A COMPARATIVE AND META-PHYLOGENETIC SYNTHESIS OF HOST SPECIFICITY IN PARASITES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Marc Jean Lajeunesse

January, 2009

A COMPARATIVE AND META-PHYLOGENETIC SYNTHESIS OF HOST SPECIFICITY IN PARASITES

Marc Jean Lajeunesse, Ph. D.

Cornell University 2009

Addressing the importance of constraint and adaptation on the ecology and evolution of host specificity in parasites is an ongoing challenge. Adaptive explanations for the evolution of host specificity predict that specialists are derived from generalists and experience greater extinction rates because of limited ability to adapt to changing conditions. Non-adaptive explanations predict that opportunities to parasitize hosts are limited—geographic barriers and endemism limit the diversity of organisms to parasitize. Here I explore these predictions by (1) developing statistics that allow for the synthesis of published information on host specificity; (2) compiling published phylogenies of parasites to synthesize macroevolutionary patterns of host specificity; (3) using comparative methods to evaluate constraints between the richness and phylogenetic relatedness of host species; and (4) using biogeographic information to test non-adaptive explanations for variation in host specificity.

Analysis of published information on host specificity requires a synthesis of comparative and meta-analytical methods. I develop a statistical framework based on the generalized theory of least squares that integrates phylogenies into meta-analysis, and outline a protocol for testing evolutionary hypotheses by contrasting neutral and adaptive models of evolution based on an Ornstein-Uhlenbeck process.

Using evolutionary meta-analysis, I synthesize 43 published phylogenies to show that macroevolutionary predictions on host specificity—where specialist

parasites are continually evolving from generalists and going extinct—are not universal for parasites. The exceptions were many, and phylogenies where generalists are derived from specialists occur as frequently as lineages with the predicted generalist-to-specialist trajectory.

Focusing on 70 species of obligate avian brood parasites, I further show that the host range (richness of host species) of brood parasites predicts host phylogenetic diversity (average shared phylogenetic distance of host species) worse than random, and that 96% of broad host ranges are invariant in host phylogenetic diversity; indicating that generalists can be as constrained as specialists in the phylogenetic diversity of hosts.

Finally, I show that geographic range size is a good predictor of host specificity in avian brood parasites—where specialist brood parasites occupy smaller geographic ranges that generalists with many host species. This relationship appears to be mediated by the broad continental endemism of independently evolved clades of brood parasites.

BIOGRAPHICAL SKETCH

I grew up in a francophone neighborhood called the Moulin à Fleur in Sudbury (Ontario, Canada). There it is common for sentences to begin in French and end in English—almost unaware, as is the Franco-Ontarian tradition. I attended Carleton University (Ottawa, Ontario) for a B.Sc. (Biology) while living in Hull (Quebec) and returned to Sudbury each summer to work as an industrial laborer in a nickel smelter and copper refinery. Still at Carleton, I transitioned to an M.Sc. (Biology) with Prof. Mark Forbes, where I travelled to Lethbridge (Alberta) for a conference and had the luxury to spend all my summers studying dragonflies and other insects at the Queen's University Biological Station (Chaffey's Locks, Ontario). In 2003, I began my Ph.D. at the University of Toronto with Prof. Anurag Agrawal, while there I presented some of my dissertation research in Ventura (California) and Portland (Oregon). I followed Anurag here to Cornell in 2004, where again I travelled to Ottawa, Montreal, Stony Brook (Long Island), the Archbold Biological Station (Florida), and Christchurch (New Zealand) for conferences or research. Throughout my dissertation, I also visited Santa Barbara (California) four times to meet with an NCEAS workgroup on meta-analysis. In the fall of 2008, I will begin a postdoctoral fellowship at NESCent (Duke University, Durham NC).

Pour Maurice et Danielle Lajeunesse, avec amour.

ACKNOWLEDGMENTS

When I first interviewed with Anurag Agrawal in Toronto, he had asked me to push his car out of a snowbank. I couldn't pull through for him—it was just too stuck and I wasn't strong enough to give it a good push. The opposite situation emerged during my thesis. I was stuck; anxiety crippled my progress. However, Anurag was able to pull through for me. He pushed, he never gave up, and for that I will forever be indebted to him. I was lucky to have him as an advisor; and I'm even luckier to have him as a lifelong friend and colleague.

Throughout my dissertation, there was also strong support from my 'Thagrawal' labmates. There was the University of Toronto crew (who I miss so much): Marc Johnson, Danush Viswanathan, Nile Kurashige, Celine Griffin and Cesar Rodriguez-Saona. I am also very thankful for the Cornell crew (who I will also miss): John Parker, Kailen Mooney, Susan Cook, Alexis Erwin, Scott McArt and Sergio Rasmann. My thesis committee here at Cornell was also inspirational: Rick Harrison, Nelson Hairston Jr. and Bryan Danforth.

In particular, I need thank Mike Stastny and Stuart Campbell (I don't know if I'll make better friends); Amy Parachnowitsch and Sara DeLeon (for just wanting to dance); Susan Rottschaefer and Chris Yourth (for all those wasted hours at Korova); and Jen Koslow, Dave Baker and the Mouchka's (for just helping me kick-it at SNEEB, Center Ithaca, BRB and Pixel). And then there is Danica Lombardozzi: who keeps me calm and comfortable in my skin, and who Anurag should give a million dollars because that's something I once thought impossible. Thank you, Danica.

I also could have never finished this thesis without the support from my parents, Maurice and Danielle Lajeunesse, and my sister Nicole. It's so difficult being away from you; and I'm always working to make you proud.

Finally, the research presented in this thesis would not have been possible without the financial assistance of Anurag Agrawal (Research Assistantships and summer support), National Sciences and Engineering Research Council of Canada (PGSB), the Cornell University Department of Ecology and Evolutionary Biology (Teaching Assistantships, Graduate Research grants, summer support), School of Graduate Studies Travel Grants, Sigma Xi (Cornell Chapter), the Andrew W. Mellon Student Research Grant, and the Cornell IGERT in Biogeochemistry and Environmental Biocomplexity Small Research Grant.

TABLE OF CONTENTS

LIST OF FIGURES

PAGE

CHAPTER ONE

CHAPTER TWO

CHAPTER THREE

CHAPTER FOUR

LIST OF TABLES

PAGE

CHAPTER ONE

CHAPTER TWO

CHAPTER THREE

CHAPTER FOUR

PREFACE

Price (1980) once argued that the evolutionary biology of parasites lacked a coherent body of theory—an issue he regarded as a serious oversight given that parasites form a large portion of the earth's biodiversity. Major advances in several ecological and evolutionary principles have since provided a basis for explaining the diversity of parasites (e.g., Mitter *et al.* 1991; Hafner and Page 1995; but see Poulin and Moran 2000), their impact on populations and communities (e.g., Anderson and May 1981; Price *et al.* 1986), and their role as a constraint on adaptive traits of hosts (e.g., Hamilton and Zuk 1982; Sheldon and Verhulst 1996). Yet addressing the importance of constraint and adaptation on the ecology and evolution of parasites themselves is an ongoing challenge.

This thesis attempts to deepen and amplify what is understood about ecological and evolutionary constraints on the host specificity of parasites. Groups of parasites form a spectrum from highly host-specific taxa to host generalists (Futuyma and Moreno 1988; Jaenike 1991), and explanations for this spectrum can be either adaptive or non-adaptive. Adaptive explanations describe specialists as being incapable of successful reproduction on alternative hosts because of fitness trade-offs or phenotypic constraints (Gould 1979; Karban 1989; Fry 1990)—both are products of selection optimizing fitness on specific host species or characteristics (Whitlock 1996). This optimization is thought to give a competitive advantage over generalists because specialists may be more effective at "finding or mining" hosts as resources (Thompson 1994). However, environmental heterogeneity and rapidly changing conditions will offset this advantage towards generalists because generaists have variation in fitness extending across a broad diversity of hosts (Futuyma and Slatkin 1983; Kassen 2002).

Non-adaptive explanations for variation in host specificity argue that opportunities to parasitize hosts are limited—geographic barriers and endemism limit the diversity of organisms available to be parasitized (e.g., Krasnov *et al.* 2005).

The above adaptive and non-adaptive explanations are in no way an exhaustive list of mechanisms that generate variation in host specificity. However, I focus on these concepts because they form the basis for macroevolutionary predictions about host specificity. These predictions specify that specialist lineages cannot revert back to more generalized states, are continually going extinct, and that new lineages are established by generalists (Mayr 1963; Koch 1980; Futuyma *et al.* 1995). The long-standing challenge has been to directly address the relative importance of past (macroevolutionary) and present-day (ecological) influences on host specificity. In the following chapters, I connect historical patterns to ecological processes in an attempt to explain the broad diversity in host specificity of parasites.

CHAPTER ONE contains the statistical background and principles on which the later chapters on host specificity are based. There I unify the comparative phylogenetic method and meta-analysis into a common statistical framework to synthesize published research from a diversity of taxa. I also extend this framework to allow for the testing of neutral and adaptive hypotheses using a model of evolution based on an Ornstein-Uhlenbeck process (Hansen 1997). CHAPTER TWO generalizes macroevolutionary patterns in host specificity across a broad diversity of parasites (e.g., avian brood parasites, phytophagous insects, monogeneans) using this statistical framework. I particularly focus on the most common prediction of the macroevolution of host specificity: that specialist parasites occupy the most derived positions within lineages and suffer greater extinction rates than generalists (Mayr 1963; Koch 1980; Kelly and Farrell 1998).

In the remaining chapters I narrow my focus with a detailed investigation of host specificity among avian brood parasites—birds that parasitize the brood rearing efforts of other birds by laying eggs in nests of different species. Avian brood parasites are uniquely appropriate for comparative questions because information on host specificity is based on nearly a century of detailed natural history observations (see Davies 2000). CHAPTER THREE explores the complex connection between host range and host phylogenetic diversity among avian brood parasites. CHAPTER FOUR analyzes geographic constraints on host specificity of avian brood parasites.

REFERENCES

Anderson, R. M., and R. M. May. 1981. The population dynamics of microparasites and their vertebrate hosts. *Philosophical Transactions of the Royal Society of London* 210:658–661.

Davies, N. B. 2000. *Cuckoos, Cowbirds and Other Cheats*. T. & A.D. Poyser.

Fry, J. D. 1990. Trade-offs in fitness on different hosts: evidence from a selection experiment with a phytophagous mite. *American Naturalist* 136:569–565.

Futuyma, D. J., and G. Moreno. 1988. The evolution of ecological specialization. *Annual Review of Ecology and Systematics* 19:207–233.

Futuyma, D. J., and M. Slatkin. 1983. *Coevolution*. Sunderland, Mass., Sinauer.

Gould, F. 1979. Rapid host range evolution in a population of the phytophagous mite *Tetranychus urticae* Koch. *Evolution* 33:791–802.

Hafner, M. S., and R. D. Page. 1995. Molecular phylogenies and host-parasite cospeciation: gophers and lice as a model system. *Philosophical Transactions of the Royal Society of London, Series* B 349:77–83.

Hamilton, W. D., and M. Zuk. 1982. Heritable true fitness and bright birds: A role for parasites? *Science* 218:384–387.

Hansen, T. F. 1997. Stabilizing selection and the comparative analysis of adaptation. *Evolution* 51:1341–1351.

Jaenike, J. 1990. Host specialization in phytophagous insects. *Annual Review of Ecology and Systematics* 21:243–273.

Karban, R. 1989. Fine-scale adaptation of herbivorous thrips to individual hot plants. *Nature* 340:60–61.

Kassen, R. 2002. The experimental evolution of specialists, generalists, and the maintenance of diversity. *Journal of Evolutionary Biology* 15:173–190.

Kelley, S. T., and B. D. Farrell. 1998. Is specialization a dead end? The phylogeny of host use in *Dendroctonus* bark beetles (Scolytidae). *Evolution* 52:1731–1743.

Koch, C. F. 1980. Bivalve species duration, areal extent and population size in a Cretaceous sea. *Paleobiology* 6:184–192.

Krasnov, B. R., R. Poulin, G. I. Shenbrot, D. Mouillot, and I.S. Khokhlova. 2005. Host specificity and geographic range in haematophagous ectoparasites. *Oikos* 108: 449–456.

Mayr, E. 1963. *Animal Species and their Evolution.* Belknap Press of Harvard University Press, Cambridge, MA.

Mitter, C., B. Farrell, and D. J. Futuyma. 1991. Phylogenetic studies of insect-plant interactions: Insights into the genesis of diversity. *Trends in Ecology & Evolution* 6:290–293.

Poulin, R., and S. Morand. 2000. The diversity of parasites. *Quarterly Review of Biology* 75:277–82.

Price, P. W., M. Westoby, B. Rice, P. R. Atsatt, R. S. Fritz, J. N. Thompson, and K. Mobley. 1986. Parasite mediation in ecological interactions. *Annual Review of Ecology and Systematics* 17:487–505.

Sheldon, B. C., and S. Verhulst. 1996. Ecological immunity: costly parasite defenses and trade-offs in evolutionary ecology. *Trends in Ecology & Evolution* 11:317–21.

Thompson, J. N. 1994. *The coevolutionary process.* University of Chicago Press, Chicago, IL.

Whitlock, M. C. 1996. The Red Queen beats the Jack-of-all-trades: the limitations on the evolution of phenotypic plasticity and niche breath. *American Naturalist* 148:S65–S77.

CHAPTER ONE

META-ANALYSIS AND THE COMPARATIVE PHYLOGENETIC METHOD

Abstract

Meta-analysis has contributed substantially to shifting paradigms in ecology and has become the requisite method for quantitatively synthesizing published research. However, a unique and emerging challenge for biology is the lack of statistical methods and protocols to synthesize studies while simultaneously accounting for phylogenetic non-independence of taxa. Phylogenetic non-independence arises from homology, the similarity of taxa due to shared ancestry, and treating related taxa as independent data violates assumptions of statistics. Given that an explicit goal of meta-analyses is to generalize across a broad range of taxa, phylogenetic non-independence may threaten the validity of conclusions drawn from such reviews. Here, I develop a statistical framework that integrates phylogenetic information into conventional meta-analysis when a) taking a weighted average effect size using fixed- and random-effects models and b) testing for homogeneity of variances. In addition, I provide a protocol for testing evolutionary hypotheses—a first for meta-analysis—by introducing methods that a) evaluate phylogenetic conservatism and b) contrast neutral and adaptive models of evolution based on an Ornstein-Uhlenbeck process. I illustrate these methods by integrating phylogenetic information in a meta-analysis that addresses the evolution of mating systems in Lepidoptera. Finally, I address several theoretical and practical issues to provide a unified statistical framework accessible to ecologists and evolutionary biologists.

Introduction

Closely related taxa are more similar in morphology, physiology, behavior and ecology than distantly related taxa (Harvey and Purvis 1991). This similarity due to shared phylogenetic history is a problem when analyzing data from a diversity of taxa because it violates two statistical assumptions. First, data are assumed to be drawn from independent samples; phylogenetic history introduces a correlated structure to data because taxa form a nested hierarchy of phylogenetic relationships (Felsenstein 1985; Maddison 1990). Second, data are assumed to be sampled from a population that has a normal distribution with a common variance; sampling data with a phylogenetic structure can yield different variance structures because lineages within phylogenies may have evolved at different rates (Pagel 1992; Pagel 1999). Given that an explicit goal of recent meta-analyses on tradeoffs (Koricheva *et al.* 2004), trophic cascades (Borer *et al.* 2005), and invasive biology (Parker *et al.* 2006) is to include a diversity of taxa and to generalize across a broad range of species, violating these assumptions may threaten the validity of conclusions drawn from such reviews.

A brief example will help to make the point. Plant resource allocation theory predicts tradeoffs among multiple anti-herbivore defenses. Yet a meta-analysis of 31 independent studies found no common pattern of defense tradeoffs among grasses, herbaceous plants and trees—despite all sharing herbivory as a source of selection (Koricheva *et al.* 2004). Although meta-analysis provided a powerful framework to test this theory—by synthesizing and exploring heterogeneity of results from multiple studies while accounting for statistical issues known to mediate the ability to detect significant outcomes—it could not provide an evolutionarily robust test of this theory. Here the test for generality of tradeoffs was potentially biased because the non-independence in the phylogenetic history of plants and how this relates to the evolution of defenses was not evaluated (see Agrawal 2007). A robust test of the

generality of such a relationship would account for phylogenetic non-independence and thus explicitly address whether defences across plant taxa were convergent solutions to herbivory.

This issue is systematic to all meta-analyses based on a diversity of taxa, and if the literature synthesized is affected by phylogenetic history, then this may in turn bias meta-analytic results to yield overstated conclusions. Here I introduce a framework that integrates phylogenetic information into meta-analysis by unifying the statistics of meta-analysis and the comparative phylogenetic method—this is possible because both are special cases of the generalized least squares theory (Hedges and Olkin 1985; Cooper and Hedges 1994; Rohlf 2001; Adams 2008). My statistical approach integrates phylogenetic information in all the traditional meta-analytical statistics, such as fixed and random-effects models for pooling effect sizes across studies and heterogeneity tests (for a general text see Hedges and Olkin 1985).

Introducing phylogenetic information into meta-analysis also significantly extends the scope of quantitative reviews: now the historical evolutionary processes that are responsible for the diversity of taxa and their responses to experimentation can be evaluated. For instance, integrating phylogenetic information provides a conceptual advantage in that a phylogenetically-independent meta-analysis can indicate evolutionary convergence—taxa experiencing similar selection have evolved similarly in response to this selection, and thus may respond in the same manner to experimentation. In fact, all ecological meta-analyses that do not explicitly include phylogenetic information assume convergence when pooling responses to experimentation across multiple taxa. Without a formal test of this assumption, however, reviewers can have a biased interpretation of the meta-analytic results because an alternative explanation—that responses to experimentation are phylogenetically constrained—is overlooked. Again, phylogenetic conservatism or

simply similarities due to shared ancestry may overestimate adaptive similarities due to convergent evolution (Schluter 2000).

In order to directly test these evolutionary hypotheses, I introduce diagnostics to evaluate phylogenetic conservatism of meta-analytic data, and statistics for contrasting neutral and adaptive models of evolution. These tools differ from the more common practice of using the statistics of comparative analysis solely as a control to avoid the 'pseudo-replication' with related taxa or to eliminate bias due to phylogenetic history (see Westoby *et al.* 1995; Ricklefs and Starck 1996). Reviewers need to think beyond simply phylogenetically-correcting meta-analyses, and should start explicitly recognizing that phylogenetic history may be an important explanatory variable for the diversity of experimental responses across taxa. I conclude with an application of this framework by re-analyzing a published meta-analysis that tests the contribution of lepidopteran mating system to explain why females have greater reproductive success when they mate with virgin males (Torres-Vila and Jennions 2005).

Pooling Studies from Multiple Taxa and the "Apples and Oranges Problem"

The "apples and oranges problem" has a long history in meta-analysis, and is a criticism of mixing studies with different conceptual and operational definitions (Wolf 1986; Lynn 1989), or more generally studies belonging to different groups (e.g., different estimates of experimentation, different metrics of study outcomes, etc.). The analogy here identifies that apples differ from oranges, and thus aggregating them as a group is not appropriate or meaningful. The "apples and oranges problem" is not a valid criticism because the rationale for aggregating studies is relative to the focus of the review: if the focus is on apples then including oranges is inappropriate, but if the focus is on fruit, then grouping apples and oranges is meaningful. The distinction is

important because it recognizes that, conceptually, objects can form a hierarchy: apples and oranges are fruit because they share similar characteristics (e.g., in an informal sense they are edible, sweet, and have seeds). This concept of hierarchy should also resonate for all ecologists and evolutionary biologists because taxa form a nested hierarchy of phylogenetic relationships. Apples and oranges have similar characteristics because of their shared ancestry among flowering plants (angiosperms). Considering the phylogenetic history of taxa is important for meta-analysis because quantitative reviews often seek to generalize as broadly as possible by pooling (mixing) all studies of a given conceptual topic—irrespective of taxonomic identity of the organism under study.

However, pooling studies from multiple taxa is only meaningful as long as the hierarchical relationship of taxa is recognized as a potential bias. A common approach to evaluate this bias is to group studies by taxonomic rank and then treat these subsets as a moderator variable (e.g., different Lepidoptera families). Moderator variables are important for meta-analysis because they help evaluate potential sources of bias, are useful to explain study heterogeneity in experimental outcomes, and to test hypotheses (Hedges and Olkin 1985). Here, a meta-analysis is repeated with subsets of studies that are parsed by taxonomic rank. Alternatively, subsets of studies with a particular taxonomic rank are excluded from analyses to evaluate their overall sensitivity of meta-analysis to the inclusion of these groups (Lajeunesse and Forbes 2002). Both these approaches are based on the assumption that if a meta-analysis is consistent among different taxonomic ranks or consistent when these groups are excluded from analysis then the overall meta-analysis may not have a significant taxonomic bias.

These approaches are useful to determine the presence of bias when mixing studies from a diversity of taxa, but do not provide a solution to account for this bias when pooling study effect sizes or testing for study heterogeneity in research

outcomes. If a moderator variable is useful for explaining variation in study outcomes, then it should be modeled explicitly in meta-analyses to increase accuracy. Given the considerable evidence that pooled effect size estimates can be biased when few studies are included in reviews (Hedges and Olkin 1985), and that this bias is not independent of the effect size metric used (Lajeunesse and Forbes 2003), then any approach that controls for potential sources of bias, such as phylogenetic history, can improve the ability of meta-analysis to detect significant overall effect sizes and allow for superior exploration of other sources of study heterogeneity.

Non-Independence and Expected Variance

Similarity of taxa due to shared ancestry is a problem for meta-analysis because it violates two statistical assumptions. First, that data are drawn from independent samples—here effect sizes from different studies may share similar magnitudes or directions because they are derived from closely related species. Second, that effect sizes are sampled from a population of research outcomes that has a normal distribution with an expected or common variance—when effect sizes are sampled from a population with an evolutionary history the variances may differ among lineages. Violating these assumptions increases the likelihood of making type I error; that is finding a significant study outcome when one does not exist (Harvey and Pagel 1991).

Phylogenetic history violates the first assumption of independence because it introduces a correlated structure to data: species form a nested hierarchy of phylogenetic relationships such that their traits and characteristics do not have an independent origin. In meta-analysis, however, the units of analysis are effect sizes—a statistical measure of the magnitude and direction of experimental outcomes (Hedges and Olkin 1985)—and unlike species traits, the relative contribution of phylogenetic

history on influencing study outcomes remains empirically unknown. However, examples can be envisioned, and beyond meta-analysis, effect sizes are emerging as an important unit of analysis in ecology and evolutionary biology (Strauss *et al.* 2008). Consider an effect size that estimates the magnitude of experimental differences using a control and treatment mean (i.e. Hedges' *d* or lnRR; see Table 2). Phylogenetic history can bias these effect sizes in two ways. First, means are derived from measurements of species traits, and these traits may be phylogenetically conserved. For example, body size is often used as a surrogate for fitness, and for many animals body size is phylogenetically conserved, where closely related species or whole lineages tend to share similar sizes. The second bias can occur if the effect sizes themselves are phylogenetically conserved—such as phenotypic responses to multiple environments or among taxa with sexual dimorphism. Here, the magnitude of differences in body size between sexes may also be due to common ancestry among taxa. It is known, for instance, that the magnitude and direction (females > males, or females < males) of size dimorphisms is dependent on the mating system of a lineage (e.g., polygamy or monogamy; Björklund 1997). Again, the mating system itself may also be a phylogenetically conserved trait. These two phylogenetic biases of effect sizes may not be independent, but in either case will contribute to an evolutionary correlated structure of study outcomes: studies using related species for experimentation may yield similar study outcomes (effect sizes).

Several statistical approaches have been developed in comparative analysis to account for phylogenetic non-independence (see Felsenstein 1985; Chererud *et al.* 1985; Maddison 1990; Graphen 1989; Martins and Garland 1991; Pagel 1997), but Felsenstein's (1985) phylogenetically-independent contrasts (PIC) remains the most widespread. An alternative using generalized least squares (GLS) is increasingly used (Pagel 1994), and it is now recognized that PIC's are a special case of the GLS

approach (Pagel 1999; Rohlf 2001). Here, I apply this GLS framework to account for phylogenetic history because meta-analysis is also modeled under the GLS family of statistics. The main advantage of this statistical framework is the full flexibility to apply different statistical models for analysis. Below, I outline these methods and the application of the GLS approach in more detail.

The second statistical assumption invalidated by phylogenetic history is that data are sampled from a normal distribution with an *expected* variance. Sampling from a phylogeny can generate data with different variances because lineages within a phylogeny may evolve at different rates or may have had different times to evolve (Harvey and Pagel 1994). Variance here is defined as the rate of evolutionary change within a phylogeny, and typically comparative analyses make assumptions about how lineages evolve to meet the statistical assumption of an expected variance. For instance, Felsenstein's (1985) approach with phylogenetically-independent contrasts assumes that evolution proceeds as a Brownian motion (BM) process (e.g., random drift), and uses information on phylogenetic branch-lengths to calculate the expected variance of change. This is because Brownian motion evolution predicts that long branches can yeild more character change, and that evolutionary rates are the same throughout the phylogeny (and thus all have the same expected variance of change; see Felsenstein 1985). Thus to satisfy both assumptions of non-independence and variance, Felsenstein's approach transforms data into a set of contrasts that have zero (phylogenetic) covariance and standardizes these contrasts to have equal (evolutionary rate) variance (using the square root of the sum of all branch-lengths as the expected variance of change).

Although BM forms the basis for nearly all phylogenetic comparative statistics (Felsenstein 1985; Cheverud *et al.* 1985; Maddison 1991; Graphen 1989; Martins and Garland 1991; Pagel 1997), it is a model that over-simplifies the process of evolution

and natural selection. For instance, BM assumes that character change is independent within each lineage, and that the character variance among lineages will increase with time (Martins 1994). However, selection is an important force in character change among taxa, and a more complete evolutionary model should account for this force. Hansen (1997) proposed a model of evolution by means of an Ornstein-Uhlenbeck (OU) process, which has two evolutionary parameters: η as the intensity of the random fluctuations in the evolutionary process (drift), and β as the strength of selection. This approach is more appropriate for evolutionary hypothesis testing because it includes natural selection as a process that contributes to the evolution of character change within a lineage. In addition, because the OU model assumes that selection occurs, and under some parameters of selection (i.e. stabilizing selection), it is expected that the variance in traits will remain bounded and constant through time—this complies with the statistical assumption of homogeneity of variances. Thus, using BM and OU to contrast different models of both neutral and adaptive evolution can provide a powerful framework for evolutionary hypothesis testing in meta-analysis. This can be achieved by contrasting the fit of different models of evolution (BM vs OU) on a collection of studies based on taxa with an estimated phylogenetic tree.

A Primer on Meta-Analysis and the Comparative Method

One purpose of meta-analysis is to statistically weight study outcomes by their inverse variance to control for within-study sampling error (see Lajeunesse and Forbes 2003). This downweighting is important because studies with large variances or small sample sizes are sensitive to sampling error—perhaps under or over estimating effect sizes. The comparative phylogenetic method, however, uses phylogenetic information during the regression of traits to control for shared ancestry of taxa (Felsenstein 1985) and to test explicit models of evolution (Pagel 1999). Here again, a weighting scheme

is used to penalize data: data based on taxa stemming from short branches on phylogenies are downweighted because they may not represent independent pieces of information (e.g., not enough time for derived characteristics to change).

One practice for accounting for phylogenetic non-independence is to perform a comparative analysis on meta-analytical data (see Verdú and Traveset 2004, 2005) or to use a phylogenetically-independent transformation and then analyze these transformed data with a conventional GLS framework (Adams 2008). This latter approach will yield a phylogenetically-independent analysis, but will transform effect sizes under evolutionary units such that direct comparisons between untransformed (non-phylogenetically corrected) and transformed meta-analyses are impossible (Butler and King 2004). This comparison is essential to determine whether including evolutionary history in the model was useful in explaining variation among effect sizes. In addition, these approaches do not yield the familiar statistics of meta-analysis: a) they may not use sample precision to weight studies when pooling study outcomes or b) use important meta-analytic statistics that are necessary to evaluate study heterogeneity.

What unites meta-analysis and the comparative phylogenetic method is that they are both special cases of the generalized theory of least squares (GLS). Statistics based on ordinary least squares, such as regression and ANOVA, have several assumptions (for a general text see Groβ 2003); but the two of interest for meta-analysis are that effect sizes (the independent variable) share a common variance (e.g., are homoskedastic) and are uncorrelated (statistically independent). A way of illustrating these assumptions using matrix notation (see Table 1.1) is as follows

$$\mathbf{E} \sim \mathcal{N}(\mathbf{X}\overline{\delta}, \sigma^2\mathbf{I}), \tag{1.1}$$

where $\mathbf{E}$ is a $k \times 1$ column vector of k number of effect sizes (δ), which are assumed

Table 1.1. A roundup of variables used in evolutionary meta-analysis.

symbol	Definition
Effect size parameters	
δ	Effect size (for examples see Table 1.2)
$\sigma^2(\delta)$	Effect size variance (see Table 1.2)
Meta-analysis	
k	Number of effect sizes for synthesis (review sample size)
$\overline{\mu}$	Unweighted mean effect size across k studies
$\overline{\delta}$	Weighted mean effect size (fixed-effects)
$\mathbf{X}$	The design matrix defined as $k \times 1$ column vector of ones (modeled to take the average of effect sizes), or information on evolutionary optima (θ)
$\mathbf{E}$	Column vector with a $k \times 1$ dimension containing all the effect sizes (δ)
Σ	Covariance matrix of $k \times k$ dimensions used to account for within-study sampling error and phylogenetic non-independence
Q	Homogeneity test of effect size variances
τ	Between-study effect size variance
$\hat{\delta}$	Random-effects mean effect size across k studies using τ
Phylogenetics	
λ	Degree of a phylogenetic signal (conservatism or inertia)
b	Branch-length distance on a phylogenetic tree
$\mathbf{P}$	Phylogenetic correlation matrix of $k \times 1$ dimensions containing all the shared b between taxa within a phylogeny
BM	Brownian-motion model of evolution for coding $\mathbf{P}$
OU	Ornstein-Uhlenbeck process used to model $\mathbf{P}$
η	Evolutionary parameter depicting the strength of random drift in OU
β	Evolutionary parameter depicting the strength of selection in OU
θ	Evolutionary optima that designates the period under which selection occurs
m	Number of hypothesized evolutionary optima (θ)
Comparative phylogenetic meta-analysis	
$\overline{\delta}^{\mathbf{P}}$	Weighted and phylogenetically-independent mean effect size across k studies
$Q_{\mathrm{R}}^{\mathbf{P}}$	Phylogenetically-independent chi-square statistic testing whether $\overline{\delta}^{\mathbf{P}} \neq 0$
$Q_{\mathrm{H}}^{\mathbf{P}}$	Phylogenetically-independent homogeneity test for $\delta_1 = \delta_2 = \ldots = \delta_k$
$\mathbf{W}$	Column vector containing the taxonomic weights of m number of groups
AIC	Akaike's information criterion used for selecting the best among competing evolutionary models

Note: Bold type and non-italicized symbols are vectors (all column vectors) or matrices, and transposed matrices are denoted with $^{\mathrm{T}}$ and inverse matrices with an exponent of –1.

to be normally distributed ($\mathcal{N}$) with an expected mean of $\mathbf{X}\bar{\delta}$ and variance of $\sigma^2\mathbf{I}$. The expected mean ($\mathbf{X}\bar{\delta}$) of **E** designates the averaging behavior of effect sizes. How effect sizes are averaged is defined in by design matrix **X** (typically a $k \times 1$ column vector of ones), and this average yields the pooled effect size ($\bar{\delta}$). The variance of effect sizes ($\sigma^2\mathbf{I}$) is known as the scalar variance-covariance matrix (i.e. var(**E**) = $\sigma^2\mathbf{I}$), and this matrix defines how effect sizes are correlated. The identity matrix (**I**) here indicates that variances are uncorrelated and share a common variance (σ^2).

The method of generalized least squares is a statistical framework that directly addresses violations due to non-independence and heteroscedasticity of data. More precisely, instances where var(**E**) $\neq \sigma^2\mathbf{I}$. These violations are explicitly modeled in a $k \times k$ covariance matrix ($\mathbf{\Sigma}$), such that the expected distribution of effect sizes is now

$$\mathbf{E} \sim \mathcal{N}(\mathbf{X}\bar{\delta}, \mathbf{\Sigma}). \quad (1.2)$$

For instance, a weighted meta-analysis uses $\mathbf{\Sigma}$ that contains the variances of each effect size $\sigma^2(\delta)$ on its main diagonal—effectively modeling a weighted least squares regression where effect sizes with large variances are penalized during the pooling of effect sizes. This codification of $\mathbf{\Sigma}$ differs from the comparative method, which uses all off-diagonal elements (covariances) of $\mathbf{\Sigma}$ to account for the correlated evolution history of taxa, thereby giving less weight to taxa that are on average more closely related to other taxa when fitting a regression line through these data (see Pagel 1997, 1999). Thus the elements of $\mathbf{\Sigma}$ can be formulated to serve both the interests of meta-analysis (weighting by study sampling error) and the comparative method (weighting by relative relatedness of taxa).

Statistical Framework

The following statistical framework for meta-analysis is divided into three sections. I first describe how to code the elements of the $\mathbf{\Sigma}$ covariance matrix to account for non-independence and heteroscedasticity. I then apply this matrix to a GLS framework for calculating a phylogenetically-independent meta-analysis of effect sizes and homogeneity tests. The final section outlines how this GLS framework can be extended to test neutral and adaptive evolutionary hypotheses with meta-analysis. It is important to note that all of the methods outlined below can be applied to any effect sizes metric as long as an estimate of their variance is known (see Table 1.2 for examples).

Covariance Matrix and Bias

Here I model both heteroscedasticity and phylogenetic correlations in a single covariance matrix ($\mathbf{\Sigma}$). The first source of bias of heteroscedasticity is modeled on the main diagonal of $\mathbf{\Sigma}$ which will contain the effect size variances $\sigma^2(\delta_i)$ for each i^{th} effect size of a collection of k studies. As in traditional meta-analysis, studies with large variances are weighted less heavily because they are imprecise estimates of the 'true' population effect size (Hedges and Olkin 1985). The second bias of phylogenetic non-independence is modeled on all the off-diagonals of $\mathbf{\Sigma}$, which contain the between study covariances (cov). Here, the covariances measure how effect sizes vary together based on their correlated phylogenetic history (as described in the $\mathbf{P}$ correlation matrix), and are calculated for each pair of effect sizes as follows:

$$\text{cov}(\delta_i, \delta_j) = \mathbf{P}_{i,j}\sqrt{\sigma^2(\delta_i)}\sqrt{\sigma^2(\delta_j)}\,. \tag{1.3}$$

Thus the symmetric covariance matrix $\mathbf{\Sigma}$ that has the following elements in its $i = 1,\ldots,k$ rows and $j = 1,\ldots,k$ columns:

Table 1.2. Variance estimates $\sigma^2(\delta)$ of effect size metrics (δ) used in ecological meta-analysis. Effect sizes and their variances are used to define the variance-covariance matrix (Σ). This matrix is then applied to GLS models that estimates phylogenetically-independent weighted mean of a collection of studies. For more details of these variances see Cooper and Hedges (1994). Here, effect sizes are estimated from a study that has a sample size (*n)*, and control (*C*) and treatment (*T*) means ($\overline{X}$). Finally, the log odds ratio is based on count data of members belonging to groups A or B. Examples of these effect sizes are found in Van Zandt and Mooper (1997) for Hedges' *d* and log(*RR*), Koricheva *et al.* (2004) for correlation coefficients, and Beirinckx *et al.* (2006) for log(*OR*).

effect size type	effect size [A]	effect size variance
Hedges' *d*	$d = \frac{\overline{X}_T - \overline{X}_C}{s}\left(1 - \frac{3}{4(n_T + n_C) - 9}\right)$	$\frac{n_T + n_C}{n_T n_C} + \frac{d^2}{2(n_T + n_C)}$
log response ratio (*RR*)	$RR = \ln\left(\frac{\overline{X}_T}{\overline{X}_C}\right)$	$\frac{s_T^2}{n_T \overline{X}_T^2} + \frac{s_C^2}{n_C \overline{X}_C^2}$
Pearson's product-moment correlation coefficient (*r*)	$r = r$	$\frac{(1 - r^2)^2}{n - 2}$
Fisher *Z* transformation of *r*	$Z(r) = \frac{1}{2}\ln\left(\frac{1 + r}{1 - r}\right)$	$\frac{1}{n - 3}$
log odds ratio (log *OR*)	$\log OR = \ln\left(\frac{n_{A_C} n_{B_T}}{n_{A_T} n_{B_C}}\right)$	$\frac{1}{n_{A_C}} + \frac{1}{n_{A_T}} + \frac{1}{n_{B_C}} + \frac{1}{n_{B_T}}$

[A] Hedges' *d* uses the pooled standard deviation (*s*) between the control and treatment; this is calculated as $s = ([(n_T - 1)s_T^2 + (n_C - 1)s_C^2]/[n_T + n_C - 2])^{½}$.

$$\boldsymbol{\Sigma}_{i,j} = \begin{bmatrix} \sigma^2(\delta_1) & \cdots & \operatorname{cov}(\delta_1, \delta_{j-1}) & \operatorname{cov}(\delta_1, \delta_j) \\ & \ddots & \vdots & \vdots \\ & & \sigma^2(\delta_{i-1}) & \operatorname{cov}(\delta_{i-1}, \delta_j) \\ & & & \sigma^2(\delta_i) \end{bmatrix}. \tag{1.4}$$

Taking the inverse of $\boldsymbol{\Sigma}$ yields a weight matrix (Groβ 2003), however for simplicity I will continue to refer to $\boldsymbol{\Sigma}$ as the covariance matrix.

Correlation Matrix and Phylogenetic History

The **P** matrix contains the correlated relationships among effect sizes due to the shared phylogenetic history of taxa. The strength of these correlations are often coded as the phylogenetic branch-length (*b*) distance between taxa—where for example, the total branch-length distance for each species are on the main diagonal, and the shared distance between species are all off-diagonals of the matrix. In effect, this approach to coding the **P** matrix assumes the purely neutral BM model of evolution. A more formal definition of $\mathbf{P}^{\mathrm{BM}}$ is as follows

$$\mathbf{P}^{\mathrm{BM}}_{i,j} = \begin{cases} b_i^{\mathrm{total}} & \text{if } i = j, \\ b_{i,j}^{\mathrm{shared}} & \text{if } i \neq j. \end{cases} \tag{1.5}$$

Here, b denotes the phylogenetic branch-length distance, where the total distance from the root to tip is b^{total}, and b^{shared} is the branch-length distance that taxon *i* shares with taxon *j* (see example of $\mathbf{P}^{\mathrm{BM}}$ in Rohlf 2001). Comparative methods based on GLS often treat this distance matrix as the covariance matrix (e.g., $\boldsymbol{\Sigma} = \mathbf{P}^{\mathrm{BM}}$) and apply it directly to GLS equations (see Rohlf 2001).

Hansen (1997) proposed an alternate model of evolution by means of an Ornstein-Uhlenbeck process (OU; Gardiner 1985). Both BM and OU are stochastic processes, but in which BM is similar to white noise, and OU is more of a 'colored' noise. Biologically, OU differs from BM because it assumes that natural selection is a

process that contributes to 'color' the evolution of character change within a lineage. Unlike the Brownian motion model which predicts a linear relationship between phylogenetic correlations and time since divergence, OU assumes that closely related clades are exponentially more similar to one another than more distantly related taxa (see Figure 1.1). This exponential relationship becomes more pronounced in clades undergoing strong selection, where strong selection acts to erase any ancestral constraints (Hansen 1997).

Briefly, the phylogenetic correlation matrix (**P**) coded under OU is

$$\mathbf{P}_{i,j}^{\mathrm{OU}} = \eta^2 \, \mathbf{C}_{i,j}, \tag{1.6}$$

where **C** is a *k* by *k* matrix with the following elements

$$\mathbf{C}_{i,j} = \begin{cases} (1 - e^{-2\beta b^{\mathrm{total}}}) / 2\beta & \text{if } i = j, \\ (e^{-2\beta (b^{\mathrm{total}} - b_{i,j}^{\mathrm{shared}})} - e^{-2\beta b^{\mathrm{total}}})) / 2\beta & \text{if } i \neq j. \end{cases} \tag{1.7}$$

Here there are two evolutionary parameters that are used to model phylogenetic correlations: the strength of random drift (η) and selection (β). Both parameters can range from zero to infinity, as long as $\eta < \beta$. Note that both η and β are explicitly related to time of divergence among taxa, such that the phylogeny used to code the $\mathbf{P}^{\mathrm{OU}}$ matrix should have a common timescale and a universal b^{total} (e.g., an ultrametric tree where all taxa are aligned at the tips). This timescale does not have to be absolute, but should at least contain information on the relative divergence times among the taxa under study. An important feature of $\mathbf{P}^{\mathrm{OU}}$ is that it converges to $\mathbf{P}^{\mathrm{BM}}$ (e.g., $\mathbf{P}^{\mathrm{OU}} \rightarrow \mathbf{P}^{\mathrm{BM}}$) when the strength of selection approaches zero ($\beta \rightarrow 0$). Later I discuss how this nestedness of $\mathbf{P}^{\mathrm{BM}}$ in $\mathbf{P}^{\mathrm{OU}}$ allows the sequential fitting of simple neutral and adaptive evolutionary models (Butler and King 2004).

