

**For: American Naturalist**

**Correspondence to:**

Thiago Fernando Rangel  
Departamento de Ecologia, ICB  
Universidade Federal de Goiás  
Cx.P: 131, 74001-970  
Goiânia, Goiás, Brasil  
thiagorangel@icb.ufg.br

Article:

**Phylogenetic Uncertainty in Macroecological Analyses**

Thiago F. Rangel <sup>1</sup>

Robert K. Colwell <sup>2</sup>

Gary R. Graves <sup>3</sup>

Carsten Rahbek <sup>4</sup>

José Alexandre F. Diniz-Filho <sup>1</sup>

1: Departamento de Ecologia, ICB, Universidade Federal de Goiás, CP 131, 74.001-970, Goiânia, GO, Brasil. TFR: thiagorangel@icb.ufg.br. JAFD-F: diniz@icb.ufg.br.

2: Department of Ecology and Evolutionary Biology, University of Connecticut, 75 North Eagleville Rd., Unit 3043, 06269-3043, Storrs, CT, United States. robert.colwell@uconn.edu

3: Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, DC 20013, United States. gravesg@si.edu

4: Center for Macroecology, Evolution and Climate, Department of Biology, University of Copenhagen, Universitetsparken 15, 2100, Copenhagen O, Denmark. crahbek@bio.ku.dk

**Key words:** autocorrelation, comparative methods, phylogenetic diversity, hummingbirds

**Running title:** Phylogenetic Uncertainty in Macroecology

Items to be included in the Expanded Online Edition:

**Online Appendix A:** References used in the compilation of maximum elevational range limit of the 304 species of hummingbirds

## Abstract

Ecologists usually rely on a single phylogenetic tree to study evolutionary processes that affect macroecological patterns. This approach ignores the fact that phylogenetic trees often leave out many extant species, include polytomies, and have involved substantial uncertainty in the choice among the best candidate trees. When the phylogenetic uncertainty hidden behind incomplete consensus trees is ignored, statistical inference and biological interpretation in macroecological analyses may be seriously distorted. We propose a new approach that maximizes the use of incomplete phylogenetic information in macroecological studies, while accounting for phylogenetic uncertainty in statistical inference. We illustrate the above issues with a clade-wide analysis of the hummingbirds showing how incomplete phylogenies may seriously distort statistical inference and biological interpretation in macroecological analyses, when phylogenetic uncertainty is ignored. When phylogenetic uncertainty is accounted for, we found no statistical relationship between body mass and maximum elevational and latitudinal range limits for 304 species of hummingbirds. We also showed that, at broad geographic scales, the phylogenetic structure of hummingbird assemblages responds linearly to temperature. Although no statistical approximation will ever substitute for a complete and robust phylogeny, our method enables ecologists to broaden the number of clades for which evolutionary studies are possible and allows the estimation and control of phylogenetic uncertainty.

## Introduction

20 Evolutionary hypotheses that seek to explain contemporary macroecological patterns lie at the heart of important current scientific controversies (e.g. Davies et al. 2005; Jablonski 2006; Arita and Vazquez-Dominguez 2008). Although macroecologists increasingly aim to take phylogeny into account in testing ecological hypotheses (Westoby 2006), phylogenetic uncertainty is routinely ignored. Meanwhile, systematists and evolutionary biologists have incorporated estimates of uncertainty in the inference of ancestral states, (Ronquist 2004), tree topology, and branch lengths (Holder and Lewis 2003), providing robust methods for reconstructing phylogenetic trees and for studying character evolution. In short, although dealing with phylogenetic uncertainty is now common practice in evolutionary biology, ecologists continue to rely on a single consensus tree (e.g. Graham et al. 2009) to study evolutionary processes that affect macroecological patterns.

30       Why does it matter? First, although comprehensive phylogenies are always the goal, phylogenetic projects generally begin with the aim of understanding the deeper structure of clades, and thus sample one or a few species from major lineages (e.g. Bleiweiss et al. 1997; McGuire et al. 2007). This is certainly a worthy objective for evolutionary questions, but is inadequate for an ecologist interested in local assemblage structure or subtle macroecological questions requiring the largest possible sample sizes and maximal use of information, while controlling for phylogenetic signal. Using a major-lineage tree to account for phylogenetic non-independence in ecological work that involves species not placed on the tree is paradoxical, as it may fail to remove the statistical non-independence among species. Second, a consensus tree, which may well include polytomies and poorly supported nodes, hides the phylogenetic uncertainty in estimating the true tree even if all  
40 species have been included in a phylogenetic study.

In this paper, we present a method, designed for macroecological applications, that relies on a resampling procedure to estimate variance due to phylogenetic uncertainty. When a partial

empirical phylogeny is available as a framework, our method permits the inclusion of all species known to belong to a clade, whether included in the empirical phylogeny or not. The stochastic placement of phylogenetically uncertain species on the empirical phylogeny also allows assessment of uncertainty due to any polytomies that it includes. Although the randomization procedure we propose is based on merging multiple sources of phylogenetic information (e.g. molecular studies, taxonomy, hybridization), it is also useful when a number of competing phylogenetic hypotheses are available (Housworth and Martins 2001). We contend that any ecological study that tests

50 evolutionary hypotheses should not only incorporate species missing from the best empirical phylogeny, using all the available phylogenetic information, but should also include all the best candidate trees estimated by the phylogenetic methods.

#### *The method: Background and strategy*

Macroecology seeks to understand processes underlying observed patterns resulting from ecological and evolutionary dynamics at broad spatial and temporal scales. The underlying processes tend to occur in an autocorrelated mode across these dimensions. For example, species cannot be considered to be statistically independent because their traits evolved from common ancestors, thereby generating phylogenetic signal (Felsenstein 1985; Harvey and Pagel 1991). Because

60 ecological assemblages represent the outcome of dynamic interactions between environment and evolution, environmental factors not only drive trait evolution, but also trigger species sorting. Thus, patterns of species abundance, range sizes and species richness are usually autocorrelated as well (Gaston 2003, Borregaard and Rahbek 2010), especially at larger geographical scales. Untangling the factors that drive autocorrelation in macroecological patterns provides important clues on how ecological and evolutionary processes interact (Legendre 1993).

Initial applications of phylogenetic comparative methods in macroecology focused on the evaluation of phylogenetic signal in data (Freckleton et al. 2002) or, more frequently, on the non-

independence of species in reference to correlated macroecological traits (Blackburn and Gaston 1998; Blackburn 2004). However, phylogenetic comparative methods are now being used to deal  
70 with more complex macroecological questions, including the evolutionary processes driving geographical patterns of biodiversity (Fjelds  and Rahbek 2006; Diniz-Filho et al. 2007, 2009; Purvis 2008). For instance, niche conservatism (Peterson et al. 1999; Wiens and Donoghue 2004) is now an important component of many macroecological models designed to explain geographical range dynamics and diversity patterns, and has been invoked as a theoretical framework for coupling geographical and phylogenetic patterns (Hawkins et al. 2006, 2007; Wiens et al. 2006; Rangel et al. 2007; Colwell and Rangel 2010). Although the conceptual link between these patterns requires further development, it is usually recognized that niche evolution can be directly associated with the evolution of particular traits, which in turn mediate (both as cause and consequence) geographical ranges and, consequently, the composition and richness of local assemblages across  
80 environmental gradients in space (Webb 2000; Webb et al. 2002; Davies et al. 2007) and time (Colwell and Rangel 2010).

Macroecological studies usually involve hundreds of species, ranging from deep time to recent evolutionary histories, from locally endemic species to those with worldwide distributions. Although there are now data on the generalized distribution of most groups of terrestrial vertebrates, species sampling for phylogenetic analysis is rarely complete for larger clades. Even when a phylogeny is available for a macroecological analysis, phylogenetic uncertainty is seldom accounted for in macroecological analyses. This uncertainty may arise from three distinct sources: (1) weak empirical support for hypothesized relationships among species in a given clade, (2) errors associated with tree topology and branch lengths, and (3) incomplete and unrepresentative sampling  
90 of known species.

