Phylogenetic Patterns of Diversity in Herbivorous Insects

by

Chloe Anna Kaczvinsky

A thesis submitted to the Graduate Faculty of
Auburn University
In partial fulfillment of the
Requirements for the Degree of
Master of Science

Auburn, Alabama
December 14, 2019

Approved by

Nathaniel Hardy, Chair, Associate Professor of Entomology
Alana Jacobson, Assistant Professor of Entomology
Robert Boyd, Professor of Biological Sciences

Abstract

Herbivorous insects comprise an extraordinary amount of the world's biodiversity, but we do not fully understand what led to the extensive diversification within this group. Popular hypotheses focus on coevolutionary dynamics between herbivorous insects and their host plant, but their predictions have not been rigorously tested. I approach this problem in two ways. First, I test a set of UCE (Ultraconserved Elements) probes for their suitability to estimate a better aphid phylogeny, finding that these probes capture enough genomic data to potentially resolve areas of uncertainty in the aphid phylogeny. This is a crucial step for testing hypotheses of diversification and speciation in this group. Next, I perform novel comparative phylogenetics in nymphalids to test the prediction that host switches in herbivorous insects are linked to increased speciation, finding instead that host switches are associated with a decrease in speciation in nymphalids. This suggests that we need to reconsider our assumptions and hypotheses about diversification in herbivorous insects.

Acknowledgements

I would like to thank, first and foremost, the chair of my committee, Dr. Nate Hardy, without whom this thesis would not exist.

I would also like to thank my committee members, Dr. Alana Jacobson and Dr. Robert Boyd, for their invaluable help and advice throughout this process.

I owe a debt of gratitude to all of the professors who gave their time and advice. Thank you, Dr. Jamie Oaks and Dr. Scott Santos for your coding help. Thank you, Dr. Charles Ray, for your help with aphid identifications. Thank you also to Dr. James Fordyce for your advice on chapter three.

Thank you to my lab mates, Gwendolyn Bird and Ricki Hamilton, who have given advice and encouragement when needed.

I have a special note of gratitude to all my friends and classmates, for their support throughout this process.

And, finally, thank you to my parents, who have always supported me, even when you aren't quite sure what it is I'm doing.

Table of Contents

List of Tables

List of Figures

Introduction and Literature Review

General Diversity in Insects

Roughly half of the 1.8 million described species are insects, and roughly half of all insect species are herbivorous (Roskov et al., 2018). This means that about a quarter of described species are herbivorous insects. Matching this species richness is incredible ecological diversity.

Herbivorous insects are found all over the world and feed on almost all species of vascular plants. They are crucial to the structure and function of natural and agricultural terrestrial ecosystems: in the latter they can cause extensive economic losses through damage to crops and the spread of plant pathogens (Blackman and Easthop, 2017). Some, like butterflies and many beetles, are external folivores, that is, they live on the exterior surface of their host plants and feed by chewing leaves (Ali and Agrawal, 2012). Others, like aphids, suck on the sugar-rich phloem of their host plants (Hardy, 2018). Herbivorous insects also vary in the diversity of plant species they consume, that is, their diet breadth. The majority of insects use only one or a few hosts, but smaller and smaller proportions of insect species have broader and broader diets, with some species able to feed on hundreds of plant families (Forister et al., 2015).

Herbivorous Insect Diversification

Why are herbivorous insect so diverse? The most popular hypotheses all take as their starting point that this diversity is due to the ecological interactions between herbivorous insects and their host plants. This is a departure from what biologists assume about speciation in most other organisms. Usually, researchers assume speciation is driven by geographic separation. In herbivorous insects, however, researchers tend to assume that speciation is driven by the evolution of host use. The exact way that host use evolution affects speciation is unknown,

though there are numerous hypotheses. One of the first and most influential is the Escape and Radiate Hypothesis, essentially, a co-evolutionary version of adaptive radiation (Ehrlich and Raven, 1964; Thompson, 1989). In adaptive radiations, a species escapes former constraints on diversity by evolving a new trait or moving to a new environment that exposes it to new ecological opportunities in the form of empty niche spaces. In Escape and Radiate, diversification is driven by reciprocal adaptive radiations, wherein a host plant lineage diversifies after escaping the pressure of insect feeding, often through the evolution of a novel defensive chemical, and an insect lineage diversifies in turn after evolving to overcome this novel defense (Erhlich and Raven, 1964; Thompson, 1989).

Other hypotheses about how host use evolution could drive species diversification in herbivorous insects include the Oscillation Hypothesis (Janz and Nylin, 2008) and the Musical Chairs Hypothesis (Hardy and Otto, 2014). The Oscillation Hypothesis proposes that host-specialist herbivorous insect populations evolve from more generalist populations without losing the possibility of reverting back to hosts in the ancestral diet. In contrast, in the Musical Chairs Hypothesis, the majority of host-use diversification comes from host-switching within specialist lineages rather than specialization within generalist lineages (Hardy and Otto, 2014). The explanation is that host-use evolution in specialists is more likely to result in reproductive isolation and subsequent speciation. At present, several of the assumptions of these hypotheses remain untested.

Hypothesis-Testing Challenges

One of the limitations in testing these hypotheses is lack of both host-use data and a robust phylogeny for a particular study group, because both are required for comparative analysis

of the connection between host-use evolution and species diversification. At present, there are no large clades of herbivorous insects for which we have both a robust phylogeny and near comprehensive host-use data (although some clades come close). Another limitation to testing hypotheses of herbivorous insect diversification is our inability to model and draw inferences from large comparative phylogenetic data sets of host-use evolution and species diversification. However, some feasible approaches to surmount these limitations have yet to be attempted.

Thesis Focal Clades

In this thesis I focus on two clades of herbivorous insect: aphids and nymphalid butterflies. Aphids are a superfamily (Aphidodea) of sap-sucking insects that feed on numerous plant families. There are extensive and robust host-use records for aphid species, especially those that feed on crop plants, but the phylogeny is almost entirely unresolved. The extensive records on host-use in aphid species suggests this could be an important model system for understanding the interrelation of host-use evolution and speciation in phytophagous insects. However, the lack of a resolved phylogeny makes comparative phylogenetics in this group functionally impossible.

The Lepidoptera, unlike aphids, have a relatively well-resolved phylogeny. They were the original study system in the seminal paper proposing the Escape and Radiate Hypothesis (Ehrlich and Raven, 1964) and have remained a model for studying herbivorous insect diversification. The Nymphalidae are a family for which we have a well-supported phylogeny for the relationships among genera and some information about species' host-use, making comparative phylogenetics feasible for understanding evolutionary patterns in these herbivorous insects. However, in comparison to aphids, host use data for nymphalids is far less complete. Despite

this lack of completeness, previous researchers have used this group to test diversification hypotheses (e.g. Fordyce, 2010) and I will do the same.

Species of aphids and nymphalids have markedly different life histories and interactions with their host plants. While aphids are relatively poor dispersers with little ability to actively select their host plants and high migration mortality (Ward et al., 1998), butterflies are excellent dispersers and have the opportunity to be quite choosey about their host plants. Whereas aphids are sap suckers (Hardy, 2018), nymphalids are external folivores (Ali and Agrawal, 2012). This is particularly important, given that different feeding guilds invoke different defensive responses from their host plant (Ali and Agrawal, 2012). By examining host-use evolution in both nymphalids and aphids, it will be easier to draw more general conclusions about host-use evolution patterns, as well as how these vary among orders and feeding guilds.

Thesis Structure

The second chapter of this thesis evaluates a target-capture-based genomic sequencing approach (Faircloth, 2017) to aphid phylogenetics. This study is a pilot study of 14 aphid species that was performed ahead of a larger phylogeny estimate of ~350 aphid species. We found that, although target capture was inefficient relative to the number of loci targeted, it still yielded more than sufficient data for robust phylogeny estimation. This is the first step to addressing the issue of a poorly resolved phylogeny for aphids and will provide a basis for future studies of aphid evolution and biology. In particular, an aphid phylogeny will allow researchers to leverage the comprehensive aphid host-use data for tests of herbivorous insect diversification hypotheses.

The third chapter of this thesis is a study of comparative phylogenetics in nymphalids. There I test whether ancestral host use shifts are correlated with increased extant diversity, as

predicted by Escape and Radiate, finding the opposite is the case. Host-use changes are actually related to significant decreases in nymphalid diversity. This result could be explained in two ways. First, our current models of the evolution of species diversification and complex phenotypes are inadequate, or, second, current assumptions about evolutionary ecology in phytophagous insects need to be revisited.

While chapter two takes a step towards estimating a better aphid phylogeny, chapter three provides towards more sophisticated and powerful ways of using comparative phylogenetic data to test hypotheses of herbivorous insect diversification. Chapter three also brings some of our basic assumptions about herbivorous insect diversification into question, making it clear that further study in additional clades is necessary. Together, these chapters provide a foundation for further studies of phylogenetic patterns of herbivorous insect evolution generally, and host-use evolution specifically.

A Test of UCE Hemipteran Baits for Aphid Phylogeny Estimation

Chloe Kaczvinsky and Nate Hardy

Introduction

Aphids are a diverse group of Hemipterans, with more than 5,000 described species (Hardy, 2018). As the primary vectors of many plant pathogens, they are of the crucial economic importance (Blackman and Easthop, 2017). This numerous group contains many unusual biological characteristics, including generational changes in reproduction, host use, and eusociality (Hardy, 2018). Understanding the evolutionary history of this clade can contribute to better knowledge of the origins and consequences of their unusual biological characteristics. Since aphids are major agricultural pests and disease vectors (Blackman and Easthop, 2017), better understanding their biology could also have important economic benefits. However, current understanding of aphid's evolutionary history is limited.

There are multiple elements of unusual biology in aphids. The first is their life cycle. Most aphids undergo cyclical parthenogenesis, an unusual life cycle in which a generation produced by sexual reproduction is followed by several generations produced asexually (Blackman and Easthop, 2017; Hardy, 2018; von Dohlen and Moran, 2000). Within the broad category of cyclical parthenogenetic reproduction, there is wide variation in life cycle features. Some aphids switch between two unrelated hosts at specific points in the life cycle. Often this host switch occurs in a single year, but some aphids have two-year host alternation cycles. Some aphids even alter the biology of their hosts to create galls, abnormal plant growths that shelter the aphids (Hardy, 2018). In some species cyclical parthenogenesis only occurs in the more temperate parts of an aphid's range and asexual reproduction only occurs closer to the tropics

(Hardy, 2018). Because a single aphid metapopulation can reproduce in different ways across its range, study of these metapopulations could shed light on the causes and consequences of reproductive mode variation. Some aphid species are eusocial and have non-reproductive social castes, including soldier aphids (Hardy, 2018). Many aphid species also have symbiotic relationships with ants, trading protection for sugar-rich excretions called honeydew (Völkl et al., 1999). Comparative studies of aphids can contribute to our understanding of the evolution of life cycle complexity, reproductive mode variation, social complexity and cooperation. However, these efforts are limited by the lack of a robust aphid phylogeny.

Unfortunately, we know little of aphid phylogeny. Most studies have focused on relationships within genera (Normack, 2000; Jousselin et al., 2013) whereas at higher taxonomic levels aphid phylogeny is almost entirely unresolved (Hardy, 2018). Here, we assess the utility for aphid phylogenetics of a new probe set designed for target enrichment of Ultra-Conserved Elements (UCEs) (Faircloth, 2017). This is a crucial first step in estimating a more comprehensive aphid phylogeny that can be useful for ecological and evolutionary studies within this group.

Current Taxonomy

Three families of aphids—the viviparous Aphididae and a sister clade of the oviparous Adelgidae and Phylloxeridae--are broadly accepted by aphid taxonomists (von Dohlen and Moran, 2000). Below the family level, however, there is less taxonomic resolution. Different aphid taxonomists recognize from between 11 to 25 subfamilies of Aphidinae (Hardy, 2018). This taxonomic uncertainty is in part due to the fact that aphids are small insects with very few morphological features that could be used to inform taxonomic classifications based on

morphology (Blackman and Easthop, 2017; von Dohlen and Moran, 2000). Moreover, some of the features used to define clades are likely similar due to convergent evolution rather than shared ancestry (von Dohlen and Moran, 2000). In sum, aphid taxonomy is a system for which molecular phylogenetics may be especially useful but up to this point little molecular taxonomic research has been performed.

Current Phylogeny

Monophyly of aphids and of each of the three aphid families is well supported; along with a sister relationship between the two oviparous families (Adelgidae and Phylloxeridae) (Hardy 2018; Johnson et al. 2019). Unfortunately, suprageneric relationships are murky at best. Most published aphid phylogenies have focused on relationships within genera (for example, Normack, 2000; von Dohlen et al., 2002; Jousselin et al., 2013; Moran et al., 2002; Yang et al., 2010; von Dohlen and Teulon, 2003). Several studies estimated relationships among aphid species as part of a comparative analysis of some aspect of aphid biology (for example, Jousselin et al., 2010; Stern, 1998; Pike et al., 2007). Six published studies have taken aim at family and subfamily relationships (von Dohlen and Moran, 2000; von Dohlen et al., 2006; Ortiz-Rivas et al., 2004; Ortiz-Rivas and Martinez-Torres, 2010; Novakova et al., 2013; Hardy et al., 2015). All but two were based on no more than a handful of loci sampled across fewer than 50 species. Novakova (2013) used 255 gene sequences from an aphid symbiont from 70 aphid species. The other (Hardy et al., 2015) used data from 7 loci sampled unevenly from ~1000 species but found little support for supragneric groupings. Clearly more loci are needed to estimate most relationships. From all these studies there is evidence for monophyly for the subfamily Lachinae

(Hardy, 2018; Ortiz-Rivas and Martinez-Torrez, 2010) but most other relationships are unresolved.

Ultra-Conserved Elements (UCE)

Researchers have recently been successful in estimating phylogenies from target enriched genomic sequencing. One prominent target is Ultra-Conserved Elements (UCEs). UCEs were first discovered as sequences in the human genome that had almost no sequence divergence from even distantly related taxa, such as rats and fugu fish (Bejerano et al., 2004). This conservation makes it possible to design molecular probes that will bind to UCE sites across a panel of divergent genomes. Moreover, the more variable stretches of DNA flanking these sites can be used to infer the history of divergences (Faircloth, 2017; Smith et al., 2014).

A newly developed set of UCE probes could allow researchers to efficiently obtain genome-scale data that could be used for aphid phylogenetics (Faircloth, 2017). However, the performance of these probes for that application is yet to be assessed. This study will serve as the initial test of the utility of these probes for aphid phylogeny estimation.

Methods

Taxon Sampling

Samples of fifteen aphid species were collected from the Southeastern US (See Table 1). These species are exemplars of several of the major aphid lineages, with representatives of the oviparous (Phylloxeridae) and viviparous (Aphididae) families, as well as five subfamilies of Aphididae, including two species from each of the most species rich subfamilies (Macrosiphini

and Aphidini). We also sampled a mealybug species to use as an outgroup for rooting the phylogeny.

Species	**Collector**	**Collection Location**
Prociphilus fraxinifolii	Dr. Charles Ray	Russel Co., AL
Monelliopsis nigropunctata	Chloe Kaczvinsky	George Washington National Forest, VA
Monelliopsis caryae	Chloe Kaczvinsky	George Washington National Forest, VA
Rhadobium porosum	Chloe Kaczvinsky	Auburn, Lee Co., AL
Aphis gossypi	Gwendolyn Bird	Auburn, Lee Co., AL
Sarucallis kahawaluokalani	Gwendolyn Bird	Auburn, Lee Co., AL
Uroleucon ambrosiae	Gwendolyn Bird	Auburn, Lee Co., AL
Sipha flava	Gwendolyn Bird	Auburn, Lee Co., AL
Hammelistes sp.	Ron Miller	Baldwin Co., AL
Phylloxera spinosa	Ricki Hamilton	Dauset Trails, Butts Co., GA
Phylloxera devastrix	Ricki Hamilton	Edisto Nature Trail, Colleton Co., SC
Phylloxera caryaecaulis (var. *caryaemagna)*	Ricki Hamilton	Auburn, Lee Co., AL
Schizaphis grammim	Chloe Kaczvinsky	Auburn, Lee Co., AL
Stemmatomerinx aricula	Dr. David Held	Auburn, Lee Co., AL
Aphis nerii	Dr. David Held	Auburn, Lee Co., AL

Table 1: Aphid Species and Collection Data

Extraction

Specimens were stored in 95% ethanol until DNA was extracted using the Omega E.Z.N.A. Tissue DNA Kit (Norcross, GA), with the following modifications. Instead of allowing pulverized specimens to macerate in the lysis mixture for three hours in a 55 °C water bath, we lysed intact specimens overnight. Specimen cuticles were removed after lysis and stored in 70% ethanol prior to slide mounting and vouchering. After extraction samples were stored in a -40 °C freezer and a ThermoFisher 2000 Nanodrop 2000 (Waltham, MA) was used to determine DNA concentration.

Vouchering

Recovered cuticles were slide mounted as vouchers when possible, but some specimens did not survive the initial extraction steps due to mechanical and chemical damage to the cuticle. Specimens of *Aphis nerii* from which DNA was extracted were not vouchered, but specimens from the same collection event were. Vouchered specimens were stained and desiccated using double stain and Essig's aphid fluid, 70% ethanol, and clove oil. They were then slide mounted using balsam resin.

Library Prep

DNA Extractions were sent to ArborBioScience (Ann Arbor, Michigan) for library prep and Illumina Sequencing using the Ultra-Conserved Element (UCE) probe set designed by Faircloth (2017) for Hemiptera, which includes 40,207 baits targeting 2,731 UCE sites. In testing, Faircloth (2017) recovered an average of 2381 loci. Samples were quantified, sonicated, and then selected for roughly 400 bps using SPRI beads. Adapters were added using an A-tailing Y-adapter method and samples were amplified for 8 cycles with 8nt sequence indexing primers. Four sample libraries were then pooled with 250 ng per sample.

Library prep was carried out by using myBaits version 4.00 (ArborBioSciences; Ann Arbor, Michigan) with 500 ng probes, using the MyBaits Library Prep Protocol, with minor variations. IDT's xGen Universal Blocking Oligonucleotides was used in place of Block A in the hybridization mixture. Hybridization was carried out at 60 °C. Reamplification of the samples proceeded for 10 cycles. Samples were sequenced using 150 bp paired-end read Illumina sequencing on a partial NovaSeq S4 lane.

Sequence Analysis

Illumina reads were processed using the Python package phyluce (Faircloth, 2016). Sequence adapters and indices were trimmed and quality control carried out with trimmomatic (Bolger et al., 2014). De novo assemblies were performed with Velvet (Zerbino and Bimey, 2010) on the Alabama Supercomputer, with kmer values at 55. Orthology assessment was done by blasting sequences against the UCE loci in the Hemipteran probes using phyluce (Faircloth, 2016). We conducted multiple sequence alignment using MAFFT (Katoh and Standley, 2013). For each alignment, we used trimAl (Capella-Gutiérrez et al., 2009) to remove all sites with gaps in more than five percent of the alignment, unless this was less than twenty-five percent of the total data, in which case the top twenty-five percent of alignments were preserved.

