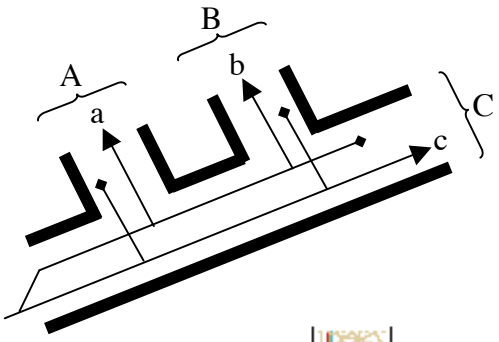
April 9, 2018**. Phylogenetic trees IX: Below the "species level"; phylogeography; dealing with reticulation**

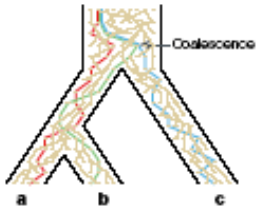
We have discussed a number of assumptions that are used in phylogenetic methods. One fundamental shared assumption shared by *all* methods we have talked about so far is that divergence (splitting of lineages) occurs, not reticulation (joining of lineages). To address the latter possibility, following the general approach in this class, we need to first consider the principles of the topic before we can figure out how to study it: ***what, if anything, is reticulation -- and how do we find it?***

A. The fractal nature of the Tree of Life.

As mentioned in earlier lectures, the tree of life is inherently fractal-like in its complexity. Look closely at one *lineage* of a phylogeny (remember: defined as a diachronic connection between an ancestor and a descendent) and it dissolves into many smaller lineages, and so on, down to a very fine scale. Thus the nature of both the *terminal units* (OTUs -- the "twigs" of the tree in any particular analysis) and the characters (hypotheses of homology, markers that serve as evidence for the past existence of a lineage) change as one goes up and down this "fractal" scale.

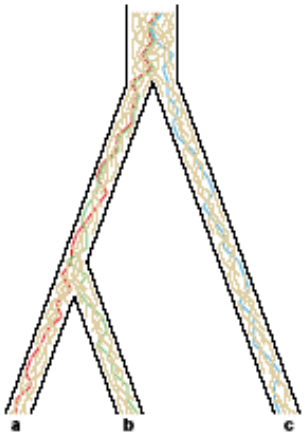
Ontologically speaking, larger-scale lineages are usually composed of smaller lineages nested inside them, and so on. Thus, epistemologically speaking, the choice of which lineage to represent in a particular analysis depends on the questions begin asked. Furthermore, lineages at these different scales potentially have different histories; in other words the smaller lineages are not always proper subsets of the larger containing ones. This is sometimes called the "gene tree / species tree" distinction, but that distinction is far too simplified; there are many nested levels of potentially incongruent lineages, not just two. Besides, there is no *a priori* "species tree" to compare gene trees to; on the contrary the "species tree" has to be inferred, largely from gene trees!





B. Is there a difference between genealogy and phylogenetics? or: Is there an important break at the "species level"?

Rosenberg and Nordborg (2002,"Genealogical trees, coalescent theory and the analysis of genetic polymorphisms," *Nature Reviews Genetics* 3: 380-390) say that there is a difference, and many workers (primarily zoologists) do make a distinction between reconstructing trees at the population level (genealogies) and at the species level and above (phylogenies). The same workers would distinguish phylogeography from phylogenetic biogeography, and separate out study of coalescence ("gene trees") from branching evolution ("species trees").



Workers following this the distinction often make use of a concept of **population**, that is left relatively unexplored (but see discussion by Millstein 2009, "Populations as individuals," *Biological Theory* 4: 267-273). What is a population? A geographically circumscribed set of organisms in the same species? [Can you see the possible problems with a definition like this?] The goal in this approach is to look at patterns of genetic similarities and difference across these "populations," examining parameters such as effective population size, migration rates, divergence times, and population growth or decline (Knowles and Maddison 2002, "Statistical phylogeography," *Molecular Ecology* 11.12: 2623-2635). See figure on last page.