Macroecologists have dealt with incomplete phylogenies in several ways. The first approach is to focus only on clades for which complete phylogenies are available, but this strategy restricts

macroecological studies to a very small number of groups (Pagel 1999) and can undermine assemblage-level studies, if only certain species can be included. Assembling supertrees from smaller, overlapping trees or based on taxonomic classification (Bininda-Emonds et al. 2002; Bininda-Emonds 2004) can fill gaps and increase the number of groups with relatively complete phylogenies. Nevertheless, even for the best studied taxa, such as mammals and birds, supertrees that span all species are relatively recent achievements (e.g. Bininda-Emonds et al. 1999; Beck et al. 2006) that are still not fully resolved.

100        A second and more radical strategy is to ignore the species that are absent from the available phylogeny under the assumption that the species included in the analysis represent an unbiased and representative sample of all species in clade. Of course, the full evolutionary history of a clade of substantial age can be described only by a phylogeny with all species, including extinct ones, although this is generally only achievable in simulated scenarios (Colwell and Rangel 2010). Estimating the degree of bias due to missing species can, in principle, be done by replicating the analysis with random sub-samples of species that are present in the phylogeny (cross-validation), but unfortunately this is not a common practice in macroevolutionary and macroecological studies.

110        Martins (1996) proposed a way to carry out phylogenetic comparative studies when the phylogenetic relationships among species are unknown. In her method, a large sample of trees is randomly generated using models of phenotypic evolution and diversification rates. Species traits are then analyzed on each of the possible trees. The mean of the squared standard error of the calculated evolutionary statistic (e.g. the correlation between two species' traits), known as  $V_s$ , estimates the true variance of the statistic (due to sample size and deviations from estimated to observed data). The variance of the evolutionary statistic calculated among the randomly generated trees, known as  $V_p$ , estimates the variance due to phylogenetic uncertainty. Abouheif (1998) warned that Martins' (1996) approach to trait correlation using fully random phylogenies converges on non-phylogenetic correlations, because of the simple, unconstrained topology of the random trees.

Housworth and Martins (2001) later adapted Martins' (1996) procedure to estimate phylogenetic uncertainty by generating samples of random trees constrained by the topology of partially specified  
120 phylogenies, including taxonomic polytomies.

Here we show how incomplete phylogenies may seriously distort statistical inference and biological interpretation in macroecological analyses, when phylogenetic uncertainty is ignored. We present a way to measure phylogenetic uncertainty when partially known phylogenies are assembled into supertrees using information from taxonomic classification or biological criteria to constrain the random placement of missing species. Based on this method, we propose an adaptation of Housworth and Martins' (2001) approach to maximize the use of incomplete phylogenetic information in macroecological studies, while accounting for phylogenetic uncertainty in statistical inference.

130 *Two examples with hummingbirds: body mass and range limits, and the phylogenetic composition of assemblages*

We illustrate our method by applying it to macroecological hypotheses for the hummingbirds (Trochilidae), a large, monophyletic clade (~330 species, the second-largest family of birds) with a rich natural history literature and ongoing phylogenetic, morphological, behavioral, and ecological research. Currently, the most complete, multilocus, molecular phylogeny published for the hummingbirds (McGuire et al. 2007) includes only 146 of those species, but it encompasses 73 of the approximately 104 currently recognized genera (Schuchmann 1999) and spans all the higher-level trochilid diversity. This level of completeness is typical of (or superior to) current phylogenetic knowledge for many comparably diverse taxa, and no doubt shares their biases and  
140 limitations.

With their hovering flight the most costly form of sustained locomotion (in Kcal/g/s) of any vertebrate (Chai and Dudley 1995), hummingbirds operate under multiple tight constraints on body mass, flight mechanics, and energetics. Climatic constraints, particularly temperature, vary geographically across the breeding season range of the family, from the lowland wet tropics to Alaska, Tierra del Fuego, and the high peaks of the Andes. For these tiny birds (median body mass about 4.8 g), the homeotherm bauplan is challenged not just by their large surface-to-volume ratio, nor simply by seasonal cold (for those species that breed outside the tropics), but even by surviving overnight (at all latitudes) (Carpenter and Hixon 1988; Hiebert 1990). Not only temperature, but also air density and the partial pressure of oxygen decline with elevation, posing a triple challenge for high elevation species (Altshuler et al. 2004a, 2004b), yet hummingbirds are abundant and diverse at high elevations in the Andes (Rahbek and Graves 2000) and throughout the Americas. A complex pattern of sexual body mass dimorphism, probably driven by balance between natural and sexual selection on males, is superimposed on and interacts with other factors affecting body mass evolution (Colwell 2000).

To demonstrate the utility of our approach to account for phylogenetic uncertainty in macroecological studies of trait correlations that rely upon phylogenies, we focus on geographical patterns of hummingbird body mass, in relation to environmental temperature and the elevational and latitudinal range limits of species. To the degree that each species' geographical distribution is an expression of its ecological niche (Hutchinson's duality, *sensu* Colwell and Rangel 2009), the evolutionary dynamics of niches (Wiens and Donoghue 2004; Pearmann et al. 2008) can be studied through patterns of species' range limits. In the case of elevational range limits of hummingbirds, a direct link between niche and morphology has long been recognized. On biomechanical grounds, Feinsinger et al. (1979) predicted that the upper and lower elevational range limit for a hummingbird species with a particular wing-disc loading (indexed by the ratio of wing length, squared, to body mass) should be limited by the minimum and maximum power output it can



achieve for hovering flight, independent of elevation. Thus, the decline in air density with increasing elevation would be offset by compensatory changes in wing-disc loading, which must therefore decline with elevation (Rand 1936; Mayr 1963). The predicted among-species elevational gradient in wing-disc loading could be accomplished, in principle, either by declining body mass with elevation (with isometry) or by constant or increasing body mass with elevation, with allometric increases in wing length. The morphological and distributional data for 38 Andean species available to Feinsinger et al. (1979) offered strong support for both the power output and wing-disc loading predictions, but found only a weak ( $r = 0.08$ ) non-significant ( $P > 0.05$ ) relation between body mass and elevation.

Neither a phylogeny nor statistical methods for correcting for non-independence were available 1979, but Altshuler et al. (2004a) confirmed the predictions and conclusions of Feinsinger et al. (1979) regarding power output and wing-disc loading, based on laboratory experiments and phylogenetically independent contrasts for 43 Andean species. Like Feinsinger et al. (1979), they also detected a weak ( $r^2 = 0.10$ ) but in this case, significant ( $P < 0.05$ , using independent contrasts) correlation between body mass and elevation along a 4,000 m elevational transect. They hypothesized that greater body mass is an adaptation to lower temperatures at higher elevations, despite requiring more extreme wing allometry and despite imposing an even greater challenge for hovering in the lower density air and low oxygen partial pressure of high elevations. We follow up here on this important but unresolved question of conflicting selective demands, with an analysis of body mass in relation to elevation and latitude for 304 species, throughout the Americas.

When multiple lineages overlap geographically and conserve different ecological traits, community structure may be constrained by limiting similarity, so that representatives of different lineages are more likely coexist locally than members of the same clade (Webb 2000). In contrast, after recent speciation events or when novel adaptations are required for life in more challenging environments, successful lineages may be overrepresented locally (e.g. Webb et al. 2002; Graham

et al. 2009). How can we distinguish these two kinds of patterns on broad geographical scales?

Measures of the phylogenetic structure of assemblages that allow inclusion of all local members of a clade are clearly more appropriate for such questions than incomplete empirical phylogenies.

Extreme environmental conditions at higher elevations (lower temperature, lower air density, and lower oxygen levels) or at higher latitudes (colder mean annual temperature, stronger daily and seasonal variation in temperature) may filter out hummingbird lineages that are unable to tolerate those conditions, whereas at lower elevations and latitudes, where environmental conditions are less extreme, the phylogenetic structure of assemblages may be more likely to be driven by species interactions (Darwin 1859; Graves 1988; Terborgh 1971). As a second example of the application of our methods, we examine the phylogenetic structure of hummingbird assemblages, on a broad geographical scale.