Before $\mathbf{P}^{\mathrm{BM}}$ and $\mathbf{P}^{\mathrm{OU}}$ can be applied to the covariance matrix (equation 1.1), a few issues need to be considered. The covariance equation used in meta-analysis

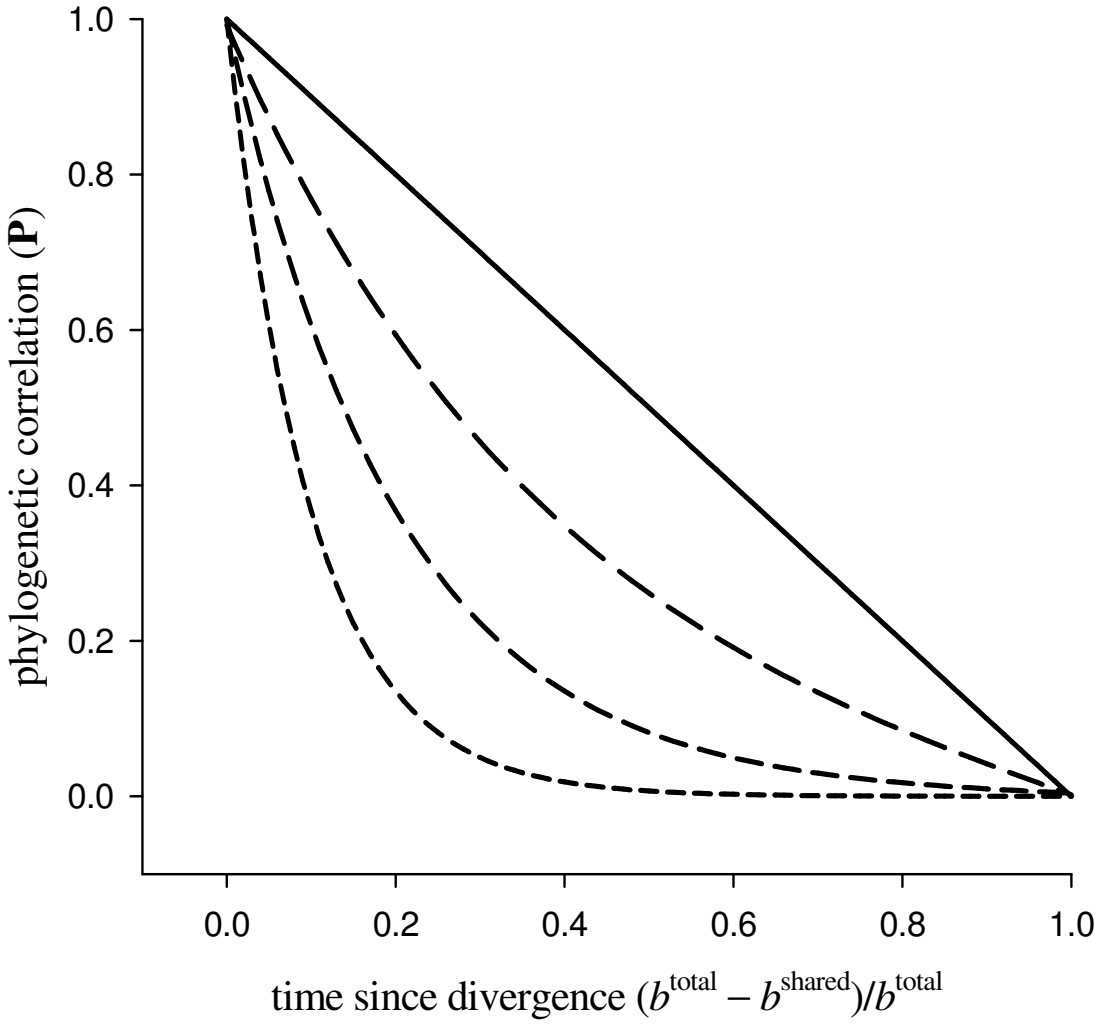


Figure 1.1. The predicted relationship between similarity among taxa (phylogenetic correlation) and their relative time since divergence for Brownian motion (solid line) and Ornstein-Uhlenbeck (dashed lines) models of evolution. The multiple dashed lines show how under an OU model the phylogenetic correlations decrease exponentially with time when taxa undergo weak to moderate to strong selection (as the last dashed line with the most pronounced curve). Because Brownian motion evolution is nested within OU, the solid line also depicts OU when $\beta = 0$.

assumes that the diagonal correlations of **P** equal one, and that all non-diagonal elements should not equal or exceed one or negative one. This assumption relating to elements of the correlation matrix is important because they contribute to computations issues of matrix inversion (Groβ 2003). This problem is also why $\mathbf{P}^{BM}$ or $\mathbf{P}^{OU}$ should not be directly applied to GLS models for meta-analysis—they would yield incomparable pooled effect sizes because the main diagonal of **P** will contain the total branch-length distance of each taxa. This diagonalization with phylogenetic branch-lengths will scale the weighting by variance among effect sizes. Generally this transformation is accepted in comparative analysis, because using **P** as the covariance matrix $\mathbf{\Sigma}$ has an effect of weighting the traits of taxa by the inverse of their total branch length distance (which is on the main diagonal). This fits with an assumption of BM evolution that long branches accumulate more change, and thus taxa with long branches (with more time for drift to change in traits) are penalized during analysis. However, this differential weighting only occurs each taxon does not share a common b^{total}.

To avoid this problem, I impose a constraint on the **P** correlation matrix, such that the branch-lengths distances among taxa are summarized as follows

$$\overline{\mathbf{P}}_{i,j} = \begin{cases} 1 & \text{if } i = j, \\ 2b_{i,j}^{\text{shared}}(b_i^{\text{total}} + b_j^{\text{total}})^{-1} & \text{if } i \neq j. \end{cases} \tag{1.8}$$

Here, the off-diagonal elements are now constrained to $0 < [2b_{i,j}^{\text{shared}}(b_i^{\text{total}} + b_j^{\text{total}})^{-1}] < 1$, such that closely related taxa will yield higher correlations, and taxa with few shared branch-lengths will have small correlations. Note that the conditions of equation (1.8) will always produce branch-length correlations ranging from zero to one—even when the phylogeny is not ultrametric (i.e. the tips of the phylogeny are not aligned or contemporaneous). Limiting correlations to this range allows for the direct

comparison between non-phylogenetically corrected and phylogenetically-corrected effect sizes when $\mathbf{\Sigma}$ is used in GLS equations (see equation 1.1). In addition, it should also be noted that because BM and OU evolution both assume a relationship with time, then only ultrametric trees (chronograms) should be used for coding the phylogenetic correlation matrix—despite equation (1.8) serving as a control to standardize the matrix to yield a correlation matrix with ones on the main diagonal. If the tree is already ultrametric, then equation (1.8) is equivalent to standardizing (dividing) the elements of the **P** matrix by b^{total} (which should be the same for all taxa). Hereafter, I will assume that all correlation matrices will be standardized with equation (1.8). This standardization also fosters the cross comparison of phylogenetically corrected effect sizes based on different phylogenies because the correction is based on relative phylogenetic correlations (expected similarity) and not on relative divergence times among taxa.

Phylogenetically-Independent Mean Effect Sizes and Variances

A key purpose of meta-analysis is to take the weighted mean of effect sizes from a collection of *k* studies. To control for phylogenetic history when taking the weighted mean, then assume that each i^{th} study has an estimated effect size δ_i with a variance of $\sigma^2(\delta_i)$, and that this collection of studies is based on a diversity of taxa for which there is a hypothesized phylogenetic history. This phylogenetic history is defined in the correlation matrix **P** (see previous section). To generate an unbiased mean effect size of these studies, we want a) to weight effect sizes by sampling error (as in traditional meta-analysis), and b) to control for the phylogenetic non-independence of effect sizes (as in comparative analysis). Now that the covariance matrix $\mathbf{\Sigma}$ is modeled to account for heteroskedacity and phylogenetic correlations (either $\overline{\mathbf{P}}^{\text{BM}}$ or $\overline{\mathbf{P}}^{\text{OU}}$), we can estimate a phylogenetically-independent weighted mean effect size ($\overline{\delta}^{\mathbf{P}}$) with the

GLS regression equation

$$\bar{\delta}^{\mathbf{P}} = (\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{E}, \tag{1.9}$$

which has a variance $\sigma^2(\bar{\delta}^{\mathbf{P}})$ of

$$\sigma^2(\bar{\delta}^{\mathbf{P}}) = (\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}. \tag{1.10}$$

Here, the design $\mathbf{X}$ is a $k \times 1$ column vector of ones and $\mathbf{E}$ is a $k \times 1$ column vector of all the *k* number of effect sizes (δ).

The design matrix **X** is the same as in regression analysis, and can be used to integrate quantitative predictors, covariates or grouping variables into meta-analysis (akin to a regression or a one-way ANOVA). In the above use, the design matrix models the global mean of effect sizes. To include a covariate (*c*) into meta-analysis (e.g., a continuous variable like migration rates or body size), then *c* is treated as an additional column in **X** where

$$\mathbf{X} = \begin{bmatrix} 1 & \cdots & 1 \\ c_1 & \ldots & c_k \end{bmatrix}^{\mathrm{T}}. \tag{1.11}$$

Here, using this design matrix in the regression equation (1.7) will yield $\bar{\delta}^{\mathbf{P}}$ as a matrix containing the intercept and the regression coefficient of the meta-regression between effect sizes and the covariate. These regression statistics are weighted by variance and controlled for phylogenetic non-independence. In the medical sciences, integrating a covariate as in the above GLS model is referred to as meta-regression. Later, I further modify this design matrix to test evolutionary hypotheses with meta-analysis.

Reductions to the main diagonal and off-diagonal elements of $\boldsymbol{\Sigma}$ will also yield other important mean effect sizes. For instance, reducing all off-diagonal elements of $\boldsymbol{\Sigma}$ to zero, such that $\boldsymbol{\Sigma} = \mathrm{diag}[\sigma^2(\delta_1),\ldots,\sigma^2(\delta_k)]$, will treat effect sizes as independent and thus generate the traditional weighted mean effect size ($\bar{\delta}$) when applied to equation (1.7) (see Hedges and Olkin 1986). A second reduction of the main diagonal elements to ones will result in $\boldsymbol{\Sigma} = \mathbf{I}$, where **I** is an Identity matrix.

Applying **I** to equation (1.7) will yield a simple arithmetic mean ($\bar{\mu}$) of effect sizes (e.g., an unweighted non-phylogenetically corrected mean effect size). In all cases, $\mathbf{\Sigma}$ must be symmetric and positive definite (for details see Groβ 2003).

Finally, there are two ways to evaluate whether $\bar{\delta}^{\mathbf{P}}$ is non-zero (i.e. $\bar{\delta}^{\mathbf{P}} \neq 0$). The first is to determine whether 95% confidence intervals (CI) around $\bar{\delta}^{\mathbf{P}}$ overlap with zero. These 95% CI are calculated as

$$95\%\ \text{CI}\ [\bar{\delta}^{\mathbf{P}} - 1.96\sqrt{\sigma^2(\bar{\delta}^{\mathbf{P}})}; \bar{\delta}^{\mathbf{P}} + 1.96\sqrt{\sigma^2(\bar{\delta}^{\mathbf{P}})}]. \quad (1.12)$$

Using 95% CI has an added advantage in that they are useful to evaluate the statistical power of meta-analysis: broad 95% CI indicates a poor ability to detect a non-zero $\bar{\delta}^{\mathbf{P}}$ should it exist (see Nakagawa and Cuthill 2007). The second approach directly tests whether $\bar{\delta}^{\mathbf{P}} \neq 0$ using the following regression test statistic

$$Q_{\mathrm{R}}^{\mathbf{P}} = (\bar{\delta}^{\mathbf{P}})^{\mathrm{T}} \mathbf{X}^{\mathrm{T}} \mathbf{\Sigma}^{-1} \mathbf{X} \bar{\delta}^{\mathbf{P}}, \text{ where if } Q_{\mathrm{R}}^{\mathbf{P}} > \chi^2_{df=1} \text{ then } \bar{\delta}^{P} \neq 0. \quad (1.13)$$

This test is based on a χ^2 distribution with one degree of freedom (*df*), such that a significant $Q_{\mathrm{R}}^{\mathbf{P}}$ indicates that there is no evidence that the observed effect size deviates from zero.

Phylogenetically-Independent Homogeneity Tests

Homogeneity statistics (Q_{H}) in meta-analysis test whether effect sizes (δ) from a collection of *k* studies share a common effect (e.g., $\delta_1 = \delta_2 = \ldots = \delta_k$; see Hedges and Olkin 1986), and are used to evaluate whether the observed variation across studies is due to within-study sampling error. More generally homogeneity tests serve two purposes: 1) to determine if it is acceptable to pool a collection of effect sizes, and 2) to determine if moderator variables are useful to explain variation among effect sizes. The second purpose is vital for research synthesis in ecology because contrasting the homogeneity between and within groups of studies is the major approach to test

hypotheses with meta-analytical data (Gurevitch and Hedges 1999; Cooper and Hedges 1994).

Again using the GLS framework, a phylogenetically-independent homogeneity test ($Q_{\mathrm{H}}^{\mathbf{P}}$) can be calculated as follows

$$Q_{\mathrm{H}}^{\mathbf{P}} = \mathbf{E}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{E} - Q_{\mathrm{R}}^{\mathbf{P}}, \text{ where if } Q_{\mathrm{H}}^{\mathbf{P}} \leq \chi^2_{df=k-1} \text{ then } \delta_1 = \delta_2 = \ldots = \delta_k. \qquad (1.14)$$

Interpretation of a non-significant $Q_{\mathrm{H}}^{\mathbf{P}}$ is straightforward: should the test reveal that $\delta_1 = \delta_2 = \ldots = \delta_k$ then there is not enough evidence to indicate that the variation observed among studies is not due to sampling error. Thus pooling effect sizes with equation (1.9) will not yield a biased mean effect size. However, when $Q_{\mathrm{H}}^{\mathbf{P}}$ is significant then this suggests that variation among effects sizes (study heterogeneity) is due to other sources than sampling error or indicates that groups of effect sizes may share different variances—these are not mutually exclusive and in either case pooling effect sizes among all studies may yield a biased mean effect size (see Cooper and Hedges 1994).

To evaluate these sources of heterogeneity, two options are available. The first is to parse studies among different moderator groups in attempt to determine the possible causes of heterogeneity, such as mating system, trophic rank or geographic region. Another common application is to parse studies by taxonomic rank to determine whether effect sizes differ among these broad groups of organisms (see Lajeunesse and Forbes 2002). Given that the above analysis is already corrected for phylogenetic history, then other grouping variables, such as behavioral or ecological characteristics, will be more informative for evaluating sources of heterogeneity among effect sizes.

However, sometimes moderator groups are not useful for explaining heterogeneity, and an unbiased mean effect sizes is still needed to assess overall magnitude of effect. In this case, a random-effects model should be used to pool

effect sizes. The random-effects model assumes that effect sizes are random constants each with their own variance (Hedges 1992; Cooper and Hedges 1994). This differs from the above approach to calculating $\overline{\delta}^{\mathbf{P}}$ because the underlying assumption of equation (1.9) is that effect sizes are unknown but have a fixed or single variance. This is referred to as the fixed-effects model for pooling effect sizes. Thus finding significant heterogeneity among effect sizes (e.g., $Q_{\mathrm{H}}^{\mathbf{P}} > \chi^2_{d.f.=k-1}$) may indicate that the fixed-model does not fit the data because it has inappropriately assumed that effect sizes share a common experimental variance—again using equation (1.9) will yield a biased mean effect size. Assuming a random-effects model may be more appropriate when there is significant variation among effect sizes—in fact, many have argued that random effect models should be the only approach to summarizing research in ecology (see Gurevitch and Hedges 1999).

To calculate a pooled effect size under the random-effects model we need to estimate the between study variance (τ) as follows

$$\tau = \begin{cases} \dfrac{Q_\tau - df}{\mathrm{tr}(\boldsymbol{\Sigma}^{-1}) - \mathrm{tr}[(\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}(\boldsymbol{\Sigma}\boldsymbol{\Sigma})^{-1}\mathbf{X}]} & \text{if } Q_\tau > df, \\ 0 & \text{if } Q_\tau \le df, \end{cases} \tag{1.15}$$

where tr(**A**) is the trace of matrix **A**. The between study variance (τ) is set to zero when Q_τ is smaller than the degrees of freedom of the meta-analysis ($df = k - m$) because by definition the between study variance cannot be negative. Here m is the column rank of **X**. Finally, Q_τ is the residual sum of squares from the weighted regression given by

$$Q_\tau = \mathbf{E}^{\mathrm{T}}[\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}]\mathbf{E}. \tag{1.16}$$

Q_τ is also important to evaluate whether $\tau = 0$ and such that if $Q_\tau \le \chi^2_{df=k-1}$, then this would indicate that including the between study variance does not significantly

improve the pooled effect size estimate. Integrating the between-study variance in equations (1.1) and (1.2) is straightforward and involves adding τ to all the variances (main diagonal) in the $\mathbf{\Sigma}$ covariance matrix, such as

$$\hat{\sigma}^2(\delta_i) = \sigma^2(\delta_i) + \tau . \quad (1.17)$$

Hereafter, all mean effect sizes and variances estimated under the random-effects model will be designated with a hat (e.g., $\hat{\delta}^{\mathbf{P}}, \hat{\mathbf{\Sigma}}$). It is important to note that under the random effects model, homogeneity tests and important diagnostics of publication bias, such as the funnel plot (see Cooper and Hedges 1994) are ineffective because they assume a fixed-effect model. Assuming a random effect model thus assumes multiple variances and thus multiple funnels in the data.

A Diagnostic for Taxonomic Bias

A diagnostic that may be useful for meta-analysis is the relative weight of particular taxonomic classes in influencing the overall pooled effect size. One heuristic used to assess overall weight or taxonomic bias is to calculate the percentage of studies derived from a specific group. For example, 70% of *k* studies were based on invertebrates. The caveat in this example is that other taxonomic groups such as vertebrates are underrepresented in the literature, and thus the overall effect size across studies is biased towards invertebrate research. However, given that individual studies are weighted by their sample precision and corrected for their shared evolutionary history (closely related groups carry less weight in the overall analysis), then a percentage that accounts for these adjustments would me more useful in evaluating the overall bias across broad taxonomic classes of organisms.

Estimating the overall contribution of different taxonomic classes on the pooled effect size is possible because equation (1.9) contains the given weights based on sampling precision and phylogenetic correlations due to shared evolutionary

history—these weights are found in the weighting matrix ($\mathbf{\Sigma}^{-1}$). Partitioning these weights among taxonomic classes will give information on the overall contribution of these classes in influencing the pooled effect size. This estimate of overall contribution (quantified as the percentage of the overall weight on the pooled effect size) for *m* number of taxonomic classes (e.g., order or family) can be calculated as

$$\mathbf{W} = \mathbf{X}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\mathbf{Z})^{-1}, \tag{1.18}$$

where $\mathbf{W}$ is a $m \times 1$ column vector containing the percentage weight of each taxonomic class (where the sum of its elements will equal 100%), and $\mathbf{Z}$ is a column vector of $k \times 1$ size containing 0.01. This matrix serves the purpose for calculating percentages. Finally, $\mathbf{X}$ is the design matrix with *m* number of taxonomic classes, where each class is a column containing one if the taxa belongs to that class and zero otherwise. For example, if a meta-analysis has three taxa and two of which belong to the same genus then the $k \times m$ design matrix (here 3×2) should be as follows

$$\mathbf{X} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{1.19}$$

Much like when computing the pooled effect size estimate (equation 1.9), different percentages can be estimated based on the design of the covariance matrix $\mathbf{\Sigma}$. Assuming that $\mathbf{\Sigma} = \mathbf{I}$ will yield the weight of each taxonomic class if each taxa (effect sizes) are evenly weighted (e.g., are not weighted by sample precision or phylogenetic correlations). Following the example above, then the relative weights for the two genera are $\mathbf{W}^{\mathbf{I}} = [66.7, 33.3]^{\mathrm{T}}$. However, assuming as in traditional meta-analysis that $\mathbf{\Sigma} = \mathrm{diag}[\sigma^2(\delta_1), \ldots, \sigma^2(\delta_k)]$, will yield the overall weighting of each class relative to their sample precision, while assuming that $\mathbf{\Sigma} = \mathbf{\Sigma}^{\mathrm{BM}}$ will further adjust these weights based on their phylogenetic correlations. Finally, interpreting the percentages in $\mathbf{W}$ is straightforward: a high percentage indicates a greater weight of the effect sizes in that

taxonomic class when pooling effect sizes.

Detecting a Phylogenetic Signal among Effect Sizes

Testing for a phylogenetic signal is good practice in comparative analyses (Freckleton *et al.* 2002), and should be part of any meta-analysis that includes information on phylogenetic history. Detecting a phylogenetic signal is a test for phylogenetic conservatism (also known as phylogenetic inertia; see Wilson 1978), and determines the degree to which related taxa tend to be more similar than distantly related species. This degree of similarity influences the strength of phylogenetic correlations (phylogenetic dependence) among taxa (**P**). When traits show little phylogenetic conservatism, they are considered evolutionarily labile and appear randomly distributed among tips of the phylogeny.

Knowing the degree to which data are phylogenetically conserved allows for a more informative interpretation of phylogenetically-independent results (Björklund 1990): strong phylogenetic signal may indicate strong bias due to phylogenetic non-independence. Unfortunately one philosophy with comparative data is to wave a 'phylogenetically-controlled' analysis should data lack a phylogenetic signal (Westoby *et al.* 1995b, 1995c). This approach should be avoided given that failing to detect a signal may be more of a statistical issue than a biological one (see Martins 2000); for example, small phylogenies are less likely to show a signal than larger phylogenies (Freckleton *et al.* 2002). Later, I describe how to add a test of phylogenetic conservatism within this framework that sequentially fits different evolutionary models—this avoids making decisions on research directions based on significance testing.

To evaluate the degree of phylogenetic conservatism (λ) in effect sizes among a collection of taxa, I first describe a manual (heuristic) approach to fitting λ to the

phylogenetic correlation matrix (**P**), and then follow with a maximum likelihood (ML) method that optimizes λ to the data (Pagel 1994, 1997). Pagel (1994) provides a formal definition of phylogenetic conservatism as the degree to which data fit the BM model of evolution—that phylogenetic correlations among effect sizes are linearly related with time of divergence between the taxa for which they are based (see Fig. 1.1). As λ nears one ($\lambda \to 1$) then effect sizes are distributed phylogenetically as expected by BM, whereas when $\lambda \to 0$ then data are randomly distributed and appear phylogenetically uncorrelated (e.g., no conservatism). Here, λ is treated as a scaling factor that transforms the phylogenetic correlations among taxa from having no signal (no phylogenetic correlations) to having a full signal (full correlations as predicted by $\mathbf{P}^{\mathrm{BM}}$). Integrating λ to the phylogenetic correlation matrix is as follows:

$$\mathbf{P}_{i,j}^{\mathrm{BM}(\lambda)} = \begin{cases} b_i^{\mathrm{total}} & \text{if } i = j, \\ \lambda b_{i,j}^{\mathrm{shared}} & \text{if } i \neq j. \end{cases} \tag{1.20}$$

One way to evaluate the contribution of λ in biasing meta-analysis is to adopt a manual 't-shirt' approach. Here, a range of small (no correlation), medium and large (full correlation) values of phylogenetic conservatism (e.g., 0, 0.5, 1) are plugged into $\mathbf{P}^{\mathrm{BM}}$. These transformed phylogenetic correlation matrices ($\mathbf{P}^{\mathrm{BM}(\lambda)}$) are then applied to meta-analysis (e.g., starting with equation 1.1). For instance, using $\mathbf{P}^{\mathrm{BM}(\lambda=0)}$ to calculate a pooled effect size will generate the traditional weighted (non-phylogenetically corrected) mean effect size ($\bar{\delta}$).

Alternatively, λ can be optimized via maximum likelihood (ML). This method finds the best value of λ that minimizes the residual sums of squares (*SSE*) of the model, and provides more information on whether the phylogenetic conservatism of effect sizes tend towards zero or one. Under assumptions of normality among effect sizes (**E**), the least squares likelihood ($\mathcal{L}$) of λ is

$$\mathcal{L}[\lambda \mid \mathbf{E}] = \frac{e^{SSE[-2\sigma^2(\bar{\mu}_{\mathrm{BM}(\lambda)})]^{-1}}}{\sqrt{[2\pi\sigma^2(\bar{\mu}_{\mathrm{BM}(\lambda)})]^k \det(\mathbf{P}^{\mathrm{BM}(\lambda)})}}, \quad (1.21)$$

where,

$$\bar{\mu}_{\mathrm{BM}(\lambda)} = (\mathbf{X}^{\mathrm{T}}[\mathbf{P}^{\mathrm{BM}(\lambda)}]^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}[\mathbf{P}^{\mathrm{BM}(\lambda)}]^{-1}\mathbf{E}, \quad (1.22)$$

$$SSE = (\mathbf{E} - \mathbf{X}\bar{\mu}_{\mathrm{BM}(\lambda)})^{\mathrm{T}}[\mathbf{P}^{\mathrm{BM}(\lambda)}]^{-1}(\mathbf{E} - \mathbf{X}\bar{\mu}_{\mathrm{BM}(\lambda)}). \quad (1.23)$$

and $\sigma^2(\bar{\mu}_{\mathrm{BM}(\lambda)}) = SSE(k-1)^{-1}$. Finally, $\bar{\mu}_{\mathrm{BM}(\lambda)}$ is the pooled (unweighted) effect size with a variance of $\sigma^2(\bar{\mu}_{\mathrm{BM}(\lambda)})$ under the BM model. Since λ is mostly a scalar of $\mathbf{P}^{\mathrm{BM}}$, the ML estimate of λ (hereafter $\hat{\lambda}$) is found based on a linear optimization method that maximizes the expression of $\mathcal{L}$. For instance, this optimization could involve calculating equation (1.21) through (1.23) for each λ value ranging from 0.0001 to 1, where the smallest value of $\mathcal{L}$ (with the least amount of error) is chosen for $\hat{\lambda}$. Once $\hat{\lambda}$ is found, then it is applied to $\mathbf{P}^{\mathrm{BM}(\hat{\lambda})}$, which is then applied to the covariance matrix ($\mathbf{\Sigma}$) for meta-analysis. Irrespective of the approach taken (manual or ML), all of the above effect sizes statistics (e.g., pooled effect size, homogeneity test) will be fitted conditionally on λ or $\hat{\lambda}$.

Estimating the Contribution of Selection and Drift

In a previous section, I outlined how to model the phylogenetic covariance matrix (**P**) using both Brownian Motion evolution and an Ornstein-Uhlenbeck process. I now extend these models into a more formal evolutionary framework to evaluate the contribution of drift (η) and selection (β) in generating variation in research outcomes. These parameters are useful for meta-analysis because they will serve as the basis for testing neutral and adaptive hypotheses for the phylogenetic patterning of effect sizes. My description of these parameters in how they fit into meta-analysis will be purely applied—for further information on background and derivation see Hansen (1997) and Butler and King (2004). I will avoid theory when possible, and present

only the equations necessary to calculate directly evolutionary parameters under different models of evolution.

As described earlier under the OU model of evolution, drift and selection are two evolutionary parameters that are useful for describing the contribution of phylogenetic history on meta-analysis (see Figure 1.1). Here is it assumed that drift acts to push effect sizes away from an adaptive optima (e.g., a peak in a fitness landscape), while selection acts to counteract this movement away from this optima. In the absence of selection ($\beta \to 0$), the OU model collapses to a Brownian motion model—where the phylogenetic patterning of data is a product of an evolutionary process with random fluctuations (drift). Estimating the contribution of selection is similar to the approach taken with evaluating a phylogenetic signal (λ): this parameter can be fit via a manual ('t-shirt') or ML approach. For instance, a range of small, medium and large values of selection (e.g., β = 0.1, 1, 10) are incorporated into $\mathbf{P}^{\mathrm{OU}}$, and these phylogenetic correlation matrices ($\mathbf{P}^{\mathrm{OU}(\lambda)}$) are then applied to meta-analysis (e.g., starting with equation 1.3).

Another criterion for choosing estimates of selection is through maximum likelihood which finds the optimal value of selection that minimizes the SSE of the model. Again, all of the effect sizes statistics (e.g., pooled effect size, homogeneity test) will be fitted conditionally on selection. Assuming that effect sizes ($\mathbf{E}$) are normally distributed, then the likelihood of selection $\mathcal{L}[\beta \mid \mathbf{E}]$ is the same as equation (21). Unfortunately, a nonlinear optimization method is required to find the ML estimate of β (hereafter $\hat{\beta}$). This optimization problem occurs because selection does not fit linearly in equation (1.17)—the Ornstein-Uhlenbeck process models evolution under exponential terms (see equation (1.5) or Figure 1.1). The statistical language R provides a nonlinear optimization function to solve this issue.

Under the OU model, the intensity of drift (η) will change the rate at which

phylogenetic correlations are lost among taxa. This can be illustrated in Figure 1.1 as the slope of the full line (BM model) converging to a null slope with increasing drift. However, Hansen (1997) proposed a simple approach to estimating drift that is conditional on selection. Here, drift is not optimized to the data, but simply reflects the residual sums of squares (*SSE*) of a GLS model based on a particular estimate of selection. Future approaches should develop methods that jointly fit both drift and selection by ML. However for simplicity, I follow Hansen's (1997) method and consider only selection as the parameter to be optimized. Finally, drift is estimated as

$$\eta = \sqrt{\frac{SSE^{\mathrm{OU}(\hat{\beta})}}{k-m}}, \tag{1.24}$$

where m is the column rank of $\mathbf{X}$. It is important to note that theoretically the drift component should be smaller than selection; otherwise it would erase any signature that selection would leave on the phylogeny.

Moderator Variables as Adaptive Optima

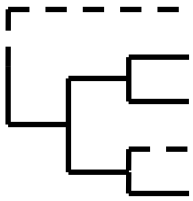
Moderator variables are important for evaluating bias and testing hypotheses with meta-analysis (Cooper and Hedges 1994). Here, I extend the scope of meta-analysis to test evolutionary hypotheses by treating moderator groupings as adaptive explanatory variables for variation in research outcomes. This is possible because the OU model for evolution assumes that traits (more specifically adaptations) are maintained near fitness optima through selection (Hansen 1997), and that the strength of this selection determines the rate at which taxa evolve from the ancestral to the primary (contemporaneous) optima (Figure 1.1). Thus selection acts to release ancestral constraints on adaptation (e.g., similarity due to shared ancestry) by pushing taxa to evolve towards a new hypothesized optimum. By hypothesizing different optima (moderator variables) throughout the history of phylogeny, we could estimate

the effect of a selective factor on the position of these optima. The strength of this selective factor is used to assess the degree to which moderator variables can serve as adaptive explanations for the phylogenetic patterning of effect sizes among taxa.

A way to model the contribution of optima on selection is by modifying the design matrix (**X**) to include multiple columns—where each column represents a different hypothesized optimum (θ), and where the elements of each column will contain the time spent within each optima relative to selection (see example in Figure 1.2). In terms of GLS modeling, this approach is similar to modifying the design matrix to conduct a one way ANOVA among multiple moderator groups. However, here the relative weight of each moderator group will vary depending on the strength of selection towards the contemporaneous optimum, and the time spent in this optimum. Using this modified **X** matrix in equation (1.3) will yield weighted averages among moderator groups; the caveat here is that more weight will be given for the taxa that have evolved under the primary optima. Estimating selection (β) via the ML equation (1.21) with this new design matrix will determine how well this moderator variable serves as a good explanation for data—where these moderator groups did not significantly contribute to selection (e.g., $\beta \to 0$) then these moderator groups are deemed not adaptively significant in the evolution of taxa.

To generate an OU model with *m* number of selective optima (θ) we can specify the design matrix **X** as follows. The $k \times m$ design matrix (**X**) is used to weight effect sizes based on their optima—again taxa with the primary optima will have their effect sizes weighted more heavily than non-optimal states taxa. The sum of the weights of the optima are determined for each taxon—these are essentially modeled as the time spent throughout the history of the clade in a particular state. An illustration of these optima on a phylogeny can be found in Table 1.2. For instance, each element of the design matrix **X** can be generalized for *m* number of evolutionary states (θ) as

a)



b)

$$\begin{bmatrix} 1-e^{-\beta b^{\theta}} & e^{-\beta b^{\theta}} \\ 0 & 1 \\ 0 & 1 \\ 1-e^{-\beta b^{\theta}} & e^{-\beta b^{\theta}} \\ 0 & 1 \end{bmatrix}$$

c)

$$\begin{bmatrix} 1-e^{-\beta (b^{\text{total}}-b^{\theta})} & e^{-\beta (b^{\text{total}}-b^{\theta})} \\ 1-e^{-\beta b^{\text{total}}} & e^{-\beta b^{\text{total}}} \\ 1-e^{-\beta b^{\text{total}}} & e^{-\beta b^{\text{total}}} \\ 1-e^{-\beta (b^{\text{total}}-b^{\theta})} & e^{-\beta (b^{\text{total}}-b^{\theta})} \\ 1-e^{-\beta b^{\text{total}}} & e^{-\beta b^{\text{total}}} \end{bmatrix}$$

Figure 1.2. Two examples of coding the design matrix (**X**) to test different evolutionary hypotheses of the impact of different moderator variables (evolutionary optima) on the phylogenetic patterning of effect sizes. The phylogeny **a)** has two evolutionary optima: taxa designed with a dashed line have the primary optima, whereas the remaining taxa have the secondary optima (full line). The design matrix **b)** is coded with the assumption that the primary optima is derived from the secondary (ancestral optima) with an unknown origin ($\theta_2^{\text{start}} = \infty$). The second design matrix **c)** assumes that there is only one optimum such that taxa without this state evolve from an unknown but rooted ancestral state ($\theta_2^{\text{start}} = b^{total}$).

$$\mathbf{X}_{i,m} = e^{-\beta\theta_i^{\text{end}}} - e^{-\beta\theta_i^{\text{start}}}, \tag{1.25}$$

where for each i^{th} taxon, θ^{start} and θ^{end} is the start and end of a period in the phylogenetic history occupied by this optimum. Let b^{θ} designate the sum of phylogenetic branch-length distances between θ^{start} and θ^{end}. By modifying b^{θ} we could give differential weight to taxa with an optimal state compared to taxa with the original (ancestral) state.

Now let us consider two hypotheses on the influence of the moderator groups in the evolution of their optima. In the first we will treat each optimum separately with an unknown ancestral state, and in the second test we will assume two optimal states where one functions as the primary (derived) and the other as the secondary (ancestral) state. When there is single optima the corresponding row vector for the i^{th} taxa in $\mathbf{X}$ is $\mathbf{x}_{i,m}^{\theta=1} = [1 - e^{-\beta(b_i^{\text{total}} - b_i^{\theta})}, e^{-\beta(b_i^{\text{total}} - b_i^{\theta})}]$, whereas all other row vectors are $\mathbf{x}_{i,m}^{\theta=0} = [1 - e^{-\beta b_i^{\text{total}}}, e^{-\beta b_i^{\text{total}}}]$. These row vectors are simplified versions of equation (1.25), where for example $\mathbf{x}_{i,1}^{\theta=1} = e^{-\beta(b_i^{\text{total}} - b_i^{\text{total}})} - e^{-\beta(b_i^{\text{total}} - b_i^{\theta})}$. These row vectors are then inserted into the design matrix

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_{1,1} & \dots & \mathbf{x}_{i,m} \end{bmatrix}^{\text{T}} \tag{1.26}$$

However, when there are two optima, where one is the primary (derived) and the other is secondary (ancestral) then taxa with the primary optima will have a row vector defines as $\mathbf{x}_{i,m}^{\theta} = [1 - e^{-\beta b_i^{\theta}}, e^{-\beta b_i^{\theta}}]$, and the remaining rows (with ancestral optima) as $\mathbf{x}_{i,m} = [0, 1]$. It is important to note that the sum of elements of each row vector of $\mathbf{X}$ must equal one, such that the global sum of all the elements of $\mathbf{X}$ will equal *k*. A final model can assume that the ancestral optimum is unknown, and that each *m* hypothesized optimum is derived from this unknown state. Here an additional column in $\mathbf{X}$ will designate this unknown optima, and all the elements of this column will equal $e^{-\beta b_i^{\text{total}}}$, the remaining columns will include the distances relative to the primary/secondary optimum to this unknown (e.g.,

$\mathbf{x}_{i,m}^{\theta} = [e^{-\beta b_i^{\text{total}}}, e^{-\beta b_i^{\text{total}}} - e^{-\beta b_i^{\theta}}, e^{-\beta b_i^{\theta}}]$; see Butler and King 2004). With these new design matrices that assume different optimal designs, the maximum likelihood equations (1.21) through (1.23) are used to estimate β.

Selecting Among Competing Evolutionary Models

The above statistical framework generates multiple evolutionary hypotheses on the phylogenetic patterning of effect sizes—from simple neutral models of evolution to more elaborate adaptive explanations with multiple adaptive optima. These models have a nested design based on an Ornstein-Uhlenbeck process (Hansen 1997; Butler and King 2004), and serve as competing hypotheses on how effect sizes evolve. Here I describe a model selection approach that evaluates the relative fit of these competing hypotheses using Akaike's information criterion (AIC). This approach differs from the more common application of likelihood-ratio tests (LRT) to contrast different evolutionary hypotheses in comparative analysis (see Pagel 1997). AIC scores will be more useful for evolutionary meta-analysis given the large number of evolutionary models that could be generated using an OU process. AIC scores also avoid statistical issues relating to the sequential testing of multiple hypotheses within a nested design: such as making assumptions on which model will serve as the null hypothesis and subsequent non-independence of contrasts due to shared null hypotheses (see Cohen 1992).

Each model predicts a different hypothesis on phylogenetic correlations among effect sizes (as indicated by **P**), and form a nested hierarchy illustrated as follows:

$$\mathbf{P}^{\text{I}} \overset{\lambda}{\subset} \mathbf{P}^{\text{BM}} \overset{\beta}{\subset} \mathbf{P}^{\text{OU}} \overset{\beta(m-1)}{\subset} \mathbf{P}^{\text{OU}(m-1)} \subset \cdots \overset{\beta(m)}{\subset} \mathbf{P}^{\text{OU}(m)}, \tag{1.27}$$

where $\subset$ indicates that the left hypothesis is a subset of the one to the right, and where the symbols above $\subset$ are evolutionary parameters (λ and β) used to parameterize the fit of the proceeding model. To evaluate the relative fit of these models, an AIC score

is estimated for each model with the following

$$\mathrm{AIC} = 2m - 2\ln(\mathcal{L}). \quad (1.28)$$

Here the likelihood estimate ($\mathcal{L}$) of each GLS model (see equation 1.21) will form the basis for model selection. This approach penalizes models with high error (low fit) in describing the data and models that use additional evolutionary parameters to describe data (e.g., $\mathbf{P}^{\mathrm{OU}(m)}$). The model that best fits the effect sizes is that with the smallest AIC score. For instance, should $\mathrm{AIC}(\overline{\mu}) < \mathrm{AIC}(\overline{\delta}^{\mathrm{BM}(\lambda=1)})$, then pooling effect sizes with a model that assumed phylogenetic correlations based on BM evolution was less effective is minimizing statistical error than a model lacking information on evolutionary history.

An Illustrative Example

Virgin males and Female Reproductive Output

Here I revisit Torres-Vila and Jennions (2005) meta-analysis on whether females have greater reproductive output if they mate with virgin males. This study is a good example where an evolutionary meta-analysis could have been useful to integrate phylogenetic information and test hypotheses because a) the authors found a family-level effect when testing for taxonomic bias, and b) found that this effect related to the mating history of the taxa (e.g., among polyandrous and monandrous lepidopterans). I first begin by re-analyzing their data by pooling effect size and heterogeneity tests, followed by an evolutionary test of their hypothesis that male mating history will have a stronger effect in polyandrous than monandrous species.

Phylogeny Construction and Analysis

A necessary aspect of meta-analysis is reporting the inclusion/exclusion criteria used to collate the group of studies for synthesis (Cooper and Hedges 1994). This helps

identify sources of bias that can affect the representation of studies included in meta-analysis. This philosophy of transparency should also crossover to the approaches and methods used to collate data for phylogenetic tree construction. Below is a sketch of my inclusion/exclusion criteria for building a tree using genetic sequence data from GenBank. My approach is likely the most exclusive way of building a tree because the following meta-analysis will be based solely on taxa with available genetic data. However, for the purposes of this paper, my approach is useful to a) draw attention to taxonomic and genetic biases in public databases, and b) to generate a phylogeny with relative branch-length distances, which is important for testing evolutionary hypotheses with meta-analysis.

Partial sequence data useful for phylogenetic construction were available for 84% of 25 lepidopteran species analyzed in the original meta-analysis ($N = 25$; see Table 1.3). These data were distributed across eight genes (COI, COII, 12S rRNA, 16S rRNA, 28S rRNA, Cyt B, NADH, EF-1 alpha), but no one gene was available for all species. A more involved analysis could make use of these data; however for simplicity I limit my analysis to the mtDNA COII gene (albeit only available for 53% of the 25 species). This gene best resolved the topology of taxa following known published relationships (Nylin *et al.* 2001). In addition, COII data were not available for *Zeiraphera canadiensis*, but information existed for another species within the same genus (*Z. diniana*). I thus substituted information between these species, given that there were no other taxa from that genus included in the original meta-analysis. These fourteen species belong to the suborder Ditrysia, and additional two species from a sister suborder Incurvarioidea were chosen as outgroups: *Prodoxus gypsicolor* (acc. AF150920) and *Greya variabilis* (AF150909). All sequences were aligned with ClustalW (Larkin *et al.* 2007) and then visually inspected for consistency. This analysis found that GenBank data from *Diatraea considerate* did not align well

Table 1.3. Lepidopteran species synthesized in Torres-Vila and Jennions (2005) meta-analysis. These species are grouped as moths (Heterocera) or butterflies and skippers (Rhopalocera), and by mating system (polyandrous or monandrous). Also presented are the original effect size data (δ = Hedges' d) and variances $\sigma^2(d)$ from Torres-Vila and Jennions (2005), and the GenBank accession numbers.

Species	Family	type	mating system	Hedges' d	$\sigma^2(d)$	COII accession no.
Busseola fusca	Noctuidae	H	P	0.469	0.082	AY320474
Chilo partellus	Crambidae	H	M	-0.041	0.052	AY320482
Choristoneura fumiferana	Tortricidae	H	M	0.137	0.048	L19098
Choristoneura rosaceana	Tortricidae	H	P	1.028	0.037	L19099
Colias eurytheme	Pieridae	R	P	1.013	0.205	AF044024
Helicoverpa armigera	Noctuidae	H	P	0.304	0.128	DQ059302
Jalmenus evagoras	Lycaenidae	R	M	0.366	0.071	DQ456502
Ostrinia nubilalis	Crambidae	H	M	0.359	0.017	AF321880
Papilio glaucus	Papilionidae	R	P	0.232	0.044	EF126474
Papilio machaon	Papilionidae	R	M	0.251	0.155	AY457593
Pieris napi	Pieridae	R	P	1.169	0.255	AF170861
Trichoplusia ni	Noctuidae	H	P	0.263	0.075	AB158623
Zeiraphera canadiensis	Tortricidae	H	M	0.016	0.067	DQ241506

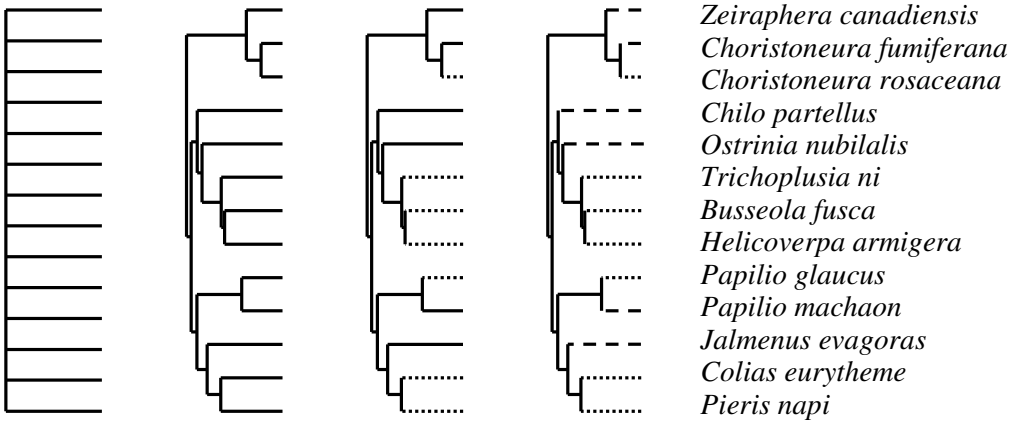
(< 20% alignment) with other sequences and thus this species was excluded from phylogenetic construction.

