

"PRINCIPLES OF PHYLOGENETICS: ECOLOGY AND EVOLUTION"

Integrative Biology 200B
University of California, Berkeley

Spring 2011
B.D. Mishler

April 12, 2011. **Comparing cladograms; supertrees; coevolution; symbiosis**

-- There are many reasons why one would want to compare cladograms, falling into three basic categories:

-- *Within an analysis of one clade, with the same OTUs*; e.g., equally or nearly equally parsimonious (or likely) trees, trees resulting from different character partitions, models of evolution, or methods of analysis, and comparisons with trees from the literature.

-- *Within an analysis of one clade, with different OTUs*; trying to come up with a general tree for all OTUs, e.g. super trees, compartmentalization.

-- *Comparing analyses of different clades*, e.g., gene family evolution, migration between populations, vicariance biogeography, host/ parasite relationships, symbiosis, community evolution, or any long-term ecological association

Methodology for comparing cladograms:

(1) consensus techniques (strict, semi-strict, majority rule, Adams) -- for finding shared signal among trees.

Strict consensus: Only monophyletic groups found in all source trees are found in the resultant tree. The tree excludes a subset of all possible trees and conversely includes a subset of possible trees, whether or not they are part of the source set, e.g. $(A(B(CD))) + (A(C(BD))) = (A(BCD))$ but this also implies $(A(D(BC)))$. In some sense the most conservative consensus.

Semistrict consensus: Only monophyletic groups found in at least **one** of the source trees and compatible (not in conflict) with all other source trees are found in the resultant tree, i.e. if a clade is never contradicted, but not always supported, then it is still included in this compromise tree. E.g. $(A(B(CD))) + (A(BCD)) = (A(B(CD)))$

Majority-rule consensus: Shows groups that appear in more than a pre-specified percentage of source trees, usually >50%. Not recommended for summary of equally-optimal trees resulting from a search.

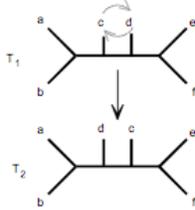
Adams Consensus: Inconsistently placed taxa are moved down to the first node that summarizes the possible topologies. N.B., groups can appear in Adams consensus that are not found in **any** source tree. Adams trees have no biological or phylogenetic interpretation, but they do point to "wildcard" taxa. Those taxa may be experimentally removed from the matrix and the resulting analysis compared to when they are included.

(2) tree-to-tree distance metrics. There are two types of approaches. One counts the number of steps needed to transform one tree into another (e.g., NNI interchange metric, partition metrics, agreement subtrees). The second represents two trees as sets of simpler structures and then measures similarity between these (e.g., quartet measures)

Transforming one tree into another

A good example of a measure defined in terms of transforming one tree into another is the nearest neighbor interchange (NNI) metric (e.g., Waterman and Smith, 1978) which measures the minimum number of NNIs required to change T_1 into T_2 . In the example below, one NNI is required to convert T_1 into T_2 , so $d_{NNI}(T_1, T_2) = 1$.

Figure 5.1
Transforming T_1 into T_2
by a single nearest
neighbor interchange of
leaves c and d



from the Component User's Guide, by Rod Page
(<http://taxonomy.zoology.gla.ac.uk/rod/cplite/title.pdf>)

(3) component analysis (more next week in biogeography) -- finding individual statements of relationship that are shared among trees, basically a node relating some taxa to the exclusion of others.



(4) maximum likelihood approaches (parametric bootstrapping) -- comparing alternative trees or alternative models of evolution for your data.

(5) representing the grouping information in separate trees as characters in a matrix (e.g., using Brooks parsimony, also called "matrix representation parsimony"). This might be used when comparing hosts and parasites, or phylogenies of different taxa that all live in the same areas of endemism. See example below and on board (see Brooks & McLennan, 1991; Brooks 1981, Syst. Zool. 30:229; Wiley 1988, Syst. Zool. 37:271; and see Kluge 1988, Syst. Zool. 37:315 for some suggested modifications)

Steps:

1. Cladogram of parasite group
2. Cladogram of host group
3. Cladogram of parasite group is taken as a completely polarized, multistate transformation series -- recoded by additive binary coding
4. make new data matrix with hosts as OTU's and parasite clades as characters
5. construct new host cladogram from this matrix
6. compare this new host cladogram (derived from cladogram of the parasite group) with the original host cladogram (derived from host characters) and with host cladograms based on other parasite groups, if possible. Congruence is taken as evidence of common cause (shared history); incongruence (homoplasy) is taken as due to separate causes (e.g., host-switching or extinction)

Problem resolution:

1. when more than one parasite occurs in given host, codes are combined (e.g., host E in example combines codes of 10 & 5)
2. when all members of parasite clade are missing from a host taxon, host coded with "?"
3. when one parasite occurs in more than one host, codes are combined (e.g., if parasite species 10 occurred in hosts A,B,& E) -- note that this is controversial -- some have suggested downweighting such species, even eliminating them from the analysis entirely.

Methodology:

-- These kinds of questions always involve comparing different cladograms, yet how to do this can be hypothesis-dependent. In addition to comparing topologies per se, some questions would require comparing branch lengths on the topologies, or at least comparing the presence or absence of specific characters. See specific examples on the following sheets.

-- Like other areas of comparative methods we have discussed, the general approach is to first define the patterns you are looking for, carefully define the causal hypothesis to be tested, then specify a null hypothesis (what you would expect if the hypothesized cause is NOT working), and finally design a test that would let you reject the null hypothesis if it is indeed false. These sorts of comparative cladogram studies are in their infancy, and you could make contributions to both methods and empirical results.