## Materials and Methods

### *Quantifying phylogenetic uncertainty*

Our method quantifies the phylogenetic uncertainty in a clade that can be attributed to three sources: (1) species known to be members of the clade, but not included in a formal phylogenetic hypothesis for the clade; (2) lack of full resolution (polytomies) in the phylogenetic hypothesis; and (3) discrepancies among phylogenetic hypotheses. The McGuire et al. (2007) molecular phylogeny for hummingbirds, for example, encompasses only 146 of the 304 hummingbird species for which we have morphological and biogeographical data (see below), and includes five polytomies.

A simple measure of phylogenetic uncertainty (of the first kind) might be calculated as the ratio between the number of species not included in the phylogeny (*phylogenetically uncertain taxa*, which we will abbreviate PUT) and the total number of species known to belong to the clade (PUT plus PKT, the *phylogenetically known taxa* included in the empirical phylogeny). For the

hummingbirds this ratio would be  $158/304 = 0.52$ . However, this simple measure considers all PUT as phylogenetically equivalent and does not take into account what is already known of the phylogenetic structure of the clade.

A better measure of phylogenetic uncertainty attributable to the missing species would be based on intelligently constructed hypothetical trees that include them. To construct the basic  
 220 framework for such hypothetical trees, we first determine, for each PUT, the *most derived consensus clade* (MDCC) that unequivocally contains each the PUT, using all available biological information and, when necessary, classification based on the best available taxonomy. Once the MDCC has been determined for each PUT, we can compute an index of phylogenetic uncertainty ( $U$ ) as

$$U = \frac{\sum h_{PUT}}{\sum h_{PUT} + \sum l},$$

where  $\sum h_{PUT}$  is the sum of the MDCC node height (distance between the node and the branch tips of the tree) of all PUT, and  $\sum l$  is the sum of lengths of all branches in the empirical phylogeny. In other words, for the purpose of computing this index of uncertainty for a hypothetical tree, we assume that the branch for each PUT arises from the base of its MDCC as a polytomy. To assess

230 how this index of phylogenetic uncertainty is structured within the phylogeny, we apply it to successive sections of the tree, moving from root to tips as in a lineages-through-time plot. For hummingbirds, we relied upon the taxonomic classification of Schuchmann (1999) to designate the most likely MDCC for each PUT.

Phylogenetic uncertainty also tends to be structured in space, because assemblage composition varies spatially in complex ways, and species-rich assemblages or poorly sampled regions may include a disproportionate number of phylogenetically uncertain species. If the geographical distributions of all species are available, our phylogenetic uncertainty index can easily

be calculated for local assemblages of species for any point in space, as we demonstrate for the hummingbirds.

240

*Modeling phylogenetic uncertainty with simulations*

The process of placing the phylogenetically uncertain taxa (PUT) in the molecular phylogenetic tree begins by randomizing the order in which the PUT are to be added to the tree. Next, for each species in the PUT queue, we identify the most derived consensus clade (MDCC) that unequivocally contains it. Within the MDCC, the species is then assigned to a random point on a randomly chosen branch. The branch length for a PUT, once it has been placed in this way, is simply the distance from the attachment point to the tip of its own branch. Once a PUT has been inserted, its own branch may serve as a potential insertion point for subsequent species assigned from the PUT queue. The algorithm iterates until each PUT has been added to the appropriate MDCC, producing a complete and fully-resolved (without polytomies) phylogeny. Multiple  
250 alternative phylogenies built with this strategy can be used for any objective that relies on phylogeny (e.g., to estimate phylogenetic diversity), as long as the number of replicates is sufficient to estimate the error of the phylogenetic measure. Cayuela et al. (2011) have recently used a similar iterative randomization approach to develop a method to account for taxonomic uncertainty (species identities) in analyses of ecological communities.

Even the best molecular phylogenetic hypotheses are not fully resolved, and the McGuire et al. (2007) phylogeny, which has five polytomies, is no exception. Our method treats these as soft polytomies, the result of phylogenetic uncertainty. We transform all these polytomies into dichotomies with the same randomization processes used to assign PUT to the empirical  
260 phylogenetic tree. Thus, for a node with three or more branches, two branches are chosen at random and reassigned to the same node. Each remaining branch from the original polytomy is added sequentially, in random order, to the clade constructed from the former polytomy at a randomly

chosen position along the length of the existing branches. Dichotomous nodes in the original phylogeny are not changed (Housworth and Martins 2001). This process guarantees that the resulting tree is fully resolved and that the phylogenetic uncertainty arising from polytomies is also taken into account when multiple randomized trees are compared.

For the hummingbird example, we used this process to generate 1000 phylogenies for all 304 hummingbird species. We then applied Martins' (1996) approach to compute confidence intervals for the analysis of hummingbird data across the randomly generated phylogenies, taking into account not only the error of the statistic due to sample size and deviations between estimated and observed data ( $V_s$ ), but also the error due to phylogenetic uncertainty ( $V_p$ ). Thus, the total error of the statistics used to calculate confidence intervals is the sum of  $V_p$  and  $V_s$ .

#### *Morphological and biogeographical data for the hummingbird examples*

The hummingbirds (Trochilidae) constitute a single, unambiguously monophyletic clade, with all extant species endemic to the New World (Rahbek and Graves 2000). From the total of about 330 described species, we compiled geographical and morphological data for 304 species. Species endemic to islands (the West Indies and the Juan Fernández Islands) were not included in the analysis. We compiled estimates of average body mass (intersexual mean) for each of the 304 species from Dunning (2007) and Schuchmann (1999). Although large intersexual and geographical variation in hummingbird body size is well documented, the averages used here are useful for broad taxonomic analyses. Based on the literature, we also recorded the maximum known elevational and latitudinal range limits for each species (see Appendix A for references). We extracted distributional data from an updated version (16 July 2010) of the comprehensive database for all land and fresh-water birds that are known to have breeding populations in the Western Hemisphere, compiled by Rahbek and Graves (2000, 2001), and mapped the geographical range of each species on a gridded map at a resolution of  $1^\circ \times 1^\circ$  (latitude-longitude). These maps represent a conservative

extent-of-occurrence estimate of the breeding range based on museum specimens, published sight records, and spatial distribution of habitats between documented records for South America.

290 Finally, for each map cell, we obtained the mean annual temperature from Hijmans et al. (2005).

*Estimating trait correlations in the context of phylogenetic uncertainty*

Correlations between organismal traits or between traits and environmental characteristics are fundamental to the inference of adaptation in macroecology. Because discounting for shared ancestry can substantially diminish sample size and limit inference (Harvey and Pagel 1991), making the most of available information is always advisable—and often critical. Our approach to assessing the role of phylogenetically uncertainty allows maximal use of information when complete phylogenies are unavailable, as we demonstrate with a hummingbird example.

Geographic variation in the size and shape of birds is influenced by abiotic factors,  
300 including temperature, humidity, air density and seasonality (e.g. Graves 1991). As an adaptive response to temperature one might expect that hummingbirds of colder regions (both high elevations and high latitudes) would have, on average, larger body mass than the lowland tropical species, as suggested by Altshuler et al. 2004. Our first step was to describe the phylogenetic signal in body mass, maximum elevational range limit, and maximum latitudinal range limit among hummingbird species. Thus, for each of the 1000 simulated phylogenies, we calculated a phylogenetic correlogram (Gittleman and Kot 1990; Diniz-Filho 2001), which measures the level of concordance between two traits on a phylogeny, for body mass and elevational range limit and body mass and latitudinal range limit. The average Moran's I coefficient among simulated phylogenies estimates the true phylogenetic autocorrelation in the traits, whereas variance of Moran's I among  
310 phylogenies is a measure of phylogenetic uncertainty.