My final dataset included thirteen species with meta-analytical data plus two additional species that served as outgroups for phylogenetic construction. A Modeltest (v. 3.5; Posada and Crandall 1998) analysis using AIC selection criteria determined that a GTR+I+G was the best nucleotide substitution model for these data, and this model was applied to a ML analysis. Here, I performed a heuristic search with TBR branch swapping, a stepwise addition starting tree, and with the ASIS stepwise addition option. This ML analysis was performed with PAUP* 4.0d10 (Swofford 2001) with a molecular clock assumption. Although these analyses were not calibrated against real time, the phylogeny still describes the relative temporal orderings of nodes (and relative intervening branch-lengths), which again are necessary to test evolutionary hypotheses.

Using this phylogenetic hypothesis with meta-analytical data, I then performed an evolutionary meta-analysis to pool effect sizes across taxa using the following models: traditional (weighted) meta-analysis (labeled as N for normal); two BM models where the first was transformed with the ML estimate for phylogenetic conservatism ($\lambda = \text{ML}$) and the second model assuming full phylogenetic correlations ($\lambda = 1$); and finally an OU model based on the ML estimate for selection ($\beta = \text{ML}$). I then repeat these analyses using monandry and polyandry as moderator variables.

I further treated these moderator variables as hypothesized adaptive optima using two multi-optima OU models (see Table 1.4). The first model assumed that the monandry is the primary optimum that selection will act to maintain, and polyandrous taxa are derived. I only report the results for this direction in the evolution of optima because opposite ordering would yield similar results. My second and most complex model assumes that both monandry and polyandry are separate primary optima

Table 1.4. Alternative phylogenetic models for effect size evolution ranging from no phylogenetic correlation, Brownian Motion (BM), to an Ornstein-Uhlenbeck process (OU) with two and three hypothesized adaptive optima. The dotted or dashed lines indicate the start and end of the primary optima for a particular mating system adaptive regime and the full line indicates the ancestral optima. In the two-optima model, the monandrous species have evolved under the primary optima whereas polyandrous species have the ancestral optima; this model is equivalent to a model that hypotheses the opposite where monandrous species have the ancestral state. The three-optima model hypothesizes that both monandrous and polyandrous species (dashed line) are separate primary regimes that are derived from shared but unknown ancestral optima. Note that an OU model with one optima is a BM model (BM~OU).

evolutionary model				
		OU		
no correlation	BM	2-optima	3-optima	species
				*Zeiraphera canadiensis* *Choristoneura fumiferana* *Choristoneura rosaceana* *Chilo partellus* *Ostrinia nubilalis* *Trichoplusia ni* *Busseola fusca* *Helicoverpa armigera* *Papilio glaucus* *Papilio machaon* *Jalmenus evagoras* *Colias eurytheme* *Pieris napi*

derived from a third unknown adaptive optimum. This latter model is important as it assumes that the origin of the mating system is unknown. Finally, for all GLS models I calculated homogeneity tests and AIC scores for model selection; these statistics will serve to assess the fit of these competing phylogenetic models on evolutionary meta-analysis.

Results

As in the original meta-analysis by Torres-Vila and Jennions (2005), females mated to virgin males had greater reproductive output than females mated with non-virgins among 13 lepidopteran species (Table 1.5). A homogeneity test revealed significant variation among these 13 taxa under a fixed effect model (Table 1.5), but this variation was largely removed by parsing taxa into monandrous ($k = 7$) or polyandrous ($k = 6$) mating systems (Table 1.5). Thus for simplicity, I will continue to assume a fixed effect model. In addition, females of polyandrous taxa tended to have greater reproductive output than monandrous females when mated to virgin males (95% CI of the two groups did not overlap; Table 1.5). This latter result was only marginal in the original meta-analysis and was dependent on the further parsing of taxa by Rhopalocera (moths) and Heterocera (butterflies) taxonomic ranks (see Torres-Vila and Jennions 2005).

Accounting for the evolutionary history of lepidoptera did not significantly change the overall meta-analysis: female reproductive output is strongly affected by male mating history (Table 1.4 and 1.5). The GLS model with the best AIC score was the traditional (non-phylogenetically corrected) meta-analysis; whereas the meta-analysis with full phylogenetic correlations (BM where $\lambda = 1$) was the least effective in minimizing error during pooling effect sizes. This was expected given that the maximum likelihood (ML) estimate of phylogenetic conservatism of effect sizes (λ)

Table 1.5. Results from integrating phylogenetic data into Torres-Vila and Jennions (2005) meta-analysis on female mating success with virgin or non-virgin males. Here k is the sample size, and parenthicals for GLS models are the values of evolutionary parameters: phylogenetic signal (λ), selection (β) and drift (η). Parameters were also optimized via maximum likelihood (ML).

			pooled effect size			homogeneity test			evolutionary parameters			
GLS evolutionary model		k	$\bar{\delta}$	UCI	LCI	Q_H	d.f.	p	λ	β	η	AIC
all studies												
N		13	0.374	0.243	0.506	23.3	12	<0.001	0.0000	-	-	13.58
BM (λ = ML)		13	0.374	0.243	0.506	23.3	12	<0.001	0.0001	-	-	15.58
BM~OU ($\beta = 0, \lambda = 1$)		13	0.398	0.228	0.568	63.4	12	<0.001	1.0000	0.0	0.53	25.99
OU (β = ML)		13	0.374	0.243	0.506	23.3	12	<0.001	-	74.0	4.59	17.58
by mating-system												
monandrous (N)		7	0.225	0.056	0.394	3.5	6	0.621	-	-	-	10.42
polyandrous (N)		6	0.604	0.395	0.814	12.1	5	0.059				
monandrous (BM~OU, $\beta = 0, \lambda = 1$)		7	0.171	–0.015	0.356	3.6	6	0.604	1.0000	0.0	0.37	18.97
polyandrous (BM~OU, $\beta = 0, \lambda = 1$)		6	0.742	0.413	0.840	11.7	5	0.069				
2 optima OU (β = ML)	monandrous	7	0.223	0.049	0.397	3.5	6	0.621	-	100.0*	4.24	16.39
	polyandrous	6	0.605	0.395	0.815	12.1	5	0.059				
3 optima OU (β = ML)	monandrous	7	0.206	0.022	0.385	3.4	6	0.757	-	77.7	3.63	18.14
	polyandrous	6	0.605	0.395	0.814	12.1	5	0.059				
	unknown ancestor	-	6.709	-	-	-	-	-				

* Selection could not be optimized and thus indicates the maximum value attainable following ML constraints. (e.g., min = 0.0001 and max = 100).

was near zero (Table 1.5), and the ML estimate of selection (β) from the OU model was very high. Both evolutionary parameters here removed all phylogenetic correlations from meta-analysis, thereby modeling the same covariance matrix as traditional meta-analysis (e.g., $\mathbf{\Sigma}^{\mathrm{OU}(\beta=\mathrm{ML})} \cong \mathbf{\Sigma}^{\mathrm{BM}(\lambda=\mathrm{ML})} \cong \mathrm{diag}[\sigma^2(\delta_1),\ldots,\sigma^2(\delta_k)]$; see Dinoz-fihlo 2001). These evolutionary models were also penalized for modeling traditional MA with evolutionary parameters (see AIC scores in Table 1.5).

A GLS model that included mating system as a moderator variable had a higher AIC score than a model lacking this variable (Table 1.5). This effect is also indirectly observed through the homogeneity tests: parsing species by mating system significantly removed all within study heterogeneity among effect sizes (Table 1.5). Thus grouping species under monandrous and polyandrous mating system explained much of the variation in effect sizes across the 13 studies. However, as in the pooled analysis across all species, the evolutionary model with full phylogenetic correlations (BM where $\lambda = 1$) was least effective in explaining the patterning of effect sizes.

Modeling mating systems as evolutionary optima also did not improve the fit of GLS models and indicated again that selection was strong—yielding an evolutionary meta-analysis that was effectively the same traditional meta-analysis (Table 1.5). These models with hypothesized adaptive optima had poor AIC scores because they explained the same amount of information as traditional meta-analysis (with a moderator variable) but with additional (ineffective) evolutionary parameters. Models with strong selection point again that mating system serves an important explanatory variable for female reproductive output—should the ML estimate of selection had been weak, then this effect would have indicated that hypothesizing mating systems as evolutionary optima did not serve as a good explanation for the patterning effect sizes.

Should modeling meta-analysis with phylogenetic information have resulted in

a better fit to the data, then analyses would have only had a small marginal effect of the weighting of taxa belonging to either Rhopalocera (moths) and Heterocera (butterflies) taxonomic ranks (Table 1.6). Perhaps this indicates that the significant effect among mating systems when grouping taxa among moths and butterflies in the original meta-analysis was not a product of having accounted for the shared ancestry among these classes. The significant effect may have been an epiphenomenon of some other unknown moderating characteristic that is closely aligned with these two taxonomic classes. Models with phylogenetic information would have also given more weight to two species from the grass moth family (Crambidae): *Chilo partellus* and *Ostrinia nubilalis* (Table 1.6). These species would be weighted more heavily in any analysis, because they have the greatest shared mean branch-length distance to all other species (see phylogeny in Table 1.4), and are thus the least phylogenetically correlated with other lepidopterans.

Discussion

Phylogenetic non-independence cannot be ignored given the current role of meta-analysis in explaining contingency in research of ecological and evolutionary concepts, and in the systematic appraisal of conservation and management practices (Casey and Myers 1998). The aim of this paper is to provide statistical methods to account for phylogenetic bias in meta-analysis and to improve inferences based on research drawn from multiple taxa. In addition, a model selection approach is emphasized to contrast competing evolutionary hypotheses on the phylogenetic patterning of effect size (sensu Bulter and King 2004). This statistical framework integrates across comparative and meta-analytic approaches based on generalized least squares theory (Hedges 1991; Pagel 1994).

For the illustrative example on female reproductive output when mating with

Table 1.6. The overall weights (**W**) of different taxonomic ranks on different the mean effect size pooled across 13 studies (see Table 1.5). These percentages (%) are derived from a raw average ($\bar{\mu}$) of effect sizes (where each study has equal weight), meta-analysis ($\bar{\delta}$; weighted by sample precision), and phylogenetically corrected meta-analysis ($\bar{\delta}^{\mathbf{P}(\mathrm{BM})}$; weighted by sample precision and phylogenetic correlations based on a BM model of evolution assuming $\lambda = 1$).

taxonomic rank		k	% taxonomic weight (**W**) on pooled effect size (k = 13)		
			$\bar{\mu}$	$\bar{\delta}$	$\bar{\delta}^{\mathbf{P}(\mathrm{BM})}$
division	Heterocera	8	61.6	76.5	77.2
	Rhopalocera	5	38.4	23.5	22.8
Family	Crambidae	2	15.4	35.2	48.6
	Lycaenidae	1	7.6	6.3	7.5
	Noctuidae	3	23.1	13.0	7.4
	Papilionidae	2	15.4	13.2	13.9
	Pieridae	2	15.4	4.0	1.4
	Tortricidae	3	23.1	28.3	21.2

virgin males, competing evolutionary models were less effective in fitting effect size data than a model lacking phylogenetic information. There are several explanations for this lack of fit. First, the phylogenetic composition of taxa form non-random paraphyletic (e.g., only certain lepidopteran families) or polyphyletic groups (e.g., only taxa where there is published research). This phylogenetic composition is likely the outcome of taxonomic and publication bias in research—where taxa from model systems are more likely to show significant positive results because more information is known to control for experimental bias (Clark and May 2002). For instance, many of the taxa included in the meta-analysis are well studied agricultural pests (see Table 1.2). Testing evolutionary hypotheses with a non-random (phylogenetic) sample will likely bias the ability for statistical tests to detect evolutionary effects should they exist—further exacerbating known biases such as sample size (Freckleton *et al.* 2002) and data type (e.g., structural versus behavioral). Given that the associated error due to taxon sampling and publication bias is largely unknown for evolutionary meta-analysis, a conservative approach that explicitly recognizes phylogenetic relationships should be used over approaches based on phylogenic elimination (see Westoby *et al.* 1995b; 1995c). Further exploration of evolutionary hypotheses using meta-analysis should not be limited if statistics fail to reveal a phylogenetic signal among effect sizes. Given that authors of reviews are unable to experimentally manipulate the number of studies for meta-analysis—a more conservative application of this evolutionary approach should reflect this limitation.

A second issue that will affect the ability for evolutionary hypotheses to explain variation in research outcomes is availability of phylogenetic information to connect evolutionary relationships among taxa. One source of bias in information is when published phylogenies are selective in the taxa used for analysis (e.g., using representative samples from a genus), and how they report this information in public

databases (e.g., Genbank). I found that only partial sequence data were available for 53% of the species included in the original meta-analysis—albeit my inclusion criterion was narrow and limited to one gene. This subsampling of the original meta-analytical data resulted in a significant mating system effect that was only detected in the original study after correcting for taxonomic ranks, and likely decreased the power to detect a phylogenetic signal given the small sample size (Table 1.3). An alternative approach could have assembled taxa based on published relationships, or perhaps used information on Linnean rankings. Freckleton *et al.* (2002) found that even phylogenies with course information can improve the description of the data and improve statistical inferences. This approach would have increased the sample size for evolutionary meta-analysis, but information on the relative branch-length distances among hypothesized divergences would be missing. This disadvantage would constrain the diversity of evolutionary hypotheses that could be used to explain the patterning of the data. If possible, both traditional meta-analysis and evolutionary meta-analysis (presumably analyzed with a smaller sample size) should be reported in concert to improve inferences of quantitative reviews.

Dissecting the composition of effect sizes can also provide information on the outcomes of evolutionary meta-analysis. For instance, the effect sizes in Torres-Vila and Jennions (2005) meta-analysis estimate the difference in lifetime fecundity between females mated with virgin and non-virgin males. Among taxa were raw data was available, both the lifetime fecundity of females mated with virgin and experienced males had strong (non-zero) phylogenetic signals: fecundity with virgins ($\hat{\lambda} = 0.71$) and with experienced males ($\hat{\lambda} = 0.8$). These strong signals match empirical data that fecundity is often highly constrained by body size (Honěk 1993) and that size is phylogenetically conserved among Lepidoptera. In addition, there is also a shared correlated evolutionary history between testis size and body size among

Lepidoptera (Gage 1996). Presumably this degree of trait conservatism would also be conserved experimentally across taxa, but the raw difference in fecundity among females mated with virgin or experienced males was not phylogenetically conserved ($\hat{\lambda} = 0.0$).

These results suggest that the mating difference in reproductive output of females is evolutionary labile and diverges rapidly and independently from constraints imposed by evolutionary history. This perhaps is expected given that mating with virgin and non-virgin males has a direct effect on fitness, and that the evolutionary meta-analysis found strong selection for increased fecundity when mating with virgin males (Table 1.5). This selection would erase the contribution of phylogenetic correlations due to shared ancestry. However, there is some evidence to indicate that mating system can serve as a constraint given the significant effect of parsing studies as monondrous and polyandrous groups (Table 1.5). However, whether differences in mating system serve as an evolutionary constraint was unclear (Table 1.5).

Polytomies in Trees: A Potential Bias for Homogeneity Tests

An additional bias can occur when there are polytomies in the phylogenetic tree. Polytomies are tree nodes that have more than two immediate descendents, and Madisson (1989) divides these into two types: one biological and the other statistical. The first is referred to as a 'hard' polytomy, and depicts a true biological event where a group of sister taxa have diverged (speciated) simultaneously. Examples of hard polytomies exist (see Walsh *et al.* 1999), but appear uncommon, and may indicate explosive (rapid) speciation events, simultaneous fragmentation of populations, or introgressive hybridization and recombination events (Hoelzer and Meinick 1994). Polytomies more commonly occur because of statistical uncertainty when resolving relationships among taxa; these are known as 'soft' polytomies. Statistical approaches

to building phylogenetic trees (e.g., Maximum Likelihood) will generate soft polytomies when they are unable to statistically break-up true speciation or bifurcation events among sister taxa.

Polytomies may influence the type I error rates of homogeneity tests because they influence the degrees of freedom (*df*) of the analysis (Purvis and Garland 1993). The degrees of freedom are a measure of the number of independent pieces of information ($k-1$ for a completely resolved tree), and polytomies in phylogenetic trees cancels some of this independence because multiple lineages are specified with shared divergences. The correct *df* for $Q_{\mathrm{H}}^{\mathbf{P}}$ tests should reflect this imprecision in the phylogenetic tree. One approach to correct the *df* is to conservatively bind the *df* to minimize type I errors (false positives outcomes). Purvis and Garland (1993) suggest a conservative *df* to be $k-p$. Here, p is total number of polytomies in the phylogeny. Should the significance of $Q_H^{\mathbf{P}}$ be lost after correcting the *df*'s for polytomies, then all conclusions drawn from homogeneity tests are inconclusive given the current resolution of the phylogeny.

Alternatively, Purvis and Garland (1993) suggest arbitrarily resolving polytomies in a tree (with the branch-length of new intervening nodes set to a very small number), where analyses are repeated and averaged across these resolutions. Unfortunately, the GLS framework summarizes phylogenetic relationships using a correlation matrix (**P**) based on branch-length distances. Consequently, arbitrary resolution of polytomies does not significantly change the branch-length distances between species (see Rolf 2004; Diniz-Filho 2001), and thus analyses may be insensitive to these resolutions. However, this approach could be used should the branch-length distances among taxa be arbitrarily set. This can occur when information on the topology of the phylogeny is available but intervening branch-length distances are not. See Purvis and Garland (1994) for multiple procedures to

arbitrarily set the branch-length distances of phylogenies.

Finally, including multiple studies based on the same taxon is also a challenge for meta-analysis. These studies are important because they serve as species-level replication, and their inclusion in meta-analysis will improve variances estimates of research outcome (Cooper and Hedges 1994). One approach could place these additional studies as a polytomy on the phylogeny. However, given the way polytomies are modeled under the GLS framework (as very strong phylogenetic correlations; $\mathbf{P}_{i,j} \rightarrow 1$), then this approach will likely significantly underweight this group in the final meta-analysis. This weighting occurs because uncorrelated groups are given precedence over correlated groups (see Table 1.6). I suggest pooling all the effect sizes derived from a single species before evolutionary meta-analysis—such that only a single effect size estimate is used for each species in the final analysis. Future research on evolutionary meta-analysis should explore the effects of polytomies on the precision of research synthesis tools.

Beyond Phylogenetic Non-Independence

Correlations used to estimate distances among effect sizes need not be limited to phylogenetic information—experimental, genetic, temporal and spatial/geographic correlations (see Adams 2008) should also be applied to the GLS framework presented in this paper. One application could be to control for the overrepresentation of single studies due to extracting multiple effect sizes from a single experimental design (Rosenthal 1994). Here effect sizes are not independent because many of the experimental comparisons are based on multiple traits measured from the same individuals. Should a study report information on how these traits are correlated (i.e. Pearson product moment correlations), then a within-study pooled effect size (controlling for non-independence among treatment effects) can be estimated using

equation (1.2) with a **P** matrix containing the correlations among traits.

A similar approach can be applied to calculate within study pooled effect size from studies based on repeated measure designs. For instance, the duration of measurement intervals in a time series analysis will dictate the independence of data, assuming that data measured in rapid intervals are more likely to be similar. The duration among intervals (as a linear distance) can be treated as correlations in the **P** matrix, where each effect size (drawn from each repeated measure) can be pooled based on these assumed temporal correlations. Simpler correlation structures could also be assumed: such as when measurements are only dependent on the previous measurement (e.g., strict autocorrelation). This avoids treating each repeated measure as a separate effect size or having to apply strict selection criteria that could loose information (e.g., extracting only the final endpoint as the study's effect size).

Conclusions and Prospectus

Meta-analysis is a retrospective endeavor—and the lessons learned from synthesizing published research should serve as a stepping point for future experiments. Testing evolutionary hypotheses with meta-analysis can help identify phylogenetic effects on research outcomes, and uncover adaptive explanations for mediating factors influcing research outcomes. This information should emphasize a comparative approach to experimentation: where designs explicitly consider phylogenetic relationships of taxa (see Webb *et al.* 2002) and conceptualize experimentation through effect sizes (e.g., Nakagawa and Cuthill 2007; Strauss *et al.* 2008). These approaches will allow for both the primary researcher and the meta-analyst to reach the broadest generalizations possible, and perhaps yield causal explanations for the diversity of ecological and evolutionary patterns observed among taxa.

Acknowledgements

I thank the NCEAS workgroup on meta-analysis for thoughtful discussion on this topic; in particular Julia Koricheva, Jessica Gurrevitch, Michael Jennions for lending his study for evolutionary analysis, and Michael Rosenberg for a generous solution to making these statistics applicable to all effect size metrics. I also thank Kerrie Mengersen and Anurag Agrawal for valuable feedback on this manuscript.

REFERENCES

Ackerly, D. D. 2000. Taxon sampling, correlated evolution, and independent contrasts. *Evolution* 54:1480–1492.

Adams, D. C. 2008. Phylogenetic meta-analysis. *Evolution* 62:567–572.

Agrawal, A. A. 2007. Macroevolution of plant defense strategies. *Trends in Ecology and Evolution* 22:103–109.

Beirinckx, K., H. Van Gossum, M. J. Lajeunesse, and M. R. Forbes. 2006. Sex biases in dispersal and philopatry: insights from a meta-analysis based on capture–mark–recapture studies of damselflies. *Oikos* 113:539–547.

Björklund, M. 1997. Are 'comparative methods' always necessary? *Oikos* 80:607–612.

Blomberg, S. P., T. Garland, and A. R. Ives. 2003. Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution* 57:717–745.

Borer, E. T., E. W. Seabloom, J. B. Shurin, K. E. Anderson, C. A. Blanchette, B. Broitman, S. D. Cooper, and B. S. Halpern. 2005. What determines the strength of a trophic cascade? *Ecology* 86:528–537.

Butler, M. A., and A. A. King. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *American Naturalist* 164:683–695.

Casey, J.M., and R. A. Myers. 1998. Near extinction of large, widely distributed fish. *Science* 281:690–692.

Cheverud, J. M., M. M. Dow, and W. Leutenegger. 1985. The quantitative assessment of phylogenetic constraints in comparative analyses: sexual dimorphism in body-weight among primates. *Evolution* 39:1335–1351.

Clark, J. A., and R. M. May. 2002. Taxonomic bias in conservation research. *Science* 297:191–192.

Cooper, H., and L. V. Hedges. 1994. *Handbook of research synthesis*. Rusell Sage Foundation, New York.

Diniz-Filho, J. A. F. 2001. Phylogenetic autocorrelation under distinct evolutionary processes. *Evolution* 55:1104–1109.

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

Freckleton, R. P., P. H. Harvey, and M. Pagel. 2002. Phylogenetic analysis and comparative data: a test and review of evidence. *American Naturalist* 160:712–726.

Gage, M. J. G. 1994. Associations between body size, mating pattern, testis size and sperm lengths across butterflies. *Proceedings of the Royal Society of London, Series* B 258:247–254.

Grafen, A. 1989. The phylogenetic regression. *Philosophical Transactions of the Royal Society of London* B 326: 119–157.

Groβ, J. 2003. *Linear regression*. Berlin: Springer.

Gurevitch, J., and L. V. Hedges. 1999. Statistical issues in conducting ecological meta-analyses. *Ecology* 80:1142–1149.

Hansen, T. F. 1997. Stabilizing selection and the comparative analysis of adaptation. *Evolution* 51:1341–1351.

Harvey, P. H., and M. D. Pagel. 1991. Oxford Series in Ecology and Evolution. Vol. 1. *The comparative method in evolutionary biology*. Oxford University Press, Oxford.

Harvey, P. H., and A. Purvis. 1991. Comparative methods for explaining adaptation. *Nature* 351:619–624.

Hedges, L.V. 1992. Meta-analysis. *Journal of Educational Statistics* 17:279–296.

Hedges, L.V., and I. Olkin 1985. *Statistical methods for meta-analysis*. Academic Press, FL.

Honěk, A. 1993. Intraspecific variation in body size and fecundity in insects: a general relationship. *Oikos* 66: 483–492.

Koricheva, J., H. Nykänen, and E. Gianoli 2004. Meta-analysis of trade-offs among plant antiherbivore defenses: are plants jacks-of-all-trades, masters of all? *American Naturalist* 163:E64–E75.

Lajeunesse, M.J., and M.R. Forbes. 2002. Host range and local parasite adaptation. *Proceedings of the Royal Society of London, Series* B 269:703–710.

Lajeunesse, M.J., and M.R. Forbes. 2003. Variable reporting and quantitative reviews: a comparison of three meta-analytical techniques. *Ecology Letters* 6:448–454.

Larkin M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez R, J. D. Thompson, T. J. Gibson, and D. G. Higgins. 2007. ClustalW and ClustalX version 2. *Bioinformatics* 23:2947–2948.

Lynn, M.R. 1989. Meta-analysis: appropriate tool for the integration of nursing research? *Nursing Research* 38:302–305.

Maddison, W. 1990. A method for testing the correlated evolution of two binary characters: are gains or losses concentrated on branches of a phylogenetic tree? *Evolution* 44:539–557.

Martins, E. P. 2000. Adaptation and the comparative method. Trends in Ecology and *Evolution* 15:296–299.

Martins, E. P., and T. Garland. 1991. Phylogenetic analyses of the correlated evolution of continuous characters: a simulation study. *Evolution* 45:534–557.

Meinick, D. J., and G. A. Hoelzer. 1994. Patterns of speciation and limits to phylogenetic resolution. *Trends in Ecology and Evolution* 9:104–107.

Nakagawa, S., and I. C. Cuthill. 2007. Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Review* 82:591–605.

Nylin, S., K. Nyblom, F. Ronquist, N. Janz, J. Belicek, and M. Källersjö. 2001. Phylogeny of *Polygonia*, *Nymphalis* and related butterflies (Lepidoptera: Nymphalidae): a total-evidence analysis. *Zoological Journal of the Linnean Society* 132:441–468.

Pagel, M. D. 1992. A method for the analysis of comparative data. *Journal of Theoretical Biology* 156:431–442.

Pagel, M. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proceedings of the Royal Society of London, Series* B 255:37–445.

Pagel, M. 1997. Inferring evolutionary processes from phylogenies. *Zoologica Scripta* 26:331–348.

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

Parker, J. D., D. E. Burkepile, and M. E. Hay. 2006. Opposing effects of native and exotic herbivores on plant invasions. *Science* 311:1459–1461.

Posada, D., and K. A. Crandall. 1998. MODELTEST: Testing the model of DNA substitution. *Bioinformatics* 14:817–818.

Ricklefs, R. E., and J. M. Starck. 1996. The application of phylogenetically independent contrasts: a mixed progress report. *Oikos* 77:167–172.

Rohlf, F. J. 2001. Comparative methods for the analysis of continuous variables: geometric interpretations. *Evolution* 55:2143–2160.

Thompson, J. D., D. G. Hoggins, and T. J. Gibson. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 26: 4673–4680.

Thompson, S.G., and J. P. T. Higgins. 2002. How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine* 21:1559–1573.

Schluter, D. 2000. *The ecology of adaptive radiations.* Oxford Univ. Press, Oxford, UK.

Strauss, S. Y., J. A. Lau, T. W. Schoener, and P. Tiffin. 2008. Evolution in ecological field experiments: implications for effect size. *Ecology Letters* 11:199–207

Swofford, D. L. 2001. PAUP: Phylogenetic analysis using parsimony (and other methods), version 4.0b8. Sinauer Associates, Sunderland, Massachusetts.

Torres-Vila, L.M., and M. D. Jennions. 2005. Male mating history and female fecundity in the Lepidoptera: do male virgins make better partners? *Behavoral Ecology and Sociobiology* 57:318–326.

Van Zandt, P. A., and S. Mopper. 1998. A meta-analysis of adaptive deme formation in phytophagous insect populations. *American Naturalist* 152:595–604.

Verdú, M., and A. Traveset. 2004. Bridging meta-analysis and the comparative method: a test of seed size effect on germination after frugivores' gut passage. *Oecologia* 138:414–418.

Verdú, M., and A. Traveset. 2005. Early emergence enhances plant fitness: a phylogenetically controlled meta-analysis. *Ecology* 86:1385–1394.

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

Westoby, M., M. R. Leishman, and J. M. Lord. 1995a. Issues of interpretation following phylogenetic correction. *Journal of Ecology* 83:892–893.

Westoby, M., M. R. Leishman, and J. M. Lord. 1995b. Further remarks on phylogenetic correction. *Journal of Ecology* 83: 727-729.

Westoby, M., M. R. Leishman, and J. M. Lord. 1995c. On misinterpreting the "phylogenetic correction". *Journal of Ecology* 83:531-534.

Wolf, F. M. 1986. *Meta-analysis: Quantitative Methods for Research Synthesis*. Sage Publications Newbury Park, CA.

CHAPTER TWO

EVOLUTIONARY DEAD-ENDS AND THE MACROEVOLUTION OF HOST SPECIFICITY IN PARASITES

Abstract

The evolutionary dead-end hypothesis predicts that specialists are derived from generalists and experience greater extinction rates because of limited ability to adapt to changing conditions. Here I synthesize 43 published phylogenies with comparative and meta-analytical methods to show that these predictions are not universal for specialist parasites, and that exceptions where generalists are derived phylogenetically from specialists occur as frequently as clades where specialists are the more derived. Using Nee's approach to estimate extinction and speciation rates from gene phylogenies, I further show that clades of parasites with few versus many host species do not differ in extinction rates, but that clades with strong specialist-to-generalist trajectories can have greater rates of cladogenesis. Collectively, these findings indicate that there are few sweeping generalizations on the macroevolution of host specificity in parasites, and that predicted constraints and dead-ends due to specialization are less common than traditionally believed.

Introduction

An often-predicted sequence in the evolution of specialization is from generalist to specialist—with a phylogenetic arrangement of specialists being derived (Cope 1896; Brues 1920; Huxley 1942; Rensch 1980; Futuyma and Moreno 1988). It is further predicted that this sequence will lead, in cases of extreme specialization, to evolutionary dead-ends (Mayr 1942; Noble and Noble 1979). These predictions imply

both irreversibility in the evolution of specialization and constraint on the evolutionary persistence of specialists; where specialist lineages cannot revert back to more generalized states, are continually going extinct, and new lineages are established by generalists (Mayr 1963; Futuyma and Moreno 1988). Macroevolutionary tests of these predictions have largely focused on the evolutionary consequences of host specificity in parasites (Poulin 2007). This repeated focus on parasites is due to their obligate trophic link to hosts (e.g., Clayton and Johnson 2003a; Morse and Farrell 2005), which presumably can constrain adaptation, evolutionary trajectories and rates of cladogenesis because opportunity to respond to changing conditions is limited to few host species (Simpson 1953; Mayr 1963; Weigmann *et al.* 1993; Futuyma *et al.* 1995).

Yet the prediction that specialist parasites are continually evolving from generalists and going extinct is misaligned with empirical observations: studies have published positive, null and negative macroevolutionary trends in host specificity for a broad diversity of parasite lineages (e.g., Funk *et al.* 1995; Janz *et al.* 2001; Singer *et al.* 2003). Previous attempts to synthesize this research using lineages of herbivorous insects could not find a universal pattern in host specificity (Nosil 2003; Nosil and Moore 2005). However, issues relating to loss of information by abbreviating specificity into generalist and specialist states, as well as a lack of proper statistical methods to account for study uncertainty, diminished efforts to provide a robust synthesis of phylogenetic trends. Here, I provide the first quantitative synthesis of the evolutionary dead-end hypothesis with 43 published phylogenies of parasites using a composite of meta-analytical and comparative phylogenetic methods (see CHAPTER ONE; but also see Hedges and Olkin 1985; Pagel 1997).

I primarily focus on whether the phylogenetic arrangement of specialists and generalists can be predicted by evolutionary transitions in host specialization and

extinction rates within and among parasitic lineages. I begin by testing the most common prediction for the evolution of specialization: host specialist parasites occupy the most derived positions within a clade (Rensch 1980; Kelley and Farrell 1998). A pattern where specialist parasites are near tips of phylogenies would be consistent with predictions that specialists are in a transitory state and/or have experienced greater extinction rates (Morse and Farrell 2005). Using meta-analysis of published morphological and molecular phylogenies of parasites, I compare the strengths of correlations between the degree to which specialists are derived and their host specificity (Thompson 1994).

This correlative approach, however, cannot determine whether specialization is evolutionarily constrained, whether rates of macroevolution differ among host specialists and generalists (transition rates between states), or provide information of speciation and extinction rates within clades. To determine the contribution of these evolutionary patterns of host specificity, I narrow my focus to 28 molecular phylogenies of parasites with information on the temporal ordering of divergence events within lineages of parasites. With this information, I estimated the rate of evolution toward specialization and generalization for each phylogeny, with the prediction that the evolutionary transition rate towards specialization will be greater than the rate towards generalization (Pagel 1997). Another plausible explanation for high rates of directional transitions in a phylogeny (or biases towards particular evolutionary trajectories) is high extinction rates (Morse and Farrell 2005). Extinction and speciation rates do not occur *pari passu*, estimating these rates separately informs how such processes influence the frequency of extant specialist/generalist taxa and their evolutionary relationships. In an attempt to provide some information on relative extinction rates, I apply Nee *et al.* (1994) maximum likelihood approach to estimate relative extinction and speciation rates from molecular phylogenies.

Methods

Phylogenetic Survey

I surveyed published phylogenies to determine whether specialization is an evolutionary dead-end for parasites. My approach is correlative, and combines various statistical methods that a) determine patterns of trait evolution, b) obtain estimates of speciation and extinction rates from a phylogenetic tree, c) summarize data from within and between data-sets (meta-analysis), and d) control for evolutionary non-independence of these datasets (comparative analyses). This statistical framework accounts for various statistical and biological biases that can confound results of meta-phylogenetic surveys (for details see following sections).

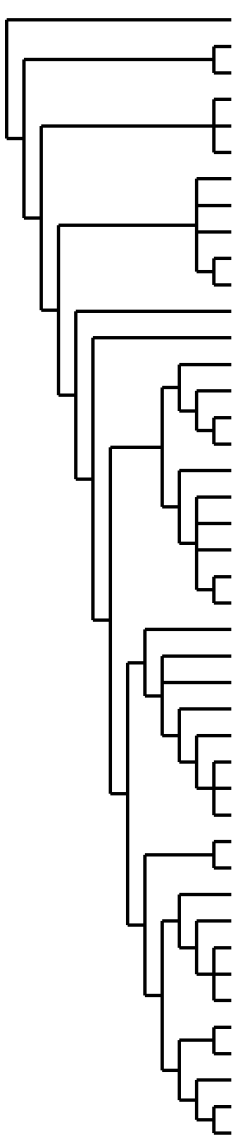
Published phylogenies were drawn from two types of research: studies that directly tested the evolutionary dead-end hypothesis (reporting a phylogeny for a lineage of parasites and information on their specificity; e.g., Morse and Farrell 2005), and studies without details on specificity focusing on the systematics of parasites (e.g., Kopf *et al.* 1998). For this latter group, information on host specificity for each parasite species was obtained from published records and/or annotated lists of host use. Here, it is unavoidable that the focus of the study (either direct tests of evolutionary dead-end or systematics) can bias the taxon sampling of parasite species within lineage, because the stringency in sampling may differ between approaches. For instance, studies focusing on evolutionary dead-ends will often rigorously categorize relationships of all extant species (with information on host specificity) within a clade (e.g., a group of species within a genus). This contrasts with studies focusing on parasite systematics which are mostly interested in resolving broad relationships among representatives within a clade (e.g. representative species from different genera). In addition, I also include phylogenies generated from communities of parasites found on particular groups of hosts. Here, the group of parasites sampled

is related directly to specific ecological interactions. These differences in taxonomic sampling may influence the ability to detect evolutionary trends such as dead-ends and the precision of methods that estimate extinction rates (Thompson 1994), given that key species informing transitions of states may be undersampled. However, inclusion of all these phylogenies was necessary to address a significant taxonomic gap in the research for evolutionary dead-ends—a field dominated by studies on herbivorous insects (see Table 2.1).

In total, I collated 43 phylogenies belonging to diverse groups of parasites (Table 2.1). These data are a broad sampling of various modes of parasitism that occur in nature—such as hematophagy, phytophagy and brood parasitism—and include parasites with narrow and broad host affiliations. Twenty-nine of these studies directly tested the dead-end hypothesis (80% of which were on herbivorous insects), while another 15 phylogenies (not focusing on the dead-end hypothesis) had adequate published information on the host specificity of parasites to test the dead-end hypothesis. An additional, three studies reported phylogenies with a paraphyletic composition based on an ecological community of parasites (e.g., Stireman and Singer 2003).

A variety of data, from genetic to morphological, and a diversity of statistical approaches (maximum likelihood, Bayesian, etc.) were used by the original studies to generate phylogenies. In addition, the group of taxa under study can further bias these approaches (e.g., influence the types of genes used). However, there has yet to be a method developed to reconcile these difficulties in summarizing research across published phylogenies. One approach is simply to report a broad survey across all phylogenies (irrespective of how the phylogenies were generated; see Table 1; here k = 43) paired with a more detailed analysis of only molecular phylogenies where analyses can be standardized (k = 28; high precision data but with fewer phylogenies).

Table 2.1. The 43 parasite phylogenies synthesized with meta-analysis. Each phylogeny, with *n* number of species, was either based on (A)llozyme, (M)olecular, or (MO)rphological data. The hypothesized phylogenetic relationship among parasitic lineages with a topology based on a composite of published information and with internode branch-lengths arbitrarily scaled following Pagel (1994).

global phylogeny	parasite phylogenies	*n*	type	source
	*Orobanche* (holoparasitic plants)	31	M	Manen *et al.* (2004)
	Molothrus (avian brood parasites)	5	M	Freeman and Zink (1995)
	Vidua (avian brood parasites)	16	M	Sorenson *et al.* (2003)
	Lamellodiscus (flatworms)	20	M	Desdevises *et al.* (2002)
	Gyrodactylus (flatworms)	19	M	Cable *et al.* (1999)
	mixed flatworms	74	M	Sasal *et al.* (1999)
	mixed pinworms	48	MO	Hugot (1999)
	Litomosoides (nematodes)	20	MO	Brant and Gardner (2000)
	Howardula/*Parasitylenchus* (nematodes)	6	M	Perlman and Jeanike (2003)
	Parasitylenchus (nematodes)	7	MO	Beveridge and Chilton (2001a)
	Cyclostrongylus (nematodes)	13	MO	Beveridge and Chilton (2001b)
	Cecidophyopsis (gall mites)	7	M	Fenton *et al.* (2000)
	Timema (walkingsticks)	17	M	Crespi and Sandoval (2000)
	Alebra (leafhoppers)	5	MO	Aguin-Pombo (2002)
	Tamalia (aphids)	5	M	Miller and Sterling (2003)
	Calophya (plant lice)	42	MO	Burckhardt and Basset (2000)
	Arytainilla/*Livilla* (plant lice)	23	M	Percy (2003)
	Dennyus (lice)	13	M	Page *et al.* (1998)
	Physconelloides (bodylice)	10	M	Clayton and Johnson (2003a)
	Austrogoniodes (chewing lice)	13	M	Banks *et al.* (2006)
	Brueelia (wing lice)	19	M	Johnson *et al.* (2002)
	Columbicola (wing lice)	21	M	Johnson *et al.* (2003)
	Columbicola (wing lice)	13	M	Clayton and Johnson (2003b)
	Dendroctonus (bark beetles)	18	M	Kelly and Farrell (1998)
	Caryedon (seed beetles)	17	M	Silvain and Delobel (1998)
	Tetraopes/*Phaea* (longhorn beetles)	14	M	Farrell and Mitter (1998)
	Stator (beetles)	21	M	Morse and Farrell (2005)
	Bruchidius/*Bruchus* (seed beetles)	32	M	Kergoat *et al.* (2004)
	Ophraella (leaf beetles)	12	M	Funk *et al.* (1995)
	Oreina (leaf beetles)	14	A	Dobler *et al.* (1996)
	Phratora (leaf beetles)	8	M	Köpf *et al.* (1998)
	Megastigmus (seed chalcids)	28	M	Auger-Rozenberg *et al.* (2006)
	Achrysocharoides (parasitoid chalcids)	67	M	Lopez-Vaamonde *et al.* (2005)
	Geusibia (fleas)	17	MO	Liang and Houyong (2005)
	Urophora (thistle flies)	11	A	Brändel *et al.* (2005)
	mixed parasitoid flies	25	M	Singer *et al.* (2003)
	Tomoplagia (gall flies)	19	M	Yotoko *et al.* (2005)
	Phytomyza (leaf-mining flies)	16	M	Scheffer and Wiegmann (2000)
	Phyllonorycter (leaf-mining moths)	15	M	Lopez-Vaamonde *et al.* (2003)
	Coleophora (moths)	32	MO	Bucheli *et al.* (2002)
	Hemileuca (moths)	26	M	Rubinoff and Sperling (2002)
	Chrysoritis (butterflies)	18	M	Rand *et al.* (2000)
	Nymphalini (butterflies)	20	M	Janz *et al.* (2001)

Here I balance the trade-off between stringency (e.g., precise within-phylogeny data; i.e. branch-lengths, temporal ordering of nodes) and quantity (number of phylogenies that can influence synthesis level analyses and power). This bias in collating data is akin to biases of traditional meta-analysis in the inclusivity of differing approaches for synthesizing findings across experimental studies (Lajeunesse and Forbes 2003).

Estimating host specificity

I limit my quantitative synthesis to parasites with information on their number of host species (hereafter referred to as host range). Host range is the most commonly available estimate of host specificity for all parasitic taxa (e.g., herbivorous insects, monogeneans, avian brood parasites; Price 1980; Poulin 1998), and represents the richness of host species and thus sheer ecological opportunity available to parasites (Price 1980). I specifically avoided characterizing parasites as either 'specialist' or 'generalist' (see Nosil 2002), because this dichotomous classification is often based on host range information and would be difficult to standardize given the broad diversity of parasites used in my study. I also avoid using coarse taxonomic rankings of host groups (e.g., number of genera or families used) when possible to approximate specificity. The definition of genus or family (or any type of Linnean ranking) is inconsistent across the broad taxa of hosts in this study (e.g., a Family of birds is quantitatively different than a Family of insects). This latter approach also assumes that taxonomic diversity is equivalent to species richness—but these two estimates of host specificity may not be monotonically related (see CHAPTER THREE). Finally, using host range as an estimate of host specificity is also aligned with evolutionary prediction of dead-ends: should specialists be limited in resource use and thus constrained in the opportunity to respond to selection, then host-species richness is

likely an important ecological constraint on this predicted evolutionary process.

Meta-analysis of the phylogenetic arrangement of specialists and generalists

Here I quantify the simplest prediction for evolutionary dead-ends: specialist parasites occupy the most derived positions within a clade. Should this phylogenetic arrangement occur within a clade, then the total number of nodes separating each parasite species from the root of the tree (hereafter clade rank) will be negatively correlated with host range. Since, each node represents a hypothesized speciation event; species categorized with a broad clade rank are a product of numerous prior speciation events (and thus are more derived). This approach is a strict topological estimate of pattern occurring within phylogenetic trees.

