bioRxiv preprint doi: https://doi.org/10.1101/040311; this version posted March 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Macroevolutionary trade-offs in plant-feeding insects

Daniel A. Peterson,[1] Nate B. Hardy,[2] and Benjamin B. Normark[1]

[1]Department of Biology and Graduate Program in Organismic and Evolutionary Biology, 221 Morrill Science Center, 611 North Pleasant Street, University of Massachusetts, Amherst, MA 01003

[2]Department of Entomology & Plant Pathology, 321 Funchess Hall, Auburn University, Auburn, Alabama 36849

Corresponding Author. Daniel Peterson, Department of Biology, 221 Morrill Science Center, 611 North Pleasant Street, University of Massachusetts, Amherst, MA 01003.

Phone: (413) 545-2844. Email: dapeters@cns.umass.edu

bioRxiv preprint doi: https://doi.org/10.1101/040311; this version posted March 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract

Most plant-feeding insects are ecological specialists restricted to one or a few closely related host-plant species. A long-standing hypothesis asserts that natural selection favors host specialization because of trade-offs between performance on alternative hosts, yet empirical evidence for such trade-offs is scarce. Here we assess trade-offs between adaptations to alternative hosts over macroevolutionary timescales in two major orders of plant-feeding insects: Lepidoptera (caterpillars) and Hemiptera (true bugs). Across 1604 caterpillar species, we find both positive and negative associations between presence on diverse host taxa. The patterns of these associations suggest that different trade-offs constrain host-use over short and long evolutionary timescales. In contrast, host-use patterns of 955 true bug species reveal uniformly positive associations between adaptations to the same host taxa 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 observe only broad trade-offs, suggesting that alternative evolutionary processes drive the high degree of ecological specialization observed in most plant-feeding insects.

bioRxiv preprint doi: https://doi.org/10.1101/040311; this version posted March 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

The ubiquity of ecological specialists (1, 2) is often attributed to evolutionary constraints on generalism (3, 4). Adaptations to one environment may come at the cost of performance in other environments (1, 3), and such trade-offs are crucial elements of most theoretical models of the evolution of specialization (5). Trade-offs are implied by negative correlations between traits in many biological systems (6), but in plant-feeding insects, performance measures on alternative host species are most often uncorrelated or positively correlated (1, 7). In fact, although plant-feeding insects are a model system for the study of ecological specialization (1, 3), the role of trade-offs in the evolution of insect host-specialization remains generally unsupported.

However, trade-offs that drive specialization in plant-feeding insects may be difficult to detect within species (8). Host-use traits are phylogenetically conserved in many insect groups (9), and genetic variation for use of novel hosts is often absent within a single population (4). These constraints on host-use variation likely stem from the fact that host use is a complex trait involving multiple interacting genetic loci (10), so the evolutionary trajectories that lead to high fitness on alternative resources may be mutually exclusive. For example, the evolution of improved performance on one host can select for a preference for that host, driving a positive feedback loop of further specialization (11). When epistatic interactions are strong, the greatest variation in host-use strategies will evolve between reproductively isolated lineages over long time scales (12), with trade-off outcomes determined by each lineage's evolutionary history (13). Analogous macroevolutionary trade-offs have been described in plants; the presence of alternative defensive strategies are negatively correlated over plant evolutionary history (14). However, it remains unknown whether this functional diversification across hosts has created macroevolutionary trade-offs for plant-feeding insects; performance on alternative hosts may instead be positively associated if generalist adaptations have been prevalent during the evolution of insect herbivores (15).

We investigated trade-offs between adaptations to diverse host taxa over long- and short-term

bioRxiv preprint doi: https://doi.org/10.1101/040311; this version posted March 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

macroevolutionary timescales in two orders of plant-feeding insects: Lepidoptera (caterpillars) and Hemiptera (true bugs). Assuming that trade-offs should produce negative correlations between observed use of alternative host taxa for individual insect species (15), we quantified the macroevolutionary relationships between use of the most common host-plant orders in North America. We assembled the pairwise correlations between use of the focal hosts to produce network graphs of host-use relationships, revealing overall patterns of host-use evolution in each insect order. Specifically, we investigated two plausible trade-off scenarios: segregation of host orders into discrete clusters and isolation of individual host orders.

Results

Host-use in Lepidoptera and Hemiptera. We obtained North American host-use records and phylogenetic data for 1604 caterpillar species and 955 bug species from online databases. 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 total interactions in the Lepidoptera dataset and 57% of total interactions in the Hemiptera dataset.

