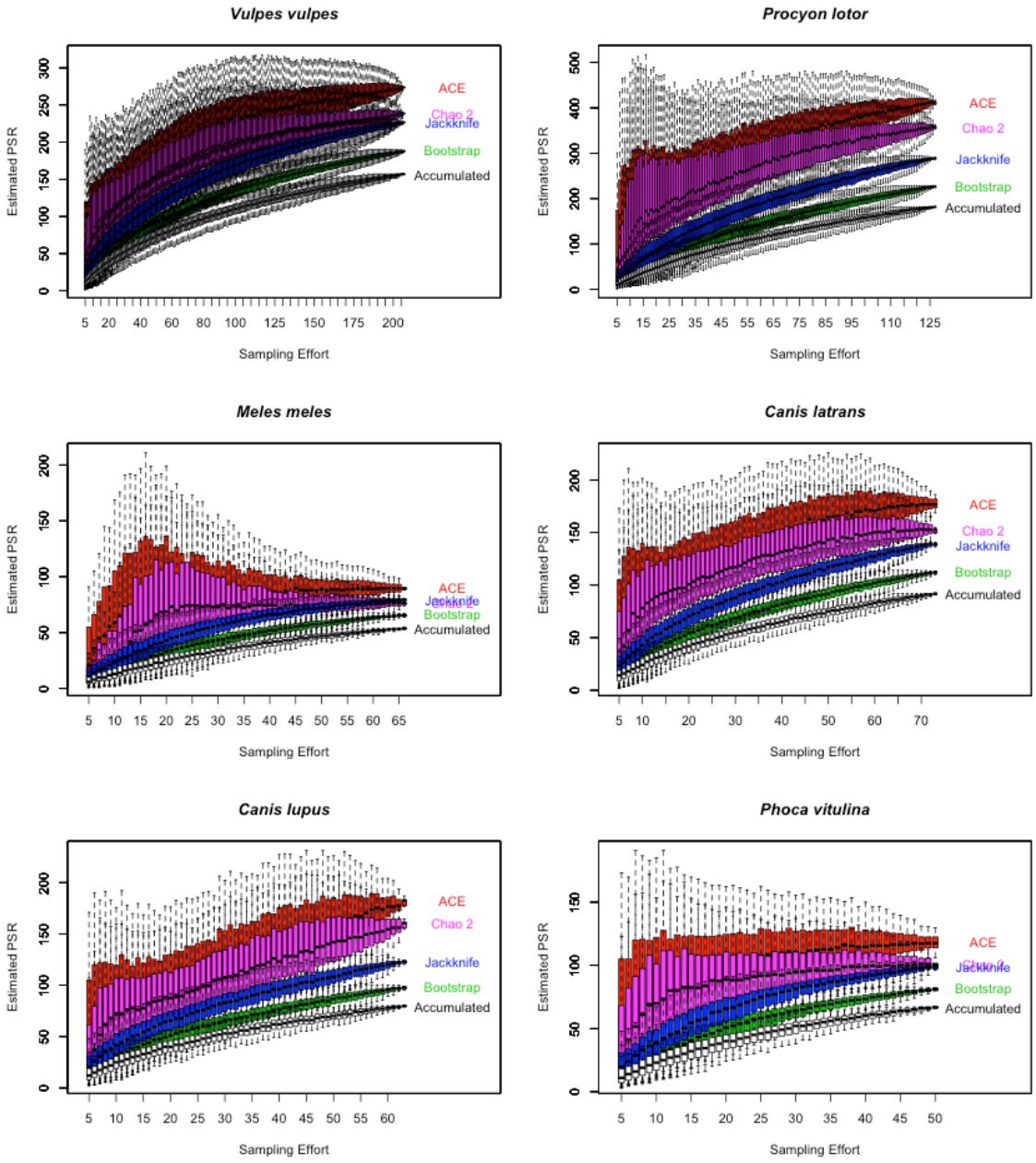


Figure S1 PSR estimated based on randomly subsampled reports using four different methods in comparison with the accumulated PSR observed in the reports.