The statistical relationship between host range and clade rank was summarized as a Pearson product-moment correlation coefficient (r). These correlations were calculated for each phylogeny (with n number of taxa) and then converted to an effect size (δ) using the following Z score transformation (Rosenthal 1991):

$$\delta = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right). \tag{2.1}$$

These effect sizes are then pooled ($\overline{\delta}$) using meta-analysis by statistically weighting each effect size by their sampling variance $\sigma^2(\delta) = (n-3)^{-1}$. Sampling error can yield correlations that over- or under-estimate the relationship between host range and clade rank—a weighting approach that penalizes correlations with high variance should control for this bias (Rosenthal 1991; Lajeunesse and Forbes 2003).

This pooled effect sizes is interpreted as follows: a negative value indicates that a group of k phylogenies show, on average, that parasites with narrow host range (specialists) are more derived; whereas a positive $\overline{\delta}$ indicates that parasites with broad host ranges (generalists) are more derived across phylogenies. To determine whether these effect sizes significantly differ from zero, ninety-five percent confidence

intervals (95% CI) were calculated using the Hedges and Olkin (1985) method. Pooled effect sizes with 95% CI overlapping zero indicates that there is no general pattern in the phylogenic arrangement of specialists and generalists across phylogenies (e.g., a random arrangement).

I also perform several diagnostics to evaluate publication bias and homogeneity of variances. One way to evaluate publication bias is to plot the estimated effect sizes (see equation 2.1) for each study with their sample size (n). A funnel distribution is expected should a) studies with small sample sizes have greater sampling error than large studies, and b) the likelihood of being published is not dependent on this sampling effect or the sign of the research outcome (e.g., null or negative outcome). If such a bias occurs then this would generate an imbalance in the funnel; this imbalance can be tested directly using Egger's test for funnel symmetry (see details in Egger *et al.* 1998). The funnel plot also assumes that effect sizes share a common variance (fixed effect)—if this true then studies should be distributed within a predicted range based on the random sampling of effect sizes from a normal distribution as follows: $\mathcal{N}(\bar{\delta}, \sqrt{\sigma^2(\bar{\delta})})$. Outliers of this predicted distribution indicate that studies do not have homogeneity of variances. Homogeneity of variances can also be tested directly using Chi-square tests with $k-1$ degrees of freedom (see CHAPTER ONE; Hedges and Olkin 1986). When data show a lack of homogeneity, then a random-effects model that assumes multiple potential variances was used to estimate the pooled effect sizes (see Hedges and Olkin 1986; Hedges 1991).

Meta-analysis and phylogenetic non-independence

The phylogenetic non-independence of parasite lineages may be an important source of bias because closely related clades may experience similar constraints associated

with host specificity—given that they likely share similar modes of parasitism (e.g., phloem or blood feeding, leaf mining), similar taxonomic classes of hosts (e.g., strictly vertebrate or plant hosts) or are restricted to a single stage in host development (e.g., parasitizing only nestlings or larvae). Moreover, parasitism among clades may not have an independent origin (e.g., plant feeding has a deep ancestry in sternorrhynchans; i.e. aphids) or may be phylogenetically related but have independently evolved parasitism (e.g., clades of avian brood parasites; see Rothstein *et al.* 2002). I controlled for phylogenetic dependence by fitting a generalized linear model that accounts for evolutionary history to meta-analytic models for summarizing findings across studies.

Controlling for evolutionary history and pooling effect sizes across clades of parasites was completed as follows (see also details in CHAPTER ONE). Note that in this study, clades of parasitism are considered as the study unit (see how these clades are phylogenetically related in Table 2.1), and that the phylogenetically corrected pooled effect size will be referred to as $\overline{\delta}^{\mathrm{P}}$. When a collection of k number of effect sizes (δ) are phylogenetically dependent and share a common effect size (fixed effect model; see Hedges and Olkin 1985), then they can be pooled and weighted by their inverse variance with this generalized least squares (GLS) equation:

$$\overline{\delta}^{\mathrm{P}} = (\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{Y}, \qquad (2.2)$$

where $^{\mathrm{T}}$ indicates the transposition and $^{-1}$ the inverse of a matrix. The components of equation (2.2) are as follows: $\mathbf{Y}$ is the $k \times 1$ column vector of effect sizes (see equation 2.1); $\mathbf{X}$ is the design matrix and is $k \times 1$ column vector of ones (modeled to take a global average); and $\boldsymbol{\Sigma}$ is the $k \times k$ variance-covariance matrix that contains the effect size variances $\sigma^2(\delta)$ on the main diagonal, and the covariance (cov) of each study due to phylogenetic correlations in all off diagonals. The covariance of each i^{th} and j^{th} pair of effect sizes are defined as

$$\text{cov}(\delta_i, \delta_j) = \mathbf{P}_{i,j}\sqrt{\sigma^2(\delta_i)}\sqrt{\sigma^2(\delta_j)}\,. \tag{2.3}$$

The strength of covariance between effect sizes is mediated by the $k \times k$ phylogenetic correlation matrix **P**, which has elements of

$$\mathbf{P}_{i,j}^{\text{BM}} = \begin{cases} b_i^{\text{total}} & \text{if } i = j, \\ \\ b_{i,j}^{\text{shared}} & \text{if } i \neq j. \end{cases} \tag{2.4}$$

Here b^{total} is the maximum branch-length distance from tip to root of a hypothesized phylogeny, and b^{shared} is the shared branch-length distance between each pair of parasite clades. This analysis will thus weight studies by variance (as in traditional meta-analysis), and downweights clades of parasites that are more phylogenetically related than unrelated clades. The variance of the $\overline{\delta}^{\text{P}}$ is calculated with $\sigma^2(\overline{\delta}^{\text{P}}) = (\mathbf{X}^{\text{T}}\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}$ and 95% confidence-intervals (CI) are calculated following Hedges and Olkin (1985).

Finally, I also estimate whether there is a phylogenetic signal across effect sizes using Pagel's (1997) maximum likelihood (ML) approach. This ML approach optimizes a scaling parameter (β) that adjusts the phylogenetic correlations among effect sizes to fit a neutral model of evolution; where $\beta \to 0$ indicates a poor fit of phylogenetic correlations to effect sizes and $\beta \to 1$ a strong fit to correlations. This neutral model is based on Brownian motion evolution which predicts that time of divergence linearly relates with the expected change within a lineage (Rohlf 2001); thus closely related lineages within a phylogeny are expected to be more similar (have a greater phylogenetic correlation in **P** than distantly related lineages. This serves as a hypothesis for predicting the relative phylogenetic conservatism of effect sizes across parasite phylogenies. The scaling parameter (β) is then integrated in the GLS equation (2.2) while pooling of effect sizes ($\overline{\delta}^{\text{P}(\beta=\text{ML})}$)—this reanalysis serves as an alternative

hypothesis on the contribution of phylogenetic correlations. The results of these analyses are then contrasted with a model that assumes no phylogenetic correlations (i.e. $\overline{\delta}^{P(\beta=0)} = \overline{\delta}$ = traditional meta-analysis) and with full (unscaled) phylogenetic correlations (e.g., $\overline{\delta}^{P(\beta=1)}$). Later, I also estimate the β of host specificity within each parasite phylogeny to evaluate the degree of conservatism across lineages. To distinguish the use of β between and within phylogenies, I designate the above phylogenetic signal across parasite phylogenies as $\beta^{between}$.

Rebuilding published phylogenies for estimating evolutionary parameters

The above meta-analysis is a synthesis of the topology of phylogenetic trees irrespective of how the initial published phylogeny was constructed. Here, I rebuild all molecular phylogenies using maximum likelihood (ML) methods in order to a) get information on the relative branch-length distances of nodes (these data are important for estimating various patterns of host range evolution), b) rebuild trees with assumptions of molecular clocks such that there is a temporal ordering of nodes (a required method to estimate extinction rates and speciation rates), and c) standardize consistency of data because published phylogenies often would only report trees based on specific types of phylogenetic analyses (e.g., a consensus tree based on maximum parsimony).

For each published molecular phylogeny, gene sequence data for each parasite species were collated from GenBank (Appendix Table 2.1) or if available, aligned sequence matrices from TreeBASE (Sanderson *et al.* 1994). When available, I also included additional parasite species in phylogenetic analyses that were not included in the original published phylogeny, should host range information be available for these species (see details in Appendix Table 2.1).

Unfortunately few studies report sequence alignment procedures, and thus the phylogenetic hypotheses used in my study will vary slightly from published phylogenies. In general, slight differences in my alignments tended only to influence the positioning of nodes near tips of trees (see below and Appendix Table 2.1). When necessary, sequences were aligned without gap restrictions using CLUSTALW multiple alignment software (vers. 1.4; Thompson *et al.* 1994). When analyses used multiple genes (see Appendix Table 2.1), these genes were aligned separately, and then concatenated for analyses (often following methods outlined in the published phylogeny). If studies did not report outgroup species for phylogenetic analysis, I would take the nearest species belonging to the closest related Family with similar genes; because degree of relatedness between outgroup and ingroup species is known to obscure the phylogenetic relationships within the ingroup (DeSalle *et al.* 1987; Pashley *et al.* 1993). More generally, I closely followed the protocols used in the original study to best replicate the topology and branch-length estimates of their tree.

Maximum likelihood (ML) analyses were performed to determine the topology and branch-length of each published phylogeny of parasites. The best nucleotide substitution model for ML analyses was chosen with MODELTEST (Appendix Table 2.1; Posada and Crandall 1998), which makes use of hierarchical likelihood ratio and AIC tests. However, if HR and AIC models differed, I chose the AIC model because it is a more conservative model selection criterion. All ML analyses were performed with PAUP* 4.0d8 (Swofford 2001) using NBR method. I performed a heuristic search with TBR branch swapping, a stepwise addition starting tree, and with the ASIS stepwise addition option. The generated tree(s) was compared with the published phylogeny, for consistency of topology and branch-length distances. I then repeated all analyses with a molecular clock assumption. Although these analyses were not calibrated against real time, the generated phylogeny still describes the

temporal orderings of nodes, which again are necessary to estimate patterns of clade extinction and speciation (see below; Nee *et al.* 1994).

Extinction and speciation rates

It is possible in principle to estimate extinction rates (μ) and speciation rates (λ) for clades that have grown according to the birth–death (speciation–extinction) process (Kendall 1948; Nee *et al.* 1994). Nee *et al.* (1994) re-parameterized Kendall's (1948) birth–death model such that the likelihood is maximized over functions of μ and λ; specifically, (μ / λ) and ($\lambda - \mu$). This re-parameterization allows for estimation of extinction and speciation rates separately. Here, ($\lambda - \mu$) is the net rate of cladogenesis or diversification of the phylogeny, and (μ / λ) is the magnitude of extinction rate relative to the speciation rate (species turnover rate), indicating the general vulnerability of clades to extinction (Nee *et al.* 1994).

I estimated ($\lambda - \mu$) and (μ / λ) for each clade using the APE (Analyses of Phylogenetics and Evolution; Paradis *et al.* 2004) package of R (vers. 2.0.0; Ihaka & Gentleman 1996). Standard deviations (SD) of ($\lambda - \mu$) and (μ / λ) were calculated using a normal approximation of maximum likelihood estimated (Nee *et al.* 1994), and 95% CI (confidence intervals) with profile likelihood estimates (for details see Nee *et al.* 1994). All analyses were performed on ultrametic ML phylogenies with branch-length and nodes adjusted according to molecular clock assumptions (see above; following Barraclough and Nee 2001). Analyses were preformed only with ingroup data.

Estimating evolutionary parameters of host specificity

The following group of analyses focuses on estimating phylogenetic parameters that are useful for describing the evolution of host specificity in parasites. The first parameter is whether a phylogenetic signal of host range exists for each parasite phylogeny—that is how well host specificity is phylogenetically conserved among closely related parasites. Here a scaling parameter (β) is estimated from the phylogeny (β here is Pagel's lambda; see Pagel 1997; Rambaut and Pagel 2001), and is tested against the null hypothesis that the parasite's host ranges are phylogenetically independent (e.g., $\beta = 0$). If $\beta \neq 0$ and $\beta > 0$, then this suggests a phylogenetic signature and that closely related species tend to have similar host ranges. The values of β will range from zero (host specificity is phylogenetically random) to values that are greater than 1.0. Values greater than one indicate that the phylogenetic patterning of host ranges have a strong phylogenetic signal, but that Brownian motion evolution is perhaps not the best model to explain this phylogenetic pattering of data (see Freckleton *et al.* 2002).

Host range evolution can also occur in a more punctual or gradual manner within lineages of parasites. The scaling parameter (κ) estimates this phyletic pattern of evolution, such that when $\kappa > 0$ and $\kappa \neq 0$, implies a more gradual change in host range specificity; and when $\kappa = 0$, a more punctual model best describes host range evolution. In addition, I also estimate the tempo of host range evolution among parasite lineages by estimating (ε) parameter. Here $\varepsilon \neq 1$ and $\varepsilon > 1$ implies that species-level adaptation in host ranges has been dominant later in the evolution of a lineage, whereas $\varepsilon < 1$ implies an early burst of radiation followed by a more stable sequence of speciation events is coincident with patterns of host range specificity. This test is important because different clades of parasitism may experience differing

strengths of selection at different times during lineage cladogenesis (e.g., herbivorous insects).

These evolutionary parameters were estimated with CONTINUOUS (1.0d13; Rambaut and Pagel 2001; Pagel 1997) for every clade with a topology and branch-lengths based on ML analyses (without molecular clocks assumptions). These analyses were only implemented with ingroup data. When several ML trees were estimated for a single clade, I estimated the above parameters for each tree and averaged their result.

Correlating effect sizes with evolutionary parameters

I use meta-regression techniques to correlate the phylogenetic arrangement of specialists and generalists (δ for each phylogeny) across parasite clades with various evolutionary parameters of the phylogenetic patterning of host ranges (e.g., phylogenetic conservatism, tempo and phyletic patterns of evolution). In addition, I correlate effect sizes with lineage extinction and speciation rates estimated following Nee *et al.* (1994). Note that each data point in these analyses represents a single parasite phylogeny.

To include an evolutionary parameter (c) into meta-analysis, c is treated as an additional column in the design matrix (**X**), where for instance

$$\mathbf{X} = \begin{bmatrix} 1 & \cdots & 1 \\ c_1 & \cdots & c_k \end{bmatrix}^{\mathrm{T}}. \qquad (2.5)$$

Using this design matrix in the GLS equation (2.2) will yield $\overline{\delta}$ as a matrix containing the intercept and the regression coefficient ($\overline{\delta}_{\mathrm{slope}}$) of the meta-regression between effect sizes and the evolutionary parameter. These regression statistics are weighted by variance (as in traditional meta-analysis) and controlled for phylogenetic non-independence. The significance of the meta-regression coefficients are tested with the

following Z test

$$Z_{\text{slope}} = \frac{\left|\overline{\delta}_{\text{slope}}\right|}{\sqrt{\sigma^2(\overline{\delta}_{\text{slope}})}}, \text{ and if } Z_{\text{slope}} \geq 1.96 \text{ then } \overline{\delta}_{\text{slope}} \neq 0. \quad (2.6)$$

See Hedges (1994) for further details on this test and meta-regression in general.

Results

Survey composition and publication bias

Studies where generalists are more derived phylogenetically were more frequent in my dataset than phylogenies where specialists tended to be more derived (Figure 2.1; Egger's test for funnel asymmetry: intercept = –1.2; t = –5.7, d.f. = 42, $p < 0.001$). In addition, a sampling effect is also prevalent across phylogenies: studies with few parasite species (small within study sample size n) were more likely to vary broadly by both direction and magnitude in the correlation between host range and clade rank (see Figure 2.1b). However, the majority of studies were centered near zero; estimating a weak correlation between host range and clade rank (Figure 2.1). These results indicate that meta-analysis is required to weight studies by sampling precision in order to yield an unbiased estimate of the overall (pooled) effect size.

The funnel symmetry test assumes that all effect sizes share a common variance (Hedges 1994). However, using simulations to estimate the predicted funnel distribution by randomly generating effect sizes with a common variance found that some of the observed effect sizes fell beyond this predicted distribution (Figure 2.1b). In addition to these findings, there was also significant heterogeneity in variances across observed effect sizes (Q_{within} = 80.1, d.f. = 42, $p < 0.001$). These results indicate that a fixed effect model may be inappropriate for pooling effect sizes across phylogenies. Thus all following analyses will apply a random effects model (assuming that studies do not share a common variance) to pooling correlations across

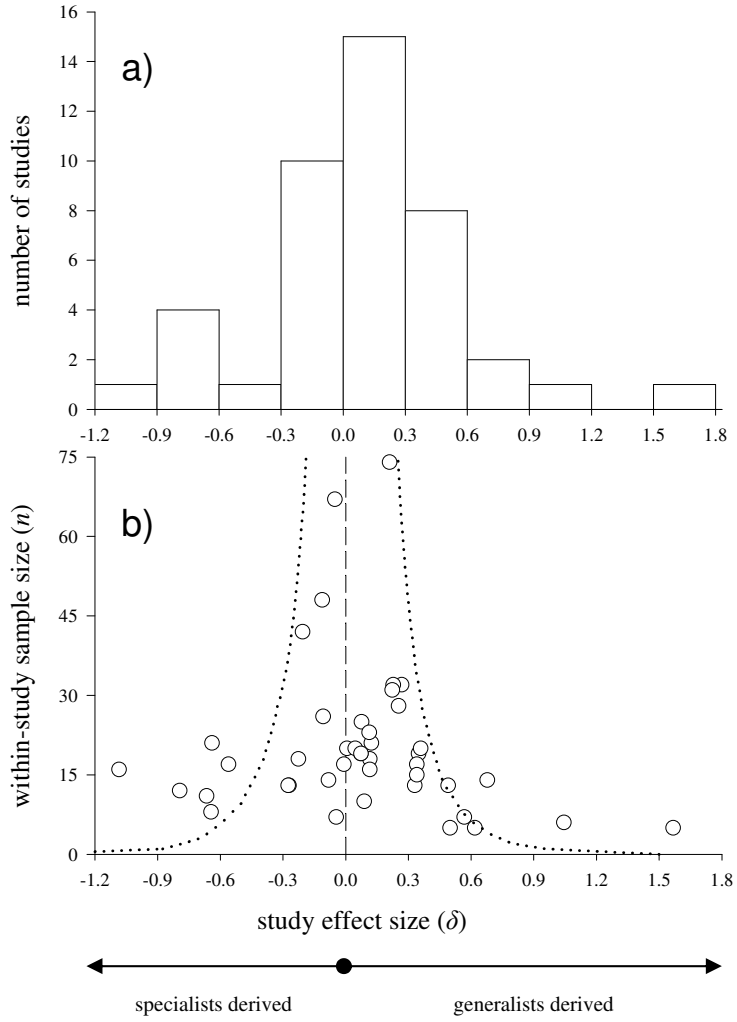


Figure 2.1. Frequency histogram (**a**) and funnel plot (**b**) of effect sizes (δ) quantifying the correlation between host specificity and parasite clade rank for 43 phylogenies (see Table 2.1). Curved dotted lines indicate the 95% CI of the predicted funnel distribution when effect sizes are randomly generated from a distribution with a mean effect size of 0.032 and standard deviation (SD) of 0.523 (these are the raw mean and SD of the 43 studies). Observed effect sizes falling outside this distribution indicate that studies may not share a common expected variance.

phylogenies (for further details see CHAPTER ONE).

Finally, there was also a significant non-zero phylogenetic signal ($\beta^{between}$) in effect sizes across the parasite phylogenies (see Table 2.1; $\beta^{between} = 0.46$, lower 95% CI = 0.03 and upper 95% CI = 0.89). Thus, there is a degree of phylogenetic dependence across effect sizes—parasite clades that are more closely related tended to have correlated effect sizes. This phylogenetic bias was integrated into meta-analysis in two ways: first by using the full phylogenetic correlations (**P** with $\beta^{between} = 1$) when pooling effect sizes, and second using only partial phylogenetic correlations using the optimized estimate of $\beta^{between}$ (**P** with $\beta^{between} = 0.46$).

Are specialists phylogenetically derived?

Pooling 43 parasite clades, specialists were not more likely to be phylogenetically derived than generalist species (see pooled effect size with 95% confidence intervals in Figure 2.2). This pattern remained when controlling for evolutionary history among lineages (see Table 2.1), and when using partial phylogenetic correlations to describe evolutionary relationship across parasite lineages (see Figure 2.2). Parsing data into various taxonomic subgroups also did not reveal any significant taxonomic bias in the phylogenetic arrangement of specialists within clades (Figure 2.2): between group differences in host–taxa (e.g., invertebrate, vertebrate, plant hosts; $Q_{between} = 0.63$, d.f. = 2, $p = 0.729$), parasite taxa (e.g., insect, nematode, monogenean or vertebrate; $Q_{between} = 0.22$, d.f. = 2, $p = 0.974$), or parasitic insects ($Q_{between} = 0.26$, d.f. = 4, $p = 0.992$).

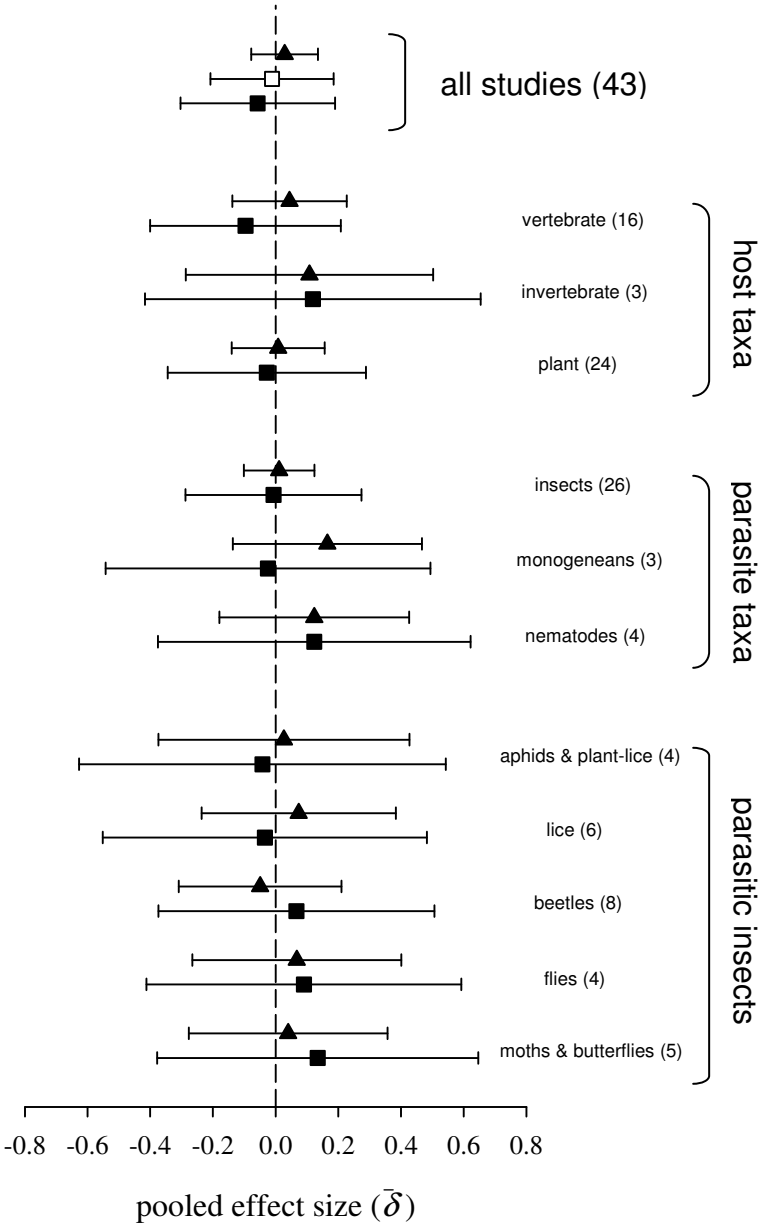


Figure 2.2. Meta-analysis of phylogenetic trends in host specificity across 43 parasite phylogenies. Symbols with 95% CI designate the approach to pooling effect sizes: traditional meta-analysis (▲), and phylogenetically-corrected meta-analysis with $\beta^{between} = 0.457$ (□) or with full phylogenetic correlations of $\beta^{between} = 1$ (■).

Do rates of extinction and cladogenesis within phylogenies predict the phylogenetic arrangement of specialists?

Extinction rate was not a good predictor of the phylogenetic arrangement of specialist and generalists across the 25 molecular phylogenies of parasites (see Figure 2.4; R^2 = 0.02, $Z = 0.88$, $p = 0.379$). When using a phylogenetically controlled meta-regression, a significant negative correlation was found: clades where specialists were more often derived tended to have greater extinction rates (see Figure 2.4; $Z = 4.02$, $p < 0.001$). However, the slope of this meta-regression was biased because only eight of the 25 phylogenies had non-zero extinction rates (Figure 2.3). Analyzing these eight studies separately did not reveal a significant trend between effect sizes and extinction rates in either traditional or phylogenetic meta-regressions ($\overline{\delta}_{\text{slope}} = -0.3$, $Z = 0.71$, $p = 0.479$; phylogenetic meta-regression: $\overline{\delta}_{\text{slope}} = -0.05$, $Z = 0.15$, $p = 0.884$). There was also no relationship between the rates of cladogenesis and the phylogenetic arrangement of specialists and generalists (Figure 2.4; $R^2 = 0.09$, $Z = 0.18$, $p = 0.856$; phylogenetically controlled meta-regression: $Z = 0.02$, $p = 0.985$). However, parasites with invertebrate hosts ($k = 3$) tended to have greater rates of cladogenesis when lineages had more derived specialists than generalists ($\overline{\delta}_{\text{slope}} = -0.02$, $Z = 1.87$, $p = 0.06$; phylogenetically controlled meta-regression: $\overline{\delta}_{\text{slope}} = -0.025$, $Z = 1.9603$, $p = 0.0499$).

Do patterns of host range evolution predict the phylogenetic arrangement of specialists?

Three phylogenetic parameters of host range evolution were estimated from each phylogeny: κ, which indicates whether host ranges evolved in a more punctual or gradual mode of evolution; ε, which indicates the rate of adaptive radiation (e.g., whether host ranges evolved deep within a clade or near tips); and finally β, which

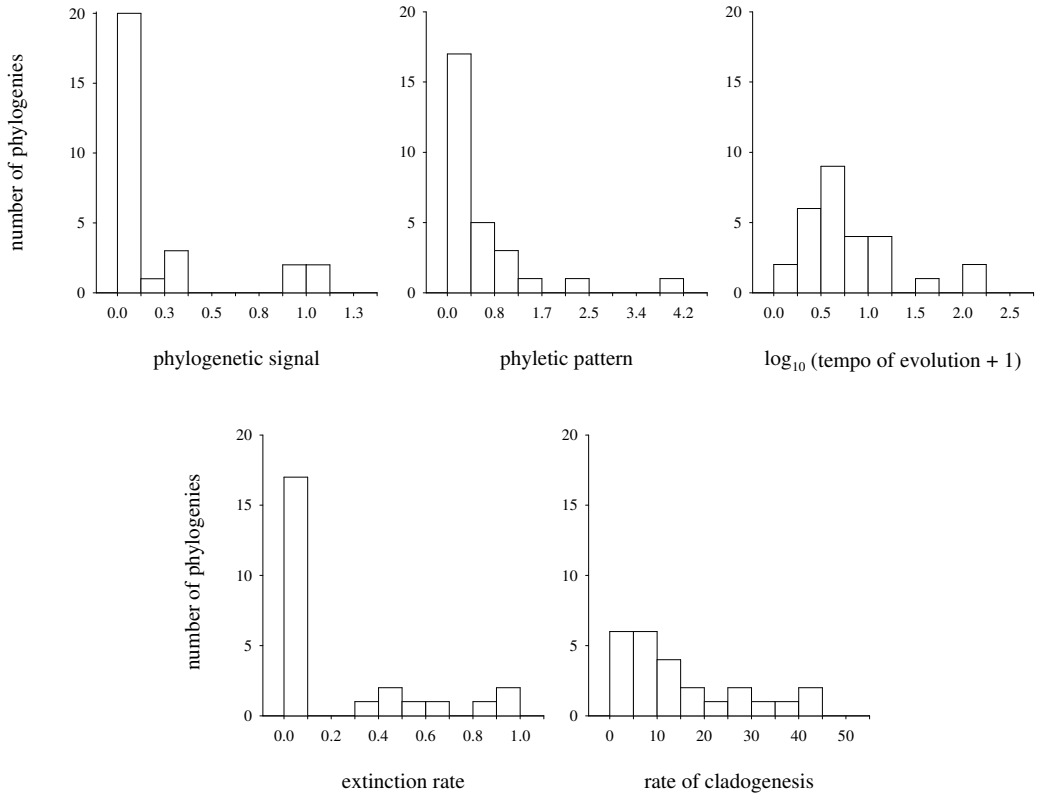


Figure 2.3. Frequency histograms of various evolutionary parameters of host specificity across molecular phylogenies of parasites (see Table 2.1). Extinction rates (μ / λ) and the rates of cladogenesis ($\lambda - \mu$) were estimated following Nee (1994) and evolutionary parameters (κ, ε and β) were estimated following Pagel's (1994) ML methods: phyletic pattern (κ) of host range evolution (i.e. punctual versus gradual), ε as the tempo of host range evolution and the rate of adaptive radiation (e.g., values greater than zero indicate slower early evolution of host range followed by quicker rates of change among closely related species), and whether there is a phylogenetic signal (β) in host ranges. ML was unable to estimate evolutionary parameters for all studies, resulting in different total numbers of phylogenies across histograms.

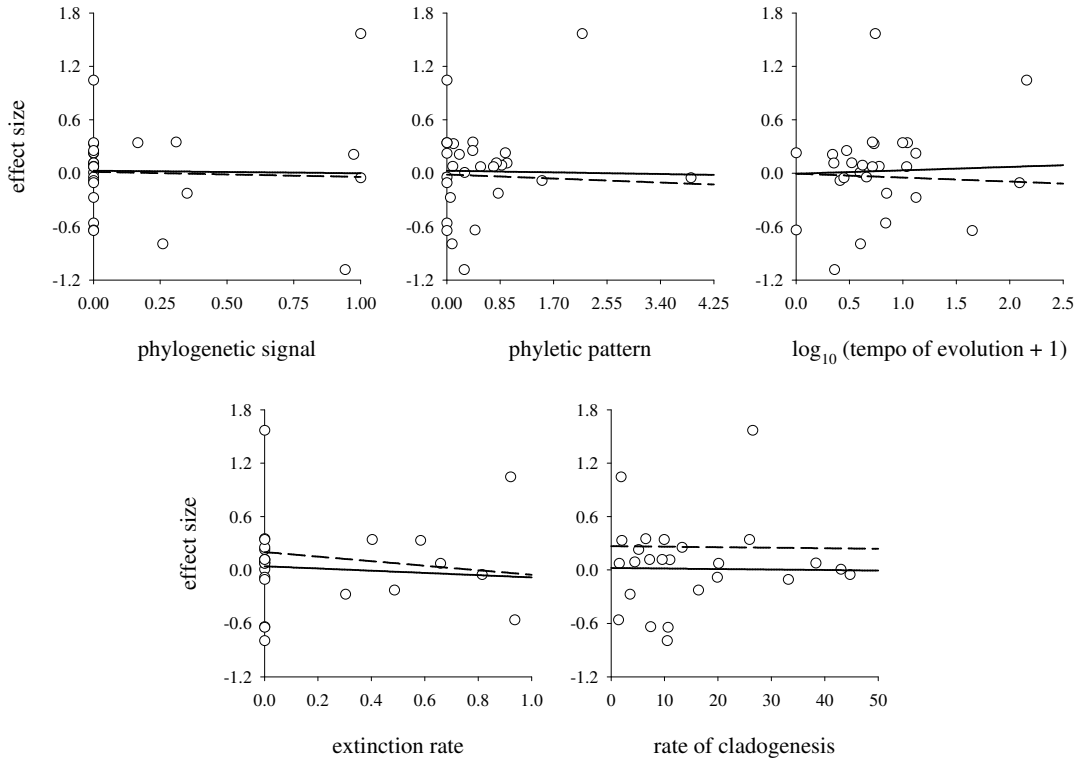


Figure 2.4. Meta-regressions of effect sizes and various evolutionary parameters of host specificity across molecular phylogenies of parasites (see Table 2.1). A negative effect size indicates that specialists were more derived than generalists. Extinction rates and the rates of cladogenesis were estimated following Nee (1994) and evolutionary parameters were estimated following Pagel's (1994) ML methods. For further details of these evolutionary parameters see Figure 2.3. Full lines indicate the slope and intercept of the weighted meta-regression across studies (see *Results* for significance tests), and dashed lines indicate the phylogenetically corrected meta-regressions across the same studies.

indicates whether there is a phylogenetic signal of host range where closely related parasites share similar specificity.

Eight of 28 phylogenies had a non-zero phylogenetic signal (β) in host range specificity (Figure 2.4), half of which were parasitic lineages of vertebrates. A meta-regression including all molecular phylogenies ($k = 28$) and only phylogenies with non-zero phylogenetic signals ($k = 8$) did not find that phylogenetic conservatism in host range explained variation in the phylogenetic arrangement of specialist and generalist parasites; all studies (Figure 2.3; $R^2 = 0.01$, $Z = 0.28$, $p = 0.779$; phylogenetically controlled meta-regression: $Z = 1.09$, $p = 0.272$) and subset with non-zero phylogenetic signals ($\overline{\delta}_{\text{slope}} = 0.03$, $Z = 0.14$, $p = 0.891$; phylogenetically controlled meta-regression: $\overline{\delta}_{\text{slope}} = 0.24$, $Z = 1.07$, $p = 0.286$). The low frequency of host range being a phylogenetically conserved indicates that host specificity for many parasites is more evolutionarily labile than expected (Miller 1987).

The majority of parasite lineages had host ranges that evolved near tips of trees rather than deeper within phylogenies—although tremendous variation was found across lineages (mean $\delta = 10.3$, LCI = –0.4, UCI = 21.1; Figure 2.4). However, this tempo of host range evolution poorly explained variation in the phylogenetic arrangement of specialists and generalists (Figure 2.3; $R^2 = 0.04$, $Z = 0.4$, $p = 0.688$; phylogenetically controlled meta-regression: $Z = 0.88$, $p = 0.386$). Parsing studies by host or parasite taxa found that these groups did not differ in meta-regressions between tempo of evolution and effect sizes (host-taxa: $Q_{\text{between}} = 0.88$, d.f. = 3, $p = 0.83$; parasite-taxa: $Q_{\text{between}} = 1.35$, d.f. = 6, $p = 0.968$).

Twenty-three of 28 parasite lineages had non-zero estimates for the mode of host range evolution: where on average phylogenies showed more gradual than punctual evolution in host ranges (mean $\kappa = 0.82$, LCI = 0.34, UCI = 1.3). However,

the mode of evolution was a poor predictor of variation among effect sizes (Figure 2.3; $R^2 = 0.04$, $Z = 0.29$, $p = 0.786$; phylogenetically controlled meta-regression: $Z = 1.91$, $p = 0.055$). Parasites with different host taxa also did not differ in meta-regressions between κ and effect sizes ($Q_{between} = 0.89$, d.f. = 3, $p = 0.828$). However, different groups of parasitic taxa differed in their relationship between the mode of evolution and the phylogenetic arrangement of specialist and generalists ($Q_{between} = 15.1$, d.f. = 6, $p = 0.019$). This significant effect was not due to the inclusion of taxonomic groups with low samples sizes (excluding taxonomic groups with $k < 2$: $Q_{between} = 10.6$, d.f. = 3, $p = 0.014$). Lineages of monogenean parasites in particular showed a significant negative correlation between the mode of evolution and effect sizes ($\overline{\delta}_{slope} = -0.73$, $Z = 2.82$, $p = 0.005$). Here, lineages where specialists were more clustered near tips had host ranges evolving more gradual rather than punctuated bursts.

Extinction and speciation rates and phylogenetic patterns in host range evolution

Irrespective of the pattern of host range evolution within clades, there was no significant relationship with estimated extinction rates (Fig. 2.4; κ: $R^2 = 0.01$; $F = 0.1$, d.f. = 23, $p = 0.712$; ε: $R^2 = 0.10$; $F = 2.3$, d.f. = 23, $p = 0.14$; β: $R^2 = 0.02$; $F = 0.5$, d.f. = 23, $p = 0.49$; phylogenetically-controlled regressions of κ: $R^2 = 0.01$; $F = 0.1$, d.f. = 6.6, $p = 0.76$; ε: $R^2 = 0.03$; $F = 0.8$, d.f. = 6.6, $p = 0.405$; and β: $R^2 = 0.02$; $F = 0.7$, d.f. = 6.6, $p = 0.43$).

Clades of parasites with high rates of cladogenesis, however, tended to show more gradual rather than punctuated evolution of host specificity (κ: $r^2 = 0.07$; $F = 1.21$, d.f. = 28, $p = 0.029$; phylogenetically-controlled regressions of κ: $r^2 = 0.07$; $F = 4.7$, d.f. = 6.6, $p = 0.073$) and also tended to have phylogenetically conserved host ranges (β: $R^2 = 0.21$; $F = 4.1$, d.f. = 28, $p = 0.0586$; phylogenetically-controlled

regressions of β: $R^2 = 0.03$; $F = 0.8$, d.f. = 6.6, $p = 0.405$). Rates of cladogenesis did not predict rates of adaptive radiation (ε: $R^2 = 0.01$; $F = 0.17$, d.f. = 28, $p = 0.6810$; phylogenetically-controlled regressions of ε: $R^2 = 0.05$; $F = 1.1$, d.f. = 6.6, $p = 0.332$).

Discussion

Theory should revise macroevolutionary predictions about the evolution of host specificity in parasites—this study indicates that specialists are not typically derived from generalists and that specialization is not an evolutionary dead-end for parasites. Analysis across 43 parasite phylogenies could not detect a central tendency for parasite evolution to be constrained by host range specificity, or whether specialist clades experienced greater extinction rates (Figures 2.1 and 2.3). These findings are difficult to interpret because they can arise from poor statistical power. However, my analysis surveyed clades of parasites across broad taxonomic diversity, and used several methods to control for statistical biases that could reduce the ability to identify patterns of host range evolution within clades. These results appear robust, and likely indicate that evolutionary dead-ends are not a common, repeated phenomenon for parasites. I did find however that clades in which generalists tended to be more derived (the opposite pattern) had greater rates of cladogenesis, indicating a potential relationship between being a generalist and belonging to a young rapidly evolving clade (as predicted by Rothstein *et al.* 2003).

A more dynamic view of host range evolution is required to explain the lack of support for directional evolution towards specialization. It does not appear that specialization limits host range evolution, since instances of specialist-to-generalist trajectories exist in nature and occur as frequently as the more often predicted generalist-to-specialist trajectory (Figure 2.1). The lack of variation in evolutionary trajectories among different groups of parasites also more likely reflects the interaction

between fitness trade-offs in the use of different hosts and ecological opportunity favoring generalization (see Jaenike 1990). In addition, host shifts are common phenomenon between related and even unrelated hosts (Dobler *et al.* 1996; Crespi and Sandoval 2000) and availability of host species (i.e. ecological opportunity) is often dependent on geographical distributions (Kelley and Farrell 1998; Rothstein *et al.* 2002). There is also a consensus that specialization is fundamental to the tempo and mode of evolution and the role of adaptation in macroevolution (Futuyma and Moreno 1988; Kelley and Farrell 1998; Janz *et al.* 2001), and the factors promoting the evolution of host specialization have received considerable attention and are numerous (e.g., genetically based trade-offs in host use, competition, predation, etc.; MacArthur and Levins 1964; Via 1985; Bernays 1989; Fry 1996; Clayton and Johnson 2002). These factors likely contribute to specialist parasites not experiencing the severe irreversibility or extinctions predicted by the dead-end hypothesis.

My study shows that host range is an evolutionary labile trait for parasites (Futuyma and Moreno 1988; Stireman 2005), because only 8 of 28 phylogenies had a non-zero phylogenetic signature for host ranges (Figure 2.4). Broad host ranges, however, do not appear to preclude a general trend for specialists to be more often phylogenetically derived (see Figure 2.1; and for example see Kelley and Farrell 1998; Janz *et al.* 2001). This lack of polarity in evolutionary trajectories and phylogenetic signature indicates that selection for host specialization does not completely account for evolutionary patterns in host use (Nosil 2002) and that specificity in terms of host ranges are not as phylogenetically conserved as often predicted (cf. Miller 1987).

Mediating factors such as competition and/or predation likely influence the range of host species available ecologically. At larger geographic scales, endemism also likely contributes to how many hosts are potential resources to parasites (i.e. host

range is limited because parasites are in contact with few ecologically compatible hosts; see CHAPTER FOUR). Alternatively, host range specificity could arise because of adaptive specialization; resulting in parasites that are incapable of successful parasitism on other (related or unrelated) sympatric host species (CHAPTER THREE). A macroevolutionary signature of adaptive specialization is the rapid speciation, in a short time interval, of a clade of parasites into several ecologically divergent states of specificity (Schluter 1996). However, in my survey, the phylogenetic position where host ranges evolved were more often early during the radiation of parasite clades (see ε in Figure 2.4), but also evolved in a more punctual rather than gradual manner (see κ in Figure 2.4). Punctual bursts of speciation are in accord with the rapid evolution of host range specificity of parasites, but deep evolution of host specificity suggests more of an evolutionary radiation rather than adaptive radiation of host ranges.

Finally, Price (1980) argued that the evolution of parasitism is a *key innovation* that can influence the rate of diversification. Thus younger clades of parasites may show marked differences in patterns of host range evolution than older clades. Older clades may have more time to (co)evolve with hosts (which in itself is expected to be a phenomena that narrows specificity; Weigmann *et al.* 1993). Here a potentially important predictor of the phylogenetic arrangement of specialist and generalists within a lineage can be the age of associations and the origin of parasitism within that lineage. For instance, the cowbird parasites *Molothrus* spp. are considered a young clade among avian brood parasites, and their broad variation in host ranges (e.g., 3 to 220 host species; Johnsgard 1998) is often explained by contrasting with the highly specific cuckoos (who are considered to have more ancient parasitic associations with hosts; Rothstein *et al.* 2002). Cuckoos tend to show greater degrees of specificity in host number (i.e. have narrow host ranges) but also in specific adaptations that can

increase parasitism success (e.g., egg mimicry of few host species). Such adaptations can occur as a product of long term evolutionary relationships with hosts. Here the age of clades may contribute significantly to observed patterns in the phylogenetic arrangement of specialists and generalists.

Likely the single greatest constraint on the macroevolution of specialist parasites is the evolution and speciation of hosts. Although many studies have tested the phylogenetic concordance between parasite and host speciation events (Page 2003), few consider the impacts of host specificity evolution in contributing to such patterns in speciation. Research should focus on synthesizing these studies to evaluate whether the likelihood of concordance between host-parasite cospeciation is independent of the mode and tempo of evolution in host specificity.

Acknowledgements

I thank Anurag Agrawal, Brian Danforth, Nelson Hairston, Richard Harrison for helpful discussion on this project. This project was supported by a Canadian NSERC post-graduate scholarship, Kathleen and Donald fellowship at the University of Toronto, Cornell University, and NSF DEB 0447550 and USDA Hatch funds (to Anurag Agrawal).