Host-use Correlations. We inferred a time-calibrated phylogeny of each insect order (Fig. 1) and used a phylogenetic meta-analysis approach to test for evolutionary trade-offs between insect use of the focal host-plant orders. For each pairwise combination of focal host-plant orders, we estimated long- and short-term evolutionary associations between use of the two host taxa by finding maximum likelihood estimates for phylogenetic and residual correlations between host-use traits in a phylogenetic mixed model (Fig. 2A-F). We recovered both positive and negative correlations between use of the focal host orders in the Lepidoptera (Fig. 2G), but mostly positive correlations in the Hemiptera (Fig. 2H).

Host-use Network Structure. We used the pairwise host-use correlations to reconstruct the overall

bioRxiv preprint doi: https://doi.org/10.1101/040311; this version posted March 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

network of host-use correlations over phylogenetic and residual timescales in each insect order. We then evaluated the structure of each network to identify clusters of evolutionarily associated host taxa as well as individually isolated host taxa. We found that the network of long-term evolutionary relationships between lepidopteran use of the focal host orders was significantly structured ($P < 0.01$), revealing two large clusters of host taxa (Fig. 3a). 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. Short-term correlations between lepidopteran use of the focal host taxa also showed significant network structure ($P < 0.01$) but on this timescale use of all angiosperm hosts formed a single cluster of mostly positive associations (Fig. 3b). However, use of Pinales was isolated from the angiosperm cluster, exhibiting a statistically significant mean pairwise correlation with use of the other focal hosts (-0.23, $P < 0.01$). In contrast, hempiteran host-use correlations indicated significant support for a single host-use cluster encompassing all focal hosts over both phylogenetic ($P < 0.01$, Fig. 3c) and residual timescales ($P < 0.01$, Fig. 3d).

Discussion

A fundamental trade-off between performance on alternative hosts remains the dominant hypothesis to explain the planet's abundance of specialized plant-feeding insects, despite the lack of empirical evidence for such trade-offs. Most studies have looked for trade-offs within insect species, but the effects of trade-offs could be more obvious over longer evolutionary timescales (8). Our results suggest that trade-offs have shaped patterns of host-use over millions of years of caterpillar evolution and diversification. However, the lack of evidence for trade-offs among plant-feeding bugs indicates that trade-offs are not universal, immutable constraints, but instead reflect historically contingent processes in individual insect lineages.

There are many differences between Lepidoptera and Hemiptera, but their fundamentally distinct relationships with host plants may be particularly important to understanding why the role of

bioRxiv preprint doi: https://doi.org/10.1101/040311; this version posted March 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

trade-offs appears to be very different in the two groups. Hemiptera are sucking insects, while Lepidoptera are generally leaf-chewers (2). These two feeding modes elicit different modes of plant defensive responses (16), and sap-sucking may be particularly amenable to generalist adaptations that circumvent host defenses (17). In contrast, Lepidoptera often rely on specialized enzymes to detoxify defensive chemicals (18), which may constrain the evolution of generalism, although super-generalist Lepidoptera do exist, possibly powered by phenotypic plasticity in enzyme expression (19).

Alternatively, defense chemistry may have little influence over long-term evolutionary relationships between plants and their insect parasites. In fact, we found that long-term evolutionary trade-offs experienced by lepidopteran lineages appear to reflect a broad-scale ecological division between plant hosts rather than a history of coevolution between plants and insects (20). The relevant differences between the two host clusters we observed are not immediately clear, but the separation of predominantly woody plant taxa (e.g. Pinales, Fagales) from predominantly herbaceous taxa (e.g. Asterales, Poales) suggests the possibility of a trade-off between host growth forms or between the habitat types where those growth forms are found (21).

Although we observed few constraints on short-term evolution of lepidopteran host-use within the focal angiosperm taxa, we did find evidence for a short-term evolutionary trade-off between their use of angiosperms and the only focal gymnosperm host order, Pinales (conifers). We found that conifers and some angiosperms are used by caterpillars with similar evolutionary backgrounds, yet a trade-off between evolutionarily labile traits tends to prevent the simultaneous use of those host taxa by individual caterpillar species. Such trade-offs between labile but mutually exclusive host-use traits may be particularly significant because they can promote rapid speciation (22), driving the adaptive radiations that have produced much of the plant-feeding insect diversity on earth (23, 24). Nevertheless, the prevalence of such constraints at smaller scales, for instance between alternative host genera or species, remains unclear given that the single short-term trade-off observed here occurred over the largest phylogenetic distance present among our focal taxa.

bioRxiv preprint doi: https://doi.org/10.1101/040311; this version posted March 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

In contrast to the patterns observed in the Lepidoptera, hemipteran host-use showed no evidence of trade-offs over long or short evolutionary timescales. This surprising result suggests that generalist adaptations that increase fitness across multiple hosts have been more important for Hemiptera than specialist adaptations to particular hosts (15). 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 (25). 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 (3). Nevertheless, costs of generalism have been difficult to document (1, 26), so the positive residual correlations we observed may instead represent evolutionary breakthroughs made possible by novel mechanisms of phenotypic plasticity or other generalist adaptations (17).