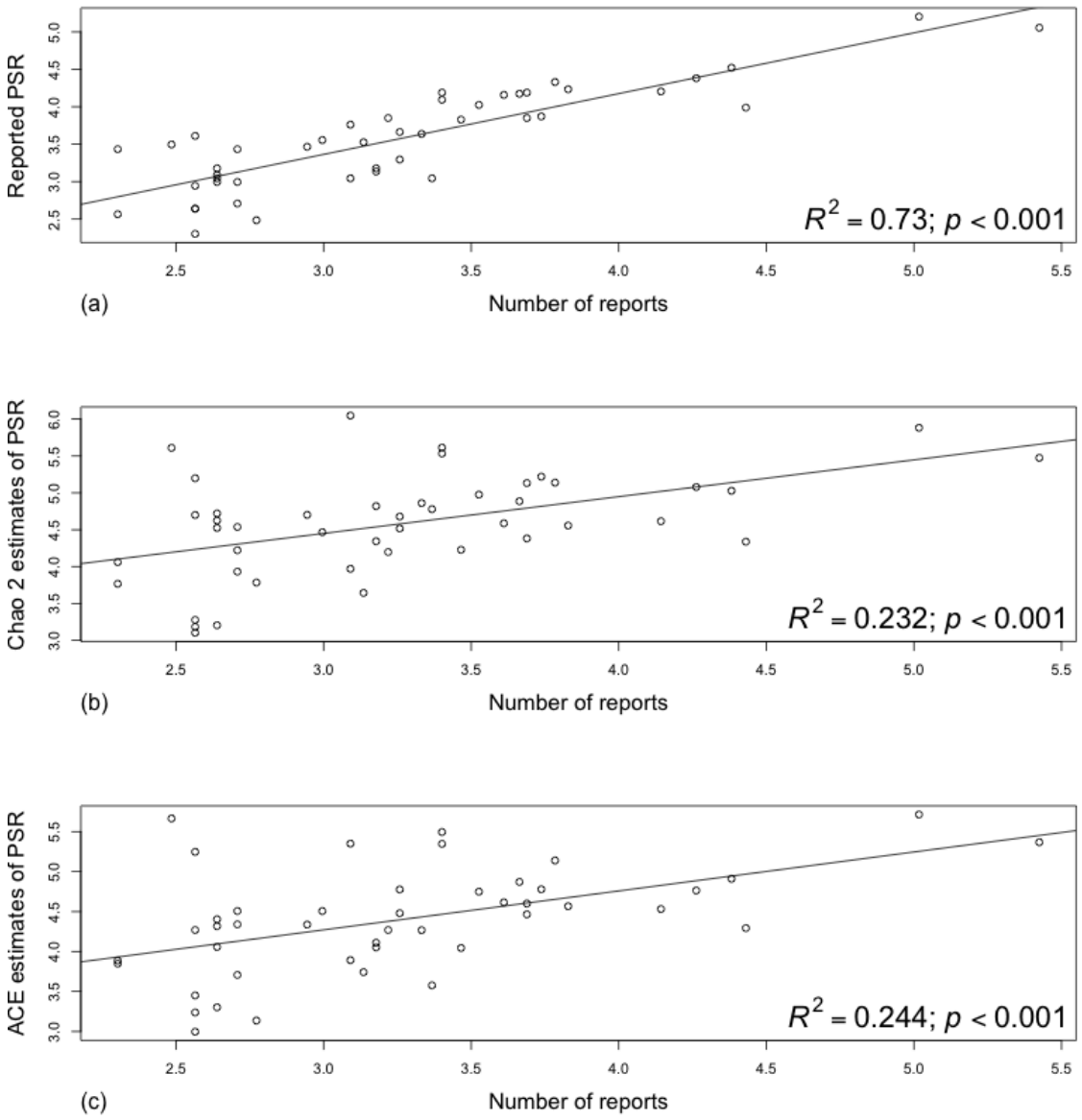


Figure S2 Relationships between the number of reports on each host species and PSR reported (a) or estimated using the Chao 2 (b) and the ACE (c) methods. All variables are log transformed.

