University of Massachusetts Amherst

ScholarWorks@UMass Amherst

Doctoral Dissertations
Dissertations and Theses

November 2017

Specialization and trade-offs in plant-feeding insects

Daniel Peterson

SPECIALIZATION AND TRADE-OFFS IN PLANT-FEEDING INSECTS

A Dissertation Presented

by

DANIEL A. PETERSON

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2017

Organismic and Evolutionary Biology

SPECIALIZATION AND TRADE-OFFS IN PLANT-FEEDING INSECTS

A Dissertation Presented

by

DANIEL A. PETERSON

Approved as to style and content by:

_____
Benjamin B. Normark, Chair

_____
Michael E. Hood, Member

_____
Laura A. Katz, Member

_____
Adam H. Porter, Member

_____
John M. Lopes, Interim Director
Interdisciplinary Graduate Programs in the
Life Sciences, CNS

ACKNOWLEDGMENTS

I am indebted to many people for their support and assistance throughout my time at UMass. First I'd like to thank Ben for his inspirational zeal for science, for introducing me to the unbelievable world of scale insects, and for all the good times we had in the field. I'd also like to thank my committee members, Laura, Adam, and Michael, for their wisdom and guidance along the way. Geoff and Nate, you guys made it fun.

This project would have been impossible without the hard work, skill, and enthusiasm of the undergraduates working in the Normark and Morse labs, and I owe them a deep debt of gratitude.

To Scott and Akiko, thanks for making it the Golden Age.

To my friends and family, thanks for being there all these years.

ABSTRACT

SPECIALIZATION AND TRADE-OFFS IN PLANT-FEEDING INSECTS

SEPTEMBER 2017

DANIEL A. PETERSON, B.A., AMHERST COLLEGE

M.S., UNIVERSITY OF WASHINGTON

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Benjamin B. Normark

The immense diversity of life on Earth has been attributed to the partitioning of available resources into ecological niches, but it is not obvious what determines the niche size of each species. For example, most plant-feeding insects consume only one or a few closely-related host-plant species despite the advantages of having a broader diet. Many researchers have therefore suggested that the evolution of broad diets in plant-feeding insects must be constrained by genetic trade-offs between adaptations to alternative host-plants. Despite its intuitive feel, however, little empirical evidence in support of the trade-off hypothesis has emerged from decades of experimental studies comparing individual performance on alternative hosts within insect populations.

Here I use a broader approach to evaluate the role of trade-offs in driving ecological specialization in plant-feeding insects. By collecting host-use data for thousands of insect species and fitting those data into long-term evolutionary models, I investigate whether trade-off constraints have left observable signatures in the present ecological niches of existing species. Chapter 1 focuses on a single family of insects, the

armored scales (Hempitera: Diaspididae), revealing that positive correlations between evolutionary changes in host performance best fit the observed patterns of diaspidid presence and absence on nearly all focal host taxa, suggesting that adaptations to particular hosts enhance rather than reduce performance on other hosts. In chapter 2, I uncovered a complex network of evolutionary interactions between caterpillar adaptations to eleven host-plant orders, indicating that different host-use trade-offs act over long- and short-term evolutionary timescales. In contrast, host-use patterns of true bugs revealed a total lack of trade-offs for the same host-plant orders over both timescales. Chapter 3 turns to armored scale insects again, this time those that we collected in systematic surveys across a large diversity of trees in two tropical rainforest habitats. Using each insect species' abundance on each tree as a proxy for host-plant performance, we found no evidence for performance trade-offs on alternative hosts despite apparent host-use specialization. Overall, these results suggest that the extreme specialization of plant-feeding insects arises from long-term, potentially nonadaptive evolutionary processes rather than simple genetic trade-offs.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

PHYLOGENETIC ANALYSIS REVEALS POSITIVE CORRELATIONS

BETWEEN ADAPTATIONS TO DIVERSE HOSTS IN A GROUP OF

PATHOGEN-LIKE HERBIVORES

1.1 Abstract

A jack of all trades can be master of none – this intuitive idea underlies most theoretical models of host-use evolution in plant-feeding insects, yet empirical support for trade-offs in performance on distinct host plants is weak. Trade-offs may influence the long-term evolution of host use while being difficult to detect in extant populations, but host-use evolution may also be driven by adaptations for generalism. Here we used host-use data from insect collection records to parameterize a phylogenetic model of host-use evolution in armored scale insects, a large family of plant-feeding insects with a simple, pathogen-like life history. We found that a model incorporating positive correlations between evolutionary changes in host performance best fit the observed patterns of diaspidid presence and absence on nearly all focal host taxa, suggesting that adaptations to particular hosts also enhance performance on other hosts. In contrast to the widely invoked trade-off model, we advocate a "toolbox" model of host-use evolution in which armored scale insects accumulate a set of independent genetic tools, each of which is under selection for a single function but may be useful on multiple hosts.

1

1.2 Introduction

The prevalence of host specialization in plant-feeding insects is often interpreted as justification for the idea that a jack of all trades can be master of none – that trade-offs in performance on distinct host taxa limit the fitness of a generalist (Futuyma & Moreno 1988; Scheirs *et al.* 2005; Singer & Stireman 2005; Forister *et al.* 2012; Barrett & Heil 2012). Trade-offs are assumed to drive specialization in most analytical models of host-use evolution (Ravigné *et al.* 2009; Nurmi & Parvinen 2011; Remold 2012), yet empirical support for trade-offs is limited. Studies of wild and experimentally selected populations of plant-feeding insects have demonstrated positive genetic correlations in performance across hosts more often than negative correlations (reviewed by Futuyma 2008; Forister et al. 2012), suggesting that performance trade-offs are not the most important factor limiting host use within species. However, trade-offs may restrict host use without leaving a signature in the extant genetic variation of a species, especially if the trade-offs are strong enough to select against generalists and drive one specialist strategy to fixation (Joshi & Thompson 1995). Moreover, studies of one aspect of herbivore fitness, such as larval feeding performance, may not detect fitness trade-offs that occur through other mechanisms like adult feeding performance (Scheirs *et al.* 2005) or predation risk (Singer & Stireman 2005).

In contrast to the trade-off model, alternative theories suggest that adaptations to one host may be neutral or even positive with respect to performance on other hosts (Forister *et al.* 2012; Gompert *et al.* 2015). In fact, generalist adaptations that increase fitness across multiple hosts are likely to spread quickly within species (Whitlock 1996;

2

Normark & Johnson 2011). An example of such an adaptation is an "effector" protein that suppresses inducible defenses early in the host plant's biochemical response pathway. The biochemistry of immune signaling is much more conserved across plants than are the particular defensive compounds produced downstream, so adaptations that repress immune responses early in the pathway are likely to be effective against a wide array of host taxa (Ali & Agrawal 2012; Barrett & Heil 2012). Most likely, the signs and magnitudes of pleiotropic interactions related to host-use adaptations are variable, with some adaptations producing trade-offs in fitness, others increasing fitness across multiple hosts, and a third category acting independently between hosts. Nevertheless, we can ask whether such interactions have been positive, negative or neutral, on average, over the evolutionary history of plant-feeding insects.

Long-term evolutionary processes affecting host use can be investigated empirically through phylogenetic analysis, and researchers have used phylogenetic comparative methods to examine evolutionary patterns in host use in many plant-feeding insects, such as butterflies (Janz *et al.* 2001; Hardy & Otto 2014), bark beetles (Kelley & Farrell 1998), leaf beetles (Futuyma *et al.* 1995), and aphids (Jousselin *et al.* 2010; Hardy *et al.* 2015). However, interpreting specialization in these and many other plant-feeding insects is complicated by the fact that both host preference and performance influence realized host use (Forister *et al.* 2012). Although the ability to use a broad range of hosts could benefit any plant-feeding insect (Futuyma & Moreno 1988), those that oviposit exclusively on preferred host plants may not experience selection to use novel hosts (Ravigné *et al.* 2009). In practice, observed limitations on host use can be attributed to

performance constraints, difficulties associated with sensory identification of and attraction to host plants, or a lack of selection for increased host breadth (Mayhew 1997; Bernays 2001). Disentangling the evolutionary effects of these three potential constraints on host-use would be challenging in most groups of herbivorous insects, and we are not aware of any phylogenetic study that has attempted to do so.

In contrast to most plant-feeding insects, armored scale insects (Hemiptera: Diaspididae) exhibit a dramatically simplified life history, thereby providing an ideal clade within which to examine the evolution of host performance. An adult diaspidid female is completely sessile: she has permanently embedded herself in her host plant and has lost all locomotory appendages (Gullan & Kosztarab 1997). Dispersal and host "selection" is therefore accomplished by the lone motile life stage that precedes feeding – first instar "crawlers" (adult males fly but do not feed; Gullan and Kosztarab 1997). When these crawlers disperse away from the maternal host-plant they do so haphazardly, typically via wind or rarely via phoresy on other insects (Magsig-Castillo *et al.* 2010), and must feed exclusively on the plant upon which they happen to land (Hill & Holmes 2009). As a result, host repertoire is overwhelmingly influenced by host performance rather than host preference, and diaspidids experience strong selection favoring the ability to use all potential host plants in the local environment. It is therefore possible to attribute limitations in the host repertoires of armored scale insects to constraints on the ability to feed, develop and reproduce on each host. Moreover, observations of diaspidid host-use are extremely reliable because adult females are found only on host plants where they

have successfully developed, and only adult females can be identified to species using published keys (Ferris 1942; Miller & Davidson 2005).

Armored scale insects have colonized a diverse set of plant species, but most diaspidid species appear to specialize on smaller subsets of host taxa (Andersen 2009; García Morales *et al.* 2015). Here we took advantage of that host-use variation and the simplicity of diaspidid natural history to assess whether the pattern of diaspidid species presences and absences on focal host taxa suggests generally positive, negative, or neutral evolutionary interactions between host-use adaptations. Specifically, we gathered host-use data from collection records of North American diaspidids to parameterize a phylogenetic model of host-use evolution in that group. We assumed that a continuous performance character determining presence or absence on each host taxon evolved randomly along each branch of the diaspidid phylogeny, but we allowed for an interaction parameter between evolutionary changes in diaspidid performance on distinct hosts. Under the trade-off model we expected to find negative interactions between evolutionary changes in diaspidid performance on distinct hosts, because adaptations to one host should reduce performance on other hosts. On the other hand, if generalist adaptations have been more important than trade-off constraints, we expected to find positive interactions between evolutionary changes in diaspidid performance on distinct hosts, because adaptations to one host should also increase performance on other hosts. Alternatively, adaptations to distinct hosts could be mostly independent of each other, leading to a lack of interactions between evolutionary changes in diaspidid performances on distinct hosts.

By integrating across presence and absence host-use data from all observed

diaspidid species, we estimated a single interaction parameter characterizing each pair of

host taxa. This parameter described the average relationship between adaptations to those

two hosts across the entire evolutionary history of armored scale insects. We did not

investigate trade-offs between extant genotypes within individual diaspidid species,

although the possible historical existence of such within-species trade-offs is implicit in

our model. Trade-offs may vary between species due to epistatic effects (Remold 2012),

but evolutionarily labile trade-offs may not constrain host-use over the long-term. Our

approach looked for broad-scale and persistent interactions between diaspidid host-use

adaptations that could play a role in structuring insect communities across ecosystems.

1.3 Methods

North American diaspidid collection data were downloaded from the Tri-Trophic

Thematic Collection Network database (http://tcn.amnh.org/). We obtained 23,810

individual specimen records and parsed them for host-plant information, creating a binary

presence/absence matrix of insect species by host plants. Armored scale insects were

considered present on all hosts for which they had at least one host record in the database.

All plant taxonomic names were standardized with the Taxonomic Name Resolution

Service (Boyle *et al.* 2013) and insect taxonomic names with ScaleNet (García Morales

et al. 2015), an online database of scale insect literature. We created separate data

matrices for host species and host genera in order to analyze host-use interactions at two

potential levels of specialization. To account for shared ancestry between the observed

scale insect species, we constructed a phylogeny of as many species as possible by combining previously published sequences with those from additional specimens (see Appendix A.1).

Analyzing comparative data in a phylogenetic context requires assuming an explicit model of evolutionary change. Here we assumed a model in which the ability of an armored scale insect to develop on a given host taxon is governed by a hidden continuous performance trait that evolves by Brownian motion, with presence on the host occurring only when the performance trait exceeds a threshold value (Felsenstein 2012). The advantage of this model over a standard discrete-trait model is that the probability of a transition between presence and absence on a given host is not constant, but is dependent on whether the insect's performance value for that host is close to, far below, or far above the threshold for that host. We believe that a continuous performance character captures the polygenic nature of host use (Forister *et al.* 2012; Barrett & Heil 2012) more realistically than a fixed-rate discrete trait model originally developed to represent single genetic loci.

We used the model described above to obtain maximum-likelihood estimates, for each pairwise combination of focal host taxa, of an interaction parameter between the inferred performance traits underlying presence on the two hosts. This parameter described the correlation between random evolutionary changes in diaspidid performance on the two hosts that would be most likely to produce the observed patterns of diaspidid host-use. Positive correlations indicated that evolutionary increases in performance on one host co-occurred with increases in performance on the other host (making co-

occurrence on the two hosts more likely). Negative correlations indicated that increases in performance on one host co-occurred with decreases in performance on the other host (making co-occurrence on the two hosts less likely). All analysis was conducted with the program ThreshML (Felsenstein 2012) using the wrapper Rthreshml from the Rphylip package (Revell & Chamberlain 2014) in the R statistical environment (R Core Team 2015). As a check on the plausibility of the model-based calculations, we also calculated, for each pairwise combination of focal hosts, the Pearson product-moment correlation between binary vectors of diaspidid presence and absence on each host. This calculation measured diaspidid co-occurrence on each pair of hosts relative to a null expectation of independent presences and absences on the two hosts, without accounting statistically for the effect of shared evolutionary history between diaspidid species.

For visual interpretation of our results, we produced network graphs in which each host taxon was represented by a node and the thickness of the line between each pair of nodes was proportional to the absolute value of the best-fit correlation value between evolutionary changes in diaspidid performance on those hosts. Positive correlations were represented as solid black lines, negative correlations as dashed red lines. Network structure was revealed by plotting distances between nodes according to the force-directed Fruchterman-Reingold algorithm as implemented in the R package igraph (Csárdi & Nepusz 2006), with the attraction between nodes proportional to the corresponding evolutionary correlation between host performance traits (negative correlations were assigned an attraction value of zero).

We tested the statistical significance of the resulting correlation values by producing a series of 1000 null datasets that simulated independent Brownian motion of a continuous character for performance on each host taxon along the diaspidid phylogeny. We converted the continuous host performance values to a binary host presence/absence character by assuming that only the diaspidid species with the highest performance values were present on that host, with the threshold set by matching the number of species using that host in the empirical data. We calculated phylogenetic and non-phylogenetic correlations of diaspidid host use on all pairwise combinations of hosts in the simulated data, producing an expected distribution of correlations under the null model of no evolutionary interaction between host taxa.

Our statistical power to detect positive or negative host-use interactions relied on observing substantial variation in host use across diaspidids for each host taxon. We therefore had little power to analyze host taxa used by only a few armored scale insect species. To focus our computational resources on analyzing hosts with the most statistical power to detect interactions, we defined focal host plant taxa as those parasitized by at least ten diaspidid species. Nevertheless, due to concern that our focal host selection criterion would bias our results, we also analyzed 1000 interactions between host pairs randomly selected (without replacement) from the entire host pool.

1.4 Results

From the 23,810 specimen records we identified 347 armored scale insect species involved in 3,379 interactions with 1,435 plant species and 3,612 interactions with 912

9

plant genera. We obtained genetic sequences for 166 of those scale insect species to infer

a phylogeny (Fig. S1). Our focal hosts were 27 host species and 64 host genera that

harbored at least 10 of the 166 armored scale species for which we had phylogenetic data.

This highly pruned subset of the data nevertheless encompassed 12% of the observed

interactions by host species and 29% by host genus, allowing 351 pairwise host-species

comparisons and 2,016 pairwise host-genus comparisons.

Our analysis of host use across the diaspidid phylogeny revealed mostly positive

associations between evolutionary changes in performance on the focal host taxa, even

between angiosperm hosts and those belonging to the distantly related cycad and conifer

clades. At the host species level, the mean evolutionary correlation between inferred host

performance traits was 0.47, and 97% of correlations were greater than zero (Fig. 1.1a).