Overall, we found strong evidence that both labile traits and long-term, historically contingent evolutionary processes produce important host-use constraints in plant-feeding insects. However, the relatively few, broad-scale trade-offs found here fail to explain the observation that a vast majority of plant-feeding insects are restricted to a single host family (2). It is clear that trade-offs can vary based on genetic context (13), so we cannot rule out the hypothesis that species-specific, idiosyncratic trade-offs drive ecological specialization. Nevertheless, although a few individual species-specific trade-offs have been documented (22), our results suggest that most microevolutionary trade-offs do not persist over macroevolutionary timescales. Trade-offs play an intuitive and possibly inescapable role in constraining performance across multiple tasks (6), yet host-specialization in plant-feeding insects is not necessarily driven by host-specific limitations on survival and growth. Alternative factors, such as neural constraints (27), mate-finding (28), or genetic drift (26) may ultimately shape the evolution of each species' ecological niche.

Materials and Methods

bioRxiv preprint doi: https://doi.org/10.1101/040311; this version posted March 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Data Collection. Lepidopteran host-use data were downloaded from the HOSTS database (nhm.ac.uk/hosts)(29), a collection of published records of worldwide caterpillar host-plants. Hemipteran host-use data were downloaded from the Tri-Trophic Thematic Collection Network database (tcn.amnh.org), a compilation of insect collection records from academic and governmental museums in the United States. 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 (30) and insect taxonomic names with the python package TaxonNamesResolver and the following reference databases: Aphid Species File (31), 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 taxon, with insects considered present on all hosts for which they had at least one host-use record. To focus computational resources on host taxa with enough statistical power to detect evolutionary host-use interactions, we restricted our 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).

We characterized shared ancestry among the North American lepidopteran and hemipteran species in our host-use dataset by estimating time-scaled phylogenies using a phyloinformatic approach (see Supplemental Materials 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 matrices for these species are available on Dryad (datadryad.org).

Statistical Analysis. Treating each pairwise comparison between host taxa individually, we set up a phylogenetic mixed model (32) for each insect order using the package MCMCglmm (33) in the R statistical framework (34). We estimated the evolutionary interaction between the pair of host-use traits using the “`random=~us(trait):Insect`” and “`rcov=~us(trait):units`” syntax to capture both phylogenetic and residual effects. Prior parameter distributions were specified as

bioRxiv preprint doi: https://doi.org/10.1101/040311; this version posted March 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

"`prior<-list(R=list(V=diag(2),nu=2), G=list(G1=list(V=diag(2),nu=2)))`". All MCMC chains were run for 10 million iterations with a burn-in of 1 million iterations, and ten chains were run for each model to test for convergence. Gelman-Rubin convergence analysis of the ten chains for each model produced potential scale reduction factors under 1.05 in every case (99% were under 1.01), suggesting that all chains successfully converged (35).

We considered two potential trade-off scenarios: isolation of individual host-use traits, and divisions of host-use traits into clusters (characterized by positive interactions within clusters and negative interactions between them). To characterize isolation of individual host-use traits, we calculated the mean correlation of the ten pairwise comparisons involving each host-use trait (separately for phylogenetic and residual correlations). To characterize associations among groups of host-use traits, we found the network structure that maximized the structure score for the entire network of host-use traits, defined as the sum of all pairwise correlations between host-use traits in the same cluster minus the sum of all pairwise correlations between host-use traits in different clusters (see Supplemental Materials for details).

We tested the statistical significance of the resulting values by producing a series of 100 null datasets that simulated independent Brownian motion of a continuous character for performance on each host order along the insect phylogenies, plus additional normally distributed residual variation with equivalent variance. 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 (36). We then calculated all pairwise correlations between use of the focal hosts, mean correlations per host-use trait, and whole-network structure as we did for the empirical data. Empirical individual host-use trait mean correlations were considered statistically significant when their absolute values exceeded the maximum absolute values of any individual mean

bioRxiv preprint doi: https://doi.org/10.1101/040311; this version posted March 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

in 95% of null datasets. Empirical network structures was considered statistically significant when their scores exceeded the network structure scores of 95% the null datasets.

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank D. Moen, A. Porter, L. Doubleday, and S. Noda for comments that improved the manuscript. This work was supported by the National Science Foundation (EF-1115191 and DEB-1258001)

Author Contributions All authors were involved in the study design and wrote the manuscript. D.A.P. and N.B.H. analyzed the data.