Tree Estimation

Phylogenetic relationships between species were estimated with RAxML (Stamatakis, 2014) using a global general time-reversible nucleotide substitution model and a heuristic search algorithm. The search used rapid hill-climbing on 100 bootstrapped data sets with CAT-approximated among site-heterogeneity. Every fifth optimal boot-strapped tree was used as a starting tree for optimization of the observed data, again using a general time reversible model, and a gamma-distributed among site rate variation.

Results

We recovered sequence data across the fifteen included species for 403 UCE loci of 2,731 targeted by the bait set (Faircloth, 2017). After contig assembly, there were 31,219 contigs for each sample on average, with the lowest being 5,341 in *Prociphilus fraxinifolii* and the

highest being 132,440 in *Phylloxera caryaecaulis*. Each assembly had between 423 (in *Schizaphis grammium*) and 1,738 (in *Sipha flava*) contigs that were more than 1,000 base pairs in length, with an average of 1077 1Kbp+ contigs. In all, we captured data from 2,168 loci. However, the majority of loci were captured for only half of the samples. Using a threshold of complete taxonomic inclusion, we ended with data for 403 loci (25% of the targeted loci), and a multiple sequence alignment supermatrix with 39.5% gap characters. Data from the 403 loci enabled us to recover phylogenetic relationships in accordance with the classification and with strong bootstrap values for all relationships but the sister relationship between *Hamamelistes* sp. and the clade encompassing species from Eristominae, Chaitophorinae, and Aphidinae. The final ML optimization likelihood was -1213155.98. The tree estimate is shown in Figure 1.

Discussion

This study demonstrates the potential efficacy of using Faircloth's (2017) UCE Hemipteran probes to estimate aphid phylogeny, especially the deeper relationships between aphid lineages that have yet to be estimated with any confidence. Many of the 2,731 UCE sites were not recovered for aphids in this test, but we had sufficient data to estimate a phylogeny. It is clear from the various quality control and trimming steps that most of the data and loci captured by this probe were captured for only some of the samples. This is consistent with Faircloth's initial tests, where of the 2,381 sites detected only 1,444 loci were present for 75% of the samples (Faircloth, 2017). Nevertheless, the 403 loci we were able to recover dwarfs previously published data sets —one of the largest of which was 7 individual loci in aphids (Hardy et al., 2017)—and were sufficient to estimate a well-supported aphid phylogeny. Not only was the size

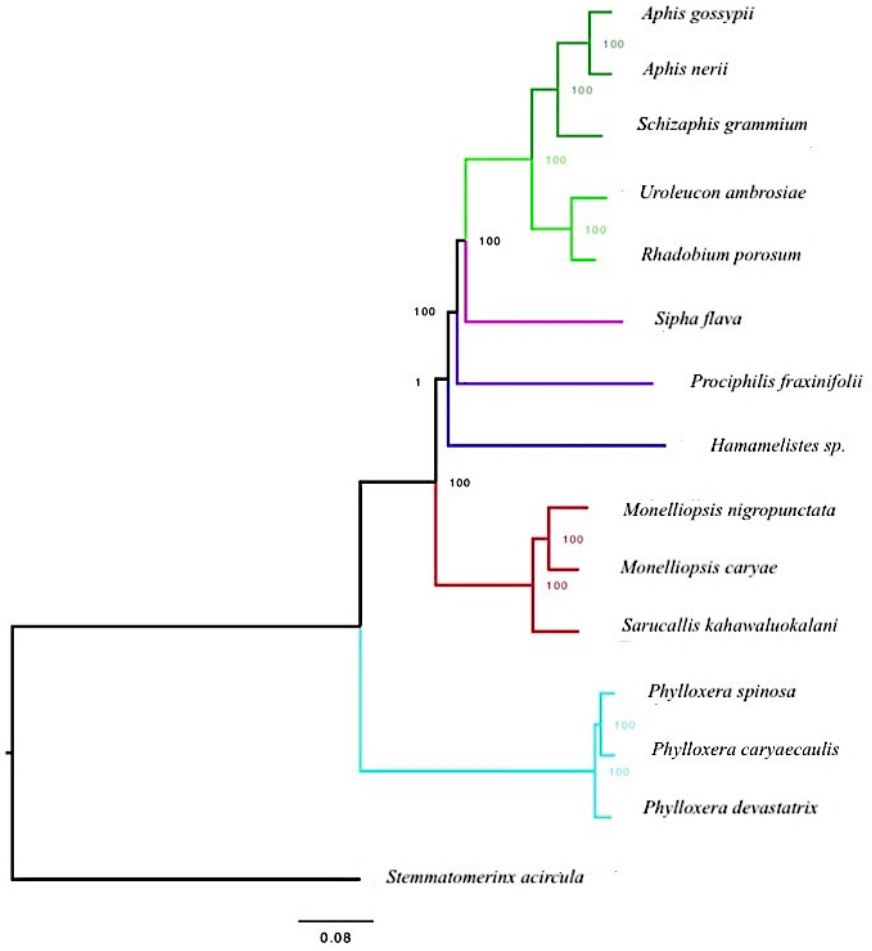


Figure 1. The maximum likelihood estimated phylogenetic tree. Numbers at the nodes are the bootstrapping support for the node and are excluded around the branch selected for rooting. Each family (for Phylloxeridae) and subfamily (for Aphididae) is color-coded and tribes in the same subfamily are different shades of the same color. Phylloxera species are in turquoise. The subfamily Calaphidinae is in red; the subfamily Hormaphidinae is in indigo; the subfamily Eristominae is in purple; the subfamily Chaitophorinae is in magenta; and the subfamily Aphidinae is in green. The tribe Macrosiphini in Aphidinae is in pale green and the tribe Aphidini in Aphidinae is in hunter green.

of the supermatrix alignment comparable to those in many target-enriched phylogenomic studies (Starrett et al., 2016; Bossert et al., 2019) but nodal support values were almost invariably high, demonstrating that even with low capture rates for the bait set, there was still sufficient data capture for phylogeny estimation.

Future Directions

This work is a pilot study; our aim was to assess the utility of the Faircloth (2017) Hemipteroid UCE probe set for aphid phylogenetics. We found that, although imperfect, it is still quite useful and a vast improvement on previous approaches. We are currently working to apply this approach at broader scale. Specifically, we aim to estimate relationships among 335 species of Nearctic aphids. We have already completed the DNA extraction and quantification and have almost completed the specimen vouchering. Our sequencing service provider has completed the library preparations, and our samples are currently in the queue for Illumina processing. Once we get these data, and use them to estimate the Nearctic aphid phylogeny, we will use that estimate in comparative analysis to identify factors that have driven aphid speciation.

Conclusions

This pilot study is a step towards building an improved aphid phylogeny. The techniques tested here bring us closer to a well-resolved picture of aphid. With a better aphid phylogeny, we can examine elements of aphid biology and evolution, such as the drivers of speciation. A better phylogeny could also be used to examine the evolutionary history of traits like eusocialism in the Ceraphidini (Stern, 1998) and the complex life cycles of cyclical parthenogenesis and host-switching found in many aphids (Hardy, 2018). In sum, a better understanding of aphid phylogeny is essential for understanding the dynamics that have led to the diversity in aphids.

Do Major Host Shifts Spark Diversification in Butterflies?

Chloe Kaczvinsky and Nate Hardy

Abstract

The Escape and Radiate Hypothesis posits that herbivorous insects and their host plants diversify through antagonistic co-evolutionary adaptive radiation. For more than fifty years it has inspired predictions about herbivorous insect macro-evolution, but only recently have the resources been generated to rigorously test those predictions. Here, we test two of these predictions with comparative phylogenetic analyses of nymphalid butterflies: that major host switches tend to increase species diversification, and that such increases will be proportional to the scope of ecological opportunity afforded by a particular novel host association. We find that in general the effect of major host-use changes on butterfly diversity is opposite what was predicted. In general, although it appears that the evolution of a few novel host associations may have caused bursts of speciation, major changes in host-use tend to be linked to significant decreases in butterfly species richness.

Keywords

Evolutionary ecology, co-evolution, species diversification, plant-insect interactions

Introduction

About half of the 1.6 million described species are insects, and about half of all insect species are herbivores (Roskov et al., 2018). How did herbivorous insects become so speciose? The Escape and Radiate Hypothesis (Ehrlich and Raven, 1964; Thompson, 1989) posits a co-evolutionary version of adaptive radiation. In this scenario, when a plant lineage evolves a new

chemical defense it escapes from its herbivores, enters a new adaptive zone, and diversifies ecologically and taxonomically. Reciprocally, when an herbivorous insect lineage evolves a counter-adaptation to a plant chemical defense, it escapes from its competitors and likewise diversifies. The Escape and Radiate Hypothesis can be traced to Ehrlich and Raven's (1964) essay on co-evolution, in which they surmise that "the fantastic diversification of modern insects had developed in large measure as the result of a stepwise pattern of co-evolutionary stages." To be clear, the main focus of Ehrlich and Raven (1964) was on co-evolution *per se* and their comments on species diversification dynamics were general and speculative. The name of this hypothesis (Escape and Radiate) was introduced by Thompson (1989). Several authors since then have sought to more fully develop the Escape and Radiate Hypothesis by making more specific and testable predictions about how host-plant co-evolution might affect the process and pattern of herbivorous insect species diversification (e.g., Fordyce, 2010; Janz, 2011; Janz and Nylin, 2008). Here, we use the name Escape and Radiate as short-hand for this extended set of predictions, several of which have yet to be tested (Futuyma and Agrawal, 2009).

As an extension of the basic theory of adaptive radiation, the Escape and Radiate Hypothesis inherits many of the same underlying assumptions. It assumes that release from constraints on diversity will result in the evolution of specialist species, rather than the niche expansion of generalists (Yoder et al., 2010). Hence, it predicts that the colonization of novel host groups will increase herbivorous insect species diversity. Implicit here is that both plant defenses and herbivorous insect diets are phylogenetically conserved (Futuyma and Mitter, 1996; Kergoat et al., 2007). Otherwise, the notion of co-evolutionary adaptive zones becomes problematic; for example, evolving to overcome one plant species' defenses would not allow an insect population to overcome the defenses of a related plant species. Another assumption that

the Escape and Radiate Hypothesis inherits from the general theory of adaptive radiation is that adaptive zones can be saturated. Hence, it predicts that the colonization of a novel host group will cause an immediate uptick and then a subsequent slowing of speciation rates as the novel adaptive zone is filled (Losos and Mahler, 2010). By extension, the Escape and Radiate Hypothesis predicts that some novel host associations should represent greater ecological opportunities and more expansive adaptive zones than others and that the dimensions of these zones should determine their effects on species diversity (Farrell and Mitter, 1994; Schluter, 2000).

Evidence for and against the Escape and Radiate Hypothesis

Evidence regarding the Escape and Radiate Hypothesis is mixed. Researchers have documented several putative cases of a plant lineage escaping from its herbivores and undergoing a subsequent burst of species diversification (Farrell et al., 1991; Futuyma and Agrawal, 2009). As for herbivorous insects, several phylogenetic studies have shown apparent links between particular host-use shifts and increased species diversity (Braby and Trueman, 2006; Futuyma and Agrawal, 2009; Wheat et al., 2007). Currently, we lack a quantitative sense for the overall phylogenetic conservation of plant chemical defenses (Agrawal, 2007), as clear evidence of phylogenetic conservatism has been found for some plant defensive chemistries (e.g., Liscombe et al., 2005; Wink and Mohamed, 2003) but not others (e.g. Wink, 2003). Likewise, host-use appears to be phylogenetically conservative in some herbivorous insects (for example butterflies; Janz and Nylin, 2008) but not others (for example mealybugs; Hardy et al., 2008). The Escape and Radiate Hypothesis would seem most applicable to groups such as butterflies for which the assumptions of phylogenetic conservation of host-use and defensive

chemistry are met in at least some host groups. To date, Fordyce's (2010) study of butterflies has been the most comprehensive test of macro-evolutionary predictions of the Escape and Radiate Hypothesis. In it, he presents evidence for temporary increases in speciation rates after the evolution of a handful of major novel host associations, classified as such *a priori*. To be sure, such bursts of speciation are as expected by the Escape and Radiation Hypothesis (Fordyce, 2010), but, as it stands, we do not know if such effects are typical or exceptional; the prediction that major host shifts will spur diversification has yet to be tested with statistical rigor.

Here, we use phylogeny-based statistical analyses of butterflies to address three key questions: (1) Do major new host associations tend to cause a burst of speciation (Fordyce, 2010)?; (2) Do major new host associations tend to cause lasting increases in species diversity (Janz and Nylin, 2008)?; and (3) Do novel host groups affording greater ecological opportunity cause greater increases in butterfly diversity (Losos and Mahler, 2010; Schluter, 2000)?

Methods

We use brush-footed butterflies (Papilionoidea: Nymphalidae) as a model system, since both their larval host associations and phylogenetic relationships are relatively well known. We worked at the level of nymphalid genera, of which 398 are extant. To test the macro-evolutionary predictions of the Escape and Radiate Hypothesis, we needed (1) reconstructions of ancestral host-use of nymphalids, (2) quantifications of the scope of ecological opportunities opened by evolving specific novel host associations (host gains, for short), and (3) a statistical approach to evaluate how major host-use changes affect speciation rates and extant species diversity. These data came from two main sources: host-use and species diversity data for nymphalid genera were obtained from Hamm and Fordyce (2015) and phylogeny data from Wahlberg (2006).

Reconstructions of Ancestral Host-Use

All analyses were performed using R (R Core Team, 2019). First, we used Dispersal Extinction Cladogenesis (DEC) models (as in Hardy, 2017) to reconstruct the phylogenetic history of the use of host orders and families. Although DEC models were initially developed to estimate ancestral geographic ranges, they are well-suited for the estimation of ancestral states for any multi-state discrete character such as host-use (Hardy, 2017). An alternative approach would have been to code the use or non-use of each nymphalid host-plant taxon as a binary trait and then use standard discrete trait models to independently reconstruct phylogenetic histories of the use of each host. However, simulations have shown that such an approach reconstructs ancestral host-use with a strong bias towards the present and tends to infer ancestors without any hosts at all – a problem that DEC estimation avoids (Hardy, 2017). In our DEC models, the host-use of each extant nymphalid genus and each ancestral nymphalid node is expressed as a combination of discrete host taxa. The feasibility of DEC modeling is limited by the size of the matrix which specifies the probabilities (or rates) of each type of possible host-use transition (Matzke, 2014). If this rate matrix is too large, computations are intractable. A full matrix, with terms for every possible combination of host plant taxa and for every possible change between those combinations, would have been unfeasible.

We took two approaches to keep rate matrices under 1,600 states (to limit analysis run times under two weeks each). First, we reconstructed the history of associations between nymphalids and their twelve most commonly used host plant orders with a maximum nymphalid genus host-breadth size of five orders. This required eliminating 23 of the most polyphagous nymphalid genera from the dataset, along with nine additional genera for which we lacked host-

use data, leaving 357 genera (89.6% of the total) for analysis. Second, we reconstructed ancestral use of host plant families over a set of nymphalid subclades using an R script (Supplemental Document 1) to divide the phylogeny into the most inclusive set of non-overlapping clades that comprised at least ten extant nymphalid genera and would result in DEC rate matrices with <1,600 states. We also excluded from consideration any host family used by only one nymphalid genus. This yielded nine nymphalid subtrees, encompassing 238 of the 398 genera and 51 host plant families. Note that this approach was not entirely inclusive and was biased against clades containing genera with especially broad host associations. The order-level reconstructions were not subject to these biases, but host-use variation at the level of families may be more biologically meaningful. Ehrlich and Raven (1964) suggested that many family-level plant taxa can be traced back to defensive innovation, and many subsequent authors have used plant family diversity as a proxy for the diversity of their defensive chemicals (e.g., Fordyce, 2010; Hardy and Otto, 2014). Hence, we conducted analyses at both taxonomic levels.

DEC estimations of ancestral host-use were performed with the R package BioGeoBEARS (Matzke, 2013). Specifically, we used the DEC* model, which excludes the ancestral null state (Massana et al., 2015) and thereby requires all ancestors to have a host, a constraint that we think better reflects biological reality (Massana et al., 2015; Matzke, 2014). (See Figure 2 for an example host-use reconstruction; figures for all reconstructions are provided in Supplemental Documents 2-10.)

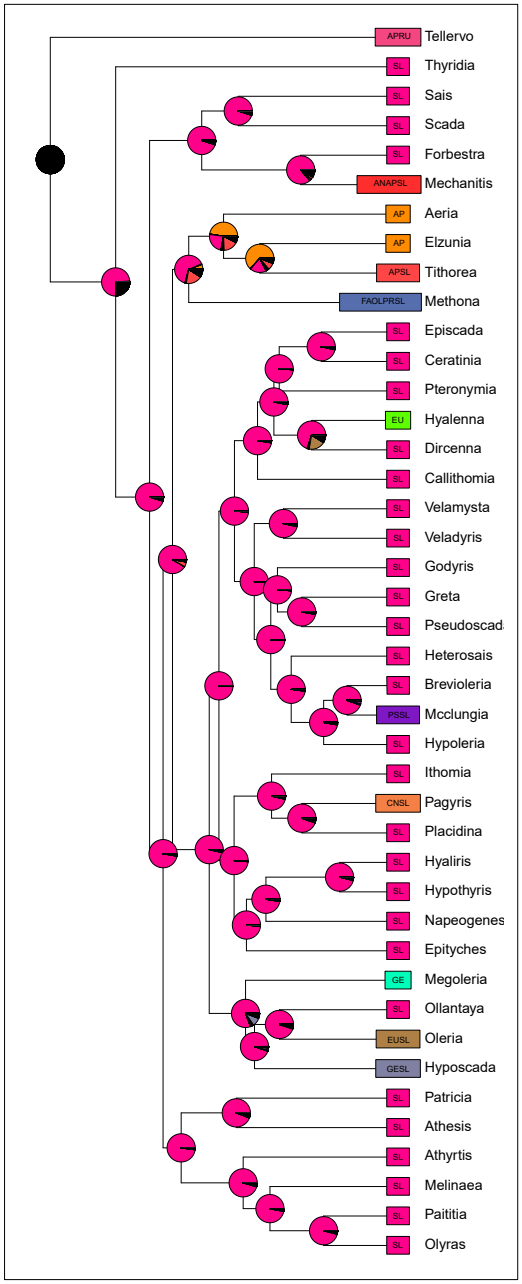


Figure 2. Example DEC* reconstruction of the use of host plant families over part of the nymphalid genus phylogeny. Each host use state has a unique color and the internal pie charts show the proportional likelihood of ancestral host use states. Host abbreviations: AP: Apocynaceae, RU: Rubiaceae, SL: Solanaceae, AN: Annonaceae, FA: Fabaceae, OL: Oleaceae, PR: Primulaceae, EU: Euphorbiaceae, PS: Passifloraceae, CN: Convolvulaceae, GE: Gesneriaceae.

Linear Model Parameterization 1 – Effects of Host-Use Changes on Diversity

We used a linear modeling approach to test predictions of the Escape and Radiate Hypothesis. We first sought to explain the variation across nymphalid lineages in species diversity and diversification rates with their phylogenetic history of host-use. For these models, we used a single predictor variable (***Host-use change)*** with three levels, indicating at each internal node if a host was gained, lost, or if no change in use occurred. This classification allowed us to test if host gains and losses tend to increase or decrease diversity.