We used Phylogenetic Eigenvector Regression (PVR, Diniz-Filho et al. 1998) to filter the phylogenetic signal in body mass and maximum elevational and latitudinal range limits. PVR begins by extracting eigenvectors from a phylogenetic distance matrix. The variation in a particular trait is then regressed against a select set of these eigenvectors, which is decomposed into a phylogenetic component ( $P$ , the estimated values of PVR), which captures the shared variation of the trait among species due to common ancestry, and a species component ( $S$ , the PVR residuals), which conveys the trait variation that is particular to each species, independent of shared ancestry (Cheverud et al. 1985; Gittleman and Kot 1990). We calculated a matrix of phylogenetic distances among species for each of the 1000 simulated hummingbird phylogenies that combined PKT and PUT, then we extracted a set of 300 eigenvectors from each distance matrix. To assess the role of including the PUT, we repeated this analysis for the PKT alone. We selected those eigenvectors that accounted for most of the phylogenetic signal in body mass (stopping once Moran's  $I$  for the residuals of PVR regression dropped below 0.05; Gittleman and Kot 1990), and separately, those that removed most of the phylogenetic signal for maximum elevational range limit and for maximum latitudinal range limit (Diniz-Filho et al. in press).

The correlation between the phylogenetically independent variation ( $S$ ) in body mass and in maximum elevational (or latitudinal) range limits can be viewed as a partial regression between these two traits, after taking into account the phylogenetic structure of the data. To estimate the phylogenetically independent correlation between each pair of traits and the error due to phylogenetic uncertainty, we replicated this analysis for each of the 1000 simulated phylogenies and for the PKT alone.

#### *Assessing phylogenetic assemblage structure in the context of phylogenetic uncertainty*

To evaluate the phylogenetic structure of hummingbird assemblages at broad geographical scales we calculated the Phylogenetic Species Variability index (PSV, Helmus et al. 2007) for each map

cell, based on the 1000 simulated phylogenies that combined PKT and PUT, as well as for PKT alone. The PSV is an inverse measure of phylogenetic relatedness among the species that comprise an assemblage, ranging from 0 (minimum phylogenetic diversity) to 1 (maximum phylogenetic diversity). To test if observed PSV values were significantly different from the null expectation that  
340 assemblages are neither phylogenetically clustered nor phylogenetically overdispersed, for each phylogeny we calculated 100 PSV values after shuffling species identities, while maintaining unchanged the species richness per cell (“null 1” in Helmus et al. 2007). Following Graham et al. (2009), we used linear correlation to measure the degree to which the spatial patterns in PSV can be explained by mean annual temperature.

## Results

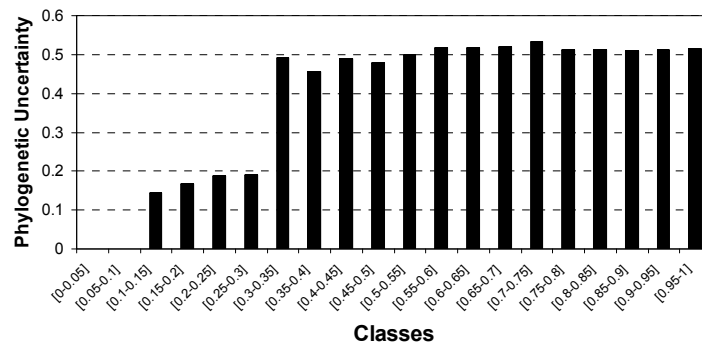
### *Quantifying phylogenetic uncertainty*

If the total length (distance from root to branch tips) of the published phylogeny of hummingbirds is scaled to 1, the sum of branch lengths of all 146 species in the phylogeny (the PKT) equals 56.47  
350 branch units. Thus, if the only phylogenetic information available for the 158 species not included in the empirical phylogeny (the PUT) is that they belong to the trochilid clade, their most distal MDCC would be the root of the phylogeny, and the sum of MDCC node heights for the PUT alone would equal 158 branch units. In this case, the phylogenetic uncertainty index would be equal 0.73, but because the higher-level taxonomy of hummingbirds is now well established (Bleiweiss et al. 1997; McGuire et al. 2007), we were able to assign all 158 PUT to at least a mid-level MDCC. Thus, taking advantage of the biological and taxonomic information compiled by Schuchmann (1999), the sum of MDCC node heights for PUT equals only 35.64 (instead of 158), which gives a phylogenetic uncertainty index of only 0.38 for the entire set of 304 species of hummingbirds. This means that the magnitude of uncertainty in our combined phylogeny that is due to PUT is not



360 greater than 38% of the information in a fully resolved phylogeny of all extant and known hummingbirds, even though the number of PUT is slightly greater than the number of PKT.

Computing the uncertainty index using the number of PUT and PKT in a lineages-through-time context, assuming that each PUT branch extends as a polytomy from the base of its MDCC to the top of the phylogeny, shows that phylogenetic uncertainty in the hummingbirds is located principally from the middle to the top of the phylogeny, which reflects the fact that most of the PUT were assigned to relatively derived MDCCs (Fig. 1).



370 Figure 1: The distribution of phylogenetic uncertainty from the root (left) to the top (right) of the hummingbird tree. Phylogenetic uncertainty is calculated as the sum of most derived consensus clade (MDCC) node heights (distance between each node and the top of the tree) for all phylogenetically uncertain taxa (PUT) in the numerator, divided by that sum plus the total lengths of all branches for the phylogenetically known taxa (PKT) in the denominator.

Phylogenetic uncertainty was also structured in geographical space. Figure 2A shows the map of species richness for all 304 hummingbird species used in this study, figure 2B shows richness of the 156 PUT, and figure 2C shows the uncertainty index calculated for hummingbird assemblages across space. The highest peaks of phylogenetic uncertainty are located in southern Mexico and northern Central America, along the Pacific coast of Ecuador and Peru, the Planalto and Atlantic forest regions of Brazil, and in eastern Amazonia. Notice, however, that those regions

380

of high phylogenetic uncertainty do not correspond exactly to the regions of highest PUT richness (Fig. 2B), as the phylogenetic uncertainty among PUT is uneven.

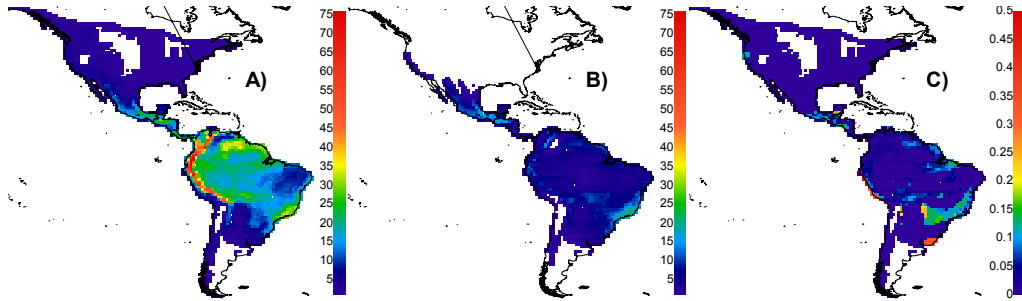


Figure 2: Maps at a scale of  $1^\circ \times 1^\circ$  latitude-longitude of (A) the richness of 304 species of hummingbirds, (B) the richness of 158 species of hummingbirds that are phylogenetically uncertain (PUT), and (C) the phylogenetic uncertainty index calculated for hummingbird assemblages .

### 390 *Estimating trait correlations: Evolution of body mass and range limits*

Phylogenetic autocorrelation in body mass declines with phylogenetic distance (Fig. 3, circles, solid line). For the first distance class (the most closely-related species), the mean Moran's  $I$  equals 0.705, variation of Moran's  $I$  among modeled phylogenies ( $V_p$ ) is 0.007, while the variation within phylogeny due to sampling error ( $V_s$ ) equals 0.016. Both variances decrease towards deep nodes in the phylogeny. Because most PUT are assigned to relatively derived MDCCs, and because the uncertainty accumulates through the phylogeny, short distance classes have larger phylogenetic variances.