Citations:

- Brooks, D.R., & D. McLennan. 1991. *Phylogeny, Ecology, and Behavior*. University of Chicago Press.
- de Vienne, D.M., Giraud, T. & Shykoff, J.A. (2007) When can host shifts produce congruent host and parasite phylogenies? A simulation approach. *Journal of evolutionary biology*, **20**, 1428-1438.
- Hafner, M. S., and R. D. M. Page. 1995. Molecular Phylogenies and Host-Parasite Cospeciation - Gophers and Lice as a Model System. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 349:77-83.
- Hafner, M. S., and S. A. Nadler. 1988. Phylogenetic Trees Support the Coevolution of Parasites and Their Hosts. *Nature* 332:258-259.
- Huelsenbeck, J. P., B. Rannala, and B. Larget. 2000. A Bayesian framework for the analysis of cospeciation. *Evolution* 54:352-364.
- Page, R. D. M. 1996. Temporal congruence revisited: Comparison of mitochondrial DNA sequence divergence in cospeciating pocket gophers and their chewing lice. *Systematic Biology* 45:151-167.
- Page, R. D. M., and M. A. Charleston. 1998. Trees within trees: phylogeny and historical associations. *Trends in Ecology & Evolution* 13:356-359.
- Page, R.D.M., editor. 2003. *Tangled trees: phylogeny, cospeciation, and coevolution*. Univ. of Chicago Press, Chicago.
- Percy, D. M., R. D. M. Page, and Q. C. B. Cronk. 2004. Plant-insect interactions: Double-dating associated insect and plant lineages reveals asynchronous radiations. *Systematic Biology* 53:120- 127.
- Jousselin, E., van Noort, S., Berry, V., Rasplus, J.Y., Rønsted, N., Erasmus, J.C. & Greeff, J.M. (2008) One fig to bind them all: host conservatism in a fig wasp community unraveled by cospeciation analyses among pollinating and nonpollinating fig wasps. *Evolution; international journal of organic evolution*, **62**, 1777-1797.
- Stone, G.N., Hernandez-Lopez, A., Nicholls, J.A., di Pierro, E., Pujade-Villar, J., Melika, G. & Cook, J.M. (2009) Extreme host plant conservatism during at least 20 million years of host plant pursuit by oak gallwasps. *Evolution; international journal of organic evolution*, **63**, 854-869.
- Weiblen, G. D., and G. L. Bush. 2002. Speciation in fig pollinators and parasites. *Molecular Ecology* 11:1573-1578.

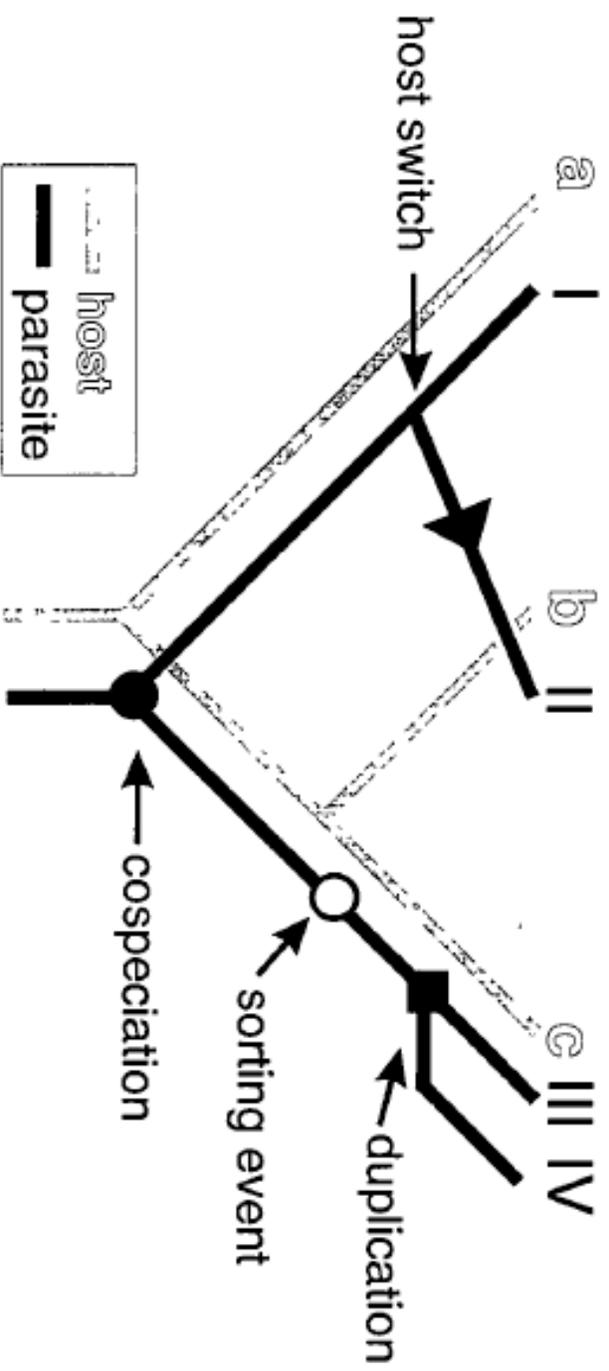


FIGURE 2. A host and parasite phylogeny showing a cospeciation event (●), a duplication (■), a host switch (▶), and a sorting event (○).