At the host genus level, the mean evolutionary correlation between inferred host

performance traits was 0.41, and 95% of correlations were greater than zero (Fig. 1.1b).

Simulation tests indicated that these results were highly statistically significant, with the

mean evolutionary correlation between inferred host performance traits and the

proportion of correlations greater than zero more extreme than any produced in 1000 null

model randomizations of the data ($P < 0.001$). The overwhelmingly positive associations

between diaspidid presences on distinct hosts were corroborated by the Pearson product-

moment correlations, although those correlations were generally lower in magnitude (Fig.

S2). Visualization of network graphs revealed no obvious network structure, with the

exception of one host genus, *Bambusa*, isolated from the other hosts (Fig. 1.1c, 1.1d).

Diaspidid performance on *Bambusa* exhibited negative evolutionary correlations with

performance on 89% of the other genera and accounted for 62% of all negative

correlations observed in the host genus comparisons. No taxon displayed such

consistently large deviations from the mean correlation value in any of the simulated null

datasets (mean deviation -0.535; $P<0.001$).

Mean correlations between host pairs randomly selected from the entire host pool

were much closer to zero, but they were positive on average (mean correlation between

host species: 0.06; between host genera: 0.09; Fig. S3).
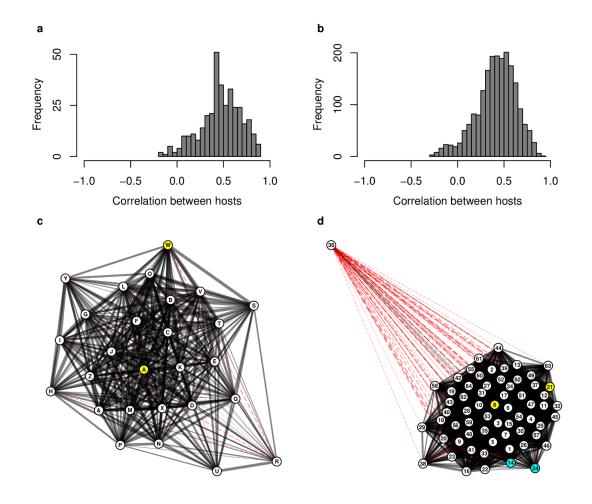
Figure 1.1. Distribution of inferred correlations between evolutionary changes in diaspidid presence on pairs of focal host plant species (a) and genera (b) under the evolutionary performance threshold model and the corresponding network graphs by host plant species (c) and genera (d). Positive interactions are represented by solid, black lines and negative correlations by dashed, red lines, with line thicknesses proportional to the magnitudes of the correlations. Network structure was plotted using the Fruchterman-Reingold algorithm, a force-directed layout method in which attraction between vertices was proportional to their correlation values (zero for negative correlations), and edge lengths were determined by minimizing total network energy. Each vertex represents a host taxon, with the following labels for host species – A: *Cycas revoluta*, B: *Cocos nucifera*, C: *Mangifera indica*, D: *Persea americana*, E: *Camellia japonica*, F: *Ligustrum japonicum*, G: *Citrus sinensis*, H: *Citrus limon*, I: *Citrus reticulata*, J: *Phoenix roebelenii*, K: *Dypsis lutescens*, L: *Laurus nobilis*, M: *Howea forsteriana*, N: *Nerium oleander*, O: *Beaucarnea recurvata*, P: *Ficus benjamina*, Q: *Persea borbonia*, R: *Carya illinoinensis*, S: *Morella cerifera*, T: *Hedera helix*, U: *Prunus persica*, V: *Ilex cornuta*, W: *Zamia pumila*, X: *Strelitzia reginae*, Y: *Citrus aurantiifolia*, Z: *Liriope muscari*, &: *Syagrus romanzoffiana* – and host genera – 1: *Quercus*, 2: *Citrus*, 3: *Ilex*, 4: *Persea*, 5: *Prunus*, 6: *Ficus*, 7: *Ligustrum*, 8: *Cycas*, 9: *Salix*, 10: *Camellia*, 11: *Cocos*, 12: *Mangifera*, 13:

12

Phoenix, 14: *Pinus*, 15: *Viburnum*, 16: *Ulmus*, 17: *Chamaedorea*, 18: *Euonymus*, 19: *Yucca*, 20: *Acer*, 21: *Zamia*, 22: *Carya*, 23: *Juglans*, 24: *Juniperus*, 25: *Magnolia*, 26: *Vaccinium*, 27: *Howea*, 28: *Nerium*, 29: *Fraxinus*, 30: *Hedera*, 31: *Strelitzia*, 32: *Laurus*, 33: *Vitis*, 34: *Dypsis*, 35: *Diospyros*, 36: *Bambusa*, 37: *Beaucarnea*, 38: *Populus*, 39: *Dracaena*, 40: *Celtis*, 41: *Rosa*, 42: *Euphorbia*, 43: *Areca*, 44: *Tillandsia*, 45: *Eugenia*, 46: *Morella*, 47: *Jasminum*, 48: *Olea*, 49: *Annona*, 50: *Liriope*, 51: *Calophyllum*, 52: *Callistemon*, 53: *Elaeagnus*, 54: *Ixora*, 55: *Syagrus*, 56: *Morus*, 57: *Hibiscus*, 58: *Buxus*, 59: *Osmanthus*, 60: *Pandanus*, 61: *Cymbidium*, 62: *Cinnamomum*, 63: *Murraya*, 64: *Asparagus*. Nodes representing taxa from non-angiosperm clades are highlighted (cycads in yellow, conifers in cyan).

1.5 Discussion

In our examination of the evolution of host use in armored scale insects, we found that a model of positive correlations between evolutionary changes in host performance best fit the observed patterns of diaspidid presence and absence on nearly all focal host taxa. Surprisingly, we found that even presence on host taxa that are extremely phylogenetically distant from the angiosperm majority, including conifers and cycads, showed mostly positive evolutionary correlations with presence on other hosts. We therefore conclude that trade-offs between adaptations to distinct host taxa do not explain the observed variation in diaspidid presence and absence on the diverse set of hosts we analyzed. Instead, generalist adaptations appear to play a significant role in shaping host use within Diaspididae.

However, despite the overwhelming trend of positive host-use correlations, presence on *Bambusa* (a genus of large, clumping bamboos), exhibited mostly negative evolutionary correlations with presence on other hosts. It is unclear why bamboos were negatively correlated with so many other genera (including several other monocots), although it is easy to imagine adaptive trade-offs involving the unique structure

13

(Parameswaran & Liese 1976; Vieira *et al.* 2002) or phenology (Franklin 2005; Nath *et al.* 2008) of bamboos. Nevertheless, another explanation for the observed pattern of negative correlations between *Bambusa* and the other genera could be that bamboos tend to grow in relatively dense, pure stands (Taylor & Zisheng 1987; Taylor *et al.* 1991), so the scale insects present in those habitats experience very weak selection for use of other hosts. The observed negative correlations may reflect patterns of short-term local adaptation or genetic drift rather than insuperable trade-offs.

By summarizing insect collection data as presence or absence on each host taxon, our analysis was unable to account for quantitative differences in survival or reproduction on distinct hosts that could allow a direct test of fitness trade-offs between host taxa. Unfortunately, the host-specific abundance data needed to accurately investigate quantitative trade-offs in host-related fitness are not presently available for Diaspididae due to the haphazard and non-quantitative nature of most scale insect collection events. We are currently conducting systematic surveys for armored scale insects at multiple high-diversity sites around the globe, the results of which may provide insight into the relationships between diaspidid abundance on each host tree species, genetic effective population size and evolutionary history.

We did not account for insect or host plant geographic distributions, a factor which might be expected to produce patterns of negative correlations between hosts whose ranges do not overlap. Nevertheless, we found few negative correlations between hosts, suggesting that geographic distributions do not limit host use for diaspidids in a substantial way. However, we only analyzed host records from within the United States –

14

a similar analysis of global host records would probably reveal negative interactions between geographically distant plant taxa.

Another caveat to our study is that we were forced (in order to maximize statistical power) to focus on plant taxa with at least ten armored scale insect species. It is possible that plant species that are involved in strong trade-offs with other host species tend to have few armored scale insect species that eat them, and that our data pruning thus biased our data against hosts that require trade-offs. We addressed this problem by analyzing data for host genera as well as host species, because more genera met the ten-diaspidid threshold. Moreover, our analysis of a random sample of all interactions, including those between hosts with fewer than ten scales, also revealed positive evolutionary interactions between host performances on average. Ultimately, of course, we cannot rule out the possibility that use of some hosts requires trade-offs with fitness on other hosts. Nevertheless, our results suggest that positive interactions between host-use adaptations are much more common than expected under a model of widespread trade-offs.

Given the largely positive evolutionary correlations between armored scale insect presence on diverse host taxa, we advocate a "toolbox" analogy to describe the evolution of host use in Diaspididae. While the trade-off model assumes fixed host-related constraints, for example, the amino acid sequence of a single enzyme that must be optimized for use with all hosts, we suggest that armored scale insects have accumulated a set of genetic tools that are optimized for particular host-related functions rather than for particular hosts. Each tool (e.g. protein or regulatory sequence) may be useful on

more than one host, thereby generating positive correlations in host use. Conflicts

between optimization of individual proteins for use on different hosts may be resolved

over evolutionary time by gene duplication and divergence. Evidence for the plausibility

of this model comes from plant-feeding Lepidoptera, in which cytochrome P450

monooxygenase enzymes have proliferated and diversified to catalyze a wide range of

plant allelochemical detoxification reactions (Berenbaum & Feeny 2008), with the

transcription of an enzyme induced by the presence of its toxic substrate in some cases

(Prapaipong *et al.* 1994; Wen *et al.* 2009). Under the toolbox model, host use is

determined by whether the armored scale insect has acquired and maintained the genes

necessary to accomplish the mechanical and biochemical tasks involved in feeding,

developing, and surviving on each host. Adapting to multiple hosts requires building up

the insect's genetic tool set, which would therefore increase the probability that it will

adapt to additional hosts.

The toolbox model implies that host specialization in armored scale insects is

driven by a combination of adaptive and non-adaptive forces. The ability to use a

particular host may often depend on the balance between selection for using that host and

the pressure of deleterious mutations in the genes required to use it (Whitlock 1996;

Remold 2012). The addition of new plant taxa to an insect's host repertoire could also be

limited by a lack of genetic variation for using novel hosts (Futuyma *et al.* 1995),

potentially due to the rarity of relevant gene duplications or other mutations that generate

host-related plasticity. Specialized diaspidids may therefore be more common in habitats

with low host diversity (e.g. bamboo forests), and in small or fragmented insect populations.

Diaspidids are clearly not stereotypical plant-feeding insects (Normark & Johnson 2011), and negative interactions between host-use adaptations may be much more important in other herbivorous groups than they are for armored scale insects. Nevertheless, our results suggest that extremely broad host ranges are not prevented by universal, unavoidable trade-offs. This conclusion supports the idea that the specialized host repertoires of most plant-feeding insects (Forister *et al.* 2015) may be driven by constraints on aspects of their ecology beyond feeding performance (Singer & Stireman 2005; Forister *et al.* 2012). Elucidating those constraints remains a challenge in most systems, but testing hypotheses about the long-term evolutionary relationships between host-related ecological traits is feasible with phylogenetic comparative methods (Hardy *et al.* 2015). We advocate this empirical, macroevolutionary viewpoint as a complementary approach to theoretical models and experimental tests of the factors influencing host use and the evolution of ecological specialization.

CHAPTER 2

MICRO- AND MACROEVOLUTIONARY TRADE-OFFS IN PLANT-FEEDING

INSECTS

2.1 Abstract

A long-standing hypothesis asserts that plant-feeding insects specialize on particular host plants because of negative interactions (trade-offs) between adaptations to alternative hosts, yet empirical evidence for such trade-offs is scarce. Most studies have looked for microevolutionary performance trade-offs within insect species, but host-use could also be constrained by macroevolutionary trade-offs caused by epistasis and historical contingency. Here we used a phylogenetic approach to estimate the micro- and macroevolutionary correlations between use of alternative host-plant taxa within two major orders of plant-feeding insects: Lepidoptera (caterpillars) and Hemiptera (true bugs). Across 1604 caterpillar species, we found both positive and negative pairwise correlations between use of eleven host-plant orders, with overall network patterns suggesting that different host-use constraints act over micro- and macroevolutionary timescales. In contrast, host-use patterns of 955 true bug species revealed uniformly positive correlations between use of the same host-plant orders over both timescales. The lack of consistent patterns across timescales and insect orders indicates that host-use trade-offs are historically contingent rather than universal constraints. Moreover, we observed few negative correlations overall despite the wide taxonomic and ecological diversity of the focal host-plant orders, suggesting that positive interactions between host-

use adaptations, not trade-offs, dominate the long-term evolution of host use in plant-feeding insects.

2.2 Introduction

Most plant-feeding insects are ecological specialists restricted to a small number of host-plant species (Forister *et al.* 2015). The prevalence of specialization is surprising given the advantages of being a generalist (including greater resource and refuge availability), and many researchers have therefore suggested that the evolution of generalism must be constrained (Futuyma & Moreno 1988; Futuyma *et al.* 1995; Scriber 2010). This constraint is usually imagined as a trade-off between adaptations to alternative hosts, whereby an increase in performance on one host comes at the cost of decreased performance on another host (Agrawal *et al.* 2010; Forister *et al.* 2012). Such trade-offs are crucial elements of most theoretical models of the evolution of specialization (Ravigné *et al.* 2009; Nurmi & Parvinen 2011; Remold 2012), and are often assumed to arise as consequences of the genetic architecture of host-use. One frequently invoked genetic model involves antagonistic pleiotropy, in which distinct alleles at a single locus have opposite fitness effects on alternative hosts (Futuyma & Moreno 1988; Scheirs *et al.* 2005; Scriber 2010; Gompert *et al.* 2015). For example, small changes to an enzyme could make it more efficient at detoxifying the secondary compounds of one plant species and less efficient at detoxifying the secondary compounds of another plant species (e.g. Li et al. 2003). Despite the intuitive appeal of antagonistic pleiotropy, however, empirical studies have generally failed to find evidence

for negative genetic correlations between performance on alternative hosts within insect species (Futuyma 2008; Forister *et al.* 2012; Gompert *et al.* 2015). Nevertheless, antagonistic pleiotropy may be difficult to detect within species because its effects can be obscured by segregating fitness variation at non-host-specific loci (Joshi & Thompson 1995). Moreover, genetic variation for use of novel hosts is often absent within a single population (Futuyma *et al.* 1995) and host-use is phylogenetically conserved in many insect groups (Futuyma & Agrawal 2009). We therefore cannot rule out the possibility that historical antagonistic pleiotropy drove the evolution of specialization in ancestral lineages of plant-feeding insects.

Although the prevalence of host-use specialization is often attributed to adaptive trade-offs, some theoretical models suggest that specialization can evolve even when adaptations to one host do not decrease performance on other hosts. Most insect species can choose which host plant they will feed on, so evolutionary feedback between the evolution of host choice and host performance could drive behavioral specialization (Ravigné *et al.* 2009; Nurmi & Parvinen 2011). For example, if a particular adaptation increases fitness on one host more than on another, individuals may evolve to feed preferentially on the host that gives them higher fitness (Fry 1996). If a non-preferred host is rarely used, selection for performance on that host will be weak, and mutation and genetic drift may eliminate the genetic tools required to use it (Whitlock 1996). In general, over long timescales, the selective environment will shape a lineage's genome, and epistatic interactions between new mutations and their genetic background will determine whether adaptations to novel hosts are possible (Weinreich *et al.* 2005; Remold

2012). We therefore expect that the evolution of host use is constrained by historical contingency and the complexity of genetic interactions. In fact, the importance of historical contingency and epistasis for the evolution of specialization has been demonstrated empirically by experimental evolution in microbial systems: trade-offs between environments can appear after significant periods of cost-free adaptation (Satterwhite & Cooper 2015) and realized trade-offs can differ between replicate lineages (Rodriguez-Verdugo *et al.* 2014). On a rugged adaptive landscape, evolutionary trajectories to alternative resource-use strategies may be mutually exclusive, and the direction taken by each lineage can depend on stochastic factors like mutation order (Elena & Lenski 2003).