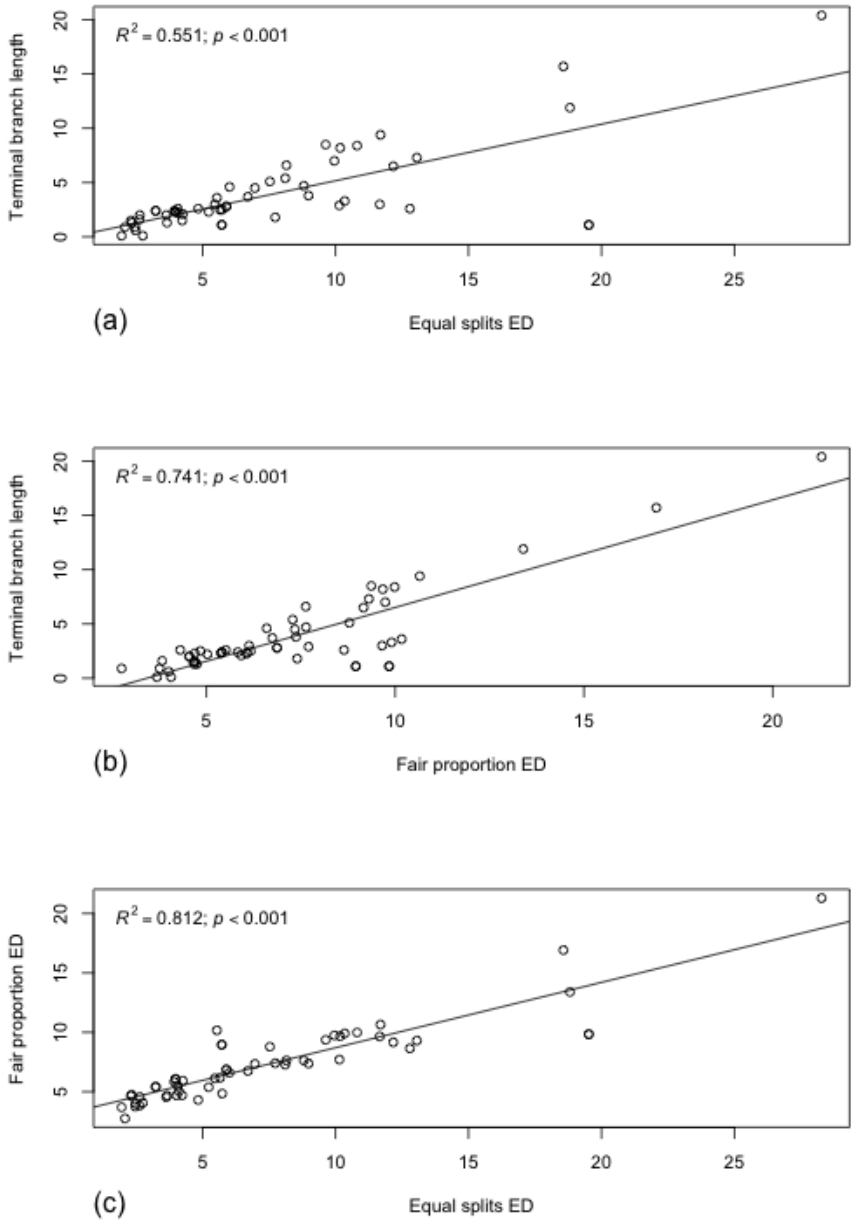


Figure S3 Relationships between the three measures of evolutionary distinctiveness (ED) included in our analyses.

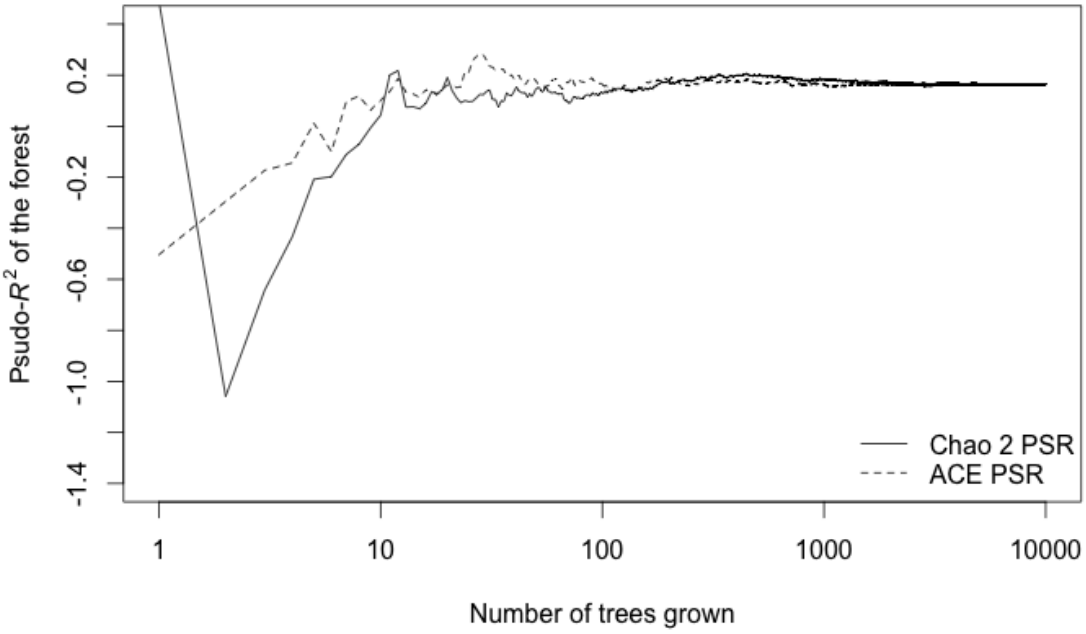


Figure S4 Psudo-R^2 of the random forests for estimated PSR. Each psudo-R^2 value is calculated for the forest comprising the corresponding number of trees that have been grown during the analysis.