We used three response variables: (1) phylogenetically-independent contrasts of speciation waiting times, (2) phylogenetically-independent contrasts of extant species diversities, and (3) gamma statistic values (Pybus and Harvey, 2000). Here we explain each.

Speciation waiting times. For each internal node in the butterfly phylogeny, we calculated the average length of the branches leading to its two immediate descendant nodes (i.e., the average waiting time for

speciation). We then performed the same calculation for the focal node's sister node. The difference in average waiting times between sister nodes was used as a phylogenetically independent contrast of speciation rates. The sign and magnitude of these contrasts can show us how the gains and losses of novel host associations tend to immediately affect speciation rates.

Extant diversity. For each internal node in the butterfly phylogeny at which a host gain occurred, we calculated the total number of extant descendant species classified in genera that are known to continue to use that host. Then, in the focal node's sister clade, we counted all of the extant species in genera that do not use the novel host. For losses, we did the inverse and counted the extant diversity only of species that did not use the lost host and compared it with descendants of the sister node that did use the lost host. Differences between these extant diversities yielded a phylogenetically independent contrast of how the evolution of a novel host association affects species diversity in the long term. These were called exclusive diversity contrasts. To account for genera removed from the phylogeny during DEC reconstructions of host family use, these calculations were performed on the corresponding nodes on the complete tree (Wahlberg, 2006). We also calculated contrasts of the *total* extant species diversities spanning each internal node; i.e., we contrasted the extant diversity of a clade descended from an ancestor that gained or lost a new host with the extant diversity of its sister clade, while ignoring whether or not extant taxa use (or do not use) that host (See Supplemental Document 11 for R scripts). We did this for all internal nodes, regardless of whether a host was lost, gained, or if no change occurred. These contrasts were called inclusive diversity contrasts. For nodes at which no change in host use occurred, all descendants were contrasted in both the inclusive and exclusive diversity contrasts. This contrast is a more direct measure of the effect of a major host-use

change *per se*, as opposed to the effects of continued occupation of or exclusion from a particular adaptive space.

Gamma statistic. This is a measure of the degree to which speciation dynamics depart from expectations of an equal-rates Markov model of phylogenetic branching (Pybus and Harvey, 2000). A value of zero corresponds to a constant diversification rate. Negative values indicate that speciation rates slow from the root to the tips, i.e., there is an early burst of speciation (Pybus and Harvey, 2000), which is expected of an adaptive radiation. Positive values of the gamma statistic are difficult to interpret. They could indicate that diversification rates accelerate from the root the tips, i.e., there is a late burst of speciation. Or they could simply be indicative of the so-called "pull of the present;" an apparent late uptick in speciation rates due to the lag between speciation and extinction under a constant-rate diversification regime. Hence, Pybus and Harvey (2000) recommend that positive values be ignored. We calculated the gamma statistic for all clades composed of at least ten nymphalid genera. (See Figure 3 for a schematic of diversification model covariates.)

Linear Model Parameterization 2 -- Effects of Ecological Opportunity

We developed four indices for the scope of the ecological opportunity opened by the evolution of a novel host association, and then used these as predictor variables to explain changes in butterfly diversity linked to evolutionary host gains. Each index is described below.

Host age was the estimated phylogenetic age of the host plant taxon (family or order). Phylogenetic ages were obtained from the TimeTree database (Kumar et al., 2017). All else being

equal – in particular rates of speciation and ecological evolution – older plant lineages should be more diverse, which may correspond to larger ecological opportunities and adaptive zones.

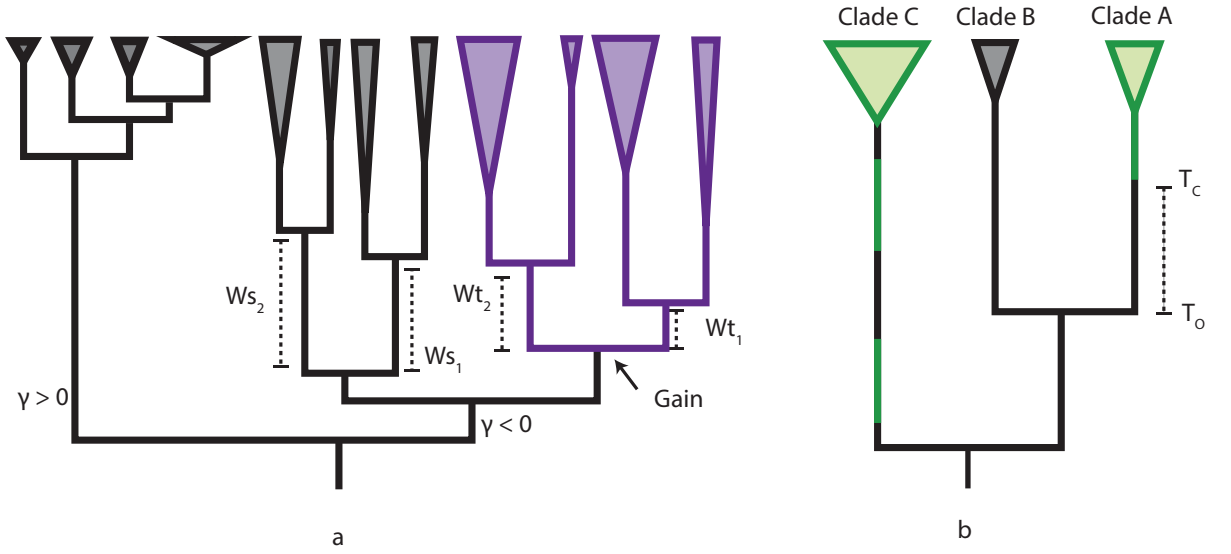


Figure 3. A schematic of diversification model covariates. a. Response variables illustrated on hypothetical butterfly phylogeny where the width of triangles at tips is proportional to extant species richness. Branches in purple correspond to lineages with a novel host association. Waiting times for speciation following the evolution of a novel host association are indicated with Wt_x. Speciation waiting times for the sister clade are indicated with Ws_x. Slowing speciation rates correspond to a γ value less than zero. By contrast, accelerating speciation rates correspond to γ values greater than zero. b. Predictor variables illustrated on hypothetical plant phylogeny. Green branches are used by a butterfly lineage, black branches are not; the width of triangles at tips is proportional to extant species richness; T_O is the stem age of a plant clade; T_C is the time at which that plant lineage was colonized by a butterfly lineage, and the difference between T_O and T_C is the early bird statistic; clade A has a more volatile host use history by butterflies than clade C.

The early adoption index was the difference between the stem age of a host taxon and the time at which it began to be used as a host by a nymphalid lineage. Thus, it measures how quickly a particular butterfly lineage colonized a new host. If earlier colonizers are exposed to less competition and more open niche space, that could amount to greater ecological opportunity.

Host diversity was simply the current species richness of the host family or order, according to the Catalogue of Life database (Roskov et al., 2018), accessed with the R-package taxize (Chamberlain and Szöcs, 2013).

Host volatility was a count of how many times a particular host group was gained over the phylogeny, divided by the host age. Previous work has demonstrated that ancestral host associations condition the probability of host switching in extant populations (Futuyma et al., 1995; Janz et al., 2001). Hence, many apparent host gains can be considered as the re-expression of a latent phenotype, and the phylogenetic history of the use of some host taxa can appear quite volatile. By contrast some hosts are seldom gained and lost. Colonization of such hosts may represent more novel niche transformations and greater ecological and evolutionary opportunities.

Model Fitting

To repeat, we sought to address two questions: Do host-use gains tend to increase diversity? and, if so, Do novel associations corresponding to greater ecological opportunities spark greater diversity gains? For the first question we examined how host-use gains and losses affected extant species diversity, speciation waiting times, and the overall dynamics of diversification (as measured with the gamma statistic). For the second question, we examined the same response variables, but only for host-gains, and tried to explain the variation in the response variables with several indices of the breadth of ecological opportunity afforded by a novel host group. Two of the response variables – speciation waiting times and extant species diversities – were phylogenetically independent contrasts. Therefore, we could use standard linear modeling methods to estimate the fixed effects of the predictor variables on their variance

(we used the built-in lm R function). To account for uncertainty in the DEC* reconstructions, we weighted each model covariate with a vector of the proportional probabilities for each estimated ancestral host-use state. For these models, as is standard for models of phylogenetically independent contrasts, we forced the regression to pass through the origin.

Values of the third response variable, the gamma statistic, were not phylogenetically independent. To account for this, we fitted linear mixed models in which phylogenetic relatedness between nodes was expressed using a pedigree structure and included this as a random effect with the Bayesian approach implemented in MCMCglmm (Hadfield, 2010). Analyses consisted of 1,000,000 MCMC iterations with a thinning interval of 100. We used the Geweke diagnostic to confirm that we had sampled sufficiently from the stationary distribution (see iix Tables 1-16 for all model results). To correct for bias in gamma statistic estimates due to incomplete sampling of branches in a genus-level analysis, we weighted each empirical gamma statistic by its distance in standard deviations from the mean gamma statistic value estimated from 100 simulations under a null birth-death model, as implemented by the MCCR test (Pybus and Harvey, 2000) in the R package phytools (Revell, 2012). Note that we incorporated these weights by using the 'mev' argument of MCMCglmm, which is intended to take a vector of effect size variances for a meta-analysis (See Supplemental Document 12 for full model specifications).

Explicit State-Dependent Early Burst models

As previously mentioned, simulation studies (Hardy, 2017) have shown that DEC models appear to accurately estimate the ancestral states of multi-state discrete traits, but only if trait states are assumed to not strongly affect species diversification rates. Violation of that

assumption can bias reconstructions (Maddison et al., 2007). Several models have been developed that can estimate the ancestral states of a discrete trait while explicitly accounting for state-dependent variation in speciation and extinction rates (for example, see Goldberg et al., 2011). Unfortunately, it is currently not feasible to fit such models to traits with as many states as host-use in nymphalids. To work around that constraint, and as a complement to our DEC-based analyses, we fit explicit state-dependent early burst models to the phylogenetic history of binary (use or non-use) traits for each nymphalid host taxon, using the fitDiscrete function in the R package geiger (Harmon et al., 2008). Then, for each host taxon, we compared the fit of the early burst model to a model with constant branching and extinction. Comparisons were made with likelihood ratio tests, using the R package extRemes (Gilleland and Katz, 2016). To be clear, both our main DEC-based analyses and the state-dependent early burst analyses are subject to known biases, but these biases are different.

Results

In our plant family-level models, both host-use gains and losses were negatively correlated with extant diversity. As explained above, we calculated contrasts of diversity in two ways for gains. In the first, which we refer to as exclusive contrasts, we compared the number of species in a focal clade that used a particular host group to the number of species in the sister clade that did not use that host group. In the second, which we refer to as inclusive contrasts, we compared the total extant species diversity (regardless of current host-use) descended from an ancestor that gained a host group to the total species diversity of its sister group. For losses, we did the inverse. In the exclusive contrasts family-level model, nodes with a host-use gain had on average 16.37 fewer extant descendant species than their sister nodes (p-value 0.042), while

nodes with a host-use loss had on average 36.94 fewer extant descendant species than their sister nodes (p-value: 0.031). In the inclusive-contrast family level model, nodes with a host-use gain had on average 20.50 fewer extant descendant species than their sister node (p-value 0.011), while nodes with a host-use loss had on average 35.88 fewer extant species than their sister node (p-value: 0.036) (see Table 2 and Figure 4). To put those values in perspective, the average summed species diversity of the focal and sister clades involved in a contrast was 69.37. In both cases, the estimated magnitude of the effects on diversity was greater for losses than for gains. Because of the paucity of internal nodes at which family-level host-use changes were reconstructed on nymphalid subtrees, we were unable to estimate the effects of ecological

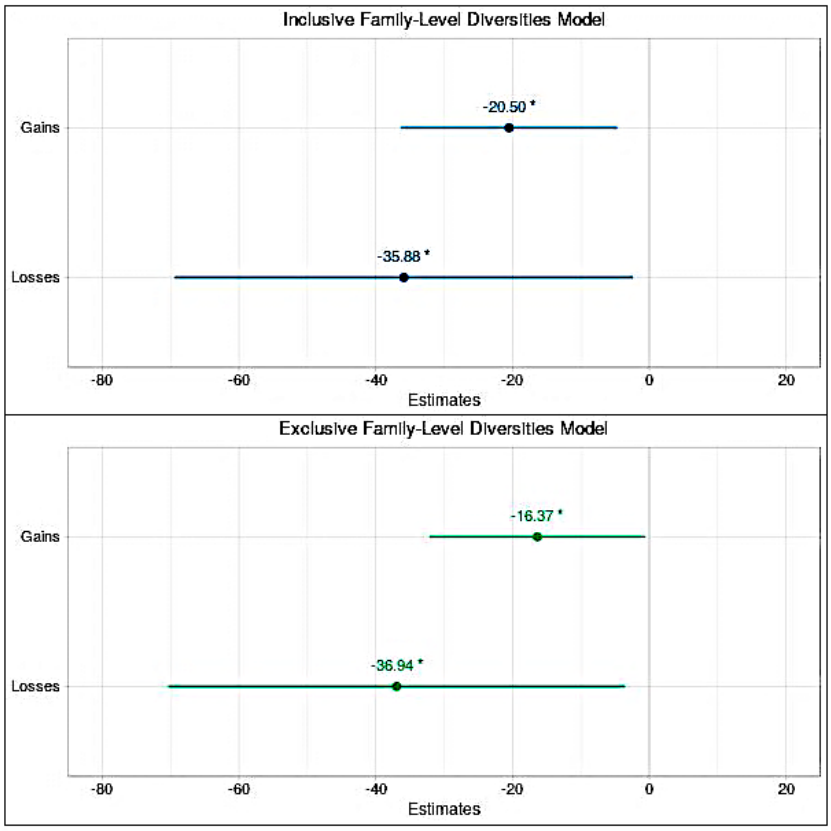


Figure 4. Plots of estimated co-efficients for significant effects shown along with 95% confidence interval

	Co-efficient	**P-value**
Exclusive Gain	-16.4	0.042
Exclusive Loss	-36.9	0.031
Inclusive Gain	-20.5	0.011
Inclusive Loss	-35.9	0.036

Table 2: Values for Host Use Change as a Function of Species Diversity (Family Models)

opportunity proxies on variation in gamma statistic values or speciation waiting times. We found no significant effects from any of the ecological opportunity proxies on variation in extant diversities (Appendix Tables 8-16).

In the order-level models, host-use change did not significantly affect diversity. Likewise, proxies for the magnitude of ecological opportunity were mostly not significantly corelated with diversity dynamics (Appendix Tables 1-8). There was one exception; gamma statistic values were positively correlated with early-adoption index values (estimated coefficient: 0.22, p-value: 0.012; See Figure 5) (See Supplemental Document 13 for a figure of the distribution of gamma statistic data points.)

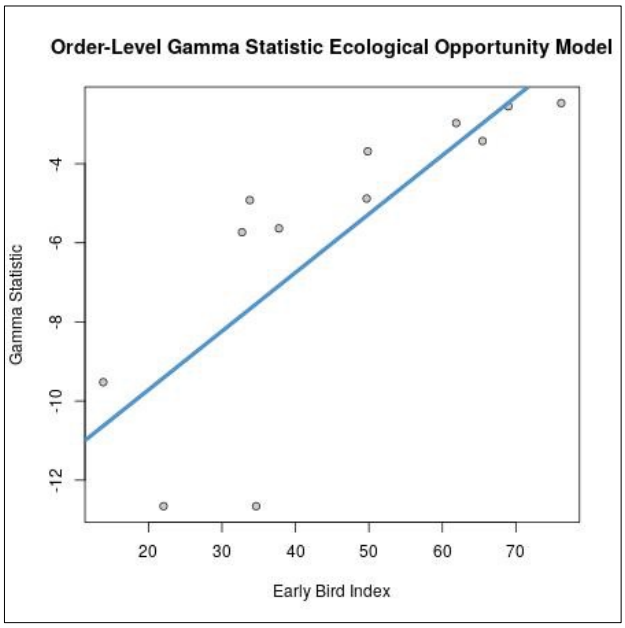


Figure 5. Plot of simplified linear model of co-efficients for order level model of gamma statistic as a function of early bird

We found that a state-dependent early burst model of phylogenetic branching was a better fit than a constant diversification rate model for just seven of 128 host plant families: Apiaceae ($p<0.00001$),Apocynaceae ($p<0.00001$), Boraginaceae ($p<0.00001$), Cyperaceae ($p<0.00001$), Euphorbiaceae ($p<0.00001$), Rosaceae ($p<0.00001$), and Verbenaceae ($p<0.00001$). For full results, see Appendix Table 17.

Discussion

Does the Escape and Radiate Hypothesis predict phylogenetic patterns of host-use and species diversity in nymphalids? For the most part, this does not seem to be the case. We found

no support for the broadest prediction, that the evolution of major new host associations tends to boost herbivorous insect diversity over short or long phylogenetic terms. To the contrary, we find evidence that novel associations tend to decrease extant butterfly species richness. It could be that major evolutionary changes in host-use are caused by major declines in fitness on ancestral hosts. Such declines could be due to a variety of factors including increased competition for diminished host plant resources, evolution of novel host defenses, invasion of host herbivore assemblages by new herbivore species, and changes in natural enemy communities (Bird et al., 2019; Kenis et al., 2009; Segraves and Anneberg, 2016). Regardless of the cause, if major evolutionary changes in herbivore diet are sparked by such calamities, we might expect such changes to be linked to long-term decreases in herbivorous insect species diversity. With reduced performance on ancestral hosts, and likely marginal performance on new hosts, growth rates and effective sizes of herbivorous insect populations could shrink along with their geographic and climatic niche ranges. This could increase the odds of extinction and decrease the odds of speciation. Thus, major evolutionary shifts in the diets of butterflies could be the result of ecological crashes more than ecological opportunities.

On the other hand, we did find evidence for some early burst dynamics linked to the evolution of the use of specific plant families. This suggests that co-evolutionary adaptive radiation may indeed have played a role in the diversification of herbivorous insects. But such events appear to have been relatively rare in the history of nymphalids; a few novel host associations have sparked explosive diversification, but most have not.

The Escape and Radiate Hypothesis also predicts that a novel host association should affect species diversity in a way that is proportional to the scope of ecological opportunity offered by the new host group. Only one of our indices of ecological opportunity explained a

significant proportion of the variation in our butterfly diversification dynamics variables. We found that gamma statistic values were more negative – indicating an earlier burst of diversification (Gavrilets and Losos, 2009) – when butterfly lineages colonized a novel host order earlier in the butterfly's evolutionary history. This is consistent with our assumption of the existence of greater empty niche space at the early stage of a plant lineage's evolutionary history. This assumption in turn is based on the assumption that origins of major plant clades are linked to events which release plants from their previous consumers. Therefore, this correlation could be interpreted as support for the concept that the magnitude of ecological opportunity is a key limitation on herbivorous insect diversity. Nevertheless, since we found that the estimated overall effect of host-use gains on butterfly species diversity was negative, this would imply that broader ecological opportunities on novel hosts reduced how much diversity was negatively affected by a major host shift, rather than lead to rapid speciation.