REFERENCES

Aguin-Pombo, D. 2002. Genetic differentiation among host-associated *Alebra* leafhoppers (Hemiptera: Cicadellidae). *Heredity* 88:415–422.

Auger-Rozenberg, A.-M., C. Kerdelhue, E. Magnoux, J. Turgeon, J.-Y. Rasplus, and A. Roques. 2006. Molecular phylogeny and evolution of host-plant use in conifer seed chalcids in the genus *Megastigmus* (Hymenoptera: Torymidae). *Systematic Entomology* 31:47–64.

Banks, J. C., R. L. Palma, and A. M. Paterson. 2006. Cophylogenetic relationships between penguins and their chewing lice. *Journal of Evolutionary Biology* 19: 156–166.

Bernays, E. A. 1989. Host range in phytophagous insects: the potential role of generalist predators. *Evolutionary Ecology* 3:299–311.

Beveridge, I., and N. B. Chilton. 2001. Co-evolutionary relationships between the nematode subfamily Cloacininae and its macropodid marsupial hosts. *International Journal for Parasitology* 31:976–996.

Brändle, M., S. Knoll, S. Eber, J. Stadler, and R. Brandl. 2005. Flies on thistles: support for synchronous speciation? *Biological Journal of the Linnean Society* 84:776–783.

Brant, S. V., and S. L. Gardner. 2000. Phylogeny of species of the genus *Litomosoides* (Nematatoda : Onchocercidae): Evidence of rampant host switching. *Journal of Parasitology* 86:545–554.

Bucheli, S., J. F. Landry, and J. Wenzel. 2002. Larval case architecture and implications of host-plant associations for North American *Coleophora* (Lepidoptera; Coleophoridae). *Cladistics* 18:71–93.

Burckhardt, D., and Y. Basset. 2000. The jumping plant-lice (Hemiptera, *Psylloidea*) associated with *Schinus* (Anacardiaceae): systematics, biogeography and host plant relationships. *Journal of Natural History* 34:57–155.

Barraclough, T. G. and S. Nee. 2001. Phylogenetics and speciation. *Trends in Ecology & Evolutionary Biology* 16:391–399.

Cable, J., P. D. Harris, R. C. Tinsley, and C. M. Lazarus. 1999. Phylogenetic analysis of *Gyrodactylus* spp. (Platyhelminthes: Monogenea) using ribosomal DNA sequences. *Canadian Journal of Zoology* 77:1439–1449.

Clayton, D. H., and K. P. Johnson. 2003. Linking coevolutionary history to ecological process: doves and lice. *Evolution* 57:2335–2341.

Cope, E. D. 1896. *The primary factors of organic evolution.* Open Court, Chicago.

Crespi, B. J., and C. P. Sandoval. 2000. Phylogenetic evidence for the evolution of ecological specialization in *Timema* walking-sticks. *Journal of Evolutionary Biology* 13:249–262.

Denno, R. F., and M. S. McClure. 1983. *Variable plants and herbivores in natural and managed systems.* Academic Press.

Desdevises, Y., S. Morand, O. Jousson, and P. Legendre. 2002. Coevolution between *Lamellodiscus* (Monogenea : Diplectanidae) and *Sparidae* (Teleostei): The study of a complex host-parasite system. *Evolution* 56:2459–2471.

Dobler, S., P. Mardulyn, J. M. Pasteels, and M. Rowell-Rahier. 1996. Reproductive biology of viviparous and oviparous species of the leaf beetle genus *Oreina*. *Entomologia Experimentalis et Applicata* 80:375–388.

Farrell, B. D., and C. Mitter. 1998. The timing of insect/plant diversification: might *Tetraopes* (Coleoptera : Cerambycidae) and *Asclepias* (Asclepiadaceae) have co-evolved? *Biological Journal of the Linnean Society* 63:553–577.

Farrell, B. D., C. Mitter, and D. J. Futuyma. 1991. Diversification at the insect-plant interface. *Bioscience* 42:34–42.

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

Fenton, B., A. N. E. Birch, G. Malloch, P. G. Lanham, and R. M. Brennan. 2000. Gall mite molecular phylogeny and its relationship to the evolution of plant host specificity. *Experimental and Applied Acarology* 24:831–861.

Freckleton, R.P., P.H. Harvey, and M. Pagel. 2002. Phylogenetic analysis and comparative data: a test and review of evidence. *American Naturalist* 160:712–726.

Freeman, S., and R. M. Zink. 1995. A phylogenetic study of the blackbirds based on variation in mitochondrial-DNA restriction sites. *Systematic Biology* 44:409–420.

Fry, J. D. 1996. The evolution of host specialization: Are trade-offs overrated? *American Naturalist* 148S:84–107.

Funk, D. J., D. J. Futuyma, G. Ortí, and A. Meyer. 1995. A history of host associations and evolutionary diversification for *Ophraella* (Coleoptera: Chrysomelidae): new evidence from mitochondial DNA. *Evolution* 49:1008–1017.

Futuyma, D. J., M. C. Keese, and D. J. Funk. 1995. Genetic constraints on macroevolution: The evolution of host affiliation in the leaf beetle genus Ophraella. *Evolution* 49:797–809.

Futuyma, D. J., and G. Moreno. 1988. The evolution of ecological specialization. *Annual. Review of Ecology and Systematics* 19:207–233.

Hedges, L. V., and I. Olkin. 1985. *Statistical Methods for Meta-Analysis.* Florida: Academic Press, Inc.

Hugot, J. P. 1999. Primates and their pinworm parasites: the Cameron hypothesis revisited. *Systematic Biology* 48:523–546.

Huxley, J. 1942. *Evolution: the modern synthesis.* George Allen and Unwin, London.

Ihaka, R., and R. Gentleman. 1996. R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 5:299–314.

Jaenike J. 1990. Host specialization in phytophagous insects. *Annual Review of Ecology and Systematics* 21:243–273.

Janz, N., K. Nyblom, and S. Nylin. 2001. Evolutionary dynamics of host-plant specialization: A case study of the tribe *Nymphalini*. *Evolution* 55:783–796.

Johnsgard, P. A. 1997. *The avian brood parasites: deception at the nest.* Oxford University Press, New York.

Johnson, K. P., R. J. Adams, R. D. M. Page, and D. H. Clayton. 2003. When do parasites fail to speciate in response to host speciation? *Systematic Biology* 52:37–47.

Kelley, S. T., and B. D. Farrell. 1998. Is specialization a dead end? The phylogeny of host use in *Dendroctonus* bark beetles (Scolytidae). *Evolution* 52:1731–1743.

Kendal, D. G. 1948. On the generalized birth-death process. *Annual Review of Mathematics and Statistics* 19:1–15.

Kergoat, G. J., A. Delobel, and, J.-F. Silvain. 2004. Phylogeny and host-specificity of European seed beetles (Coleoptera, *Bruchidae*), new insights from molecular and ecological data. *Molecular Phylogenetics and Evolution* 32:855–865.

Köpf, A., N. E. Rank, H. Roininen, R. Julkunen-Tiitto, J. M. Pasteels, and Tahvanainen, J. 1998. The evolution of host-plant use and sequestration in the leaf beetle genus *Phratora* (Coleoptera: Chrysomelidae). *Evolution* 52:517–528.

Lajeunesse, M. J., and M. R. Forbes. 2003. Variable reporting and quantitative reviews: a comparison of three meta-analytical techniques. *Ecology Letters* 6:448–454.

Lanyon, S. M. 1992. Interspecific brood parasitism in blackbirds (Icterinae): a phylogenetic approach. *Science* 255:77–79.

Liang, L., and W. Houyong. 2005. Morphological phylogeny of Geusibia Jordan, 1932 (Siphonaptera: Leptopsyllidae) and the host–parasite relationship with pikas. *Systematic Parasitology* 61:65–78.

Lopez-Vaamonde, C., H. C. J. Godfray, S. A. West, C. Hansson, and J. M. Cook. 2005. The evolution of host use and unusual reproductive strategies in *Achrysocharoides* parasitoid wasps. *Journal of Evolutionary Biology* 18:1029–1041.

Lopez-Vaamonde, C., H., C. J. Godfray, and J. M. Cook. 2003. Evolutionary dynamics of host-plant use in a genus of leaf-mining moths. *Evolution* 57: 1804–1821.

Manen, J.-F., C. Habashi, D. Jeanmonod, J.-M. Park, and G. N. Schneeweiss. 2004. Phylogeny and intraspecific variability of holoparasitic *Orobanche* (Orobanchaceae) inferred from plastid rbcL sequences. *Molecular Phylogenetics and Evolution* 33:482–500.

Mayr, E. 1942. *Systematics and the Origin of Species.* Columbia University Press, New York.

Mayr, E. 1963. *Animal Species and their Evolution.* Belknap Press of Harvard University Press, Cambridge, MA.

MacArthur, R. H., and R. Levins. 1964. Competition, habitat selection and character displacement in a patchy environment. *Proceedings of the National Academy of Sciences* USA 51:1207–1210.

Miller, J. S. 1987. Host-plant relationships in the Papilionidae (Lepidoptera): Parallel cladogenesis or colonization? *Cladistics* 3:105–120.

Moran, N. 1988. The evolution of host-plant alternation in aphids: evidence for specialization as a dead end. *American Naturalist* 132:681–706.

Morse, G. E., and B. D. Farrell. 2005. Ecological and evolutionary diversification of the seed beetle genus *Stator*. *Evolution* 59:1315–1333.

Muller, A. 1996. Host-plant specialization in western palearctic anthidiine bees (Hymenoptera: Apoidea; Megachilidae). *Ecological Monographs* 66:235–257.

Nee, S., E. H. Holmes, R. M. May, and P. H. Harvey. 1994. Extinction rates can be estimated from molecular phylogenies. *Philosophical Transactions: Biological Sciences* 344:77–82.

Noble, E. R., and G. A. Noble. 1976. *Parasitology: The Biology of Animal Parasites.* Lea & Febiger, Philadelphia, PA.

Nosil, P. 2002. Transition rates between specialization and generalization in phytophagous insects. *Evolution* 56:1701–1706.

Nosil, P., and A. O. Mooers. 2005. Testing hypotheses about ecological specialization using phylogenetic trees. *Evolution* 59:2256–2263.

Page, R. D. M. 2003. *Tangled trees: phylogeny, cospeciation, and coevolution.* The University of Chicago Press, Chicago.

Page, R. D. M., and M. A. Charleston. 1998. Trees within trees: phylogeny and historical associations. *Trends in Ecology & Evolution* 13:356–359.

Pagel, M. D. 1992. A method for the analysis of comparative data. *Journal of Theoretical Biology* 156:431–442.

Pagel, M. D. 1997. Inferring evolutionary processes from phylogenies. *Zoologica Scripta* 26:331–348.

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

Paradis, E., and J. Claude. 2002. Analysis of comparative data using generalized estimating equations. *Journal of Theoretical Biology* 218:175–185.

Percy, D. M. 2003. Radiation, diversity, and host-plant interactions among island and continental legume-feeding psyllids. *Evolution* 57:2540–2556.

Perlman, S. J., and J. Jaenike. 2003. Infection success in novel hosts: An experimental and phylogenetic study of *Drosophila*-parasitic nematodes. *Evolution* 57:544–557.

Price, P. W. 1980. *The Evolutionary Biology of Parasites.* Princeton University Press. Princeton, New Jersey.

Posada, D., and K. A. Crandall. 1998. MODELTEST: Testing the model of DNA substitution. *Bioinformatics* 14:817–818.

Poulin, R. 2007. *Evolutionary Ecology of Parasites.* Princeton University Press, Princeton, NJ.

Rand, D. B., A. Heath, T. Suderman, and N. E. Pierce. 2000. Phylogeny and life history evolution of the genus *Chrysoritis* within the Aphnaeini (Lepidoptera : Lycaenidae), inferred from mitochondrial cytochrome oxidase I sequences. *Molecular Phylogenetics and Evolution* 17:85–96.

Rambaut, A. and M. Pagel. 2001. *Continuous*. University of Oxford.

Rensch, B. 1980. Historical development of the present synthetic neo-Darwinism in Germany. pp. 284–303 in E. Mayr and W. Provine, eds. *The evolutionary synthesis: perspectives on the unification of biology.* Harvard Univ. Press, Cambridge, MA.

Rohlf, F. J. 2001. Comparative methods for the analysis of continuous variables: geometric interpretations. *Evolution* 55:2143–2160.

Rosenthal, R. 1991. *Meta-analytic procedures for social research.* Sage, Newbury Park.

Rothstein, S. I., M. A. Patten, and R. C. Fleischer. 2002. Phylogeny, specialization, and brood parasite-host coevolution: some possible pitfall of parsimony. *Behavioral Ecology* 13:1–10.

Rubinoff, D., and F. A. H. Sperling. 2002. Evolution of ecological traits and wing morphology in *Hemileuca* (Saturniidae) based on a two-gene phylogeny. *Molecular Phylogenetics and Evolution* 25:70–86.

Sanderson, M. J., M. J. Donoghue, W. Piel, and T. Eriksson. 1994. TreeBASE: a prototype database of phylogenetic analyses and an interactive tool for browsing the phylogeny of life. *American Journal of Botany* 81:183.

Sasal, P., S. Trouve, C. Muller-Graf, and S. Morand. 1999. Specificity and host predictability: a comparative analysis among monogenean parasites of fish. *Journal of Animal Ecology* 68:437–444.

Scheffer, S. J., and B. M. Wiegmann. 2000. Molecular phylogenetics of the holly leaf miners (Diptera : Agromyzidae : *Phytomyza*): Species limits, speciation, and dietary specialization. *Molecular Phylogenetics and Evolution* 17:244–255.

Schluter, D. 1996. Ecological causes of adaptive radiation. *American Naturalist* 148:S40–S64.

Silvain, J.-F., and A. Delobel. 1998. Phylogeny of west african *Caryedon* (Coleoptera: Bruchidae): Congruence between molecular and morphological data. *Molecular Phylogenetics and Evolution* 9:533–541.

Simpson, G. 1953. *The major features of evolution.* Columbia Univ. Press, New York.

Sorenson, M. D., K. M. Sefc, and R. B. Payne. 2003. Speciation by host switch in brood parasitic indigobirds. *Nature* 424:928–931.

Stireman, J.O. III. 2005. The evolution of generalization? Host range in parasitoid flies. *Journal of Evolutionary Biology* 18:325–336.

Stireman, J. O., and M. S. Singer. 2003. What determines host range in parasitoids? An analysis of a tachinid parasitoid community. *Oecologia* 135:629–638.

Swofford, D. L. 2001. PAUP: Phylogenetic analysis using parsimony (and other methods), version 4.0b8. Sinauer Associates, Sunderland, Massachusetts.

Thompson, J. N. 1994. *The coevolutionary process.* University of Chicago Press, Chicago, IL.

Thompson, J. N. 1998. Rapid evolution as an ecological process. *Trends in Ecology & Evolution* 13:329–332.

Via, S. 1989. Field estimation of variation in host plant use between local populations of pea aphids from two crops. *Ecological Entomology* 14:357–364.

Via, S. 1990. Ecological genetics and host adaptation in herbivorous insects: the experimental study of evolution in natural and agricultural systems. *Annual Review of Entomology* 35:421–446.

Wiegmann, B. M., C. Mitter, and B. Farrell. 1993. Diversification of carnivorous parasitic insects: extraordinary radiation or specialized dead end? *American Naturalist* 142:737–754.

Yotoko, K. S. C., P. I. Prado, C. A. M. Russo, and V. N. Solferini. 2005. Testing the trend towards specialization in herbivore–host plant associations using a molecular phylogeny of *Tomoplagia* (Diptera: Tephritidae). *Molecular Phylogenetics and Evolution* 35:701–711.

Zar, J. H. 1985. *Biostatistical analysis, Second Edition.* Prentice-Hall, Englewood cliffs NJ.

APPENDIX Table 2.3. Published gene phylogenies of parasites used in meta-phylogenetic analyses. Sample size of phylogenies is designated by *n*. Genes used for analyses are: (A) cytochrome c oxidase I, (B) leucine tRNA, (C) cytochrome c oxidase II, (D) 12S ribosomal RNA, (E) 18S ribosomal DNA, (K) internal transcribed spacer 1, (L) 5.8S ribosomal RNA, (M) internal transcribed spacer 2, (J) 28S ribosomal RNA, (F) cytochrome b, (G) NADH dehydrogenase subunit 6, (H) tRNA glutamine, (I) 5' half of the control region, (N) elongation factor 1 alpha, (O) large subunit of ribulose-bisphosphate carboxylase. Model specifications indicate the best nucleotide substitution model determined by MODELTEST. Maximum likelihood (ML) phylogenies with branch-lengths are presented in parenthetical format, and when multiple ML are provided. GENBANK accession numbers with * are outgroup species, and TREEBASE study accession numbers for aligned matrices are presented in [square] brackets. Sources include both published phylogenies and host ranges.

parasite group	***n***	genes	model specification	source(s)	accession numbers	hypothesized phylogeny in parenthetical (NEWICK) format
Achrysocharoides (parasitoids on microlepidopterans; Insecta: Eulophidae)	13	J + F	(GTR+I+G)	Lopez-Vaamonde *et al.*. (2005)	[S1254]	((((((('Achrysocharoides_acerianus':0.003372,(('Achrysocharoides_butus':0.003403,'Achrysocharoides_latreillii_327':0.000676,'Achrysocharoides_latreillii_328':0.000678):0.001342,'Achrysocharoides_latreillii_326':0):0.005645):0.007314,'Achrysocharoides_sp':0.008139):0.000566,'Achrysocharoides_buekkensis':0.008366):0.004217,((('Achrysocharoides_cilla_ex_Corylus':0.000916,'Achrysocharoides_cilla_ex_Viburnum':0.003923):0.001412,('Achrysocharoides_splendens_84':0.000756,'Achrysocharoides_splendens_ex_Alnus':0.000690):0.004447):0.002458,'Achrysocharoides_cilla_ex_Quercus':0.002851):0.008571):0.004386,((((('Achrysocharoides_atys':0.007311,'Achrysocharoides_suprafolius':0.010727):0.003142,('Achrysocharoides_niveipes':0.007684,('Achrysocharoides_zwoelferi_300':0,'Achrysocharoides_zwoelferi_324':0,'Achrysocharoides_zwoelferi_325':0,'Achrysocharoides_zwoelferi_329':0,'Achrysocharoides_zwoelferi_330':0.000672):0.004733):0.010132):0.002373,'Achrysocharoides_carpini':0.010699):0.004018,'Achrysocharoides_insignitellae':0.019981):0.006740):0.002143,(('Chrysocharis_nepherus':0.029848,'Chrysocharis_sp1':0.015556):0.004861,('Chrysocharis_sp2':0.042988,'Chrysocharis_sp3':0.051713):0.002850):0.023418,('Kratoysma_gliricidiae_320':0.001734,'Kratoysma_gliricidiae_321':0.008426):0.024786);

Arytainilla/Livilla (plant lice; Insecta: Psyllidae)	23	B+I	(GTR+I+G)	Percy (2003)	[S1024]	((((('A_devia':0.141811,('A_nubivaga':0.028921,'A_vittata':0.041307):0.045639):0.097395,((((('A_diluta':0.052931,((((((('A_dividens':0.032846,('A_modica':0.005847,'A_hupalupa':0.001426):0.016650):0.019540,('A_menceyat':0.065624,'A_ochrita':0.039797):0.008785):0.019328,(((((('A_equitans':0.101954,'A_gomerae':0.024591):0.001841,'A_canarien':0.035884):0.008011,'A_occident':0.027592):0.022003,('A_incuba':0.040157,'A_umbonata':0.021871):0.018181):0.014837,'A_pileolat':0.064757):0.006180):0.007216,('A_nigralin':0.061770,'A_probosci':0.035589):0.027749):0.008803,'A_fortunat':0.064540):0.010481,'A_romeria':0.098571):0.008852):0.014350,'A_prognata':0.074573):0.019305,'L_monosper':0.262541):0.041299,'Al_serpent':0.180681):0.027859):0.075187,('Caco_alate':0.121307,'Caco_mali':0.136798):0.095507):0.314528,('Aciz_holli':0.134600,'Aciz_uncat':0.130317):0);
Austrogoniodes (chewing lice; Insecta: Philopteridae)	14	A + C	(K81uf+I+G)	Banks *et al.*. (2006)	(AY229899,AY229920), (AY229900,AF491758), (AY229901,AY229921), (AY229906,AY345910), (AY229903,AY229925), (AY229905,AF491757), (AY229907,AY229926), (AY229908,AY229927), (AY229909,AY229928), (AY229910,AY229929), (AY229911,AY229930), (AY229912,AF491754), (AY229913,AY229931), (AY229898 ,AY229919)*	('N_demersus':0.992542,(('A_antarcti':0.868669,'A_gressitt':0.237085):0.132772,('A_concii':0.103297,('A_hamilton':0.006674,('A_keleri':0.004256,'A_macquari':0.004381):0.006219):0.023507):0.224953):0.060885,((((('A_bifascia':0.035893,'A_demersus':0.046476):0.072450,('A_cristati':0.197770,'A_vanalphe':0.102461):0.033741):0.100005,('A_brevipes':0.172801,'A_mawsoni':0.164640):0.350358):0.027648,'A_watersto':0.181760):0.021809);
Bruchidius/Bruchus (seed beetles; Insecta: Bruchidae)	31	A +D + F	(GTR+I+G)	Kergoat *et al.*. (2004)	(AY390639, AY390671, AY390703), (AY390640, AY390672, AY390704), (AY390641, AY390673, AY390705), (AY390642, AY390674, AY390706), (AY390643, AY390675, AY390707), (AY390644, AY390676, AY390708), (AY390645, AY390677, AY390709), (AY390646, AY390678, AY390710), (AY390647, AY390679, AY390711), (AY390648, AY390680, AY390712), (AY390649, AY390681, AY390713), (AY390650, AY390682, AY390714), (AY390651, AY390683, AY390715), (AY390652, AY390684, AY390716), (AY390653, AY390685, AY390717), (AY509806, AY509809, AY509812), (AY390654, AY390686, AY390718), (AY390657, AY390689, AY390720), (AY390655, AY390687, AY390719), (AY390656, AY390688, NULL), (AY390658, AY390690, AY390721), (AY390659, AY390691, AY390722), (AY390660, AY390692, AY390723), (AY509807, AY509810, AY509813), (AY390661, AY390693, AY390724), (AY390662, AY390694, AY390725), (AY390663, AY390695, AY390726), (AY390664, AY390696, AY390727), (AY390666, AY390698, AY390729), (AY390667, AY390699, AY390730), (AY509808, AY509811, AY509814), (AY390638, AY390670, AY390702), (AY390636, AY390668,AY390700)*, (AY390637, AY390669, AY390701)*	(((('Br_bigutta':0.380801,('Br_pauper':0.285548,'Br_unicolo':0.173876):0.058787):0.150642,(((((((('Br_bimacul':0.019766,'Br_nanus':0.037710):0.112396,('Br_fulvico':0.072689,'Br_picipes':0.113825):0.053761):0.004904,('Br_dispar':0,'Br_varius':0.037232):0.135921):0.026924,'Br_varipic':0.198909):0.048084,('Br_pusillu':0.177453,'Br_seminar':0.159618):0.082511):0.018326,((('Br_caninus':0.072712,'Br_margina':0.119875):0.160247,'Br_cineras':0.292557):0.026834,(('Br_lividim':0.198366,'Br_villosu':0.147796):0.096004,(('Br_pygmaeu':0.023152,'Br_sericat':0.010039):0.092167,'Br_trifoli':0.159587):0.135173):0.031820):0.032188):0.029342,'S_sp':0.447596):0.016467):0.020490,(((((' B_affinis':0.061368,'B_viciae':0.052830):0.041146,('B_dentipes':0.005646,'B_rufimanu':0.009085):0.076049):0.036307,('B_loti':0.093709,('B_luteicor':0.022416,'B_rufipes':0.008336):0.065556):0.078953):0.085501,(('B_brachial':0.184828,'B_laticoll':0.195388):0.048109,('B_tristicu':0.124883,'B_tristis':0.119388):0.046124):0.037504):0.029732,'Pal_gilvus':0.270690):0.037375,'Pac_cardo':0.745482);

Brueelia (avian lice; Insecta: Philopteridae)	14	A+C	(GTR+I+G)	Johnson *et al.*. (2002)	(AY149411, AY149410), (AY149417, AY149387), (AY149415, AY149385), (AY149429, AY149399), (AY149430, AY149400), (AY149421, AY149391), (AY149422, AY149392), (AY149423, AY149393), (AY149426, AY149396), (AY149427, AY149397), (AY149424, AY149394), (AY149425, AY149395), (AY149418, AY149388), (AY149420, AY149390), (AY149431, AY149401), (AY149432, AY149402), (AY149434, AY149404), (AF447199, AF348871), (AF320457, AY149405), (AY149435, AY149406), (AY149436, AF356717)*, (AF278667, AF278649)*	((((((((('B_moriona':0.068877,(('B_1':0.060316,'B_2':0.078090):0.010767,'B_3':0.085002):0.053864):0.022822,('B_laticeps':0.106884,'B_4':0.136204):0.006512):0.015060,'B_5':0.162821):0.008411,(('B_6':0.021846,'B_7':0.052094):0.070910,((('B_8':0.108704,'B_11':0.041214):0.016159,('B_9':0.050004,'B_10':0.064549):0.036259):0.022290,'B_12':0.079737):0.004713):0.042906):0.015497,('F_analoide':0.092802,'P_sp':0.241981):0.038583):0.018988,'B_marginel':0.181511):0.125963,'N_longiros':0.326146):0.121473,'P_cubanus':0.252382):0.076874,('Q_punctatu':0.234668,'S_lari':0.279442):0.090371):0.345222,'R_columbia':0);
Caryedon (seed beetles; Insecta: Bruchidae)	16	B	(TVM+G)	Silvain & Delobel (1998)	AF004114, AF004115, AF004116, AF004117, AF004118, AF004119, AF004120, AF004121, AF004122, AF004123, AF004124, AF004125, AF004126, AF004127, AF004128, AF004129, AF004131*	(((((((('C_acaciae':0.045850,('C_longispi':0,'C_serratus':0):0.012006):0.017820,(('C_excavatu':0.059690,'C_mauritan':0.000001):0.023663,'C_sahelicu':0.016066):0.021138):0.042195,(('C_crampeli':0,'C_dialii':0):0.019340,'C_pallidus':0.029271):0.091911):0.042261,('C_albonota':0.040260,('C_fathalae':0.036986,'C_macropte':0.015428):0.224513):0.042712):0.044734,'C_longipen':0.059509):0.023999,('C_fuligino':0.027193,'C_lunatum':0.024733):0.021302,'C_immacula':0.091460):0.372897,'Call_macul':0);
Cecidophyopsis (gall mites; Acari: Eriophyidae)	8	G	(TrN+G)	Fenton *et al.*. (2000)	AJ297574, AJ297575, AJ297571, AJ297572, AJ297569, AJ297570, AJ297576, AJ297577, AJ297573*	((((((('C_ribis':0,'C_spicata':0.003881):0.005911,'C_selachod':0.005948):0.009620,('C_alpina':0,'C_aurea':0.001554):0.023793):0.021697,('C_grossula':0.001604,'C_sp':0.002379):0.020251):0.081073,'C_psilaspi':0.155043):0.592085,'P_gracilis':0);
Chrysoritis (butterflies; Insecta: Lycaenidae)	21	A	(GTR+I+G)	Rand *et al.*. (2000)	AF279231, AF279237, AF279240, AF279218, AF279241, AF279222, AF279229, AF279244, AF279230, AF279224, AF279243, AF279228, AF279242, AF279236, AF279234, AF279235, AF279225, AF279239, AF279238, AF279217, AF279227, AF279221*, AF279223*, AF279220*, AF279232*, AF279226*, AF279233*, AF279219*	(((((('A_pierus':0.107473,(Targyropla:0.046819,'T_malagrid':0.082719):0.048872):0.012397,'A_argyrasp':0.117073):0.028242,('P_clavum':0.073450,'T_sardonyx':0.064078):0.063058):0.061934,'C_leroma':0.123122):0,((((((('C_aridus':0.022253,('C_balli':0.010094,'C_oriental':0.028577):0.002077,('C_brooksi':0.006111,('C_daphne':0.008700,('C_nigrica1':0.007814,'C_nigrica2':0.001178):0.004584):0.001669,'C_palmus':0.010844):0.001844):0.003821,'C_thysbe':0.031335):0.045966,((((('C_coetzeri':0.006205,('C_cottrell':0.008684,'C_zeuxo':0.005047):0.004126):0.002308,'C_zonarius':0):0.052816,'C_felthami':0.114189):0.016128,('C_pyroeis1':0.001361,'C_pyroeis2':0):0.063097):0.018009):0.008078,'C_chrysant':0.051657):0.021462,('C_dicksoni':0.075904,'C_oreas':0.068104):0.010186):0.011186,('C_chrysaor':0.041683,'C_lycegene':0.041639):0.010168):0.031128);
Columbicola (avian winglice; Insecta: Philopteridae)	10	A+B+C	(TIM+G)	Clayton & Johnson (2003)	(AF278614, AF278629, AF190419), (AF278615, AF278630, AF190421), (AF278616, AF278631, AF190417), (AF278618, AF278633, AF190423), (AF278617, AF278632, AF190422), (AY151010, AY151022, AY151016), (AF278626, AF278641, AF190410), (AF278627, AF278642, AF190411), (AF278624, AF278639, AF190424), (AF278628, AF278643, AF190425), (AF348872, AF385025, AF189140)*	(chiniri:1.34567,(gracilicap:0.191393,(baculoides:0.362105,('1passerina':0.037196,'2passerina':0.044533):0.122977):0.053412):0.045718,(extinctus:0.06344,('2macrourae':0.07809,(adamsi:0.515668,('3macrourae':0.034622,('1macrourae':0.092562,'4macrourae':0):0.044732):0.036947):0.006256):0.047963):0.034194);

Dendroctonus (bark beetles; Insecta: Scolytidae)	18	A	(GTR+I+G)	Kelley & Farrell (1998)	AF067987, AF067994, AF068004, AF067988, AF067986, AF068001, AF068000, AF068002, AF067997, AF067993, AF375315, AF067991, AF067998, AF067996, AF067989, AF067985, AF375318, AF067990, U82583*, U82236*	(((((('D_ponderos':0.032539,'D_jeffreyi':0.102227):0.095640,((('D_vitei':0.117152,'D_mexicanu':0):0.094542,('D_adjunctu':0.076664,('D_approxim':0.122443,'D_brevicom':0.090187):0.036288):0.034372):0.019871,'D_frontali':0.103266):0.039767):0.041443,(((('D_valens':0.109957,'D_rhizopha':0.060363):0.043774,'D_terebran':0.113034):0.037122,(('D_micans':0.085951,'D_punctatu':0.046829):0.027896,('D_rufipenn':0.014544,'D_murrayan':0):0.095035):0.078686):0.007734):0.024277,'D_armandi':0.261470):0.021537,('D_simplex':0.097469,'D_pseudots':0.068075):0.032599):0.285375,'T_minor':0.184525,'O_erosus':0.576959);
Dennyus (avian lice; Insecta: Menoponidae)	15	E	(K81uf+I+G)	Page *et al.*. (1998)	U96432, U96417, U96415, U96416, U96419, U96420, U96422, U96421, U96423, U96425, U96426, U96427, U96429, U96431, U96430, U96434*, U96433*	(('D_hirundin':0.511512,'D_cypsiuri':0.176289):0,('D_vauxi':0.371739,(((('D_d_distin':0.012366,'D_d_timjon':0.013853):0.063291,(('D_somadika':0.127973,'D_c_carljo':0.096478):0.034774,((('D_c_forres':0.014048,'D_c_foster':0.035706):0.111333,'D_singhi':0.105096):0.009452):0.011218):0.219977,((((('D_thompson':0.015678,'D_collinsi':0.003291):0.079525,'D_adamsae':0.031319):0.056908,'D_wellsi':0.132618):0.034464,('D_simberlo':0.060594,'D_wraggi':0.181768):0.194953):0.168170):0.113203):0.171807);
Gyrodactylus (fish parasites; Monogenea: Diplectanidae)	12	G	(HKY+G)	Cable *et al.*. (1999)	AJ001839, AJ011410, AJ001846, AY061976, AJ132985, AJ001842, AJ001845, AJ011411, AJ001841, AJ132259, AJ001847, U58097*, AJ001843*	('Gyrd_galli':0,(((('G_arcuatus':0.297749,'G_rarus':0.110079):0.231470,('G_bullatar':0.273991,'G_turnbull':0.281746):0.189882):0.158043,(('G_kobayash':0.084653,'G_gurleyi':0.123374):0.116526,((('G_pungitii':0.016210,'G_rogatens':0.042684):0.047849,((('G_gasteros':0.089277,'G_derjavin':0.067663):0.012603,'G_salaris':0.052699):0.033620):0.108101):0.491816):1.343116);
Hemileuca (moths; Insecta: Saturniidae)	20	A+C	(GTR+I+G)	Rubinoff & Sperling (2002)	(AF388006, AY040115), (AF388007, AF388010), (AF388016, AY040119), (AF388022, AF388004), (AF388005, AY040114), (AF388002, AY040113), (AF388003, AF388013), (AY040117, AF388014), (AF388012, AF391177), (AY040116, AF391175), (AY040105, AF391169), (AY040099, AF474013), (AF387983, AY040107), (AF474008, AF474009), (AF474016, AF391174), (AY040104, AF474015), (AF391173, AY040103), (AF391176, AY040106), (AF388024, AY040120), (AF387987, AY040109), (AF387989, AF387997)*	(((((((('H_electra1':0.004391,(('H_e_mojave':0.002634,'H_e_clio':0.002634):0.000762,'H_e_rubra':0.003396):0.000995):0.006954,'H_electra2':0.011345):0.030491,(((('H_juno':0.013318,'H_neumogen':0.013318):0.021362,'H_burnsi':0.034681):0.002879,('H_maia':0.006414,'H_nevadens':0.006414):0.031145):0.004276):0.002238,('H_grotei':0.029961,'H_stonei':0.029961):0.014113):0.006252,((('H_eglanter':0.014423,'H_hera':0.014423):0.010702,'H_nuttali':0.025125):0.015466,('H_chinatie':0.017429,'H_griffini':0.017429):0.023162):0.009734):0.017250,(((('H_tricolor':0.013746,'H_penninsu':0.013746):0.007785,'H_hualapai':0.021531):0.003477,'H_oliviae':0.025008):0.042566):0.040748,('C_velda':0.086884,'P_lasiocam':0.086884):0.021438):0.020108,('M_nibasa':0.058400,'A_cecrops':0.058400):0.070031):0;
Howardula/ Parasitylenchus (fruit-fly nematodes; Nematoda: Allantonematidae)	9	D	(GTR+I+G)	Perlman & Jeanike (2003); Perlman *et al.*. (2003)	AF519226, AF519227, AF519224, AF519225, AF519223, AF519222, AF519232, AF519231, AF519230, AF519233, AF519234, AF202157*, AF202164*	(((((('H_neocosmi':0.002077,'H_cf_neoco':0.000661):0.004944,(('H_aoronymp':0.001026,'H_cf_aoron':0.001088):0.023804,('H_sp_SPB':0.004361,'H_sp_SPF':0.012749):0.006437):0.002446):0.063738,'H_sp_SPA':0.156246):0.013738,'H_sp_SPMA':0.057871):0.010882,('H_sp_SPPS':0.034310,'H_sp_SPB':0.024374):0.011111):0.049455,'H_dominick':0.201364):0.138440,('P_magnicau':0.073199,'S_radicico':0.067187):0);
Lamellodiscus (fish parasites; Monogenea: Diplectanidae)	20	D	(SYM+I+G)	Desdevise *et al.*. (2002)	AY038187, AY038188, AY038189, AJ276441, AF294956, AY038190, AJ276440, AY038191, AY038192, AY038193, AY038194, AF294957, AY038195, AY038196, AY038197, AF294954, AY038198, AF294955, AJ276442, AF294953, AJ276439*, AJ228793*, AF294952*	(((((L_baeri':0.025312,'L_erythrin':0.014724):0.006810,(('L_ergensi':0.006289,'L_fraternu':0):0.009013,'L_knoeffle':0.020424):0.001544,'L_ignoratu':0.024239):0.010346,((('L_bidens':0.002093,'L_hilii':0.006317):0.004176,'L_gracilis':0.002069,'L_mirandus':0.008368):0.003420,'L_parisi':0.037463):0.013099,'L_imperviu':0.005613):0.019089):0.007338,(((((L_coronatu':0.017155,'L_elegans':0.041547):0.012085,'L_furcosus':0.002743):0.006327,'F_echeneis':0.031406):0.003903,('L_drummond':0.017124,('L_mormyri':0.002740,'L_verberis':0.018205):0.004016):0.005173):0.002310,'L_virgula':0.018918):0.013275):0.048396,'D_aequans':0.103484,('P_ardens':0.125630,'D_minor':0.222370):0.052385);

Megastigmus (conifer chalcids; Insecta: Torymidae)	26	F	(GTR+I+G)	Auger-Rozenberg *et al.*. (2006)	AY898663, AY898691, AY898669, AY898664, AY898681, AY898699, AY898702, AY898695, AY898671, AY898689, AY898672, AY898696, AY898670, AY898684, AY898676, AY898678, AY898701, AY898686, AY898680, AY898694, AY898703, AY898697, AY898688, AY898667, AY898682, AY898668, AY898704*, AY898705*	(((((amicorum:0.000008,((atlanticus:0.028332,wachtli:0.012024):0.028746,(bipunctatu:0.021385,pingii:0.056747):0.019653):0.023960):0.085645,cryptomeri:0.092517):0.007119,((((((atedius:0.035075,tsugae:0.022921):0.021326,(likiangens:0.014320,strobilobi:0.049446):0.047634):0.002429,(((borriesi:0.028297,pictus:0.017722):0.002416,((lasiocarpa:0.041832,((milleri:0.040571,pinus:0.028287):0.003798,specularis:0.028433):0.002609):0.014999,rafni:0.046056):0.003543):0.002585,hoffmeyeri:0.023701,spermotrop:0.067569):0.004057):0.005743,((pinsapinis:0.014551,suspectus:0.024483):0.019440,schimitsch:0.027349):0.034142):0.016122,thyoides:0.079086):0.018762,(dorsalis:0.137637,stigmatiza:0.163700):0.189345):0.015504):0.033783,rosae:0.089398):0.413817,'T_azureus':0.189253,'T_sp':0.214284);
Molothrus (cowbirds; Aves: Icteridae)	5	E	(TVM+I)	Freeman & Zink (1995)	AF089040, AF290172, AF089043, AF089044, AF089060, AF290173*, AF290165*	((((('M_aeneus':0.002328,'S_oryzivor':0.025898):0.003958,('M_ater':0.007519,'M_bonarien':0.008937):0.010769):0.006263,'M_rufoaxil':0.048965):0.034695,'A_phoenice':0.083849,'D_oryzivor':0.151480);
Monogeneans (MIXED; Monogenea: Monopisthocotylea/ Polyopisthocotylea)	10	G	(GTR+G)	Sasal *et al.*. 1999	AF311718, AF311702, AF311714, AF311712, AF131722, AF311719, AF382053, AF131710, AY841871, AF131711, AF286908*	('Giga_magna':0.650396,(((('Plec_obscu':0.318793,((('Atri_acarn':0.035871,'Spar_chrys':0.037194):0.015358,'Micr_eryth':0.057911):0.030167,'Pyra_holli':0.059404):0.019223):2.186890,'Furn_echen':2.241e-008):0.069961,('Ligo_mugil':0.107115,'Ancy_mogur':0.099632):0.248405):0.100255,'Caps_marti':0.465904);
Ophraella (leaf beetles; Insecta: Chrysomelidae)	12	A	(GTR+I+G)	Funk *et al.*. (1995A; 1995B);	U20684, U20689, U20695, U20686, U20712, U20707, U20717, U20701, U20716, U20704, U20706, U20714, U20681*, U20679*	((((((('O_arctica':0.002460,'O_bilineat':0.005728):0.005314,('O_artemisi':0.011866,'O_nuda':0.013786):0.043304):0.003667,'O_communa':0.006716):0.081409,('O_notulata':0.043967,'O_slobodki':0.117289):0.037608):0.023510,(('O_conferta':0.013667,'O_sexvitta':0.017754):0.059672,'O_cribrata':0.056943,'O_notata':0.098731):0.072415):0.029148,'O_pilosa':0.165891,'Monoxia_sp':0.147350);
Orobanche sect. *Orobanche* (parasitic plants; Magnoliopsida: Orobanchaceae)	31	O	(GTR+I+G)	Manen *et al.*. (2004)	AY582181, AY582182, AF161798, AY582183, AY582184, AY582185, AY582186, AY582187, AY582189, AY582190, AY582191, AF090349, AY582197, AY582199, AY582204, AF078682, AY582203, AY582206, AY582208, AY582228, AY582231, AY582240, AF130336, AY582250, AY582275, AY582254, AY582255, AY582257 AY582260, AY582269, AY582272, AF026837*, AF026838*, AF026839*	((((((('O_a_lycoct':0.019487,(('O_caryophy':0.018884,'O_teucrii':0.145514):0.012416,'O_lutea':0.021648):0.006102):0.000905,((('O_alsatica':0.004919,'O_bartling':0.000801):0.006911,(((((('O_amethys1':0.003935,'O_hederae3':0):0.001571,'O_hederae6':0):0.001570,('O_amethys2':0,'O_hederae2':0):0.003131):0.001562,(('O_a_campe1':0,'O_a_campe2':0,'O_minor1':0.000806,('O_minor2':0.004599,'O_minor5':0.002120):0.000703,'O_minor3':0,'O_minor4':0.000777):0.001425,('O_crenata':0.006879,'O_transcau':0.001615):0.000974):0.002157):0.001514,'O_pubescen':0.002558):0.012503):0.002773,'O_spec':0.013866):0.035542):0.003353,('O_gracilis':0.029519,((('O_cf':0.000793,'O_r_genis1':0.000779,'O_r_genis2':0):0.003117,'O_rigens1':0):0.000729,'O_rigens2':0.000808):0.006431):0.013712):0.005562,'O_coerules':0.005557):0.000526,('O_cernua':0.004366,'O_cumana':0.019574):0.114614):0.032517,('S_pectinat':0.012752,('S_asiatica':0.005302,'S_gesnerio':0.007394):0.015275):0);
Phratora (leaf beetle; Insecta: Chrysomelidae)	5	B, not rep. used 12S	(K81uf+G)	Köpf *et al.*. (1998)	AY027721, AY027720, AY027697, AY027719, AY027709, AY027701*	((((('P_tibialis':0.000008,('P_laticoll':0.067382,'P_vitellin':0.014317):0.013523):0.075801,'G_viridula':0.098391):0.052454,'C_vigintip':0.025777):0.051439,'L_aenea':0);