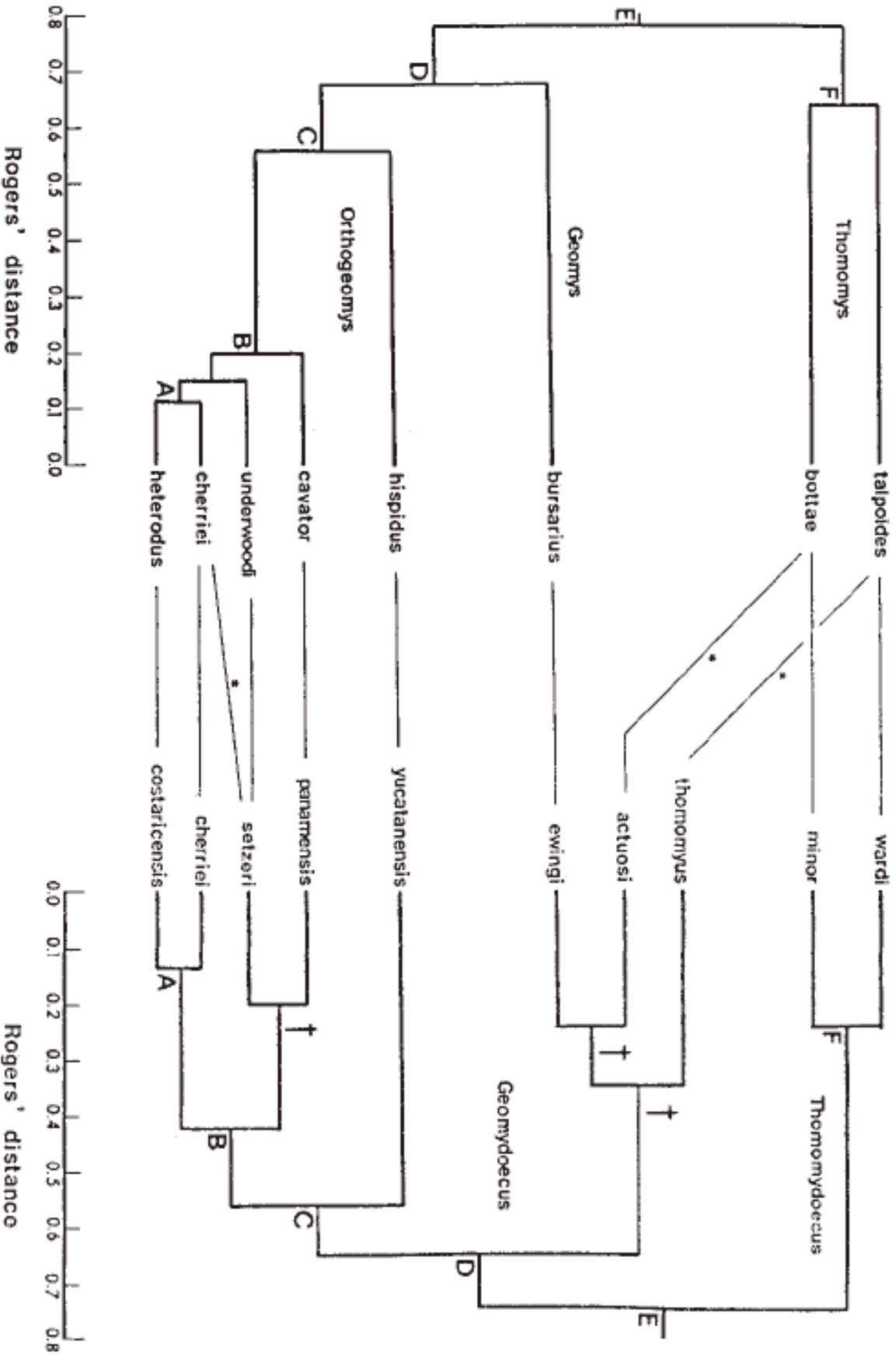
3 kinds of speciation events for parasite