Maximum elevational range limit is also structured through the phylogeny (Fig. 3, squares, dashed line). In the first distance class, mean Moran's  $I$  equals 0.473,  $V_s$  is 0.101, while  $V_p$  equals 400 0.165, again indicating that variance due to phylogenetic uncertainty is larger than variance due to sampling error. Large phylogenetic confidence intervals at intermediate distance classes are caused

by polytomies in the empirical phylogeny (McGuire et al. 2006) located at the base of the coquettes and brilliants clades, which together form the Andean clade. Maximum latitudinal range limit is less structured phylogenetically than body mass and maximum elevational range limit, as only the closely related hummingbirds tend to have similar northern and southern range limits.

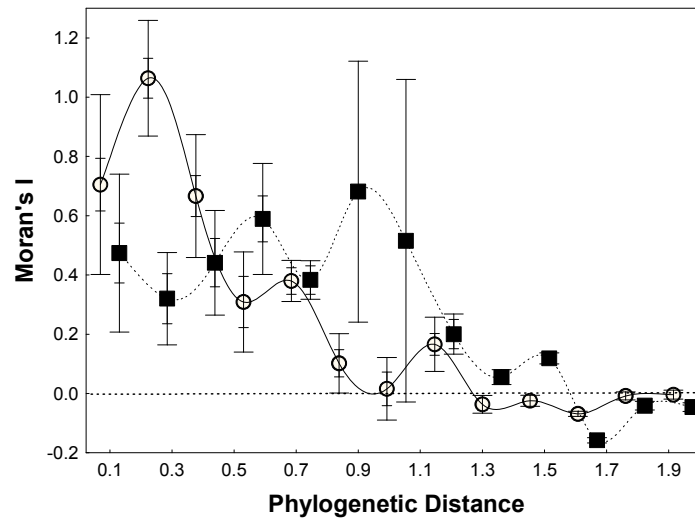


Figure 3: Phylogenetic autocorrelation of mean body mass (open circles, continuous line) and maximum elevational range limit (squares, dashed line) as function of phylogenetic distance. The inner error bars show the standard confidence interval of Moran's I statistic, calculated using  $V_s$ . The outer error bars represent the confidence interval that combines  $V_s$  and  $V_p$ .

When PUT are not taken into account, Pearson's correlation coefficient between the  $S$  components (residuals of the PVR regressions) of body mass and the  $S$  components of maximum elevational range limit was 0.166, with marginal statistical significance ( $P$ -value = 0.045). Permutations of the position of PUT in the phylogeny yielded 1000 values of Pearson's correlations between the same traits, yielding an average correlation across the permutations of  $0.107 \pm 0.068$  (standard deviation). However, the distribution of the permuted values represents only the phylogenetic variance ( $V_p = 0.005$ ), as each correlation coefficient also has its own estimated variance due to sampling error ( $V_s = 0.003$ ). When both sources of variance ( $V_p$  and  $V_s$ ) are included, the confidence interval for the correlation between body mass and elevation ranged from -

0.067 to 0.282, bracketing zero, thus not significant. The correlation between body mass and maximum latitudinal range limit, computed in the same way, was also not statistically significant (average Pearson's  $r = 0.027$ , C.I. 95%:  $-0.124 - 0.169$ ), nor was the correlation between elevational and latitudinal range limits (average Pearson's  $r = 0.046$ , C.I. 95%:  $-0.084 - 0.176$ ). Thus, latitudinal patterns of body size are clearly not confounding the result of primary interest, the relationship between body mass and elevation. We conclude that, when phylogenetic and geographical autocorrelation are controlled, phylogenetic uncertainty is taken into account, and latitude eliminated as a confounding factor, we were unable to reject the null hypothesis of no relationship between body mass and maximum elevational range limit for the 304 hummingbird species, although when only PKT are considered, a marginally significant positive correlation is found. Of course, if there was no phylogenetic uncertainty the confidence interval would be narrower, which might lead to a different conclusion. However, about half (48.9%) of the correlations are not significant in the distribution of 1000 correlation coefficients. Because all correlation coefficients are very low and  $V_s$  is much larger than  $V_p$ , evidence that the relationship between body mass and maximum elevational range limit, if such relationship indeed exists, is certainly very weak, regardless of the uncertainty in the phylogeny.

After accounting for phylogenetic autocorrelation, figure 4A shows that species occurring in the southern Andes have a strong tendency toward larger body mass. This pattern is somewhat influenced by the giant hummingbird (*Patagona gigas*), a phylogenetically isolated species (Altshuler et al. 2004b) with substantially larger body mass than any other species (more than two standard deviations greater than the second largest hummingbird, Colwell 2000).

Figure 4B shows a clear pattern of phylogenetically independent evolution of maximum elevational range limits in high elevation biomes especially in the Andes and in montane Mexico and western North America. On the other hand, hummingbird assemblages in Amazonia and

Atlantic rainforest tend to have lower maximum elevational range limits than expected based their ancestors.

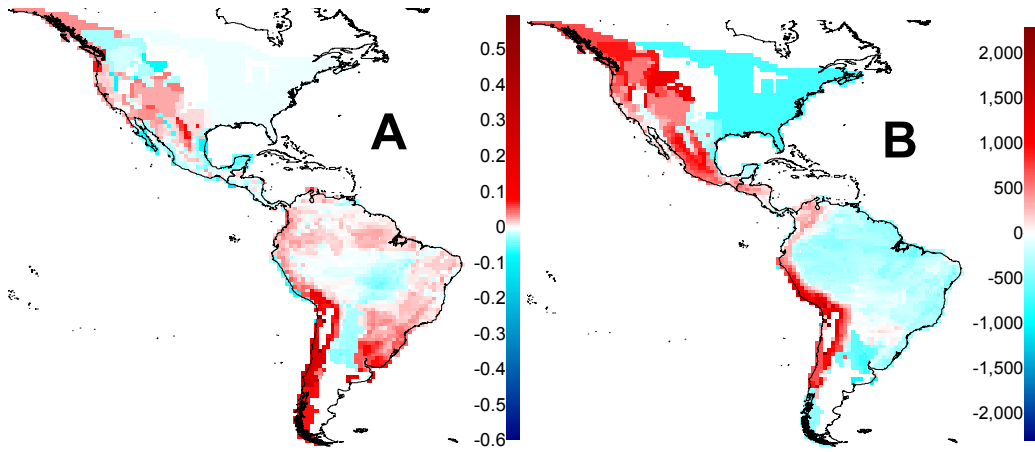


Figure 4: Maps of  $S$  (phylogenetically independent) components of species, indicating (A) the amount of deviation of body mass and (B) maximum elevational range limit from those of their ancestors, averaged among the species that occur in each grid cell. In (A), blue cells indicate species assemblages with smaller and red cells larger body mass than expected based on the distribution of their ancestors. In (B), blue cells indicate species assemblages with lower and red cells higher maximum elevational range limits than expected.

#### *Spatial patterns in phylogenetic assemblage structure*

As expected, ignoring species missing from phlogenies (PUT) strongly affects the analysis and results in studies of phylogenetic assemblage structure (Fig. 5), clearly reflecting geographical patterns of bias in phylogenetic sampling. However, Phylogenetic Species Variability (PSV) proved surprisingly robust to the details of how PUT are assigned to the simulated phylogenies

(phylogenetic uncertainty): the average difference between the lowest and highest PSV values calculated for each cell, for simulated phylogenies, was only 0.006 (potential PSV values range from 0 to 1). Moreover, variation in PSV among simulated phylogenies is not structured in space (Moran's  $I$  of the map of standard deviation of PSV values shows no significant spatial autocorrelation). When all 304 species of hummingbirds are included (PKT and PUT), a clear pattern of phylogenetic overdispersion emerges across the Amazon Basin and Atlantic Rainforest,

with phylogenetic clustering across Central and North America and on the west slope of the Peruvian Andes, reflecting the limited number of major clades that have reached these areas (Fig. 5). However, the phylogenetic clustering in large areas of the Amazon Basin, the Cerrado, and the Atlantic rainforest and the phylogenetic overdispersion in much of Central America are revealed only when both PUT and PKT are included in the analysis. The spatial pattern in the phylogenetic structure of hummingbird assemblages, as measured by PSV using all species, is explained well by a linear model that includes only mean annual temperature and range in elevation ( $r^2 = 0.63$ ), suggesting that phylogenetic structure of assemblages is, at least in part, driven by environmental factors, supporting the findings by Graham et al. (2009).

