

## LETTERS

# Ecological interactions are evolutionarily conserved across the entire tree of life

José M. Gómez<sup>1</sup>, Miguel Verdú<sup>2</sup> & Francisco Perfectti<sup>3</sup>

Ecological interactions are crucial to understanding both the ecology and the evolution of organisms<sup>1,2</sup>. Because the phenotypic traits regulating species interactions are largely a legacy of their ancestors, it is widely assumed that ecological interactions are phylogenetically conserved, with closely related species interacting with similar partners<sup>2</sup>. However, the existing empirical evidence is inadequate to appropriately evaluate the hypothesis of phylogenetic conservatism in ecological interactions, because it is both ecologically and taxonomically biased. In fact, most studies on the evolution of ecological interactions have focused on specialized organisms, such as some parasites or insect herbivores<sup>3–7</sup>, belonging to a limited subset of the overall tree of life. Here we study the evolution of host use in a large and diverse group of interactions comprising both specialist and generalist acellular, unicellular and multicellular organisms. We show that, as previously found for specialized interactions, generalized interactions can be evolutionarily conserved. Significant phylogenetic conservatism of interaction patterns was equally likely to occur in symbiotic and non-symbiotic interactions, as well as in mutualistic and antagonistic interactions. Host-use differentiation among species was higher in phylogenetically conserved clades, irrespective of their generalization degree and taxonomic position within the tree of life. Our findings strongly suggest a shared pattern in the organization of biological systems through evolutionary time, mediated by marked conservatism of ecological interactions among taxa.

Shared ancestry may produce ecological similarity, with closely related species having similar ecological niches<sup>8,9</sup>. This idea may be traced back to Darwin's famous statement that the struggle for existence is most severe among related species because they have similar phenotypes and niche requirements<sup>10</sup>. Interspecific interactions comprise a substantial part of the niche of most species<sup>11</sup>. Conventional wisdom suggests that two closely related species should be more likely to interact with similar organisms than would species that are remotely related, because the phenotypic traits that regulate the interactions are often phylogenetically conserved<sup>3–7</sup>. It is thereby widely assumed that, as with other niche components, ecological interactions are evolutionarily conserved<sup>2,7</sup>.

We explored this idea by compiling information from 116 clades belonging to seven kingdoms (Euryarchaeota, Bacteria, Excavata, Chromalveolata, Fungi, Plantae and Animalia) from the three cellular domains (Archaea, Bacteria and Eukarya) and RNA and DNA viruses (Supplementary Data, appendix 1). We chose these systems because (1) they contain all types of ecological interaction, from antagonism (for example, endophytic herbivory, folivory and parasitism) to mutualism (for example, pollination, mycorrhiza, seed dispersal and nitrogen fixing); (2) by exploring organisms from disparate portions of the tree of life, our data set avoids taxonomic and systematic biases; (3) they comprise a wide range of generalization/

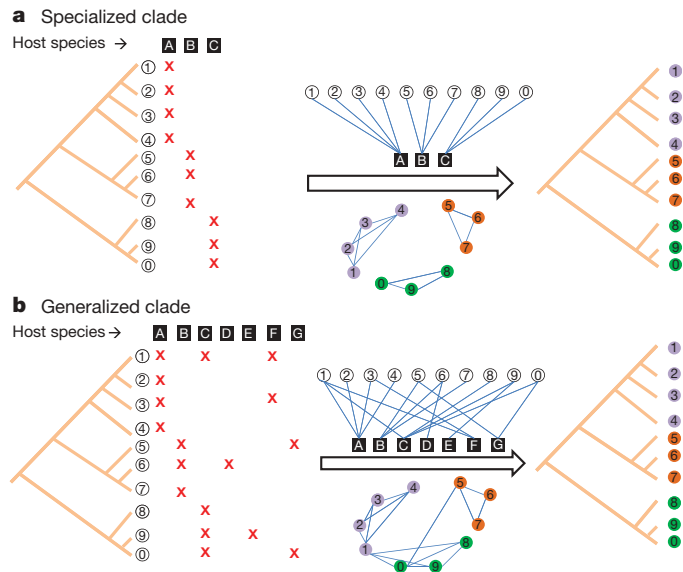
specialization degree; (4) reliable records of interacting organisms (hereafter referred to as hosts for the sake of simplicity) are available for these clades; and (5) phylogenetic trees are available (Supplementary Data, appendix 1). We have used genus as our target clade level because it is the taxonomic level at which interaction-mediated speciation mostly manifests<sup>2,7</sup> (Methods).