If historical contingency and epistasis constrain the evolution of host-use in plant-feeding insects, adaptations to one set of hosts may reduce the probability of adapting to another set of hosts, driving specialization over long evolutionary timescales. Analogous macroevolutionary trade-offs have been described in plants; alternative defensive strategies tend to be negatively correlated over plant evolutionary history (Campbell & Kessler 2013; Johnson *et al.* 2014). It remains unknown, however, whether the diversification of host plant defenses has created trade-offs for plant-feeding insects.

Although trade-offs could arise from either genetic architecture or historical contingency, each of these mechanisms could instead produce positive interactions between use of distinct hosts. A single mutation might improve performance on multiple hosts, for instance by improving an effector protein that inhibits a defensive pathway conserved across multiple plant species (Barrett & Heil 2012). Similarly, the appearance

21

of a new enzyme class could create short-term trade-offs as the enzyme is calibrated to different hosts, but long-term performance benefits across multiple hosts after gene duplication (e.g. cytochrome P450 monooxygenases; Li et al. 2003). It is also possible that the genetic factors affecting performance on alternative hosts are independent, experiencing purely neutral interactions on both micro- and macroevolutionary timescales.

One way to investigate the importance of evolutionary interactions between traits is to map the traits onto empirical phylogenies of extant species and ask whether the traits are correlated over the evolutionary history of the focal group (Maddison & FitzJohn 2015). Negative correlations across species suggest trade-offs (Shoval *et al.* 2012), although correlations alone cannot distinguish between mechanistic constraints and associations shaped by selection pressure (Agrawal *et al.* 2010). However, recently developed statistical methods allow the partitioning of correlations between species traits into phylogenetic and residual components (Hadfield & Nakagawa 2010). Macroevolutionary interactions driven by historical contingency in ancestral lineages should be apparent in correlations between traits over phylogenetic timescales, while microevolutionary interactions should be captured by residual variation – the evolution that has happened independent of the species' shared ancestry. Phylogenetic analyses therefore allow characterization of positive, negative, and neutral interactions between traits over both short and long evolutionary timescales (Fig. 2.1).
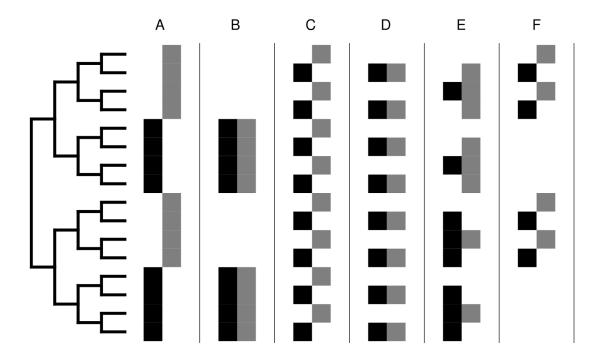
Figure 2.1. Phylogenetic and residual evolutionary correlations between traits. Hypothetical scenarios of evolutionary correlation between herbivore presence on two hosts: (A) negative phylogenetic correlation, (B) positive phylogenetic correlation, (C) negative residual correlation, (D) positive residual correlation, (E) negative phylogenetic and positive residual correlations, (F) positive phylogenetic and negative residual correlations. In each example, black squares on the left indicate which species in the herbivore phylogeny are present on host 1, and gray squares on the right indicate which species are present on host 2.

Here we used phylogenetic methods to investigate interactions between adaptations to diverse host taxa over micro- and macroevolutionary timescales in two orders of plant-feeding insects: Lepidoptera (caterpillars) and Hemiptera (true bugs). Using digitized insect host-use records from North America, we estimated pairwise evolutionary correlations between use of common host-plant orders across hundreds of species in each insect order. We then combined the pairwise correlations into network graphs, revealing overall patterns of host-use evolution. We expected that use of the focal

hosts would be mostly negatively correlated or clustered into discrete functional groups if specialization in plant-feeding insects is driven by widespread trade-offs between adaptations to different hosts. A distinction between micro- and macroevolutionary trade-offs could be made by asking whether the negative correlations appeared in the insects' residual or phylogenetic host-use variation. On the other hand, if specialization is not caused by trade-offs between adaptations to alternative hosts, we expected that correlations between host-use traits would be neutral or positive, with little overall network structure.

2.3 Methods

Lepidopteran host-use data were downloaded from the HOSTS database (nhm.ac.uk/hosts; Robinson et al. 2015), a worldwide collection of published records of caterpillars successfully reared on host-plants. Hemipteran host-use data were downloaded from the Tri-Trophic Thematic Collection Network database (tcn.amnh.org), a catalog of field-collected insect specimens and their associated host-plants at academic museums in the United States. These datasets differ in the nature of the host-use records (published rearing records vs. field observations), but each represents the best available host-use data for that insect order. For both datasets, we restricted our analysis to records from North America (all localities labeled USA, Canada, Mexico or Nearctic). All plant taxonomic names were standardized with the Taxonomic Name Resolution Service (Boyle *et al.* 2013) and insect taxonomic names with the python package TaxonNamesResolver and the following reference databases: Aphid Species File (Favret

2015), Integrated Taxonomic Information System (itis.gov), and Catalogue of Life (catalogueoflife.org). We created binary presence/absence matrices of lepidopteran and hemipteran species by host plant order, with insects considered present on all hosts for which they had at least one host-use record. To verify that potentially erroneous single observations of insect-by-host-order interactions were not driving our results, we also analyzed a second set of presence/absence matrices in which insects were considered present on a host-plant order only if they were observed feeding on at least two genera in that order.

To focus computational resources on host taxa with enough statistical power to detect evolutionary host-use interactions, we restricted our main analyses to focal host orders used by at least 100 insect species in one insect order (~10% of the total focal insect species per order). However, we also categorized host-use among host plant families that met the same cutoff to verify that any correlations we observed between use of the focal host orders were not artifacts created by lumping diverse plant families within order-level host-use traits.

We estimated time-scaled phylogenies for the North American lepidopteran and hemipteran species in our host-use dataset using a phyloinformatic approach (see Appendix A.2 for details). Phylogenetic data were not available for all species in the host-use dataset, but there was an overlap of host-use and phylogenetic data for 1604 lepidopteran species and 955 hemipteran species. Phylogenies and host-use datasets for these species are available in the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.m0n46.

We used the insect-species-by-plant-order presence/absence data to investigate whether our focal host-use traits (presence/absence on each plant order) were positively or negatively correlated across the insect species. These correlations quantified whether insect species present on plant order A were more or less likely to be present on plant order B than expected by chance. For each insect order (Hemiptera and Lepidoptera) and each pairwise comparison between host-use traits, we set up a phylogenetic mixed model (Hadfield & Nakagawa 2010) with a logit link function (to accommodate binary data) using the package MCMCglmm (Hadfield 2010) in the R statistical environment (R Core Team 2015). We estimated both phylogenetic and residual correlations between the two host-use traits using the "`random=~us(trait):insect`" and "`rcov=~us(trait):units`" syntax (Hadfield 2010). Prior parameter distributions were specified as "`prior<-list(R=list(V=diag(2),nu=2), G=list(G1=list(V=diag(2),nu=2)))`", and the mean of the posterior distribution was taken as the final estimate for each parameter. All MCMC chains ran for 10 million iterations with a burn-in of 1 million iterations, and we evaluated the convergence of ten chains for each model. Gelman-Rubin convergence analysis of each model's ten chains produced potential scale reduction factors under 1.10 in every case (96% were under 1.01), suggesting that all chains successfully converged (Gelman & Rubin 1992).

After separately estimating all pairwise evolutionary correlations between the focal host-use traits, we evaluated two emergent properties of the host-use network as a whole. First, we calculated the mean of all correlations involving each host-use trait to

summarize whether presence on that host tended to be positively or negatively correlated with presence on all other hosts. Second, we asked whether the host-use traits could be grouped into clusters that had positive correlations within them and negative correlations between them. To identify the most strongly supported clusters, we used a distance matrix calculated from the pairwise correlations between host-use traits to produce a dendrogram of associations between the traits. Agglomerative hierarchical clustering was performed with the "centroid" method of the hclust function in the R package fastcluster (Müllner 2013). After obtaining the dendrogram, we evaluated all possible cluster divisions produced by pruning the dendrogram at a single "level" (from broadest, with all host-use traits in a single cluster, to narrowest, with each host-use trait in its own cluster). The support for a given set of clusters was defined as the sum of all correlations between host-use traits in the same cluster minus the sum of all correlations between host-use traits in different clusters. Thus, positive correlations within clusters and negative correlations between clusters increased the support score, while negative correlations within clusters and positive correlations between clusters reduced the support score. The set of cluster divisions with the highest support score was chosen as the best characterization of network structure.

We tested the statistical significance of the empirical mean host-use trait correlations and overall network structure scores by comparing them to those calculated for 100 null datasets. Each null dataset was generated by simulating independent Brownian motion of a continuous character for performance on each focal host order along the insect phylogenies, plus an equal amount of normally distributed residual

27

variation in the performance values. We converted the resulting continuous host

performance values to a binary host presence/absence character by assuming that only the

insect species with the highest performance values for each host taxon were present on

that host, with the threshold set by matching the number of species using that host in the

empirical data (Felsenstein 2012). We then calculated all pairwise correlations between

the host-use traits, mean correlations per host-use trait, and whole-network structure as

we did for the empirical data. Empirical host-use trait mean correlations and network

structure scores were considered statistically significant when a Z-test indicated less than

a 5% chance of a value as extreme as the empirical value being randomly sampled from

the distribution of that parameter's null dataset values (which were approximately

normally distributed).

2.4 Results

We obtained North American host-use records and phylogenetic data for 1604

caterpillar species and 955 bug species (Fig. 2.2). Eleven host-plant orders met our

prevalence cut-off of 100 species from one insect order, and each of them met the cut-off

for both Hemiptera and Lepidoptera: Asterales, Caryophyllales, Ericales, Fabales,

Fagales, Lamiales, Malpighiales, Pinales, Poales, Rosales, and Sapindales. Interactions

with these focal host-plant orders accounted for 77% of all insect-species-by-plant-order

interactions in the Lepidoptera dataset and 57% in the Hemiptera dataset. Fewer host-

plant families met the 100-insect prevalence cutoff – for Lepidoptera: Asteraceae,

Betulaceae, Ericaceae, Fabaceae, Fagaceae, Pinaceae, Poaceae, Rosaceae, Salicaceae, and

Sapindaceae; for Hemiptera: Asteraceae, Fabaceae, Fagaceae, and Rosaceae.
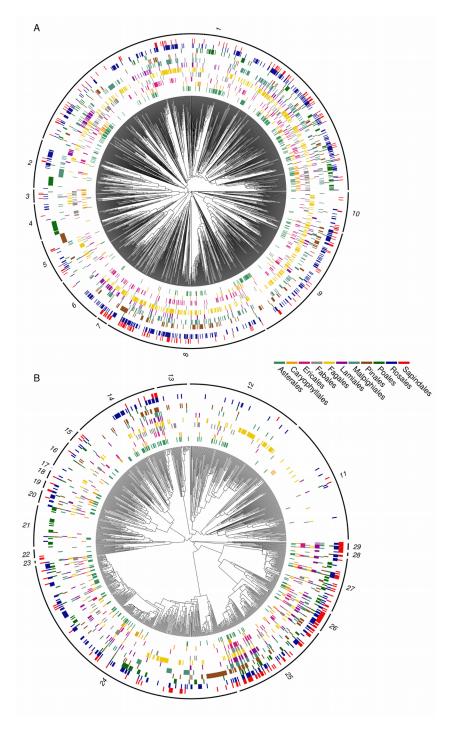
Figure 2.2. Maps of host-use traits on insect phylogenies. For each host-plant order, colored blocks indicate which insect species have been observed on that host. Insect

species with no hosts shown were observed only on non-focal hosts or had no host-use information associated with their locality records (Hemiptera only). Insect families (and one superfamily) are indicated around the phylogenies as follows: (A) Lepidoptera – 1: Noctuoidea, 2: Nymphalidae, 3: Lycaenidae, 4: Hesperiidae, 5: Pyralidae, 6: Sphingidae, 7: Saturniidae, 8: Geometridae, 9: Tortricidae, 10: Gracillariidae. (B) Hemiptera – 11: Cicadellidae, 12: Membracidae, 13: Cicadidae, 14: Miridae, 15: Tingidae, 16: Pentatomidae, 17: Scutelleridae, 18: Coreidae, 19: Rhopalidae, 20: Lygaeidae, 21: Delphacidae, 22: Fulgoridae, 23: Flatidae, 24: Aphididae, 25: Diaspididae, 26: Coccidae, 27: Pseudococcidae, 28: Psylloidea, 29: Aleyrodidae.

We recovered both positive and negative correlations between use of the focal host orders in the Lepidoptera, but mostly positive correlations in the Hemiptera (Fig. 2.3). The network of phylogenetic correlations between lepidopteran use of the focal host orders was significantly structured ($Z = 7.08$, $P < 0.0001$), revealing two large clusters of host taxa (Fig. 2.4a). Cluster membership was phylogenetically diverse: the gymnosperm order Pinales (conifers) and monocot order Poales (grasses) were each affiliated with a different set of eudicot orders. Residual correlations between lepidopteran use of the focal host taxa also showed significant network structure ($Z = 9.86$, $P < 0.0001$) but on this timescale use of all angiosperm hosts formed a single cluster of mostly positive associations (Fig. 2.4b). Use of Pinales was isolated from the angiosperm cluster, exhibiting a statistically significant negative mean pairwise correlation with use of the other hosts (-0.22, $Z = -3.17$, $P = 0.0015$). In contrast, hemipteran host-use correlations indicated significant support for a single host-use cluster encompassing all focal hosts over both phylogenetic ($Z = 11.90$, $P < 0.0001$; Fig. 2.4c) and residual timescales ($Z = 23.18$, $P < 0.0001$; Fig. 2.4d).

The patterns of host-use correlations found in the more conservative dataset (with each insect observed using at least two genera in each host-plant order) produced results

nearly identical to those of the original analysis, though statistical power was reduced

(Fig. S4). Moreover, our analysis of correlations between use of the focal host-plant

families also corroborated the network of evolutionary host-use associations revealed at

the host order level (Fig. S5).
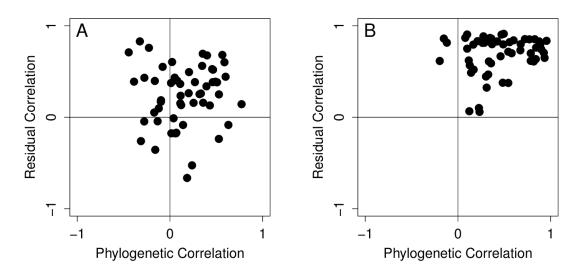
Figure 2.3. Empirical phylogenetic correlation by residual correlation plots of all 55
pairwise combinations of the focal host orders for Lepidoptera (A) and Hemiptera (B).
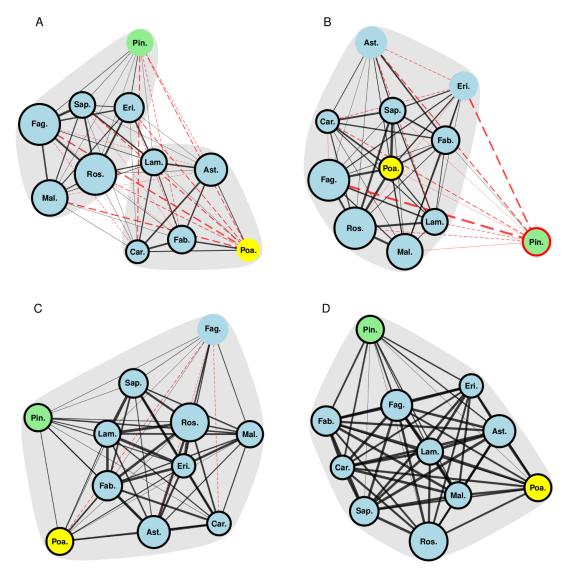
Figure 2.4. Network graphs of pairwise host-use trait correlations. (A) Lepidoptera – phylogenetic correlations. (B) Lepidoptera – residual correlations. (C) Hemiptera – phylogenetic correlations. (D) Hemiptera – residual correlations. Each vertex represents a host order, with vertex area proportional to the number of insects that were observed on that host. Positive interactions between presence on a pair of hosts are represented by solid, black lines and negative correlations by dashed, red lines, with line thickness proportional to the magnitude of the correlation. Network spatial structure was determined using the Kamada-Kawai (1989) algorithm, a force-directed layout method in which "repulsion" between vertices was proportional to the inverse of one plus the correlation values between the respective hosts. Vertices are labeled with the following abbreviations – Ast.: Asterales, Car.: Caryophyllales, Eri.: Ericales, Fab.: Fabales, Fag.: Fagales, Lam.: Lamiales, Mal.: Malpighiales, Pin.: Pinales, Poa.: Poales, Ros.: Rosales, Sap.: Sapindales. Vertices are colored by taxonomic group – eudicots: blue, monocots:

yellow, conifers: green. Statistically significant clusters ($P < 0.05$) are indicated by gray bubbles. Individual host orders with mean correlations of significantly higher magnitude than expected ($P < 0.05$) are indicated by bold vertex outlines (black for positive means, red for negative means).

