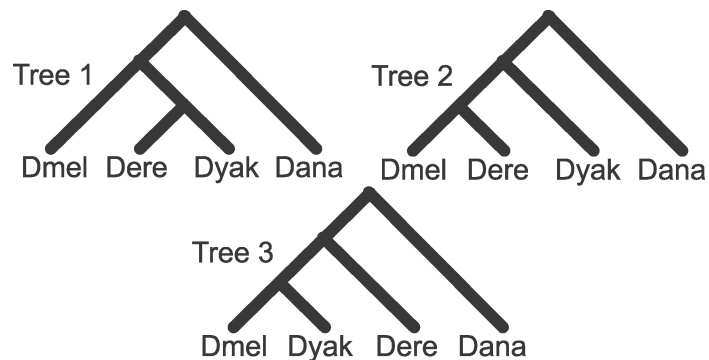


Roadmap

- Gene trees versus species trees
 - *Drosophila* case study
 - Possible solutions
- Bizarre species boundaries
- Co-speciation

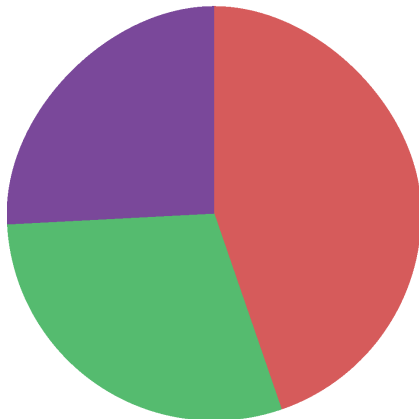
A case study

- Pollard et al. (2006) PLOS Genet
- Whole-genome sequencing of *Drosophila* species
- Previous studies gave all 3 possible trees for *D. melanogaster*, *D. erecta*, and *D. yakuba*
- *D. ananassae* is known to be an outgroup to these
- Data: 9000 genes present in all 3 species