The host range of the studied clades, calculated as the average number of organisms interacting with each species of that clade, varied from 1 (extreme specialization) to 11.2. Nevertheless, within most systems there were species interacting with many hosts (up to 50) coexisting with species interacting with very few hosts (Supplementary Data, appendix 1). Of the studied clades, 58% ( $N = 67$ ) were specialist (host range,  $<1.5$  hosts per species) and 42% ( $N = 49$ ) were generalist (host range,  $\geq 1.5$  hosts per species). Specialization depended significantly on taxonomic affiliation: 95% of the viruses but only 53% of the eukaryotes and 48% of the prokaryotes were specialist. No other system characteristic affected specialization degree (Supplementary Table 1), suggesting that our distribution of host range across genera was not biased by the sampling intensity of the original data set.

Tracking the evolutionary history of specialized interactions is not difficult, and has been performed for different kinds of interaction. Because in extremely specialized clades the host range is very narrow, it is easy to identify the host for each species in the phylogeny and to quantify host shifts and host conservatism. Specialized interactions are conserved when there is non-independence in host use among species within a clade owing to their phylogenetic relatedness. This can be tested by estimating the degree of phylogenetic signal, which is the tendency for related species to resemble each other in interaction patterns more than species randomly drawn from the phylogenetic tree do<sup>12</sup>. However, the use of this approach becomes increasingly difficult, to the point of becoming unfeasible, as the diversity of organisms interacting with the focal clade increases.

Generalist species interact with many other species, and therefore form networks of interacting organisms. Network analysis has been successfully used to analyse complex ecological interactions<sup>13,14</sup>. On the basis of the pattern of shared interactions, species can be grouped in compartments or modules. Species are tightly linked if they share a high proportion of interactions, that is, if they are ecologically similar. Groups of species interacting with similar organisms form modules within the general network<sup>15,16</sup>. Significant modularity emerges in a network when distinct groups of species closely share links with each other more than with species in other modules<sup>17</sup>. Using a network approach, we explore the evolution of ecological interactions by tracking the changes in module affiliation across the phylogenies (Fig. 1). Recent studies have quantified the effect of phylogenetic structure on the dynamics of ecological communities<sup>18,19</sup>. However, rather than taking the standard perspective on building networks, we use clade-oriented

<sup>1</sup>Departamento de Ecología, Universidad de Granada, E-18071 Granada, Spain. <sup>2</sup>Centro de Investigaciones sobre Desertificación, Consejo Superior de Investigaciones Científicas-Universidad de Valencia-Generalitat Valenciana, E-46470 Valencia, Spain. <sup>3</sup>Departamento de Genética, Universidad de Granada, E-18071 Granada, Spain.



**Figure 1 | How to study the evolution of both specialized and generalized interactions.** **a**, In specialized clades, each species (numbered tips in the phylogeny) interacts with a single host. Species are grouped according to the host used. Phylogenetic conservatism is determined by mapping such groups onto the phylogeny. **b**, In contrast, grouping species according to host use is more complex in generalized clades. Network analysis allows the detection of modules, which are groups of species sharing more hosts among themselves than they do with species in other modules. Phylogenetic conservatism of host use is determined by mapping such modules onto the phylogeny. Consequently, this method allows the exploration of the evolutionary conservatism of ecological interactions in all types of system, irrespective of their degree of generalization or host specificity.

networks (that is, groups of phylogenetically related species sharing a common ancestor but not necessarily co-occurring in the same locality). For extremely specialized clades, those showing high host specificity, each module contains the group of species interacting with the same host (Fig. 1a). Exploring phylogenetic conservatism in module ascription is analogous in these types of system to exploring phylogenetic conservatism in host use using the standard methodology. In clades in which species interact with more than one host, detecting modules of species sharing similar hosts allows for the exploration of phylogenetic conservatism even when it is not possible to group species according to their exact equivalence in host use (Fig. 1b). From this perspective, ecological interactions are conserved if a phylogenetic signal occurs in module affiliation.