This study has several limitations. Most importantly, we worked within the limits of the current models for reconstructing the phylogenetic history of a complex trait like host-use. As mentioned above, it is currently not possible to reconstruct the ancestral states of a highly multi-state discrete trait while accounting for state-dependent variation in speciation and extinction rates (Maddison et al., 2007). To work around this constraint, we took two approaches, each with its own shortcoming. In the first, we reconstructed host-use evolution with state-independent diversification DEC* models and then performed *post-hoc* analyses of diversity patterns. In the second, we fit explicitly state-dependent models of early burst diversification, but on a series of binary host-use characters. The shortcoming of the first approach is the assumption of state-independent diversification. The shortcoming of the second is that when a multi-state trait such as host-use is modeled as a series of binary traits we tend to reconstruct ancestors without hosts

(Hardy, 2017). The picture of nymphalid diversification that emerges – in which co-evolutionary adaptive radiation plays a relatively small role, and major host-use shifts may more often denote times of ecological calamity than opportunity – appears at least to be consistent across approaches. But until more powerful comparative methods have been developed, it will be difficult to make stronger conclusions. We were also limited by crude and indirect proxies of the ecological opportunity attached to a novel host association. As we learn more about ecological speciation and community assembly in herbivorous insects, we may find ways of more accurately quantifying historical ecological opportunity.

Initially, the results presented here may appear to contradict those of Hardy and Otto (2014), who found that butterfly lineages that transition more frequently between monophagy and polyphagy tend to speciate more rapidly. This apparent contradiction is especially problematic if the rate of diet-breadth oscillation is used as an index of the overall rate of host-use evolution, or of the odds that a butterfly lineage will gain or lose a novel host association. The apparent incongruence between this study and Hardy and Otto (2014) could occur because our inferences of ancestral host use and diversification dynamics were warped by the shortcomings in our models. Alternatively, it could mean that the rate of transition between monophagy and polyphagy is not a good index of the overall rate of host-use evolution. Major host groups could be mostly gained and lost by generalist lineages; in that case, rapid transitions between monophagy and polyphagy could be linked to rapid speciation, even if major host-use changes tend to depress diversity. The surprising evolutionary patterns we find here could indicate that our approach was inadequate, or could reflect that the evolutionary history of host-use and speciation in butterflies is complex and poorly understood.

We framed our analyses as tests of the Escape and Radiate Hypothesis. In our view this is consistent with its current usage and connotations, but to be sure, those connotations have evolved since Ehrlich and Raven (1964). In fact, its earliest formulations were vague enough so as to make it nigh impossible to falsify. Subsequent work (e.g., Fordyce, 2010; Janz and Nylin, 2008; Winkler et al., 2009) sharpened its predictions while also blurring its attributions. Hence, although our findings do not support the concept that major host colonization events spark adaptive radiation of herbivorous insects, one could argue that this is within the realm of expectations under the Escape and Radiate Hypothesis. In that case, however, the Escape and Radiate Hypothesis might not be very useful for understanding macro-evolutionary patterns of herbivorous insects.

In sum, our tests yield evidence against a key prediction of the Escape and Radiate Hypothesis: that the evolution of major new host associations will tend to boost species diversity of herbivorous insect lineages. We find some evidence that the scope of ecological opportunity afforded by a novel host association governs its effects on diversification dynamics. But since the overall effect of a major new association appears to be negative, it seems that new ecological opportunities function to primarily diminish negative effects on diversification, rather than to inflate positive effects. To be sure, this evidence is conditioned on the adequacy of the models we have used to infer ancestral host use and diversification dynamics, but it is consistent with another recent comparative phylogenetic study which reported that patterns in the networks of evolutionary associations between butterflies and their host plants are inconsistent with co-evolutionary adaptive radiation (Braga et al., 2018). For half a century, the Escape and Radiate Hypothesis has inspired evolutionary ecologists' explanations for the incredible diversity of herbivorous insects. As researchers continue to build our capacity to model herbivorous insect

diversification, we will be able to perform more powerful tests of the Escape and Radiate Hypothesis's predictions. In the meantime, considering there is evidence against it, we may find inspiration and motivation for alternative explanations.

References

Ali, J. G. and A.A. Agrawal, 2012. Specialist Versus Generalist Insect Herbivores and Plant Defense. *Trends in Plant Science* 17, 293–302. https://doi.org/10.1016/j.tplants.2012.02.006

Agrawal, A. A. 2007. Macroevolution of Plant Defense Strategies. *Trends in Ecology and Evolution* 22:103–109.

Bejerano, G., M. Pheasant, I. Makunin, S. Stephen, W.J. Kent, J.S. Mattick, and D. Haussler, 2004. Ultraconserved Elements in the Human Genome. *Science* 304, 1321–1325. https://doi.org/10.1126/science.1098119

Bird, G., C. Kaczvinsky, A. Wilson, and N. B. Hardy. 2019. When Do Herbivorous Insects Compete? A Phylogenetic Meta-analysis. *Ecology Letters* 22:875–883.

Blackman, R. L. and V. F. Eastop, 2017. Taxonomic Issues, in: van Emden, H.F., Harrington, R. (Eds.), *Aphids as Crop Pests*. CABI, Boston, Massachusetts, pp. 1–36. https://doi.org/10.1079/9781780647098.0001

Bossert, S., E. A. Murray, E. A. B. Almeida, S. G. Brady, B. B. Blaimer, and B.N. Danforth, 2019. Combining Transcriptomes and Ultraconserved Elements to Illuminate the Phylogeny of Apidae. *Molecular Phylogenetics and Evolution* 130, 121–131. https://doi.org/10.1016/j.ympev.2018.10.012

Braby, M. F. and J. W. H. Trueman. 2006. Evolution of Larval Host Plant Associations and Adaptive Radiation in Pierid Butterflies. *Journal of Evolutionary Biology* 19:1677–1690.

Braga, M. P., P. R. Guimarães, C. W. Wheat, S. Nylin, and N. Janz. 2018. Unifying Host-Associated Diversification Processes Using Butterfly–Plant Networks. *Nature Communications* 9.

Capella-Gutiérrez, S., J. M. Silla-Martínez, and T. Gabaldón, 2009. trimAl: A Tool for Automated Alignment Trimming in Large-Scale Phylogenetic Analyses. *Bioinformatics* 25, 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Chamberlain, S. A. and E. Szöcs. 2013. taxize: Taxonomic Search and Retrieval in R. *F1000Research* 2.

Emmanuelle, J., G. Gwenaelle, and C.D.A. Armelle, 2010. Evolutionary Lability of a Complex Life Cycle in the Aphid Genus Brachycaudus. *BMC Evolutionary Biology* 10. https://doi.org/10.1186/1471-2148-10-295

Ehrlich, P. R. and P. H. Raven. 1964. Butterflies and Plants: A Study in Coevolution. *Evolution* 18:586–608.

Faircloth, B.C., 2016. PHYLUCE is a Software Package for the Analysis of Conserved Genomic Loci. *Bioinformatics* 32, 786–788. https://doi.org/10.1093/bioinformatics/btv646

Faircloth, B. C., 2017. Identifying Conserved Genomic Elements and Designing Universal Bait Sets to Enrich Them. *Methods in Ecology and Evolution* https://doi.org/10.1111/2041-210X.12754

Farrell, B. D., D. E. Dussourd, and C. Mitter. 1991. Escalation of Plant Defense: Do Latex and Resin Canals Spur Plant Diversification? *The American Naturalist* 138.

Farrell, B. D. and C. Mitter. 1994. Adaptive Radiation in Insects and Plants: Time and Opportunity. *American Zoologist* 34:57–69.

Fordyce, J. A. 2010. Host Shifts and Evolutionary Radiations of Butterflies. *Proceedings of the Royal Society B: Biological Sciences* 277:3735–3743.

Forister, M. L., V. Novotny, A. K. Panorska, L. Baje, Y. Basset, P. T. Butterill, L. Cizek, P. D. Coley, F. Dem, I. R. Diniz, P. Drozd, M. Fox, A. E. Glassmire, R. Hazen, J. Hrcek, J. P.

Jahner, O. Kaman, T. J. Kozubowski, T. A. Kursar, O. T. Lewis, J. Lill, R. J. Marquis, S. E. Miller, H. C. Morais, M. Murakami, H. Nickel, P. Nicholas, R.E. Ricklefs, M. S. Singer, A. M. Smilanich, J. O. Stireman, S. Villamarín-Cortez, S. Vodka, M. Volf, D. L. Wagner, T. Walla, G. D. Weiblen, and L.A. Dyer, 2015. The Global Distribution of Diet Breadth in Insect Herbivores. *PNAS* 112 (2), 442–447. https://doi.org/10.1073/pnas.1423042112

Futuyma, D. J. and A. A. Agrawal. 2009. Macroevolution and the Biological Diversity of Plants and Herbivores. *PNAS* 106:18054–18061.

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 C. Mitter. 1996. Insect-Plant Interactions: The Evolution of Component Communities. *Philosophical Transactions of the Royal Society B: Biological Sciences* 351.

Gavrilets, S. and J. B. Losos. 2009. Adaptive Radiation: Contrasting Theory with Data. *Science* 323:732–738.

Gilleland, E. and R. W. Katz. 2016. extRemes 2.0: An Extreme Value Analysis Package in R. *Journal of Statistical Software* 72:1–31.

Goldberg, E., L. T. Lancaster, and R. H. Ree. 2011. Phylogenetic Inference of Reciprocal Effects between Geographic Range Evolution and Diversification. *Systematic Biology* 60:451–465.

Hadfield, J. 2010. MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. *Journal of Statistical Software* 33:1–22.

Hamm, C. A. and J. A. Fordyce. 2015. Patterns of Host Plant Utilization and Diversification in the Brush-Footed Butterflies. *Evolution* 69:589–601.

Hardy, N. B. 2017. Do Plant-Eating Insect Lineages Pass through Phases of Host-Use Generalism during Speciation and Host Switching? Phylogenetic Evidence. *Evolution* 71:2100–2109.

Hardy, N. B., 2018. The Biodiversity of Sternorrhyncha: Scale Insects, Aphids, Psyllids, and Whiteflies, in: Foottit, R.G., H. Adler, P. (Eds.), Insect Biodiversity: Science and Society, II. John Wiley & Sons, Hoboken, New Jersey, pp. 591–626. https://doi.org/10.1002/9781118945582.ch20

Hardy, N. B., P. J. Gullan, and C. J. Hodgson. 2008. A Subfamily-Level Classification of Mealybugs (Hemiptera: Pseudococcidae) Based on Integrated Molecular and Morphological Data. *Systematic Entomology* 33:51–71.

Hardy, N. B. and S. P. Otto. 2014. Specialization and Generalization in the Diversification of Phytophagous Insects: Tests of the Musical Chairs and Oscillation Hypotheses. *Proceedings of the Royal Society B: Biological Sciences* 281.

Hardy, N. B., D. A. Peterson, and C. D. von Dohlen, 2015. The Evolution of Life Cycle Complexity in Aphids: Ecological Optimization or Historical Constraint? *Evolution* 69, 1423–1432. https://doi.org/10.1111/evo.12643

Harmon, L. J., J. T. Weir, C. D. Brock, R. E. Glor, and C. Wendell. 2008. GEIGER: Investigating Evolutionary Radiations. *Bioinformatics* 24:129–131.

Janz, N. 2011. Ehrlich and Raven Revisited: Mechanisms Underlying Codiversification of Plants and Enemies. *Annual Review of Ecology, Evolution, and Systematics*42:71–89.

Janz, N., K. Nyblom, and S. Nylin. 2001. Evolutionary Dynamic of Host-Plant Specialization: A Case Study of the Tribe Nymphalini. Evolution (N. Y). 55:783–796.

Janz, N. and S. Nylin. 2008. The Oscillation Hypothesis of Host-Plant Range and Speciation. In Kelley Tillman (eds.) *Specialization, Speciation, and Radiation: The Evolutionary Biology of Herbivorous Insects* Berkeley, CA : University of California Press, 203–215.

Jousselin, E., A. Cruaud, G. Genson, F. Chevenet, R.G. Foottit, and A. Cœur d'acier, 2013. Is Ecological Speciation a Major Trend in Aphids? Insights from a Molecular Phylogeny of the Conifer-Feeding Genus Cinara. *Frontiers in Zoology* 10. https://doi.org/10.1186/1742-9994-10-56

Katoh, K. and D. M. Standley, 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* 30, 772–780. https://doi.org/10.1093/molbev/mst010

Kenis, M., M. A. Auger-Rozenberg, A. Roques, L. Timms, C. Péré, M. J. W. Cock, J. Settele, S. Augustin, and C. Lopez-Vaamonde. 2009. Ecological Effects of Invasive Alien Insects. *Biological Invasions* 11:21–45.

Kergoat, J., A. Delobel, M. Tuda, and K. Anton. 2007. Defining the Limits of Taxonomic Conservatism in Host – Plant Use for Phytophagous Insects: Molecular Systematics and Evolution of Host – Plant Associations in the Seed-Beetle Genus Bruchus Linnaeus (Coleoptera: Chrysomelidae: Bruchinae). *Molecular Phylogenetics and Evolution* 43:251–269.

Kumar, S., G. Stecher, M. Suleski, and S. B. Hedges. 2017. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Molecular Biology and Evolution* 34(7):1812-1819, doi: 10.1093/molbev/msx116.

Liscombe, D. K., B. P. Macleod, N. Loukanina, O. I. Nandi, and P. J. Facchini. 2005. Evidence for the Monophyletic Evolution of Benzylisoquinoline Alkaloid Biosynthesis in Angiosperms. *Molecular Biology and Evolution* 66:1374–1393.

Losos, J. B., and D. L. Mahler. 2010. Adaptive Radiation: The Interaction of Ecological Opportunity, Adaptation, and Speciation. Pp. 381–420 *in* M.A. Bell, D. J. Futuyma, W.F. Eanes, and J.S. Levinton, M.A. Bell, D.J. Futuyma, W.F. Eanes, ed. *Evolution Since Darwin: The First 150 Years*. Sinauer Associates.

Maddison, W. P., P. E. Midford, and S. P. Otto. 2007. Estimating a Binary Character's Effect on Speciation and Extinction. *Systematic Biology* 56:701–710.

Massana, K. A., J. M. Beaulieu, N. J. Matzke, and B. C. O'Meara. 2015. Non-null Effects of the Null Range in Biogeographic Models: Exploring parameter Estimation in the DEC Model. *bioRxiv*, doi: 10.1101/026914.

Matzke, N. J. 2013. BioGeoBEARS: BioGeography with Bayesian (and Likelihood) Evolutionary Analysis in R Scripts. R Package version 0.2, doi: 10.1016/j.bbcan.2007.10.004.

Matzke, N. J. 2014. Model Selection in Historical Biogeography Reveals that Founder-Event Speciation Is a Crucial Process in Island Clades. *Systematic Biolgoy* 63:951–970.

Moran, N. A., M. E. Kaplan, M. J. Gelsey, T. G. Murphy, and E. A. Scholes, 2002. Phylogenetics and Evolution of the Aphid Genus *Uroleucon* Based on Mitochondrial and Nuclear DNA Sequences. *Systematic Entomology* 24, 85–93. https://doi.org/10.1046/j.1365-3113.1999.00076.x

Normark, B. B., 2000. Molecular Systematics and Evolution of the Aphid Family Lachnidae. *Molecular Phylogenetics and Evolution* 14. https://doi.org/10.1006/mpev.1999.0699

Nováková, E., V. Hypša, J. Klein, R. G. Foottit, C. D. von Dohlen, and N. A. Moran, 2013. Reconstructing the Phylogeny of Aphids (Hemiptera: Aphididae) Using DNA of the Obligate Symbiont *Buchnera aphidicola*. *Molecular Phylogenetics and Evolution* 68, 42–54. https://doi.org/10.1016/j.ympev.2013.03.016

Ortiz-Rivas, B. and D. Martínez-Torres, 2010. Combination of Molecular Data Support the Existence of Three Main Lineages in the Phylogeny of Aphids (Hemiptera: Aphididae) and the Basal Position of the Subfamily Lachninae. *Molecular Phylogenetics and Evolution* 55, 305–317. https://doi.org/10.1016/j.ympev.2009.12.005

Ortiz-Rivas, B., A. Moya, D. Martínez-Torres, 2004. Molecular Systematics of Aphids (Homoptera: Aphididae): New Insights from the Long-Wavelength Opsin Gene. *Molecular Phylogenetics and Evolution* 30, 24–37. https://doi.org/10.1016/S1055-7903(03)00155-6

Pike, N., J. A. Whitfield, W. A. Foster, 2007. Ecological Correlates of Sociality in *Pemphigus* Aphids, with a Partial Phylogeny of the Genus. *BMC Evolutionary Biology* 7. https://doi.org/10.1186/1471-2148-7-185

Pybus, O. G. and P. H. Harvey. 2000. Testing Macro-Evolutionary Models Using Incomplete Molecular Phylogenies. *Proceedings of the Royal Society B: Biological Sciences* 267:2267–2272.

Revell, L. J. 2012. phytools: An R Package for Phylogenetic Comparative Biology (and Other Things). *Methods in Ecology and Evolution* 3:217–223.

Roskov Y., L. Abucay, T. Orrell, D. Nicolson, N. Bailly, P. M. Kirk, T. Bourgoin, R. E. DeWalt W. Decock, A. De Wever, E. Nieukerken, and J. P. L. van Zarucchi 2018. *Species 2000 & ITIS Catalogue of Life, 2018 Annual Checklist*.

Schluter, D. 2000. The Ecology of Adaptive Radiation. Oxford University Press, Oxford.

Segraves, K. A., and T. J. Anneberg. 2016. Species Interactions and Plant Polyploidy. *American Journal of Botany* 103:1326–1335.

Smith, B. T., M. G. Harvey, B. C. Faircloth, T. C. Glenn, and R. T. Brumfield, 2014. Target Capture and Massively Parallel Sequencing of Ultraconserved Elements for Comparative Studies at Shallow Evolutionary Time Scales. *Systematic Biology* 63, 83–95. https://doi.org/10.1093/sysbio/syt061.

Stamatakis, A., 2014. RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics* 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Starrett, J., S. Derkarabetian, M. Hedin, R. W. Bryson, J. E. McCormack, and B. C. Faircloth, 2017. High Phylogenetic Utility of an Ultraconserved Element Probe Set Designed for Arachnida. *Molecular Ecology Resources* 17, 812–823. https://doi.org/10.1111/1755-0998.12621

Stern, D. L., 1998. Phylogeny of the Tribe Cerataphidini (Homoptera) and the Evolution of the Horned Soldier Aphids. *Evolution* 52, 155–165. https://doi.org/10.2307/2410930

R Core Team. 2019, R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*, doi: 10.1007/978-3-540-74686-7.

Thompson, J. N. 1989. Concepts of Coevolution. *TREE* 4:179–183.