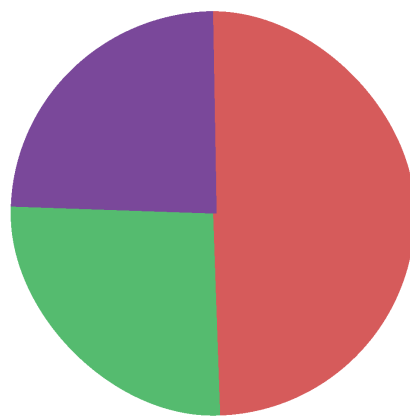


Whole-genome support for the three trees

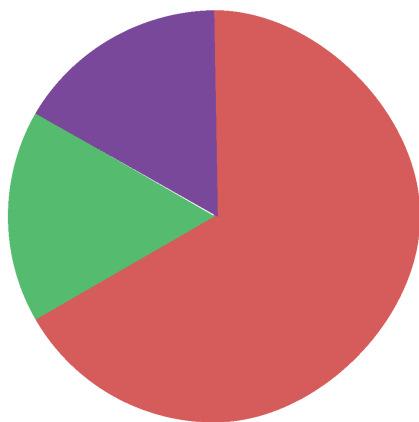
A Nucleotides



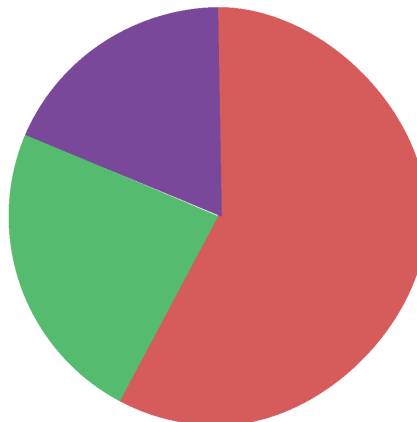
B Amino Acids



C Indels



D Genes



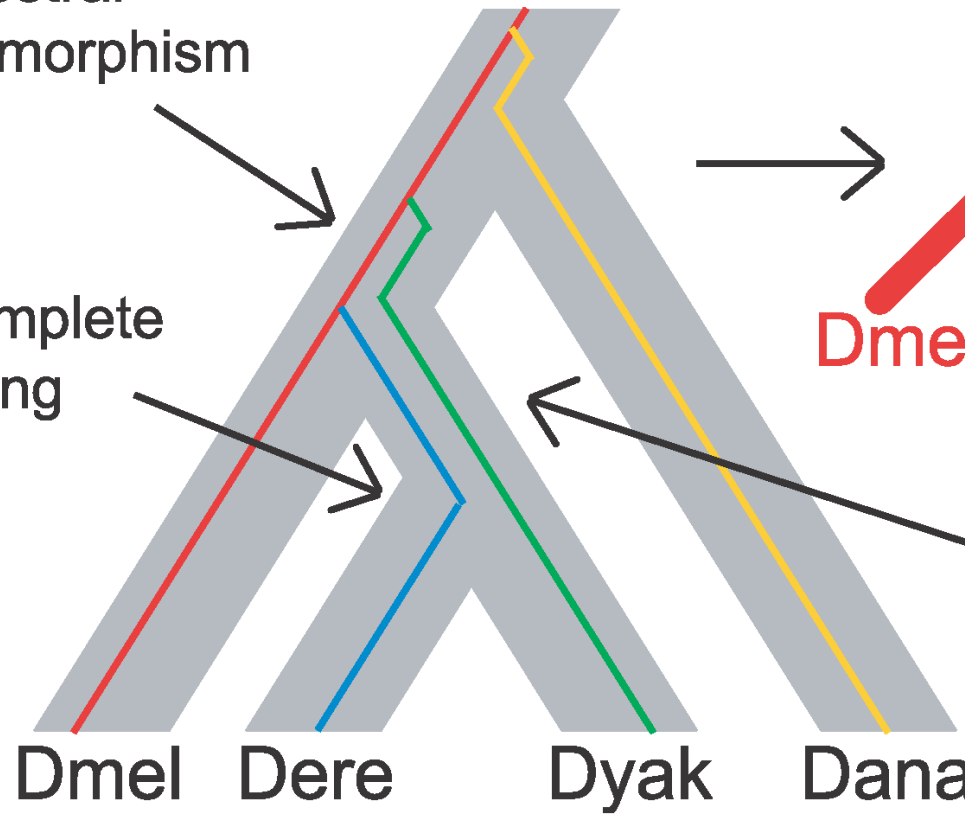
■ Tree 1 ■ Tree 2 ■ Tree 3

Discussion

- Brainstorm:
 - What causes the discordant trees?
 - What could we do about it?