sorting; extinction; failure to speciate; 'miss the boat'; missed collection

Pocket gophers

Host-parasite associations

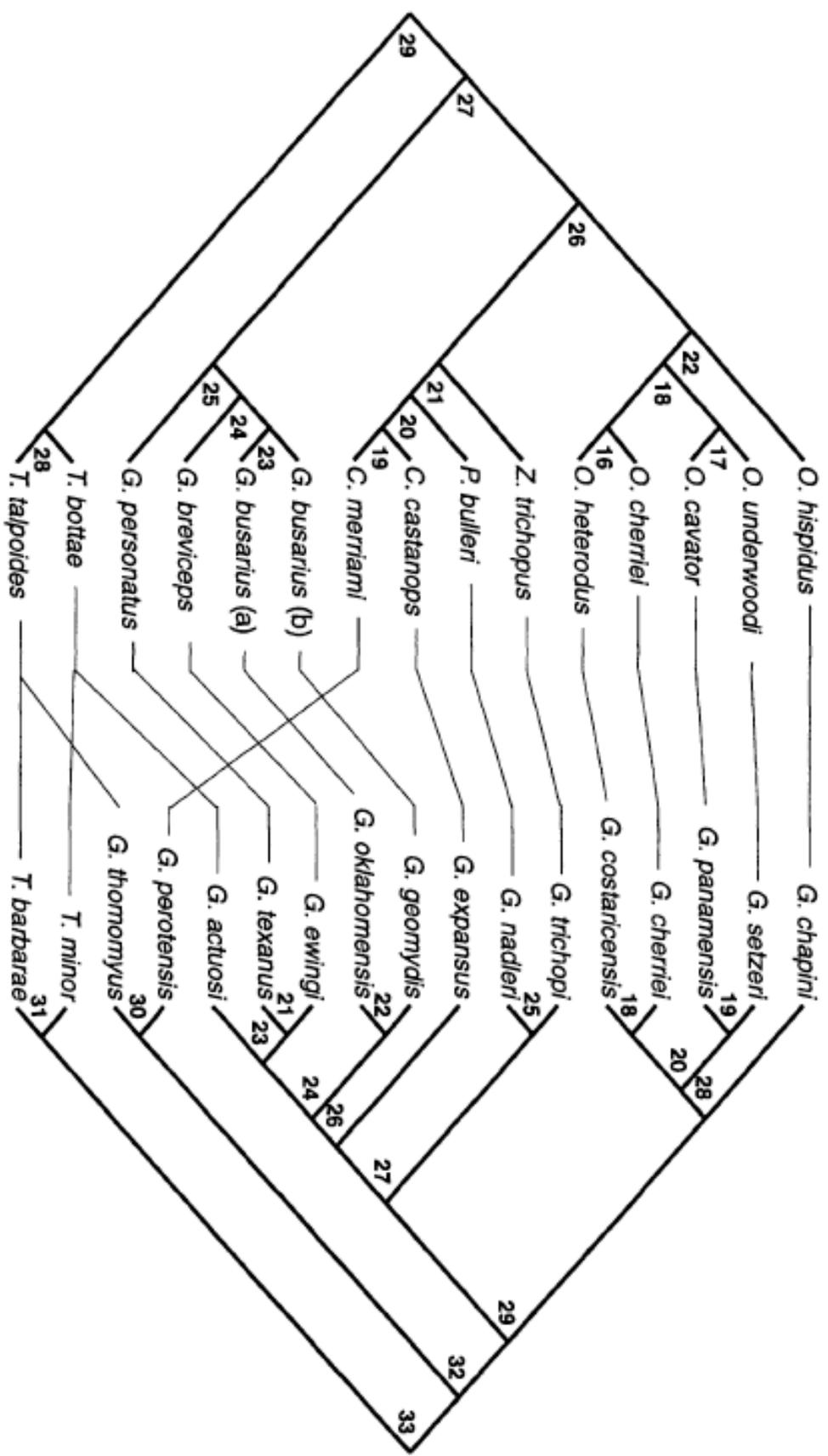
Chewing lice



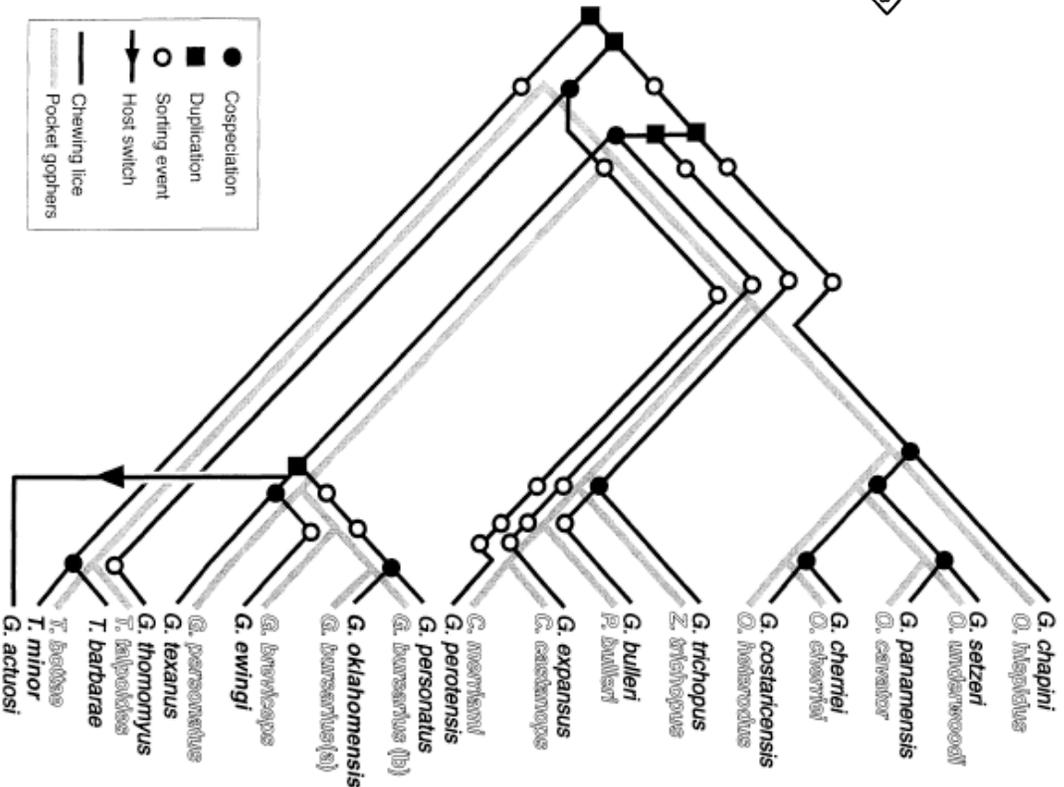
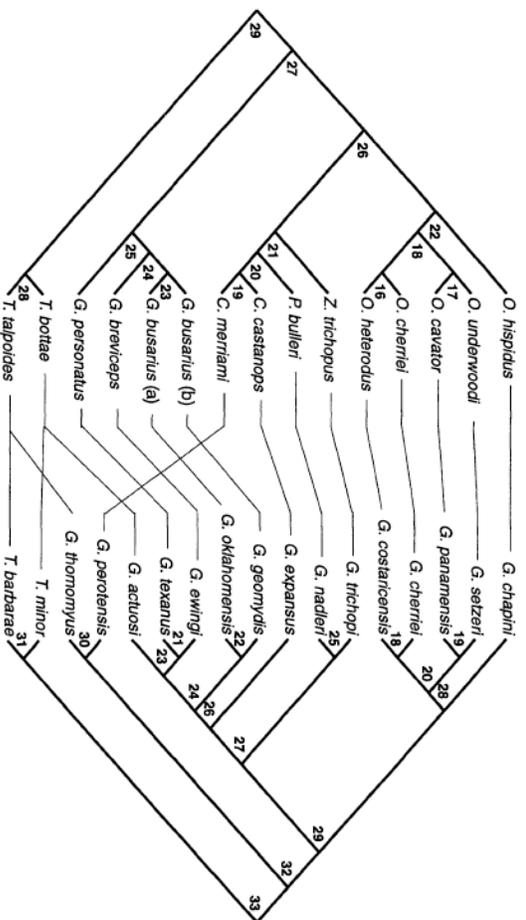
(a)