Phyllonorycter (leaf-miner moths; Insecta: Gracillariidae)	77	H	(GTR+I+G)	Lopez-Vaamonde *et al.*. (2003)	[S921]	((((((((((((((((((('P_ulmifoliella':0.007357,('P_strigulatella':0.014516,('P_sp._5':0.007819,('P_sp._4':0.172788,'P_emberizaepennella':0.014287):0.071493):0.014853):0.004684):0.002452,'P_rajella':0.015413):0.005571,'P_anderidae':0.013270):0.008522,(('P_baldensis':0.021902,'P_scopariella':0.012021):0.008723,('P_parvifoliella':0.015855,'P_stantoniella':0.006874):0.050074):0.011825):0.003365,((('P_heringiella':0.078876,'P_salictella':0):0.002208,'P_viminiella':0):0.002225,('P_brevilineatella':0.018941,'P_salicicolella':0,'P__dubitella':0.002294):0.005494):0.005049):0.002473,(('P_lautella':0.032746,'P_viminetorum':0.035675):0.006962,'P_ulicicolella':0.064341):0.002750):0.003342,('P_hilarella':0.017034,'P_nigrescentella':0.032030,(('P_platanoidella':0.007675,'P_sylvella':0.008250):0.015139,'P_geniculella':0.015922):0.014938):0.002341):0.001415,((((P_sp._2':0.018375,'P_distentella':0.002164):0.013361,'P_roboris':0.024836):0.021626,'P_millierella':0.109088):0.015824,'P_insignitella':0.005040):0.024000):0.008085,((((P_stettinensis':0.042823,'P_nicelii':0.035091):0.031097,('P_cavella':0.026637,('P_kleemannella':0.004413,'P_froelichiella':0.004564):0.021560):0.026314):0.007393,'P_issikii':0.050963):0.007930,'P_harrisella':0.041976):0.003381,((((((P_sorbi':0.002247,'P_leucographella':0.002245):0.020646,((((P_muelleriella':0.039356,'P_ilicifoliella':0.044636):0.004153,'P_sp._3':0.009988):0.007476,(('P_messaniella':0,'P_quercifoliella':0):0.007803,'P_sp._1':0.012099):0.010985):0.013645,'P_joviella':0.014332):0.011193,(((P_aemula':0,'P_coryli':0.002242):0.004503,'P_esperella':0.004651):0.009173,'P_tenerella':0.011758):0.006991):0.003091,'P_corylifoliella':0.041753):0.003076,(('P_crataegella':0.009262,('P_gerasimovi':0.009213,'P_cydoniella':0.002283,'P_pomonella':0.005021):0.007584,(('P_oxyacanthae':0.005224,'P_mespilella':0):0.002530,'P_blancardella':0,'P_elmaella':0):0.005025):0.016428,'P_junoniella':0.005679):0.021615):0.004001,'P_lantanella':0.040804):0.001929,('P_heegeriella':0.083331,'P_saportella':0.011907,((((P_sagitella':0.050392,'P_comparella':0.000170):0.064227,'P_populifoliella':0.026789):0.048515,'P_maestingella':0.056977):0.001459,('P_cerasicolella':0.015498,'P_spinicolella':0):0.021754):0.008151):0.014541):0.007649):0.003753,'P_platani':0.045023):0.022769,'P_quinqueguttella':0.074114):0.005072,'P_grewiella':0.106243):0.007879,'P_loxozona':0.204616):0.006011,(((P_scabiosella':0.012398,'P_trifasciella':0.015898):0.025419,'P_tristrigella':0.045706):0.011821,'P_schreberella':0.055521):0.010785):0.043684,(((C_guttifinitella':0.164651,('C_sp._2':0.017668,'C_sp._1':0.086271):0.016509):0.068500,'C_hamadryadella':0.058747):0.006760,'C_ohridella':0.054657):0.048845):0.084968,('Acrocerops_brongniardella':0.232661,'Dialectica_sp.':0.209631):0.081114):0.015283,((((C_populetorum':0.126542,'Calybites_auroguttella':0.085166):0.067516,('C_falconipennella':0.055969,'C_azaleella':0.010526):0.070445):0.090327,'Bucculatrix_ulmella':0.313284):0.052881,('Callisto_denticulella':0.067315,'Parornix_carpinella':0.089382):0.108646):0.023840):0.031211,'Phyllocnitis_citrella':0.606527):0.204087,('Adela_sp._2':0.154647,'Adela_sp._1':0.190107):0);
Physconelloides (avian bodylice; Insecta: Philopteridae)	13	A+B+C	(GTR+I+G)	Clayton & Johnson (2003)	(AF278648, AF278666, AY273876), (AF278650, AF278668, AY273877), (AF278651, AF278669, AY273879), (AF278649, AF278667, AF414766), (AF278657, AF278675, AY273885), (AF278658, AF278676, AY273886), (AF278656, AF278674, AY273884), (AF278655, AF278673, AY273882), (AF348842, AF348668, AY273883), (AF348841, AF348648, AY273880), (AF278654, AF278672, AY273881), (AF278661, AF278679, AY273887), (AF278660, AF278678, AY273888), (AF545707, AF320404, AY314859)*	((((((1ceratocep:0.049606,2ceratocep:0.039501):0.011610,cubanus:1.304856):0.017410,3ceratocep:0.019573):0.037203,((1eurysema:0.060448,eurysema3:0.055497):0.025258,2eurysema:0.088854):0.034671):0.017196,((1anolaimae:0.043533,2anolaimae:0.041030):0.021423,(1spenceri:0.066523,2spenceri:0.050441):0.014434):0.068551):0.014951,(wisemani:0.085495,zenaidurae:0.038620):0.040261,Goniodes:0.383527);

Phytomyza (leafminers; Insecta: Agromyzidae)	16	B	(GTR+I+G)	Scheffer & Wiegmann (2000)	AF230217, AF230241, AF276827, AF276828, AF276829, AF276830, AF276831, AF276832, AF276833, AF276834, AF276835, AF276836, AF276837, AF276838, AF276839, AF276840, AF276841, AF276842, AF276843, AF276844, AF276845, AF276846, AF276847, AF276848, AF276849, AF276850, AF276851, AF276852, AF276853, AF276854, AF276869, AF276855, AF276856, AF276857, AF276858, AF276859, AF276860, AF276861, AF276862, AF276863, AF276864, AF276865, AF276866, AF276867, AF276868	(('L_huidobre':0.261422,'C_fasciata':0.165478):0,('P_flavicor':0.104144,('P_aconiti':0.122962,((('P_plantagi':0.100648,'P_sp':0.239363):0.028497,'P_aquilegi':0.136758):0.014497,('P_erigerop':0.159783,('P_ilici':0.118915,('P_n_sp':0.143385,('P_glabrico':0.119864,('P_jucunda':0.132994,('P_vomitori':0.026661,((('P_vertici1':0.025931,('P_n_sp2':0.016009,'P_verticil':0.008973):0.011697):0.024244,(('P_n_sp5':0.024792,'P_n_sp3':0.021139):0.027133,('P_ilicicol':0.037891,('P_n_sp1':0.016216,'Ph_opacae':0.015676):0.008870):0.008633):0.002258):0.002650,(('P_nemopant':0.008831,'P_sp2':0.021023):0.026365,'P_ditmani':0.048908):0.004245):0.013188):0.032859):0.028140):0.025385):0.010223):0.039698):0.005327):0.014166):0.032143):0.043452);
Stator (beetles; Insecta: Chrysomelidae)	22	A + N	(GTR+I+G)	Morse & Farrell (2005)	(AY997368, AY997314), (AY997369, AY997315), (AY997370, AY997316), (AY997371, AY997317), (AY997372, AY997318), (AY997374, AY997320), (AY997376, AY997322), (AY997395, AY997341), (AY997396, AY997342), (AY997397, AY997343), (AY997399, AY997345), (AY997400, AY997346), (AY997403, AY997349), (AY997404, AY997350), (AY997406, AY997352), (AY997407, AY997353), (AY997408, AY997354), (AY997409, AY997355), (AY997411, AY997357), (AY997415, AY997361), (AY997417, AY997363), (AY997418, AY997364)*, (AY997419, AY997365)*, (AY997366, AY997312)*, (AY997367, AY997313)*	(((((('S_aegrotus':0.150961,('S_maculato':0.057833,'S_trisigna':0.047733):0.051419):0.033843,('S_monachus':0.091951,'S_vittatit':0.136240):0.033709):0.028400,(('S_beali':0.030746,'S_limbatus':0.026320):0.080292,((('S_furcatus':0.048820,'S_tigrensi':0.058974):0.023057,('S_generali':0.009726,'S_pacarae':0.005888):0.143555):0.006403,'S_testudin':0.116693):0.027646):0.052470):0.021807,((('S_bottimer':0.091291,'S_chalcode':0.094831):0.011116,('S_mexicanu':0.130655,'S_subaeneu':0.121938):0.030474):0.115808,(('S_chihuahu':0.002953,'S_pygidial':0.001701):0.065965,'S_sordidus':0.052645):0.122450):0.027370,('S_pruininu':0.185669,'S_vachelli':0.171345):0.049349):0.023454):0.061852,(('Se_breveap':0.219859,'Se_morosus':0.166834):0.087695,'B_rufimanu':0.235478):0.036347,'M_insolitu':0.284645);
tachinids (MIXED; Insecta: Tachinidae)	12	D	(TVM+I+G)	Singer *et al.*. (2003)	AF366652, AF366653, AF366650, AF366689, AF366660, AF366656, AF366669, AF366659, AF366666, AF366692, AF366676, AF366677, AJ558185*	('Cal_dubia':0.020665,((((('Apl_thecla':0.015837,((('Hyph_hypha':0.026657,'Exor_mella':0.015287):0.000130,('Winth_rufo':0.006736,'Lesch_adus':0.006741):0.015450):0.001008,('Carc_recli':0.012917,'Chaet_mont':0.021936):0.003939):0.006400):0.004761,('Austro_sp':0.010497,'Chet_tachi':0.008017):0.023741):0.002844,'Drino_inco':0.023749):0.003336,'Lesp_archi':0.033338):0.008636,'Siph_sp.':0.016443);
Tetraopes/*Phaea* (milkweed beetles; Insecta: Cerambycidae)	17	A	(GTR+G)	Farrell and Mitter (1998; 2001)	AF267471, AF267472, AF267473, AF267474, AF267475, AF267476, AF267477, AF267478, AF267479, AF267480, AF267481, AF267482, AF267483, AF267467, AF267468, AF267469, AF267466*, AF267465*	((('T_discoide':0.033874,'T_umbonatu':0.095221):0.013807,(((('T_melanuru':0.015278,'T_quinquem':0.001719):0.034937,'T_texanus':0.051877):0.019750,(('T_annulatu':0.031740,'T_pilosus':0.030661):0.019983,(('T_tetropth':0.014425,'T_mandibul':0.009593):0.013254,('T_basalis':0.021373,(('T_femoratu':0.010340,'T_sublaevi':0.010142):0.016996,'T_varicorn':0.025052):0.009355):0.020503):0.008381):0.025458):0.010991,(('P_maryanna':0.070334,'P_mirabili':0.128859):0.018676,'P_biplagia':0.061659):0.029296):0.030368):0.110594,('E_laevis':0.164887,'M_gigas':0.284741):0);
Timema (walkingsticks; Insecta: Timematidae)	22	A	(GTR+I+G)	Sandoval *et al.*. (1998), Crespi & Sandoval (2000), Law & Crespi (2002)	AF005331, AF005334, AF005345, AF005344, AF005343, AF005332, AF005340, AF005337, AF005339, AF005333, AF410147, AF410150, AF410104, AF005335, AF005336, AF410151, AF005342, AF005341, AF410021, AF005338, AF410081, AF005330, AF005346*, AF005347*, AY176057*	(Blattella:0.305078,((T_nakipa:0.053318,(T_boharti:0.106003,(T_podura2:0.020665,(T_geneviev:0.032877,(T_podura1:0.031286,(T_bartmani:0.018756,(T_chumash:0.008507,T_tahoe:0.021059):0.001834):0.002998):0.011011):0.005369):0.040734):0.16075):0.153876,((((T_landelse:0.044832,((T_califor3:0.012395,(T_poppensi:0,T_shepardi:0.002614):0.025283):0.023019,(T_califor2:0,T_califor1:0.008931):0.025244):0.015117):0.009376,(T_knulli:0.006094,(T_douglasi:0.016185,T_petita:0):0.002477):0.005451):0.151906,(T_cristina:0.022636,T_monikens:0.042939):0.161822):0.111493,(T_nevadens:0.324569,(T_dorothea:0.142497,T_ritensis:0.210834):0.042505):0.072285):0.050282):0.344714,(Baculum:0.338121,Anisomorph:0.175894):0.701703);

Tomoplagia (flies on plants Genus/tribes; Insecta: Tephritidae)	19	C + 16S	(TIM+I+G)	Yotoko *et al.*. (2005)	(AY164603,AY165295), (AY164607,AY165298), (AY164599,AY165293), (AY164606,AY165297), (AY164597,AY165289), (AY164622,AY165307), (AY164598,AY165291), (AY164608,AY165299), (AY164615,AY165304), (AY164614,AY165303), (AY164612,AY165302), (AY164610,AY165301), (AY164589,AY165285), (AY164621,AY165306), (AY164591,AY165286), (AY164618,AY165305), (AY164595,AY165288), (AY164602,AY165294), (AY164605,AY165296), (NC_000857,NC_000857)*, (NC_001322,NC_001322)*	((argentinie:0.056894,(((((pseudopeni:0.022252,(fiebrigi:0.015881,reimoseri:0.011620):0.014727):0.024550,bicolor:0.048875):0.002961,(minuta:0.032868,((voluta:0.008729,rupestris:0.008087):0.025509,(sp3:0.019529,sp2:0.019400):0.004486):0.006646):0.005417):0.004874,(((grandis:0.026155,achromopte:0.038116):0.004618,tripunctat:0.026472):0.010863,trivittata:0.052610):0.017411):0.007180,((rudolphi:0.100986,brasiliens:0.047521):0.005071,(costalimai:0.055653,(sp1:0.021935,incompleta:0.031841):0.026428):0.014036):0.014465):0.010703):0.051446,'C_capitata':0.089139,'D_yakuba':0.097996);
Vidua (parasitic finches; Aves: Icteridae)	15	F	(HKY+G)	Sorenson *et al.*. (2003)	AF407105, AF407109, AF407108, AY324263, AY324264, AF407107, AY322651, AY322820, AY865483, AY322795, AY322819, AY322692, AY322676, AY322837, AY322835, AY324242*, AY324254*, AY324249*	(('A_imberbis':0.238790,((('V_macroura':0.043938,(('V_paradisa':0.006609,'V_obtusa':0.001787):0.014693,'V_oriental':0.003554):0.053886):0.013466,'V_hypocher':0.048304):0.012382,((('V_chalybea':0.001837,'V_codringt':0,'V_funerea':0.000916):0.005710,('V_wilsoni':0,('V_larvatic':0.000913,'V_nigeriae':0.000913,'V_raricola':0):0.000915):0.001731):0.010833,('V_regia':0.016567,'V_fischeri':0.022396):0.008758):0.024026):0.071002):0.197786,'E_paludico':0.066683,('P_afra':0.086421,'C_monteiri':0.103469):0.038535);

CHAPTER THREE

HOST PHYLOGENETIC DIVERSITY SATURATES AT BROAD HOST RANGES IN GENERALIST PARASITES

Abstract

Generalist parasites are widely thought to be opportunists, unconstrained by the ecology and evolution of their hosts—in contrast to specialists, which often have a few, closely related hosts that are phenotypically similar[1,2,3]. Here I show that generalist avian brood parasites, with hundreds of host species, are equally constrained in the phylogenetic diversity of hosts as more specialized brood parasites with few host species, and that this constraint is limited to groups of hosts that are more often closely related than would be expected by chance. Remarkably, 96% of the variation in host ranges among brood parasites is invariant in host phylogenetic diversity. This finding challenges the conventional wisdom that generalists are insensitive to the relatedness of hosts, and indicates a strong, but previously unknown, limitation to host specificity: even the most generalized parasite is restricted to a subset of host species that are relatively narrow in their phylogenetic diversity.

Introduction, Results and Discussion

Closely related species are often more phenotypically similar than distantly related species[2], and for specialist parasites, this phenotypic distance among species can constrain the range of hosts on which they can parasitize[2]. Here, host species that are closely related to hosts to which specialists are already adapted may be more alike as resources than distantly related hosts[3,4], and in nature, it is repeatedly observed that specialist parasites with few host species often parasitize a group of closely related

hosts[1]. Generalists, in contrast, are assumed to be more constrained by geographical limitations in the richness of potential host species[2,3] and the susceptibility of these hosts to parasitism[4] because ecological interactions can occur with hosts belonging to broad taxonomic classes. Despite these observations, it remains untested whether the relatedness of host species (or phylogenetic diversity) is also quantitatively limited within or among generalist parasites. Or more simply, are generalist parasites with many host species also constrained to a relatively narrow level of host phylogenetic diversity?

I tested this hypothesis by determining the relationship between host range and host phylogenetic diversity for 70 of the 95 known species of obligate avian brood parasite—birds that parasitize parental care by laying eggs in nests of other bird species[4]. Host range is the conventional ecological index of host specificity and is measured as the richness of known host species, whereas host phylogenetic diversity (*PD*) is an evolutionary index that quantifies the relatedness of hosts by taking the average length of their shared evolutionary pathways[3,4]. To estimate the *PD* of each host range, I used a genus-level phylogeny of all avian hosts[5] and calculated the average of their shared phylogenetic history with $PD = c^{-1}\sum\sum_{i<j} BL_{ij}$, where $c = 2N(N-1)$, N is the total number hosts in the phylogeny, and *BL* is the phylogenetic branch-length distance from the i^{th} species to the shared node of the j^{th} host species. Finally, avian brood parasites are an important model to test hypotheses on host specificity because there is a working hypothesis of the phylogenetic history of all birds[7], and because host ranges are based on annotated host records from nearly a century of detailed natural history observations[7].

I found that host phylodiversity scales positively with host range when host ranges are narrow, but rapidly saturates at broad host ranges (Figure 3.1). This finding points to a previously unknown asymptotic relationship between the richness and

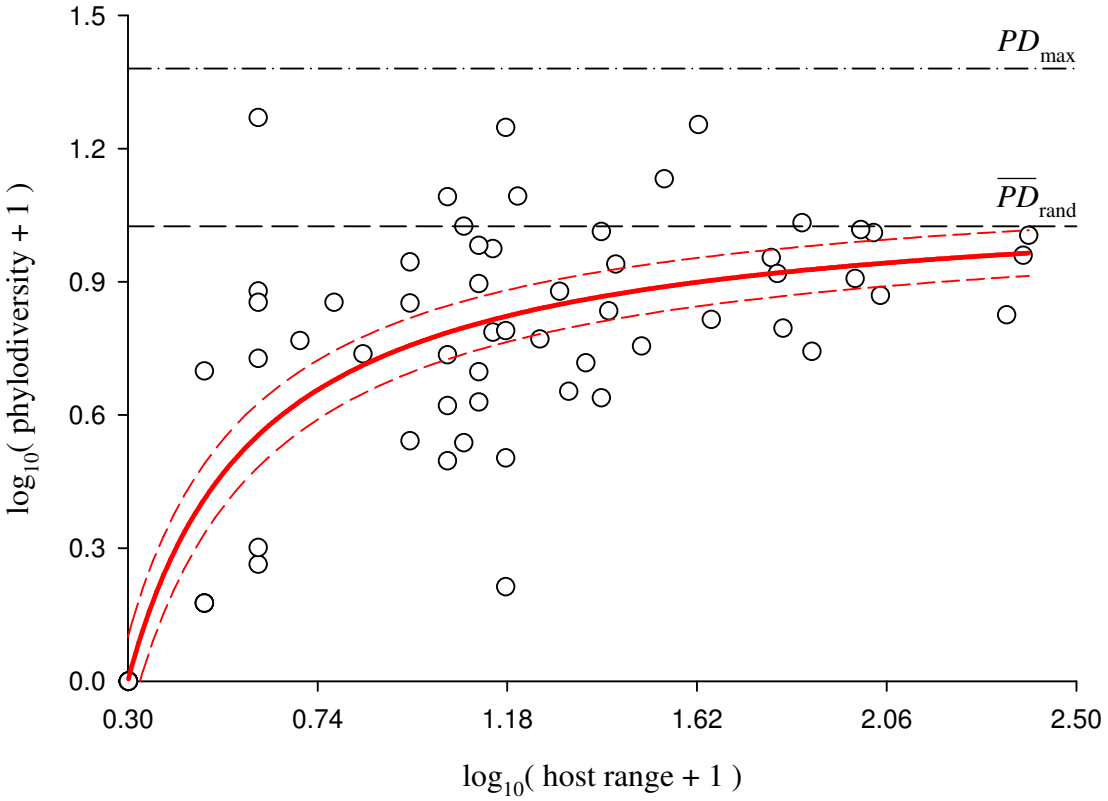


Figure 3.1. Log_{10}–$\log_{10}$ plot of the relationship between host phylodiversity (*PD*) and host range (host species richness) of 70 species of avian brood parasite. Host range and *PD* vary broadly among these parasites, and range from 1 to 244 host species and 0 to 17.6 *PD*, respectively. The solid red curve describes the relationship between host range and *PD* as defined by a reciprocal regression ($y = 1.10 - 0.33\, x^{-1}$, $R^2 = 0.72$, $F = 170.3$, d.f. = 69, $p < 0.001$), and the dashed red curves are the 95% confidence intervals for this regression. The horizontal dash-dot line is the theoretical maximum of host phylodiversity ($PD_{max} = 23$), and the dashed line is the average *PD* for a phylogenetically random group of host species ($\overline{PD}_{rand} = 9.59$)[5].

phylogenetic relatedness of hosts. Using a generalization of the rainbow test[5] to evaluate the threshold at which host phylodiversity reaches saturation, I found that brood parasites with 11 or more host species do not significantly differ in *PD* (minimum null, reciprocal regression: $R^2 = 0.09$, $F = 3.4$, d.f. = 35, $p = 0.07$). The inclusion of specialist parasites with one host species and a *PD* of zero were not necessary to describe this saturation effect (reciprocal regression excluding specialists: $R^2 = 0.32$, $F = 25.7$, d.f. = 56, $p < 0.001$). These results indicate that there is no apparent change in host phylogenetic diversity for brood parasites with broad host ranges (i.e. with 11 to nearly 250 host species).

One potential horizontal asymptote for this saturation effect is the total branch-length distance of the host phylogeny ($PD_{\max}$). $PD_{\max}$ is the ceiling to the range of *PD* values that can be estimated from the host phylogeny. However, *PD* saturates far below this limit (Figure 3.1). Another asymptote could occur if host ranges are a phylogenetically random group of hosts. If this is the case, then sampling theory would predict that estimates of host phylodiversity would be distributed symmetrically around the mean random phylodiversity ($\overline{PD}_{rand}$) in the shape of a funnel[5]—with greater variation in *PD* at more narrow than broad host ranges. In contrast to this null expectation, host phylodiversity at broad host ranges saturates significantly below $\overline{PD}_{rand}$ (Figure 3.1), where the majority of observed *PD* values are distributed asymmetrically below this expected random host phylodiversity (Figure 3.1; Egger's test for funnel asymmetry[5]: $a = -1.4$, 95% LCI = –1.7, 95% UCI = –1.1, $t = -9.2$, $p < 0.001$). Deviation from symmetry and lack of saturation at $\overline{PD}_{rand}$ further indicates that host ranges are often constrained to groups of related hosts, and that brood parasites with phylogenetically random hosts are uncommon.

Host specificity also differs among avian brood parasites because clades of parasitism are independently evolved[5,6] and endemic to continents that differ in

richness and phylogenetic diversity of potential avian hosts[7]. Simulating $\overline{PD}_{rand}$ separately for each continent did reveal a phylogeographic structure[5], but this geographic bias to $\overline{PD}_{rand}$ is expected given that the shape and structure of the host phylogeny are epiphenomena of endemic radiations of brood parasites[6]. Despite these issues, the saturation of host phylodiversity at broad host ranges appears less sensitive to endemic radiations because excluding clades of brood parasite or entire continents from the overall analysis did not reveal bias[2,5]. This insensitivity was echoed by the lack of a significant contribution of the shared ancestry among avian brood parasites (phylogenetically-independent reciprocal regression: $R^2 = 0.55$, $F = 84.3$, d.f. = 19.54, $p < 0.001$). Thus the overall relationship between host range and host phylodiversity remains: avian brood parasites endemic to specific continents still parasitized host species that were more closely related then expected by chance[5].

Overall, 96% of the variation in the host ranges among avian brood parasites is invariant with host phylogenetic diversity, and the magnitude of observed differences in host specificity among generalists is non-random and biased by the phylogenetic non-independence of hosts. Thus, despite the 'jack of all trades' reputation of some generalist parasites, these finding indicate that even the most generalized parasites can be restricted to a subset of hosts bounded by a relatively narrow phylogenetic diversity. More generally, should saturation of host phylodiversity at broad host ranges be a widespread phenomenon among parasites (i.e. flukes, herbivorous insects, parasitic plants), then a redefinition of host specificity is warranted and should emphasize the quantitative relationship between the richness and the phylogenetic relatedness of hosts.

REFERENCES

1. Price, P. W. Evolution 31, 405–420 (1978).
2. Felsenstein, J. Am. Nat. 125, 1–15 (1985).
3. Poulin, R. and Mouillot, D. Parasitology 126, 473–480 (2003).
4. Johnsgard, P. A. The avian brood parasites: deception at the nest. Oxford University Press, New York (1997).
5. see APPENDIX.
6. Sibley, C. G., and Ahlquist, J. E. Phylogeny and classification of birds: a study in molecular evolution. Yale University Press, New Haven (1990).
7. Davies, N. B. Cuckoos, Cowbirds and other Cheats. T. and A. D. Poyser, London (2000).

APPENDIX

Host Range and Host Phylodiversity

Host ranges and host species of avian brood parasites were obtained from published and annotated lists (Johnsgard 1997; Lowther 2004a, 2004b, 2005a, 2005b, 2005c, 2005d). Host phylodiversity (*PD*) is an estimate of phylogenetic relatedness of host species and is defined by

$$PD = \frac{\sum\sum_{i<j} BL_{ij}}{2N(N-1)}. \qquad (3.A1)$$

where N is the total number of host species in the host phylogeny and *BL* is the branch-length from i^{th} species to the node shared with j^{th} host species (Clarke and Warwick 2001; Poulin and Mouillot 2003).

To standardize the scale of *PD*, I estimated *PD* for each species of brood parasite using a single phylogenetic tree of all host genera (N = 498) based on the topology and internode branch-lengths of Sibley and Ahlquist (1990) DNA-DNA hybridization analysis and classification of birds (see Appendix Table 3.1). The node depth of this phylogeny was 23 (see Appendix Table 3.1), and this depth was used as the maximum branch length for the host phylogeny, such that *PD* could range from a maximum of 23 (PD^{max}; with a host range of two of the most distantly related hosts on that tree) to zero (PD^{min}). PD^{min} is given to specialist parasites with only a single host species (following Poulin *et al.* 2003). Bird genera were treated as polytomies to their respective families, and had internode *BL* scaled to one. Parasites having many host species within a single genus had half the *BL* scored for that genus (i.e. a branch length distance of 0.5).

$\overline{PD}_{rand}$ is the average host phylodiversity for a phylogenetically random assortment of host species. $\overline{PD}_{rand}$ was estimated by randomly sampling host species from the host phylogeny (Appendix Table 3.1) for the full range of observed host

Appendix Table 3.1. Hypothesized phylogenetic relationships of avian brood parasites (N = 70) and host genera (N = 498).

phylogeny in parenthetical (Newick) format

avian brood parasites

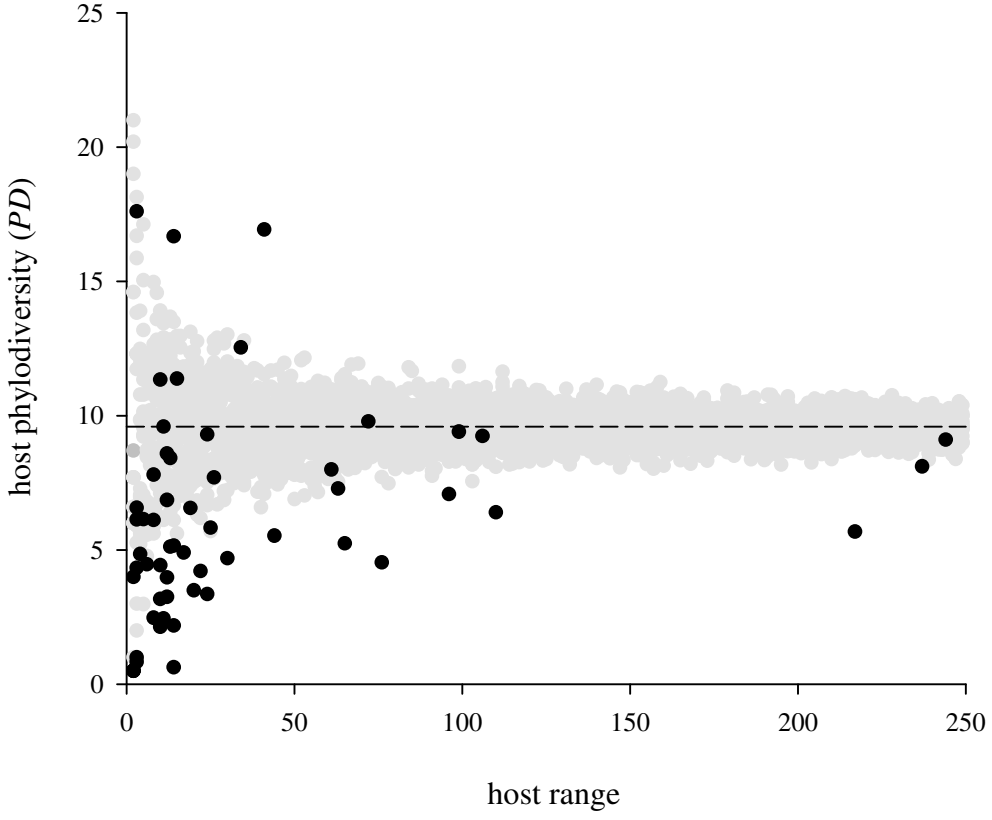
(*Heteronetta_atricapilla*,(((*Tapera_naevia*,(*Dromococcyx_pavoninus*,*Dromococcyx_phasianellus*)),((*Scythrops_novaehollandiae*,(*Eudynamys_cyanocephala*,*Eudynamys_scolopacea*,*Eudynamys_taitensis*)),(((*Clamator_coromandus*,*Clamator_glandarius*),(*Oxylophus_levaillanti*,*Oxylophus_jacobinus*)),(*Pachycoccyx_audeberti*,((*Chrysococcyx_basalis*,*Chrysococcyx_caprius*,*Chrysococcyx_crassirostris*,*Chrysococcyx_cupreus*,*Chrysococcyx_klaas*,*Chrysococcyx_lucidus*,*Chrysococcyx_maculatus*,*Chrysococcyx_osculans*,*Chrysococcyx_rufomerus*,*Chrysococcyx_xanthorhynchus*,(*Chrysococcyxminutillus*,*Chrysococcyx_m_russatus*)),(*Surniculus_lugubris*,((*Cacomantis_flabelliformis*,(*Cacomantis_merulinus*,*Cacomantis_variolosus*)),(*Cuculus_clamosus*,*Cuculus_gularis*,*Cuculus_micropterus*,*Cuculus_rochii*,*Cuculus_sparveroides*,(*Cuculus_saturatus*,(*Cuculus_fugax_hyperythrus*,*Cuculus_f_nisicolor*)),(*Cuculus_solitarius*,(*Cuculus_poliocephalus*,(*Cuculus_pallidus*,*Cuculus_canorus*))))))))))),(((*Prodotiscus_insignis*,*Prodotiscus_regulus*,*Prodotiscus_zambesiae*),(*Indicator_conirostris*,*Indicator_indicator*,*Indicator_meliphilus*,*Indicator_minor*,*Indicator_variegatus*)),((*Molothrus_rufoaxillaris*,(*Scaphidura_oryzivora*,(*Molothrus_aeneus*,(*Molothrus_ater*,*Molothrus_bonariensis*)))),(*Anomalospiza_imberbis*,(*Vidua_macroura*,((*Vidua_togoensis*,(*Vidau_paradisaea*,*Vidau_obtusa*),(*Vidua_orientalis*,*Vidau_interjecta*)),(*Vidua_hypocherina*,((*Vidua_regia*,*Vidua_fischeri*),(*Vidua_chalybeata*,((*Vidua_funerea*,*Vidua_purpurascens*),(*Vidua_wilsoni*,(*Vidua_larvaticola*,*Vidua_raricola*))))))))))));

host genera

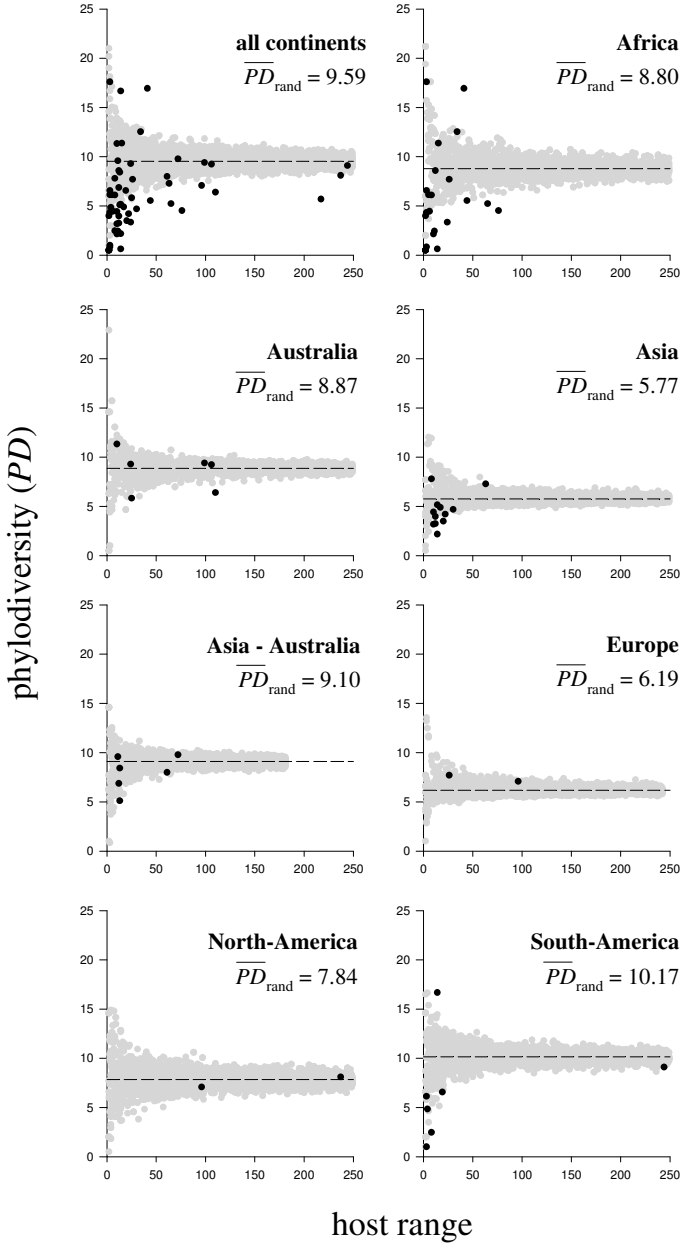
((((*Dendrocygna*:4.7,((*Anas*:1,*Aythya*:1,*Netta*:1):1,*Coscoroba*:2):2.7):6.3,*Chauna*:11):6.8,((*Gallus*:1,*Phasianus*:1):1,*Tympanuchus*:2):15.8):5.1,(((*Campethera*:1,*Colaptes*:1,*Dendropicos*:1,*Geocolaptes*:1,*Jynx*:1,*Melanerpes*:1):9.4,(*Gymnobucco*:1,*Lybius*:1,*Pogoniulus*:1,*Stactolaema*:1,*Trachyphonus*:1,*Tricholaema*:1):9.4):10.8,(((*Colius*:1,*Urocolius*:1):18.4,(*Coccyzus*:18.6,(*Cyanoramphus*:18,((((*Columbina*:1,*Geopelia*:1,*Metriopelia*:1,*Stigmatopelia*:1,*Streptopelia*:1,*Zenaida*:1):14.7,(((*Fulica*:1,*Paradirallus*:1,*Rallus*:1):13,*Aramus*:14):1,((((*Falco*:1,*Milvago*:1,*Rostrhamus*:1):9.1,(*Accipiter*:1,*Buteo*:1):9.1):1.2,(*Tachybaptus*:9.8,(*Nycticorax*:7.3,(*Ajaia*:1,*Plegadis*:1):6.3):2.5):1.5):2.3,(((*Larus*:1,*Sterna*:1):6.7,*Charadrius*:7.7):2.8,(*Actitis*:1,*Bartramia*:1,*Phalaropus*:1,*Tringa*:1):9.5):3.1):1.4):0.7):0.8,(*Dasyornis*:14.6,(*Acanthiza*:1,*Acanthornis*:1,*Aphelocephala*:1,*Calamanthus*:1,*Gerygone*:1,*Hylacola*:1,*Origma*:1,*Pycnoptilus*:1,*Pyrrholaemus*:1,*Sericornis*:1,*Smicrornis*:1):13.6,(*Mohoua*:1,*Chthonicola*:1):13.6,(*Dulus*:7.7,((((*Amytornis*:1,*Malurus*:1,*Stipiturus*:1):5.1,((*Acanthagenys*:1,*Acanthorhynchus*:1,*Anthochaera*:1,*Anthornis*:1,*Certhionyx*:1,*Conopophila*:1,*Entomyzon*:1,*Epthianura*:1,*Grantiella*:1,*Lichenostomus*:1,*Lichmera*:1,*Manorina*:1,*Meliphaga*:1,*Melithreptus*:1,*Myzomela*:1,*Notiomystis*:1,*Philemon*:1,*Phylidonyris*:1,*Plectorhyncha*:1,*Prosthemadra*:1,*Ramsayornis*:1,*Xanthomyza*:1):3.1,*Pardalotus*:4.1):2):0.2,(*Pomatostomus*:4.5,((*Corvinella*:1,*Lanius*:1):3,((*Cyclarhis*:1,*Hylophilus*:1,*Vireo*:1):2.7,*Daphoenositta*:3.7,(*Dryoscopus*:1,*Laniarius*:1,*Tchagra*:1,*Telophorus*:1):2.7,(*Falcunculus*:1,*Oreoica*:1):2.7,*Dicrurus*:3.7,*Rhipidura*:3.7,(*Batis*:1,*Platysteira*:1,*Prionops*:1):2.7,(*Colluricincla*:1,*Pachycephala*:1):2.7,(*Campephaga*:1,*Coracina*:1,*Lalage*:1,*Oriolus*:1,*Pericrocotus*:1,*Ptiloris*:1,*Sphecotheres*:1):2.7,(*Grallina*:1,*Hypothymis*:1,*Monarcha*:1,*Myiagra*:1,*Terpsiphone*:1):2.7,*Aegithina*:3.7,*Psophodes*:3.7,(*Artamus*:1,*Cracticus*:1,*Gymnorhina*:1,*Strepera*:1):2.7,(*Corcorax*:1,*Corvus*:1,*Cyanocitta*:1,*Cyanocorax*:1,*Cyanopica*:1,*Garrulus*:1,*Pica*:1,*Ptilostomus*:1,*Urocissa*:1):2.7):0.3):0.5):1.8):0.3,(*Climacteris*:1,*Cormobates*:1):5.6):1.1,(((*Dumetella*:1,*Mimus*:1,*Oreoscoptes*:1,*Toxostoma*:1):4.5,*Chloropsis*:5.5,(*Bombycilla*:1,*Phainopepla*:1):4.5,(*Cinclus*:4.6,(((*Bradornis*:1,*Cyanoptila*:1,*Cyornis*:1,*Dioptrornis*:1,*Eumyias*:1,*Ficedula*:1,*Horizorhinus*:1,*Melaenornis*:1,*Muscicapa*:1,*Myioparus*:1,*Myrmecocichla*:1,*Niltava*:1,*Sigelus*:1):1,(*Culicicapa*:1,*Drymodes*:1,*Eopsaltria*:1,*Heteromyias*:1,*Melanodryas*:1,*Microeca*:1,*Petroica*:1,*Poecilodryas*:1,*Tregellasia*:1):1,*Polioptila*:2,(*Brachypteryx*:1,*Catharus*:1,*Hylocichla*:1,*Ixoreus*:1,*Monticola*:1,*Myadestes*:1,*Myiophonus*:1,*Sialia*:1,*Turdus*:1,*Zoothera*:1):1,(*Cercomela*:1,*Cercotrichas*:1,*Cinclidium*:1,*Cochoa*:1,*Copsychus*:1,*Cossypha*:1,*Enicurus*:1,*Erithacus*:1,*Hodgsonius*:1,*Luscinia*:1,*Oenanthe*:1,*Phoenicurus*:1,*Pinarornis*:1,*Pogonocichla*:1,*Rhyacornis*:1,*Saxicola*:1,*Swynnertonia*:1,*Tarsiger*:1,*Thamnolaea*:1):1):2,(*Acridotheres*:1,*Cinnyricinclus*:1,*Lamprotornis*:1,*Onychognathus*:1,*Spreo*:1,*Sturnus*:1):3):0.6):0.9):1.1,(((*Sitta*:4.9,*Certhia*:4.9,(*Campylorhynchus*:1,*Cistothorus*:1,*Salpinctes*:1,*Thryomanes*:1,*Thryothorus*:1,*Troglodytes*:1):3.9):0.8,((*Baeolophus*:1,*Parus*:1,*Poecile*:1):4.5,((*Aegithelos*:1,*Psaltriparus*:1):4.3,((*Hirundo*:1,*Hylochelidon*:1,*Petrochelidon*:1,*Phaeoprogne*:1,*Progne*:1,*Riparia*:1,*Tachycineta*:1):4,(*Regulus*:4.6,((*Andropadus*:1,*Hemixos*:1,*Hypsipetes*:1,*Ixos*:1,*Phyllastrephus*:1,*Pycnonotus*:1,*Spizixos*:1):3.4,((*Apalis*:1,*Camaroptera*:1,*Cisticola*:1,*Prinia*:1,*Scotocera*:1,*Spiloptila*:1):3.3,((*Speirops*:1,*Zosterops*:1):3,((*Abroscopus*:1,*Acrocephalus*:1,*Bradypterus*:1,*Cettia*:1,*Eremomela*:1,*Hippolais*:1,*Locustella*:1,*Orthotomus*:1,*Phylloscopus*:1,*Seicercus*:1,*Sphenoeacus*:1,*Sylvietta*:1,*Tesia*:1,*Urosphena*:1):1,(*Garrulax*:1,*Liocichla*:1):1,(*Cinclorhamphus*:1,*Eremiornis*:1,*Megalurus*:1):1):2,(*Auriparus*:1,*Pholidornis*:1):3,(*Actinodura*:1,*Alcippe*:1,*Babax*:1,*Chamaea*:1,*Heterophasia*:1,*Illadopsis*:1,*Leiothrix*:1,*Macronous*:1,*Malacocincla*:1,*Malacopteron*:1,*Napothera*:1,*Neomixis*:1,*Panurus*:1,*Paradoxornis*:1,*Pellorneum*:1,*Phyllanthus*:1,*Pnoepyga*:1,*Pomatorhinus*:1,*Pteruthius*:1,*Rhopocichla*:1,*Rimator*:1,*Spelaeornis*:1,*Stachyris*:1,*Timalia*:1,*Turdoides*:1,*Yuhina*:1):3,*Sylvia*:4):0.3):0.1):0.2):0.4):0.3):0.2):0.2):0.3,((*Alauda*:1,*Calandrella*:1,*Eremophila*:1,*Galerida*:1,*Lullula*:1):4.3,(((((*Cardinalis*:1,*Guiraca*:1,*Passerina*:1,*Pheucticus*:1,*Saltator*:1):1,(*Aimophila*:1,*Amaurospiza*:1,*Ammodramus*:1,*Amphispiza*:1,*Arremon*:1,*Arremonops*:1,*Atlapetes*:1,*Calamospiza*:1,*Calcarius*:1,*Chondestes*:1,*Coryphospingus*:1,*Diuca*:1,*Donacospiza*:1,*Emberiza*:1,*Emberizoides*:1,*Embernagra*:1,*Gubernatrix*:1,*Junco*:1,*Loxigilla*:1,*Melophus*:1,*Melospiza*:1,*Melozone*:1,*Miliaria*:1,*Oryzoborus*:1,*Paroaria*:1,*Passerculus*:1,*Passerella*:1,*Phrygilus*:1,*Piezorhina*:1,*Pipilo*:1,*Pooecetes*:1,*Poospiza*:1,*Rhodospingus*:1,*Saltatricula*:1,*Sicalis*:1,*Spizella*:1,*Sporophila*:1,*Tiaris*:1,*Zonotrichia*:1):1,(*Basileuterus*:1,*Dendroica*:1,*Geothlypis*:1,*Helmitheros*:1,*Icteria*:1,*Limnothlypis*:1,*Mniotilta*:1,*Myioborus*:1,*Oporornis*:1,*Parula*:1,*Protonotaria*:1,*Seiurus*:1,*Setophaga*:1,*Vermivora*:1,*Wilsonia*:1):1,(*Conirostrum*:1,*Cypsnagra*:1,*Dacnis*:1,*Habia*:1,*Hemithraupis*:1,*Piranga*:1,*Ramphocelus*:1,*Rhodinocichla*:1,*Schistochlamys*:1,*Tachyphonus*:1,*Tangara*:1,*Thlypopsis*:1,*Thraupis*:1):1,(*Agelaioides*:1,*Agelaius*:1,*Agelasticus*:1,*Amblycercus*:1,*Amblyramphus*:1,*Cacicus*:1,*Chrysomus*:1,*Curaeus*:1,*Dives*:1,*Dolichonyx*:1,*Euphagus*:1,*Gnorimopsar*:1,*Gymnomystax*:1,*Icterus*:1,*Macroagelaius*:1,*Psarocolius*:1,*Pseudoleistes*:1,*Quiscalus*:1,*Sturnella*:1,*Xanthocephalus*:1,*Xanthopsar*:1):1,(*Amandava*:1,*Coccopygi*:1,*Estrilda*:1,*Granatina*:1,*Lagonosticta*:1,*Passer*:1,*Petronia*:1,*Pytilia*:1,*Sporopipes*:1):1,(*Lonchura*:1,*Neochmia*:1,*Poephila*:1,*Stagonopleura*:1,*Taeniopygia*:1):1,*Prunella*:2,(*Amblyospiza*:1,*Anaplectes*:1,*Euplectes*:1,*Malimbus*:1,*Ploceus*:1,*Plocepasser*:1,*Pseudonigrita*:1,*Quelea*:1):1,(*Anthus*:1,*Motacilla*:1):1):2.7,(*Aethopyga*:1,*Anthreptes*:1,*Arachnothera*:1,*Dicaeum*:1,*Nectarinia*:1):3.7):0.2,(*Eophona*:1,*Fringilla*:1,*Poliospiza*:1,*Pyrrhula*:1,*Serinus*:1,*Uragus*:1,*Carduelis*:1,*Carpodacus*:1,*Coccothraustes*:1,*Coccothrautes*:1,*Leucosticte*:1,*Loxia*:1):3.9):0.4):0.7):0.6):1.1):6.9,(((*Phytotoma*:1,*Tityra*:1,*Xenopsaris*:1):7.7,(*Arundinicola*:1,*Contopus*:1,*Elaenia*:1,*Empidonax*:1,*Empidonomus*:1,*Fluvicola*:1,*Hemitriccus*:1,*Hymenops*:1,*Knipolegus*:1,*Machetornis*:1,*Muscigralla*:1,*Muscisaxicola*:1,*Myiarchus*:1,*Myiodynastes*:1,*Myiophobus*:1,*Myiornis*:1,*Myiozetetes*:1,*Pachyramphus*:1,*Phelpsia*:1,*Pitangus*:1,*Poecilotriccus*:1,*Pseudocolopteryx*:1,*Pyrocephalus*:1,*Rhynchocyclus*:1,*Satrapa*:1,*Sayornis*:1,*Serpophaga*:1,*Suiriri*:1,*Todirostrum*:1,*Tyrannus*:1,*Xolmis*:1):7.7,((*Dysithamnus*:1,*Sakesphorus*:1,*Taraba*:1,*Thamnophilus*:1):7.4,((*Anumbius*:1,*Asthenes*:1,*Certhiaxis*:1,*Coryphistera*:1,*Cranioleuca*:1,*Furnarius*:1,*Leptasthenura*:1,*Limnoctites*:1,*Limnornis*:1,*Phacellodomus*:1,*Philydor*:1,*Phleocryptes*:1,*Pseudoseisura*:1,*Schoeniophylax*:1,*Synallaxis*:1):6,*Xiphorhynchus*:7,(*Rhinocrypta*:1,*Rhinopomastus*:1):6):1.4):0.3):2,*Eurylaimus*:10.7):3.9):1.9):0.9,(*Archilochus*:16.2,(*Apus*:1,*Micropus*:1):15.2):1.2):0.6):0.6,*Centropus*:18.6):0.8):0.5,(((*Merops*:14.6,(*Halcyon*:1,*Ispidina*:1):13.6):1.3,*Coracias*:15.9):2.4,(*Upupa*:9.9,*Phoeniculus*:9.9):8.4):1.6):1.3):1.7);