For each clade we built a network (Fig. 1), including as nodes only the species for which host use is accurately known. In these networks, species were linked when they shared at least one host. We then used simulated annealing to establish significant modules within each network<sup>15,16</sup>. A modular network is one in which the clade is made up of species that can be grouped according to their affinity in host use. All but six clades were significantly modular (Supplementary Table 2). The modules of all studied clades, both specialist and generalist, differed significantly in identity and composition of the host assemblage ( $P < 0.01$  for all systems, from multivariate analyses of variance based on species composition dissimilarity<sup>20</sup>; Supplementary Table 3). Furthermore, modules within a network did not differ among themselves in number of hosts ( $P > 0.05$  for most systems; Supplementary Table 3), meaning that the emergence of modules in generalized clades was not due to differences in host range across species. Together, these results show that modules describe distinct and discrete interactive niches both in specialized and in generalized clades.

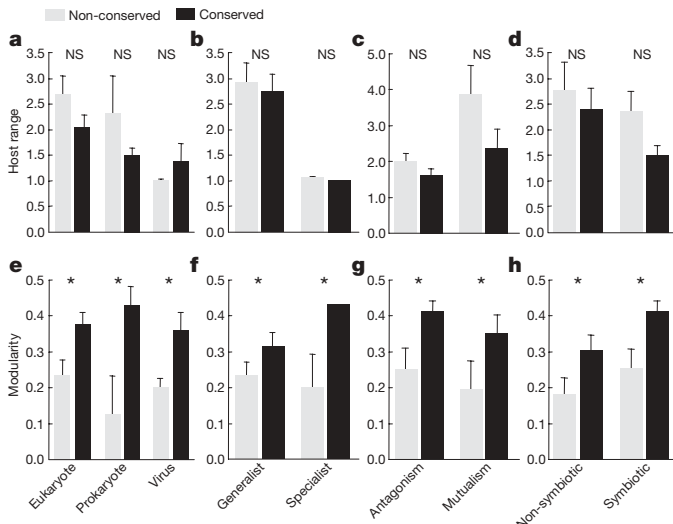
The number of modules in a given network may be considered a measure of the diversity of interactive niches occupied by the genus. The number of modules per network ranged from 2 to 20. It was

affected by the interaction intimacy, with symbiotic genera having more modules ( $6.2 \pm 0.5$ ,  $N = 75$  clades) than did non-symbiotic ones ( $4.2 \pm 0.4$ ,  $N = 40$  clades). The number of modules was negatively related to host range, even after controlling for number of species in the clade (Supplementary Table 4). This means that the number of distinct interacting niches is higher in specialized genera than in generalized ones. Domain also affected the number of modules per system, with viruses and prokaryotes having more modules per clade than did eukaryotes (Supplementary Tables 4 and 5). This may reflect a trend towards greater diversification of ecological niches and specialization in microorganisms.

For each system, we calculated the modularity index,  $M$ , which estimates how clearly delimited the modules are<sup>16</sup>. This index decreases when the fraction of between-module links increases in the total network. In the context of ecological interactions, this means that  $M$  is negatively related to the proportion of species that belong to different modules but share hosts. Consequently, it can be used as an estimate of the between-module differentiation in host use. Low values of  $M$  indicate no differentiation because many hosts are shared between different modules, and high values of  $M$  indicate high differentiation because the modules do not use common hosts. The extreme situation is exemplified by those genera in which most modules are completely isolated, without any links with the remaining modules (Supplementary Fig. 1). In our data set, modularity ranged from 0 to 0.833 (Supplementary Appendix 1). The value of  $M$  was higher in specialist clades ( $0.390 \pm 0.026$ ) than in generalist ones ( $0.232 \pm 0.030$ ). There was indeed a significant negative relationship between modularity and host range across clades (Supplementary Table 4). This means that between-module differentiation in host use decreases with the generalization of the clades. In fact, in community networks modularity is expected to increase with host specificity<sup>17</sup>. Similarly,  $M$  was higher in symbiotic ( $0.363 \pm 0.030$ ) clades than in non-symbiotic ones ( $0.239 \pm 0.034$ ; Supplementary Table 3), probably because symbionts tend to be more specialized (host range,  $1.77 \pm 0.17$ ) than do non-symbionts (host range,  $2.95 \pm 0.40$ ) ( $F = 9.63$ , d.f. = 1,108,  $P = 0.002$ , from one-way analysis of variance) and share fewer hosts between modules.