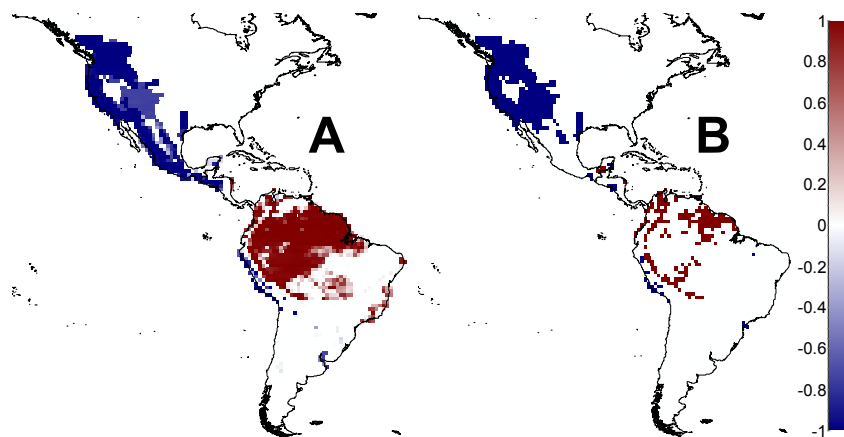


Figure 5: Hemispheric variation in phylogenetic species variability (PSV) for hummingbird assemblages, calculated for (A) 304 species (PKT and PUT) and (B) 149 species (PKT only). Regions where observed PSV values were significantly higher or lower than expected are indicated, by red (phylogenetically overdispersed) or blue (phylogenetically clustered), respectively.

## Discussion

### *Phylogenetic uncertainty and macroecological patterns*

The recent availability of phylogenies for large clades has facilitated the first analyses of geographical patterns in functional morphology, ecology and phylogenetic community structure at continental scales (e.g. Fjeldså 1994; Fjeldså and Rahbek 2006). In addition to the conceptual

difficulty of untangling the ecological and evolutionary processes underlying observed patterns, the study of phylogenetic macroecology poses several methodological challenges. Among the most important is phylogenetic uncertainty caused by ambiguous tree topology and branch lengths and from incomplete phylogenetic sampling. In this study we show how comprehensive, simulated  
490 phylogenies can be applied to macroecological questions using taxonomic and ecological information about the species that are absent from the empirical phylogeny.

Complete and rigorously constructed phylogenies are required to accurately study evolutionary processes, and will never be fully replaced by methods that combine partially known phylogenies with external sources of information about the phylogenetic relationships among species (e.g. taxonomic classification). Although the exact placement of species within an empirical phylogeny is best estimated by molecular studies, uncertainty in the placement of phylogenetically unstudied species can be greatly reduced by incorporating non-molecular information (e.g., morphology and behavior). Moreover, species missing from empirical phylogenies are unlikely to be random. Instead, they are often characterized by small geographical ranges, small populations,  
500 locations that are geographically remote or inaccessible for political reasons, or are classified as threatened or extinct. Sampling bias is probably the most serious source of error in statistical inference, especially when the statistical analysis of the sample, a partially known phylogeny, is used to make inference about ecological or evolutionary processes that affect regional assemblages or the entire clade. Monte Carlo computational procedures (e.g. bootstrapping, jackknifing, cross-validation) offer an alternative way to detect and control possible sampling biases of phylogenetic trees.

510 *Phylogenetic patterns in hummingbird body mass and elevational range limits*

Hummingbird body mass and maximum elevational range limit proved to be phylogenetically highly structured (fig. 3). Closely related species tend to have similar body masses and elevational range limits. These findings provide substantial support for evolutionary conservatism of body mass in hummingbirds, as well as fidelity to ancestral elevational range following allopatric speciation (Graves 1985). Our results are also consistent with previous studies that suggested hummingbirds originated in the tropical lowlands, subsequently invading high elevation Andean and North American habitats (Bleiweiss 1998; McGuire et al. 2007).

Our analyses show that a significant relationship between body mass and elevation, as documented by Altshuler et al. (2004b) for species on a local Andean elevational gradient, fails to  
520 emerge from an analysis that encompasses the entire hummingbird clade over the Americas, even after accounting for a possible confounding effect of latitude. The discrepancy may lie in the fact that high-latitude hummingbirds in North America are exclusively latitudinal migrants, some covering thousands of km on their migration routes, relying in most cases on limited resources along the route, which may favor small body mass, despite the cool or cold nighttime temperatures in their breeding ranges (Schuchmann 1999). Several species are also known to migrate south from the tropics to the Austral temperate zone (Rappole and Schuchmann 2003). Feinsinger et al. (1979) and Alshuler et al. (2004b) suggested that the phylogenetic signal in wing length may be more evolutionarily labile than body mass, and that it shows a stronger relationship with maximum elevational range limit. In this way, lineages with relatively conserved body mass could gain access  
530 to a wide range of elevations through the allometric, adaptive adjustment of wing length.



*Phylogenetic structure of hummingbird communities*

Species-level phylogenies also allow community ecologists to investigate how evolutionary and ecological processes drive assemblage composition. Because many traits are conserved among related species during the diversification of lineages, closely related species tend to be more similar than expected at random. If key traits that define the ecological niche of a species tend to be conserved, competition between closely related species should be more intense, therefore affecting  
540 assemblage composition (Darwin 1859; Lack 1973). In contrast, sites with extreme environmental conditions can potentially select lineages that exhibit rare physiological tolerances or occupy highly differentiated ecological niches, and at these sites we should expect to observe phylogenetic clustering (Brown and Kodric-Brown 1979).

Based on the composition of 189 local hummingbird communities in Ecuador, Graham et al. (2009) found that closely related species tend to coexist only at higher elevations (phylogenetic clustering), whereas communities at lower elevations are usually composed of more distantly related species (phylogenetic overdispersion). This finding was interpreted as evidence that only a few lineages have the physiological, morphological, and behavioral adaptations necessary for flight and thermoregulation at extremely high elevations (above 4000m), whereas at lower elevations,  
550 where physiological constraints to hummingbird survival are less severe, competition among species is perhaps the main driver of assemblage composition (Graham et al. 2009, Parra et al. 2010).

The coarser spatial resolution used in this study did not allow us to detect the same pattern found by Graham et al. (2009) in Ecuador, but there is a clear indication that hummingbird assemblages are phylogenetically clustered in the northern Andes (Chapman 1917; Chapman et al. 1926). Finding phylogenetic clustering in Central and North America is hardly surprising, as many major South American tropical clades have not reached these regions. The finding of overdispersion

at this coarse geographical scale in Amazonia, on the other hand, is noteworthy and has few parallels (but see Gotelli et al. 2010; Fjeldså and Rahbek 2006).

560

### *Concluding Remarks*

The simulation procedure described here maximizes the integration of non-molecular phylogenetic information into incomplete molecular phylogenies, while accounting for phylogenetic uncertainty and sampling bias. This procedure may also be used in statistical tests of ecological and evolutionary hypothesis when multiple candidate phylogenies are available. We also propose an index that estimates the magnitude of uncertainty when combining heterogeneous sources of phylogenetic information. We provide several examples using the largest phylogenetic dataset compiled thus far for hummingbirds, and show how sampling bias may affect inferences about ecological and evolutionary processes. Although this procedure will never substitute for a complete  
570 and robust phylogeny, it enables ecologists to broaden the number of clades for which evolutionary studies are possible and allows the estimation and control of phylogenetic uncertainty.

### **Acknowledgements**

Several concepts in this paper were inspired by the phylogenetics graduate course (EEB5349), taught by Dr. Paul Lewis, at the University of Connecticut. TFR is supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, 564718/2010-6). RKC is supported by the National Science Foundation (USA, DEB-0639979 and DBI-0851245). CR acknowledges the Danish National Research Foundation for support to the Center for Macroecology, Evolution and Climate. JAFD-F has been continuously supported by CNPq through several research grants.