A bit of history: population genetics has always placed special emphasis on identity of genes by descent -- why? It goes back to the concept of *replication* we discussed previously. To study a process of natural selection, you need to know what the replicators are, as well as the interactors. So knowing identity of genes isn't enough, you need to know that they are related, Knowing how things are related is essential to testing functional hypotheses. Population geneticists developed important methods to study gene trees, independently from the methods developed by systematists to study species trees. But the two traditions have themselves exhibited hybrid vigor in the last decade!

If it is acknowledged that branching and reticulation occur at many nested levels in the Tree of Life (even if the balance between the two processes shifts towards more reticulation as you go to finer scales), then we have to take into account the possibility of both processes at all levels, and it isn't useful to consider genealogy as different from phylogeny. It's all a matter of scale, and always a case of comparing trees. The appropriate methods for studying history remain phylogenetic (not distance-based) above and below the "species level." Questions about genetic distance, gene frequencies, etc. are of course important in population genetics, but should not be confused with questions about the history of genes and "populations."

C. So what are the twigs on phylogenetic trees meant to represent?

Thus, even if one wanted to try to avoid problems of hybridization by using only semaphoronts in a data matrix, one would still need to pay attention to the same issues of scale. On the theory side, one still needs to decide conceptually which lineages are being represented by what semaphoronts. Given the fractal nature of the problem as discussed above, there are semaphoronts nested inside of semaphoronts.

On the empirical side, it is nearly impossible in practice to use single semaphoronts as terminals rather than compositely-coded OTUs that have data taken from a number of semaphoronts. For one thing, not all semaphoronts bear all the characters; there may be juvenile specializations or sexual dimorphism present in a lineage. Some specimens will be missing reproductive organs or other key features. Different genes will often be sequenced from different individuals. Furthermore, data are often taken from the literature, (e.g., a previously published ultrastructural analysis) or from a database (e.g., another lab's gene sequence), in cases where no reference can be made to an original semaphoront (e.g., if no voucher specimen was deposited in a museum). Thus, data are virtually always compiled from studies of different individual organisms considered to represent the same terminal lineage. OTUs are nearly always composites in practice; their composition varying depending on the scale of analysis.

This touches on the species debate, which we have dealt with earlier. I just point out here that the fractal scaling of nested lineages includes those well below the traditional species level. Thus, species are not different from taxa at any other level; they are not "privileged" OTUs to be accepted *a priori* -- they need to be investigated and justified like any other.

Twigs on a given phylogenetic tree are hypotheses of non-recombining OTU's, but of course the process of defining such OTUs relies on character evidence and perhaps prior attempts at tree building (reciprocal illumination as discussed in the first couple weeks of class). ***If you take this view, then you could say that many (or most) cases of reticulation at one level are actually just incongruence of non-reticulate trees at a finer scale.***

D. Types of reticulation

Some terminology on types of reticulation:

horizontal transmission -- movement of some little bit of the genome laterally, perhaps via a virus.

introgression -- leakage of some genes from one lineage to another through occasional interbreeding, perhaps asymmetric.

hybrid speciation -- origin of new "species" (i.e., lineage) from two parent lineages that remain extant, usually by allopolyploidy.

reticulation proper (blending) -- merging of two lineages completely into one, perhaps by secondary contact between two reproductively compatible groups.

E. Can we study reticulation with phylogenetics?

Reticulation is the *bête noire* for cladistics, as initially recognized by Hennig. There are a number of different sources of homoplasy (incongruency between certain character distributions and the cladogram based on maximum parsimony), such as adaptive convergence, gene conversion, developmental constraints, mistaken coding, lineage sorting, reticulation, etc. The last named factor is the most problematical because it involves the fundamental model of reality underlying phylogenetic analysis -- the other factors are cases of mistaken hypotheses of homology, whereas "homoplastic" character distributions due to reticulate evolution involve true homologies whose mode of transmission is not tree-like. Hennig and later Nixon and Wheeler (1990, "An amplification of the phylogenetic species concept," *Cladistics* 6: 211-223) were correct in focusing on reticulation and the problems it causes for cladistics. However, the problem posed by reticulation for the species question is more complicated than their "solution" (i.e., their perceived abrupt cessation of interbreeding at the species level), for the following pair of reasons:

(1) just as barriers to reticulation are often not complete, reticulation is not a complete barrier to phylogenetic analysis. We can reconstruct relationships in the face of some amount of reticulation (how much is not yet clear, but is amenable to study). For example, McDade (1992, "Hybrids and phylogenetic systematics II. The impact of hybrids on cladistic analysis," *Evol*. 46: 1329-1346.) has shown that incorporating a few known hybrids in an analysis of "good" species does not seriously affect the cladistic topology of the "good" species. Of course, the hybrids cannot be placed correctly in a reticulate position solely via cladistic analysis, but the relationships of the non-hybrids may be perfectly reconstructable. McDade actually gives rules predicting what a hybrid taxon should do in a cladistic analysis; thus there may be a self-correcting mechanism here as there is with other sources of homoplasy: even major convergence (e.g., in cave animals) can be uncovered via cladistic

analysis. As with convergence, where the application of phylogenetic analysis provides the only rigorous basis we have for identifying homoplasy and thus demonstrating non-parsimonious evolution, the only way we can identify reticulation on the basis of character analysis alone is through the application of phylogenetics, followed by the examination of homoplasy to attempt to discover its source (see discussion by Vrana & Wheeler 1992, "Individual organisms as terminal entities: laying the species problem to rest," *Cladistics* 8: 67-72). As was pioneered by Slatkin and Maddison (1989, "A cladistic measure of gene flow inferred from the phylogenies of alleles," *Genetics* 123: 603-613), cladistic analysis of non-recombining genes can even be used to measure gene flow between populations. ***Phylogenetic analysis can be used to study reticulation -- in fact, it is the main tool for such studies.***

(2) reticulate relationships range from intense (in panmictic, sexually reproducing groups where are individual relationships are exclusively reticulate) to less intense (in spatially or temporally subdivided groups where both reticulate and divergent relationships exist, facultatively and/or obligatorily, among individuals). Furthermore, there is no consistently clear demarcation between reticulate and branching relationships. Hybridization takes place between clades of various patristic/cladistic degrees of relatedness. There is no sharp distinction between sexually vs. asexually reproducing populations in a great many organisms. Bacteria exchange genetic material in a variety of ways. Diatoms, cladocerans and rotifers commonly undergo many asexual generations with occasional sexual generations occurring in response to environmental change; some lineages within these groups can be obligately asexual. In many diatoms, only part of a single clonal lineage can become sexual at any given time. Other forms of reticulation occur throughout nature. Rare, high level hybridizations may occur among very divergent lineages, such as among genera of orchids; viral-mediated lateral transfer of genetic material is suspected at much higher levels. ***Reticulation is not a species-specific problem.***

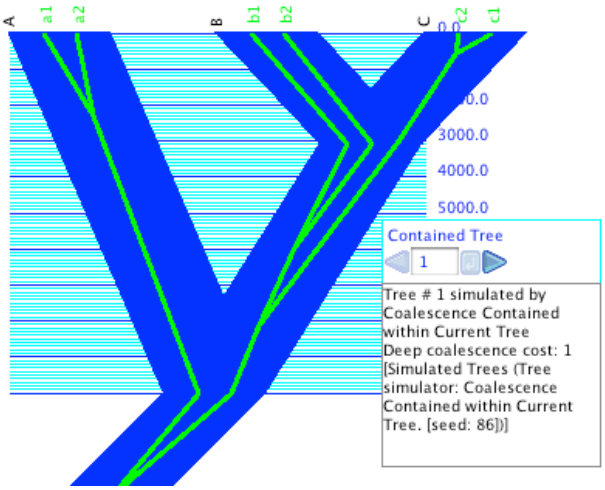
Thus, just as there may be no largest cladistic unit for which reticulation is impossible, there may be no smallest "irreducible" cladistic unit within which no further diverging phylogenetic patterns occur; ontologically speaking, we are dealing with a fractal pattern. When one looks at a lineage closely, one sees a pattern of divergence of lineages within (and some reticulation, perhaps increasingly greater as one looks at less inclusive lineages). Asexuals are the most extreme case; cladistic structure will go down below the organism-level (more on this below). This fractal pattern of reticulation and branching is a severe problem for phylogenetic inference by any means. But as argued above, phenomena such as symbiosis are discovered as incongruency between organismal and character phylogenies. Massive convergence in one character system is discovered by incongruence between that system and other characters. By presuming that synapomorphy is equivalent to strict taxic homology of sister groups, phylogenetic analysis implies that homoplasy is explainable by all other processes including reticulation. Lacking other information, reticulation must always be presumed to be a possible explanation for homoplasy.