2.5 Discussion

Most models of the evolution of ecological specialization assume negative interactions (trade-offs) between adaptations to different environments (Ravigné et al. 2009), but such interactions could also be neutral or positive (Gompert *et al.* 2015; Peterson *et al.* 2015). Here we used a statistical, phylogenetic approach to estimate the micro- and macroevolutionary correlations between use of eleven common host plant orders in both caterpillars and true bugs. Our results suggest that distinct micro- and macroevolutionary trade-offs constrain host-use in caterpillars, but use of all focal hosts are positively correlated on both timescales in true bugs. Overall, positive interactions between host-use adaptations appear to be more common than trade-offs in these plant-feeding insects.

We found some support for the idea that microevolutionary constraints (e.g. antagonistic pleiotropy) can produce host-use trade-offs in plant-feeding insects: lepidopteran presence on angiosperms was negatively correlated with presence on conifers over a short-term, phylogenetically independent timescale. This pattern suggests that caterpillar species tend to be found on either angiosperm or conifer hosts (not both), yet they can shift between these alternative host-plant clades over relatively short evolutionary timescales. Such trade-offs between labile but mutually exclusive host-use traits are particularly significant because they can promote rapid speciation (Nosil *et al.*

2002) and adaptive radiations (Farrell 1998; Janz *et al.* 2006). In this case, microevolutionary constraints appear to reflect ancient phylogenetic divergence between host clades (Soltis *et al.* 2011). A similar pattern of microevolutionary trade-offs between use of phylogenetically distant hosts has been observed in networks of ecological interactions between fleas and their mammal hosts (Hadfield *et al.* 2014) and pollinators and their plant hosts (Rafferty & Ives 2013). Nevertheless, the prevalence of such constraints in plant-feeding insects, for instance between alternative host genera or species, remains unclear given that the single microevolutionary trade-off observed here occurred over the largest phylogenetic distance present among our focal host-plant taxa.

Most theoretical work on host-use evolution has focused on microevolutionary trade-offs, but we found that host-use constraints can also act over longer, macroevolutionary timescales. Over the phylogeny of the Lepidoptera, we observed a negative correlation between presence on hosts in two large, taxonomically diverse clusters. Interestingly, the clusters appeared to divided by morphology rather than phylogeny, with predominantly woody plant taxa (e.g. Pinales, Fagales) in one cluster and predominantly herbaceous taxa (e.g. Asterales, Poales) in the other. This pattern could reflect a long-term trade-off for lepidopteran lineages between use of alternative host growth forms or the habitats where those growth forms are found (Futuyma 1976; Janz & Nylin 1998). However, it is difficult to attribute macroevolutionary patterns to any particular mechanism. The phylogenetic correlations we detected here could be driven by any number of processes, including the accumulation of epistatic interactions (Weinreich *et al.* 2005; Remold 2012), evolutionary feedback between host performance

34

and host choice (Whitlock 1996; Ravigné et al. 2009), or geographic specificity of plant and insect lineages (Hadfield et al. 2014). Regardless, host-specificity in the Lepidoptera is clearly influenced by macroevolutionary processes that may be undetectable within a single insect population.

In contrast to the patterns observed in the Lepidoptera, the Hemiptera showed mostly positive associations between use of all focal host taxa over both micro- and macroevolutionary timescales. This surprising result suggests that adaptations to one host also increase fitness, on average, on all other hosts (Peterson et al. 2015). Moreover, hemipteran generalism appears completely unrestrained by host taxonomy even over very long timescales, leading to the evolution of both super-generalist species and clades where generalist strategies are common (Normark & Johnson 2011). However, we do not account for differences in fecundity between specialist and generalist insects on particular hosts; it may be that generalists usually have lower fitness – i.e. they are jacks of all trades but masters of none (Futuyma and Moreno 1988). Nevertheless, costs of generalism have been difficult to document (Forister et al. 2012; Gompert et al. 2015), so the positive residual correlations we observed may instead represent evolutionary breakthroughs made possible by novel mechanisms of phenotypic plasticity or other generalist adaptations (Barrett and Heil 2012).

There are many differences between Lepidoptera and Hemiptera (and between the two datasets analyzed here), but their fundamentally distinct relationships with host plants may be important to understanding why evolutionary interactions between host-use traits appear to be different in the two groups (Pires & Guimarães 2012). Hemiptera are

sucking insects, while Lepidoptera are generally leaf-chewers (Forister *et al.* 2015).

These two feeding modes elicit different modes of plant defensive responses (Ali &

Agrawal 2012), and sap-sucking may be particularly amenable to generalist adaptations

that circumvent host defenses (Barrett and Heil 2012). In contrast, Lepidoptera often rely

on specialized enzymes to detoxify defensive chemicals (Berenbaum & Feeny 2008),

which may constrain the evolution of generalism, although generalist Lepidoptera do

exist, possibly powered by phenotypic plasticity in enzyme expression (Yu *et al.* 1979; Li

et al. 2002).

Overall, the relatively few, broad-scale trade-offs found here fail to explain the

prevalence of specialization in plant-feeding insects, which are often restricted to hosts in

a single plant family or genus (Forister et al. 2015). Our main analysis grouped hosts by

order, obscuring potential variation within orders in defensive strategies; host plant

families or genera with strong or physiologically unique defenses may be more likely to

produce trade-offs for plant-feeding insects than host plant taxa with weaker or more

common defenses. However, our analysis of evolutionary correlations between use of

common host-plant families revealed results nearly identical to those for the focal host-

plant orders. Moreover, a previous study of genus-level host-use in the large hemipteran

family Diaspididae found positive correlations between use of all hosts but one within a

network of 64 taxonomically diverse host genera (Peterson et al. 2015), indicating that

greater taxonomic resolution does not necessarily reveal trade-offs between host-use

traits.

We also took a broad approach in looking for correlations between host-use traits across whole insect orders, thereby overlooking any idiosyncratic trade-offs that may arise from the unique natural history of individual insect species. Species-specific trade-offs have been documented (e.g. Nosil et al. 2002), yet our results suggest that few microevolutionary trade-offs constrain host-use across large numbers of insect species. Thus, although trade-offs may emerge at any time due to novel epistatic interactions (Remold 2012; Satterwhite and Cooper 2015), the fact that generalist species frequently escape such trade-offs suggests that long-term evolutionary interactions between host-use traits are dominated by positively pleiotropic or neutral adaptations.

It is important to note that our conclusions reflect the particular data we analyzed. Research and publication bias have undoubtedly influenced the insect species that appear in our host-use databases, as well as those for which genetic sequences are available. It may be that insect pests, which are often polyphagous, are over-represented, which could bias our analyses in favor of positive correlations. Moreover, our analyses are restricted to North American insects, which tend to be more polyphagous than tropical insects (Dyer et al. 2007; Forister et al. 2015; but see Hardy et al. 2015 for a counterexample in the Hemiptera). Another limitation of our approach is that we could only analyze host-use trait relationships between host plant taxa used by at least 10% of the insect species present in our databases. Thus, we were not able to look for trade-offs between hosts that are used by few Hemiptera and Lepidoptera. Ultimately, we can conclude that persistent trade-offs between taxonomically broad host groups do not appear to be the most important factor limiting host breadth in North American Lepidoptera or Hemiptera, but

more focused studies of well characterized insect families or genera will be necessary to evaluate whether trade-offs act over taxonomic scales or ecological axes not considered here.

Trade-offs play an intuitive and possibly inescapable role in constraining performance across multiple tasks (Shoval *et al.* 2012), but performance limits may not define the ecological niches of plant-feeding insects. Alternative factors, such as mate-finding (Hawthorne & Via 2002), natural enemies (Singer & Stireman 2005), or neural constraints in host identification (Bernays 2001), may shape the evolution of each species' ecological niche. Host-range may also be limited by genetic drift even if adaptive interactions between host-use traits are positive or neutral (Gompert et al. 2015). Specialization-by-drift might be particularly significant in a geographical context, as interactions between host-range and geographical range can strongly affect the host-use selection pressures experienced by an insect lineage (Janz & Nylin 2008). In the absence of much evidence for negative interactions between host-use adaptations in plant-feeding insects, we should consider neutral models both for the structure of ecological networks (Canard *et al.* 2014) and for how those networks evolve over time.

CHAPTER 3

ARMORED SCALE INSECT HOST-USE IN TWO TROPICAL RAINFORESTS REVEALS SPECIALIZATION WITHOUT TRADE-OFFS

3.1 Abstract

Most plant-feeding insects are ecological specialists in spite of the apparent advantages to being a generalist. This conundrum might be explained by adaptive trade-offs between performance on alternative hosts, yet evidence of such trade-offs has been difficult to find. Another hypothesis is that specialization is not adaptive, but rather the result of host-use trait loss and subsequent phylogenetic conservation of the reduced host breadth. Here we assessed the evidence for trade-offs and phylogenetic constraints in the evolution of host-plant use by conducting systematic surveys of host-use by armored scale insects (Hemiptera: Diaspididae) in two tropical rainforest plots. We found strong evidence for specialization in the host-ranges of individual diaspidid species and in the phylogenetic conservation of host-use across species, yet we found no evidence for performance trade-offs on alternative hosts. Instead, host-specialization in armored scale insects may represent the nonadaptive consequences of phylogenetic inertia. The phylogenetic and biogeographic history of insect species are therefore essential for understanding their ecological interactions.

3.2 Introduction

The immense diversity of life on Earth has been attributed to the partitioning of available resources into ecological niches (Hutchinson 1959), yet niche width varies greatly among species. In plant-feeding insects, a classic system for the study of ecological specialization, niches can range from a single host-plant species to sets of more than 100 host-plant families (Normark & Johnson 2011). In fact, although ecological specialists are prevalent, a continuous distribution of host breadths is observed in most insect taxa and habitats, with generalists rare but not absent (Forister *et al.* 2015). How this diversity of ecological niches is maintained, and what forces cause a species to become a specialist or a generalist, remains a conundrum for biologists.

One intuitive hypothesis holds that trade-offs between adaptations to alternative resources drive specialization (Futuyma & Moreno 1988). In fact, most theoretical models of specialization assume that rival alleles at individual genetic loci produce opposite fitness effects in alternative environments (antagonistic pleiotropy; Ravigné et al. 2009) even though empirical evidence for that kind of genetic trade-off is scarce (Futuyma 2008; Forister *et al.* 2012). However, trade-offs between alleles at host-related loci may be difficult to detect within individual populations because they can be hidden by inter-individual fitness variation at other loci (Joshi & Thompson 1995). Moreover, trade-offs could also arise from epistatic interactions between alleles at different loci, suggesting that a lineage's genomic background and evolutionary history influence its susceptibility to trade-offs (Remold 2012; Rodriguez-Verdugo *et al.* 2014; Celorio-Mancera *et al.* 2016).

Despite a historical focus on adaptive explanations for ecological specialization, niche breadth may be strongly influenced by non-adaptive processes (Futuyma *et al.* 1995; Gompert *et al.* 2015). In fact, theoretical spatial models have demonstrated that adaptation to particular plant resources is not necessary to produce a distribution of apparent specialists and generalists resembling the distributions observed in natural communities (Forister & Jenkins 2017). Nonadaptive processes are especially likely to reduce host breadth in insect species that are composed of many small, geographically isolated populations (Gompert *et al.* 2015; Hardy *et al.* 2016). Ultimately, adaptive and non-adaptive factors likely interact over multiple timescales to produce the observed empirical distribution of niche widths. Understanding host-breadth in plant-feeding insects therefore requires examination of short-term adaptive constraints along with long-term phylogenetic trends (Peterson *et al.* 2016).

Host-use trade-offs in plant-feeding insects have traditionally been investigated by comparing the performance of different insect genotypes within a population on multiple host-plant taxa, but a comparison of host-use traits across multiple insect species can offer a complementary perspective on the processes driving host-use evolution (Funk *et al.* 1995; Futuyma 2010; Hardy & Otto 2014; Peterson *et al.* 2015, 2016). In particular, a comparative approach can illuminate whether specialist species benefit from any performance advantages (relative to generalist species) that offset the disadvantages of their limited host repertoire. While comparing any individual specialist-generalist pair can be confounded by the evolutionary history of those two species, examining a large number of species with broadly distributed host range sizes should reveal the overall

relationship between host range size and performance on individual hosts. If specialization is an adaptive response to host-use trade-offs, we would expect that specialists, on average, outperform generalists on any given host. On the other hand, if specialization is a non-adaptive outcome of phylogenetic inertia, we would not expect to find any relationship between host range and host performance. Although such a high-level macroevolutionary approach does not capture the mechanistic details of the evolutionary pressures faced by any one species, it provides a direct test of competing adaptive and non-adaptive hypotheses for the evolutionary processes responsible for the prevalence of specialization in plant-feeding insects.

The armored scale insect family (Hemiptera: Diaspididae), which includes more than 2400 described species distributed throughout the world (García Morales *et al.* 2015), provides an ideal system in which to integrate long-term evolutionary and short-term ecological data concerning host-use in plant-feeding insects. Diaspidids are particularly abundant on woody hosts, and they have a simple, pathogen-like life history in which new host trees are colonized by wingless, wind-dispersed first-instar larvae (Gullan & Kosztarab 1997). Potential for host-choice is therefore quite limited, and host-range is mostly determined by host performance (Hill & Holmes 2009). These constraints create strong selection to use all locally present host-plants, an evolutionary situation quite different from the canonical insect herbivore that easily evolves selective host-use. Trade-offs and other adaptive constraints on host-use in diaspidids should therefore be readily apparent by comparing patterns of host-use between species. In fact, we have previously shown that trade-offs between presence on alternative hosts are not supported

42

by host-use observations associated with museum specimens (Peterson *et al.* 2015).

Nevertheless, that analysis did not test for quantitative performance trade-offs, which

require some measurement of performance on alternative hosts. Moreover, it remains

unclear whether armored scale insects specialize on particular host plants over small

geographic scales, given that their undirected dispersal should select strongly for the

ability to use all locally present hosts.

Here we quantified host-use in natural populations of diaspidids at tropical

rainforest sites on two continents. By systematically searching for diaspidids across all

canopy tree species present in each habitat and sequencing the collected insect specimens

at three genetic loci, we investigated the distribution of 172 diaspidid species across 138

tree species. We tested for ecological specialization in the use of particular host taxa by

comparing our empirical observations to those expected from a null model of random

host-use. Specialization was assessed at three taxonomic levels: host-tree species, genus,

and family.

To investigate whether any observed specialization could be attributed to

performance trade-offs between use of alternative host plants, we estimated the

relationship between diaspidid host-range and mean abundance per host. We treated

abundance as a proxy for performance on each host because the presence of sessile adult

female diaspidids and second-instar juveniles indicates successful feeding and growth on

that individual host tree (Hill & Holmes 2009). Moreover, most diaspidids settle on their

natal host plant, producing a patchy distribution of colonies on individual trees across the

habitat (Gullan & Kosztarab 1997). Diaspidid density appears to be primarily regulated

by parasitoids (Reeve & Murdoch 1986; Murdoch *et al.* 2005), which often decimate an entire local patch of diaspidids, while competition within or between diaspidid species appears to be weak due to their low overall density. The mean number of individuals produced in a local patch before it is found and destroyed by parasitoids is therefore a holistic and practical proxy for diaspidid performance on each host type in a natural setting. If host-use trade-offs are strong, we expected that generalist diaspidids would be less abundant than specialists when found on any given host tree.