Völkl, W., J. Woodring, M. Fischer, M. W. Lorenz, and K. H. Hoffmann, 1999. Ant-Aphid Mutualisms: The Impact of Honeydew Production and Honeydew Sugar Composition on Ant Preferences. *Oecologia* 118, 483–491. https://doi.org/10.1007/s004420050751

Von Dohlen, C.D., U. Kurosu, and S. Aoki, 2002. Phylogenetics and Evolution of the Eastern Asian-Eastern North American Disjunct Aphid Tribe, Hormaphidini (Hemiptera: Aphididae). *Molecular Phylogenetics and Evolution* 23, 257–267. https://doi.org/10.1016/S1055-7903(02)00025-8

Von Dohlen, C. D. and N. A. Moran, 2000. Molecular Data Support a Rapid Radiation of Aphids in the Cretaceous and Multiple Origins of Host Alternation. *Biological Journal of the Linnean Society* 71, 689–717. https://doi.org/10.1006/bijl.2000.0470

Von Dohlen, C. D., Rowe, C. A., and O. E. Heie, 2006. A Test of Morphological Hypotheses for Tribal and Subtribal Relationships of Aphidinae (Insecta: Hemiptera: Aphididae) Using DNA Sequences. *Molecular Phylogenetics and Evolution* 38, 316–329. https://doi.org/10.1016/j.ympev.2005.04.035

Von Dohlen, C. D. and D. A. J. Teulon, 2003. Phylogeny and Historical Biogeography of New Zealand Indigenous Aphidini Aphids (Hemiptera, Aphididae): An Hypothesis. *Annals of the Entomological Society of America* 96, 107–116. https://doi.org/10.1603/0013-8746(2003)096[0107:pahbon]2.0.co;2

Wahlberg, N. 2006. That Awkward Age for Butterflies: Insights from the Age of the Butterfly Subfamily Nymphalinae (Lepidoptera: Nymphalidae). *Systematic Biolgoy* 55:703–714.

Ward, S. A., S. R. Leather, J. Pickup, and R. Harrington. 1998. Mortality During Dispersal and the Cost of Host-Specificity in Parasites: How Many Aphids Find Hosts? *Journal of Animal Ecology* 67, 763-773.

Wheat, C. W., H. Vogel, U. Wittstock, M. F. Braby, D. Underwood, and T. Mitchell-Olds. 2007. The Genetic Basis of a Plant Insect Coevolutionary Key Innovation. *PNAS* 104:20427–20431.

Wink, M. 2003. Evolution of Secondary Metabolites from an Ecological and Molecular Phylogenetic Perspective. *Phytochemistry* 64:3–19.

Wink, M., and G. I. A. Mohamed. 2003. Evolution of Chemical Defense Traits in the Leguminosae: Mapping of Distribution Patterns of Secondary Metabolites on a Molecular Phylogeny Inferred from Nucleotide Sequences of the rbc L gene. *Biochemical Systematics and Ecology* 31:897–917.

Winkler, I. S., C. Mitter, and S. J. Scheffer. 2009. Repeated Climate-Linked Host Shifts Have Promoted Diversification in a Temperate Clade of Leaf-Mining Flies. *PNAS* 106:18103–18108.

Yang, Z. X., X. M. Chen, N. P. Havill, Y. Feng, H. Chen, 2010. Phylogeny of Rhus Gall Aphids (Hemiptera: Pemphigidae) Based on Combined Molecular Analysis of Nuclear EF1α and Mitochondrial COII Genes. *Entomological Science* 13, 351–357. https://doi.org/10.1111/j.1479-8298.2010.00391.x

Yoder, J. B., E. Clancey, S. Des Roches, J. M. Eastman, L. Gentry, W. Godsoe, T. J. Hagey, D. Jochimsen, B. P. Oswald, J. Robertson, B. A. J. Sarver, J. J. Schenk, S. F. Spear, and L. J. Harmon. 2010. Ecological Opportunity and the Origin of Adaptive Radiations. *Journal of Evolutionary Biology* 23:1581–1596.

Zerbino, D. R. and E., Birney, 2008. Velvet: Algorithms for De Novo Short Read Assembly Using de Bruijn Graphs. *Genome Research* 18, 821–829. https://doi.org/10.1101/gr.074492.107

Appendices

Chapter 3

Appendix 1
#Script that chops a large tree into subtrees with a host family and polyphagy number that makes the DEC estimation tractable

```
library(BioGeoBEARS)
library(phytools)
library(ape)
library(phangorn)

tree <- read.tree('Wahl_Beast_New4.tre')
d <- read.csv('LepMatrix.csv') #this is the host use/non use matrix
rownames(d) <- d$X
d$X <- NULL

sub.trees <- list()

#modify tree so there are no tip without data and the most polyphagous are trimmed
for (t in tree$tip.label){
   if (t %in% rownames(d)) {
      print('Data Found')
   } else {
   tree <- drop.tip(tree, t)
   }
}

#move from the root tipward
#need function to find the root
get.root <- function(tree){
   tips <- length(tree$tip.label)
   root <- tips + 1
   root
}

#tree = name of tree object, node = starting node
#returns a list of tree objects
chop <- function(tree, node) {
   kids <- Children(tree, node)
   hosts <- data.frame()
   for(n in kids){
      hosts <- data.frame()
      #gets all child nodes of n
      kd <- Descendants(tree, n)
```

```
      #need to get just the tips, so all values of kd less than the root
      tips <- kd[[1]][kd[[1]]< root]
      #make a data frame of just the host info for the tree tips
      for (t in tips){
         #print(t)
         name <- tree$tip.label[t]
         #print(name)
          if (is.na(d[name,1])){
             print('NA')
          } else {
         row <- d[name,]
         hosts <- rbind(hosts, row)
          }
      }
      #editing hosts info to make names the rownames and remove from first column
      hosts <- hosts[,-1]
      #find total host #
      #filter hosts to just colums with at least one 1 in them
      num <- colSums(hosts)
      t.hosts <- as.data.frame(num)
      t.hosts <- filter(t.hosts, num > 0)
      total.hosts <- nrow(t.hosts)
      #find max polyphagy
      poly <- rowSums(hosts)
      max.poly <- max(poly)
      #run these through the BioGeoBEARS function that checks matrix size
      #include_null_range is false, because we aren't allowing no host
      matrix <- numstates_from_numareas(numareas=total.hosts, maxareas=max.poly,
include_null_range=FALSE)
      if (matrix < 1600 & n > root) {
      ntree <- extract.clade(tree, n)
      write.tree(ntree, file = 'SubTrees.tre', append = TRUE)
      print('Yay! Tree!')
         #chop off tree at this node and add it to a list of phylo objects
      } else {
         #call chop with n as the node to start with
         chop(tree, n)
      }
   }
}

root <- get.root(tree)
trees <- chop(tree, root)
```

Appendix 2 a

```
#the code for the exclusive calculation of extant diversity and other variables
#note: results from different trees (created by Appendix 1) will need to be appended together
before finding the corresponding gamma stat and volatility metrics and getting final model
results

library(ape)
library(geiger)
library(TreeSim)
library(phytools)
library(phangorn)
library(taxize)


##Load data#############################################################################
#note: slight name change to make file name more informative
family.host.use <- read.csv('LepMatrix_Families.csv') #this is the host use/non use matrix
rownames(family.host.use) <- family.host.use$X
family.host.use$X <- NULL

order.host.use <- read.csv('LepMatrix_Orders.csv')
rownames(order.host.use) <- order.host.use$X
order.host.use$X <- NULL

e <- read.csv('Nym_covariates_2.csv') #genus names are under X, diversities under Nsp
rownames(e) <- e$X
e$X <- NULL

z <- read.csv('Fam_Ages.csv') #order/family under fams, ages under ages
z$X <- NULL
rownames(z) <- z$fams
z$fams <- NULL

data <- read.table('range_probabilities_star_T9.txt')

tree <- read.tree('./Trees_tre/T9.tre')

key <- read.csv('Tree_Key.csv')

#load a reference tree and match it to the main tree to get unfiltered diversity counts
ref <- read.tree('Wahl_Beast_New4.tre')
match <- as.data.frame(matchLabels(ref,tree))
#######################################################################################

##Define Functions#####################################################################
```

```
#need function to find the root
get.root <- function(arg1){
   #get total number of tips
   tips <- length(arg1$tip.label)
   #the first node after the tips is the root
   root <- tips + 1
   root
}

#steps to get a colid from Catologue of Life and then corresponding downstream diversity counts
get.diversity <- function(arg1) {
   #get colid to make calling the diversity faster
   col.id <- get_colid(arg1, kingdom='Plantae')
   #run through catalog of life database to get all extant species, then add up the number of
species for a diversity count
   response <- downstream(col.id, downto = 'species', db = 'col', intermediate=F)
   #brackets weird formating stuff to actually get down to the data we want (because output of
downstream is lists of lists of lists of data frames)
   spp.count <- length(response[[1]][[1]])
}

#inputs the code for an order and returns the full order name from Tree_Key
what.host <- function(arg1){
   code <- key[key$V2==arg1,]
   arg <- code['V1']
   ht <- as.character(arg[,'V1'])
   return(ht)
}

#functionally the inverse of what.host, inputs a host and gets the corresponding key
what.key <- function(arg1){
   host <- key[key$V1==arg1,]
   k <- host['V2']
   key <- as.character(k[,'V2'])
   return(key)
}

#a function to get the focal sister diversity from the original tree (rather than the pruned one) and
include only the genera that haven't gained the focal host at some later date (for a more precise
comparison of diversities)

##include them only if they include the host
get.div <- function(node.number, host, node.status){
      sis.tips<- numeric()
      foc.tips<- numeric()
```

```
    tip <- ''
    host.list <- character()
    genus <- ''
    genus.wout.host <- character() #this variable name doesn't make sense anymore
    genus.w.host <- character() # this variable name doesn't make sense anymrore
if (node.number == root) {
    return(0)
} else {
    #gets the tips for the focal group
    kids <- Descendants(tree, node.number)
    #gets the sister node
    sis <- Siblings(tree, node.number, include.self=FALSE)
    #and the sister tips
    nieces <- Descendants(tree, sis)
    tips <- c(kids[[1]][1], nieces[[1]][1])
    # select a tip from each
    #this will generate a list of all the tips on tree with the corresponding labels
    labels <- tree$tip.label
    #this gives the name of the tip
    t <- labels[tips[1]]
    #this gives the row with the corresponding labels
    tip <- match[t,]
    #select the value row for the ref tree
    ft <- tip[,'tr1']
    #this gives the corresponding sister tip from the main tree
    s <- labels[tips[2]]
    stip <- match[s,]
    st <- stip[,'tr1']
    mtips <- c(ft, st)
    # find the MRCA from the two tips
    mrca <- findMRCA(ref, tips=mtips, type="node")
    #get the two child nodes
    groups <- Children(ref, mrca)
    #select the sister node
    test <- Descendants(ref, groups[1])
    test2 <- Descendants(ref, groups[2])
    if (ft %in% test[[1]]){
        foc <- groups[1]
        sister <- groups[2]
    } else if (ft %in% test2[[1]]) {
        foc <- groups[2]
        sister <- groups[1]
    } else {
        print('You have a problem.')
        }
```

```
      #get the Diversity of the focal node and the sister node
      foc.tips <- unlist(Descendants(ref, foc, type='tips'))
      sis.tips <- unlist(Descendants(ref, sister, type='tips'))
      for (t in foc.tips){
        #gets the genus name for each tip for later reference, will work with tip number for the
next bit
        genus <- ref$tip.label[t]
        #get the row for the genus of the tip in the host use matrix
        host.info <- family.host.use[genus,]
        host.info[is.na(host.info)] <- 0
        #find the hosts that aren't 0, i.e. those that use the host (matrix is binary)
        host.use <- which(host.info != 0)
        host.taxa <- names(host.info)[host.use]
        family.key <- key[key$V1 %in% host.taxa,]
        host.list <- as.character(unlist(family.key$V2))
        if (node.status == 'gain'){
            print(host)
            print(node.status)
          if (host %in% host.list){
            genus.w.host <- c(genus.w.host, genus)
          }
        } else if (node.status == 'loss'){
          if (host %ni% host.list){

            genus.w.host <- c(genus.w.host, genus)
          }
        } else {
          genus.w.host <- c(genus.w.host, genus)
        }
      }
      for (s in sis.tips){
        #gets the genus name for each tip for later reference, will work with tip number for the
next bit
        genus <- ref$tip.label[s]
        host.info <- family.host.use[genus,]
        host.info[is.na(host.info)] <- 0
        #find the hosts that aren't 0, i.e. those that use the host (matrix is binary)
        host.use <- which(host.info != 0)
        if (length(host.use) > 0){
          host.taxa <- names(host.info)[host.use]
          order.key <- key[key$V1 %in% host.taxa,]
          host.list <- as.character(unlist(order.key$V2))
          if (node.status == 'gain'){
            if (host %ni% host.list){
              genus.wout.host <- c(genus.wout.host, genus)
            }
```

```
            if (node.status == 'loss'){
               if (host %in% host.list){
                  genus.wout.host <- c(genus.wout.host, genus)
               }
            }
            } else {
               genus.wout.host <- c(genus.wout.host, genus)
            }
         }
      }

      #get species counts for genera that use host
      #reset sp. count to clear any data from an earlier function call
      focal.sp.count <- 0
      sister.sp.count <- 0
      #loop through the genera and get the species in each genus from our data on diversity
      for (g in genus.w.host){
         #gets species count from our data
         genus.sp.count <- e[g,]$Nsp
         #adds the species in the new genus to the total number of species
         if (!is.na(genus.sp.count)){
            focal.sp.count <- focal.sp.count + genus.sp.count
         }
      }
      for (g in genus.wout.host){
         #gets species count from our data
         genus.sp.count <- e[g,]$Nsp
         if (!is.na(genus.sp.count)){
            sister.sp.count <- sister.sp.count + genus.sp.count
         }
      }
   focal.divs <- focal.sp.count
   sis.divs <- sister.sp.count
   return(c(focal.divs, sis.divs))
   }
}

get.waiting.time <- function(node) {
   kids <- Children(tree, node)
   #check not a tip, if it is set wait.time to 0
   if (length(kids) == 0) {
      wait.time <- 0
      grand.wt <- 0
   } else {
      wt1 <- which(tree$edge[,2] == kids[1])
      w1 <- tree$edge.length[wt1]
```

```
      wt2 <- which(tree$edge[,2] == kids[2])
      w2 <- tree$edge.length[wt2]
      wait.time <- (w1 + w2)
      grandkids <- Children(tree, kids[1])
      gkids <- Children(tree, kids[2])
      grandkids <- c(grandkids, gkids)
      g.test <- length(grandkids)
      if (g.test == 0) {
         grand.wt <- 0
      } else if (g.test == 2){
         grand.wt <- 0
      } else {
         gwt1 <- which(tree$edge[,2] == grandkids[1])
         g1 <- tree$edge.length[gwt1]
         gwt2 <- which(tree$edge[,2] == grandkids[2])
         g2 <- tree$edge.length[gwt2]
         gwt3 <- which(tree$edge[,2] == grandkids[3])
         g3 <- tree$edge.length[gwt3]
         gwt4 <- which(tree$edge[,2] == grandkids[4])
         g4 <- tree$edge.length[gwt4]
         g.wt <- (g1 + g2 + g3 + g4)
         grand.wt <- (wait.time + g.wt)
      }
   }
   return(c(wait.time, grand.wt))
}

################################################################################

##Proceedures#####################################################################

host <- ''
gain.loss <- ''
wait.time <- 0
host <- ''
kids <- 0
mom <- 0
mom.host <- ''
sis.wait.time <- 0
focal.divs <- 0
sis.divs <- 0
early.bird <- 0
mom.hosts <- ''
host.age <- 0
grand.wt <- 0
sis.grand.wt <- 0
```

```
j.age <- 0

bts <- branching.times(tree)

output <- data.frame(Node=numeric(), Host.All=character(), Host.Name=character(), Host.Divs
=numeric(), Host.Age = numeric(), Focal.Divs=numeric(), Sis.Divs= numeric(),
Mom.Host.All=character(), Mom.Host=character(), Status=character(), Wait.Time=numeric(),
Sis.Wait.Time=numeric(), Grand.Wait.Time=numeric(), Sis.Grand.Wt=numeric(),
Early.Bird=numeric(), Node.Age=numeric(), Est.Prob=numeric())

`%ni%` = Negate(`%in%`)

root <- get.root(tree)

for (r in 1:nrow(data)) {
   print(r)
   if (r != root){
      row <- data[r,]
      host <- which.max(row)
      ##just have to pull a new column with the probability for the which.max host
      prob <- data[r,host]
      host1 <- colnames(data)[host]
      hosts <- strsplit(host1, "(?<=.{2})", perl=TRUE)
      hosts <- unlist(hosts)
      j.age <- bts[as.character(r)]
      names(j.age) <- NULL
      #because tips will return a NA
      if (is.na(j.age)){
         j.age <- 0
      }
      mom <- Ancestors(tree, r, type='parent')
      mom.row <- data[mom,]
      mom.host1 <- which.max(mom.row)
      mom.host.data <- colnames(data)[mom.host1]
      #make sure it isn't the root node
      mom.hosts <- character()
      if (length(mom.host1) ==! 0) {
         mom.host.data <- colnames(data)[mom.host1]
         mom.hosts <- strsplit(mom.host.data, "(?<=.{2})", perl=TRUE)
      } else {
         mom.host <- 'root'
         mom.hosts <- 'root'
         }
      mom.hosts <- unlist(mom.hosts)
      ##get sister node
      sis <- Siblings(tree, r, include.self=FALSE)
```

```
#get the waiting times
foc.waiting.time <-get.waiting.time(r)
wait.time <- foc.waiting.time[1]
grand.wt <- foc.waiting.time[2]
sis.waiting.time <- get.waiting.time(sis)
sis.wait.time <- sis.waiting.time[1]
sis.grand.wt <- sis.waiting.time[2]

event.count <- 0
none.count <- 0
gain.loss <- 'none'
##First do gains
for (h in hosts){
   if (h %ni% mom.hosts){
      early.bird <- 0
      gain.loss <- 'gain'
      event.count <- event.count + 1
      name <- what.host(h)
      host.age <- z[name,]$ages
      divs <- z[name,]$fdiv
      if (r < root) {
         early.bird <- host.age
      } else {
         early.bird <- host.age - j.age
      }
      diversities <- get.div(r, h, gain.loss)
      focal.divs <- diversities[1]
      sis.divs <- diversities[2]
      outy <- data.frame (Node= r, Host.All=host1, Host.Name=h, Host.Divs=divs,
Host.Age = host.age, Focal.Divs = focal.divs, Sis.Divs= sis.divs, Mom.Host.All=mom.host.data,
Mom.Host=NA, Status=gain.loss, Wait.Time=wait.time, Sis.Wait.Time=sis.wait.time,
Grand.Wait.Time=grand.wt, Sis.Grand.Wt=sis.grand.wt, Early.Bird=early.bird, Node.Age=j.age,
Est.Prob=prob)
      output <- rbind(output, outy)
   }
}
##Next do losses
for (m in mom.hosts){
   if (m %ni% hosts){
      gain.loss <- 'loss'
      event.count <- event.count + 1
      diversities <- get.div(r, m, gain.loss) ##IMPORTANT: Pass the lost host name not the
last value for h
      focal.divs <- diversities[1]
      sis.divs <- diversities[2]
```

```
            outy <- data.frame (Node= r, Host.All=host1, Host.Name=NA, Host.Divs=divs,
Host.Age = host.age, Focal.Divs = focal.divs, Sis.Divs= sis.divs, Mom.Host.All=mom.host.data,
Mom.Host=m, Status=gain.loss, Wait.Time=wait.time, Sis.Wait.Time=sis.wait.time,
Grand.Wait.Time=grand.wt, Sis.Grand.Wt=sis.grand.wt, Early.Bird=early.bird, Node.Age=j.age,
Est.Prob=prob)
            output <- rbind(output, outy)
          }
      }
      ##Last do the non-events (just once per node -- had been repeating them before)
      if (event.count == 0){
            h <- NA
            divs <- NA
            host.age <- NA
            mom.host.data <- NA
            m <- NA
            diversities <- get.div(r, h, gain.loss)
            focal.divs <- diversities[1]
            sis.divs <- diversities[2]
            outy <- data.frame (Node= r, Host.All=host1, Host.Name=h, Host.Divs=divs,
Host.Age = host.age, Focal.Divs = focal.divs, Sis.Divs= sis.divs, Mom.Host.All=mom.host.data,
Mom.Host=m, Status=gain.loss, Wait.Time=wait.time, Sis.Wait.Time=sis.wait.time,
Grand.Wait.Time=grand.wt, Sis.Grand.Wt=sis.grand.wt, Early.Bird=early.bird, Node.Age=j.age,
Est.Prob=prob)
            output <- rbind(output, outy)
      }
   event.count <- 0
   }
}


write.csv(output, file = 'GL_exclusive_T9.csv', row.names = FALSE)
```