Author Information The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.A.P. (dapeters@cns.umass.edu)

Literature Cited

1. Forister ML, Dyer LA, Singer M, Stireman JO, Lill JT (2012) Revisiting the evolution of ecological interactions. *Ecology* 93:981–991.

2. Forister ML et al. (2015) The global distribution of diet breadth in insect herbivores. *Proc Natl Acad Sci U S A* 112:442–7.

3. Futuyma DJ, Moreno G (1988) The evolution of ecological specialization. *Annu Rev Ecol Syst* 19:207–233.

4. 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.

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

6. Shoval O et al. (2012) Evolutionary trade-offs, Pareto optimality, and the geometry of

bioRxiv preprint doi: https://doi.org/10.1101/040311; this version posted March 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

phenotype space. *Science* 336:1157–1160.

7. Futuyma DJ (2008) in *Specialization, Speciation, and Radiation: The Evolutionary Biology of Herbivorous Insects*, pp 136–147.
8. Joshi A, Thompson JN (1995) Trade-offs and the evolution of host specialization. *Evol Ecol* 9:82–92.
9. Futuyma DJ, Agrawal AA (2009) Macroevolution and the biological diversity of plants and herbivores. *Proc Natl Acad Sci U S A* 106:18054–18061.
10. Remold S (2012) Understanding specialism when the jack of all trades can be the master of all. *Proc R Soc B Biol Sci* 279:4861–4869.
11. Fry J (1996) The evolution of host specialization: are trade-offs overrated? *Am Nat* 148:S84–107.
12. Satterwhite RS, Cooper TF (2015) Constraints on adaptation of *Escherichia coli* to mixed-resource environments increase over time. *Evolution* 69:2067–2078.
13. 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. *Proc Natl Acad Sci* 111:12121–12126.
14. Campbell SA, Kessler A (2013) Plant mating system transitions drive the macroevolution of defense strategies. *Proc Natl Acad Sci* 110:3973–3978.
15. Peterson DA et al. (2015) Phylogenetic analysis reveals positive correlations between adaptations to diverse hosts in a group of pathogen-like herbivores. *Evolution* 69:2785–2792.
16. Ali JG, Agrawal AA (2012) Specialist versus generalist insect herbivores and plant defense. *Trends Plant Sci* 17:293–302.
17. Barrett LG, Heil M (2012) Unifying concepts and mechanisms in the specificity of plant-enemy interactions. *Trends Plant Sci* 17:282–292.
18. Berenbaum MR, Feeny PP (2008) in *Specialization, Speciation, and Radiation: The Evolutionary Biology of Herbivorous Insects*, pp 3–19.
19. Li X, Schuler MA, Berenbaum MR (2002) Jasmonate and salicylate induce expression of herbivore cytochrome P450 genes. *Nature* 419:712–715.
20. Ehrlich PR, Raven PH (1964) Butterflies and plants: a study in coevolution. *Evolution* 18:586–608.
21. Futuyma DJ (1976) Food plant specialization and environmental predictability in Lepidoptera. *Am Nat* 110:285–292.
22. Nosil P, Crespi BJ, Sandoval CP (2002) Host-plant adaptation drives the parallel evolution of reproductive isolation. *Nature* 417:440–3.

bioRxiv preprint doi: https://doi.org/10.1101/040311; this version posted March 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

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

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

26. Gompert Z et al. (2015) The evolution of novel host use is unlikely to be constrained by tradeoffs or a lack of genetic variation. *Mol Ecol* 24:2777–2793.

27. Bernays EA (2001) Neural limitations in phytophagous insects: implications for diet breadth and evolution of host affiliation. *Annu Rev Entomol* 46:703–727.

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

29. Robinson GS, Ackery PR, Kitching IJ, Beccaloni GW, Hernández LM (2015) HOSTS - a database of the world’s Lepidopteran hostplants. *Nat Hist Museum, London*.

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

31. Favret C (2015) Aphid species file. Available at: Aphid.SpeciesFile.org [Accessed January 1, 2015].

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

33. Hadfield JD (2010) MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. *J Stat Softw* 33:1–22.

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

35. Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. *Stat Sci* 7:457–472.

36. Felsenstein J (2012) A comparative method for both discrete and continuous characters using the threshold model. *Am Nat* 179:145–156.