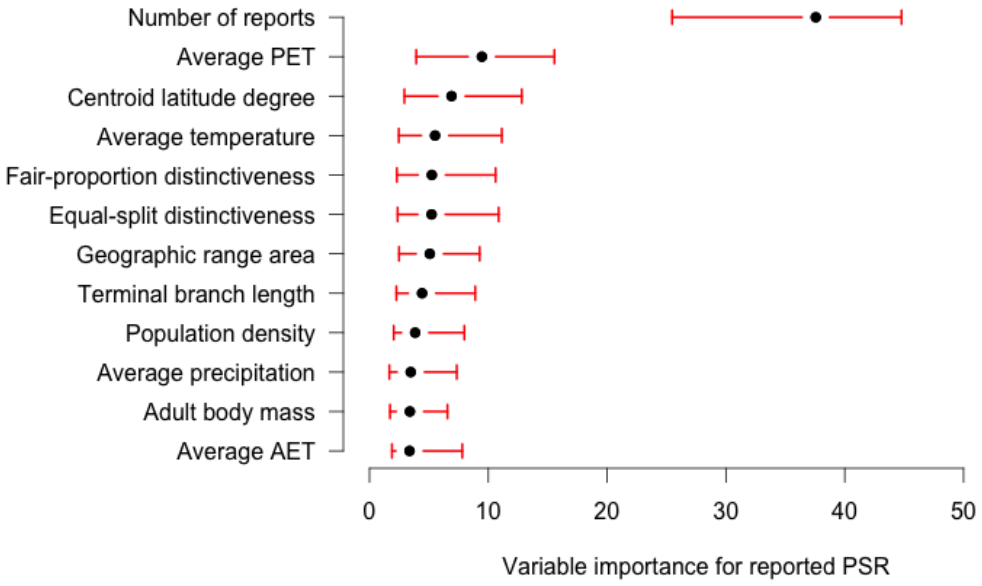


Figure S5 Variable importance determined by the random forest analysis for observed parasite species richness (PSR). The importance of each variable is quantified as the increase of node impurity (residual sum of squares) if the variable is excluded, where a high increase indicates high importance of the variable. The 90% confidence intervals (indicated by error bars) are generated from bootstrapped data.

APPENDIX D

SUPPORTING INFORMATION FOR CHAPTER 5

Table S1 The Spearman's rank correlations between parasite assemblage similarity and different measures of host evolutionary distance, geographic overlap, and ecological similarity.

	J		**CJ**	
Explanatory variables	**ρ**	**p-value**	**ρ**	**p-value**
Time since divergence	-0.177	<0.001	-0.220	<0.001
Number of divergence events	-0.245	<0.001	-0.347	<0.001
Genetic distance	-0.122	<0.001	-0.149	<0.001
Geographic overlap	0.100	<0.001	0.162	<0.001
Geographic overlap proportion	0.102	<0.001	0.161	<0.001
Adult body mass difference	-0.171	<0.001	-0.253	<0.001
Ecological dissimilarity	-0.246	<0.001	-0.312	<0.001

Table S2 The Spearman's rank correlations between parasite assemblage dissimilarity of different taxonomic groups and different measures of host evolutionary distance, geographic overlap, and ecological similarity. The strongest correlations for each group are shown in bold. The sample size (number of host pair-wise combinations) is indicated for each group of parasites.

		J		CJ	
Parasite type	**Explanatory variable**	**ρ**	**p-value**	**ρ**	**p-value**
Arthropod	*Time since divergence*	-0.083	0.232	-0.177	<0.001
(n = 208)	*Number of divergence events*	-0.102	0.141	-0.163	0.001
	Genetic distance	-0.032	0.645	**-0.196**	**<0.001**
	Geographic overlap	0.274	<0.001	0.193	<0.001
	Geographic overlap proportion	**0.282**	**<0.001**	0.195	<0.001
	Adult body mass difference	0.113	0.116	0.110	0.034
	Ecological dissimilarity	-0.073	0.297	0.092	0.072
Bacteria	*Time since divergence*	-0.008	0.876	-0.168	<0.001
(n = 428)	***Number of divergence events***	-0.027	0.581	**-0.228**	**<0.001**
	Genetic distance	-0.021	0.659	-0.125	<0.001
	Geographic overlap	-0.001	0.987	0.034	0.394
	Geographic overlap proportion	-0.003	0.944	0.034	0.401
	Adult body mass difference	-0.075	0.124	-0.145	<0.001
	Ecological dissimilarity	**-0.179**	**<0.001**	-0.151	<0.001
Helminth	***Time since divergence***	**-0.268**	**<0.001**	-0.322	<0.001
(n = 703)	***Number of divergence events***	-0.261	<0.001	**-0.441**	**<0.001**
	Genetic distance	-0.263	<0.001	-0.263	<0.001
	Geographic overlap	0.199	<0.001	0.213	<0.001
	Geographic overlap proportion	0.204	<0.001	0.217	<0.001
	Adult body mass difference	-0.176	<0.001	-0.102	0.009
	Ecological dissimilarity	-0.236	<0.001	-0.143	<0.001
Protozoa	*Time since divergence*	-0.133	<0.001	-0.092	0.002
(n = 822)	*Number of divergence events*	-0.091	<0.001	-0.091	0.003
	Genetic distance	-0.164	<0.001	**-0.123**	**<0.001**
	Geographic overlap	0.007	0.852	0.019	0.527
	Geographic overlap proportion	0.007	0.835	0.019	0.536
	Adult body mass difference	0.152	<0.001	0.021	0.486
	Ecological dissimilarity	**0.207**	**<0.001**	0.029	0.328
Virus	*Time since divergence*	-0.344	<0.001	-0.292	<0.001
(n = 1047)	*Number of divergence events*	-0.276	<0.001	-0.204	<0.001
	Genetic distance	-0.130	<0.001	-0.120	<0.001
	Geographic overlap	-0.002	0.938	0.088	0.004
	Geographic overlap proportion	0.001	0.983	0.085	0.006
	Adult body mass difference	-0.260	<0.001	-0.221	<0.001
	Ecological dissimilarity	**-0.345**	**<0.001**	**-0.326**	**<0.001**

Table S3 Generalized linear models for predicting parasite assemblage corrected Jaccard index (CJ) of different parasite taxonomic groups between carnivore host pairs. Pseudo-R2 is calculated as 1-(residual deviance / null deviance). To avoid zeros in the data, we add 1 to all the variables before log transformation.