Pocket gophers

Chewing lice



(a) Pocket gophers Chewing lice



● Cospeciation
 ■ Duplication
 ○ Sorting event
 ▲ Host switch
 — Chewing lice
 - - - Pocket gophers

FIGURE 3. A possible reconstruction of the history of the gopher-lice association that postulates 10 cospeciation events, five duplications (in situ speciation of the lice on the same host), 20 sorting events (instances where louse lineages have been lost or remain undetected), and a single host switch (by *Geomydoceus actiosi*).

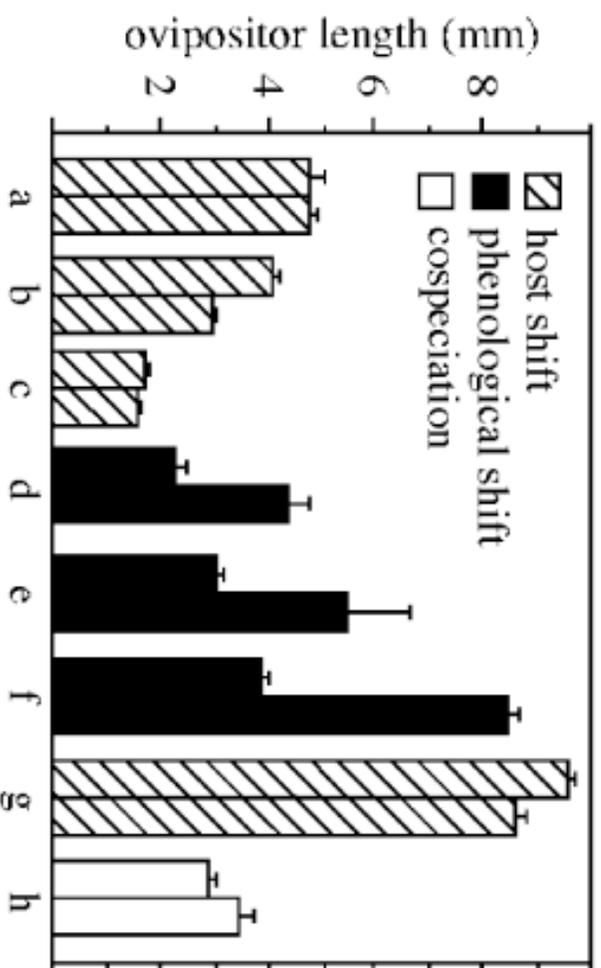


Fig. 3 Sister group comparisons of ovipositor length in the fig parasite, *Apocryptophagus*, under three alternative modes of speciation. (a–h) refer to sister groups in Fig. 2. Sister species (h) attacking sister hosts, *F. adenosperma* and *F. ochrochlora*, represent cospeciation. In contrast (a–c, g) are sister species on nonsister hosts, indicating potential instances of host switching. Closest relatives attacking the same host are represented by (d–f). We predict less divergence in ovipositor length between sister species in cases of cospeciation and host switching than in the case of a phenological shift, where divergence results from a shift in the relative timing of oviposition. Greater ovipositor length divergence between sister species (d–f) indicates a relatively large shift in the timing of oviposition, as the thickness of the fig wall increases during development (Kerdelhue & Rasplus 1996).