Ancestral
Polymorphism

Incomplete
Sorting



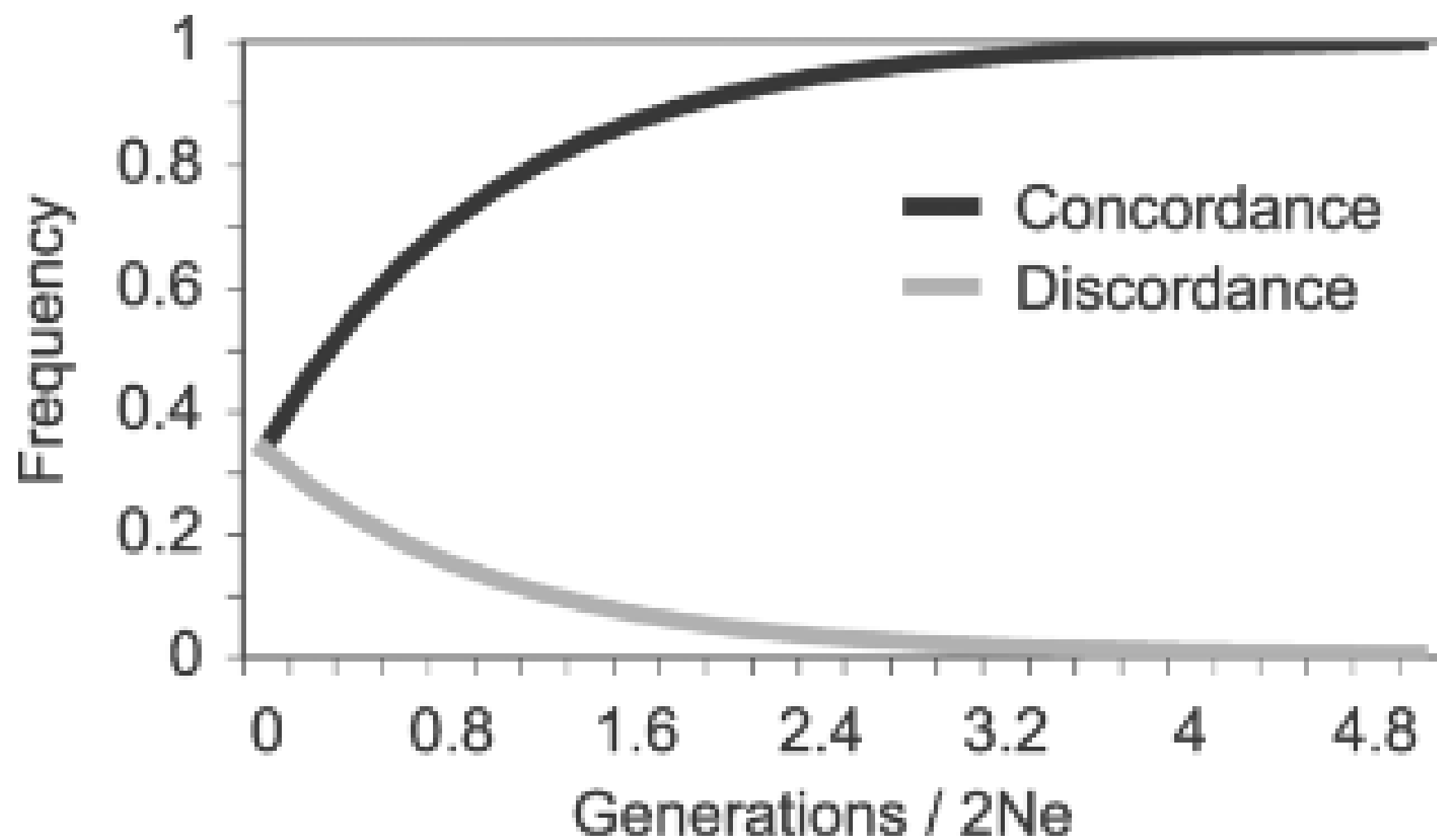
Tree 2 Genealogy



Polymorphisms
Maintained Btwn
Speciation Events

Incomplete lineage sorting AKA ancestral polymorphism

- Can be analyzed using the coalescent:
 - Chance higher if ancestral population is large
 - Chance higher if speciations were close together
- Often said to be a problem with young species
 - If both daughter species survive, they will be like this forever
 - Probably rarer at longer time distances due to extinction
 - May be examples in mammalian radiation



Other tested explanations

- Didn't make much difference:
 - Phylogeny method (parsimony vs. likelihood)
 - Mutational model
 - * More complex models fit the data better but generated *more* disagreement on the tree!
 - Inclusion of other species
- Limit the analysis to genes on which all models and species combos agreed on the tree:
 - Incongruity reduced but not eliminated
- Bootstrap estimates of the reliability of single-gene trees
 - Many genes that supported trees 2 and 3 did so strongly
 - Discordant trees are apparently not just noisy

Spatial structure of data supporting the different trees

- Blocks of sites that agreed on a single tree:
 - Around 8 kb for trees 1 and 2
 - Around 2 kb for tree 3
 - Similar to *Drosophila* LD extent of “a few kb”
- Three significantly long blocks of tree 3 support (250-700 kb)
- Weak negative correlation between block size and *D. melanogaster* recombination rate

Still more ideas?

- Long-branch attraction?
 - High-mutation regions no more discordant than low-mutation ones
 - Tree 1 is not the long-branches preference tree anyway....
- Genomic GC content?
 - *D. erectus* and *D. yakusa* have lower GC than the others
 - This would *increase* support for tree 1
- What about hybridization? Maybe 2 or 3 of these trees are really correct?