		All hosts			***Overlapped hosts***		
Response variable	**Variables**	**Coefficient**	***p*-value**	**Pseudo-R^2**	**Coefficient**	***p*-value**	**Pseudo-R^2**
Arthropod log CJ	*Number of divergence events*	-0.005 ± 0.001	0.011	0.061	-0.006 ± 0.003	0.030	0.059
	Log Geographic overlap	0.061 ± 0.011	< 0.001		0.006 ± 0.003	0.026	
Bacteria log CJ	*Number of divergence events*	-0.007 ± 0.001	< 0.001	0.062	-0.008 ± 0.002	< 0.001	0.102
	log Ecological similarity	-0.058 ± 0.023	0.013		-0.090 ± 0.034	0.008	
Helminth log CJ	*Number of divergence events*	-0.011 ± 0.001	< 0.001	0.186	-0.011 ± 0.001	< 0.001	0.238
	log Geographic overlap	0.003 ± 0.001	< 0.001		0.005 ± 0.001	< 0.001	
Protozoa log CJ	*Genetic distance*	-0.307 ± 0.075	< 0.001	0.015	-0.367 ± 0.122	0.003	0.023
Virus log CJ	*Time since divergence*	-0.004 ± 0.0004	< 0.001	0.172	-0.001 ± 0.0003	< 0.001	0.139
	Genetic distance	0.961 ± 0.175	< 0.001		Not included	N/A	
	log Ecological similarity	-0.102 ± 0.015	< 0.001		-0.096 ± 0.026	< 0.001	

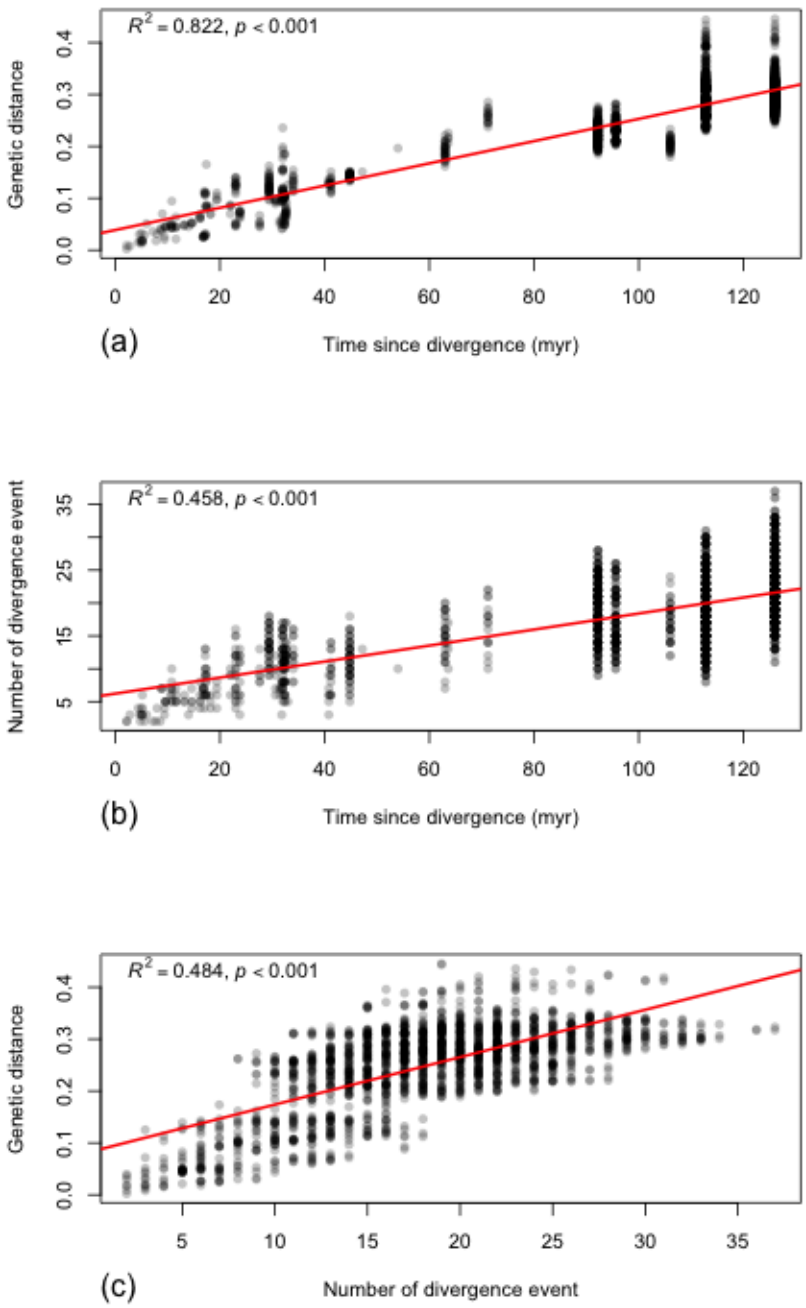


Figure S1 Relationships between the three measures of pair-wise host host phylogenetic distances.

