

April 29 and May 1, 2020. **Comparing cladograms; cospeciation; coevolution; symbiosis**

**I. Comparing trees.** There are many reasons why one would want to compare trees, 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.

-- *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. We will cover these later in the class, but for now just be aware that most of the same principles obtain.

**Methodology for comparing phylogenies (some of this is review of Kip's Feb 24 lecture):**

(1) *consensus techniques* (strict, semi-strict, majority rule, Adams) -- the goal being 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.

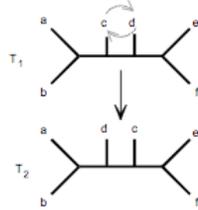
(2) *tree-to-tree distance metrics*. There are two types of approaches. One counts the number of steps needed to transform one tree into another (e.g., NNI interchange metric, partition metrics, agreement subtrees). The second represents two trees as sets of simpler structures and then measures similarity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

The Robinson Foulds metric is a commonly used metric; defined as the sum of [the number of partitions of data implied by the first tree but not the second tree] plus [the number of partitions of data implied by the second tree but not the first tree]. It is also called the symmetric difference metric (Robinson, D. R., Foulds, L. R. 1981. *Mathematical Biosciences*, **53**, 131-147).

**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>)

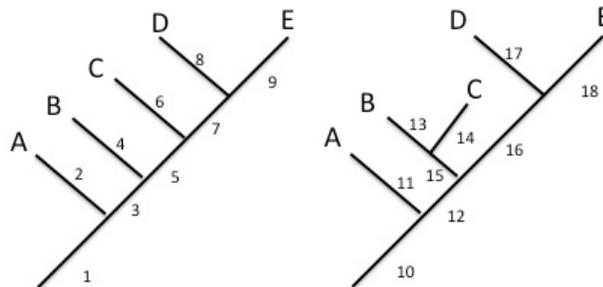
(3) *component analysis* (more in the biogeography lectures) -- finding individual statements of relationship that are shared among trees, basically a node relating some taxa to the exclusion of others.



(4) *Maximum likelihood approaches* (e.g., parametric bootstrapping) -- comparing alternative trees or alternative models of evolution for your data (e.g., Efron, et al. 1996. Proceedings of the National Academy of Sciences 93: 13429).

(5) *Brooks parsimony*, i.e., 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 endemcity. [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.]

Simple example:  
two trees being  
compared/combined  
for the same taxa



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
A	1	1	0	0	0													
B	1	0	1	1	0													
C	1	0	1	0	1													
D	1	0	1	0	1													
E	1	0	1	0	1													

*In class exercise: fill in the rest of this character matrix. Then when you are all done, what topology would the matrix support?*

(6) *Supertrees* are one of the frequent applications of tree comparisons, in this case attempting to combine different trees of the same larger clade that were developed from different sets of OTUs. In the simplest case, detailed phylogenies of individual genera are stitched together using a backbone phylogeny of a family that might have one representative of each genus. For a more analytical approach, Brooks parsimony can be used (branches in the separate trees are represented in a data matrix for analysis).

(7) *Supermatrices* are at the opposite end of the spectrum from the supertree approach -- these are so-called total evidence analyses that concatenate all the data types into one matrix. Note that different models can be used for different partitions (i.e., different genes, morphological data, etc.).

## II. "Coevolution" and "Symbiosis"

How so we define coevolution, exactly? It can be over-broadly defined to include *any* interaction between different lineages; that makes it basically synonymous with plain ol' "evolution." So most people take it to mean a *long-term* interaction between lineages -- but how long-term? Janzen (1980) defined it in terms of trait evolution: "evolutionary change in a trait of the individuals of one population in response to a trait of the individuals of a second population, followed by an evolutionary response by the second population to the change in the first." (this definition essentially follows the original meaning from Ehrlich and Raven 1968).

Coevolution can occur in the context of many different ecological interactions. The interactions can be mutually beneficial, or beneficial to one lineage and either neutral or detrimental to the other. The term symbiosis refers to a *close and prolonged ecological relationship* between the individuals of two (or more) different species, and can involve mutualism, parasitism or other interactions.

---

Name of Interaction	Effect of Interaction	
	Species A	Species B
Competition	-	-
Predation/Herbivory/Parasitism	+	-
Mutualism	+	+
Commensalism/Facilitation	0	-

---

If coevolution is defined to be a relationship that is long enough to continue through one or more divergences of the lineages involved, then it can be studied directly using phylogenetic comparative methods. Prolonged coevolution can lead to *cospeciation*, which will appear as phylogenetic congruence between two or more systems of lineages.