To explore how evolutionarily conserved ecological interactions are, we statistically tested whether phylogenetically related species were more prone to belonging to the same module than would be expected randomly<sup>21</sup> (that is, we tested for phylogenetic signal of ecological interactions). We found that over 83% of the specialist clades showed a significant phylogenetic signal for host use (Supplementary Data, appendix 1). Furthermore, 52% of the generalist clades also showed a significant phylogenetic signal. In fact, we found host range to have no effect on the probability of there being significant phylogenetic conservatism in ecological interaction (Fig. 2 and Supplementary Table 6). Similarly, the occurrence of conservatism in ecological interactions did not depend on the sign of the interaction (Supplementary Table 6), as 69% of antagonistic systems and 59% of mutualistic systems had a significant phylogenetic signal. That is, parasites and predators have the same probability of having phylogenetically conserved ecological interactions as do pollinators, seed dispersers or mycorrhizae.

There was a slight tendency of symbiotic systems to have more-conserved interactions (71% of the symbiotic systems had a significant phylogenetic signal) than do non-symbiotic systems (57% had a significant phylogenetic signal), although this difference was not statistically significant (Supplementary Table 6). Similarly, there was a tendency for phylogenetic conservatism to be more frequent in viruses (85%) and prokaryotes (80%) than in eukaryotes (59%), although this difference was also nonsignificant. Finally, the occurrence of a phylogenetic signal in our data set was significantly related to the number of species studied per clade<sup>5,18</sup> (Supplementary Table 6). In fact, if we remove from our data set those clades with fewer than 20 species, we find that 87% of specialist clades ( $N = 32$  systems) and 68% of generalist clades ( $N = 33$  systems) had a significant phylogenetic