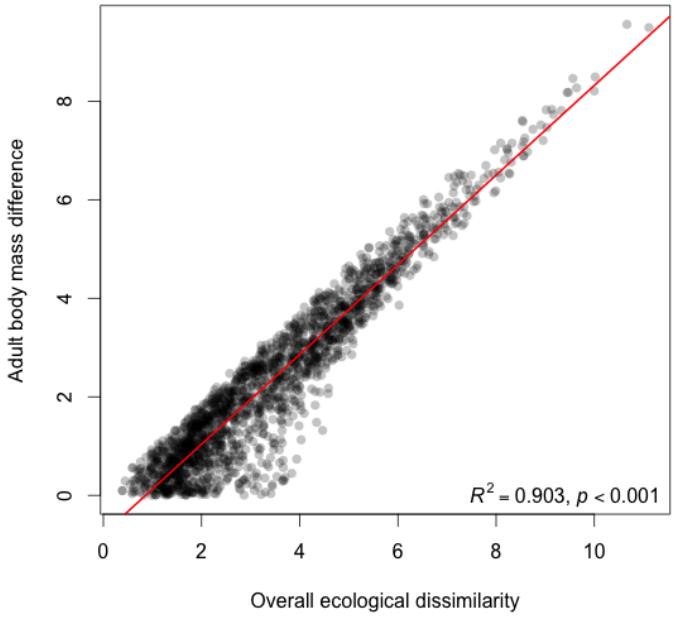


Figure S2 The relationship between adult body mass (g) difference (after log transformation) and overall ecological dissimilarity for carnivore host species pairs.

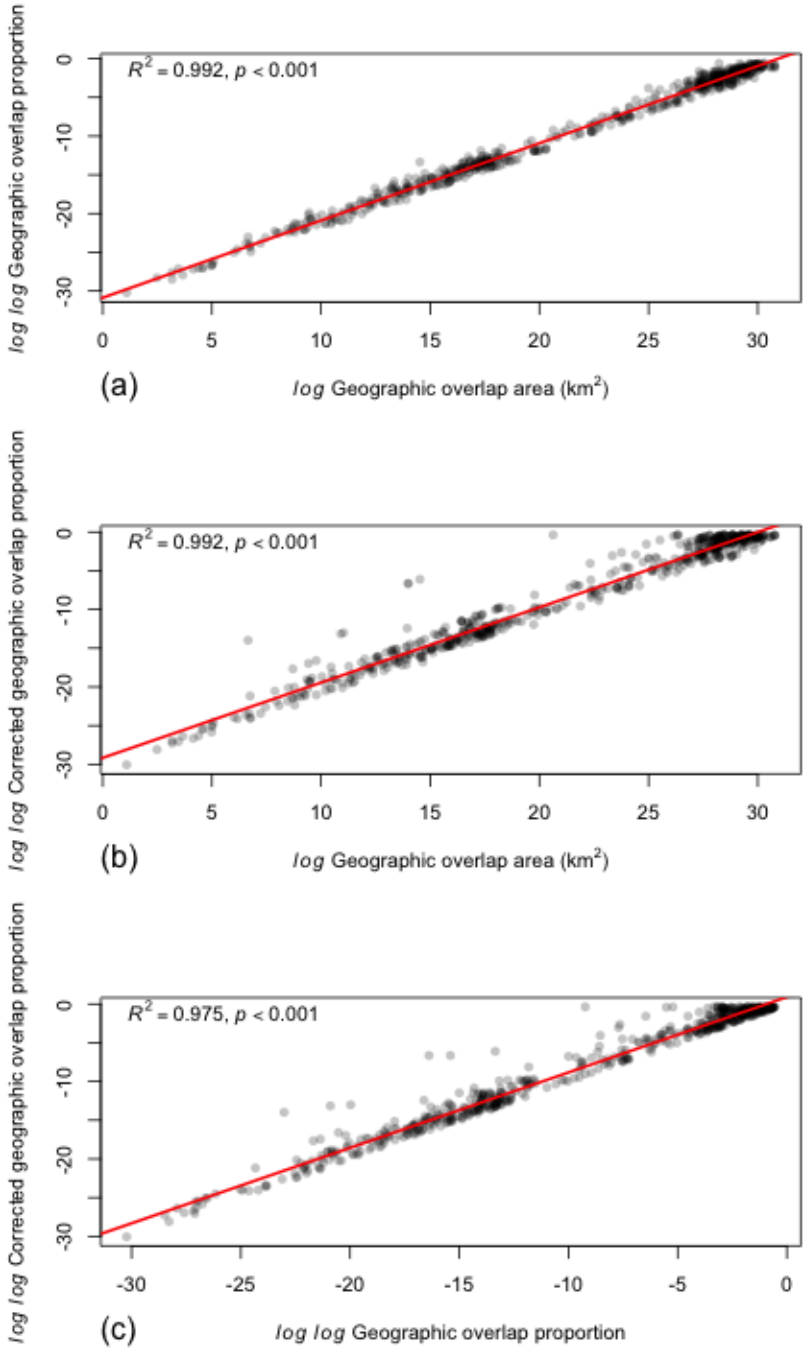


Figure S3 The relationships between our three measures of pair-wise host geographic range overlaps.