Figure Legends

Fig. 1. 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

bioRxiv preprint doi: https://doi.org/10.1101/040311; this version posted March 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

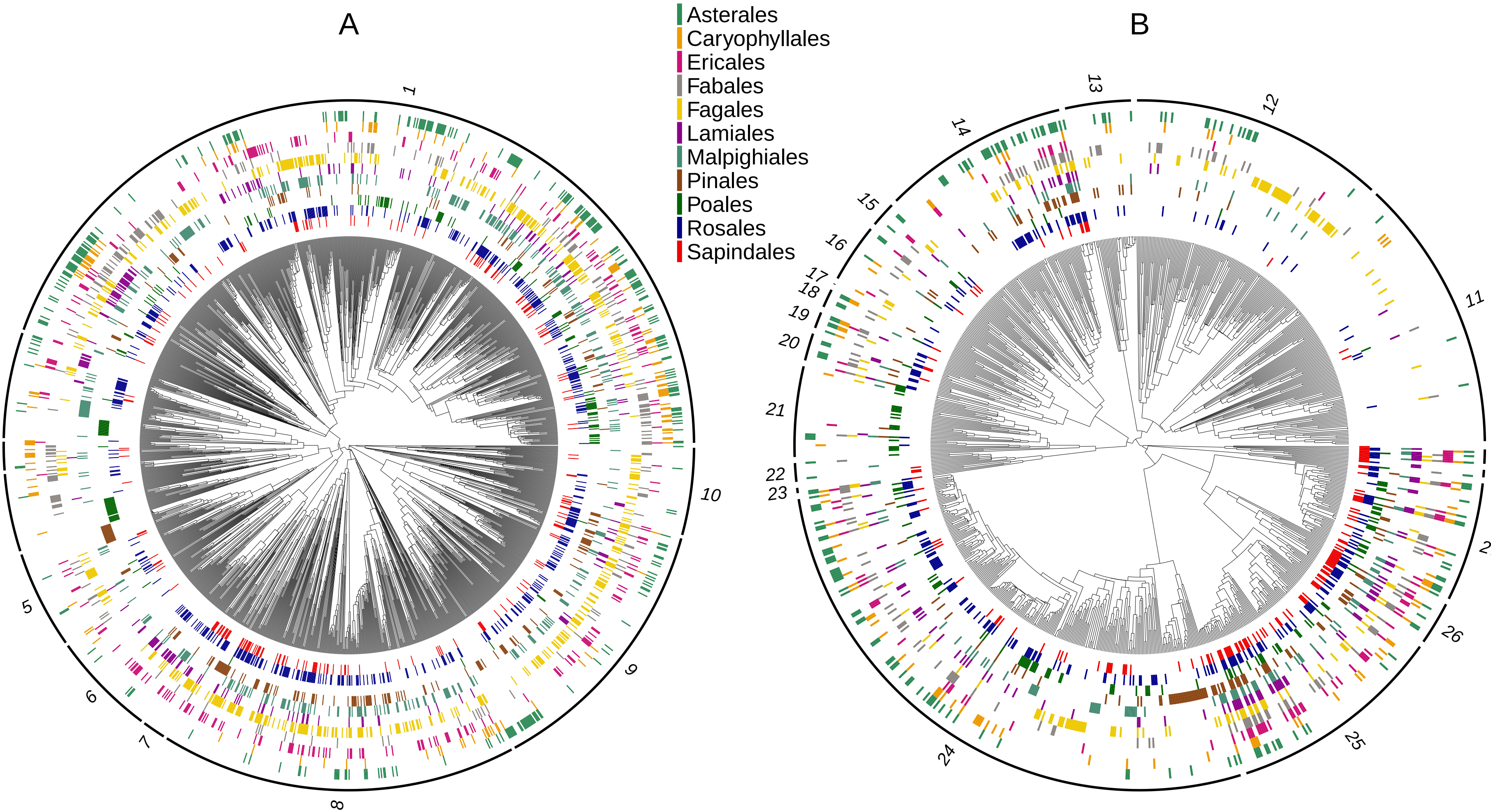
observed only on non-focal hosts or had no host-use information associated with their locality records (Hemiptera only). Insect families (and one superfamily) are indicated around the phylogenies as follows: (A) Lepidoptera – 1: Noctuoidea, 2: Nymphalidae, 3: Lycaenidae, 4: Hesperiidae, 5: Pyralidae, 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.

Fig. 2. 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, red squares on the left indicate which species in the herbivore phylogeny are present on host 1, and blue squares on the right indicate which species are present on host 2. (G) Empirical phylogenetic correlation by residual correlation plots of all 55 pairwise combinations of the focal host orders for Lepidoptera and (H) Hemiptera.