Appendix 2b

```
#the code for the inclusive calculation of extant diversity and other variables
#note: results from different trees (created by Appendix 1) will need to be appended together
before finding the corresponding gamma stat and volatility metrics and getting final model result

library(ape)
library(geiger)
library(TreeSim)
library(phytools)
library(phangorn)
library(taxize)
```

```
##Load data#############################################################################
family.host.use <- read.csv('LepMatrix_Families.csv') #this is the host use/non use matrix
rownames(family.host.use) <- family.host.use$X
family.host.use$X <- NULL

order.host.use <- read.csv('LepMatrix_Orders.csv')#host use/non-use matrix for orders
rownames(order.host.use) <- order.host.use$X
order.host.use$X <- NULL

e <- read.csv('Nym_covariates_2.csv') #genus names are under X, diversities under Nsp
rownames(e) <- e$X
e$X <- NULL

#list of the ages of host families
z <- read.csv('Fam_Ages.csv') #order under fams, ages under ages
z$X <- NULL
rownames(z) <- z$fams
z$fams <- NULL

data <- read.table('range_probabilities_star_T9.txt')

tree <- read.tree('./Trees_tre/T9.tre')

key <- read.csv('Tree_Key.csv')

#load a reference tree and match it to the main tree to get unfiltered diversity counts
ref <- read.tree('Wahl_Beast_New4.tre')
match <- as.data.frame(matchLabels(ref,tree))
#######################################################################################

##Define Functions#####################################################################

#need function to find the root
get.root <- function(arg1){
   #get total number of tips
   tips <- length(arg1$tip.label)
   #the first node after the tips is the root
   root <- tips + 1
   root
}

#steps to get a colid from Catologue of Life and then corresponding downstream diversity counts
get.diversity <- function(arg1) {
   #get colid to make calling the diversity faster
   col.id <- get_colid(arg1, kingdom='Plantae')
```

```
   #run through catalog of life database to get all extant species, then add up the number of
species for a diversity count
   response <- downstream(col.id, downto = 'species', db = 'col', intermediate=F)
   #brackets weird formating stuff to actually get down to the data we want (because output of
downstream is lists of lists of lists of data frames)
   spp.count <- length(response[[1]][[1]])
}

#inputs the code for an order and returns the full order name from Tree_Key
what.host <- function(arg1){
   code <- key[key$V2==arg1,]
   arg <- code['V1']
   ht <- as.character(arg[,'V1'])
   return(ht)
}

#functionally the inverse of what.host, inputs a host and gets the corresponding key
what.key <- function(arg1){
   host <- key[key$V1==arg1,]
   k <- host['V2']
   key <- as.character(k[,'V2'])
   return(key)
}

#a function to get the focal sister diversity from the original tree (rather than the pruned one) and
include only the genera that haven't gained the focal host at some later date (for a more precise
comparison of diversities)

##include them only if they include the host
get.div <- function(node.number, host, node.status){
      sis.tips<- numeric()
      foc.tips<- numeric()
      tip <- ''
      host.list <- character()
      genus <- ''
      genus.wout.host <- character()
      genus.w.host <- character() #list of genera that use the host
   if (node.number == root) {
      return(0)
   } else {
      #gets the tips for the focal group
      kids <- Descendants(tree, node.number)
      #gets the sister node
      sis <- Siblings(tree, node.number, include.self=FALSE)
      #and the sister tips
      nieces <- Descendants(tree, sis)
```

```
tips <- c(kids[[1]][1], nieces[[1]][1])
# select a tip from each
#this will generate a list of all the tips on tree with the corresponding labels
labels <- tree$tip.label
#this gives the name of the tip
t <- labels[tips[1]]
#this gives the row with the corresponding labels
tip <- match[t,]
#select the value row for the ref tree
ft <- tip[,'tr1']
#this gives the corresponding sister tip from the main tree
s <- labels[tips[2]]
stip <- match[s,]
st <- stip[,'tr1']
mtips <- c(ft, st)
# find the MRCA from the two tips
mrca <- findMRCA(ref, tips=mtips, type="node")
#get the two child nodes
groups <- Children(ref, mrca)
#select the sister node
test <- Descendants(ref, groups[1])
test2 <- Descendants(ref, groups[2])
if (ft %in% test[[1]]){
   foc <- groups[1]
   sister <- groups[2]
} else if (ft %in% test2[[1]]) {
   foc <- groups[2]
   sister <- groups[1]
} else {
   print('You have a problem.')
   }

#get the Diversity of the focal node and the sister node
foc.tips <- unlist(Descendants(ref, foc, type='tips'))
sis.tips <- unlist(Descendants(ref, sister, type='tips'))
for (t in foc.tips){
   #gets the genus name for each tip for later reference, will work with tip number for the
next bit
   genus <- ref$tip.label[t]
   #get the row for the genus of the tip in the host use matrix
   host.info <- family.host.use[genus,]
   host.info[is.na(host.info)] <- 0
   #find the hosts that aren't 0, i.e. those that use the host (matrix is binary)
   host.use <- which(host.info != 0)
   host.taxa <- names(host.info)[host.use]
   order.key <- key[key$V1 %in% host.taxa,]
```

```
        host.list <- as.character(unlist(order.key$V2))
        genus.w.host <- c(genus.w.host, genus)
      }
      for (s in sis.tips){
        #gets the genus name for each tip for later reference, will work with tip number for the
next bit
        genus <- ref$tip.label[s]
        host.info <- family.host.use[genus,]
        host.info[is.na(host.info)] <- 0
        #find the hosts that aren't 0, i.e. those that use the host (matrix is binary)
        host.use <- which(host.info != 0)
        if (length(host.use) > 0){
          host.taxa <- names(host.info)[host.use]
          order.key <- key[key$V1 %in% host.taxa,]
          host.list <- as.character(unlist(order.key$V2))
          genus.wout.host <- c(genus.wout.host, genus)
        }
      }

      #get species counts for genera that use host
      #reset sp. count to clear any data from an earlier function call
      focal.sp.count <- 0
      sister.sp.count <- 0
      #loop through the genera and get the species in each genus from our data on diversity
      for (g in genus.w.host){
        #gets species count from our data
        genus.sp.count <- e[g,]$Nsp
        #adds the species in the new genus to the total number of species
        if (!is.na(genus.sp.count)){
          focal.sp.count <- focal.sp.count + genus.sp.count
        }
      }
      for (g in genus.wout.host){
        #gets species count from our data
        genus.sp.count <- e[g,]$Nsp
        if (!is.na(genus.sp.count)){
          sister.sp.count <- sister.sp.count + genus.sp.count
        }
      }
    focal.divs <- focal.sp.count
    sis.divs <- sister.sp.count
    return(c(focal.divs, sis.divs))
    }
}

get.waiting.time <- function(node) {
```

```
   kids <- Children(tree, node)
   #check not a tip, if it is set wait.time to 0
   if (length(kids) == 0) {
      wait.time <- 0
      grand.wt <- 0
   } else {
      wt1 <- which(tree$edge[,2] == kids[1])
      w1 <- tree$edge.length[wt1]
      wt2 <- which(tree$edge[,2] == kids[2])
      w2 <- tree$edge.length[wt2]
      wait.time <- (w1 + w2)
      grandkids <- Children(tree, kids[1])
      gkids <- Children(tree, kids[2])
      grandkids <- c(grandkids, gkids)
      g.test <- length(grandkids)
      if (g.test == 0) {
         grand.wt <- 0
      } else if (g.test == 2){
         grand.wt <- 0
      } else {
         gwt1 <- which(tree$edge[,2] == grandkids[1])
         g1 <- tree$edge.length[gwt1]
         gwt2 <- which(tree$edge[,2] == grandkids[2])
         g2 <- tree$edge.length[gwt2]
         gwt3 <- which(tree$edge[,2] == grandkids[3])
         g3 <- tree$edge.length[gwt3]
         gwt4 <- which(tree$edge[,2] == grandkids[4])
         g4 <- tree$edge.length[gwt4]
         g.wt <- (g1 + g2 + g3 + g4)
         grand.wt <- (wait.time + g.wt)
      }
   }
   return(c(wait.time, grand.wt))
}

################################################################################

##Proceedures#####################################################################

host <- "
gain.loss <- "
wait.time <- 0
host <- "
kids <- 0
mom <- 0
mom.host <- "
```

```
sis.wait.time <- 0
focal.divs <- 0
sis.divs <- 0
early.bird <- 0
mom.hosts <- ''
host.age <- 0
grand.wt <- 0
sis.grand.wt <- 0
j.age <- 0

bts <- branching.times(tree)

output <- data.frame(Node=numeric(), Host.All=character(), Host.Name=character(), Host.Divs
=numeric(), Host.Age = numeric(), Focal.Divs=numeric(), Sis.Divs= numeric(),
Mom.Host.All=character(), Mom.Host=character(), Status=character(), Wait.Time=numeric(),
Sis.Wait.Time=numeric(), Grand.Wait.Time=numeric(), Sis.Grand.Wt=numeric(),
Early.Bird=numeric(), Node.Age=numeric(), Est.Prob=numeric())

`%ni%` = Negate(`%in%`)


##check the written file, may also want to append family.divs to Fam_Ages.csv

root <- get.root(tree)

for (r in 1:nrow(data)) {
   print(r)
   if (r != root){
      row <- data[r,]
      host <- which.max(row)
      prob <- data[r,host]
      ##just have to pull a new column with the probability for the which.max host
      host1 <- colnames(data)[host]
      hosts <- strsplit(host1, "(?<=.{2})", perl = TRUE)
      hosts <- unlist(hosts)
      j.age <- bts[as.character(r)]
      names(j.age) <- NULL
      #because tips will return a NA
      if (is.na(j.age)){
         j.age <- 0
      }
      mom <- Ancestors(tree, r, type='parent')
      mom.row <- data[mom,]
      mom.host1 <- which.max(mom.row)
      mom.host.data <- colnames(data)[mom.host1]
      #make sure it isn't the root node
```

```
mom.hosts <- character()
if (length(mom.host1) ==! 0) {
   mom.host.data <- colnames(data)[mom.host1]
   mom.hosts <- strsplit(mom.host.data, "(?<=.{2})", perl = TRUE)
} else {
   mom.host <- 'root'
   mom.hosts <- 'root'
   }
mom.hosts <- unlist(mom.hosts)
##get sister node
sis <- Siblings(tree, r, include.self=FALSE)

#get the waiting times
foc.waiting.time <-get.waiting.time(r)
wait.time <- foc.waiting.time[1]
grand.wt <- foc.waiting.time[2]
sis.waiting.time <- get.waiting.time(sis)
sis.wait.time <- sis.waiting.time[1]
sis.grand.wt <- sis.waiting.time[2]

##The structure here was funky. Maybe that was causing poblems?
##I reworked it.
event.count <- 0
none.count <- 0
gain.loss <- 'none'
##First do gains
for (h in hosts){
   if (h %ni% mom.hosts){
      early.bird <- 0
      gain.loss <- 'gain'
      event.count <- event.count + 1
      name <- what.host(h)
      host.age <- z[name,]$ages
      divs <- z[name,]$fdiv
      if (r < root) {
         early.bird <- host.age
      } else {
         early.bird <- host.age - j.age
      }
      diversities <- get.div(r, h, gain.loss)
      focal.divs <- diversities[1]
      sis.divs <- diversities[2]
      outy <- data.frame (Node= r, Host.All=host1, Host.Name=h, Host.Divs=divs,
Host.Age = host.age, Focal.Divs = focal.divs, Sis.Divs= sis.divs, Mom.Host.All=mom.host.data,
Mom.Host=NA, Status=gain.loss, Wait.Time=wait.time, Sis.Wait.Time=sis.wait.time,
```

```
Grand.Wait.Time=grand.wt, Sis.Grand.Wt=sis.grand.wt, Early.Bird=early.bird, Node.Age=j.age,
Est.Prob=prob)
            output <- rbind(output, outy)
         }
      }
      ##Next do losses
      for (m in mom.hosts){
         if (m %ni% hosts){
            gain.loss <- 'loss'
            event.count <- event.count + 1
            diversities <- get.div(r, m, gain.loss) ##IMPORTANT: Pass the lost host name not the
last value for h
            focal.divs <- diversities[1]
            sis.divs <- diversities[2]
            outy <- data.frame (Node= r, Host.All=host1, Host.Name=NA, Host.Divs=divs,
Host.Age = host.age, Focal.Divs = focal.divs, Sis.Divs= sis.divs, Mom.Host.All=mom.host.data,
Mom.Host=m, Status=gain.loss, Wait.Time=wait.time, Sis.Wait.Time=sis.wait.time,
Grand.Wait.Time=grand.wt, Sis.Grand.Wt=sis.grand.wt, Early.Bird=early.bird, Node.Age=j.age,
Est.Prob=prob)
            output <- rbind(output, outy)
         }
      }
      ##Last do the non-events (just once per node -- had been repeating them before)
      if (event.count == 0){
            h <- NA
            divs <- NA
            host.age <- NA
            mom.host.data <- NA
            m <- NA
            diversities <- get.div(r, h, gain.loss)
            focal.divs <- diversities[1]
            sis.divs <- diversities[2]
            outy <- data.frame (Node= r, Host.All=host1, Host.Name=h, Host.Divs=divs,
Host.Age = host.age, Focal.Divs = focal.divs, Sis.Divs= sis.divs, Mom.Host.All=mom.host.data,
Mom.Host=m, Status=gain.loss, Wait.Time=wait.time, Sis.Wait.Time=sis.wait.time,
Grand.Wait.Time=grand.wt, Sis.Grand.Wt=sis.grand.wt, Early.Bird=early.bird, Node.Age=j.age,
Est.Prob=prob)
            output <- rbind(output, outy)
      }
   event.count <- 0
   }
}


write.csv(output, file = 'GL_inclusive_T9.csv', row.names = FALSE)
```

Appendix 3

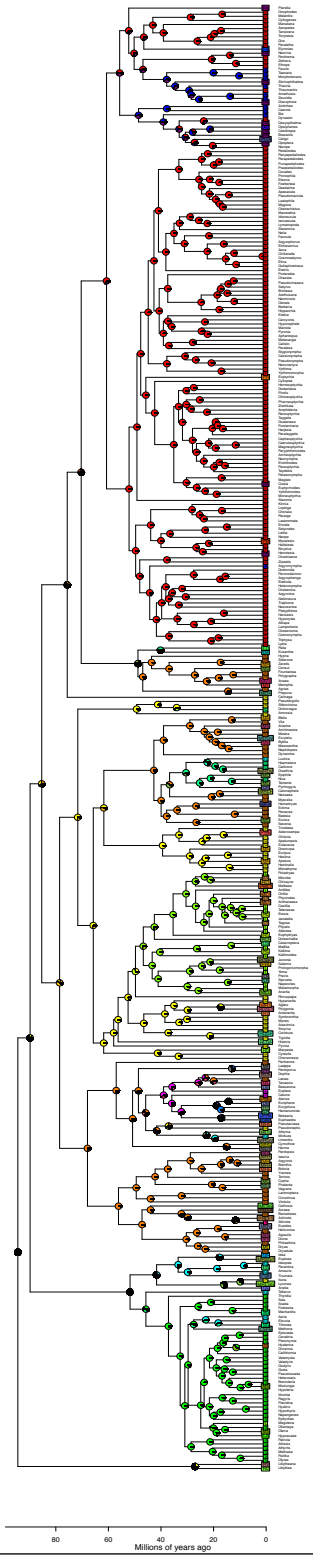
##This is the script to run the final linear models, it needs an input of a data table with information for the response and predictor variables and a phylogeny (in pedigree format) that can be used to control for phylogeny with MCMCglmm

```
library(scales)
library(MCMCglmm)

#this is the phylogeny (in pedigree format) to constrain gamma stat
ped <- read.csv('Ordered_Ped.csv')

#this is the data for all node coverage
d <- read.csv('./Order/Gain_Loss_strict_o_final.csv')


#rename statuses so that 'none' is the reference status
d$Status <- as.character(d$Status)
d$Status[d$Status == 'none'] <- 'a'
d$Status[d$Status == 'gain'] <- 'b'
d$Status[d$Status == 'loss'] <- 'c'
d$Status <- factor(d$Status)

##subset data and set up phylogenetically independent contrasts
#extant diversity contrasts
d$Div.Diff <- d$Focal.Divs - d$Sis.Divs

d$host.vol <- d$host.vol / d$Host.Age

#this subsets events to just gains
div.gain <- subset(d, Status=='b')


#exclude any nodes where the focal or sister is a tip (b/c it will introduce inaccurate values into the model)
wait <- subset(d, Sis.Wait.Time != 0)
wait <- subset(wait, Wait.Time != 0)

#waiting time contrasts
wait$Wait.Diff <- wait$Wait.Time - wait$Sis.Wait.Time

#waiting time contrasts just for gainsge
wait.gain <- subset(wait, Status=='b')


#get gamma data subsetted
```

```
gamma <- subset(d, gamma.stat != 0)
gamma.gain <- subset(gamma, Status=='b')


###Run the three models
#use the none nodes as the reference
#diversity contrasts, node status model
m1 <- lm(d$Div.Diff ~ d$Status)

#waiting time, node status model
m2 <- lm(wait$Wait.Diff ~ wait$Status)


#model looking at diversity contrast and our predictor variables
md <- lm(Div.Diff ~ Early.Bird + host.vol + Host.Age + Host.Divs -1, data=div.gain)

#model looking at waiting time contrast and our predictor variables
mw <- lm(Wait.Diff ~ Early.Bird + host.vol + Host.Age + Host.Divs -1, data=wait.gain)

##Set up pedigree constraint for gamma stat- Bayesian model fit
priorT <- list(R=list(V=1,nu=0.02),G=list(G1=list(V=1,nu=0.02,alpha.V=1000)))

#gamma stat, node status model controlling for phylogeny
m3 <- MCMCglmm(gamma.stat ~ Status, random=~animal, data=gamma, prior=priorT,
pedigree=ped, nit=1000000, thin=100)

#model looking at gamma stats and our predictor variables
mg <- MCMCglmm(gamma.stat ~ Early.Bird + host.vol + Host.Age + Host.Divs,
random=~animal, data=gamma.gain, prior=priorT, pedigree=ped, nit=1000000, thin=100)
```