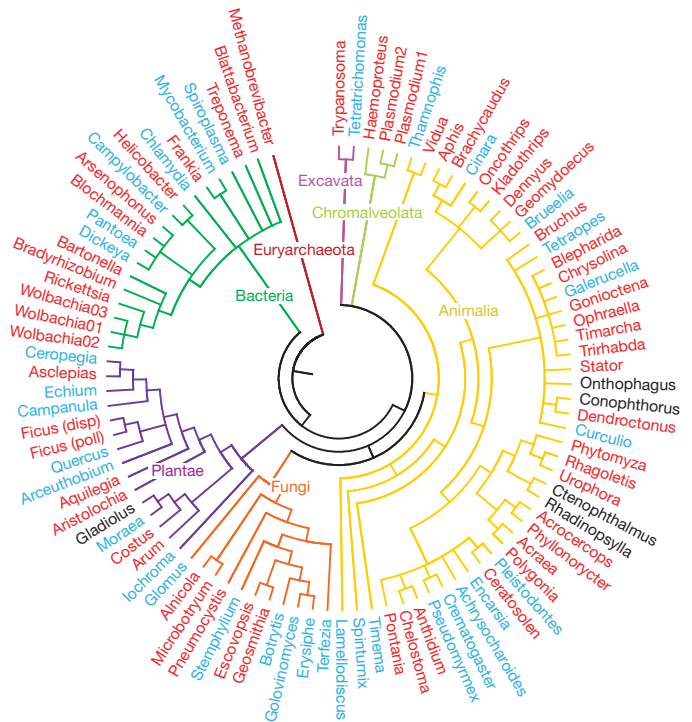


**Figure 2 | Differences between phylogenetically conserved and non-conserved clades in average host range per clade and modularity.** a–d, The occurrence of phylogenetic conservatism was not related to host range, indicating that, irrespective of the type of organism or interaction, the degree of generalization (quantitatively estimated as host range) was independent of the phylogenetic conservatism of the interactions. e–h, In contrast, the modularity index was consistently higher in phylogenetically conserved interactions, indicating that in these systems there was a decrease in the sharing of hosts among modules, irrespective of the type of organism. (average  $\pm$  s.e.m.,  $n = 110$ ,  $*P < 0.01$ ; NS, not significant).

signal. Most previous studies have stressed the ubiquity of phylogenetic conservatism in host use in specialist (mostly symbiotic) systems, from RNA viruses to herbivorous insects<sup>3–7,22–24</sup>. In agreement with this traditional view, several studies have pointed out that co-cladogenesis and phylogenetic conservatism in ecological interaction disappears when generalist species are included in the analyses<sup>7,24</sup>. Our study indicates, however, that ecological interactions are also conserved in generalist (both symbiotic and non-symbiotic) clades. Evolutionary conservatism in ecological interactions is a recurrent phenomenon across the entire tree of life (Fig. 3).

Clades with a significant phylogenetic signal in ecological interactions also had higher values of modularity, and this occurred in all kinds of organism (virus, prokaryote and eukaryote) and interaction (Fig. 2e–h). These findings indicate that clades with higher evolutionary conservatism in their ecological interactions also have stronger differentiation in host use among modules. This means that species belonging to the same module share few hosts with species from other modules in conserved systems, whereas in non-conserved systems species belonging to different modules tend to share some hosts. This probably occurs because the use of a specific host assemblage requires particular adaptations. In clades in which modules are conserved, species retain ancestral traits that influence their ecological interactions<sup>7,23</sup>, constraining the present and future capacity to use alternative hosts from other modules. In contrast, in non-conserved systems most traits involved in host use are likely to represent new adaptations. In this scenario, the species could possess adaptations for using alternative and disparate hosts. It is remarkable that this relationship was also found for generalist clades (Fig. 2f), despite the fact that modularity is negatively related to host range. That is, although generalist species usually share some hosts among different modules, among-module differentiation in the composition of their host assemblages is higher in evolutionarily conserved interactions. Conservatism in ecological interactions is associated with high host-use differentiation both in generalist and specialist organisms.

Our study has demonstrated that phylogenetic conservatism in ecological interactions is a general pattern occurring in many taxa belonging to very separate branches of the entire tree of life, from



**Figure 3 | Ecological interactions are evolutionarily conserved across the entire tree of life.** The phylogenetic tree shows the evolutionary relationships between the studied genera and the phylogenetic conservatism of the ecological interactions mapped onto the tips. Red names represent genera with conserved ecological interactions (that is, with a significant phylogenetic signal) and blue names represent genera with non-conserved ecological interactions (that is, with a nonsignificant phylogenetic signal). Black names represent genera in which the phylogenetic signal could not be tested because modularity was not significant or an accurate phylogeny was not available. Viruses were excluded because they are polyphyletic and cannot be accurately located within the tree.

viruses to animals, and in most types of interaction, from specialized symbiotic antagonisms to generalized non-symbiotic mutualisms. The same rules seem to drive the evolution of most ecological interactions and strongly contribute to the organization of biodiversity on the Earth.

## METHODS SUMMARY

We constructed bipartite networks ( $N = 116$  systems) of species belonging to the same genus and their known hosts. Species were then connected through the co-occurrence of interactions. We subsequently converted the bipartite networks into unipartite networks according to shared interactions. The modularity level and the number of modules per network were determined using an algorithm based on simulated annealing<sup>15,16</sup>. This algorithm identifies modules, which are groups of species having most of their links within their own module, with an accuracy of 90% (ref. 16). The modules were validated statistically by permutational multivariate analyses of variance based on species composition dissimilarity (using the function 'adonis' in the R package VEGAN)<sup>20</sup>. We determined phylogenetic conservatism in host use in each system by estimating the significance of the phylogenetic signal following ref. 25. The character 'host use' was the module to which the species was ascribed by the annealing algorithm. We mapped the evolution of host use onto published phylogenetic trees.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 2 February; accepted 22 April 2010.  
Published online 2 June 2010.