## References

- Abouheif, E. 1998. Random trees and the comparative method: a cautionary tale. *Evolution* 52:1197-1204.
- Altshuler, D., Dudley, R. and McGuire, J. 2004a. Resolution of a paradox: Hummingbird flight at high elevation does not come without a cost. *Proceedings of the National Academy of Sciences of the United States of America* 101:17731-17736.
- Altshuler, D. L., Stiles, F. G. and Dudley, R. 2004b. Of hummingbirds and helicopters: hovering costs, competitive ability, and foraging strategies. *The American Naturalist* 163:16-25.
- Arita, H. and Vazquez-Dominguez, E. 2008. The tropics: cradle, museum or casino? A dynamic null model for latitudinal gradients of species diversity. *Ecology Letters* 11:653-663.
- 590 Beck, R. M. D., Bininda-Emonds, O. R. P., Cardillo, M. Liu, F.-G., R. and Purvis, A. 2006. A higher-level MRP supertree of placental mammals. *Evolutionary Biology* 6:93-107.
- Bininda-Emonds, O. R., Gittleman, J. L. and Purvis, A. 1999. Building large trees by combining phylogenetic information: a complete phylogeny of the extant Carnivora (Mammalia). *Biological Reviews* 74: 143-175.
- Bininda-Emonds, O. R. P. 2004. The evolution of supertrees. *Trends in Ecology and Evolution* 19:315-322.
- Bininda-Emonds, O. R. P., Gittleman, J. L. and Steel, M. A. 2002. The (super) tree of life: procedures, problems and prospects. *Annual Reviews of Ecology and Systematics*. 33:265-289.
- 600 Blackburn, T. M. 2004. Methods in Macroecology. *Basic and Applied Ecology* 5:401-412.
- Blackburn, T. M. and Gaston, K. J. 1998. Some methodological issues in Macroecology. *The American Naturalist* 151: 68-83.
- Bleiweiss R., Kirsch, J. A. W. and Matheus, J. C. 1997. DNA hybridization evidence for the principal lineages of hummingbirds (aves: trochilidae). *Molecular Biology and Evolution* 14:325-243.
- Bleiweiss, R. 1998. Tempo and mode of hummingbird evolution. *Biological Journal of the Linnean Society* 65:63-76.
- Borregaard, M. K. and Rahbek, C. 2010. Causality of the relationship between geographic distribution and species abundance. *The Quarterly Review of Biology* 85:3-25.
- 610 Brown, J. H. and Kodric-Brown, A. 1979. Convergence, competition and mimicry in a temperate community of hummingbird-pollinated flowers. *Ecology* 60:1022-1035.
- Carpenter, F. L. and Hixon, M. A. 1988. A new function for torpor: fat conservation in a wild migrant hummingbird. *The Condor* 90:373-378.
- Cayuela, L., de la Cruz, M. and Ruokolainen, K. 2011. A method to incorporate the effect of taxonomic uncertainty on multivariate analyses of ecological data. *Ecography* 34:94-102.
- Chai, P. and R. Dudley. 1995. Limits to vertebrate locomotor energetics suggested by hummingbirds hovering in heliox. *Nature* 377: 722-725.
- Chapman, F.M. 1917. The distribution of bird-life in Colombia: a contribution to a biological survey of South America. *Bulletin of the American Museum of Natural History* 36:1-736.

- 620 Chapman, F.M. 1926. The distribution of bird-life in Ecuador. Bulletin of the American Museum of Natural History 55:1-784.
- Cheverud, J. M., Dow, M. M. and Leutenegger, W. 1985. The quantitative assessment of phylogenetic constraints in comparative analyses: Sexual dimorphism in body weight among primates. *Evolution* 39:1335-1351.
- Colwell, R. K. 2000. Rensch's rule crosses the line: convergent allometry of sexual size dimorphism in hummingbirds and flower mites. *The American Naturalist* 156:495-510.
- Colwell, R. K. and Rangel, T. F. 2009. Hutchinson's duality: The once and future niche. *Proceedings of the National Academy of Sciences of the United States of America* 106:19651-19658.
- 630 Colwell, R. K. and Rangel, T. F. 2010. Modelling quaternary range shifts and richness on tropical elevational gradients. *Philosophical Transactions of the Royal Society, B* 365: 3695-3707.
- Davies, T. J., R. Grenyer, and J. L. Gittleman. 2005. Phylogeny can make the mid-domain effect an inappropriate null model. *Biology Letters* 1:143-146.
- Davies, T.J., Meiri, S., Barraclough, T. G. and Gittleman, J. L. 2007. Species co-existence and character divergence across carnivores. *Ecology* 10:146-152.
- Darwin, C. 1859. *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. John Murray. London.
- Diniz-Filho, J. A. F., de Sant'Ana, C. E. R. and Bini, L. M. 1998. An eigenvector method for estimating phylogenetic inertia. *Evolution* 52:1247-1262.
- 640 Diniz-Filho, J. A. F. 2001. Phylogenetic autocorrelation under distinct evolutionary processes. *Evolution* 55:1104-1109.
- Diniz-Filho, J. A. F., Bini, L. M., Rodríguez, M. Á., Rangel, T. F. L. V. B. and Hawkins, B. A. 2007. Seeing the forest for the trees: partitioning ecological and phylogenetic components of Bergmann's rule in European Carnivora. *Ecography* 30:598-608.
- Diniz-Filho, J. A. F., Rodríguez, M. Á., Bini, L. M., Olalla-Tarraga, M. Á., Cardillo, M., Nabout, J. C., Hortal, J. and Hawkins, B. A. 2009. Climate history, human impacts and global body size of Carnivora (Mammalia: Eutheria) at multiple evolutionary scales. *Journal of Biogeography* 36:2222-2236.
- Diniz-Filho, J. A. F., Bini, L. M., Rangel, T. F., Morales-Castilla, I., Olalla-Tarraga, M., Rodríguez, M. Á. And Hawkins, B. A. (2011). On the selection of phylogenetic eigenvectors for ecological analyses. *Ecography in press*.
- 650 Dunning, J. B. 2007. *CRC handbook of avian body masses*. 2<sup>nd</sup> ed. CRC press.
- Feinsinger, P., Colwell, R. K., Terborgh, J. and Chaplin, S. B. 1979. Elevation and the morphology, flight energetics, and foraging ecology of tropical hummingbirds. *American Naturalist* 113:481-497.
- Felsenstein, J. 1985. Phylogenies and the comparative method. *American Naturalist* 125:1-15.
- Fjeldså, J. 1994. Geographic patterns for relict and young species of birds in Africa and South America and implications for conservation priorities. *Biodiversity and Conservation* 3: 207-226.
- 660 Fjeldså, J. and Rahbek, C. 2006. Diversification of Tanagers, a species rich bird group, from lowlands to montane regions in South America. *Integrative and Comparative Zoology* 46: 72-81