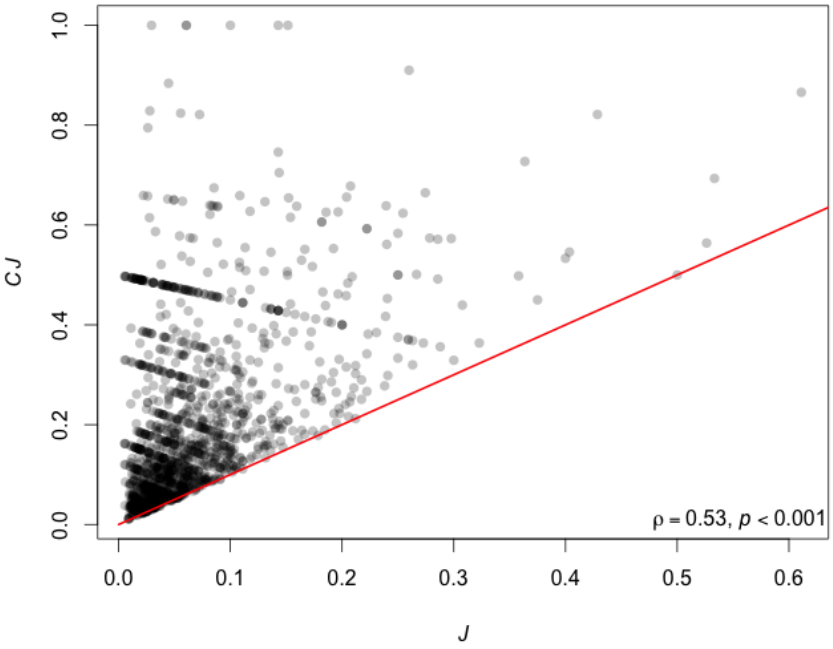


Figure S4 Relationships between the Jaccard index (*J*) and corrected Jacchard index (*CJ*) of parasite assemblages between pairs of carnivore host species, in comparison with a one-to-one line (red).