- Price, P. W. *Macroevolutionary Theory on Macroecological Patterns* (Cambridge Univ. Press, 2003).
- Thompson, J. N. *The Geographic Mosaic of Coevolution* (Univ. Chicago Press, 2005).
- Sasal, P., Desdevises, Y. & Morand, S. Host-specialization and species diversity in fish parasites: phylogenetic conservatism? *Ecography* 21, 639–643 (1998).

4. Jackson, A. P. & Charleston, M. A. A cophylogenetic perspective of RNA–virus evolution. *Mol. Biol. Evol.* **21**, 45–57 (2004).
5. Holmes, E. C. *The Evolution and Emergence of RNA Viruses* (Oxford Univ. Press, 2009).
6. Gilbert, G. S. & Webb, C. O. Phylogenetic signal in plant pathogen–host range. *Proc. Natl Acad. Sci. USA* **104**, 4979–4983 (2007).
7. Winkler, I. S. & Mitter, C. in *Specialization, Speciation and Radiation: The Evolutionary Biology of Herbivorous Insects* (ed. Tilmon, K. J.) 240–263 (Univ. California Press, 2007).
8. Freckleton, R. P., Harvey, P. H. & Pagel, M. Phylogenetic analysis and comparative data: a test and review of evidence. *Am. Nat.* **160**, 712–726 (2002).
9. Wiens, J. J. & Graham, C. H. Niche conservatism: integrating evolution, ecology, and conservation biology. *Annu. Rev. Ecol. Evol. Syst.* **36**, 519–539 (2005).
10. Darwin, C. *On the Origin of Species* 78–79 (Murray, 1859).
11. Chase, J. M. & Leibold, M. A. *Ecological Niches: Linking Classical and Contemporary Approaches* 19–45 (Univ. Chicago Press, 2003).
12. Revell, L. J., Harmon, L. J. & Collar, D. C. Phylogenetic signal, evolutionary process, and rate. *Syst. Biol.* **57**, 591–601 (2008).
13. Dunne, J. A., Williams, R. J. & Martinez, N. D. Food-web structure and network theory: the role of connectance and size. *Proc. Natl Acad. Sci. USA* **99**, 12917–12922 (2002).
14. Bascompte, J. & Jordano, P. Plant–animal mutualistic networks: the architecture of biodiversity. *Annu. Rev. Ecol. Evol. Syst.* **38**, 567–593 (2007).
15. Guimerà, R. & Amaral, L. A. N. Functional cartography of complex metabolic networks. *Nature* **433**, 895–900 (2005).
16. Guimerà, R. & Amaral, L. A. N. Cartography of complex networks: modules and universal roles. *J. Stat. Mech.* P02001 (2005).
17. Olesen, J. M., Bascompte, J., Dupont, Y. L. & Jordano, P. The modularity of pollination networks. *Proc. Natl Acad. Sci. USA* **104**, 19891–19896 (2007).
18. Rezende, E. L., Lavabre, J. E., Guimarães, P. R. Jr, Jordano, P. & Bascompte, J. Non-random coextinctions in phylogenetically structured mutualistic networks. *Nature* **448**, 925–928 (2007).
19. Rezende, E. L., Albert, E. M., Fortuna, M. A. & Bascompte, J. Compartments in a marine food web associated with phylogeny, body mass, and habitat structure. *Ecol. Lett.* **12**, 779–788 (2009).
20. Oksanen, J. *Multivariate Analysis of Ecological Communities in R: Vegan Tutorial* (R Project for Statistical Computing, 2008).
21. Blomberg, S. P., Garland, T. & Ives, A. R. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* **57**, 717–745 (2003).
22. Pride, D. T., Wassenaar, T. M., Ghose, C. & Blaser, M. J. Evidence of host–virus co-evolution in tetranucleotide usage patterns of bacteriophages and eukaryotic viruses. *BMC Genomics* **7**, 8 (2006).
23. Price, P. W. in *Specialization, Speciation and Radiation: The Evolutionary Biology of Herbivorous Insects* (ed. Tilmon, K. J.) 174–187 (Univ. California Press, 2007).
24. Refrégier, G. *et al.* Cophylogeny of the anther smut fungi and their caryophyllaceous hosts: prevalence of host shifts and importance of delimiting parasite species for inferring cospeciation. *BMC Evol. Biol.* **8**, 100 (2008).
25. Maddison, W. P. & Slatkin, M. Null models for the number of evolutionary steps in a character on a phylogenetic tree. *Evolution* **45**, 1184–1197 (1991).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank J. Bascompte, J. Bosch, A. González-Megías, P. Jordano, M. Lineham, M. Méndez, I. Reche, E.V. Schupp and S. Strauss for comments on a previous draft, R. Guimerà for kindly providing NETCARTO software, and B. Krasnov, C. Mitter, L. Navarro, J. Ollerton and J. M. Pleguezuelos for providing access to their data set. This work was funded by the Spanish Ministry of Science (J.M.G., M.V. and F.P.) and by the Junta de Andalucía (J.M.G. and F.P.).