However, because dispersal and colonization of empty patches is so important in a metapopulation of ephemeral patches (Metz & Gyllenberg 2001; Ronce 2007), we also estimated the relationship between diaspidid host-range and the proportion of host tree individuals colonized. If trade-offs arise from differences in the ability to colonize new host individuals or maintain living colonies, we expected that generalists would be present on a smaller proportion of trees belonging to each host taxon (even if they were present on more host trees overall).

3.3 Methods

We surveyed diaspidids at two tropical rainforest sites: San Lorenzo National Park, Panama and Lambir Hills National Park, Malaysia (on the island of Borneo). At each site, we used a crane to access the forest canopy, spending 20 person-minutes at each focal tree visually searching for diaspidids on all foliage accessible from a single location in the canopy of that tree. We were not able to search all the trees in each plot, so we used pre-existing databases of the trees at each site to divide the identified tree

individuals (those over 10 cm dbh) into sampling "rounds" of one randomly selected individual per tree species. We did not sample any tree individual more than once, so species with only one individual were present only in the first round of sampling, those with two individuals were present in the first two rounds, and so on. This protocol allowed us to sample across the full diversity of host taxa, while also getting multiple independent samples from common host types.

All plant material that appeared to be infested by diaspidids was placed in plastic bags in the field and brought back to the lab where individual live diaspidids were confirmed under magnification, cut out of the leaf or twig, and preserved in 95% ethanol. Specimens were subsequently sorted to life stage and second-instars and adult females were regarded as evidence of successful establishment. We then extracted DNA from all adult females and second-instars using Qiagen DNeasy Blood & Tissue kits (Qiagen, Valencia, CA) following the procedure outlined in Normark et al. (2014).

We amplified three genetic regions previously used for phylogenetic inference in diaspidids: the nuclear protein-coding gene elongation factor 1 α (EF1α), the D2 and D3 expansion segments of the large ribosomal subunit rDNA gene (28S), and a region of mitochondrial DNA encompassing portions of cytochrome c oxidase I and II (COI-II). PCR primers and protocols followed Andersen et al. (2010) and Gwiazdowski et al. (2011). PCR products were visualized using 1.5% agarose gels with SYBRsafe (Invitrogen, Carlsbad, CA, USA) and successful reactions were purified with Exo SAP-IT enzymatic digestion (Affymetrix, Cleveland, OH, USA). Sanger sequencing of the

PCR products was completed by Macrogen (Cambridge, MA, USA) or Eton Biosciences (San Diego, CA, USA).

To determine species identities and phylogenetic relationships within our diaspidid samples, we constructed a molecular phylogeny of all collected individuals. Alignments of each genetic locus were produced and iteratively refined using PASTA (Mirarab *et al.* 2014). Each alignment was then trimmed to include only sites with non-gap sequence for at least 80% of specimens (Capella-Gutiérrez *et al.* 2009), and we inferred a phylogeny for each locus using the GTR+CAT model in RAxML (Stamatakis 2014). The three single-locus alignments were then concatenated into a single alignment, from which we also inferred a phylogeny with RAxML. We delimited species in the final concatenated phylogeny by checking for the presence of each clade in at least two gene trees. All clades shared between gene trees, and not contradicted by the third gene tree, were considered evolutionarily independent. Species were provisionally defined as the largest clades within which no smaller clades were independent. This approach implements the genealogical concordance method of species delimitation (following Gwiazdowski et al. 2011). Next, we calculated the minimum branch-length divergence between these evolutionarily independent clades to estimate a maximum within-species branch-length distance. Any specimens separated by this distance were considered distinct species, which allowed us to determine whether isolated pairs of specimens should be considered conspecific. Finally, a pruned, ultrametric phylogeny with one tip per diaspidid species was produced using treePL (Smith & O'Meara 2012) for subsequent

phylogenetic analysis, calibrated with a diaspidid ancestral divergence date of 50-75

million years ago (Vea & Grimaldi 2016).

We looked for evidence of host-use specialization by diaspidids in two ways, and

at three levels of host taxonomy (species, genus, and family). First, we asked whether

there was evidence for specialization in the distribution of host taxa among the individual

host trees associated with each diaspidid species. If diaspidids are specialized for

particular hosts, the diversity of host taxa across the individual trees on which they are

observed should be lower than that expected under a model of random host-use with

respect to host taxa. We quantified host-tree diversity using Simpson's reciprocal

diversity index of host taxa, which can be thought of as the number of host taxa with a

correction for uneven distribution. To simulate data under the null model, we produced

1000 null datasets by randomly permuting the associations between diaspidid species and

individual host trees and again calculating the mean reciprocal diversity index of host

taxa. Second, we asked whether there was evidence for specialization when comparing

host-use between diaspidid species – specifically, whether use of each host taxon was

correlated with the diaspidid phylogeny. We calculated the phylogenetic signal of

diaspidid presence/absence on each host taxon along the diaspidid species phylogeny by

estimating the proportion of variance in the response variable explained by the insect

phylogeny (Hadfield & Nakagawa 2010). The empirical values for phylogenetic signal

were then compared to those calculated under a null model. In this case, null datasets

were produced by randomly swapping associations between diaspidid species and host

taxa until the associations were thoroughly shuffled (the number of random swaps was 10

times the overall number of associations). This null model preserved the empirical

distribution of generalist and specialist diaspidid species while randomizing the host-

taxon-by-diaspidid-species associations. *P*-values for the empirical phylogenetic signal

values were calculated using a *Z*-test against each parameter's null dataset values (which

were approximately normally distributed). We corrected for multiple comparisons by

assigning statistical significance according to a false discovery rate (FDR; Benjamini &

Hochberg 1995) of 0.05. The FDR procedure was conducted separately for each host-

taxon level because these analyses were not independent, and must be interpreted as

alternative configurations of the same data.

We investigated the strength of performance trade-offs between use of alternative

hosts by calculating the correlation between diaspidid host range (number of host taxa

used) and mean abundance per host tree on which at least one armored scale insect

individual was found. If performance trade-offs are strong, we expected that generalist

diaspidids would be less abundant than specialists on any given host tree. We also

investigated the relationship between host range and the proportion of host trees

colonized because patch occupancy rate may be a better indicator of performance than

local abundance in a metapopulation of discrete colonies (Gyllenberg & Metz 2001). For

each response variable separately (abundance per host and proportion of hosts colonized),

we fit a generalized linear model with host range as the predictor variable using the glm()

function in the R statistical environment (R Core Team 2015). For the local abundance

model, the response variable was the number of diaspidid individuals identified per host

tree, assuming a Poisson distribution. For the metapopulation performance model, the

response variable was the probability that an individual tree within each host taxon would be colonized by a diaspidid species, assuming a binomial distribution (and excluding host taxa with fewer than 3 trees surveyed). Both models only incorporated data for host-taxon-by-diaspidid associations with at least one record; thus there were no zeros in the data used to parameterize either model. This approach allowed us to avoid penalizing specialists for pursuing the potentially valid ecological strategy of maximizing performance on particular host taxa. To assess the statistical significance of each model's results, we compared the model parameters estimated from the empirical data with those from 1000 null datasets produced by randomly permuting the empirical data.

3.4 Results

In Panama, the first of two collection sites, we surveyed 90 trees over 3 rounds, representing 53 species, 48 genera, and 29 families. We found live diaspidids on 75 trees, yielding 380 second instar and adult female specimens. At least two loci were successfully amplified for 184 specimens, belonging to 53 species according to our method of species delimitation (Fig. S6; Table 3.1). In Malaysia, we surveyed 211 trees over 20 rounds, including 85 species, 48 genera, and 27 families. We found live diaspidids on 102 trees, yielding 480 second instar and adult female specimens. At least two loci were successfully amplified for 255 specimens, belonging to 119 species (Fig. S7).

Table 3.1: Trimmed genetic sequence alignment data for diaspidid specimens collected in Panama and Malaysia.

Location	Locus	Number of Specimens	Alignment Length (bp)	Proportion Missing	Proportion Variable sites	Proportion Parsimony Informative
Panama	28S	290	520	0.02	0.59	0.53
Panama	COI_II	212	745	0.00	0.74	0.68
Panama	EF1a	206	747	0.04	0.53	0.48
Malaysia	28S	359	611	0.04	0.55	0.47
Malaysia	COI_II	271	734	0.02	0.86	0.80
Malaysia	EF1a	234	798	0.08	0.53	0.45

In Panama, 29 of the 53 diaspidid species were only found on one host species, but 12 diaspidid species were observed on 3 or more different host plant species, including one extreme generalist observed on 18 host species in 14 families (Fig. 3.1). In Malaysia, 106 of the 119 diaspidid species were found on a single host plant species, and no diaspidid species was observed on more than three host plant species.
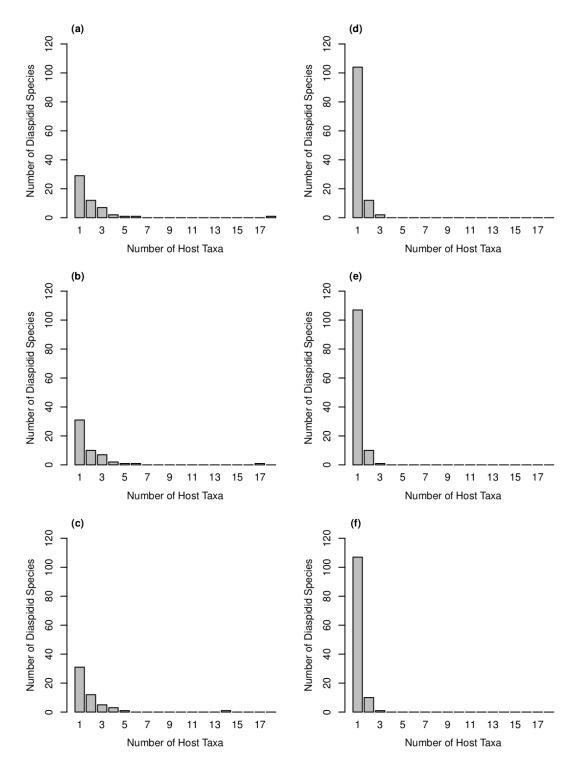
Figure 3.1. Histograms of the number of diaspidid species with each host-range size. Results are divided by location and host taxonomic level: a) Panama, species; b) Panama, genus; c) Panama, family; d) Malaysia, species; e) Malaysia, genus; f) Malaysia, family.

We found strong evidence for host-use specialization, both as a reduction of taxon diversity within the observed hosts of individual diaspidid species compared to a null model and as phylogenetic signals of host-use across diaspidid species. The Simpson's reciprocal diversity index of host taxa within each diaspidid species' observed host trees was lower than expected at all host taxonomic levels and in both locations, and this result was statistically significant at all host taxon levels and locations except at the host species level in Panama (Table 3.2). Phylogenetic signal was significantly different from its null expectation for 19 host taxa, and in all of those cases the signal was greater than expected (Fig. 3.2). Phylogenetic signal for host taxon use was generally higher at the Malaysia site (mean 0.61) than at the Panama site (mean 0.45), and 16 of the 19 host taxa showing a significant increase over their null phylogenetic signal expectations did so among Malaysian diaspidids.

Table 3.2: Mean Simpson's reciprocal diversity index (1/D) of individual host trees colonized by each diaspidid species for both sampling locations and all three host taxonomic levels.

Location	Taxon Level	Emp. 1/D	Null 1/D	Z	P
Panama	Species	3.162	3.321	-1.449	0.147
Panama	Genus	3.008	3.295	-2.721	0.007
Panama	Family	2.671	2.983	-2.887	0.004
Malaysia	Species	1.643	2.087	-9.902	< 0.001
Malaysia	Genus	1.461	1.955	-6.400	< 0.001
Malaysia	Family	1.461	1.785	-3.472	< 0.001
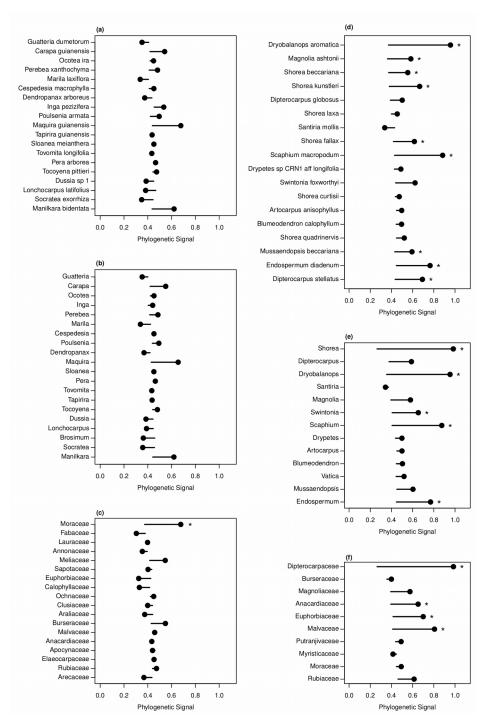
Figure 3.2. Phylogenetic signal of presence on each host taxon used by at least three diaspidid species. The dot indicates the empirical value, and the line is drawn to its expected value under the null model. Empirical values significantly different from the expected value (with a false discovery rate of 0.05) are marked with an asterisk. Results are divided by location and host taxonomic level: a) Panama, species; b) Panama, genus; c) Panama, family; d) Malaysia, species; e) Malaysia, genus; f) Malaysia, family.

Despite the prevalence of host-use specialization among diaspidids at our two sites, we found no evidence for trade-offs between performance on alternative hosts. Specialists were not more abundant than generalists per tree; the number of live adult or second instar female diaspidids of a particular species found on each tree was not significantly correlated with the host-range of that diaspidid species ($P > 0.9$ in all tests; Fig. 3.3; Table 3.3). Moreover, we did not observe the negative correlation between diaspidid host range and the proportion of host trees colonized expected under a metapopulation trade-off model. In fact, we observed the opposite pattern: diaspidids with larger observed host ranges were observed on a higher proportion of the trees in their host taxa than were diaspidids apparently specialized on those taxa (Fig. 3.4; Table 3.4).

Table 3.3: Statistical results from the models relating abundance per host to the local host range of each diaspidid species.

Location	Taxon Level	Slope	Z	P
Panama	Species	0.000	0.059	0.953
Panama	Genus	0.000	0.070	0.944
Panama	Family	0.001	0.074	0.941
Malaysia	Species	0.007	0.098	0.922
Malaysia	Genus	0.006	0.115	0.909
Malaysia	Family	0.006	0.111	0.912

Table 3.4: Statistical results from the models relating proportion of colonized host tree individuals to the local host range of each diaspidid species.

Location	Taxon Level	Slope	Z	P
Panama	Species	0.030	2.616	0.009
Panama	Genus	0.036	3.125	0.002
Panama	Family	0.052	2.908	0.004
Malaysia	Species	0.361	2.077	0.038
Malaysia	Genus	0.455	1.867	0.062
Malaysia	Family	0.765	5.381	< 0.001
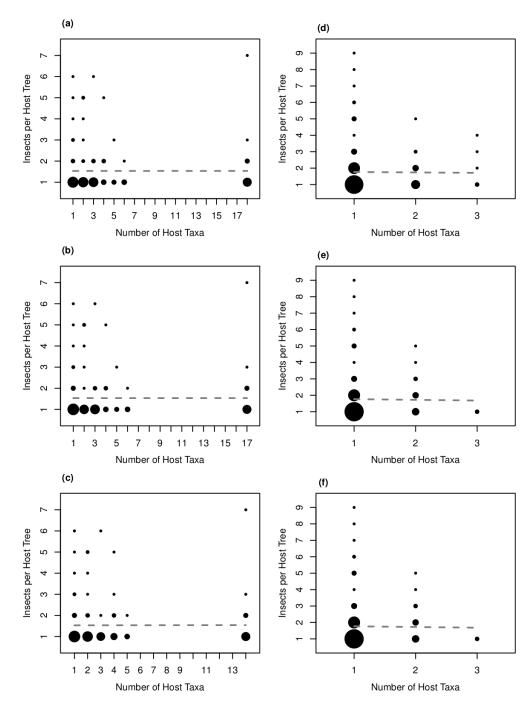
Figure 3.3. Scatter plot of the number of diaspidid individuals per colonized host by the local host-range of that diaspidid species. Dot area is proportional to the number of data points at that location. Results are divided by location and host taxonomic level: a) Panama, species; b) Panama, genus; c) Panama, family; d) Malaysia, species; e) Malaysia, genus; f) Malaysia, family. None of these relationships (as fitted by a linear model, dashed line) was statistically different from expectations under a null model (all $P > 0.9$).