ranges (e.g., from 2 host species to 250 hosts). Appendix Figure 3.1 shows the funnel distribution of $\overline{PD}_{rand}$ for each host range, where at narrow host ranges *PD* estimates tend to show greater variation in both magnitude and direction (e.g., either groups of closely related hosts or groups of distantly related hosts). In addition, $\overline{PD}_{rand}$ was estimated for each continent, as many brood parasites are endemic to particular continents can differ in richness and phylogenetic diversity of hosts. These simulations indicate that $\overline{PD}_{rand}$ can vary broadly among continents (see Appendix Figure 3.2), but generally, observed *PD* values among brood parasites are often still distributed asymmetrically towards groups of closely related hosts (i.e. African brood parasites still falling below $\overline{PD}_{rand}$).

The symmetry of the distribution around $\overline{PD}_{rand}$ also indicates that random sampling at narrow host ranges can produce estimates of *PD* with broad variation at narrow host ranges, with values falling far above or below $\overline{PD}_{rand}$ (see Appendix Figure 3.1). Funnel symmetry is often used in meta-analysis to test for sampling biases that occur between an experiment's sample size and observed effect size, and how the two interact to influence publication bias (Egger *et al.* 1997). I tested for symmetry by modifying Egger's regression test for funnel symmetry for meta-analysis. Specifically, for each parasite species with a *PD* and host range (*HR*), symmetry was determined by testing whether the intercept (a) is non-zero for the regression $x = \ln(PD + 1/HR + 1)$ versus $y = \ln(1/HR + 1)$. A significant intercept indicates that *PD* values among host ranges are not distributed symmetrically; a pattern expected by chance should host ranges of parasites have consisted of a random sampling of hosts.



Appendix Figure 3.2. Funnel plot of phylodiversity (*PD*) values estimated by randomly sampling hosts from the host phylogeny (Appendix Table 3.1) for the full range of observed host ranges (from 2 to 250 host species). The dashed black line indicates $\overline{PD}_{rand}$ = 9.59, the phylogenetically random mean of *PD* across host ranges. *PD* values generated by simulations are in light grey (15 replicates for each host range), whereas the *PD* values in black are the *PD* values calculated from the host ranges of avian brood parasites (*N* = 70 species).



Appendix Figure 3.3. Funnel plots of *PD* values for specific continent regions where avian brood parasites are endemic. Estimates are similar to Appendix Figure 3.2. Dashed white lines indicates $\overline{PD}_{rand}$, the mean phylogenetically random *PD* across host ranges. *PD* values generated by simulations are in black (15 replicates for each host range), whereas the *PD* values for host ranges of avian brood parasites belonging to specific continents are in light grey.

Conventional and Phylogenetically Controlled Analyses

I first determined whether host range predicts *PD* using an inverse first-order polynomial regression (reciprocal regression). I also contrasted various regression models (e.g., second-order polynomials), but found that the reciprocal regression best explained the relationship between host range and *PD* with the least amount of error. Host ranges and *PD* were $\log_{10}(x + 1)$ transformed prior to analyses to standardize variances (Harvey 1982).

I then assumed that brood parasites are part of a hierarchical, branching phylogeny (see below and Appendix Table 3.1), and that host ranges and *PD* may have a correlated evolutionary history. I first used phylogenetically independent contrasts to test for confounds in host specificity due to shared ancestry of brood parasites (Felsenstein 1985). Independent contrasts were calculated and analyzed with the PDTREE module of the Phenotypic Diversity Analysis Programs (PDAP; Garland *et al.* 1999). Regression with independent contrasts was forced through the origin (Felsenstein 1985). I accounted for polytomies in the brood parasite phylogeny by conservatively bounding the degrees of freedom of the regression: as the total number of nodes minus one (following Purvis and Garland 1993). This analysis found that host range and host phylodiversity do share a correlated evolutionary history (phylogenetic independent contrasts: $R^2 = 0.32$, $F = 32.4$, d.f. = 48, $p < 0.001$).

A more comparable analysis is to repeat the reciprocal regression while accounting for the brood parasite evolutionary history. I used Pagel's GLS method to control for evolutionary, and fit this inverse first order polynomial to host ranges against host phylodiversity ($y = 1.09 - 0.31\ x^{-1}$, $R^2 = 0.55$, $F = 84.3$, d.f. = 19.54, $p < 0.001$). Degrees of freedom (d.f.) were adjusted following suggestions by Paradis and Claude (2002). Contrasting the uncorrected and phylogenetically-corrected reciprocal regressions found that 18% of the variation between host range and *PD* were

explained by parasite evolutionary history.

Hypothesized Phylogenetic Relationship of Brood Parasites

My hypothesized evolutionary relationship of avian brood parasites (Appendix Table 3.1) was constructed by combining species relationships based on molecular and morphological phylogenies. Briefly, phylogenetic relationships between broad clades of brood parasite were based on Sibley and Ahlquist (1990). Cowbird species with Freeman and Zink (1995); cuckoos with Aragón *et al.* (1999), Hughes (2000) and Johnson *et al.* (2000); and indigobirds with Sorenson *et al.* (2003). Species lacking phylogenetic information were grouped as polytomies at the root of their genus. Pagel's (1992) method was used to transform internode *BL* of this phylogeny in order to minimize heteroscedasticity of phylogenetically independent contrasts (Garland 1992; Díaz-Uriarte and Garland 1996).

Clade and Continental Biases in Host Specificity

The relationship between host range and host phylodiversity does not appear to be significantly biased due to the inclusion of species from broad continental regions or species from independently evolved clades of avian brood parasitism (Appendix Table 3.2).

Appendix Table 3.2. Regression results of host range versus host phylodiversity when excluding species of avian brood parasite belonging to particular clades or continent. All results are based on inverse first order polynomial regressions (reciprocal regression).

	traditional regression			
	R^2	*F*	d.f.	*p*
excluded continents				
Africa (*N* = 37)	0.77	113.4	1,34	<0.001
Asia (*N* = 15)	0.71	136.4	1,55	<0.001
Asia–Australia (*N* = 7)	0.72	154.6	1,61	<0.001
Australia (*N* = 6)	0.74	172.1	1,62	<0.001
Europe (*N* = 2)	0.73	179.3	1,66	<0.001
North America (*N* = 2)	0.72	173.4	1,66	<0.001
South America (*N* = 3)	0.74	171.5	1,61	<0.001
excluded clades				
Anatidae (*N* = 1) (parasitic duck)	0.73	184.1	1,67	<0.001
Cuculidae (*N* = 43) (old world cuckoos)	0.69	71.6	1,31	<0.001
Fringillidae (*N* = 5) (cowbirds)	0.73	167.7	1,63	<0.001
Indicatoridae (*N* = 8) (honeyguides)	0.79	224.8	1,60	<0.001
Neomorphidae (*N* = 3) (Am. ground cuckoos)	0.73	177.6	1.65	<0.001
Passeridae (*N* = 16) (Afri. parasitic finches)	0.50	52.2	1,52	<0.001

APPENDIX REFERENCES

Aragón, S., A. P. Møller, J. J. Soler, and M. Soler. 1999. Molecular phylogeny of cuckoos supports a polyphyletic origin of brood parasitism. *Journal of Evolutionary Biology* 12:495–506.

Clarke, K. R., and R. M. Warwick. 2001. A further biodiversity index applicable to species lists: Variation in taxonomic distinctness. *Marine Ecology Progress Series* 216:265–278.

Egger, M., Davey-Smith, G., Schneider, M., and C. Minder. 1997. Bias in meta-analysis detected by a single, graphical test. *British Medical Journal* 315:629–34.

Freeman, S., and R. M. Zink. 1995. A phylogenetic study of the blackbirds based on variation in mitochondrial DNA restriction sites. *Systematic Biology* 44:409–420.

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

Garland, T., Jr., P. E. Midford, and A. R. Ives. 1999. An introduction to phylogenetically based statistical methods, with a new method for confidence intervals on ancestral states. *American Zoologist* 39:374–388.

Hughes, J. M. 2000. Monophyly and phylogeny of cuckoos (Aves, Cuculidae) inferred from osteological characters. *Zoological Journal of the Linnean Society* 130:263–307.

Johnsgard, P. A. 1997. *The avian brood parasites: deception at the nest.* Oxford University Press, New York.

Johnson, K. P., S. M. Goodman, and S. M. Lanyon. 2000. A phylogenetic study of the Malagasy *Couas* with insights into cuckoo relationships. *Molecular Phylogenetics and Evolution* 14:436–444.

Lowther, P. E. 2004a. Host list of avian brood parasites – 1 – Anatidae. Ver. 6 Aug. *http://fm1.fieldmuseum.org/aa/Files/lowther/HList.pdf*

Lowther, P. E. 2004b. Host list of avian brood parasites – 3 – Cuculiformes; Neomorphidae. Ver. 28 Oct. *http://fm1.fieldmuseum.org/aa/Files/lowther/NWcList.pdf*

Lowther, P. E. 2005a. Host list of avian brood parasites – 3 – Cuculiformes; Cuculidae. Ver. 29 Jul. *http://fm1.fieldmuseum.org/aa/Files/lowther/OWcList.pdf*

Lowther, P. E. 2005b. Host list of avian brood parasites – 4 – Piciformes; Indicatoridae. Ver. 28 Jul. *http://fm1.fieldmuseum.org/aa/Files/lowther/IList.pdf*

Lowther, P. E. 2005c. Host list of avian brood parasites – 5 – Passeriformes; Viduidae. Ver. 10 Feb. *http://fm1.fieldmuseum.org/aa/Files/lowther/VList.pdf*

Lowther, P. E. 2005d. Lists of victims and hosts of the parasitic cowbirds (*Molothrus*), ver. 16 Mar. *http://fm1.fieldmuseum.org/aa/Files/lowther/CBList.pdf*

Martins, E. P., and T. Garland, Jr. 1991. Phylogenetic analyses of the correlated evolution of continuous characters: a simulation study. *Evolution* 45:534–557.

Paradis, E., and J. Claude. 2002. Analysis of comparative data using generalized estimating equations. *Journal of Theoretical Biology* 218:175–185.

Poulin, R., and D. Mouillot. 2003. Parasite specialization from a phylogenetic perspective: a new index of host specificity. *Parasitology* 126:473–483.

Pagel, M. D. 1992. A method for the analysis of comparative data. *Journal of Theoretical Biology* 156:431–442.

Purvis, A., and T. Garland, Jr. 1993. Polytomies in comparative analyses of continuous characters. *Systematic Biology* 42:569–575.

Sibley, C. G., and J. E. Ahlquist. 1990. *Phylogeny and classification of birds: a study in molecular evolution.* Yale University Press, New Haven.

Sorenson, M. D., and R. B. Payne. 2001. A single ancient origin of brood parasitism in African finches: implication for host-parasite coevolution. *Evolution* 55:2550–2567.

CHAPTER FOUR

HOST SPECIFICITY SCALES WITH GEOGRAPHIC RANGE IN OBLIGATE AVIAN BROOD PARASITES

Abstract

Bird species that are obligate brood parasites have either few or several to many avian host species. Success in parasitizing the brood-rearing efforts of individuals of other species should determine host species range. How well the obligate brood parasites are adapted to their host species is thought to relate to the evolutionary time of association, the costs of parasitism to the hosts, and fitness tradeoffs to the parasite in use of different host species. Less well studied are the macroecological determinants of host specificity among brood parasites particularly in a phylogenetic and biogeographical context. Here, I show that host specificity scales with geographic range among 76 of the 95 known obligate avian brood parasites. The strength of this relationship varies significantly among independent evolutionary clades of brood parasitism, some of which include only species endemic to particular continents. Generally, obligate brood parasites with small geographic ranges tend to have few host species, whereas those with larger breeding ranges can have either few or many host species. These results underscore the phylogeographic determinants and constraints of host use in the evolution of obligate brood parasitism in birds.

Introduction

MacArthur (1972) theorized that habitat specialization may be key determinant of a species geographic range, and predicted that specialists should occupy smaller ranges

than generalists. A broad range of studies have tested this hypothesis and generally find that habitat specialization negatively correlates with geographic range size (Glazier 1980; Brown 1984; Eeley and Foley 1999; Lehman 2004). This negative correlation may be due to specialization to one or few ecological conditions may limit ability in alternate habitats (Brown 1984; Fry 1996). Given that habitats are often unevenly distributed and range in abundance geographically, specialization to any one habitat could result in a narrow range in a species' geographic distribution (Hanski and Gyllenberg 1997). Alternatively, species that are broadly distributed may have multiple resources simply due the greater ecological opportunity of these resources.

Price (1980) further argued that such processes should extend trophically to parasites because host species are habitats/resources—which are also unevenly distributed and vary in abundance geographically (e.g., Watters 1992; Lively and Jokela 1996). In fact, there is already some evidence that the geographic distributions of parasites follow host geographic distributions (see Loder *et al.* 1998; Brändle *et al.* 2002). For instance, Krasnov *et al.* (2005) found that there is a high likelihood that the geographic distributions of generalist parasites of rodents likely reflect the combined distribution their host species. The aim of this study is to provide macroecological and phylogenetic context to the broad variation in host specificity among obligate avian brood parasites (Davies 2000). I primarily focus on the link between host specificity and the geographic range size of brood parasites—a link that has only been evaluated or discussed to a lesser extent in other parasite-host systems (see Loder *et al.* 1998; Brändle *et al.* 2002; Krasnov *et al.* 2005). Finding a correlation between host specificity and geographic range of parasites may thus also reflect classical island biogeographic and spatial heterogeneity theory: species richness and diversity are predicted to scale with geographic range sizes (Preston 1962; Simpson 1964; MacArthur and Wilson 1963). Should host species diversity and richness be predicted

by geographical ranges of parasites, then these early theories may have even greater generality than was initially envisioned (Price 1980).

Avian brood parasites are well suited to test correlations between host specificity and geographic range because of the availability if comprehensive data on the ecology of host use among brood parasites and their hosts (see Johnsgard 1997), and because the initial formulation of MacArthur's (1972) hypothesis was influenced by observed patterns in avian richness. In addition, I also evaluate the potential for phylogenetic constraints on host specificity, and whether such constraints are broadly mediated by continental boundaries (Wiens and Donoghue 2004). Obligate avian brood parasitism has independently evolved seven times (Sorenson and Payne 2001), and some of these independent lineages are endemic to specific continents (see Table 4.1). These independent origins within separate continents can thus provide information about how unique geographic characteristics—such as continental size or perhaps diversity of bird species within continents—could have influenced patterns in host use. For instance, there is a disparity in avian diversity and richness between and within continents (Soler and Møller 1996; Avise 2000; Jetz and Rahbek 2001; Jetz and Rahbek 2002; Graves and Rahbek 2005), and this disparity may set the upper limit of potential hosts for brood parasites. This prediction is especially relevant if there has been a long history of parasitizing specific taxonomic groups of avian hosts that are contained within historic biogeographical areas (Davies *et al.* 1996; Rothstein *et al.* 2001). Although it is difficult to causally link geographic distributions and adaptations for ability to use multiple hosts, a phylogenetic analysis combined with an examination of continental trends provides a novel test of true associations between the ecological traits and evolutionary history of avian brood parasites.

Methods

Estimates of Host Specificity and Geographic Range Size

To examine the relationship between host specificity and geographic range, I collated information on total number of known host species (host range) and the geographic distributions of 70 species of brood parasites (Table 4.1a and 4.1b); as reported in Johnsgard (1997) and annotated lists of host species by Lowther (2004a, 2004b, 2005a, 2005b, 2005c, 2005d). These species belong to all known clades of avian brood parasitism (e.g., cowbirds, honeyguides, Old and New World cuckoos, parasitic finches, and a parasitic fowl), and represent all species of obligate brood parasite with detailed information on field patterns in host use. In addition to these 70 species, data on host use and geographic distributions were also available for six subspecies (see Table 4.1a and 4.1b), and I include these data in my analyses assuming that the ecological processes influencing patterns of host use and geographic range are the same for subspecies as in species.

I also use the average phylogenetic relatedness of host species (hereafter host phylodiversity or *PD*) as an alternate estimate of host specificity. This estimate is important because it gauges the phylogenetic diversity of host species, which may in turn (much like host richness) be influenced by broad continent constraints on the phylogenetic diversity of birds. Following Clarke and Warwick (2001) and Poulin and Mouillot (2003), I define host phylodiversity as $PD = C\sum\sum_{i<j} BL_{ij}$, where $C = (2N(N-1))^{-1}$, N is the number of host species in the phylogeny, and *BL* is the branch-length from i^{th} host tip to j^{th} host tip on the phylogeny.

The phylogenetic tree used to estimate *PD* was based on Sibley and Ahlquist's (1990) broad classification, and had internode *BL* scaled following their DNA-DNA hybridization method, in addition to an arbitrary scaling of internodes following Pagel (1992). To standardize the scale of *PD* across brood parasites, I used all host species

Table 4.1a. The hypothesized phylogenetic relationship (based on combined molecular and morphological data; see *Methods*), host range (total number of known avian host species), geographic location (continent), and geographic distribution (km ^{2}) of avian brood parasites used in analyses. Species marked with 'J' are Japanese populations.

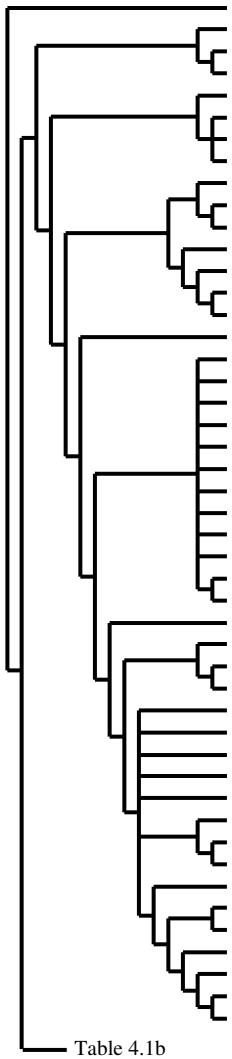
phylogeny	species	host range	continent	geographic range size
	*Heteronetta atricapilla* (black-headed duck)	12	South America	1398000
	Tapera naevia (striped cuckoo)	20	South America	13140000
	Dromococcyx pavoninus (pavonine cuckoo)	5	South America	8275000
	Dromococcyx phasianellus (pheasant cuckoo)	3	South America	9807000
	Scythrops novaehollandiae (channel-billed cuckoo)	9	Asia–Australia	1692000
	Eudynamys cyanocephala (Australian koel)	21	Australia	1712000
	Eudynamys scolopacea (Asian koel)	14	Asia–Australia	4014000
	Eudynamys taitensis (long-tailed koel)	8	Australia	255000
	Clamator coromandus (chestnut-winged cuckoo)	23	Asia	2310000
	Clamator glandarius (great spotted cuckoo)	19	Africa	12500000
	Clamator glandarius (European population)	6	Europe	428000
	Oxylophus levaillanti (Levaillant's cuckoo)	9	Africa	8463000
	Oxylophus jacobinus jacobinus (pied cuckoo)	10	Asia	1734000
	Oxylophus jacobinus pica	36	Africa	13821000
	Oxylophus jacobinus serratus	16	Africa	1797000
	Pachycoccyx audeberti (thick-billed cuckoo)	3	Africa	3686000
	Chrysococcyx basalis (Horsfield's bronze cuckoo)	97	Australia	4504000
	Chrysococcyx caprius (dideric cuckoo)	44	Africa	17963000
	Chrysococcyx crassirostris (pied bronze cuckoo)	1	Asia–Australia	73000
	Chrysococcyx cupreus (African emerald cuckoo)	18	Africa	11103000
	Chrysococcyx klaas (Klaas's cuckoo)	39	Africa	14120000
	Chrysococcyx lucidus (shining bronze-cuckoo)	84	Australia	1241000
	Chrysococcyx maculatus (Asian emerald cuckoo)	12	Asia	1299000
	Chrysococcyx osculans (black-eared cuckoo)	23	Australia	2126000
	Chrysococcyx rufomerus (green-cheeked bronzed cuckoo)	1	Asia	11000
	Chrysococcyx xanthorhynchus (violet cuckoo)	10	Asia	1382000
	Chrysococcyx minutillus (little bronze cuckoo)	14	Asia–Australia	1302000
	Chrysococcyx minutillus russatus	23	Asia–Australia	673000
	Surniculus lugubris (Drongo cuckoo)	20	Asia	2902000
	Cacomantis flabelliformis (fan-tailed cuckoo)	81	Asia–Australia	1825000
	Cacomantis merulinus (plaintive cuckoo)	14	Asia	2448000
	Cacomantis variolosus (brush cuckoo)	58	Asia–Australia	2200000
	Cuculus clamosus (black cuckoo)	19	Africa	9443000
	Cuculus gularis (African cuckoo)	2	Africa	12931000
	Cuculus micropterus (Indian cuckoo)	11	Asia	5223000
	Cuculus rochii (Madagascar cuckoo)	4	Africa	543000
	Cuculus sparveroides (large hawk cuckoo)	28	Asia	2310000
	Cuculus saturatus (oriental cuckoo)	16	Asia	16740000
	Cuculus fugax hyperythrus (Hodgson's hawk cuckoo, J)	10	Asia	4787000
	Cuculus fugax nisicolor (Indian population)	20	Asia	4787000
	Cuculus solitarius (red-chested cuckoo)	22	Africa	12870000
	Cuculus poliocephalus (lesser cuckoo)	22	Asia	2184000
	Cuculus poliocephalus (J)	8	Asia	376000
	Cuculus pallidus (pallid cuckoo)	111	Australia	7618000
	Cuculus canorus canorus (common cuckoo)	100	Europe	25046000
	Cuculus canorus bangsi	4	Africa	201000
	Cuculus canorus telephonus	28	Asia	4825000

Table 4.1b

Table 4.1b. Continuation of the hypothesized phylogenetic relationship of avian brood parasites (based on combined molecular and morphological data; see *Methods*). See Table 4.1a for further details.

phylogeny	Species	host range	continent	geographic range size
Table 4.1a	*Prodotiscus insignis* (Cassin's honeyguide)	4	Africa	2265000
	Prodotiscus regulus (Wahlberg's honeyguide)	5	Africa	3278000
	Prodotiscus zambesiae (green-backed honeyguide)	9	Africa	1912000
	Indicator conirostris (thick-billed honeyguide)	4	Africa	1729000
	Indicator indicator (greater honeyguide)	49	Africa	10374000
	Indicator meliphilus (pallid honeyguide)	3	Africa	695000
	Indicator minor (lesser honeyguide)	30	Africa	10614000
	Indicator variegatus (scaly-throated honeyguide)	12	Africa	1856000
	Molothrus rufoaxillaris (screaming cowbird)	3	South America	2121000
	Scaphidura oryzivora (giant cowbird)	8	South America	9370000
	Molothrus aeneus (bronzed cowbird)	77	North America	704000
	Molothrus ater (brown-headed cowbird)	220	North America	7478000
	Molothrus bonariensis (shiny cowbird)	201	South America	12965000
	Anomalospiza imberbis (parasitic weaver)	10	Africa	5149000
	Vidua macroura (pin-tailed whydah)	13	Africa	12833000
	Vidua togoensis (togo paradise whydah)	1	Africa	442000
	Vidau paradisaea (eastern paradise whydah)	1	Africa	6943000
	Vidau obtusa (broad-tailed paradise whydah)	1	Africa	3832000
	Vidua orientalis (northern paradise whydah)	1	Africa	3290000
	Vidau interjecta (long-tailed paradise whydah)	1	Africa	1537000
	Vidua hypocherina (steel-blue whydah)	3	Africa	1344000
	Vidua regia (queen whydah)	3	Africa	1816000
	Vidua fischeri (straw-tailed whydah)	1	Africa	2146000
	Vidua chalybeata (village indigobird)	1	Africa	9443000
	Vidua funerea (variable indigobird)	1	Africa	1936000
	Vidua purpurascens (dusky indigobird)	1	Africa	2665000
	Vidua wilsoni (pale-winged indigobird)	2	Africa	2028000
	Vidua larvaticola (baka indigobird)	1	Africa	101000
	Vidua raricola (jambandu indigobird)	1	Africa	95000

of brood parasites and compiled a single genus-level phylogeny to estimate phylodiversity for each parasite species; this phylogeny contained 498 genera belonging to 81 families as classified by Sibley and Ahlquist (1990). However, estimating *PD* from a genus-tree gives equal scoring to host species belonging to the same genus, in such cases I gave these hosts have half the *BL* score for those closely related host species. The node depth of this phylogenetic tree was 21 (Table 4.1a and 4.1b), such that phylodiversity could potentially range from 21 (the theoretical maximum of *PD*) to zero (absolute specialists with a single host species; Poulin *et al.* 2004). Unfortunately, the annotated lists of host species (Lowther 2004a, 2004b, 2005a, 2005b, 2005c, 2005d) did not distinguish the identity of host species for subspecies of brood parasites—thus I could not estimate *PD* for subspecies. In total, *PD* was estimated for 70 species of brood parasite.

For estimates of geographic range size, I analyzed Johnsgard's (1997) breeding distribution maps using the pixel statistic software tpsDig (Rohlf 2001). I restricted data on geographic distributions to breeding ranges, rather than to total geographic distributions, because breeding ranges represent the areas where parasites interact and parasitize hosts. For each map, I determined the number of pixels within the area(s) representing the breeding distribution, and the number within the continent(s) containing this distribution. The ratio between these two area counts was then scaled to known continent/land mass sizes. Although coarse, these distribution estimates show large interspecific differences in size between brood parasites and thus are informative for broad cross-species analyses (Gaston and Blackburn 2000).

Conventional and Phylogenetically Controlled Analyses

I approached my comparative study in two ways. I begin with a conventional analysis that assumes host range (or host phylodiversity) and geographic range size are not

evolutionary correlated, and that all species of brood parasites have unrelated evolutionary histories. To determine whether geographic range can predict host specificity I first use simple linear regression, followed by ANCOVA testing for continental patterns in host ranges (or host phylodiversity), and correcting for geographic range size (non-significant interactions were excluded from final models to increase power of tests; Zar 1985). Differences between continents (i.e. area size or avian diversity; Howard and Moore 1991) may have defined the limitations of geographic range sizes and specificity of parasites. It was therefore important to test for such effects, since most brood parasites in these analyses are endemic to particular continents. Host ranges, host phylodiversity and geographic range sizes were log (x + 1) transformed prior to analyses to standardize variances (Harvey 1982).

My next approach assumes that species are part of a hierarchical, branching phylogeny, and that host specificity and geographic range may have a correlated evolutionary history. This analysis requires a phylogenetic hypothesis on the evolutionary history of species of avian brood parasites. My hypothesized evolutionary relationship among brood parasites (Tables 4.1a and 4.1b) was constructed by combining molecular and morphological phylogenies. Combining phylogenetic sources maximizes the number of species used in analyses, while also resolving some soft polytomies (which can affect statistical tests; see below). Briefly, broad groupings were based on Sibley and Ahlquist (1990). Cowbirds were based on Freeman and Zink's (1995); cuckoos on Aragón *et al.* (1999), Hughes (2000) and Johnson *et al.* (2000); and indigobirds and viduidines on Sorenson *et al.* (2003). Species of brood parasite lacking phylogenetic information were grouped as polytomies at the root of their genus.

The disadvantage of mixing phylogenetic sources is the lack of a common currency to define internode branch lengths of the phylogeny. Four transformations

were used to arbitrarily set these branch-lengths. These branch-length transformations generally have an effect of removing heteroscedasticity that can significantly decrease the power of phylogenetically independent contrasts (details outlined below; Garland 1992; Díaz-Uriarte and Garland 1996). Contrasting results across multiple transformation methods further controls for Type I errors due to particular biases imposed by any one transformation (Díaz-Uriarte and Garland 1996).

Specifically, I used Pagel's (1992) method (species lined up with internode branch segments set to one; see topology in Table 4.1a and 4.1b), Grafen's (1989) method (species lined up with internode length proportional to the number of species derived from it), and Nee's method (Purvis 1995; species lined up with internode length set equal to log number of species derived from it). These three methods assume a gradual Brownian motion model of evolution (as assumed in phylogenetically independent contrasts; Felsenstein 1985), whereas my final method follows the speciation model for evolutionary change, with all internode lengths set constantly to one (Martins and Garland 1991).

Thus in the second set of analyses, I controlled for evolutionary non-independence between brood parasites with phylogenetically-independent contrasts (PIC; Felsenstein 1985). Independent contrasts were calculated and analyzed with the PDTREE module of the Phenotypic Diversity Analysis Programs (PDAP; Garland *et al.* 1999; Garland and Ives 2000). As in the first set of analyses, I tested whether geographic range sizes could predict host specificity—although here the regression using PIC's were forced through the origin (following Garland *et al.* 1992). I also accounted for effects of polytomies in our phylogeny by conservatively bounding the degrees of freedom of PIC regressions: as the total number of tree nodes minus one (following Purvis and Garland 1993).

The conventional ANCOVA testing for continental effects did not control for

phylogenetic differences between species. Testing for mean differences between categorical variables (e.g., continent) while correcting for shared ancestry is more difficult than analyses of simple evolutionary correlations (Felsenstein 1985). The difficulty lies in obtaining the correct degrees of freedom for ANCOVA that account for the phylogenetic non-independence of species (Garland *et al.* 1993). I thus use a method developed by Garland *et al.* (1993) that uses Monte Carlo methods to simulate traits evolving along the brood parasite phylogeny to determine phylogenetically correct *F* statistics for my conventional ANCOVA. These phylogenetically corrected *F* values become the new critical values of the tests (details in Garland *et al.* 1993).

Because this method simulates the evolution of traits to obtain statistics, and that the actual way species characteristics evolve is unknown, I report results using a range of simple and complex evolutionary models. A similar result using different models would imply that a particular model of evolution did not significantly bias results. These four models simulate the evolution of host range and geographic distribution along my hypothesized phylogenetic tree (Tables 4.1a and 4.1b) under different evolutionary parameters (detailed below). All simulations outlined below were completed using the PDSIMUL module of PDAP. Specifically, for all models, I used the actual variances of my data to set the expected variances of the tips in my phylogeny (PDSIMUL default), and removed the correlation between host ranges and geographic distributions (since I aimed to test for effects of geographic ranges). I also used the estimated root node values obtained from an independent-contrast analysis of log-transformed data to set the initial and final means of the models. These analyses were based on PIC estimated from a phylogeny with Pagel's transformation of the internode-branch lengths.

The four evolutionary models used in the phylogenetically corrected ANCOVA were as follows. The first two were gradual and speciational models of

Brownian motion evolution (e.g. without limits, correlations or strengths in the evolution of host ranges and geographic distributions). The gradual model assumed character change of traits was greater for longer branches, and the speciational model assumed all changes in traits occur in association with speciation events. The third model was similar to the gradual Brownian motion evolution, but had limitations to the evolution of traits (e.g., host range limits: 1–227 host species, host phylodiversity limits: 0–21 and geographic range limits: 104 km^2–2.6 x 107 km^2; each being the lower–upper ranges observed across species of brood parasites; see Tables 4.1a and 4.1b). This model had correlated evolutionary changes between traits based on an independent contrast analysis of the raw data. The final simulation is based on Ornstein-Uhlenbeck evolution (OU model) which builds on the previous model (e.g., trait limits and correlated evolutionary changes). The OU model simulates natural selection acting as a 'spring', stabilizing the motion of a population wandering back and fourth on an adaptive peak (optima), under the influence of genetic drift (see CHAPTER ONE). Adaptive optima in this model were set equal to final means (estimated root node values). All simulation data were analyzed with the PDANOVA module of PDAP, and for further details of these models see Garland *et al.* (1993). I use the 95% quantile of *F* values from 10,000 simulation runs as the phylogenetically corrected critical values. PDAP does not calculate or provide estimates for least square means (LS) and 95% confidence intervals (CI) from the phylogenetically corrected ANCOVA. LS means and 95% CI were thus estimated with the above simulation data from PDAP, and should be considered as conservative estimates of these statistical metrics.

Results

Species of brood parasites with small distributions were more likely to have narrow host ranges than brood parasites with larger geographic range sizes (Figure 4.1a; R^2 = 0.18, $F_{[1,75]} = 16.3$, $p < 0.001$). Much of the variation found between geographic range size and host specificity (Figure 4.1a) is associated with biases imposed by the shared ancestry of brood parasites, because correcting for evolutionary history strengthened this positive correlation (Figure 4.1b; using Pagel's (1992) tree-branch transformation: PIC $R^2 = 0.27$, $F_{[1,54]} = 27.5$, $p < 0.001$; Grafen's (1989): PIC $R^2 = 0.24$, $F_{[1,54]} = 22.7$, $p < 0.001$; Nee's (Purvis 1995): PIC $R^2 = 0.23$, $F_{[1,54]} = 21.7$, $p < 0.001$; constant speciation: PIC $R^2 = 0.23$, $F_{[1,54]} = 21.5$, $p < 0.001$). In general, it appears that size of geographic ranges and host specificity of avian brood parasites do share a correlated evolutionary history, but more generally these traits are ecologically functionally related.

Avian brood parasites with broad geographic ranges also have a greater number of phylogenetically distant host species than parasites with narrow distributions (Figure 4.1b; $R^2 = 0.12$, $F_{[1,69]} = 9.2$, $p = 0.003$). However, these data appear particularly effected by parasitic finches, which represent the majority of species clustered near the narrow spectrum of host phylodiversity (13 of 19 species; Figure 4.1c). Accounting for the evolutionary history of brood parasites had an effect of removing much of the bias by parasitic finches, but resulted in the phylogenetic correlation explaining less variation between host phylodiversity and geographic distribution (Figure 4.1d; Pagel's (1992) branch transformation: PIC $R^2 = 0.13$, $F_{[1,48]} = 10.0$, $p = 0.003$; Grafen's (1989): PIC $R^2 = 0.03$, $F_{[1,48]} = 1.8$, $p = 0.186$; Nee's (Purvis 1995): PIC $R^2 = 0.07$, $F_{[1,48]} = 5.1$, $p = 0.029$; constant speciation: PIC R^2 = 0.16, $F_{[1,48]} = 13.0$, $p < 0.001$).

Correlations between host specificity and geographic range also varied

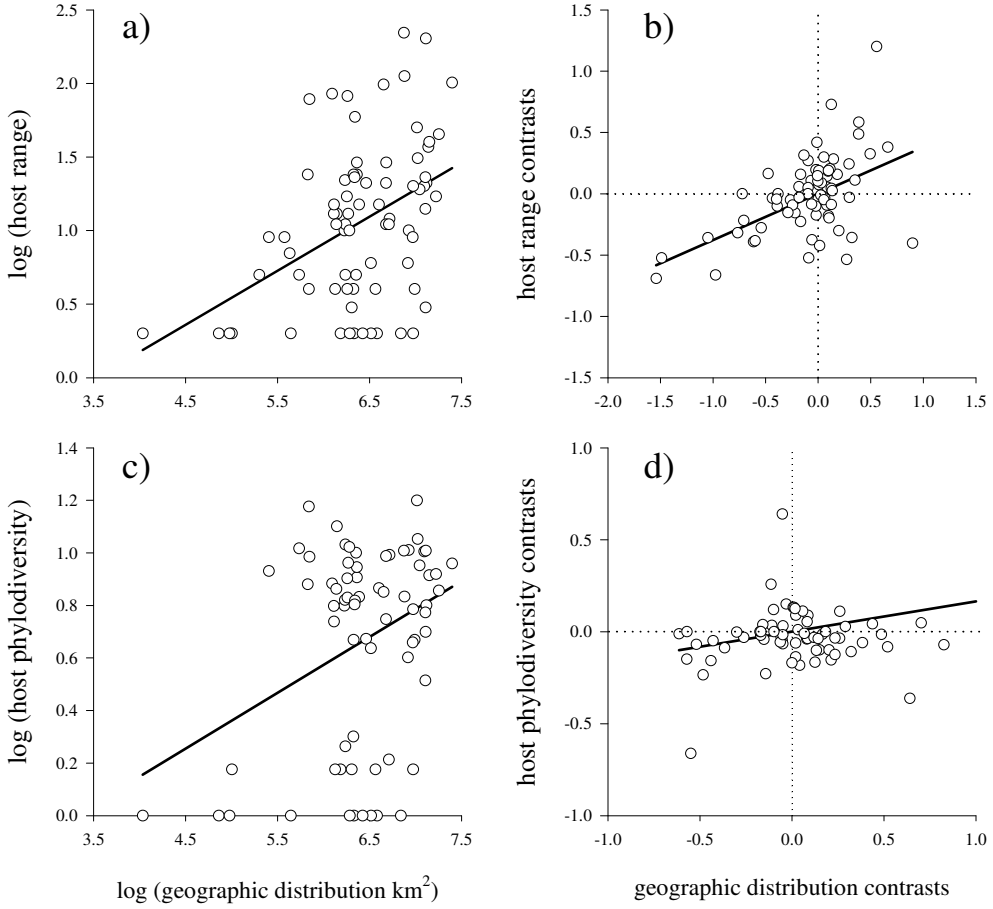


Figure 4.1. Pairwise relationship between host specificity and geographic range size of avian brood parasites. **(a)** Analysis on raw data between geographic range and host range (N = 76), **(b)** corresponding relationship after accounting for evolutionary non-independence using Felsenstein's (1985) independent contrasts (N = 75), **(c)** analysis on raw data between geographic range and host phylodiversity (N = 70), and **(d)** after accounting for evolutionary non-independence using independent contrasts (N = 69). Phylogenetically independent contrasts of (b) and (d) were based on a phylogenetic tree (see Tables 4.1a and 4.1b) transformed using Pagel's (1992) method for adjusting internode branch-lengths.

between families of brood parasites (Table 4.2). Specifically for old world cuckoos (Cuculidae), geographic distributions were more useful in predicting host specificity (both host range and phylodiversity) than in other families, even after correcting for shared ancestry of parasites. Honeyguides (Indicatoridae) also had a strong relationship between geographic ranges and host ranges (Table 4.2), but when factoring out phylogenetic history, this relationship did not hold. Host phylodiversity also scaled strongly with geographic range size in American ground cuckoos, where cuckoos with large geographic ranges had a much broader range of phylogenetically distant hosts than did narrowly distributed species of brood parasites (Table 4.2). However, the host ranges of African parasitic finches (Passeridae) do not appear to scale geographically. The lack of correlation here may be due to these finches having very narrow host ranges (ca. 1–3 host species).