Appendix 4

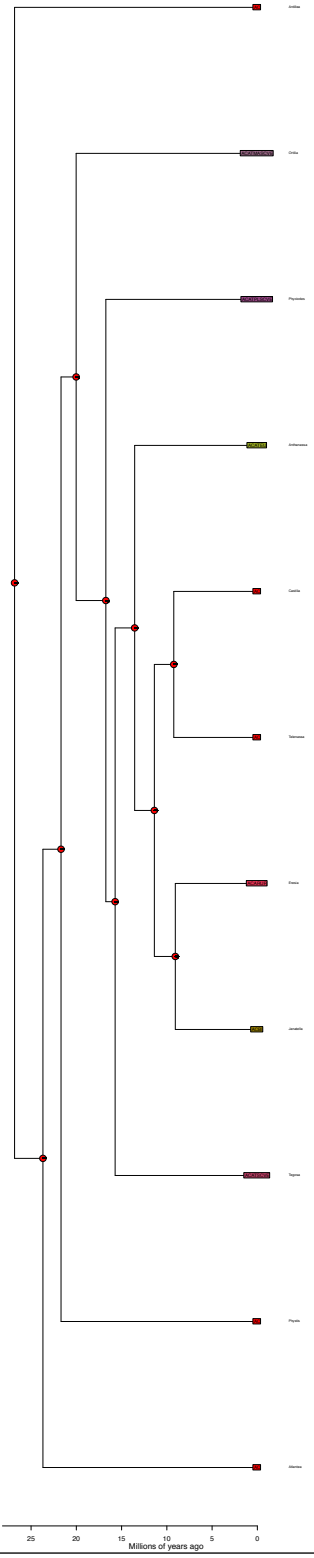
Order-Level DEC* Reconstruction



DEC* Order-Level Reconstruction

Appendix 5

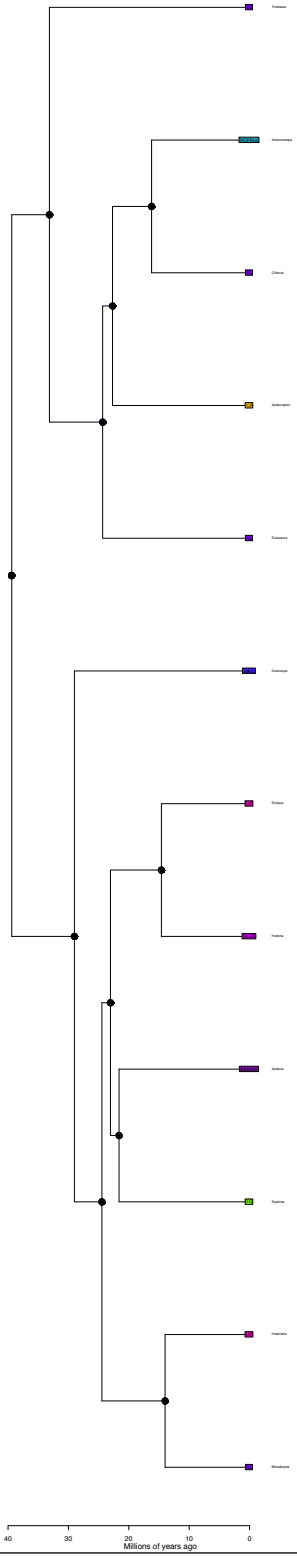
Family-Level Subtree 2 DEC* Reconstruction



DEC* Family-Level Reconstruction

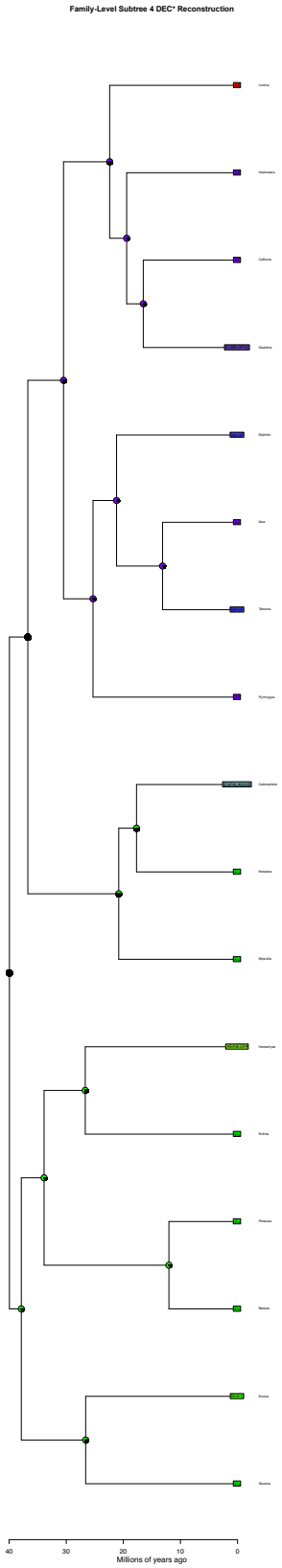
Appendix 6

Family-Level Subtree 3 DEC* Reconstruction



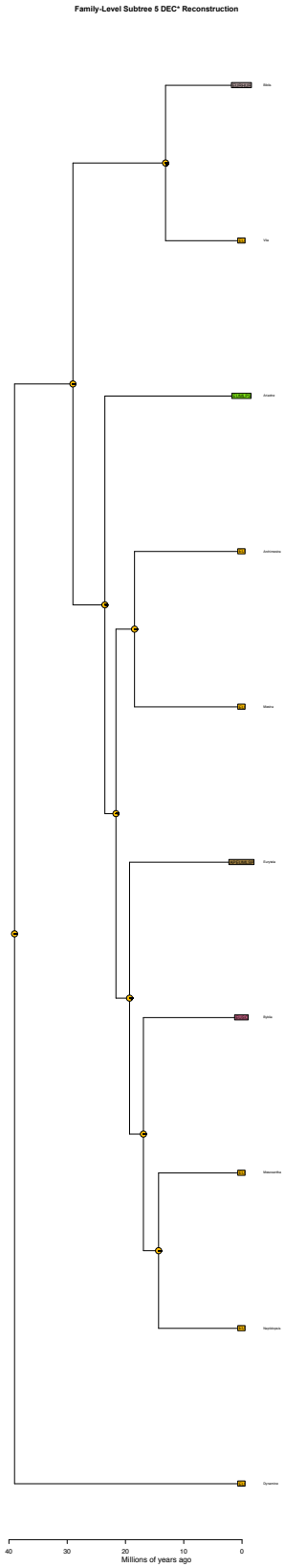
DEC* Family-Level Reconstruction

Appendix 7



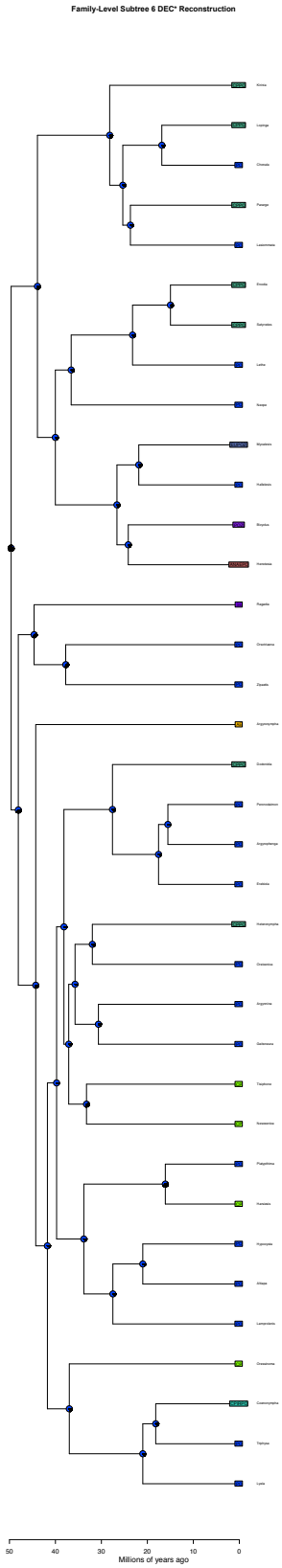
DEC* Family-Level Reconstruction

Appendix 8



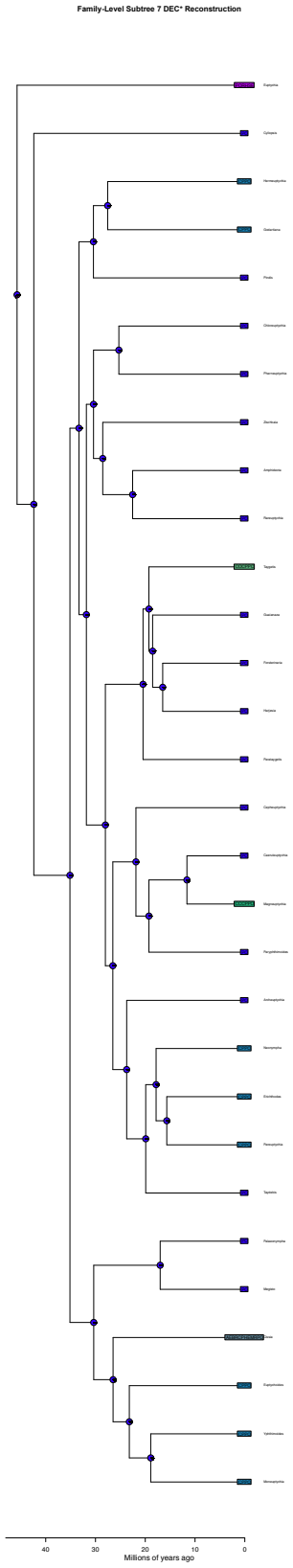
DEC* Family-Level Reconstruction

Appendix 9



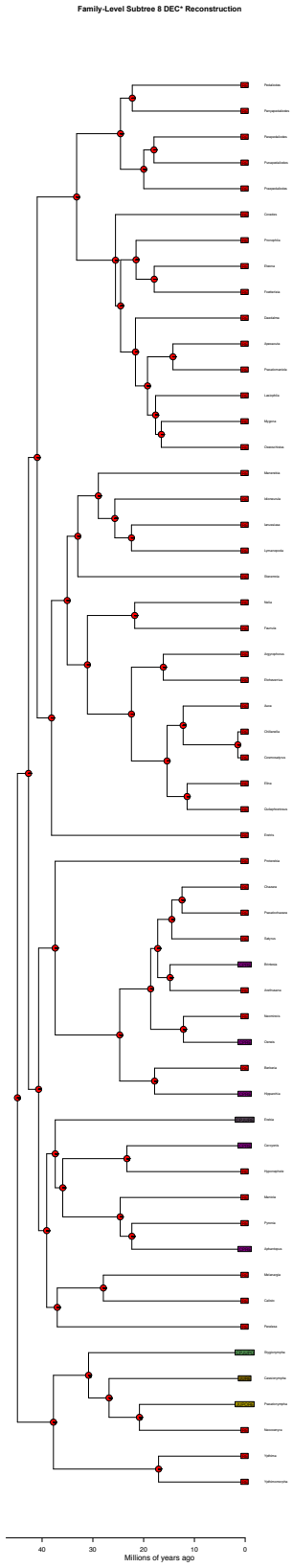
DEC* Family-Level Reconstruction

Appendix 10



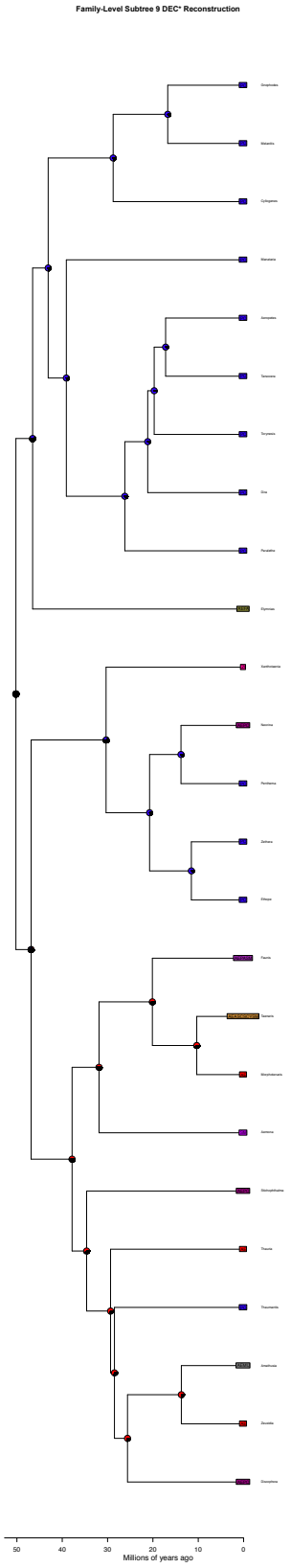
DEC* Family-Level Reconstruction

Appendix 11



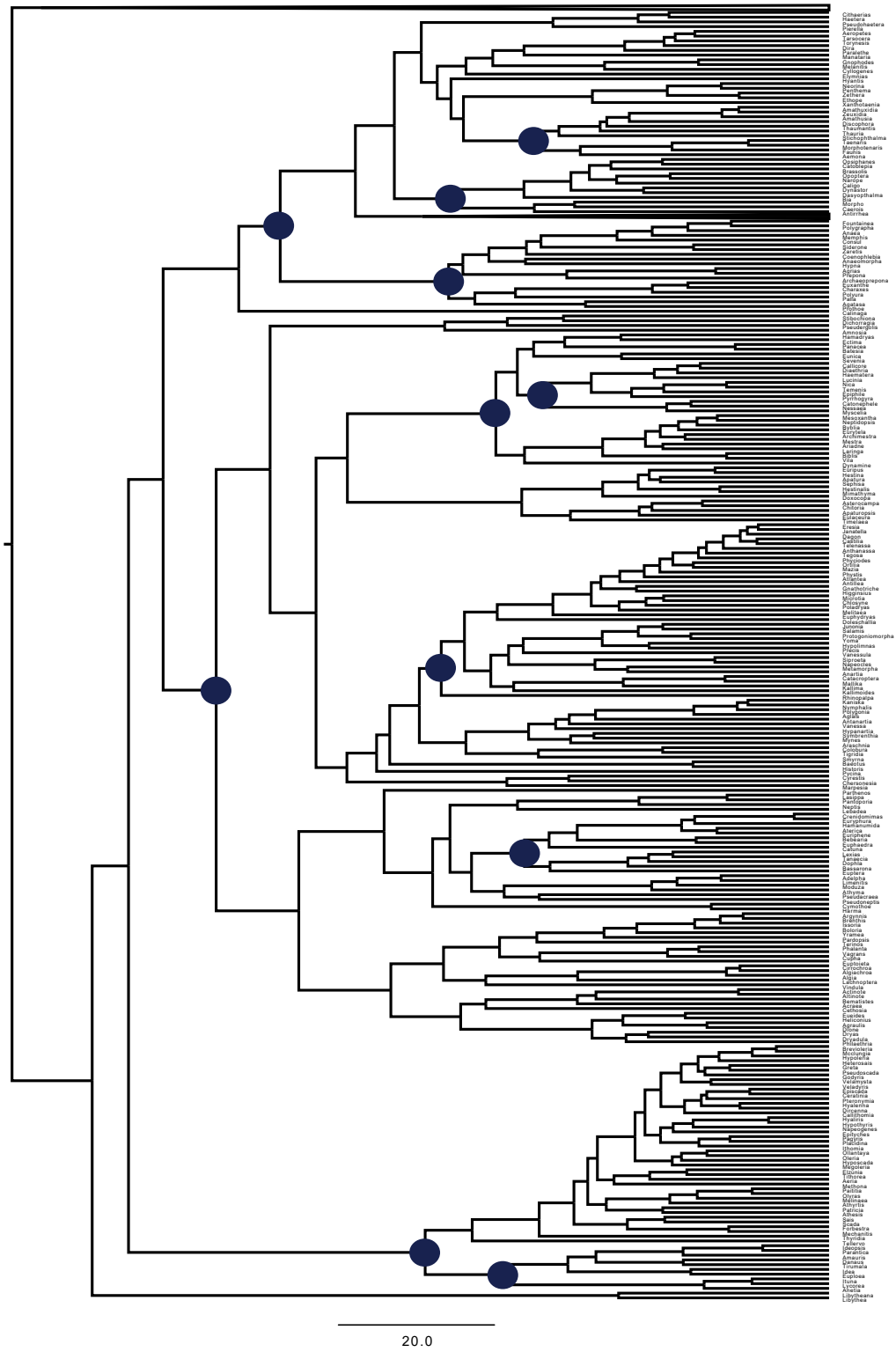
DEC* Family-Level Reconstruction

Appendix 12



DEC* Family-Level Reconstruction

Appendix 13



A figure showing the distribution of gamma nodes across the phylogeny in the order-level model.