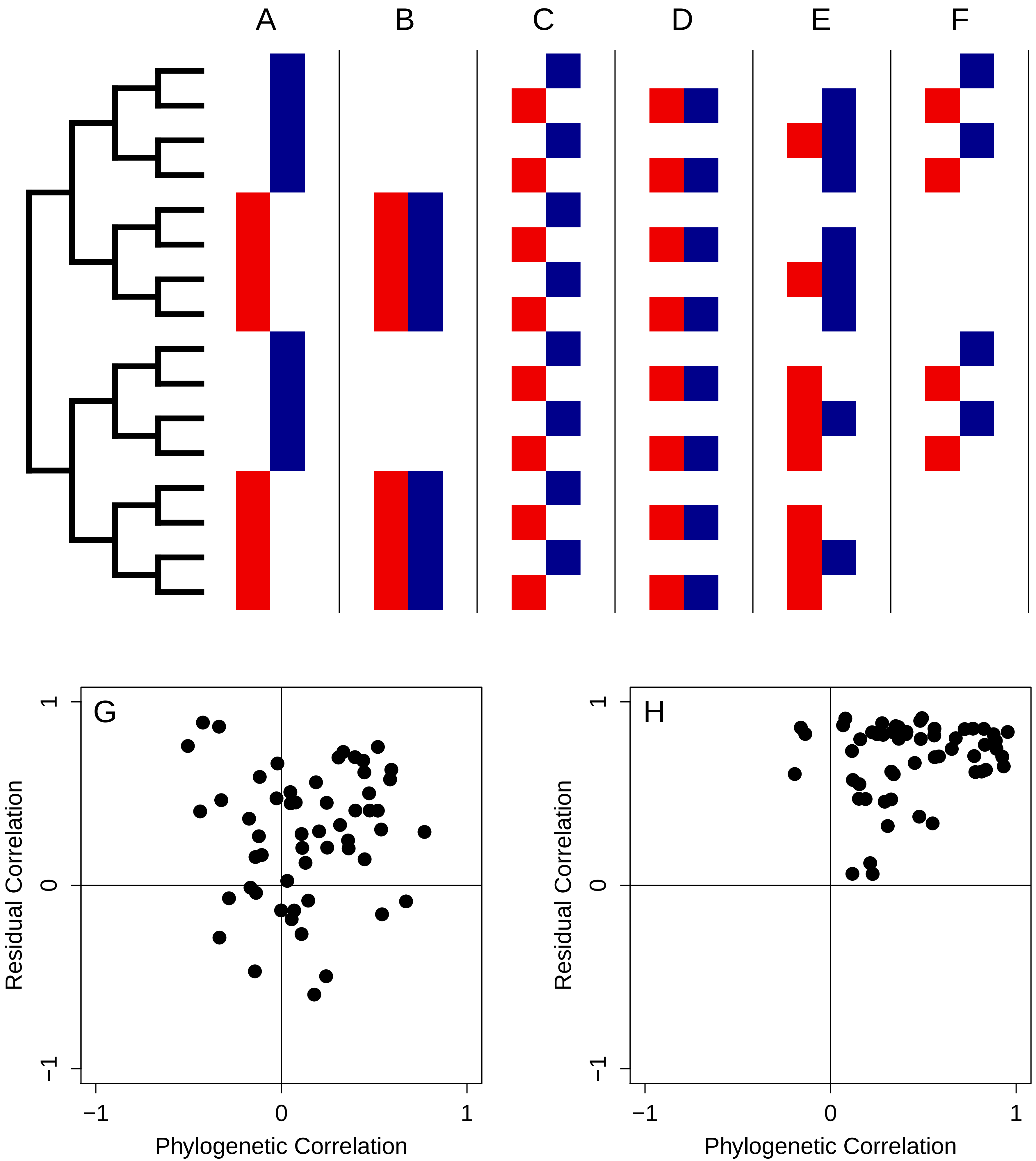
Fig. 3. Network graphs of inferred pairwise host-use correlations. (A) Lepidoptera – phylogenetic correlations. (B) Lepidoptera – residual correlations. (C) Hemiptera – phylogenetic correlations. (D) Hemiptera – residual correlations. Each vertex represents a host order, with vertex area proportional 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 algorithm (47), a force-directed layout method in which “repulsion” between