**Author Contributions** J.M.G., M.V. and F. P. designed the study, J.M.G. compiled the data set and performed the analysis of host use, M.V. performed the phylogenetic analyses, J.M.G. wrote a first version of the manuscript and all authors contributed to the final draft.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to J.M.G. ([jmgreyes@ugr.es](mailto:jmgreyes@ugr.es)).



## METHODS

**The data set.** Our data set includes 116 genera belonging to seven kingdoms, as described. Because genus is the taxonomic level at which ecological interaction-mediated speciation mostly manifests, most macro-coevolutionary theories predict that coevolution generates the appearance of new species (diversifying coevolution), that radiation occurs within some genera because species escape from antagonistic interactions (escape-and-radiate coevolution) or that some lineages track the evolution of other lineages (sequential evolution). In all cases, the evolution of interactions is most apparent between related species, usually belonging to the same genus. In addition, using taxonomic levels above genus would make it difficult to explore the evolution of host use within specific interactions, as different types of interaction could appear within the same clade. The studied clades were also very diverse in their species number, ranging between 8 (*Terfezia*) and 2,400 species (*Onthophagus*). The list of these clades and their source references are included in Supplementary Information. We obtained host-use information from an average of  $63.3 \pm 4.3\%$  of the species belonging to the studied genera, with 36% of the clades being fully documented (meaning that we obtained information from all the species in the genus). As expected, completeness of the data set was negatively related to the number of species in a clade ( $1.70 - 0.27\log(N_s)$ ;  $N_s$ , number of species per system;  $R^2 = 0.68$ ,  $N = 116$  systems,  $t = 12.03$ ,  $P < 0.0001$ , from log-linear regression), because species-rich genera have traditionally been less intensely studied than have those with low numbers of species.

The taxonomic resolution of the hosts depended on the available literature for each studied system. It was family in 57 systems, genus in 26 systems, species in 12 systems, order in 7 systems, section in 5 systems, subfamily in 4 systems, class in 3 systems and tribe in 2 systems. To check for potential biases produced by the taxonomic resolution of the hosts, we performed all the statistical analyses including host taxonomic resolution as covariate. This inclusion of taxonomic resolution did not change the results of the analyses.

**Clade-oriented network analysis.** Network analysis has been successfully applied to the study of ecological communities in recent decades<sup>14</sup>. In this study, we have extended the application of network tools to the study of phylogenetically related clades. We constructed bipartite networks of species belonging to the same genus and their known hosts throughout their distribution ranges. Species were then connected through the co-occurrence of interactions. We subsequently converted the bipartite networks into unipartite networks according to shared interactions. Consequently, we obtained networks that connected the species of our focal systems on the basis of their similarity in the range of host species used. All network analyses were done with PAJEK<sup>26</sup>.