Assuming we want to discover reticulation by objective means (Vrana and Wheeler, 1992), it will be important to pay further attention to the problem of reticulation. Were phylogenetic analysis to be attempted on individuals within a panmictic group, consensus cladograms would presumably be nearly completely unresolved. This would be the correct result: there is little or no cladistic structure to reconstruct in such cases. An unproven

assumption in such cases of intense reticulation among OTUs is that there would be a disproportionate number of nearly most parsimonious trees. One might also expect to observed non-random distributions of homoplastic characters (concerted homoplasy) in cases of hybridization. How modes of reticulation actually affect character distributions on cladograms is an productive avenue for empirical and theoretical investigations.

F. Coalescence theory, the study of non-recombining gene genealogies. An exciting area of current research, bridging between systematics and population genetics. Studying properties of gene trees with respect to each other and to simulated trees, can give insights into processes of natural selection, biogeography, population parameters, etc.

For example, gene trees coalescing more recently than random could indicate directional selection, or bottlenecks. Conversely, gene trees coalescing more deeply than random could indicate stabilizing selection.



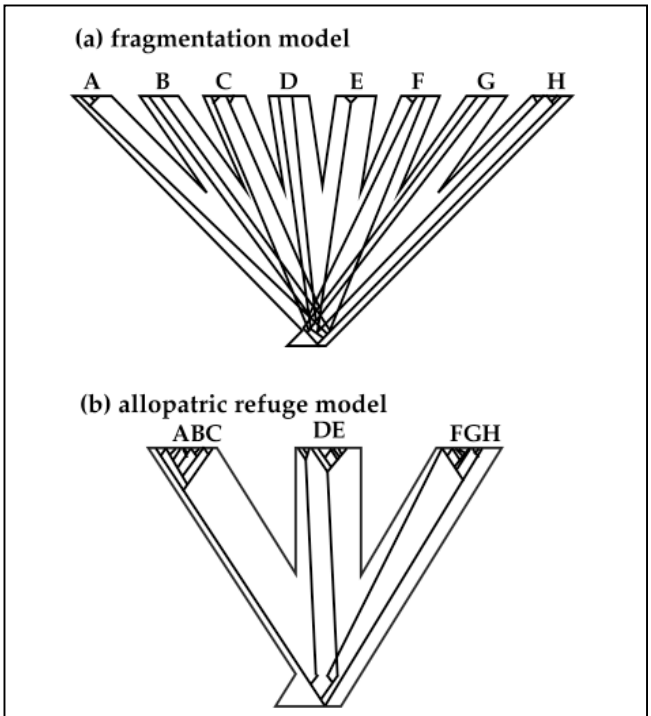
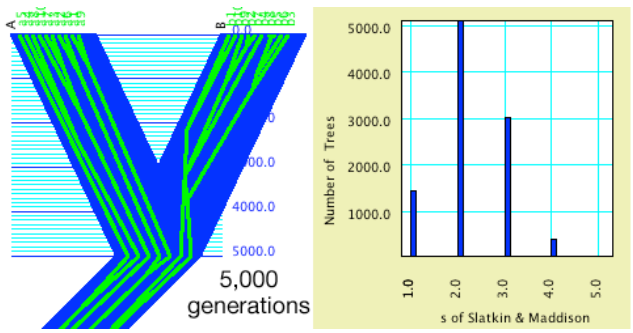
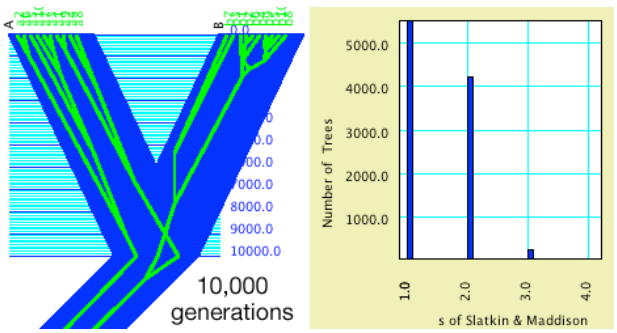
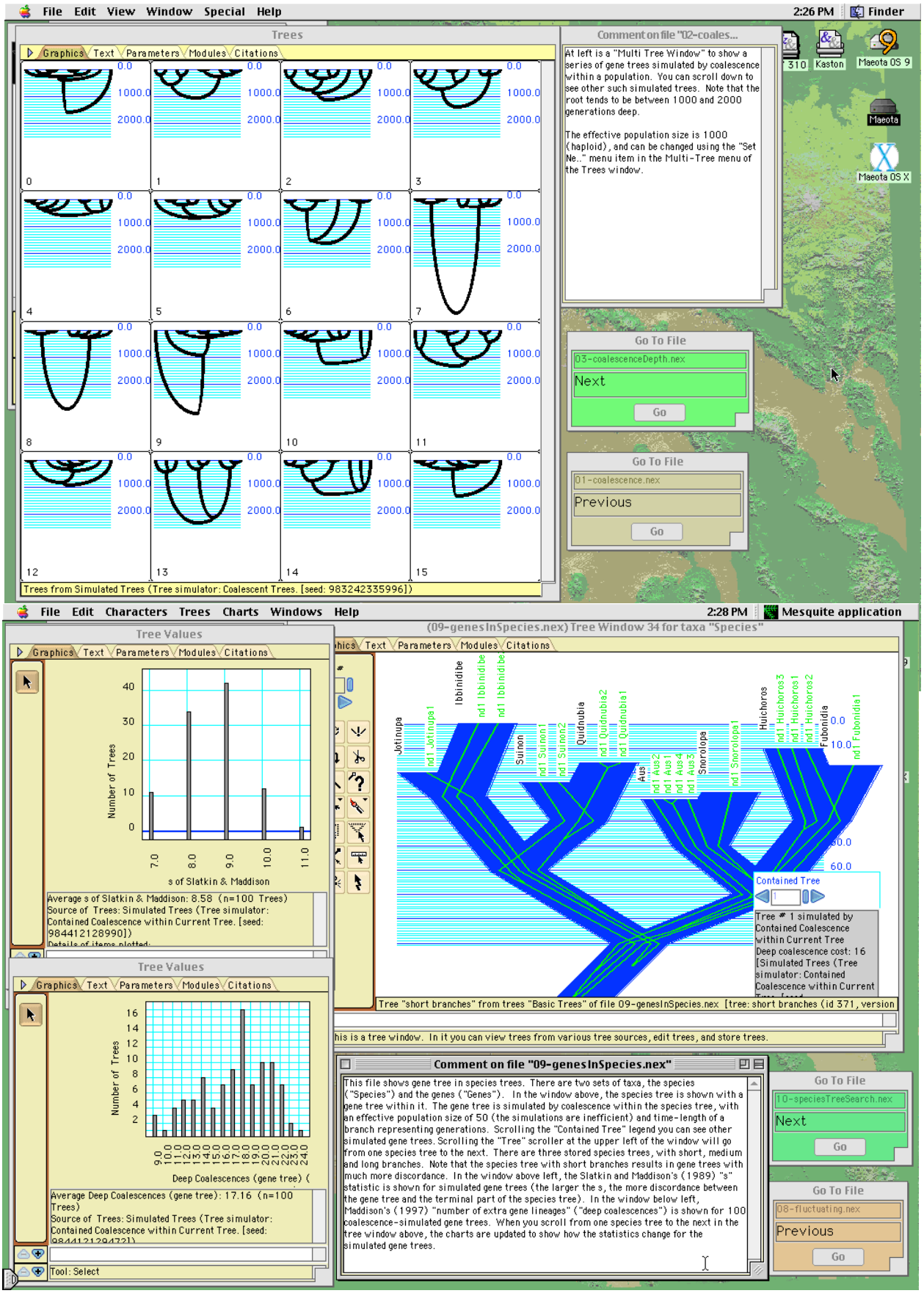