- Freckleton, R. P., Harvey, P. H. and Pagel, M. 2002. Phylogenetic analysis and the comparative data: a test and review of evidence. *American Naturalist* 160:712-726.
- Gaston, K. J. 2003. The structure and dynamics of geographic ranges. Oxford University Press, Oxford.
- Gittleman, J. L. and Kot, M. 1990. Adaptation: Statistics and a null model for estimating phylogenetic effects. *Systematic Zoology* 39:227-241.
- 670 Gotelli, N. J., Graves, G. R., and Rahbek, C. 2010. Macroecological signals of species interactions in the Danish avifauna. *Proceedings of the National Academy of Sciences* 107:5030-5035.
- Graham, C. H., Parra, J. L., Rahbek, C. and McGuire, J. A. 2009. Phylogenetic structure in tropical hummingbird communities. *Proceedings of the National Academy of Sciences of the United States of America* 106:19673-19678.
- Graves, G. R. 1985. Elevational correlates of speciation and intraespecific geographic variation in plumage in Andean forest birds. *The Auk* 102: 556-579.
- Graves, G. R. 1988. Linearity of geographic range and its possible effect on the population structure of Andean birds. *The Auk* 105: 47-52.
- Graves, G. R. 1991. Bergmann's rule near the equator: Latitudinal clines in body size of an Andean passerine bird. *Proceedings of the National Academy of Sciences of the United States of America* 88:2322-2325.
- 680 Harvey, P. H. and Pagel, M. 1991. The comparative method in evolutionary biology. Oxford University Press, Oxford.
- Hawkins, B. A., Diniz-Filho, J. A. F., Jaramillo, C. A. and Soeller, S. A. 2006. Post-eocene climate change, niche conservatism, and the latitudinal diversity gradient of New World birds. *Journal of Biogeography* 33:770-780.
- Hawkins, B. A., Diniz-Filho, J. A. F., Jaramillo, C. A. and Soeller, S. A. 2007. Climate, niche conservatism, and the global bird diversity gradient. *American Naturalist* 170:S16-S27.
- Helmus, M. R., Bland, T. J., Williams, C. K. and Ives, A. R. 2007. Phylogenetic measures of biodiversity. *American Naturalist* 169:E68-E83.
- 690 Hiebert, S. M. 1990. Energy costs and temporal organization of torpor in the Rufous Hummingbird (*Selasphorus rufus*). *Physiological Zoology* 63:1082-1097.
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. and Jarvis, A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25:1965-1978.
- Holder, M. and Lewis, P. O. 2003. Phylogeny estimation: traditional and Bayesian approaches. *Nature Reviews – Genetics* 4: 275-284.
- Housworth, E. A. and Martins, E. P. 2001. Random sampling of constrained phylogenies: conducting phylogenetic analyses when the phylogeny is partially known. *Systematic Biology* 50:628-639.
- 700 Jablonski, D., Roy, K. and Valentine, J. W. 2006. Out of the tropics: evolutionary dynamics of the latitudinal diversity gradient. *Science* 314:102.
- Lack, D. 1973. The numbers of species of hummingbirds in the West Indies. *Evolution* 27:326-337.
- Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology* 74:1659-1673.

- Martins, E. P. 1996. Conducting phylogenetic comparative studies when the phylogeny is not known. *Evolution* 50:12-22.
- Mayr, E. 1963. *Animal species and evolution*. Belknap Press, Cambridge, MA.
- McGuire, J. A., Witt, C. C., Altshuler, D. L. and Remsen Jr, J. V. 2007. Phylogenetic systematics and biogeography of hummingbirds: Bayesian and maximum likelihood analyses of partitioned data and selection of an appropriate partitioning strategy. *Systematic Biology* 56:837-856.
- Pagel, M. D. 1999. Inferring the historical patterns of biological evolution. *Nature* 401:877-884.
- Parra, J. L., McGuire, J. A. and Graham, C. H. 2010. Incorporating clade identity in analyses of phylogenetic community structure: an example with hummingbirds. *American Naturalist* 176: 573-587.
- Pearmann, P. B., Guisan, A., Broennimann, O. and Randin, C. F. 2008. Niche dynamics in space and time. *Trends in Ecology and Evolution* 23:149-158.
- Peterson, A. T., Soberón, J. and Sánchez-Cordero, V. 1999. Conservatism of ecological niches in evolutionary time. *Science* 285:1265-1267.
- Purvis, A. 2008. Phylogenetic approaches to study extinction. *Annual Reviews of Ecology, Evolution and Systematics* 39:301-319.
- Rahbek, C. and Graves, G. R. 2000. Detection of macro-ecological patterns in South American hummingbirds is affected by spatial scale. *Proceedings of the Royal Society of London, B* 267:2259-2265.
- Rahbek, C. and Graves, G. R. 2001. Multiscale assessment of patterns of avian species richness. *Proceedings of the National Academy of Sciences of the USA* 98:4534-4539.
- Rand, A. S. 1936. Altitudinal variation in new guinea birds. *American Museum Novitates* 890:1-14.
- Rangel, T. F. L. V. B., Diniz-Filho, J. A. F. and Colwell, R. K. 2007. Species richness and evolutionary niche dynamics: a spatial pattern-oriented simulation experiment. *American Naturalist* 170:602-616.
- Rappole, J. and Schuchmann, K. 2003. Ecology and evolution of hummingbird population movements and migration. Page 39 in Berthold, P., Gwinner, E. and Sonnenschein, E. (editors). *Avian migration*. Springer.
- Ronquist, F. 2004. Bayesian inference of character evolution. *Trends in Ecology and Evolution* 19:475-481.
- Schuchmann, K.L. 1999. Family Trochilidae. Pages 468-680 in J. del Hoyo, A. Elliott, and J. Sargatal, editors. *Handbook of the Birds of the World*. Vol. 5. Barn-owls to Hummingbirds. Lynx Edicions, Barcelona.
- Terborgh, J. 1971. Distribution on environmental gradients: theory and preliminary interpretation of distributional patterns in the avifauna of the Cordillera Vilcambaba, Peru. *Ecology* 52, 23-40.
- Webb, C. O. 2000. Exploring the phylogenetic structure of ecological communities: an example for rain forest trees. *American Naturalist* 156:145-155.
- Webb, C. O., Ackerly D. D., McPeck M. A., Donoghue M. J. 2002. Phylogenies and community ecology. *Annual Reviews of Ecology & Systematics* 33:475–505.

Westoby, M. 2006. Phylogenetic ecology at world scale: a new fusion between ecology and evolution. *Ecology* 87:163-165.

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

750 Wiens, J. J., Graham, C. H., Moen, D. S., Smith, S. A. and Reeder, T. W. 2006. Evolutionary and ecological causes of the latitudinal diversity gradient in hylid frogs: treefrog trees unearth the roots of high tropical diversity. *American Naturalist* 168:579-596.

## Online Appendix A: References used in the compilation of maximum elevational range limit of the 304 species of hummingbirds

- Cortés-Diago, A., Ortega, L. A., Mazariegos-Hurtado, L. and Weller, A.-A. 2007. A new species of *Eriocnemis* (Trochilidae) from southwest Colombia. *Ornithologia Neotropical* 18:161-170.
- Fjeldså, J. and N. Krabbe. 1990. *Birds of the High Andes*. Zoological Museum, University of Copenhagen, Denmark.
- 760 Graves, G. R. 1985. Elevational correlates of speciation and intraspecific geographic variation in plumage in Andean forest birds. *Auk* 102:556-579.
- Hilty, S. L. 2003. *Birds of Venezuela*. Christopher Helm, London, UK.
- Hilty, S. L., and Brown, W. L. 1986. *A guide to the birds of Colombia*. Princeton University Press, Princeton, New Jersey.
- Howell, S. N. G. and Webb, S. 1995. *The birds of Mexico and northern Central America*. Oxford University Press, Oxford, UK.
- Ridgely, R. S., and Greenfield, P. J. 2001. *The birds of Ecuador: status, distribution, and taxonomy*. Cornell University Press, Ithaca, New York.
- Ridgely, R. S., and Gwynne, J. A. 1989. *A guide to the birds of Panama*. Princeton University Press, Princeton, New Jersey.
- 770 Schuchmann, K.-L. 1999. Family Trochilidae. Pages 468-680 in J. del Hoyo, A. Elliott, and J. Sargatal, editors. *Handbook of the Birds of the World*. Vol. 5. Barn-owls to Hummingbirds. Lynx Edicions, Barcelona.
- Schulenberg, T. S., Stotz, D. F., Lane, D. F., O'Neill, J. P. and Parker III, T. A. 2007. *Birds of Peru*. Princeton University Press, Princeton, New Jersey.
- Stiles, F. G. 1996. A new species of emerald hummingbird (Trochilidae, *Chlorostilbon*) from the Sierra de Chiribiquete, southeastern Colombia, with a review of the *C. mellisugus* complex. *Wilson Bulletin* 108:1-27.
- Stiles, F. G., and Skutch, A. F. 1989. *A guide to the birds of Costa Rica*. Christopher Helm,
- 780 London.
- Wetmore, A. 1968. *The birds of the Republic of Panama*. Part 2. Columbidae (Pigeons) to Picidae (Woodpeckers). *Smithsonian Miscellaneous Collections* 150:1-605.