Today: build upon making/selecting best tree, apply now from species trees to gene trees

Different evolutionary rates caused by:

Stabilizing and diversifying selection

Population size – genetic drift happens faster with smaller populations (neutral genes have bigger impact in this scenario)

Mismatch can be caused by:

HGT (bacterial conjugation/transformation, viruses, plant hybridization/ploidy)

ILS – incomplete lineage sorting. Gene tree does not match with species tree due to multiple gene loss/gain events. Happens more often during short time periods, thinner branches (smaller population). Coalescence happens less frequently when ILS occurs

Coalescent events – points back in time where branches meet

Depth – doesn’t matter

Gene duplication/deletion

Maximum parsimony attempts to minimize these events

muN(1/N) = mu mu = mutation rate, N = pop size

Branch length doesn’t matter for genes

Maximum parsimony with genes:

To try to connect gene and species trees:

Count # of branches/deep coalescences

Bootstrap to find best species tree

Try all possible trees with above

Maximum likelihood with genes:

II loci possible gene trees [P(sequences| gene tree), pi is operator (like sigma but for multiplication), P = probability, | = given (this)

Compare gene and species trees, particularly # of coalescent events

If they don’t match up – anomalous tree. Mashing them together and counting tend to give the wrong answer, however!

Felsenstein zone – area where more data is collected, the more it supports the wrong answer

When creating a mathematical model to solve a problem:

-test range of your variables

-make sure you use variables that apply to the examples you use. Account for events that could occur in the data

-size of data (ex. how many genes)

-test significance of your results (confidence)

-have a way to deal with “messy” data, since this will happen with real data sets

-compare to a previous model that does similar/same goal

Heled and Drummond Paper

Changes formula from Maddison paper by adding integral, prior term

Model for priors does not account for: extinction events, path length

Hyperprior is used – they have to use estimates to determine them

Compared to BEST model at end

Solis-Lemus and Ane Paper

Quartet = four taxa on an unrooted tree. Can be chosen in any order, written (ab|cd). For trees with more than four taxa, can combine quartet results to get a larger result. Some may not be possible given your tree

Identifiability – can there be anything meaningful gained from the data

Julia is an upcoming programming language kind of like R