Pollard et al. recommendation

- “[M]ethods that can infer the most likely species tree using an entire genome in a single calculation, considering lineage sorting explicitly.”
- StarBEAST (*BEAST) is the closest approach I am currently aware of
 - Heled and Drummond 2010, Ogilvie and Drummond 2016
 - Co-infer individual gene genealogies and species tree
 - At the edge of feasibility with modern computers
- Assumptions of StarBEAST and relatives:
 - Individual genes have trees
 - Incongruence is caused by lineage sorting, so depends on time between splits and N_e
 - No hybridization (species tree is a tree)

A few more comments on genealogy samplers

- Basic idea similar to Monte Carlo integration
- By making small steps, improve acceptance (at cost of perhaps missing whole regions of the distribution)
- Getting them to “mix” (search efficiently) is a black art

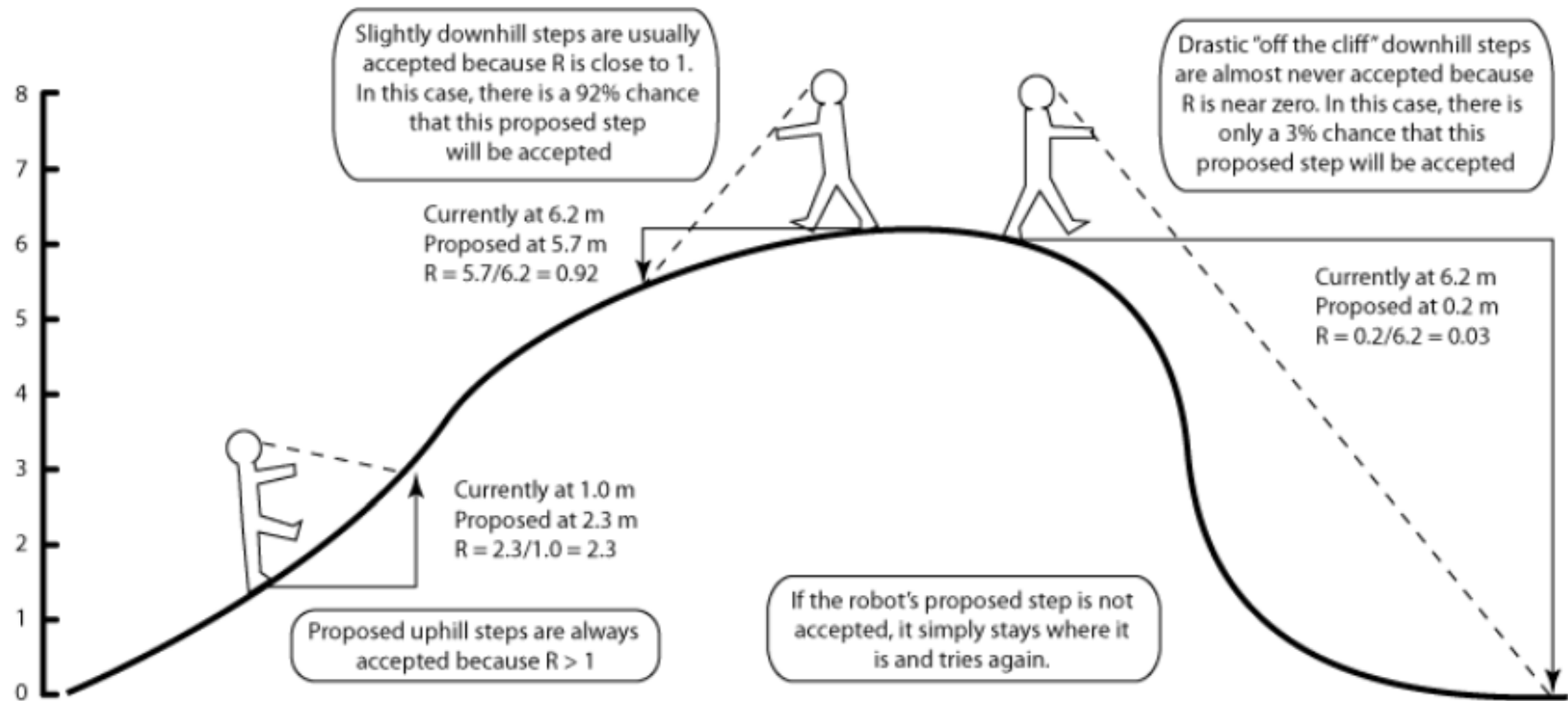


Illustration of MCMC method process (Lewis, 2011)