56
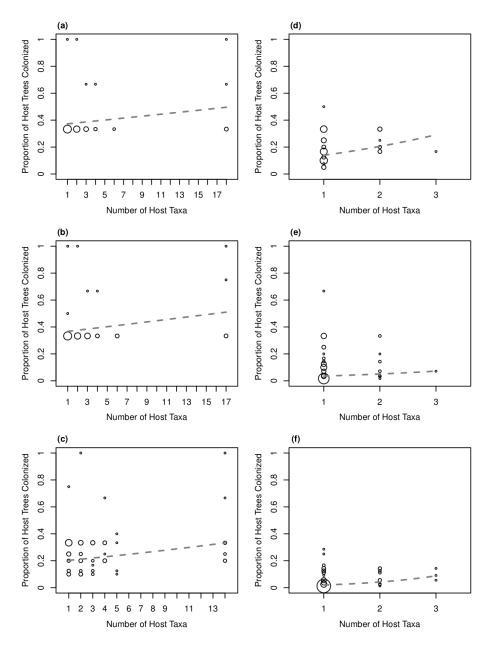
Figure 3.4. Scatter plot of the proportion of host trees colonized by the local host-range of that diaspidid species. Each observed host-taxon-by-diaspidid-species interaction is plotted independently, although host taxa with fewer than three tree individuals surveyed were excluded from this analysis. Circle area is proportional to the number of data points at that location. Results are divided by location and host taxonomic level: a) Panama, species; b) Panama, genus; c) Panama, family; d) Malaysia, species; e) Malaysia, genus; f) Malaysia, family. All fitted slopes (dashed lines) were positive and all were statistically significant ($P < 0.05$), except in Malaysia by host genus ($P = 0.062$).

3.5 Discussion

Through our systematic surveys of armored scale insect host-use in two tropical rainforest canopy habitats, we found evidence that most armored scale insect species use only a small proportion of the host-plant taxa present in their local environment. However, this distribution does not appear to be adaptive, as specialists were no more abundant on their host plants than were generalists. If adaptive performance trade-offs are not the main force driving specialization in this group, host breadth may instead be limited by non-adaptive processes playing out over macroevolutionary timescales (Gompert *et al.* 2015; Hardy *et al.* 2016; Forister & Jenkins 2017).

It is important to note that our sampling of host trees was very limited relative to the number of diaspidid species observed at each site (particularly in Malaysia), and that our observed host-ranges may therefore be much smaller than the true host-ranges for many of these species. Nevertheless, we observed no apparent specialists that were particularly abundant on their individual host trees, suggesting either that there are no true specialists, or that trade-offs do not have a strong effect on host performance. In fact, the highest single-tree abundance of any diaspidid species was claimed by the most extreme generalist found in Panama, with a host range of 18 plant species in 14 families. We therefore conclude that performance trade-offs are not strong in this system, and are unlikely to drive the evolution of specialization. This conclusion corroborates our previous phylogenetic analysis of host-use evolution in North American diaspidids (Peterson *et al.* 2015).

Another striking result from this study was that many host-use traits displayed a strong phylogenetic signal across the diversification of diaspidid species. Although this finding matches the phylogenetic conservatism of host-use observed in many comparative studies of plant-feeding insects, including butterflies (Janz *et al.* 2001), beetles (Kelley & Farrell 1998), and aphids (Hardy *et al.* 2015), phylogenetic conservatism of host specialization takes on increased significance when it is uncoupled from directed dispersal between hosts. For diaspidids, nearly all of which colonize new hosts haphazardly via wind (Magsig-Castillo *et al.* 2010), a phylogenetic signal of specialization indicates a constraint on the evolution of host-use – an inability to adapt to additional hosts despite a significant fitness cost (Hill & Holmes 2009). Moreover, we found no evidence for performance trade-offs between alternative hosts in this study or in our previous phylogenetic work (Peterson *et al.* 2015), suggesting that the phylogenetic host-use constraints observed among armored scale insects persist in the face of strong selection for increased host-range. Nevertheless, a few diaspidid species are known to have huge host-ranges and broad geographic distributions (Normark & Johnson 2011), indicating that the generalist lifestyle is a biological possibility. Thus, our results suggest that while host-use constraints may not be adaptive, they can define host-use for insect lineages over micro- and macroevolutionary timescales.

The lack of evidence for adaptive specialization in diaspidids suggests an important role for nonadaptive processes in host-use evolution – if diaspidid populations are small and isolated from each other, they may rapidly lose the functional alleles required to use hosts that are absent from their environment. The rarity of evolutionary

transitions from specialist to generalist is supported by our recent global study of diaspidid host-use, which found that generalist lineages appear to be the long-lived "trunks" of the diaspidid phylogeny, with specialist lineages representing short-lived, dead-end branches (Hardy *et al.* 2016). The "specialization-as-dead-end" model has received mixed support in other groups (Day *et al.* 2016), but it seems clear that non-adaptive processes can be a significant constraint on host-range in plant-feeding insects. Although we observed here what appear to be radiations of specialist species among Malaysian diaspidids, those lineages are likely trapped in the local habitat due to their dependence on particular hosts. In fact, the water barriers around the island of Borneo may offer protection from invasive generalist species, allowing the specialized local fauna to persist in spite of their ecological disadvantages and increasing the rate of gene loss through drift. The prevalence of dipterocarps within the Bornean forest also increases the likelihood that dipterocarp specialists can persist. Overall, the geographic history of a diaspidid lineage may be the most important determinant of what hosts it can use.

Our results can also help to illuminate the complexity of host-use traits in plant-feeding insects (Forister *et al.* 2012; Barrett & Heil 2012). We found that specialization in armored scale insects occurs at all three of the host-taxonomic levels that we considered (species, genus, and family), suggesting that the genomic architecture of host-use traits is both complex and hierarchical. Use of multiple hosts is often associated with close phylogenetic relationships among those hosts (Gilbert & Webb 2007; Krasnov *et al.* 2012), yet such results in flying insects may reflect host-preference or ease of host recognition more than host performance (Bernays 2001). The particular natural history of

diaspidids depends upon host-performance much more than host-preference in determining realized host-use, so the importance of multiple taxonomic levels implies that performance on any given host likely depends on a large number of traits of various effect sizes. In fact, the involvement of many genetic loci in plant-insect interactions is corroborated by both ecological (Singer & Stireman 2005) and genetic (Remold 2012) theory, yet the actual mechanistic basis of host performance is not well understood in scale insects or even in the closely related aphids. Previous work suggests that effector/inhibitor dynamics play a significant role in determining feeding success for sucking insects (Hogenhout & Bos 2011; Ali & Agrawal 2012), but how evolutionary dynamics in those interactions have produced the patterns of host-use that we observe in the wild remains obscure.

APPENDIX A

SUPPLEMENTARY METHODS

Diaspidid Phylogeny Reconstruction

We constructed a phylogeny of Diaspididae using Genbank sequences published by Morse & Normark (2006) and Andersen et al. (2010) as well as additional sequences obtained from specimens held by the University of Massachusetts Insect Collection.

Individual specimens with small amounts of host plant material were preserved in 100% ethanol and stored at -20 °C freezer until further processed. Each specimen was subjected to a joint molecular/morphological preparation protocol that results in genomic DNA from a single individual specimen that corresponded to a permanent slide-mount of the cuticle. Total genomic DNA from individual specimens was isolated using Qiagen DNeasy Blood & Tissue kit (Qiagen, Valencia, California). In order to facilitate digestion, each insect was punctured with a 000 entomological pin on its abdominal region before being placed in the lysis solution. Upon digestion, individual cuticles were retrieved from the lysate using a sterile micropipette tip and slide mounted as voucher specimens for later morphological identification. The remainder of the Qiagen protocol was followed, except at the first elution only 60μl of Buffer AE was used.

Polymerase chain reaction (PCR) was performed to amplify regions of the mitochondrial genes cytochrome c oxidase I and II (COI and COII, ~800 bp), nuclear protein-coding gene elongation factor 1-alpha (EF1α, ~1150 bp), and the D2 and D3 expansion segments of the large subunit ribosomal RNA (28S, ~800 bp). Standard amplification protocols were followed using oligonucleotide primers as reported by

Morse & Normark (2006) and Andersen et al. (2010). Amplification products were visualized using 1.5% agarose gels stained with SYBRsafe™ (Invitrogen, Carlsbad, California) in 1X TBE and purified with an ExoSAP-IT PCR Product Clean-Up enzyme digest (Affymetrix, Cleveland, Ohio).

PCR products were sequenced directly using an ABI-3130XL Genetic Analyzer at University of Massachusetts Genomics Resource Laboratory. Sequence fragments were assembled and edited using Sequencher 4.9 (Gene Codes Corporation, Ann Arbor, Michigan). Datasets for all genes were trimmed at the 5' and 3' ends such that there was no missing data. No insertions or deletions were present in the coding regions of COI and COII, which were aligned using amino acid sequences. Four separate introns of EF1α marked by starting position (GT) and ending position (AG) were removed from the analysis and exons were aligned based on the amino acid sequences. Hyper-variable regions of D2 and D3 expansion segments of 28S were unalignable without reliable homology assessment, and therefore removed from the analysis and the remaining conserved regions were aligned using MAFFT (Katoh *et al.* 2005). We used RAxML v8.0.23 (Stamatakis 2014) in rapid hill-climbing mode to estimate the diaspidid phylogeny, using a GTR model of nucleotide substitution plus CAT approximation of among-site rate variation with 25 categories.

Lepidoptera and Hemiptera Phylogeny Reconstruction

Phylogenetic datasets were assembled from published DNA sequence data using the PHLAWD megaphylogeny pipeline (Smith *et al.* 2009). The Lepidoptera dataset

consisted of 7470 sites sampled from 9 loci across 1604 species. The Hemiptera dataset

comprised 9,015 sites sampled from 15 loci across 955 species. For each taxon

(Lepidoptera and Hemiptera), we used the NCBI taxonomy as a constraint tree in a ML

search under a GTR nucleotide substitution model with CAT-approximated among-site

rate variation, with model parameters estimated independently for each locus. Using

RAxML (Stamatakis 2014), we estimated phylogenies from 100 non-parametric

bootstrap replicates of the multiple sequence alignment supermatrix and then used every

fifth bootstrap tree as the starting tree for optimization of the empirical data. We scaled

the branch lengths of the ML tree to time using Penalized Likelihood, assuming an

autocorrelated model of among-lineage rate variation and selecting a value for the

smoother parameter with a cross-validation procedure (treePL; Smith and O'Meara

2012).

We calibrated the Lepidoptera divergence time estimates with three constraints:

the age of the root, which corresponds to the crown node of Ditrysia, was fixed at 140

Ma, to match the the published estimate of Wahlberg et al. (2013). Based on estimates of

the ages of fossil taxa (Sohn *et al.* 2012), a uniform constraint with a minimum age of

100 and a maximum age of 140 Ma was placed on the crown node of Gracillariidae

(Labandeira *et al.* 1994), and a uniform constraint with a minimum age of 56 and a

maximum age of 140 Ma was placed on the crown node of Nymphalidae (Kristensen &

Skalski 1998). We calibrated Hemiptera divergence time estimates with five constraints.

The age of the Hemiptera root was fixed 291 Ma, to match the published median estimate

of Misof et al. (2014). A minimum age of 240 and maximum age of 291 Ma were

imposed on the crown node of Aphidomorpha (Szwedo & Nel 2011). A minimum age of 99 and a maximum age of 291 were placed on the crown node of Pemphiginae and that of Coccidae (Kononova 1977; Vea & Grimaldi 2015). A minimum age of 142 and a maximum age of 291 were placed on the crown node of Heteroptera (Popov 1971).

APPENDIX B

SUPPLEMENTARY FIGURES

Figure S1. Phylogeny of armored scale insect species inferred with RaxML.
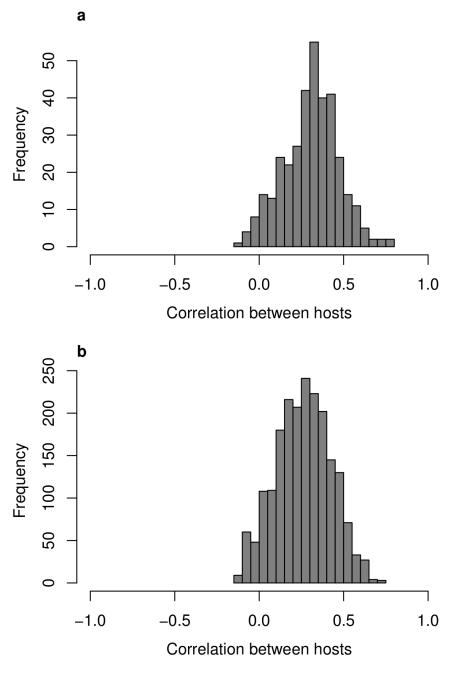
Figure S2. Distribution of correlations between diaspidid presences and absences on pairs of focal host plant species (a) and genera (b) using the Pearson product-moment correlation calculation (uncorrected for phylogeny).
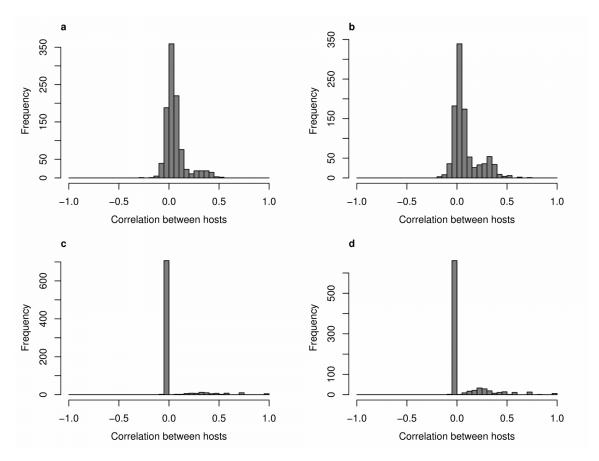
Figure S3. Distribution of phylogenetic correlations between diaspidid presences and absences on 1000 randomly selected pairs of host plant species (a) and genera (b) and Pearson product-moment correlations between diaspidid presences and absences on the same pairs of host plant species (c) and genera (d).

69
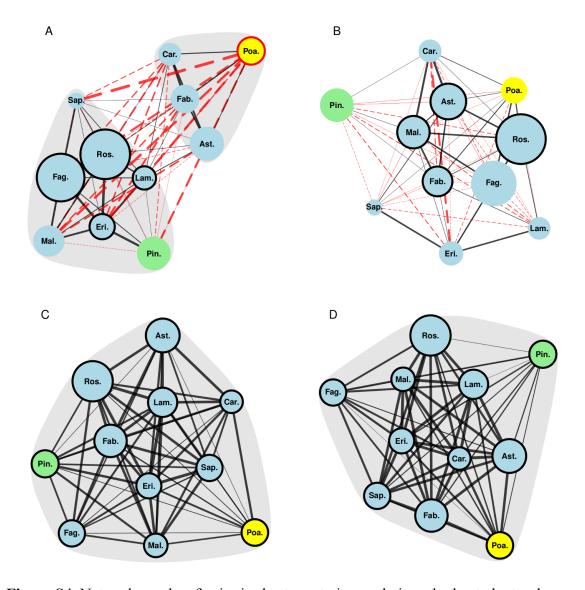
Figure S4. Network graphs of pairwise host-use trait correlations, by host plant order, calculated using plant-insect interaction matrices in which insect species were counted as using a given host order only if they were observed on at least two genera in that order. (A) Lepidoptera – phylogenetic correlations. (B) Lepidoptera – residual correlations. (C) Hemiptera – phylogenetic correlations. (D) Hemiptera – residual correlations. Each vertex represents a host order, with vertex area proportional to the number of insects that were observed on that host. Positive interactions between presence on a pair of hosts are represented by solid, black lines and negative correlations by dashed, red lines, with line thickness proportional to the magnitude of the correlation. Network spatial structure was determined using the Kamada-Kawai (1989) algorithm, a force-directed layout method in which "repulsion" between vertices was proportional to the inverse of one plus the correlation values between the respective hosts. Vertices are labeled with the following abbreviations – Ast.: Asterales, Car.: Caryophyllales, Eri.: Ericales, Fab.: Fabales, Fag.:

Fagales, Lam.: Lamiales, Mal.: Malpighiales, Pin.: Pinales, Poa.: Poales, Ros.: Rosales, Sap.: Sapindales. Vertices are colored by taxonomic group – eudicots: blue, monocots: yellow, conifers: green. Statistically significant clusters ($P < 0.05$) are indicated by gray bubbles. Individual host orders with mean correlations of significantly higher magnitude than expected ($P < 0.05$) are indicated by bold vertex outlines (black for positive means, red for negative means).