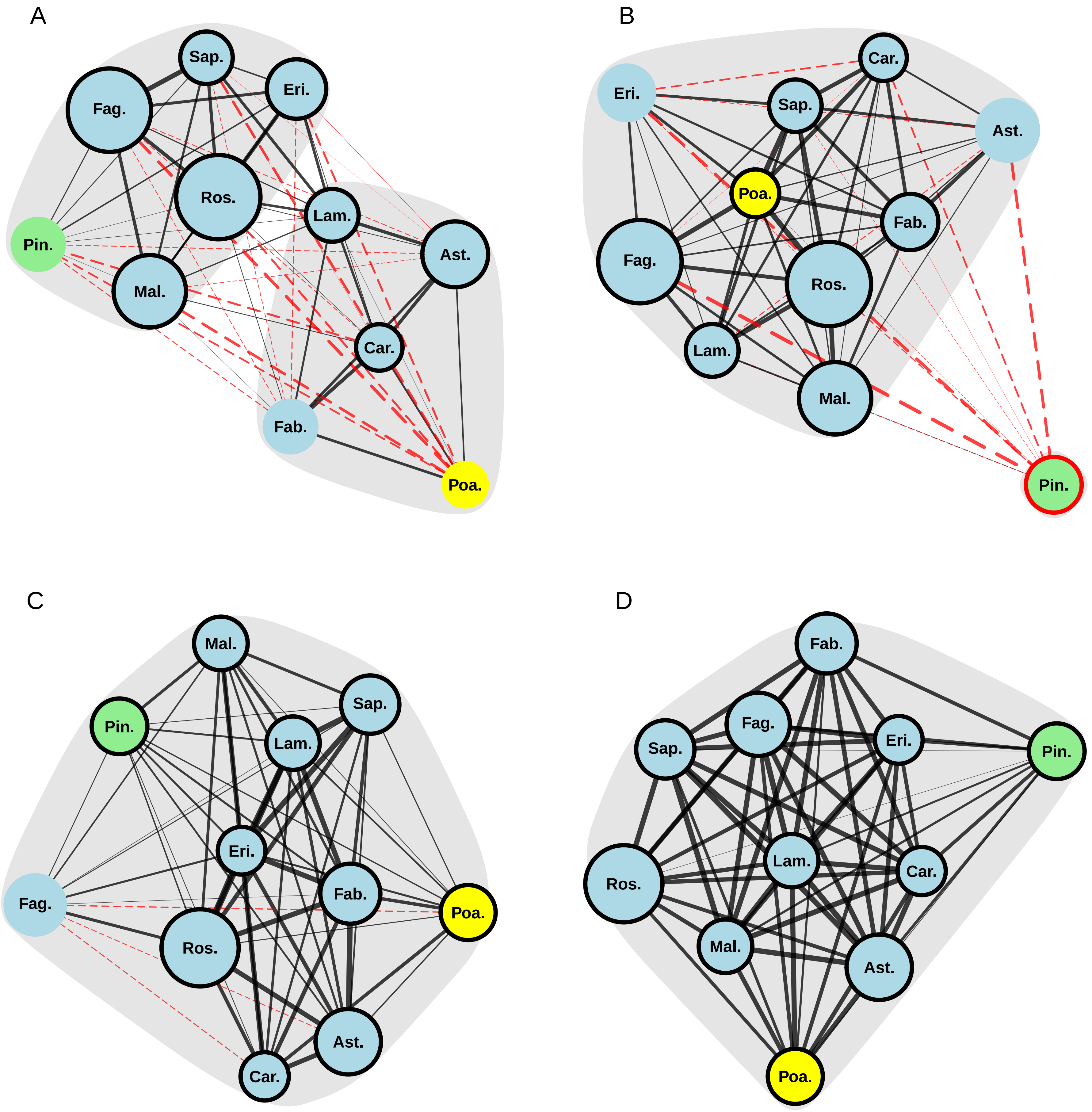
bioRxiv preprint doi: https://doi.org/10.1101/040311; this version posted March 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 modules ($P < 0.05$) are indicated by grey 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).


A
B
Asterales
Caryophyllales
Ericales
Fabales
Fagales
Lamiales
Malpighiales
Pinales
Poales
Rosales
Sapindales
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29

certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.




A
Sap.
Eri.
Fag.
Ros.
Lam.
Pin.
Ast.
Mal.
Car.
Fab.
Poa.
B
Car.
Eri.
Sap.
Ast.
Poa.
Fab.
Fag.
Ros.
Lam.
Mal.
Pin.
C
Mal.
Sap.
Pin.
Lam.
Eri.
Fag.
Fab.
Poa.
Ros.
Ast.
Car.
D
Fab.
Fag.
Eri.
Sap.
Pin.
Lam.
Ros.
Car.
Mal.
Ast.
Poa.

bioRxiv preprint doi: https://doi.org/10.1101/040311; this version posted March 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Supplementary Methods

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 4 constraints: The age of the root, which corresponds to the crown node of Apoditrysia, was fixed at 140 Ma, to match the the published estimate of Wahlberg et al. (2013). Uniform constraint with a minimum age of 65 Ma and a maximum age of 140 were placed on the crown nodes of Gracillariidae, Noctuidae, and Nymphalidae based on the minimum age estimates of fossil taxa (Kozlov 1988; Martins-Neto 1989). 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 140 and maximum age of 291 Ma were imposed on the crown node of Aphidomorpha (Brodie 1845). 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 and Grimaldi 2015). A minimum age of 197 and a maximum age of 291 were placed on the crown node of Heteroptera (Giebel 1856).