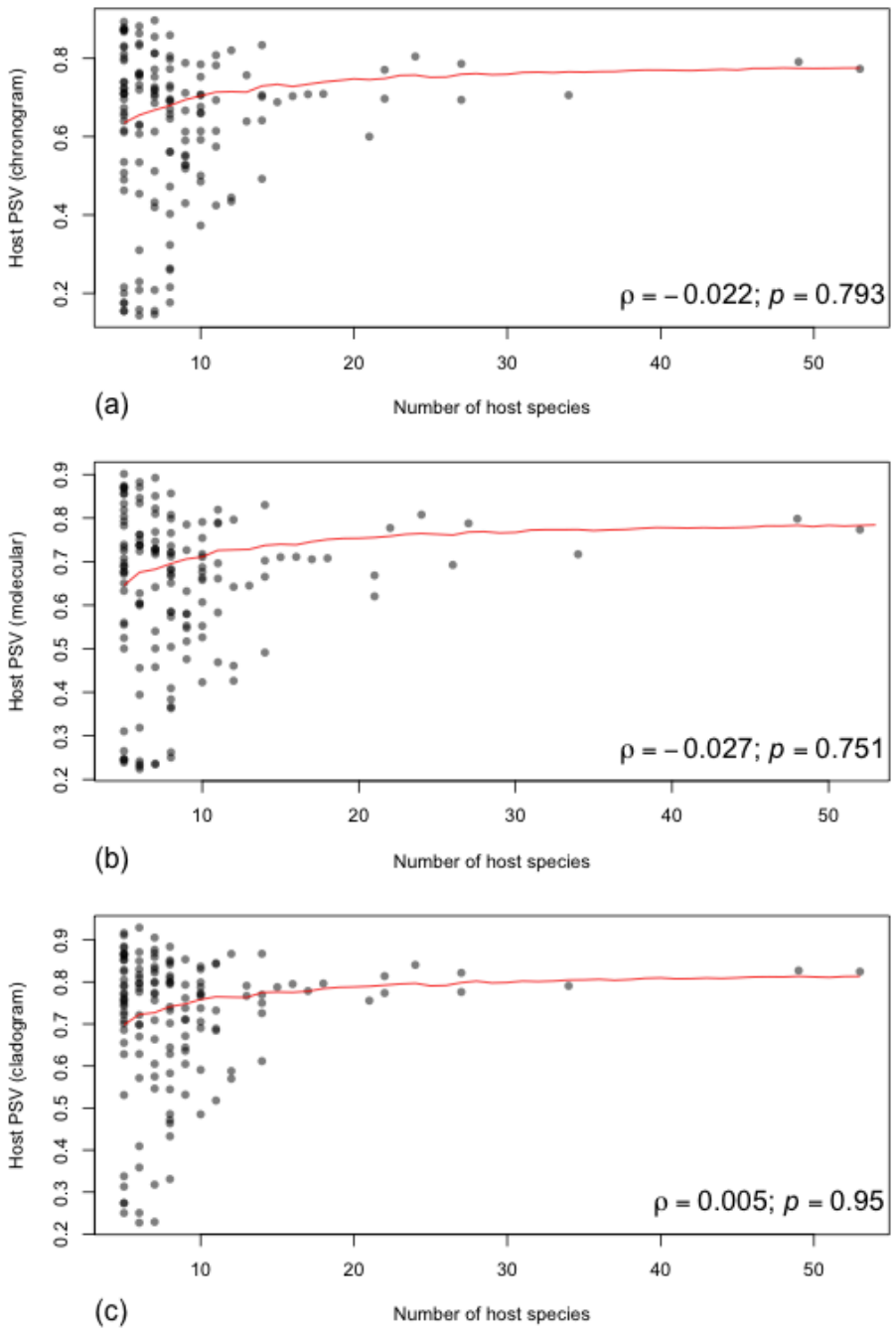


Figure S5 Relationships between host species number and host PSV calculated based on three different host phylogenetic trees: (a) a chronogram of host evolutionary history, (b) a molecular tree, and (c) a cladogram of host evolutionary history. The Spearman ranking-order correlations between the number of host species and host PSV were indicated. The red line shows the lower 5% quantile of PSV calculated from randomly selected 5 to 53 host species (1000 simulation for each level of host numbers), therefore, the points below the red line indicate parasite species whose host ranges are significantly constrained by host phylogeny.

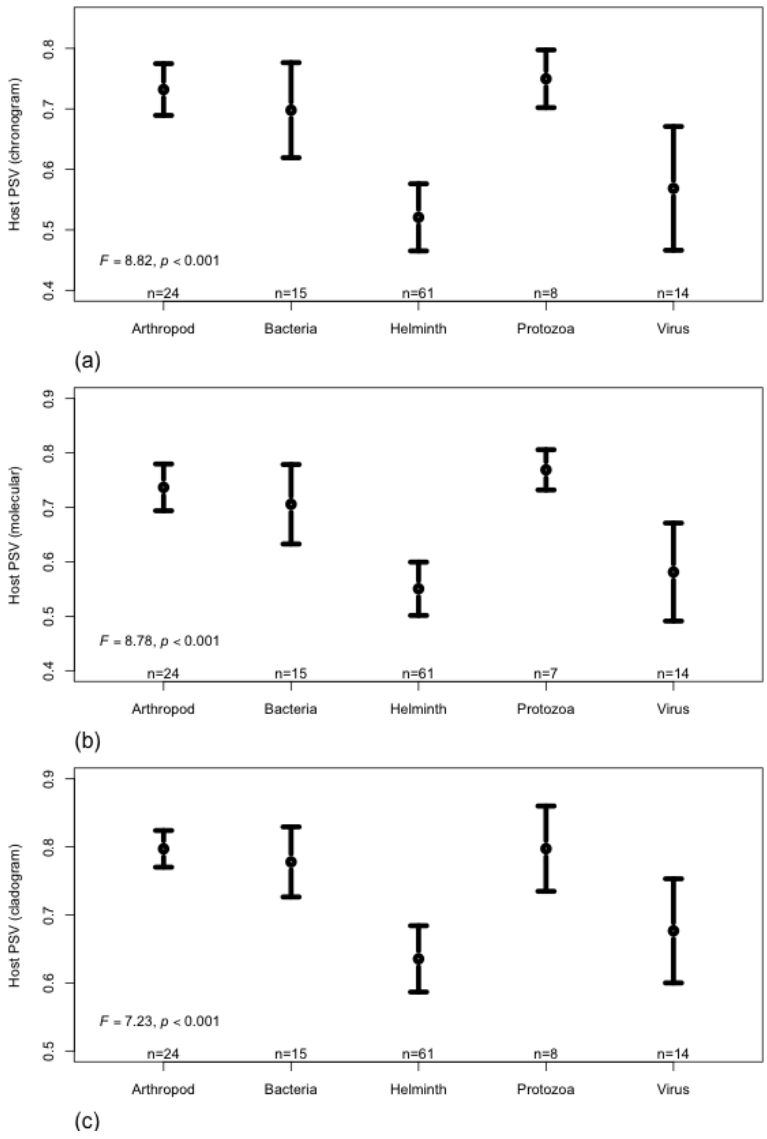


Figure S6 Mean host PSV in different groups of parasites. Host PSV were calculated based on three different host phylogenetic trees: (a) a chronogram of host evolutionary history, (b) a molecular tree representing host genetic distance, and (c) a cladogram of host evolutionary history. The 95% confidence intervals of the means are indicated by the error bars, and the number of parasite species in each group is shown below the bars.