Continents may limit both host specificity and geographic distributions of species of brood parasites, as they vary in geographic size and in richness of potential avian hosts (Howard and Moore 1991). Species of brood parasites from different continents did have distinct host ranges, after correcting for geographic distributions (Table 4.3; Figure 4.2a). Species of brood parasites from Australia and North America generally had more host species than parasites inhabiting other continents (e.g., South America or Africa; Figure 4.2a). However, phylogenetic corrections decreased the observed variation of host specificity across continents, such that apparent continental differences in host use (as seen in Figure 4.2a) were not as prominent (Figure 4.2b). Controlling for phylogeny in this analysis also may have accounted for broad correlations between speciation events and dispersal across large geographic scales (a pattern common with birds; Chesser and Zink 1994).

Species of brood parasites also differed continentally in the range of

Table 4.2. Correlations between host specificity and geographic range size between independently evolved clades of brood parasites, using conventional regressions and phylogenetically independent contrasts (Felsenstein 1985). Degrees of freedom (d.f.) of independent contrasts were conservatively adjusted to account for polytomies (see *Methods*). Only one species represented the Anatidae family and were not included in analyses. In all analyses, regressions were positively correlated. Significant correlations are indicated in bold, and abbreviated as follows: (*) $p < 0.05$; (**) $p < 0.005$; (***) $p < 0.0005$.

	host range					host phylodiversity				
	conventional		independent contrasts			conventional		independent contrasts		
	R^2	F [B]	R^2	F	d.f.	R^2	F [B]	R^2	F	d.f.
Cuculidae (N [A]) (old world cuckoos)	**0.31**	**18.4*** **	**0.32**	**19.6** **	**26**	**0.34**	**17.6*** **	**0.41**	**24.4*** **	**20**
Fringillidae ($N = 5$) (cowbirds)	0.05	0.2	0.01	0.1	3	0.03	0.1	0.01	0.1	3
Indicatoridae ($N = 8$) (honeyguides)	**0.74**	**17.2***	0.41	4.2	2	0.04	0.2	0.03	0.2	2
Neomorphidae ($N = 3$) (new world cuckoos)	0.67	2.1	0.37	0.6	1	**0.99**	**918.6***	**0.99**	**530.4***	**1**
Passeridae ($N = 16$) (African finches)	0.16	2.6	0.15	2.5	13	0.03	0.4	0.01	0.1	13

superscripts [A]: $N = 43$ for host range and $N = 37$ phylodiversity, [B]: d.f. = $N - 1$.

Table 4.3. ANCOVA testing for continental differences in host ranges among 76 avian brood parasites, and correcting for geographic range sizes of brood parasites. Critical values for *F* statistics and significance levels (*p*) are presented for conventional ANCOVA and four different phylogenetically corrected ANCOVA (each based on different models of evolution). These corrected ANCOVA re-estimate critical values based the phylogenetic relationships of brood parasites (Tables 4.1a and 4.1b 1). New significance levels are recomputed by comparing corrected critical values to *F* values from the conventional ANCOVA.

					phylogenetically corrected ANCOVA							
		conventional ANCOVA			Brownian motion gradual		Brownian motion speciational		gradual limits		Ornstein-Uhlenbeck	
source of variation	d.f.	*F*	critical value	*p*	critical value	*p*	critical value	*p*	critical value	*p*	critical value	*p*
continent	6	11.1	2.23	<0.001	8.53	0.014	9.57	0.026	12.20	0.074	10.21	0.035
geographic range size	1	43.6	3.99	<0.001	19.61	0.003	24.91	0.008	46.31	0.057	45.40	0.054
total explained	7	13.8	2.14	<0.001	7.15	<0.001	6.52	<0.001	6.59	<0.001	5.29	<0.001
error	68											
total	75											
interaction *	6	0.4	2.25	0.85	3.88	0.965	3.46	0.953	4.18	0.971	2.83	0.925

* Interaction effects between continent and geographic range size are presented separately, as these were all non-significant, and thus were excluded from ANCOVA models.

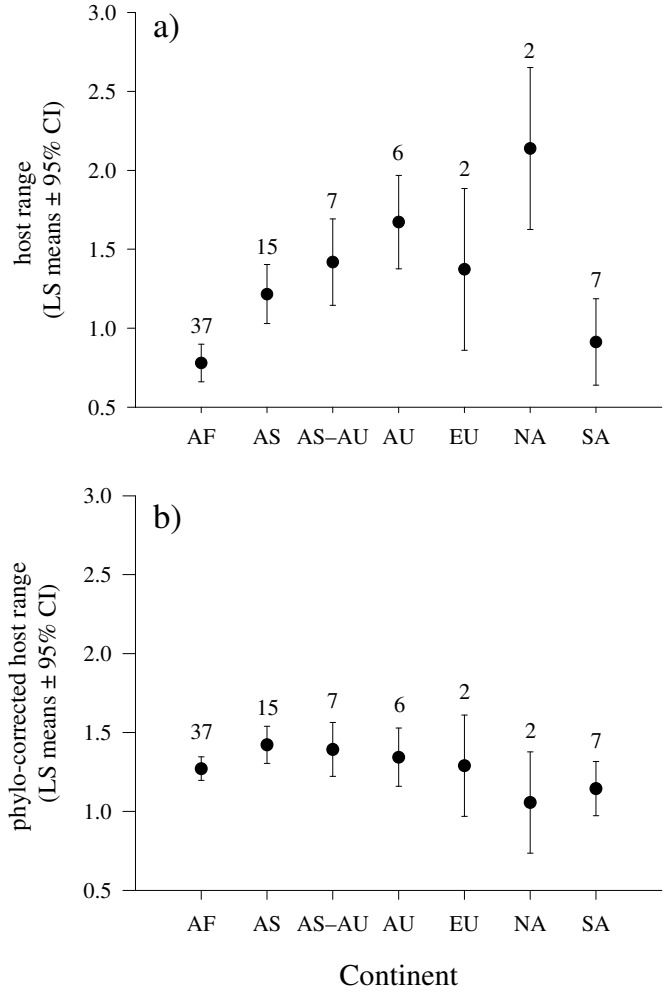


Figure 4.2. Continental differences in host ranges of avian brood parasites: **(a)** results from a conventional, non-phylogenetically corrected, ANCOVA treating geographic range size as a covariate (see Table 4.3; least square means (LS) and 95% confidence intervals (CI) were calculated following Zar 1985), **(b)** phylogenetically corrected ANCOVA treating geographic range size of brood parasites as a covariate (results based on Ornstein-Uhlenbeck model of evolution; see Table 4.3). Continent abbreviations are as follows: Africa (AF), Asia (AS), islands between Asia–Australia (AS–AU), Australia (AU), Europe (EU), North America (NA) and South America (SA). The numbers of species for each continental region are indicated above 95% CI.

phylogenetically distinct host species (ANCOVA; covariate/geographic range: $F_{[1,69]}$ = 14.4, $p < 0.001$, continent: $F_{[6, 69]} = 2.8$, $p = 0.019$). However, only North American and African brood parasite species showed marked differences in phylodiversity, with the former group of species having the greatest and the latter group having the lowest host phylodiversity score. All other continents had brood parasites that were intermediate in the ranges of phylogenetically distinct host species. Analysis with phylogenetically corrected ANCOVA found that these continental differences did not remain after correction for avian broad parasite evolutionary history under all evolutionary models). I thus only report OU evolutionary model results: covariate/geographic range: conventional ANCOVA critical value (CV) = 4.0, phylogenetically corrected ANCOVA CV = 5.5, $p < 0.001$; continent: conventional CV = 2.2, phylogenetically corrected CV = 10.7, $p = 0.99$).

Discussion

Where avian hosts have been comprehensively studied and specific data are available, nearly 83% of 76 avian brood parasite species have more than one host species (Tables 4.1a and 4.1b). The history of host associations and specialization across clades of brood parasites might be entirely ascribable to ecological agents of selections (e.g., host abundance or resistance, predation, competition, etc.). However, my results imply that this is not entirely the case. I found a positive correlation with host specificity (in terms of both number and phylogenetic diversity of hosts) and the size of the geographic distribution of brood parasites. I also found that these traits share a correlated evolutionary history, which is in part not independent of the continent where such evolutionary and ecological processes occur. My results suggest either that a broad use of host species results in greater geographic distributions of species of brood parasites, or a large geographical distribution range can assist exposure to an

increased number and diversity of potential host species.

What is also clear from my results is that many brood parasites with few host species can also have broad distributions. This can occur when brood parasites specialize on few widely distributed host species (e.g., African cuckoo, *Cuculus gularis* with its broadly distributed hosts: the yellow-billed shrike, *Corvinella corvina* and the fork-tailed drongo, *Dicrurus adsimilis*), or if there is a low richness of sympatric avian species that make compatible hosts. These brood parasites are not limited by constraints of small landmass sizes *per se* because they are endemic to large continents (Jetz and Rahbek 2001; Graves and Rahbek 2005). This variation across specialist brood parasites can also occur if host species also have limited geographic ranges (e.g., perhaps hosts that are habitat specialists), or age of host associations (Rothstein *et al.* 2002). Many of the specialists in this study are also endemic to small islands (e.g., New Guinea/Indonesia; i.e. green-cheeked bronzed cuckoo, *Chrysococcyx rufomerus*), or even biogeographical islands (e.g., viduidine species). For species of brood parasites limited to landmasses with strong area size constraints, narrow host specificity may be a product of exposure to a limited host pool, strong inter-specific competition (Ney-Nifle *et al.* 2005), or is pleisomorphic (Rothstein *et al.* 2002).

Brown and Lomolino (1998) argue that specific details on resource use, spatial variation in environments and dispersal are key to interpreting the causes of limits to species geographic ranges. The range of resources used by avian brood parasites is determined by biotic (habitat preferences and host species use) and abiotic factors (area size, latitudinal gradients). Adult habitat preferences (e.g. breeding and feeding ecology) can also certainly constrain geographic ranges of brood parasites, preventing species from using perfectly suitable host species in alternate habitats (Johnsgaard 1999). Dispersal can also permit species of brood parasites to occupy geographic

ranges where niche requirements are not met (use of host species that cannot sustain brood parasite populations; e.g., satellite or sink hosts; Holt 1996; Pulliam 2000), or constrain use of perfectly suitable hosts (potential host species separated by geographic or ecological barriers). Alternatively, geographical range expansion of the parasite species can be, to a large extent, constrained by host dispersal (Boulinier *et al.* 2001; McCoy *et al.* 2003). In turn, avian host dispersal is tied to their degree of habitat specialization or specific resource requirements, which is known to constrain colonization and expansion of new areas (Ricklefs 2003; Ricklefs 2005).

Both host specificity and geographic distributions are labile traits, yet accounting for evolutionary history of brood parasites did indicate that these traits share a correlated evolutionary history. Geographic range of species of brood parasites can potentially reflect the evolution of host specificity, but also dispersal and intensity of coevolutionary interactions with hosts. In general, it is predicted that the dynamism in species range involves the evolution of species traits (e.g., through host selection); and although environmental (non host related) change also influences these dynamics, it is certainly not required (Kirkpatrick and Barton 1997). Coevolutionary intensity of parasites and their host(s) likely strongly influences traits of parasites that allow for narrow and broad host species range. Ultimately the issue of trade-offs is essential to explain these evolutionary processes (Brown 1984; Via and Hawthorn 2001), and the degree in specialization of brood parasites is a function of how selection favors use of particular hosts over others in different areas (Singer 1984) and how this varies both ecologically and phylogenetically.

Coevolutionary dynamics between brood parasites and hosts can lead to geographic range expansions or contractions, but also influence patterns of speciation. Prolonged interaction and use of particular host species may lead to allopatric cospeciation, where brood parasites remain associated with hosts as these hosts evolve

(Soler and Møller 1990; Marchetti *et al.* 1998). Changes in host use by brood parasites also could be mediated by shifts in host resistance, where alterations in host susceptibility may drive brood parasites to use fewer or alternate host species (Rothstein 2001). Behavioral imprinting of parasitic young on novel hosts can lead to assortative mating with other parasites reared on similar hosts, which may lead to sympatric isolation and perhaps speciation (Sorenson *et al.* 2003). Given these interactions, it is perhaps not surprising that there is also evidence that age of associations between brood parasites and their hosts can significantly influence the degree of specialization (see Rothstein *et al.* 2002), but also the differences in richness of species among clades of brood parasitism (e.g., cowbirds vs. cuckoos).

I found broad continental differences in the host ranges but not host phylogenetic diversity of brood parasites, and such continental differences appear significantly influenced by evolutionary history. Molecular evidence suggests seven independent origins of obligate avian brood parasitism (Sorenson and Payne 2001), and many of these independent clades are endemic to specific continents (e.g., viduidine finches and American ground cuckoos). Accounting for evolutionary history removed some of clades biases, but did not significantly remove host range differences among continents. Continental differences in avian richness are well known (Avise 2000), and this diversity should set the upper limit in potential host species used by brood parasites. It has yet to be determined whether the pool of potential host species used by brood parasites is particularly biased by the endemism of particular host clades. For cowbirds, there is definite bias to use passerine species of host, which are incredibly diverse in North America (Avise 2000).

My study suggests a strong relationship between host specificity and geographic distributions of avian brood parasites, and that large-scale geography as well as phylogeny can mediate host specificity of these parasites. Large geographic

distributions can expose parasites to an increasing number of potential hosts, and if such hosts are compatible (ecologically and/or phylogenetically), than perhaps the brood parasite may expand its host range and integrate these avian species as hosts. Finally, the interaction between specificity and geographic ranges may extent trophically to parasites, and thus parasites may have broader community consequences then originally envisioned (MacArthur 1972).

Acknowledgements

I thank Anurag Agrawal, Irby Lovette, Jaroslav Picman, Kailen Mooney and the Cornell Laboratory of Ornithology for helpful comments and discussion about this topic. This research was funded with NSERC and Cornell University.

REFERENCES

Aragón, S., A. P. Møller, J. J. Soler, and M. Soler. 1999. Molecular phylogeny of cuckoos supports a polyphyletic origin of brood parasitism. *Journal of Evolutionary Biology* 12:495–506.

Avise, J. C. 2000. Phylogeography: the history and formation of species. Harvard University Press, Cambridge, MA.

Boulinier, T., K. D. McCoy, and G. Sorci. 2001. Dispersal and parasitism. Pp. 169–179 in J. Clobert, E. Danchin, A. Dhondt and J. D. Nichols, eds. Dispersal. Oxford Univ. Press, Oxford, U.K.

Brändle, M., S. Öhlschläger, and R. Brandl. 2002. Range sizes in butterflies: correlation across scales. *Evolutionary Ecology Research* 4:993–1004.

Briskie , J. V., S. G. Sealy, and K. A. Hobson. 1992. Behavioral defenses against avian brood parasitism in sympatric and allopatric host population. *Evolution* 46:334–340.

Brown, J. H. 1984. On the relationship between abundance and distribution of species. *American Naturalist* 124:255–247.

Brown, J. H., and M. V. Lomolino. 1998. *Biogeography*. Sinauer Associates, Sunderland, MA.

Chesser, R. T., and R. M. Zink. 1994. Mode of speciation in birds: a test of Lynch's method. *Evolution* 48:490–497.

Clarke, K. R., and R. M. Warwick. 2001. A further biodiversity index applicable to species lists: Variation in taxonomic distinctness. *Marine Ecology Progress Series* 216:265–278.

Davies, N. B. 2000. *Cuckoos, Cowbirds and other Cheats.* T. and A. D. Poyser, London.

Davies, M., M. L. Brooke, and A. Kacelnik. 1996. Recognition errors and probability of parasitism determine whether reed warblers should accept or reject mimetic eggs. *Proceedings of the Royal Society of London* B 263:925–931.

Diaz-Uriarte, R., and T. Garland, Jr. 1996. Testing hypotheses of correlated evolution using phylogenetically independent contrasts: sensitivity to deviations from Brownian motion. *Systematic Biology* 45:27–47.

Eeley, H.A.C., and R.A. Foley. 1999. Species richness, species range size and ecological specialisation among African primates: Geographical patterns and conservation implications. *Biodiversity and Conservation* 8:1033–1056.

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

Freeman, S., and R. M. Zink. 1995. A phylogenetic study of the blackbirds based on variation in mitochondrial DNA restriction sites. *Systematic Biology* 44:409–420.

Friedmann, H. 1968. The evolutionary history of the avian genus *Chrysococcyx*. Smithsonian Inst. Press, Washington.

Fry, J. D. 1996. The evolution of host specialization: Are trade-offs overrated? *American Naturalist* 148:S84–S107.

Garland, T., Jr., A. W. Dickerman, C. M. Janis, and J. A. Jones. 1993. Phylogenetic analysis of covariance by computer simulation. *Systematic Biology* 42:265–292.

Garland, T., Jr., P. H. Harvey, and A. R. Ives. 1992. Procedures for the analysis of comparative data using phylogenetically independent contrasts. *Systematic Biology* 41:18–32.

Garland, T., Jr., P. E. Midford, and A. R. Ives. 1999. An introduction to phylogenetically based statistical methods, with a new method for confidence intervals on ancestral states. *American Zoologist* 39:374–388.

Garland, T., Jr., and A. R. Ives. 2000. Using the past to predict the present: confidence intervals for regression equations in phylogenetic comparative methods. *American Naturalist* 155:346–364.

Gaston, K. J., and T. M. Blackburn. 2000. *Pattern and process in macroecology*. Blackwell Science, Oxford.

Glazier, D. S. 1980. Ecological shifts and the evolution of geographically restricted species of North American *Peromyscus* (mice). *Journal of Biogeography* 7:63–83.

Grafen, A. 1989. The phylogenetic regression. *Philosophical Transactions of the Royal Society of London* B 326:119–157.

Graves, G. R., and C. Rahbek. 2005. Source pool geometry and the assembly of continental avifaunas. *Proceedings of the National Academy of Sciences* USA USA 102: 7871–7876.

Hanski, I., and M. Gyllenberg. 1997. Uniting two general patterns in the distribution of species. *Science* 275:397–400.

Harvey, P. H. 1982. On rethinking allometry. *Journal of Theoretical Biology* 95:37–41.

Holt, R. D. 1996. Adaptive evolution in source-sink environments: Direct and indirect effects of density-dependence on niche evolution. *Oikos* 75:182–192.

Hosoi, S. A., and S. I. Rothstein. 2000. Nest desertion and cowbird parasitism: evidence for evolved responses and evolutionary lag. *Animal Behaviour* 59:823–840.

Howard, R., and A. Moore. 1991. *A complete checklist of birds of the world.* Second Edition. Harcourt, London.

Hughes, J. M. 1996. Phylogenetic analysis of the cuckoos (Aves, Cuculidae) using behavioral and ecological characters. *Auk* 113:10–22.

Hughes, J. M. 2000. Monophyly and phylogeny of cuckoos (Aves, Cuculidae) inferred from osteological characters. *Zoological Journal of the Linnean Society* 130:263–307.

Jetz, W., and C. Rahbek. 2001. A two-dimensional geometric constraints model explains much of the species richness pattern in African birds. *Proceedings National Academy of Science* USA 98:5661–5666.

Jetz, W., and C. Rahbek. 2002. Geographic range size and determinants of avian species richness. *Science* 297:1548–1551.

Johnsgard, P. A. 1997. *The avian brood parasites: deception at the nest*. Oxford University Press, New York.

Johnson, K. P., S. M. Goodman, and S. M. Lanyon. 2000. A phylogenetic study of the Malagasy *Couas* with insights into cuckoo relationships. *Molecular Phylogenetics and Evolution* 14:436–444.

Kirkpartick, M., and N. H. Barton. 1997. Evolution of a species' range. *American Naturalist* 150:1–23.

Krasnov, B. R., R. Poulin, G. I. Shenbrot, D. Mouillot, and I.S. Khokhlova. 2005. Host specificity and geographic range in haematophagous ectoparasites. *Oikos* 108: 449–456.

Langmore, N. E., S. Hunts, and R. M. Kilner. 2003. Escalation of a coevolutionary arms race through host rejection of brood parasitic young. *Nature* 422:157–160.

Lehman, S. M. 2004. Biogeography of the primates of Guyana: effects of habitat use and diet on geographic distribution. *International Journal of Primatology* 25:1225–1242.

Lively, C. M., and J. Jokela. 1996. Clinal variation for local adaptation in a host-parasite interaction. *Proceedings of the Royal Society London* B 263:891–897.

Loder, N., K. J. Gaston, P. H. Warren, and H. R. Arnold. 1998. Body size and feeding specificity: macrolepidoptera in Britain. *Biological Journal of the Linnean Society* 63:121–139.

Lotem, A., and S. I. Rothstein. 1995. Cuckoo-host coevolution: from snapshots of an arms race to the documentation of microevolution. *Trends in Ecology and Evolution* 10:436–437.

Lowther, P. E. 2004a. Host list of avian brood parasites – 1 – Anatidae. Ver. 6 Aug. http://fm1.fieldmuseum.org/aa/Files/lowther/HList.pdf

Lowther, P. E. 2004b. Host list of avian brood parasites – 3 – Cuculiformes; Neomorphidae. Ver. 28 Oct. http://fm1.fieldmuseum.org/aa/Files/lowther/NWcList.pdf

Lowther, P. E. 2005a. Host list of avian brood parasites – 3 – Cuculiformes; Cuculidae. Ver. 29 Jul. http://fm1.fieldmuseum.org/aa/Files/lowther/OWcList.pdf

Lowther, P. E. 2005b. Host list of avian brood parasites – 4 – Piciformes; Indicatoridae. Ver. 28 Jul. http://fm1.fieldmuseum.org/aa/Files/lowther/IList.pdf

Lowther, P. E. 2005c. Host list of avian brood parasites – 5 – Passeriformes; Viduidae. Ver. 10 Feb. http://fm1.fieldmuseum.org/aa/Files/lowther/VList.pdf

Lowther, P. E. 2005d. Lists of victims and hosts of the parasitic cowbirds (Molothrus), ver. 16 Mar. http://fm1.fieldmuseum.org/aa/Files/lowther/CBList.pdf

MacArthur, R. H. 1972. *Geographical ecology*. Harper and Row, New York.

MacArthur, R. H., and E. O. Wilson. 1963. An equilibrium theory of insular zoogeography. *Evolution* 17:373–387.

Marchetti, K., H. Nakamura, and H. L. Gibbs. 1998. Host-race formation in the common cuckoo. *Science* 282:471–472.

Martins, E. P., and T. Garland, Jr. 1991. Phylogenetic analyses of the correlated evolution of continuous characters: a simulation study. *Evolution* 45:534–557.

Mayfield, H. F. 1965. The brown-headed cowbird with old and new hosts. *Living Bird* 4:12–28.

McCoy, K. D., T. Boulinier, C. Tirard, and Y. Michalakis. 2003. Host-dependent genetic structure of parasite populations: differential dispersal of seabird tick host races. *Evolution* 57:288–296.

Ney-Nifle, M., C. Bernstein, J. C. Reboreda, and A. Kacelnik. 2005. Population dynamics and avian brood parasitism: persistence and invasions in a three-species system. *Journal of Animal Ecology* 74:274–284.

Pagel, M. D. 1992. A method for the analysis of comparative data. *Journal of Theoretical Biology* 156:431–442.

Peterson, A. T., J. Soberon, and V. Sanchez-Cordero. 1999. Conservatism of ecological niches in evolutionary time. *Science* 285:1265–1267.

Poulin, R., and D. Mouillot. 2003. Parasite specialisation from a phylogenetic perspective: a new index of host specificity. *Parasitology* 126:473–483.

Preston, F. W. 1962. The canonical distribution of commonness and rarity, Part 1. *Ecology* 43:185–215.

Pulliam, H. R. 2000. On the relationship between niche and distribution. *Ecology Letters* 3:349–361.

Purvis, A. 1995. A composite estimate of primate phylogeny. *Philosophical Transactions of the Royal Society of London* B 348:405–421.

Purvis, A., and T. Garland, Jr. 1993. *Polytomies in comparative analyses of continuous characters.* Systematic Biology 42:569–575.

Ricklefs, R. E. 2003. Global diversification rates of passerine birds. *Proceedings of the Royal Society of London* B 270:2285–2291.

Ricklefs, R. E. 2005. Small clades at the periphery of passerine morphological space. *American Naturalist* 164:651–659.

Rolfh, F. J. 2001. tpsDig 1.3. SUNY, Stonybrook.

Rothstein, S. I. 1990. A model system for coevolution: avian brood parasitism. *Annual Review of Ecology and Systematics* 21:481–508.

Rothstein, S. I. 2001. Relic behaviours, coevolution and the retention versus loss of host defences after episodes of avian brood parasitism. *Animal Behaviour* 61:95–107.

Rothstein, S. I., M. A. Patten, and R. C. Fleischer. 2002. Phylogeny, specialization, and brood parasite-host coevolution: some possible pitfalls of parsimony. *Behavioral Ecology* 13:1–10.

Sibley, C. G., and J. E. Ahlquist. 1990. *Phylogeny and classification of birds: a study in molecular evolution*. Yale University Press, New Haven.

Sibley, C. G., and B. L. Monroe, Jr. 1990. *Distribution and taxonomy of birds of the world*. Yale University Press, New Haven.

Simpson, G. G. 1964. Species density of North American recent mammals. *Systematic Zoology* 13:57–73.

Singer, M. C. 1983. Determinants of multiple host use by a phytophagous insect population. *Evolution* 37:389–403.

Soler, M., and A. P. Møller. 1990. Duration of sympatry and coevolution between the great spotted cuckoo and its magpie host. *Nature* 343:748–750.

Soler, M., and A. P. Møller. 1996. A comparative analysis of the evolution of variance in appearance of eggs of European passerines in relation to brood parasitism. *Behavioral Ecology* 7:89–94.

Soler, J. J., J. G. Martinez, M. Soler, and A. P. Møller. 1999. Genetic and geographic variation in rejection behavior of cuckoo eggs by European magpie populations: an experimental test of rejecter-gene flow. *Evolution* 53:947–956.

Sorenson, M. D., and R. B. Payne. 2001. A single ancient origin of brood parasitism in African finches: implication for host-parasite coevolution. *Evolution* 55:2550–2567.

Sorenson, M.D., K.M. Sefc, and R.B. Payne. 2003. Speciation by host switch in brood parasitic indigobirds. *Nature* 424:928–931.

Strausberger, B. M., and M. V. Ashley. 1997. Community-wide patterns of parasitism of a host "generalist" brood-parasitic cowbird. *Oecologia* 112:254–262.

Via, S., and D. J. Hawthorne. 2002. The genetic architecture of ecological specialization: correlated gene effects on host use and habitat choice in pea aphids. *American Naturalist* 159:S76–S88.

Watters, G. 1992. Unionids, fishes, and the species-area curve. *Journal of Biogeography* 19:491–490.

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

Zar, J. H. 1985. *Biostatistical analysis*, Second Edition. Prentice-Hall, Englewood cliffs NJ.

CURRICULUM VITAE

Marc J. Lajeunesse

Department of Ecology and Evolutionary Biology
Cornell University, Corson Hall
Ithaca, New York 14853 USA
mjl63@cornell.edu
http://www.people.cornell.edu/pages/mjl63/

CURRENT POSITION .....

Post Doctoral Fellow at NESCent (National Evolutionary Synthesis Center), *Duke University*, Durham NC, USA.

EDUCATION .....

Ph.D. (Ecology) 2008, *Cornell University*, Ithaca NY, USA.
M.Sc. (Biology, Distinction) 2002, *Carleton University*, Ottawa ON, Canada.
B.Sc. (Biology, Honours) 2000, *Carleton University*, Ottawa ON, Canada.

PUBLICATIONS and *SEMINARS* .....

a) *Peer reviewed papers*

Agrawal, A.A., **Lajeunesse, M.J.** and Fishbein, M. (2008) Evolution of latex and its constituent defensive chemistry in milkweeds (*Asclepias*): a phylogenetic test of plant defense escalation. *Entomologia Experimentalis et Applicata* **128**, 126–138.

Lajeunesse, M.J. (2007) Ectoparasitism of damselflies by water mites in Central Florida. *Florida Entomologist* **90**, 643–649.

Beirinckx, K., Van Gossum, H., **Lajeunesse, M.J.** and Forbes, M.R. (2006) Sex biases in dispersal and philopatry: insights from a meta-analysis based on capture–mark–recapture studies of damselflies. *Oikos* **113**, 539–547.

Johnson, M.T.J., **Lajeunesse, M.J.** and Agrawal, A.A. (2006) Genotypic diversity in plant populations shapes arthropod community composition. *Ecology Letters* **9**, 24–34.

Lajeunesse, M.J., Forbes, M.R. and Smith, B.P. (2004) Species and sex biases in ectoparasitism of dragonflies by mites. *Oikos* **106**, 501–508.

Lajeunesse, M.J. and Forbes, M.R. (2003) A comparison of structural size and condition in two female morphs of the damselfly *Nehalennia irene* (Hagen) (Zygoptera : Coenagrionidae). *Odonatologica* **32**, 281–287.

Lajeunesse, M.J. and Forbes, M.R. (2003) Variable reporting and quantitative reviews: a comparison of three meta-analytical techniques. *Ecology Letters* **6**, 448–454.

a) *Peer reviewed papers (continued)*

Lajeunesse, M.J. and Forbes, M.R. (2002) Host range and local parasite adaptation. *Proceedings of the Royal Society of London, Series* B. **269**, 703–710.

b) *Book chapters in review*

Lajeunesse, M.J. Power statistics for meta-analysis: Tests for mean effects, homogeneity and moderator variables. In: Meta-analysis for ecology and evolution. Eds. J. Gurevitch and J. Koricheva. Princeton University Press.

Lajeunesse, M.J., Jennions, M.D. and Rosenberg, M.S. Phylogenetic history and meta-analysis. In: Meta-analysis for ecology and evolution. Eds. J. Gurevitch and J. Koricheva. Princeton U. Press.

Lajeunesse, M.J. and Schmid, C. Imputation and addressing gaps in the literature. In: Meta-analysis for ecology and evolution. Eds. J. Gurevitch and J. Koricheva. Princeton.

Lau, J., Rothstein, H.R., **Lajeunesse, M.J.** and Mengersen, K. Progress of meta-analysis in medical and social sciences compared to ecology. In: Meta-analysis for ecology and evolution. Eds. J. Gurevitch and J. Koricheva. Princeton U. Press.

Lau, J., Mengersen, K. and **Lajeunesse, M.J.** Combination of meta-analysis and other quantitative and qualitative methods of research synthesis. In: Meta-analysis for ecology and evolution. Eds. J. Gurevitch and J. Koricheva. Princeton U. Press.

Curtis, P.S., Mengersen, K., **Lajeunesse, M.J.**, Rothstein, H.R. and Stewart, G.B. Quality of data, extracting data from papers, managing databases. In: Meta-analysis for ecology and evolution. Eds. J. Gurevitch and J. Koricheva. Princeton U. Press.

Gurevitch, J., Myers, R.A, Mengersen, K., Rosenberg, M.S., **Lajeunesse, M.J.** and Rothstein, H.R. Effect sizes: Issues and Challenges. In: Meta-analysis for ecology and evolution. Eds. J. Gurevitch and J. Koricheva. Princeton U.

Lortie, C.J., Lau, J. and **Lajeunesse, M.J.** Graphical presentation of meta-analytical results. In: Meta-analysis for ecology and evolution. Eds. J. Gurevitch and J. Koricheva. Princeton U. Press.

d) *Invited departmental seminars and symposia talks*

Lajeunesse, M.J. (Aug. 2008) *International Society for Behavioral Ecology* 2008 (Ithaca, NY). Congress symposia: "Advances in statistical philosophy and experimental design in behavioral ecology" organized by László Z. Garamszegi (University of Antwerp) and Shinichi Nakagawa (University of Waikato). "Meta-analysis and the comparative phylogenetic method"

Lajeunesse, M.J. (Jan. 2006) *Laurentian University, Biology Department.* Sudbury ON, Canada. "Host range evolution in avian brood parasites".

Lajeunesse, M.J. (Aug. 2002) *Carleton University, Department of Biology.* Ottawa ON, Canada. "Local adaptation and evolutionary potential of multi-host parasites".

e) *Conference presentations*

Lajeunesse, M.J. (2008) Meta-analysis and the comparative phylogenetic method. *Tree Thinking Symposia*, 2008. Ithaca, NY, USA.

Lajeunesse, M.J. (2007) Generalists are not generalists: host relatedness constrains parasite specificity. *Evolution 2007.* Christchurch, New Zealand.

Forbes M.R., Robb, T., **Lajeunesse, M.J.** and Smith, B.P. (2006) Generalist mites in the face of selection for specificity. *XIIth International Congress of Acarology*. Amsterdam, The Netherlands.

Lajeunesse, M.J. and Forbes, M.R. (2006) Host range scales with geographic range in avian brood parasites. *ICOPA XI: 11th International Congress of Parasitology.* SECC, Glasgow, Scotland.

Lajeunesse, M.J. (2006) Specialization is not an evolutionary dead-end: a meta-phylogenetic approach. *Evolution 2006.* Stony Brook, LI, USA.

Lajeunesse, M.J. (2005) Constraints on correlated evolution of body size and specialization in avian brood parasites. *ESA Ecology Conference.* Montreal, QC, Canada.

Lajeunesse, M.J. (2005) Specialization in avian brood parasites: constraints on correlated evolutionary history of body size but not geographic range. *Laboratory of Ornithology Seminar Series.* Cornell University, Ithaca, NY, USA.

Lajeunesse, M.J. (2005) Constraints on correlated evolution of body size and specialization in avian brood parasites. *29th Annual Ecology and Ethology Colloquium* . Cornell University, Ithaca, NY, USA.

Lajeunesse, M.J. (2004) Specialization is not a cul-de-sac for phytophagous and parasitic taxa. *ESA Ecology Conference.* Portland, OR, USA.

Lajeunesse, M.J. (2004) Specialization is not a cul-de-sac for phytophagous and parasitic taxa. *28th Annual Ontario Ecology* and *Ethology Colloquium.* Missasauga, ON, Canada.

Lajeunesse, M.J. (2004) Is specialisation a cul-de-sac for herbivorous and parasitic taxa? *Gordon research conference: plant–herbivore interactions.* Ventura, CA, USA.

Lajeunesse, M.J. and Forbes, M.R. (2002) Host range and local adaptation in parasites. *Canadian Society of Zoologists Annual Meeting.* Lethbridge, Alberta, Canada.

Lajeunesse, M.J. and Forbes, M.R. (2002) Host range and local parasite adaptation. *26th Annual Ontario Ecology* and *Ethology Colloquium.* Kingston, Ontario, Canada.

f) *Invited workgroups, article and grant reviews*

Invited Workgroups.

- NCEAS (*National Center for Ecological Analysis and Synthesis*) Workgroup on "Meta-analysis in Ecology: Lessons, Challenges and Future" lead by J. Koricheva (Royal Holloway University of London) and J. Gurevitch (State University of New York at Stony Brook), Santa Barbara CA, 2006–2008. our website: http://www.nceas.ucsb.edu/meta

f) *Invited workgroups, article and grant reviews – continued*

- CEE (*Collaboration for Environmental Evidence*) Methods Workgroup on exploring new statistical approaches to data synthesis and methodological approaches to reducing bias in datasets. We provide advice and support to the CEE on the development of the methodologies of systematic reviews. Currently, the CEE Methods Group is based at the *Centre for Evidence Based Conservation* (CEBC), Bangor University, UK.

Invited Article Reviewer. *American Naturalist*, *Behavioral Ecology and Sociobiology, Biological Journal of the Linnean Society, Ecology, Ecological Applications, Ecological Entomology*, *Ecology Letters*, *Environmental Entomology*, *Evolution*, *Experimental and Applied Acarology, Journal of Insect Behaviour*, *Oikos*, *Parasitology* and *Science*.

Invited Grant Reviewer. NSF IGERT review committee for Biogeochemistry and Environmental Biocomplexity research grants: reviewed proposals and allocated $40,000 US among graduate student applicants. – Mar. '06.

GRANTS, SCHOLARSHIPS and ***AWARDS*** .....

a) Grants

- School of Graduate Studies Travel Grant (Cornell University) – '07, $650 US
- School of Graduate Studies Travel Grant (Cornell University) – '06, $150 US
- Andrew W. Mellon Student Research Grant (College of Agriculture and Life Sciences, Cornell University) – '06, $1,200 US
- NSF IGERT in Biogeochemistry and Environmental Biocomplexity Small Research Grant – '05, $2,000 US
- Sigma Xi (Cornell Chapter) Research Support Grant– '05, $600 US
- Department of Ecology and Evolutionary Biology Summer Research Grant (2; Cornell University) – '05 and '06, $1,118 US

b) Scholarships

- National Evolutionary Synthesis Center (NESCent) Postdoctoral Fellowship (Duke University, Durham NC) – $78,000 US + $10,000 research ('08 - '10)
- National Sciences and Engineering Research Council of Canada (NSERC) Postdoctoral Scholarship – $80,000 CND ('08, declined)
- Kathleen and Donald Strong Hull Fellowship (University of Toronto) – '03, $5,900 CND
- NSERC University of Toronto Fellowship – '03, $2,500 CND
- NSERC Postgraduate Scholarship B – $42,000 CND ('03 - '04; '05, declined)
- Ontario Graduate Student (OGS) Postgraduate Scholarship – $30,000 CND ('03, declined)

c) Awards

- American Society of Naturalists, Graduate Travel Award – '07, $800 US

c) Awards - continued

- Teaching Award for "Persistent Dedication in Facilitating the Cornell University Teaching Assistant Development Workshops" (Centre for Learning and Teaching, Cornell University) – '06, $100 US
- University Medal for Outstanding Graduate Work, Master's Level (given to one graduating Master's degree among all fields; Carleton University) – '02
- Canadian Society of Zoologists Travel Award (Lethbridge AB) – '02, $200
- Best B.Sc. Honours Thesis poster in Dept. Biology (Carleton University) – '00, $150

TEACHING EDUCATION .....

- *Internship in Education* (EDUC 620, Dept. of Education, Cornell University). A fifteen-week course on learning and applying constructivist models for teaching undergraduate students (supervised by David Way). – '06.
- *Writing in the Majors Seminar* (WRIT 701, John S. Knight Institute for Writing in the Disciplines, Cornell University). A six-week course on teaching strategies in advanced writing instruction of undergraduates (supervised by Keith Hjortshoj). – '06.
- *Graduate Student Professional Development Workshop* (College of Agriculture and Life Sciences, Cornell University) – '05.

TEACHING EXPERIENCE .....

- **Teaching assistant.** *Evolutionary Biology, Writing in the Majors* (2nd year course; Spring '08 and Spring '06), Cornell University. I currently have 12 students (17 in 2006) for which I have the responsibly to assign their entire course grade.
- **Teaching assistant.** *Evolutionary Biology* (2nd year course; Fall '07 and Fall '05), Cornell University.
- **Teaching facilitator.** *Graduate Student Professional Development Workshop* (College of Agriculture and Life Sciences, Cornell University). – Aug. '06, Jan. '07, Sept. '07, Feb. '07. Co-ordinating microteaching scenarios for new teaching assistants.
- **Lab demonstrator.** *Introductory Biology* (1st year course; Fall/Spring '03/'04), University of Toronto.
- **Lab co-ordinator and demonstrator.** *Animal Behaviour* (3rd year course; Fall '02), Carleton University. I assisted 23 students in conceiving, designing, analysing, and writing individual research projects on behaviour.
- **Teaching assistant.** *Animal Physiology* (3rd year course; Spring '00 and Fall '01) Carleton University.
- **Teaching assistant.** *Introductory Biology* (1st year course; Spring '01), Carleton University.

PROFESSIONAL EMPLOYMENT .....

- **Research Assistant.** Dr. Anurag Agrawal (Aug '06 – May '07), Cornell University. Phyletic and comparative analyses of *Asclepias* milkweeds.
- **Research Assistant.** Dr. Mark Forbes (May '03 – Aug. '03), Carleton University. Co-ordinated and assisted field projects on ecological immunology of damselflies.
- **Research Technician.** Dr. Mark Forbes (Aug. '02 – Feb. '03), Carleton University. Organised and co-ordinated projects on meta-analysis and data analysis.
- **Research Assistant.** Dr. Mark Forbes (Carleton University) and Dr. Maria Servia (University of Santiago de Compostela, Spain), May '02 – July '02. Assisted and organised field projects on fluctuating and directional asymmetry of damselflies and dragonflies.
- **Research Assistant.** Dr. Mark Forbes (May '00 – Sept. '00), Carleton University. Assisted field projects on ecological immunology of amphibians (w/ D. McCruer) and damselflies (w/ C.P. Yourth), and developed techniques for assessing innate immune-ability of insects.

VOLUNTEER and *COMMITTEE POSITIONS* .....

- Organized and started an three-week course on insect diversity for a first-grade class in Fall Creek Elementary as part of the *Graduate Student School Outreach Project* (GSSOP, Cornell University) – April to May '07.
- Invited teaching facilitator for the *Graduate Student Development Workshops* (Centre for Learning and Teaching, Cornell University). I lead a workshop entitled: "Dealing with students in office hours" and "Facilitating learning in the laboratory". – Mar. '07 and Sept. '07.
- Presenter of plant-insect interactions display during the 3rd Annual Insectapalooza (Department of Entomology public open house, Cornell University) – Oct. '06.
- Invited teaching facilitator for the *Graduate Student Development Workshop* (Centre for Leaning and Teaching, Cornell University). I lead a workshop entitled: "Managing stress and achieving balance". – Sept. '06.
- Organiser of the 2007/2008, 2006/2007 and 2005/2006 EEB *Lunch Bunch* Seminar Series (50 seminars) for the Department Ecology and Evolutionary Biology, Cornell University (Ithaca, NY). – Sept. '05 to June '08.
- Taught a C++ programming workshop to 7 graduate students in the Departments of Zoology and Botany at the University of Toronto (3 hours a week, for 3 months). – Apr. '04 to Aug. '04.
- Presenter of parasitism and entomology research during the '00, '01, '02 and '03 Open House at the *Queen's University Biological Station* (QUBS; Chaffey's Locks, ON).
- Undergraduate representative for interviewing applicants to new faculty positions (Biology Department, Carleton University). – Jan. '00 to Mar. '00.