Network Structure Analysis

To characterize the host-use association network structure, we produced a dendrogram of host-use associations from a Euclidean distance matrix calculated from the pairwise correlations between presence on the focal hosts. Agglomerative heirarchical clustering was performed by the hclust function in the R package fastcluster (Müllner 2013). After obtaining the dendrogram of host-use associations, we evaluated all possible module divisions produced by pruning the dendrogram at a single branch height. The fit of a given module structure was defined as the sum of all correlations between host-use traits in the same module minus the sum of all correlations between host-use traits in different modules. Thus, positive correlations within modules and negative correlations between modules increased the fit score, while negative correlations within modules and positive correlations between modules reduced the fit score. The module structure with the highest fit score was chosen as the best characterization of network structure.

Literature Cited

Brodie, P. B. 1845. A History of the Fossil Insects in the Secondary Rocks of England. John Van Voorst.

Giebel, C. G. 1856. Die Insekten und Spinnen. *in* Fauna der Vorwelt mit steter Breucksichtigung der lebenden Thiere, Brodhaus, Leipzig.

bioRxiv preprint doi: https://doi.org/10.1101/040311; this version posted March 9, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Kononova, E. L. 1977. New aphid species (Homoptera, Aphidinea) from the Upper Cretaceous deposits of the Taymyr. Rev. d'Entomologie l'URSS 56:588–600.

Kozlov, M. V. 1988. Papilionida. Pp. 22–55 *in* Ponomarenko, ed. Cretaceous Biosynotic crisis in the evolution of insects.

Martins-Neto, R. G. 1989. Novos insectos terciarios du Estado de São Paulo. Rev. Bras. Geociencias 19:375–386.

Misof, B., S. Liu, K. Meusemann, R. S. Peters, A. Donath, C. Mayer, P. B. Frandsen, J. Ware, T. Flouri, R. G. Beutel, O. Niehuis, M. Petersen, F. Izquierdo-Carrasco, T. Wappler, J. Rust, A. J. Aberer, U. Aspock, H. Aspock, D. Bartel, A. Blanke, S. Berger, A. Bohm, T. R. Buckley, B. Calcott, J. Chen, F. Friedrich, M. Fukui, M. Fujita, C. Greve, P. Grobe, S. Gu, Y. Huang, L. S. Jermiin, A. Y. Kawahara, L. Krogmann, M. Kubiak, R. Lanfear, H. Letsch, Y. Li, Z. Li, J. Li, H. Lu, R. Machida, Y. Mashimo, P. Kapli, D. D. McKenna, G. Meng, Y. Nakagaki, J. L. Navarrete-Heredia, M. Ott, Y. Ou, G. Pass, L. Podsiadlowski, H. Pohl, B. M. von Reumont, K. Schutte, K. Sekiya, S. Shimizu, A. Slipinski, A. Stamatakis, W. Song, X. Su, N. U. Szucsich, M. Tan, X. Tan, M. Tang, J. Tang, G. Timelthaler, S. Tomizuka, M. Trautwein, X. Tong, T. Uchifune, M. G. Walzl, B. M. Wiegmann, J. Wilbrandt, B. Wipfler, T. K. F. Wong, Q. Wu, G. Wu, Y. Xie, S. Yang, Q. Yang, D. K. Yeates, K. Yoshizawa, Q. Zhang, R. Zhang, W. Zhang, Y. Zhang, J. Zhao, C. Zhou, L. Zhou, T. Ziesmann, S. Zou, X. Xu, H. Yang, J. Wang, K. M. Kjer, and X. Zhou. 2014. Phylogenomics resolves the timing and pattern of insect evolution. Science 346:763–767.

Müllner, D. 2013. fastcluster: Fast hierarchical, agglomerative clustering routines for R and Python. J. Stat. Softw. 53:1–18.

Smith, S. A., J. M. Beaulieu, and M. J. Donoghue. 2009. Mega-phylogeny approach for comparative biology: an alternative to supertree and supermatrix approaches. BMC Evol. Biol. 9:37.

Smith, S. A., and B. C. O'Meara. 2012. treePL: divergence time estimation using penalized likelihood for large phylogenies. Bioinformatics 28:2689–2690.

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

Vea, I. M., and D. A. Grimaldi. 2015. Diverse new scale insects (Hemiptera: Coccoidea) in amber from the Cretaceous and Eocene with a phylogenetic framework for fossil Coccoidea. Am. Museum Novit. 3823:1–15.

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