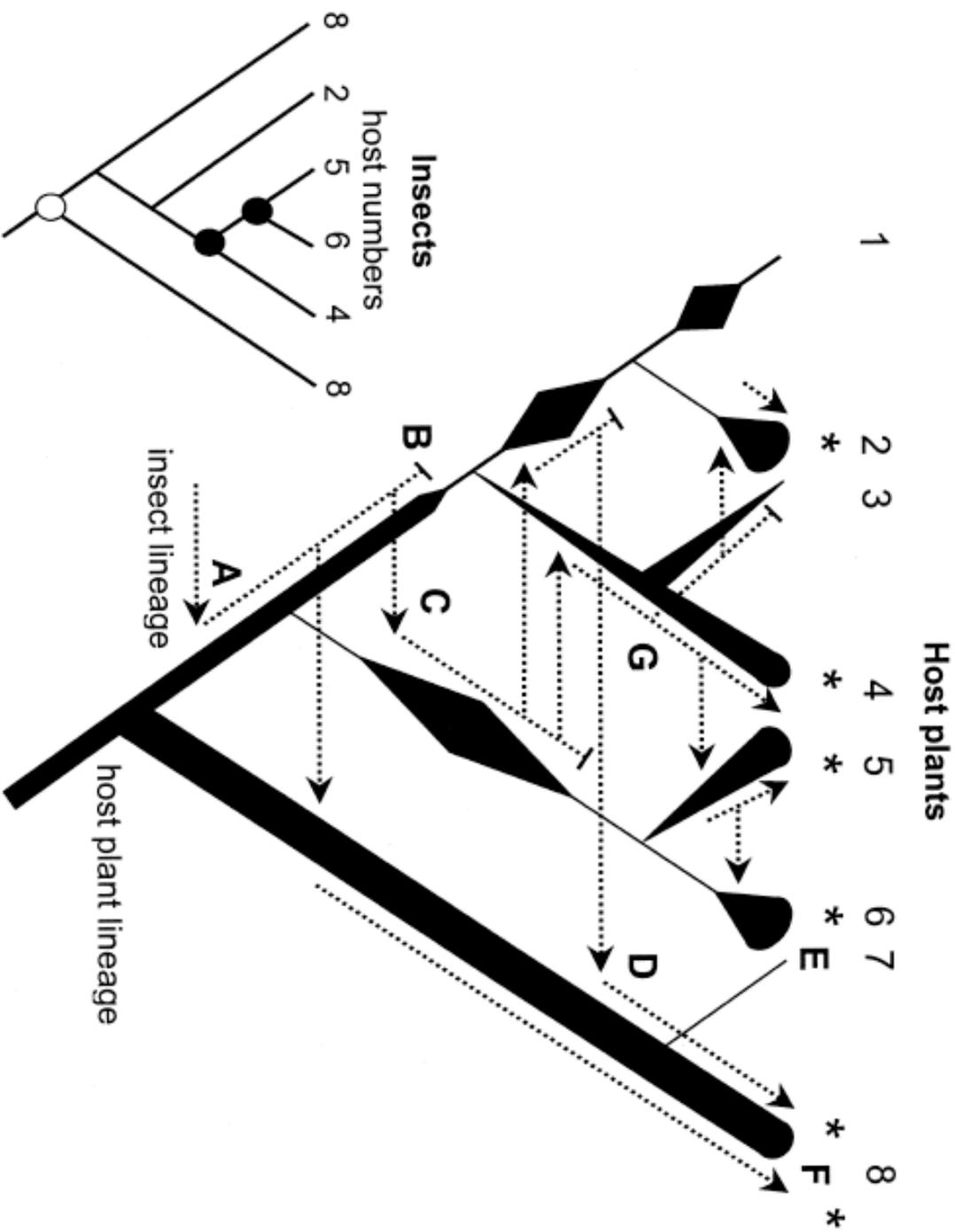
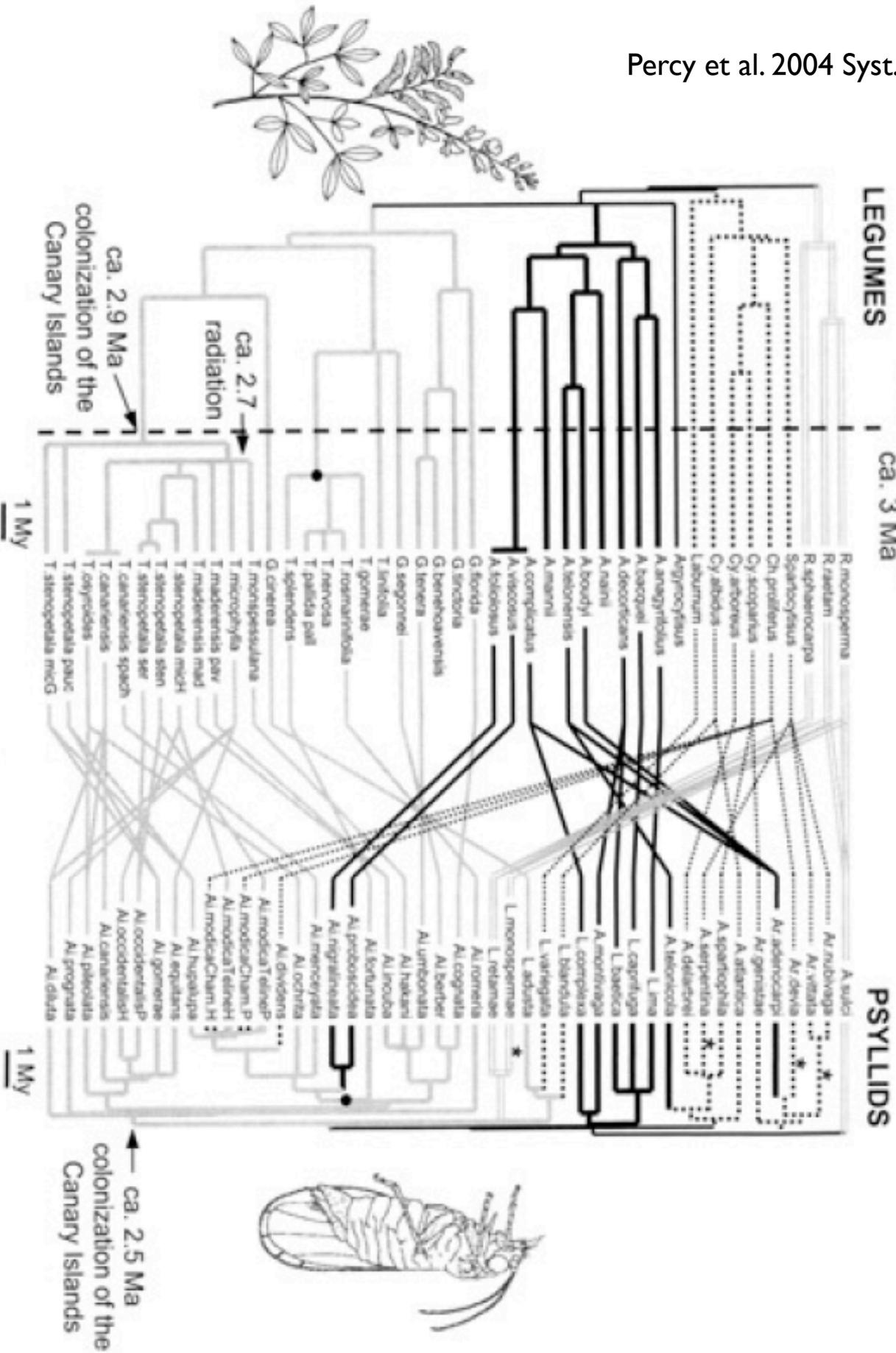