Cospeciation represents another generalization of the phylogeny/ homology relationship (which could be considered the "coevolution" of organism lineages

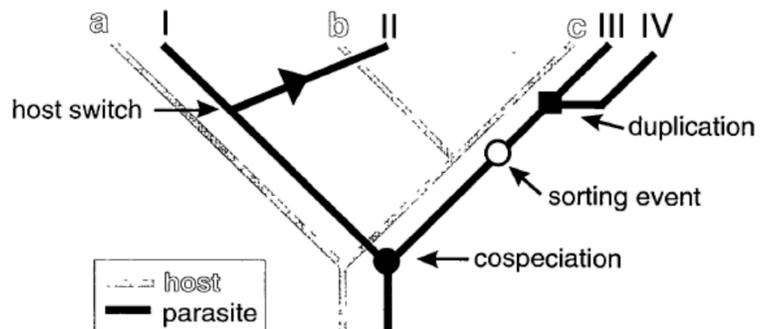


FIGURE 2. A host and parasite phylogeny showing a cospeciation event (●), a duplication (■), a host switch (▶), and a sorting event (○).

and their characters), like other examples we have discussed before such as vicariance biogeography (which is organism/ earth coevolution) or gene family evolution (which is gene/genome coevolution).

Some cool emerging applications in this area include phylogenetic approaches to microbiomes and ecological networks

### **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.

### **Some of the literature:**

- Althoff, David M., Kari A. Segraves, and Marc TJ Johnson. "Testing for coevolutionary diversification: linking pattern with process." *Trends in ecology & evolution* 29.2 (2014): 82-89.
- Bascompte J, Jordano P, Olesen JM. Asymmetric coevolutionary networks facilitate biodiversity maintenance. *Science*. 2006 Apr 21;312(5772):431-3.
- Brooks, D.R., & D. McLennan. 1991. *Phylogeny, Ecology, and Behavior*. University of Chicago Press.
- Charleston, MA and DL Robertson. 2002. Preferential host switching by primate lentiviruses can account for phylogenetic similarity with the primate phylogeny. *Syst Biol*, 51(3):528-535.
- 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.
- Drury JP, Grether GF, Garland T, Morlon H. 2017. An assessment of phylogenetic tools for analyzing the interplay between interspecific interactions and phenotypic evolution. *Systematic biology*.
- 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.
- Huelsenbeck, J. P., B. Rannala, and B. Larget. 2000. A Bayesian framework for the analysis of cospeciation. *Evolution* 54:352-364.
- 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.
- Kembel SW, Eisen JA, Pollard KS, Green JL. The phylogenetic diversity of metagenomes. *PLoS One*. 2011 Aug 31;6(8):e23214

- Martiny JB, Jones SE, Lennon JT, Martiny AC. Microbiomes in light of traits: a phylogenetic perspective. *Science*. 2015 Nov 6;350(6261):aac9323.
- Merkle D, Middendorf M. 2005. Reconstruction of the cophylogenetic history of related phylogenetic trees with divergence timing information. *Theor Biosci*. 123:277–299.
- O'Dwyer JP, Kembel SW, Green JL. Phylogenetic diversity theory sheds light on the structure of microbial communities. *PLoS computational biology*. 2012 Dec 20;8(12):e1002832.
- 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.
- Pérez-Valera E, Goberna M, Faust K, Raes J, García C, Verdú M. Fire modifies the phylogenetic structure of soil bacterial co-occurrence networks. *Environmental microbiology*. 2017 Jan 1;19(1):317-27.
- Refrégier, G., Le Gac, M., Jabbour, F., Widmer, A., Shykoff, J. A., Yockteng, R., ... & Giraud, T. (2008). Cophylogeny of the anther smut fungi and their caryophyllaceous hosts: prevalence of host shifts and importance of delimiting parasite species for inferring cospeciation. *BMC Evolutionary Biology*, 8(1), 100
- 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.
- Rezende EL, Lavabre JE, Guimarães PR, Jordano P, Bascompte J. Non-random coextinctions in phylogenetically structured mutualistic networks. *Nature*. 2007 Aug;448(7156):925.