Figure S5. Network graphs of pairwise host-use trait correlations, by host plant family. (A) Lepidoptera – phylogenetic correlations. (B) Lepidoptera – residual correlations. (C) Hemiptera – phylogenetic correlations. (D) Hemiptera – residual correlations. Each vertex represents a host order, with vertex area proportional to the number of insects that were observed on that host. Positive interactions between presence on a pair of hosts are represented by solid, black lines and negative correlations by dashed, red lines, with line thickness proportional to the magnitude of the correlation. Network spatial structure was determined using the Kamada-Kawai (1989) algorithm, a force-directed layout method in which "repulsion" between vertices was proportional to the inverse of one plus the correlation values between the respective hosts. Vertices are labeled with the following abbreviations – Ast.: Asteraceae, Bet.: Betulaceae, Eri.: Ericaceae, Fab.: Fabaceae, Fag.: Fagaceae, Pin.: Pinaceae, Poa.: Poaceae, Ros.: Rosaceae, Sal.: Salicaceae, Sap.: Sapindaceae. Vertices are colored by taxonomic group – eudicots: blue, monocots:

yellow, conifers: green. Statistically significant clusters ($P < 0.05$) are indicated by gray bubbles. Individual host orders with mean correlations of significantly higher magnitude than expected ($P < 0.05$) are indicated by bold vertex outlines (black for positive means, red for negative means).

Figure S6. Phylogeny of diaspidid species sampled in Panama. Labels indicate the host plant species (and families) used by each diaspidid species.

Shorea curtisii (Dipterocarpaceae)
Magnolia ashtonii (Magnoliaceae)
Santiria mollis (Burseraceae)
Artocarpus anisophyllus (Moraceae)
Myristica gigantea (Myristicaceae)
Dacryodes incurvata (Burseraceae)
Magnolia ashtonii (Magnoliaceae)
Magnolia ashtonii (Magnoliaceae)
Magnolia ashtonii (Magnoliaceae)
Endospermum diadenum (Euphorbiaceae)
Endospermum diadenum (Euphorbiaceae)
Shorea laxa (Dipterocarpaceae)
Shorea laxa (Dipterocarpaceae)
Dryobalanops aromatica (Dipterocarpaceae)
Shorea kunstleri (Dipterocarpaceae)
Shorea crassa (Dipterocarpaceae)
Santiria griffithii (Burseraceae)
Swintonia sp CRN1 (Anacardiaceae)
Swintonia foxworthyi (Anacardiaceae)
Santiria mollis (Burseraceae)
Swintonia foxworthyi (Anacardiaceae)
Magnolia ashtonii (Magnoliaceae)
Allanthospermum borneense (Ixonanthaceae)
Pouteria malaccensis (Sapotaceae)
Durio lanceolatus (Malvaceae)
Gironniera hirta (Cannabaceae)
Santiria laevigata (Burseraceae)
Gironniera hirta (Cannabaceae)
Shorea laxa (Dipterocarpaceae)
Dipterocarpus globosus (Dipterocarpaceae)
Swintonia foxworthyi (Anacardiaceae)
Dialium sp CRN2 aff trifoliolatum (Fabaceae)
Mastixia sp CRN1 aff pentandra (Cornaceae)
Artocarpus anisophyllus (Moraceae)
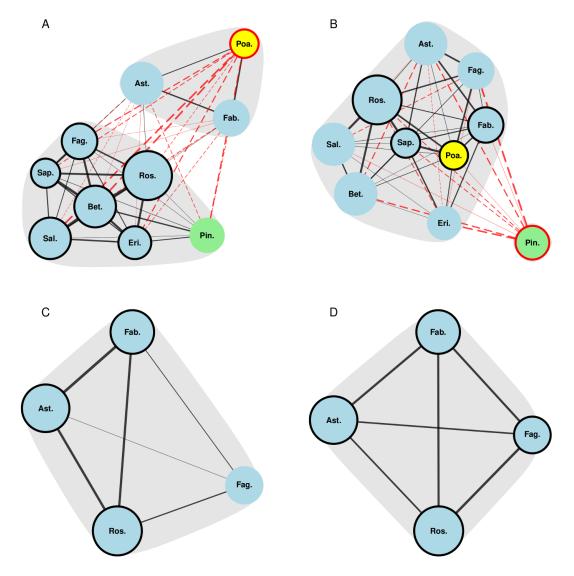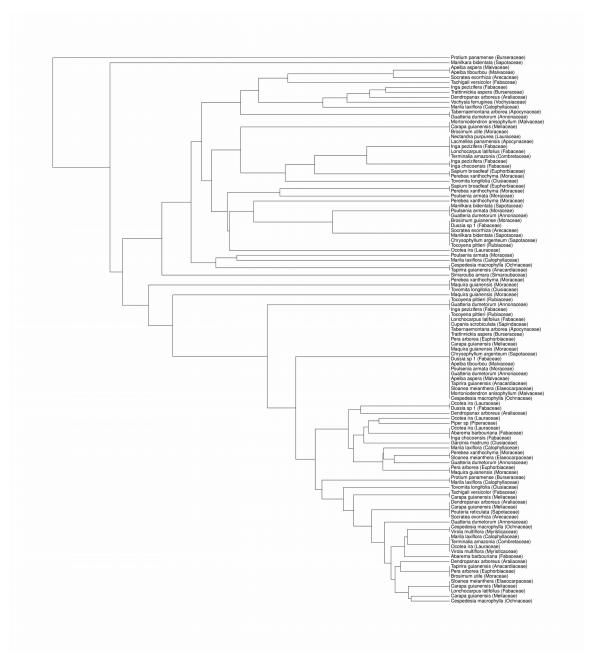Dipterocarpus globosus (Dipterocarpaceae)
Swintonia foxworthyi (Anacardiaceae)
Dipterocarpus globosus (Dipterocarpaceae)
Dipterocarpus stellatus (Dipterocarpaceae)
Magnolia ashtonii (Magnoliaceae)
Artocarpus anisophyllus (Moraceae)
Canarium apertum (Burseraceae)
Shorea acuta (Dipterocarpaceae)
Shorea sp cf ovata (Dipterocarpaceae)
Shorea fallax (Dipterocarpaceae)
Shorea beccariana (Dipterocarpaceae)
Shorea curtisii (Dipterocarpaceae)
Shorea sp cf ovata (Dipterocarpaceae)
Calophyllum sp aff soulattri (Calophyllaceae)
Dipterocarpus globosus (Dipterocarpaceae)
Shorea beccariana (Dipterocarpaceae)
Drypetes sp CRN1 aff longifolia (Putranjivaceae)
Santiria laevigata (Burseraceae)
Shorea fallax (Dipterocarpaceae)
Shorea laxa (Dipterocarpaceae)
Shorea kunstleri (Dipterocarpaceae)
Dacryodes incurvata (Burseraceae)
Gymnacranthera bancana (Myristicaceae)
Shorea beccariana (Dipterocarpaceae)
Shorea fallax (Dipterocarpaceae)
Vatica oblongifolia (Dipterocarpaceae)
Vatica micrantha (Dipterocarpaceae)
Shorea fallax (Dipterocarpaceae)
Santiria mollis (Burseraceae)
Santiria griffithii (Burseraceae)
Scaphium macropodum (Malvaceae)
Blumeodendron calophyllum (Euphorbiaceae)
Xanthophyllum brevipes (Polygalaceae)
Ctenolophon parvifolius (Ctenolophonaceae)
Santiria grandiflora (Burseraceae)
Blumeodendron calophyllum (Euphorbiaceae)
Palaquium sp CP2 black–rusty (Sapotaceae)
Santiria mollis (Burseraceae)
Gymnacranthera bancana (Myristicaceae)
Scaphium macropodum (Malvaceae)
Magnolia ashtonii (Magnoliaceae)
Scaphium macropodum (Malvaceae)
Diospyros dictyoneura (Ebenaceae)
Endiandra sp aff srobiculata (Lauraceae)
Dipterocarpus stellatus (Dipterocarpaceae)
Dipterocarpus stellatus (Dipterocarpaceae)
Scaphium macropodum (Malvaceae)
Endospermum diadenum (Euphorbiaceae)
Allanthospermum borneense (Ixonanthaceae)
Magnolia ashtonii (Magnoliaceae)
Scaphium macropodum (Malvaceae)
Calophyllum sp aff soulattri (Calophyllaceae)
Magnolia ashtonii (Magnoliaceae)
Mussaendopsis beccariana (Rubiaceae)
Mussaendopsis beccariana (Rubiaceae)
Shorea kunstleri (Dipterocarpaceae)
Shorea beccariana (Dipterocarpaceae)
Drypetes sp CRN1 aff longifolia (Putranjivaceae)
Shorea kunstleri (Dipterocarpaceae)
Shorea quadrinervis (Dipterocarpaceae)
Shorea beccariana (Dipterocarpaceae)
Shorea excelliptica (Dipterocarpaceae)
Shorea crassa (Dipterocarpaceae)
Shorea excelliptica (Dipterocarpaceae)
Dryobalanops aromatica (Dipterocarpaceae)
Dryobalanops aromatica (Dipterocarpaceae)
Swintonia acuta (Anacardiaceae)
Mussaendopsis beccariana (Rubiaceae)
Santiria mollis (Burseraceae)
Vatica oblongifolia (Dipterocarpaceae)
Blumeodendron calophyllum (Euphorbiaceae)
Dipterocarpus globosus (Dipterocarpaceae)
Shorea fallax (Dipterocarpaceae)
Dipterocarpus globosus (Dipterocarpaceae)
Shorea laxa (Dipterocarpaceae)
Shorea acuta (Dipterocarpaceae)
Shorea beccariana (Dipterocarpaceae)
Shorea beccariana (Dipterocarpaceae)
Shorea kunstleri (Dipterocarpaceae)
Drypetes sp CRN1 aff longifolia (Putranjivaceae)
Shorea quadrinervis (Dipterocarpaceae)
Shorea beccariana (Dipterocarpaceae)
Shorea kunstleri (Dipterocarpaceae)
Shorea curtisii (Dipterocarpaceae)
Drypetes sp CRN1 aff longifolia (Putranjivaceae)
Shorea kunstleri (Dipterocarpaceae)
Shorea quadrinervis (Dipterocarpaceae)
Shorea bullata (Dipterocarpaceae)
Swintonia sp CRN1 (Anacardiaceae)
Dryobalanops aromatica (Dipterocarpaceae)
Dryobalanops aromatica (Dipterocarpaceae)
Dryobalanops aromatica (Dipterocarpaceae)
Dryobalanops aromatica (Dipterocarpaceae)
Dryobalanops aromatica (Dipterocarpaceae)
Dryobalanops aromatica (Dipterocarpaceae)
Dryobalanops aromatica (Dipterocarpaceae)
Dipterocarpus globosus (Dipterocrarpaceae)
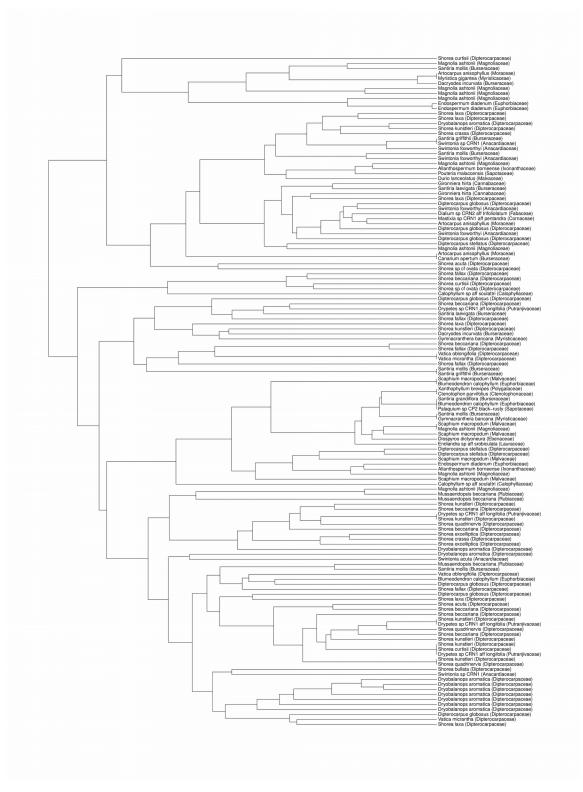Vatica micrantha (Dipterocarpaceae)
Shorea laxa (Dipterocarpaceae)

Figure S7. Phylogeny of diaspidid species sampled in Malaysia. Labels indicate the host plant species (and families) used by each diaspidid species.

BIBLIOGRAPHY

Agrawal AA, Conner JK, Rasmann S (2010) Tradeoffs and Negative Correlations in Evolutionary Ecology. In: *Evolution Since Darwin: the First 150* , pp. 243–268.

Ali JG, Agrawal AA (2012) Specialist versus generalist insect herbivores and plant defense. *Trends in Plant Science*, **17**, 293–302.

Andersen JC (2009) A phylogenetic analysis of armored scale insects, based upon nuclear, mitochondrial, and endosymbiont gene sequences. University of Massachusetts, Amherst.

Andersen JC, Wu J, Gruwell ME *et al.* (2010) A phylogenetic analysis of armored scale insects (Hemiptera: Diaspididae), based upon nuclear, mitochondrial, and endosymbiont gene sequences. *Molecular Phylogenetics and Evolution*, **57**, 992–1003.

Barrett LG, Heil M (2012) Unifying concepts and mechanisms in the specificity of plant-enemy interactions. *Trends in Plant Science*, **17**, 282–292.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B*, **57**, 289–300.

Berenbaum MR, Feeny PP (2008) Chemical mediation of host-plant specialization: the papilionid paradigm. In: *Specialization, Speciation, and Radiation: The Evolutionary Biology of Herbivorous Insects* , pp. 3–19.

Bernays EA (2001) Neural limitations in phytophagous insects: implications for diet breadth and evolution of host affiliation. *Annual Review of Entomology*, **46**, 703–727.

Boyle B, Hopkins N, Lu Z *et al.* (2013) The taxonomic name resolution service: an online tool for automated standardization of plant names. *BMC Bioinformatics*, **14**, 16.

Campbell SA, Kessler A (2013) Plant mating system transitions drive the macroevolution of defense strategies. *Proceedings of the National Academy of Sciences*, **110**, 3973–3978.

Canard EF, Mouquet N, Mouillot D *et al.* (2014) Empirical evaluation of neutral interactions in host-parasite networks. *The American Naturalist*, **183**, 468–79.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**, 1972–1973.

Celorio-Mancera M de la P, Wheat CW, Huss M *et al.* (2016) Evolutionary history of host use, rather than plant phylogeny, determines gene expression in a generalist butterfly. *BMC Evolutionary Biology*, **16**, 59.

Csárdi G, Nepusz T (2006) The igraph software package for complex network research. *InterJournal Complex Systems*, **1695**, 1–9.

Day EH, Hua X, Bromham L (2016) Is specialization an evolutionary dead-end? Testing for differences in speciation, extinction and trait transition rates across diverse phylogenies of specialists and generalists. *Journal of Evolutionary Biology*, **29**, 1257–1267.

Dyer LA, Singer MS, Lill JT *et al.* (2007) Host specificity of Lepidoptera in tropical and temperate forests. *Nature*, **448**, 696–699.

Elena SF, Lenski RE (2003) Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nature reviews. Genetics*, **4**, 457–69.

Farrell BD (1998) "Inordinate fondness" explained: why are there so many beetles? *Science*, **281**, 555–559.