Models Examining Speciation Rates and Diversity After Host-Use Changes and Ecological Opportunities Associated with Host-Use Gains

Appendix Table 1: Waiting Time Difference as a Function of Status (Order-Level)

Residuals:

Min	1Q	Median	3Q	Max
-57.435	-11.974	0.283	12.420	57.374

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.1141	1.5860	-0.072	0.943
Gains	-0.3356	7.1692	-0.047	0.963
Loss	-1.9161	6.8626	-0.279	0.780

Residual standard error: 20.85 on 230 degrees of freedom
Multiple R-squared: 0.0003434, Adjusted R-squared: -0.008349
F-statistic: 0.03951 on 2 and 230 DF, p-value: 0.9613

Appendix Table 2: Waiting Time Difference Ecological Opportunity Model (Order-Level)

Residuals:

Min	1Q	Median	3Q	Max
-26.836	-4.861	-1.279	6.023	32.918

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
Early Bird	4.570e-01	3.171e-01	1.441	0.166
Host Volatility	-4.693e+01	1.305e+02	-0.360	0.723
Host Age	-2.882e-01	2.766e-01	-1.042	0.311
Host Diversity	4.748e-04	8.498e-04	0.559	0.583

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 15.86 on 19 degrees of freedom
Multiple R-squared: 0.112, Adjusted R-squared: -0.07496
F-statistic: 0.5991 on 4 and 19 DF, p-value: 0.6678

Appendix Table 3: Diversity Difference as a Function of Status- Exclusive (Order-Level)
Residuals:

Min	1Q	Median	3Q	Max
-2349.98	-17.63	1.07	24.23	2210.95

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
None	1.633	10.293	0.159	0.874
Gain	-17.858	20.978	-0.851	0.395
Loss	-24.651	39.163	-0.629	0.529

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 238 on 804 degrees of freedom
Multiple R-squared: 0.001241, Adjusted R-squared: -0.001243
F-statistic: 0.4996 on 2 and 804 DF, p-value: 0.6069

Appendix Table 4: Diversity Difference Ecological Opportunity Model - Exclusive (Order-Level)

Residuals:

Min	1Q	Median	3Q	Max
-1245.32	-17.98	12.99	35.09	845.31

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
Early Bird	1.130e+00	1.268e+00	0.891	0.374
Host Volatility	-1.966e+02	2.596e+02	-0.757	0.450
Host Age	-9.289e-01	1.325e+00	-0.701	0.484
Host Diversity	1.747e-04	1.258e-03	0.139	0.890

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 153.2 on 187 degrees of freedom
Multiple R-squared: 0.01613, Adjusted R-squared: -0.004911
F-statistic: 0.7667 on 4 and 187 DF, p-value: 0.5482

Appendix Table 5: Diversity Difference as a Function of Status- Inclusive (Order-Level)

Residuals:

Min	1Q	Median	3Q	Max
-2349.98	-19.06	-0.63	25.59	2210.95

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
None	1.633	10.444	0.156	0.876
Gain	-23.444	21.286	-1.101	0.271
Loss	5.168	39.738	0.130	0.897

---Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 241.5 on 804 degrees of freedom
Multiple R-squared: 0.0016, Adjusted R-squared: -0.0008837

F-statistic: 0.6442 on 2 and 804 DF, p-value: 0.5254

Appendix Table 6: Diversity Difference Ecological Opportunity Model- Inclusive (Order-Level)

Residuals:

Min	1Q	Median	3Q	Max
-1302.86	-15.31	20.19	38.05	885.21

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
Early Bird	-2.005e+00	1.423e+00	-1.409	0.160
Host Volatility	-1.178e+02	2.913e+02	-0.404	0.686
Host Age	2.006e+00	1.487e+00	1.349	0.179
Host Diversity	-2.478e-04	1.411e-03	-0.176	0.861

---Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 171.9 on 187 degrees of freedom
Multiple R-squared: 0.02432, Adjusted R-squared: 0.003449
F-statistic: 1.165 on 4 and 187 DF, p-value: 0.3276

Appendix Table 7: Gamma Statistic as a Function of Status (MCMCglmm Model) (Order-Level)

Iterations = 3001:999901
Thinning interval = 100
Sample size = 9970

DIC:-8.289565

G-structure: ~animal

	post.mean	l-95% CI	u-95% CI	eff.samp
animal	5.206	3.791	6.915	9970

R-structure: ~units

	post.mean	l-95% CI	u-95% CI	eff.samp
units	0.01247	0.001141	0.03441	9211

Location effects: Gamma Stat ~ Status

	post.mean	l-95% CI	u-95% CI	eff.samp	pMCMC
(Intercept)	-3.9303	-4.6992	-3.1696	10544	<1e-04 ***
Gain	-0.2000	-1.4545	1.1361	9970	0.763
Loss	-0.1756	-1.4337	1.0796	9970	0.772

Geweke Diagnostic:
Fraction in 1st window = 0.1
Fraction in 2nd window = 0.5

(Intercept)	Statusb	Statusc
-0.4081	-0.8405	-0.4737

Appendix Table 8: Gamma Statistic Ecological Opportunity Model (MCMCglmm Model) (Order-Level)

Iterations = 3001:999901
Thinning interval = 100
Sample size = 9970

DIC: 16.40472

G-structure: ~animal

	post.mean	l-95% CI	u-95% CI	eff.samp
animal	5.192	1.098e-12	14.03	9970

R-structure: ~units

	post.mean	l-95% CI	u-95% CI	eff.samp
units	1.597	0.001316	6.353	9970

Location effects: gamma.stat ~ Early.Bird + Host Volatility + Host Age + Host Diversity

	post.mean	l-95% CI	u-95% CI	eff.samp	pMCMC
(Intercept)	2.816e+00	-1.622e+01	2.208e+01	11927	0.7286
EarlyBird	2.188e-01	6.728e-02	3.704e-01	9970	0.0134 *
Host Volatility	4.277e+00	-2.433e+01	3.144e+01	9970	0.7157
Host Age	-1.903e-01	-4.004e-01	1.841e-02	11261	0.0680 .
Host Diversity	-4.949e-05	-2.494e-04	1.267e-04	9970	0.5430

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Geweke Diagnostic:
Fraction in 1st window = 0.1
Fraction in 2nd window = 0.5

(Intercept)	Early.Bird	Host Volatility	Host Age	Host Diversity
1.8328	0.7729	-1.3619	-1.7799	0.3771

Appendix Table 9: Waiting Time Difference as a Function of Status (Family-Level)

Residuals:

Min	1Q	Median	3Q	Max
-55.505	-11.884	0.608	12.991	46.967

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.6385	1.7522	-0.364	0.716
Gains	10.3272	17.5839	0.587	0.558
Loss	3.7037	13.1261	0.282	0.778

Residual standard error: 18.18 on 131 degrees of freedom
Multiple R-squared: 0.003197, Adjusted R-squared: -0.01202
F-statistic: 0.2101 on 2 and 131 DF, p-value: 0.8108

Appendix Table 10: Waiting Time Difference Ecological Opportunity Model (Family-Level)

Residuals:
ALL 4 residuals are 0: no residual degrees of freedom!
Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
Early Bird	0.43170	NA	NA	NA
Host Volatility	-1.71051	NA	NA	NA
Host Age	-0.44402	NA	NA	NA
Host Diversity	0.00487	NA	NA	NA

---Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: NaN on 0 degrees of freedom
Multiple R-squared: 1, Adjusted R-squared: NaN
F-statistic: NaN on 4 and 0 DF, p-value: NA

Appendix Table 11: Diversity Difference as a Function of Status- Exclusive (Family-Level)

Residuals:

Min	1Q	Median	3Q	Max
-638.13	-13.13	-0.13	19.30	631.87

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
None	3.347	4.229	0.791	0.4291
Gain	-16.365	8.019	-2.041	0.0418 *
Loss	-36.938	17.034	-2.169	0.0306 *

---Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 78.14 on 528 degrees of freedom
Multiple R-squared: 0.01464, Adjusted R-squared: 0.0109
F-statistic: 3.922 on 2 and 528 DF, p-value: 0.02039

Appendix Table 12: Diversity Difference Ecological Opportunity Model – Exclusive (Family-Level)

Residuals:

Min	1Q	Median	3Q	Max
-321.56	-5.20	5.05	19.44	100.50

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
Early Bird	-0.4555273	1.6073393	-0.283	0.777
Host Volatility	50.8440592	38.1510208	1.333	0.185
Host Age	0.3806447	1.6048426	0.237	0.813
Host Diversity	-0.0007372	0.0007305	-1.009	0.315

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 55.55 on 131 degrees of freedom
Multiple R-squared: 0.08809, Adjusted R-squared: 0.06024
F-statistic: 3.164 on 4 and 131 DF, p-value: 0.01612

Appendix Table 13: Diversity Difference as a Function of Status- Inclusive (Family-Level)

Residuals:

Min	1Q	Median	3Q	Max
-611.14	-12.35	-1.01	19.15	630.46

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
None	3.347	4.242	0.789	0.4304
Gain	-20.497	8.043	-2.549	0.0111 *
Loss	-35.878	17.085	-2.100	0.0362 *

---Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 78.37 on 528 degrees of freedom
Multiple R-squared: 0.01798, Adjusted R-squared: 0.01426
F-statistic: 4.835 on 2 and 528 DF, p-value: 0.008306

Appendix Table 14: Diversity Difference Ecological Opportunity Model- Inclusive (Family-Level)

Residuals:

Min	1Q	Median	3Q	Max
-315.101	-10.199	6.786	21.615	104.233

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
Early Bird	-0.2979736	1.6486391	-0.181	0.8569

Host Volatility	32.4692473	39.1312915	0.830	0.4082
Host Age	0.2257809	1.6460782	0.137	0.8911
Host Diversity	-0.0012417	0.0007493	-1.657	0.0999 .

---Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 56.98 on 131 degrees of freedom
Multiple R-squared: 0.1107, Adjusted R-squared: 0.08357
F-statistic: 4.078 on 4 and 131 DF, p-value: 0.003789

Appendix Table 15: Gamma Statistic as a Function of Status (MCMCglmm Model) (Family-Level)

Iterations = 3001:999901
Thinning interval = 100
Sample size = 9970

DIC: 9.296147

G-structure: ~animal

	post.mean	l-95% CI	u-95% CI	eff.samp
animal	0.4068	1.889e-07	1.279	7243

R-structure: ~units

	post.mean	l-95% CI	u-95% CI	eff.samp
units	0.132	0.001445	0.5009	9495

Location effects: Gamma Stat ~ Status

	post.mean	l-95% CI	u-95% CI	eff.samp	pMCMC
(Intercept)	-3.7639	-4.3648	-3.2051	9970	<1e-04 ***
Gain	0.9105	-1.5147	3.3517	9977	0.449
Loss	0.8769	-0.9921	2.8689	10301	0.362

Geweke Diagnostic:

Fraction in 1st window = 0.1
Fraction in 2nd window = 0.5

(Intercept)	Statusb	Statusc
0.06493	-0.02890	0.41701

Appendix Table 16: Gamma Statistic Model (MCMCglmm Model) (Family-Level)

Iterations = 3001:999901
Thinning interval = 100
Sample size = 9970

DIC: 15.62889
G-structure: ~animal

	post.mean	l-95% CI	u-95% CI	eff.samp
animal				

R-structure: ~units

	post.mean	l-95% CI	u-95% CI	eff.samp
units				

Location effects: X0 ~ Early.Bird

	post.mean	l-95% CI	u-95% CI	eff.samp	pMCMC
(Intercept)	NA	NA	NA	NA	NA
Earl y Bird	NA	NA	NA	NA	NA
Host Volatility	NA	NA	NA	NA	NA
Host Age	NA	NA	NA	NA	NA
Host Diversity	NA	NA	NA	NA	NA

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Appendix Table 17: Geiger Likelihood Models

Host	Early Burst Likelihood	Null Likelihood	P-Value
Acanthaceae	-24.2705889166749	-24.4756559805568	1
Achariaceae	-6.80384015822068	-6.82545068855855	1
Amaranthaceae	-30.8217134313346	-33.4444478391127	1
Anacardiaceae	-34.8693627693718	-35.7662127144548	1
Annonaceae	-47.6983901313256	-47.7073509146823	1
Apiaceae	-262.702781432219	-6.79364646956203	2.55E-113
Apocynaceae	-83.970976114268	-62.2157567766798	4.22E-11
Aquifoliaceae	-18.701618428438	-18.8102560659757	1
Araceae	-14.5980476059492	-14.603123091087	1
Araliaceae	-6.55692632510564	-6.57810108152929	1
Araucariaceae	-6.93094951394479	-7.30260109705276	1
Arecaceae	-69.7115690138446	-69.769321530938	1
Aristolochiaceae	-12.4817130751343	-13.4698431237797	1
Asparagaceae	-22.1808653509207	-23.6966448728518	1
Asteraceae	-67.4035878782108	-67.773112229075	1
Berberidaceae	-6.92220757967273	-7.17499856502485	1
Betulaceae	-26.6029339190823	-29.3232600080215	1

Bignoniaceae	-34.8711261705137	-36.3031774972318	1
Boraginaceae	-262.702781432218	-26.2210235137557	7.26E-105
Brassicaceae	-17.5022156418202	-19.4169952117311	1
Bromeliaceae	-17.4019664014372	-17.8436178199958	1
Burseraceae	-12.3048329912621	-12.3979018524393	1
Calophyllaceae	-6.89279064975089	-7.01725583377264	1
Cannabaceae	-28.470004563345	-28.9418629270413	1
Cannaceae	-6.88593921398943	-6.99489573452801	1
Capparaceae	-6.68367580108751	-6.68393692895186	1
Caprifoliaceae	-36.2472205256356	-36.7901343455852	1
Caricaceae	-6.93562612174448	-8.00939866131714	1
Caryocaraceae	-6.89279062023153	-7.01725583377264	1
Caryophyllaceae	-6.93094951396764	-7.30260109705276	1
Casuarinaceae	-6.91800013899589	-7.14056717170542	1
Celastraceae	-24.5534575377609	-24.9445791389883	1
Chrysobalanaceae	-9.52930722550898	-11.6875200443071	1
Clethraceae	-1.38629462213216	-5.73203587136837	1
Clusiaceae	-34.8619350614944	-35.4572025497014	1
Combretaceae	-26.6052869233509	-26.8822891699412	1
Commelinaceae	-22.1808363360853	-22.1872694951667	1
Connaraceae	-12.3901457461315	-12.5726751287719	1
Convolvulaceae	-42.5129360296963	-43.8613258671962	1
Costaceae	-12.4820355390037	-13.568766898775	1
Crassulaceae	-12.3497384391058	-12.4572566104912	1
Cucurbitaceae	-11.5179450989476	-11.6605185066158	1
Cycadaceae	-12.482035595545	-13.8107082549629	1
Cyclanthaceae	-12.4817073678076	-13.4505691420422	1
Cyperaceae	-262.702781432218	-102.88318605514	1.74E-71
Dichapetalaceae	-12.479699791311	-13.0033091805488	1
Dilleniaceae	-12.4810190894205	-13.7589457466328	1
Dioscoreaceae	-6.9355719499748	-7.95786493250361	1
Dipterocarpaceae	-6.93492591056769	-7.58170471853563	1

Ebenaceae	-6.92025085903114	-7.15806434882404	1
Elaeocarpaceae	-6.89279062324096	-7.01725583377264	1
Ericaceae	-22.1804343354287	-24.9629373096855	1
Erythroxylaceae	-16.2128019005946	-16.8608974602591	1
Euphorbiaceae	-262.702781432219	-100.344058386746	1.36E-72
Fabaceae	-98.0211968133319	-98.639803044315	1
Fagaceae	-30.821329372525	-33.8409568304083	1
Garryaceae	-6.92773033332384	-7.23998943410325	1
Geraniaceae	-12.1261475637004	-12.1375813278419	1
Gesneriaceae	-17.4891850144841	-18.5532704674518	1
Grossulariaceae	-14.2437622150418	-14.3289923091088	1
Heliconiaceae	-26.3281409559336	-26.7255633018473	1
Hernandiaceae	-6.92252612161765	-7.17815932133376	1
Humiriaceae	-6.93504307531246	-7.62137303026753	1
Hypoxidaceae	-6.9307425281913	-7.29691749509513	1
Icacinaceae	-12.4808515256176	-13.3863085518521	1
Iridaceae	-12.4780533138332	-13.413011272161	1
Juglandaceae	-6.77360595501313	-6.78344785339499	1
Juncaceae	-12.2509321218185	-12.2985775963058	1
Lacistemataceae	-6.77360595495654	-6.78344785339499	1
Lamiaceae	-26.5984523132069	-28.6104079522117	1
Laminaceae	-6.73085202521106	-6.73252402695871	1
Lauraceae	-46.2156951340432	-45.2558838088449	0.165898265096917
Lecythidaceae	-6.93094951431532	-7.30260109705276	1
Liliaceae	-12.4822664176156	-14.57806183266	1
Linaceae	-6.68561363052249	-6.68581105918832	1
Loranthaceae	-6.93441109400682	-7.48380570737111	1
Lythraceae	-6.9226115726712	-7.17874402988691	1
Malpighiaceae	-38.7795977926754	-39.6588721969522	1
Malvaceae	-89.328849701096	-90.9415349019118	1
Marantaceae	-30.6597451388631	-31.3198987317011	1
Melastomataceae	-17.5001219468834	-18.9093702233436	1

Meliaceae	-27.0320957584463	-27.3827137993738	1
Melianthaceae	-12.4605719076862	-12.7333980040201	1
Menispermaceae	-21.5618574001633	-21.5621194326342	1
Monimiaceae	-22.1638369349807	-22.6806337340504	1
Moraceae	-62.6909090617541	-62.7145784949122	1
Musaceae	-33.602661890997	-33.7152352340827	1
Myrtaceae	-22.1757515880637	-22.5066367462364	1
Neckeraceae	-6.39971589230197	-6.47572158989062	1
Nyctaginaceae	-9.06259944595364	-11.0294199840162	1
Ochnaceae	-30.8165730133795	-32.6931880681823	1
Oleaceae	-26.6027905407189	-27.5705513020729	1
Onagraceae	-12.4608341526092	-13.0113438923752	1
Orchidaceae	-6.3702240818603	-6.45898136037853	1
Pandanaceae	-12.4769129863842	-13.3314677502581	1
Papaveraceae	-6.78125808774484	-6.79364646956203	1
Passifloraceae	-36.464501881233	-36.7912871843928	1
Phrymaceae	-6.92057344658009	-7.16069446736797	1
Phyllanthaceae	-1.38629462213216	-5.73203587136837	1
Piperaceae	-38.7740735707876	-40.0392505528978	1
Plantaginaceae	-38.6792479280904	-39.7981969479569	1
Poaceae	-104.282067331842	-104.363319882621	1
Polygonaceae	-26.6028081446915	-26.8777721274971	1
Polypodiaceae	-6.93220977514693	-7.3407239669079	1
Portulacaceae	-12.3497384670579	-12.4572566104912	1
Primulaceae	-30.8213877036606	-34.1548439292658	1
Proteaceae	-10.9508138204275	-12.0147237317404	1
Pteridaceae	-6.89436932227414	-7.0230947160647	1
Ranunculaceae	-1.38629462213216	-5.73203587136837	1
Restionaceae	-12.4353504828232	-12.7419008744681	1
Rhamnaceae	-56.6038842251682	-60.0545419503126	1
Rhizophoraceae	-6.92261157268639	-7.17874402988691	1
Rosaceae	-63.1341643099507	-55.0112997048474	5.56E-05

Rubiaceae	-50.7551051791666	-50.7583654413512	1
Rutaceae	-34.6714507795428	-35.2679226063573	1
Sabiaceae	-12.3863091782338	-12.5597554301038	1
Salicaceae	-76.1727885026932	-76.510966964234	1
Sapindaceae	-61.1099744227157	-61.2457360890633	1
Sapotaceae	-42.5223430421839	-44.345028652158	1
Saxifragaceae	-6.93414539721374	-7.45278698042378	1
Scrophulariaceae	-49.3739645740233	-50.2247274977083	1
Selaginellaceae	-11.4515469247845	-11.5874016927412	1
Simaroubaceae	-1.38629462213217	-5.73203587136837	1
Smilacaceae	-25.956329866151	-26.2951885632985	1
Solanaceae	-42.6087401725715	-43.1376521891618	1
Surianaceae	-6.92261157391639	-7.17874402988691	1
Theaceae	-6.90534845499983	-7.06777937358849	1
Typhaceae	-12.4820337799824	-14.2031982571009	1
Ulmaceae	-73.0465754509317	-73.1401209199552	1
Urticaceae	-88.0350410908002	-88.0555673055626	1
Verbenaceae	-262.702781432205	-78.0242702596041	2.58E-82
Violaceae	-35.4508362741492	-35.6237792430673	1
Vitaceae	-1.38629462213216	-5.73203587136837	1
Vochysiaceae	-6.93516147948022	-7.67143530421434	1
Xyridaceae	-6.9342605667635	-7.46447514772497	1
Zingiberaceae	-30.6253427877809	-31.3765646160151	1

The log-likelihood of early-burst fitted model and a non-transformed model (both equal-rates models) for each host family (as fitted in Geiger) and the p-value of the log likelihood ratio comparing the two models. A p-value of <0.05 indicates the early-burst model is significantly better than the null, non-transformed model.