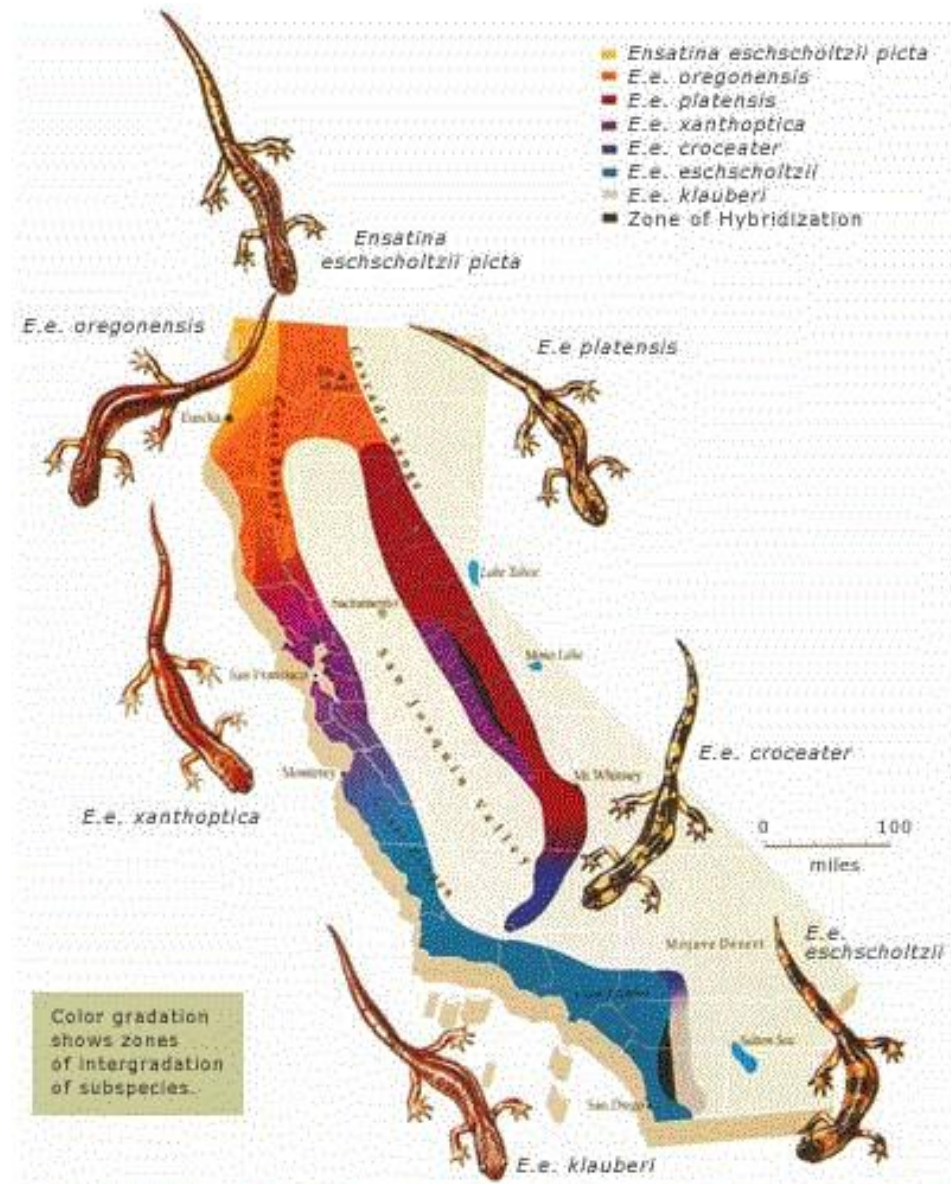
Slide and metaphor due to Paul Lewis

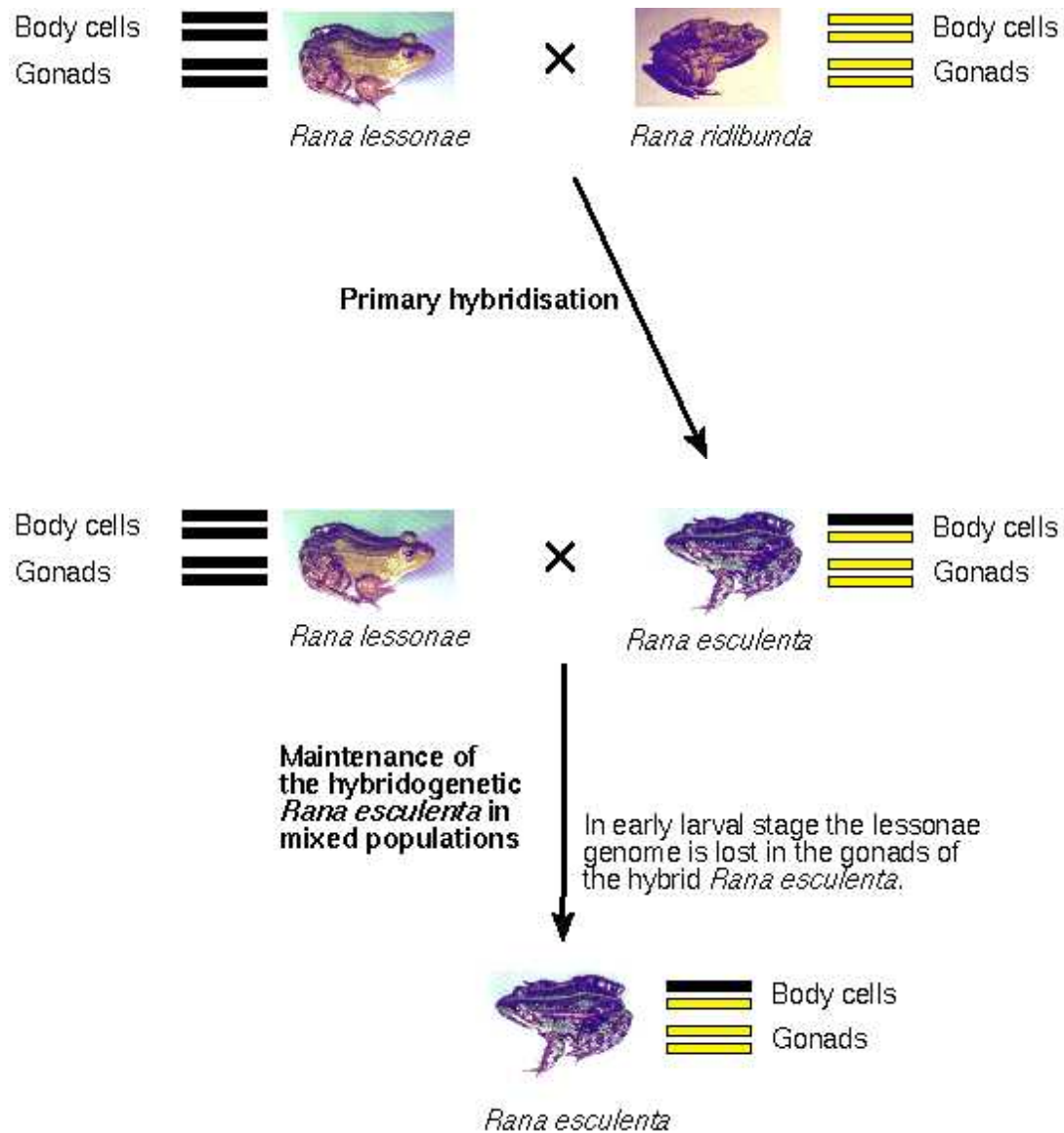
Various species tree issues and ideas

- No bright line between species
- REALLY no bright line between some species
- Coordinated speciation in host and parasite