**Modularity analysis.** The modularity level and the number of modules per network was determined using an algorithm (in the software NETCARTO) based on simulated annealing and provided by R. Guimerà<sup>15,16</sup>. This algorithm identifies modules with an accuracy of 90% (ref. 16). For each network, we calculated the index of modularity,  $M$  (a measure of the extent to which species have more links within their modules than would be expected if linkage was random), as

$$M = \sum_{s=1}^r \left( \frac{l_s}{L} - \left( \frac{d_s}{2L} \right)^2 \right)$$

where  $r$  is the number of modules in the network,  $L$  is the number of links in the network,  $l_s$  is the number of links between nodes in module  $s$  and  $d_s$  is the sum of

degrees of the nodes in module  $s$  (refs 15, 16). Because random networks also may have strong modularity<sup>27</sup>, we explored whether our networks were significantly more modular than random networks by running the same simulated annealing algorithm in 100 random networks with the same species degree distribution as the empirical one<sup>15</sup>. This method produces a modularity index that is a measure of the degree to which the network is organized into clearly delimited modules.

**Statistical validation of modularity.** The modules were validated statistically by permutational multivariate analyses of variance using distance matrices (using the function 'adonis' in VEGAN<sup>20</sup>), which test whether element similarity (that is, similarity between species as a function of host use and similarity between host taxa as a function of their interacting organisms) was significantly higher within than between modules. The function 'adonis' partitions dissimilarities for the sources of variation, and uses permutation tests to inspect the significances of those partitions. Dissimilarity was calculated as a Bray–Curtis distance.

**Phylogenetic conservatism.** Phylogenetic conservatism in host use was determined in each system by estimating the significance of the phylogenetic signal following ref. 25. This test estimates whether the minimum number of evolutionary steps in a character on a phylogenetic tree is lower than the chance result. This was determined under a null model in which data were reshuffled 1,000 times across the tips of the phylogeny. The character 'host use' was the module to which the species was ascribed by the annealing algorithm. Module was considered as an unordered, multistate factor. We mapped the evolution of host use onto published phylogenetic trees. If a phylogeny was not available but DNA sequences were available in GenBank (*Timarcha* and *Alnicola*), we inferred the tree on the basis of Bayesian methods. Ultimately, we obtained the phylogenies of 111 genera. We recovered the topology but not the branch lengths of these 111 phylogenetic trees because the phylogenetic signal test we used is based on parsimony and branch lengths are therefore not necessary. We did not use likelihood approaches to take advantage of the information on branch lengths because of the limitation associated with the number of states of the character being high<sup>28</sup>. This limitation results from the fact that likelihood approaches need to estimate the rates at which a discrete character makes transitions among its possible states as it evolves through time<sup>27</sup>. With the high number of states encountered in most of our phylogenies (mean, 6; range, [2, 18]), the number of transition rates to estimate is extremely large. All tests were performed using MESQUITE 2.71<sup>29</sup>.

**Statistical analyses.** We used generalized linear models to test the effect of several characteristics of the studied system on their host range, modularity and phylogenetic conservatism. In these models, we included as explanatory variables the type of interaction (mutualistic versus antagonistic), intimacy (symbiotic versus non-symbiotic), classical taxonomic domain (eukaryote, prokaryote and virus), host range (when not used as a dependent variable) and sample size (the number of species appearing in the phylogenies).

26. de Nooy, W., Mrvar, A. & Batagelj, V. *Exploratory Social Network Analysis with Pajek: Structural Analysis in the Social Sciences* (Cambridge Univ. Press, 2005).

27. Guimerà, R., Sales-Pardo, M. & Amaral, L. A. N. Modularity from fluctuations in random graphs and complex networks. *Phys. Rev. E* **70**, 025101 (2004).

28. Pagel, M. Inferring the historical patterns of biological evolution. *Nature* **401**, 877–884 (1999).

29. Maddison, W. P. & Maddison, D. R. Mesquite: a modular system for evolutionary analysis. *Mesquite* (<http://mesquiteproject.org>) (2009).