Favret C (2015) Aphid species file. http://aphid.speciesfile.org.

Felsenstein J (2012) A comparative method for both discrete and continuous characters using the threshold model. *The American Naturalist*, **179**, 145–156.

Ferris GF (1942) *Atlas of the Scale Insects of North America*. Stanford University Press, Palo Alto, CA.

Forister ML, Dyer LA, Singer MS, Stireman JO, Lill JT (2012) Revisiting the evolution of ecological specialization, with emphasis on insect-plant interactions. *Ecology*, **93**, 981–991.

Forister ML, Jenkins SH (2017) A neutral model for the evolution of diet breadth. *The American Naturalist*, **190,** e40-e54.

Forister ML, Novotny V, Panorska AK *et al.* (2015) The global distribution of diet breadth in insect herbivores. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 442–7.

Franklin DC (2005) Vegetative phenology and growth of a facultatively deciduous bamboo in a monsoonal climate. *Biotropica*, **37**, 343–350.

Fry J (1996) The evolution of host specialization: are trade-offs overrated? *The American Naturalist*, **148**, S84-107.

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

Futuyma DJ (1976) Food plant specialization and environmental predictability in Lepidoptera. *The American Naturalist*, **110**, 285–292.

Futuyma DJ (2008) Sympatric speciation: norm or exception? In: *Specialization, Speciation, and Radiation: The Evolutionary Biology of Herbivorous Insects*, pp. 136–147.

Futuyma DJ (2010) Evolutionary constraint and ecological consequences. *Evolution*, **64**, 1865–84.

Futuyma DJ, Agrawal AA (2009) Macroevolution and the biological diversity of plants and herbivores. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 18054–18061.

Futuyma DJ, Keese M, Funk DJ (1995) Genetic constraints on macroevolution: the evolution of host affiliation in the leaf beetle genus *Ophraella*. *Evolution*, **49**, 797–809.

Futuyma DJ, Moreno G (1988) The evolution of ecological specialization. *Annual Review of Ecology and Systematics*, **19**, 207–233.

García Morales M, Denno B, Miller DR *et al.* (2015) ScaleNet: A literature-based model of scale insect biology and systematics.

Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**, 457–472.

Gilbert GS, Webb CO (2007) Phylogenetic signal in plant pathogen-host range. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 4979–83.

Gompert Z, Jahner JP, Scholl CF *et al.* (2015) The evolution of novel host use is unlikely to be constrained by trade-offs or a lack of genetic variation. *Molecular Ecology*, **24**, 2777–2793.

Gullan PJ, Kosztarab M (1997) Adaptations in Scale Insects. *Annual Review of Entomology*, **42**, 23–50.

Gwiazdowski RA, Vea IM, Andersen JC, Normark BB (2011) Discovery of cryptic species among North American pine-feeding Chionaspis scale insects (Hemiptera : Diaspididae). *Biological Journal of the Linnean Society*, **104**, 47–62.

Gyllenberg M, Metz J (2001) On fitness in structured metapopulations. *Journal of Mathematical Biology*, **43**, 545–60.

Hadfield JD (2010) MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. *Journal of Statistical Software*, **33**, 1–22.

Hadfield JD, Krasnov BR, Poulin R, Nakagawa S (2014) A tale of two phylogenies: comparative analyses of ecological interactions. *The American Naturalist*, **183**, 174–87.

Hadfield JD, Nakagawa S (2010) General quantitative genetic methods for comparative biology: Phylogenies, taxonomies and multi-trait models for continuous and categorical characters. *Journal of Evolutionary Biology*, **23**, 494–508.

Hardy NB, Otto SP (2014) Specialization and generalization in the diversification of phytophagous insects: tests of the musical chairs and oscillation hypotheses. *Proceedings of the Royal Society B*, **281**, 20132960.

Hardy NB, Peterson DA, von Dohlen CD (2015) The evolution of life cycle complexity in aphids: Ecological optimization or historical constraint? *Evolution*, **69**, 1423–1432.

Hardy NB, Peterson DA, Normark BB (2016) Nonadaptive radiation: pervasive diet specialization by drift in scale insects? *Evolution*, **70**, 2421–2428.

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

Hill MG, Holmes T (2009) An analysis of latania scale (*Hemiberlesia lataniae*) crawler settlement behaviour on kiwifruit leaves and bark. *New Zealand Plant Protection*, **62**, 56–62.

Hogenhout SA, Bos JIB (2011) Effector proteins that modulate plant-insect interactions. *Current Opinion in Plant Biology*, **14**, 422–428.

Hutchinson GE (1959) Homage to Santa Rosalia or why are there so many kinds of animals? *The American Naturalist*, **93**, 145–159.

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

Janz N, Nylin S (1998) Butterflies and plants: a phylogenetic study. *Evolution*, **52**, 486–502.

Janz N, Nylin S (2008) The oscillation hypothesis of host-plant range and speciation. In: *Specialization, Speciation, and Radiation: The Evolutionary Biology of Herbivorous Insects* , pp. 203–215.

Janz N, Nylin S, Wahlberg N (2006) Diversity begets diversity: host expansions and the diversification of plant-feeding insects. *BMC Evolutionary Biology*, **6**, 4.

Johnson MTJ, Ives AR, Ahern J, Salminen JP (2014) Macroevolution of plant defenses against herbivores in the evening primroses. *New Phytologist*, **203**, 267–279.

Joshi A, Thompson JN (1995) Trade-offs and the evolution of host specialization. *Evolutionary Ecology*, **9**, 82–92.

Jousselin E, Genson G, Coeur d'acier A (2010) Evolutionary lability of a complex life cycle in the aphid genus *Brachycaudus*. *BMC Evolutionary Biology*, **10**, 295.

Katoh K, Kuma K, Toh H, Miyata T (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic acids research*, **33**, 511–8.

Kelley ST, Farrell BD (1998) Is specialization a dead end? The phylogeny of host use in Dendroctonus bark beetles (Scolytidae). *Evolution*, **52**, 1731–1743.

Kononova EL (1977) New aphid species (Homoptera, Aphidinea) from the Upper Cretaceous deposits of the Taymyr. *Revue d'Entomologie de l'URSS*, **56**, 588–600.

Krasnov BR, Fortuna MA, Mouillot D *et al.* (2012) Phylogenetic signal in module composition and species connectivity in compartmentalized host-parasite networks. *The American Naturalist*, **179**, 501–511.

Kristensen NP, Skalski AW (1998) Phylogeny and palaeontology. In: *Lepidoptera, Moths and Butterflies, Vol. 1: Evolution, Systematics, and Biogeography. Handbook of Zoology, Vol. IV Arthropoda: Insecta, Part 35* (ed Kristensen NP), pp. 7–25. Walter de Gruyter, Berlin & New York.

Labandeira CC, Dilcher DL, Davis DR, Wagner DL (1994) Ninety-seven million years of angiosperm-insect association: paleobiological insights into the meaning of coevolution. *Proceedings of the National Academy of Sciences of the United States of America*, **91**, 12278–12282.

Li X, Schuler MA, Berenbaum MR (2002) Jasmonate and salicylate induce expression of herbivore cytochrome P450 genes. *Nature*, **419**, 712–715.

Li W, Schuler MA, Berenbaum MR (2003) Diversification of furanocoumarin-metabolizing cytochrome P450 monooxygenases in two papilionids: specificity and substrate encounter rate. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 14593–14598.

Maddison WP, FitzJohn RG (2015) The unsolved challenge to phylogenetic correlation tests for categorical characters. *Systematic Biology*, **64**, 127–136.

Magsig-Castillo J, Morse JG, Walker GP *et al.* (2010) Phoretic dispersal of armored scale crawlers (Hemiptera: Diaspididae). *Journal of Economic Entomology*, **103**, 1172–1179.

Mayhew PJ (1997) Adaptive patterns of host-plant selection by phytophagous insects. *Oikos*, **79**, 417–428.

Metz J, Gyllenberg M (2001) How should we define fitness in structured metapopulation models? Including an application to the calculation of evolutionarily stable dispersal strategies. *Proceedings of the Royal Society B: Biological Sciences*, **268**, 499–508.

Miller DR, Davidson JA (2005) *Armored Scale Insect Pests of Trees and Shrubs*. Cornell University Press, Ithaca, NY.

Mirarab S, Nguyen N, Warnow T (2014) PASTA: ultra-large multiple sequence alignment. *Research in Computational Molecular Biology*, 177–191.

Misof B, Liu S, Meusemann K *et al.* (2014) Phylogenomics resolves the timing and pattern of insect evolution. *Science*, **346**, 763–767.

Morse GE, Normark BB (2006) A molecular phylogenetic study of armoured scale insects (Hemiptera: Diaspididae). *Systematic Entomology*, **31**, 338–349.

Müllner D (2013) fastcluster: fast hierarchical, agglomerative clustering routines for R and Python. *Journal of Statistical Software*, **53**, 1–18.

Murdoch WW, Briggs CJ, Swarbrick S (2005) Host suppression and stability in a parasitoid-host system: experimental demonstration. *Science*, **309**, 610–613.

Nath AJ, Das G, Das AK (2008) Vegetative phenology of three bamboo species in subtropical humid climate of Assam. *Tropical Ecology*, **49**, 85–89.

Normark BB, Johnson NA (2011) Niche explosion. *Genetica*, **139**, 551–564.

Normark BB, Morse GE, Krewinski A, Okusu A (2014) Armored scale insects (Hemiptera: Diaspididae) of San Lorenzo National Park, Panama, with descriptions of two new species. *Annals of the Entomological Society of America*, **107**, 37–49.

Nosil P, Crespi BJ, Sandoval CP (2002) Host-plant adaptation drives the parallel evolution of reproductive isolation. *Nature*, **417**, 440–3.

Nurmi T, Parvinen K (2011) Joint evolution of specialization and dispersal in structured metapopulations. *Journal of Theoretical Biology*, **275**, 78–92.

Parameswaran N, Liese W (1976) On the fine structure of bamboo fibres. *Wood Science and Technology*, **10**, 231–246.

Peterson DA, Hardy NB, Morse GE *et al.* (2015) Phylogenetic analysis reveals positive correlations between adaptations to diverse hosts in a group of pathogen-like herbivores. *Evolution*, **69**, 2785–2792.

Peterson DA, Hardy NB, Normark BB (2016) Micro- and macroevolutionary trade-offs in plant-feeding insects. *The American Naturalist*, **188**, 640–650.

Pires MM, Guimarães PR (2012) Interaction intimacy organizes networks of antagonistic interactions in different ways. *Journal of The Royal Society Interface*, rsif20120649.

Popov YA (1971) Historical development of the infraorder Nepomorpha (Heteroptera). *Trudy Paleontologicheskogo Instituta Akademii Nauk SSSR*, **129**, 1–121.

Prapaipong H, Berenbaum MR, Schuler MA (1994) Transcriptional regulation of the *Papilio polyxenes* CYP6B1 gene. *Nucleic Acids Research*, **22**, 3210–3217.

R Core Team (2015) R: A language and environment for statistical computing.

Rafferty NE, Ives AR (2013) Phylogenetic trait-based analyses of ecological networks. *Ecology*, **94**, 2321–2333.

Ravigné V, Dieckmann U, Olivieri I (2009) Live where you thrive: joint evolution of habitat choice and local adaptation facilitates specialization and promotes diversity. *The American Naturalist*, **174**, E141–E169.

Reeve J, Murdoch W (1986) Biological control by the parasitoid *Aphytis melinus*, and population stability of the California red scale. *The Journal of Animal Ecology*, **55**, 1069–1082.

Remold S (2012) Understanding specialism when the jack of all trades can be the master of all. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 4861–4869.

Revell LJ, Chamberlain SA (2014) Rphylip: An R interface for PHYLIP. *Methods in Ecology and Evolution*, **5**, 976–981.

Robinson GS, Ackery PR, Kitching IJ, Beccaloni GW, Hernández LM (2015) HOSTS - a database of the world's lepidopteran hostplants. *Natural History Museum, London*.

Rodriguez-Verdugo A, Carrillo-Cisneros D, Gonzalez-Gonzalez A, Gaut BS, Bennett AF (2014) Different tradeoffs result from alternate genetic adaptations to a common environment. *Proceedings of the National Academy of Sciences*, **111**, 12121–12126.

Ronce O (2007) How does it feel to be like a rolling stone? Ten questions about dispersal evolution. *Annual Review of Ecology, Evolution, and Systematics*, **38**, 231–253.

Satterwhite RS, Cooper TF (2015) Constraints on adaptation of *Escherichia coli* to mixed-resource environments increase over time. *Evolution*, **69**, 2067–2078.

Scheirs J, Jordaens K, De Bruyn L (2005) Have genetic trade-offs in host use been overlooked in arthropods? *Evolutionary Ecology*, **19**, 551–561.

Scriber JM (2010) Integrating ancient patterns and current dynamics of insect-plant interactions: Taxonomic and geographic variation in herbivore specialization. *Insect Science*, **17**, 471–507.

Shoval O, Sheftel H, Shinar G *et al.* (2012) Evolutionary trade-offs, Pareto optimality, and the geometry of phenotype space. *Science*, **336**, 1157–1160.

Singer MS, Stireman JO (2005) The tri-trophic niche concept and adaptive radiation of phytophagous insects. *Ecology Letters*, **8**, 1247–1255.

Smith SA, Beaulieu JM, Donoghue MJ (2009) Mega-phylogeny approach for comparative biology: an alternative to supertree and supermatrix approaches. *BMC Evolutionary Biology*, **9**, 37.

Smith SA, O'Meara BC (2012) treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics*, **28**, 2689–2690.

Sohn J-C, Labandeira C, Davis D, Mitter C (2012) An annotated catalog of fossil and subfossil Lepidoptera (Insecta: Holometabola) of the world. *Zootaxa*, **3286**, 1–132.

Soltis DE, Smith SA, Cellinese N *et al.* (2011) Angiosperm phylogeny: 17 genes, 640 taxa. *American Journal of Botany*, **98**, 704–30.

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–3.

Szwedo J, Nel A (2011) The oldest aphid insect from the Middle Triassic of the Vosges, France. *Acta Palaeontologica Polonica*, **56**, 757–766.

Taylor AH, Reid DG, Zisheng Q, Jinchu H (1991) Spatial patterns and environmental associates of bamboo (*Bashania fangiana* Yi) after mass-flowering in southwestern China. *Bulletin of the Torrey Botanical Society*, **118**, 247–254.

Taylor AH, Zisheng Q (1987) Culm dynamics and dry matter production of bamboos in the Wolong and Tangjiahe giant panda reserves, Sichuan, China. *Journal of Applied Ecology*, **24**, 419–433.

Vea IM, Grimaldi DA (2015) Diverse new scale insects (Hemiptera: Coccoidea) in amber from the Cretaceous and Eocene with a phylogenetic framework for fossil Coccoidea. *American Museum Novitates*, **3823**, 1–15.

Vea IM, Grimaldi DA (2016) Putting scales into evolutionary time: the divergence of major scale insect lineages (Hemiptera) predates the radiation of modern angiosperm hosts. *Scientific Reports*, **6**, 23487.

Vieira RC, Gomes DMS, Sarahyba LS, Arruda RCO (2002) Leaf anatomy of three herbaceous bamboo species. *Brazilian Journal of Biology*, **62**, 907–922.

Wahlberg N, Wheat CW, Peña C (2013) Timing and patterns in the taxonomic diversification of Lepidoptera (butterflies and moths). *PLoS ONE*, **8**, 1–8.

Weinreich DM, Watson RA, Chao L (2005) Perspective: sign epistasis and genetic constraint on evolutionary trajectories. *Evolution*, **59**, 1165–1174.

Wen Z, Zeng R Sen, Niu G, Berenbaum MR, Schuler MA (2009) Ecological significance of induction of broad-substrate cytochrome P450s by natural and synthetic inducers in *Helicoverpa zea*. *Journal of Chemical Ecology*, **35**, 183–9.

Whitlock MC (1996) The red queen beats the jack-of-all-trades: the limitations on the evolution of phenotypic plasticity and niche breadth. *American Naturalist*, **148**, S65–S77.

Yu SJ, Berry RE, Terriere LC (1979) Host plant stimulation of detoxifying enzymes in a phytophagous insect. *Pesticide Biochemistry and Physiology*, **12**, 280–284.