Ring species

- *Ensatina* salamanders
- Each adjacent pair of populations in this ring can interbreed
- Populations from the far ends of the ring cannot
- Black-winged gulls have a circumpolar ring





(these species are now in genus *Pelophylax*)

Hybridogenic species

- An adaptation in *Pelophylax lessonae* will be present in *P. esculenta* but not transmitted from there
- An adaptation in *P. esculenta* will be inherited only by *P. esculenta*
- *P. esculenta* has distinctive morphology and behavior

A haploid mammalian half-genome

- The *P. ridabunda* genome copy:
 - No longer being replenished from *P. ridabunda*
 - No meaningful recombination (never sees another copy)
- Can you predict the outcome of forcing two *P. esculenta* to reproduce?

Ambystoma platineum

- All-female species; sperm needed only to activate the egg
- Usually triploid (3N) but sometimes 4N, 5N
- In preparation for meiosis they double their chromosomes one time more than usual
- Is this self-fertilization or cloning?



Amazon salamanders

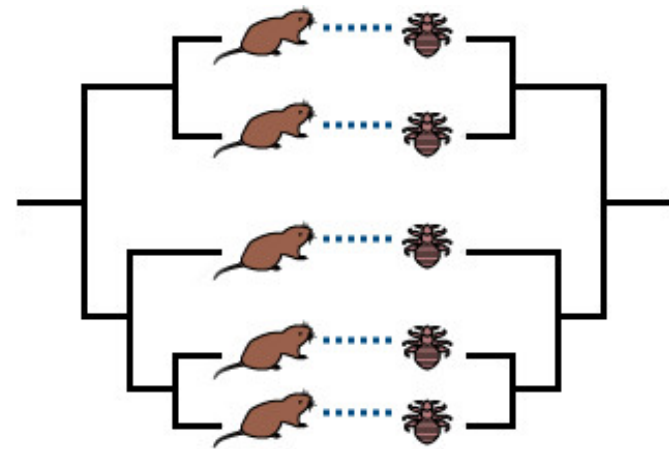
- They may carry chromosomes from the sexual species *A. jeffersonianum*, *A. laterale*, *A. tigrinum* or *A. texanum*
- Many have chromosomes from multiple sexual species
- Often the mtDNA is from a different species than the nuclear chromosomes
- Probably a few sperm sneak through to increase ploidy—they are not 100% asexual

Amazon salamanders

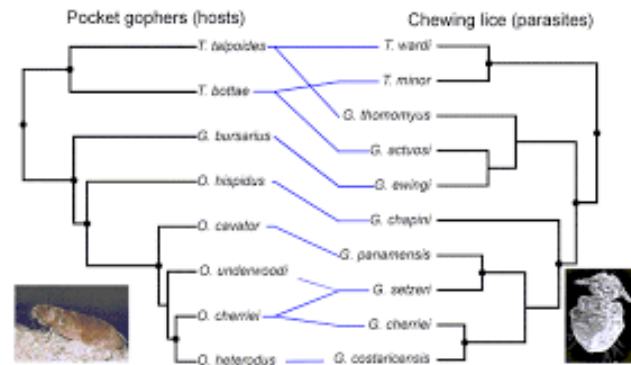
- Genome-wide, *A. platineum* animals are not particularly related to each other
- The only thing they share is (hypothetically) a group of genes which lead to the all-female phenotype
- Does the existence of *A. platineum* cast doubt on the validity of the sexual species?
- Can *A. platineum* itself be considered a species at all?