FIGURE 1. A hypothetical fluctuating host plant lineage (right), illustrating the factors that may influence how an insect lineage interacts with a fluctuating plant lineage and an interpretation of the resultant phylogenetic patterns (left). This model takes into account the susceptibility factors for host switching suggested by examination of the ecology of Canary Island psyllids and their legume hosts. Host abundance is indicated by line width. Examples of different events are labeled: A = colonization of preexisting host lineage by insect lineage; B = loss of insect due to extinction; C = increase in host abundance after bottleneck leads to vacant host, usually filled by near host switching from related plants; D = abundant host and geographical proximity may lead to wide host switching; E = rare plants are unlikely to be colonized; F = temporarily stable host lineages may sustain multiple insect lineages; G = occasional cospeciation events. The asterisks indicate extant insect taxa, whose phylogeny is shown (left). Given the two phylogenies, certain events would be misinterpreted in a TreeMap analysis: the solid circles are cospeciation events, and the open circle is a duplication event (speciation of parasites in situ on the host).

colonization of the
Genisteae by psyllids:



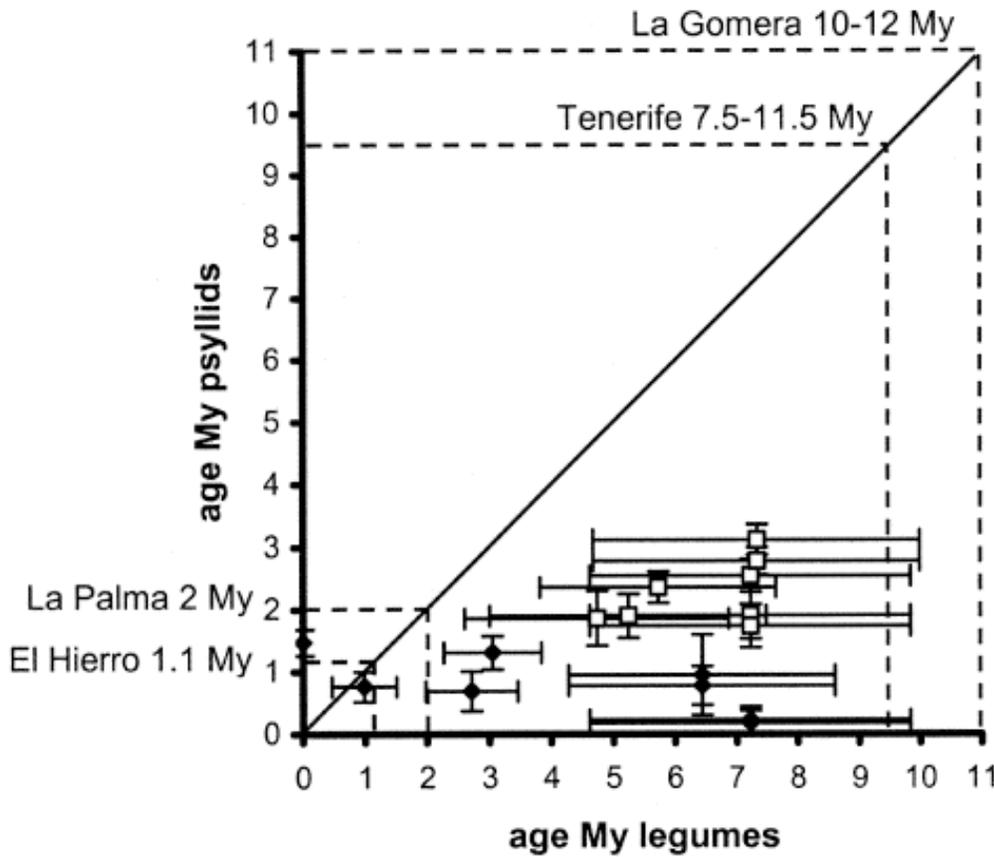


FIGURE 3. Comparative ages of 16 putative cospeciation events suggested by TreeMap (\square = continental; \blacklozenge = island). Reconciled psyllid and legume phylogenies, without considering dates of divergence, produce associated legume nodes that are much older than the putatively synchronous psyllid nodes. This plot shows the cospeciation events determined by TreeMap, which optimizes a psyllid node onto the most recent common ancestor of all its host plants (i.e., no assumption of host switching). All but one of the results are temporally implausible. The dates plotted represent the means and SDs of 100 bootstrap replicates.