Fig. 3 The two phylogeographic hypotheses Knowles (2001b) tested with coalescent simulations to determine if contemporary sky-island populations of montane grasshoppers were derived from (a) a single ancestral population as represented by a fragmentation model, or (b) multiple ancestral populations, using an allopatric refuge model (groupings of populations were identified from previous phylogeographic analyses (see Figs 3 and 4, Knowles 2001b).

from Knowles and Maddison 2002






File Edit View Window Special Help
2:26 PM
Finder
Trees
Graphics Text Parameters Modules Citations
Trees from Simulated Trees (Tree simulator: Coalescent Trees. [seed: 983242335996])
Comment on file "02-coales...
At left is a "Multi Tree Window" to show a series of gene trees simulated by coalescence within a population. You can scroll down to see other such simulated trees. Note that the root tends to be between 1000 and 2000 generations deep.
The effective population size is 1000 (haploid), and can be changed using the "Set Ne.." menu item in the Multi-Tree menu of the Trees window.
Go To File
03-coalescenceDepth.nex
Next
Go
Go To File
01-coalescence.nex
Previous
Go
310
Kaston
Maeota OS 9
Maeota
Maeota OS X
File Edit Characters Trees Charts Windows Help
2:28 PM
Mesquite application
Tree Values
Graphics Text Parameters Modules Citations
Number of Trees
s of Slatkin & Maddison
Average s of Slatkin & Maddison: 8.58 (n=100 Trees)
Source of Trees: Simulated Trees (Tree simulator: Contained Coalescence within Current Tree. [seed: 984412128990])
Tree Values
Graphics Text Parameters Modules Citations
Number of Trees
Deep Coalescences (gene tree) (
Average Deep Coalescences (gene tree): 17.16 (n=100 Trees)
Source of Trees: Simulated Trees (Tree simulator: Contained Coalescence within Current Tree. [seed: 984412129472])
Tool: Select
(09-genesInSpecies.nex) Tree Window 34 for taxa "Species"
Text Parameters Modules Citations
Jotinupa
Ibbinidibe
Suinon
Quidnubia
Aus
Snorolopa
Huichoros
Fubonidia
Contained Tree
Tree # 1 simulated by Contained Coalescence within Current Tree
Deep coalescence cost: 16
[Simulated Trees (Tree simulator: Contained Coalescence within Current
Tree "short branches" from trees "Basic Trees" of file 09-genesInSpecies.nex [tree: short branches (id 371, version
his is a tree window. In it you can view trees from various tree sources, edit trees, and store trees.
Comment on file "09-genesInSpecies.nex"
This file shows gene tree in species trees. There are two sets of taxa, the species ("Species") and the genes ("Genes"). In the window above, the species tree is shown with a gene tree within it. The gene tree is simulated by coalescence within the species tree, with an effective population size of 50 (the simulations are inefficient) and time-length of a branch representing generations. Scrolling the "Contained Tree" legend you can see other simulated gene trees. Scrolling the "Tree" scroller at the upper left of the window will go from one species tree to the next. There are three stored species trees, with short, medium and long branches. Note that the species tree with short branches results in gene trees with much more discordance. In the window above left, the Slatkin and Maddison's (1989) "s" statistic is shown for simulated gene trees (the larger the s, the more discordance between the gene tree and the terminal part of the species tree). In the window below left, Maddison's (1997) "number of extra gene lineages" ("deep coalescences") is shown for 100 coalescence-simulated gene trees. When you scroll from one species tree to the next in the tree window above, the charts are updated to show how the statistics change for the simulated gene trees.
Go To File
10-speciesTreeSearch.nex
Next
Go
Go To File
08-fluctuating.nex
Previous
Go