Cospeciation

- Host species and parasite species often speciate together
- Species trees of the two groups will look very similar
- Example: gophers and gopher lice
- Reproductive isolation of hosts may isolate parasites
- Adaptation of hosts may spur adaptation of parasites (or vice versa)



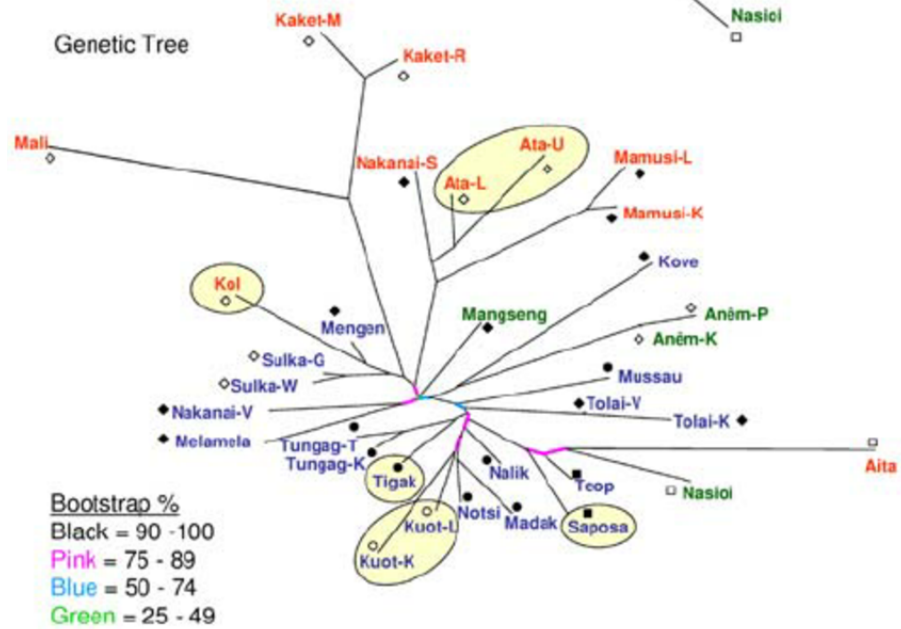
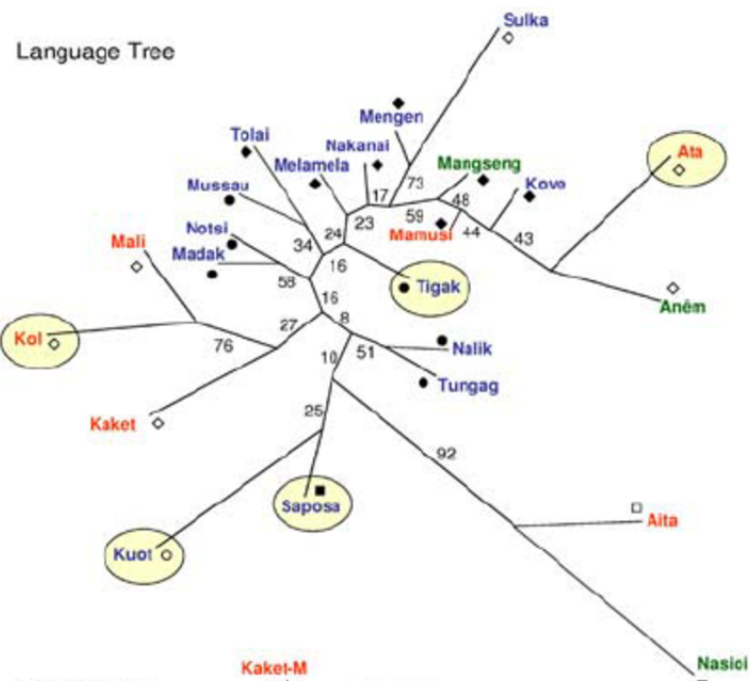
Idealized schematic



Real data

Linguistic trees?

- A relationship tree among languages might mirror relationships among populations
- Problems:
 - Population “tree” not necessarily a tree
 - Language “tree” not necessarily a tree either
 - Establishing homology in words is difficult and subjective: may be biased by preconceptions of the tree
- Next slide from Hunley et al. 2008, “Genetic and linguistic coevolution in Northern Island Melanesia”



Blue=coastal
 Green=intermediate
 Red=inland

Wednesday

- Kin selection:
 - Relationship coefficient (reprise)
 - Altruism
- Group selection
 - “Greenbeard genes”