FIGURE 2. A tanglegram of the insect–host–plant relationships of a group of legume-feeding psyllids (right) and their Genisteae (Broom) hosts (left). Four main legume lineages are distinguished: *Retama* (double gray line), *Cytisus* (dotted black line), *Adenocarpus* (thick black line), and *Genista* (thick gray line). The psyllids on these hosts are indicated by the same type of branches. Most of the psyllids on each host group are from a single psyllid group, but some psyllid–legume associations have involved a wide host switch between legume groups. The dates on the psyllid tree are much younger than the dates on the legume tree, indicating a colonizing, not a cospeciating, psyllid lineage. Calibration points are indicated by solid circles. Of several independent colonizations of the Canary Islands by legumes and psyllids, only one could have been contemporaneous (around 2.5 and 2.9 MY ago), with a single potential cospeciation event during the radiation of legumes and psyllids in the Canary Islands (Fig. 3). The evolution of other island psyllid lineages (indicated by asterisks) substantially postdate the evolution of their island legume hosts. Psyllid genera: *Arytaina* (Ar.), *Arytainilla* (A.), *Arytainis* (Ai.), *Lirilla* (L.); legume genera: *Chamaecytisus* (Ch.), *Cytisus* (Cy.), *Genista* (G.), *Telme* (T.).

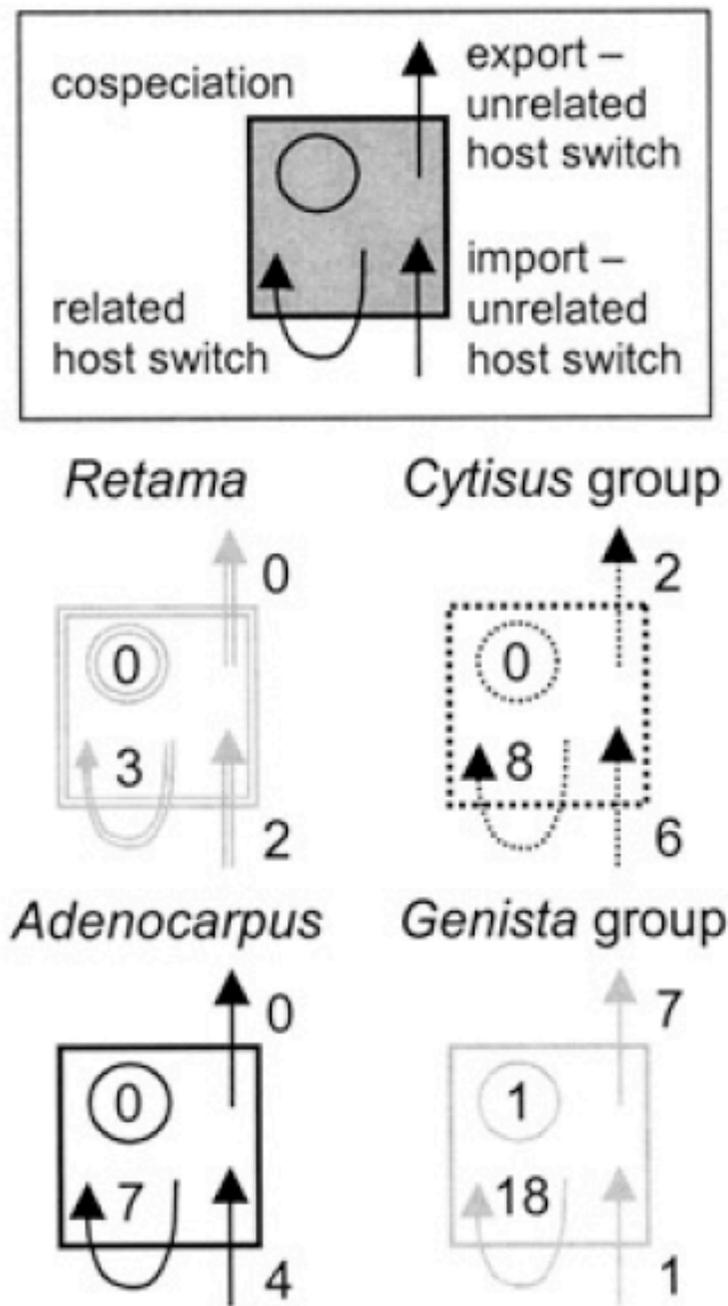


FIGURE 4. The levels of different types of host switching and cospeciation implied for the four major legume lineages, on the assumption that the presence of asynchronous nodes rules out cospeciation over most of the association. The legume lineages are coded as in Figure 2: *Retama* (double gray line), *Cytisus* (dotted black line), *Adenocarpus* (thick black line), and *Genista* (thick gray line). The only possible cospeciation events are found in the *Genista* group, abundant in the Canary Islands and Madeira, where there has been a recent radiation of both legumes and psyllids. The majority of the host switching (61%) apparently occurred between related hosts (within major legume lineages) and produced some shared phylogenetic patterns that can be misinterpreted as cospeciation.