

Today's Lecture

- PhastCons

Detecting sequence conservation with PhyloHMMs

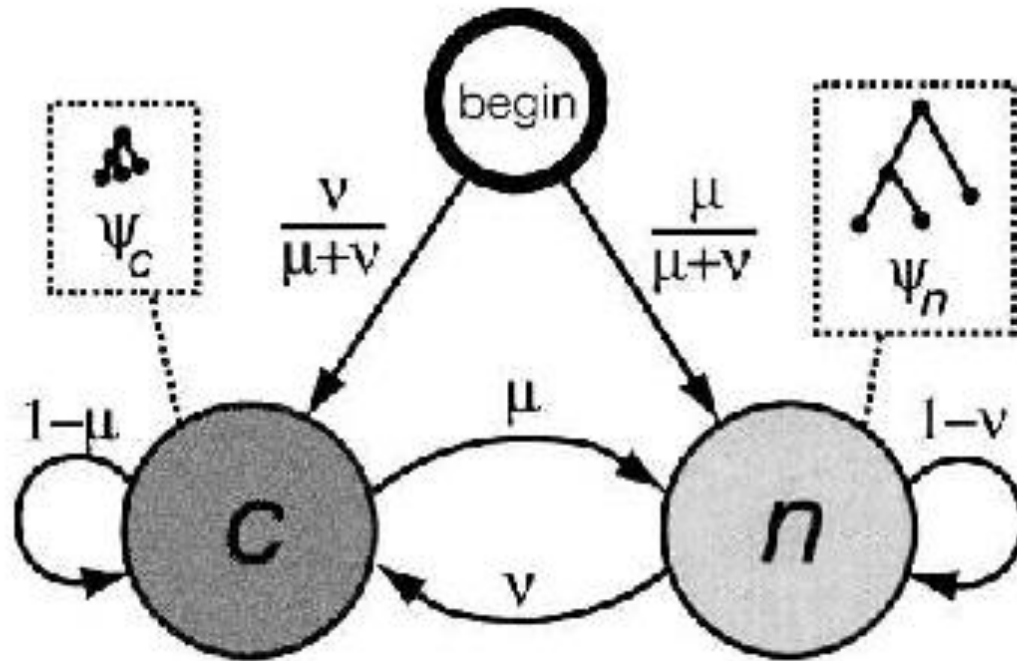
- PhyloHMMs: Yang 1995; Felsenstein & Churchill 1996
- Siepel A. *et al.* (2005): Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* 15:1034-50
 - basis of PhastCons conservation scores (UCSC genome browser)

- Goal: starting from multiple genome sequence alignment, identify
 - conserved regions (regions under purifying selection),against background of
 - neutrally evolving regions

PhastCons PhyloHMM

- model:
 - 2-state HMM
 - c**: conserved state
 - n**: neutral (or nonconserved) state
 - emitted **symbols** are *alignment columns*
 - emission **probabilities** based on *phylogenetic tree* relating sequences
 - discussed in Genome 541, or molecular phylogeny course
 - gaps in alignment treated as *missing data*

PhastCons PhyloHMM



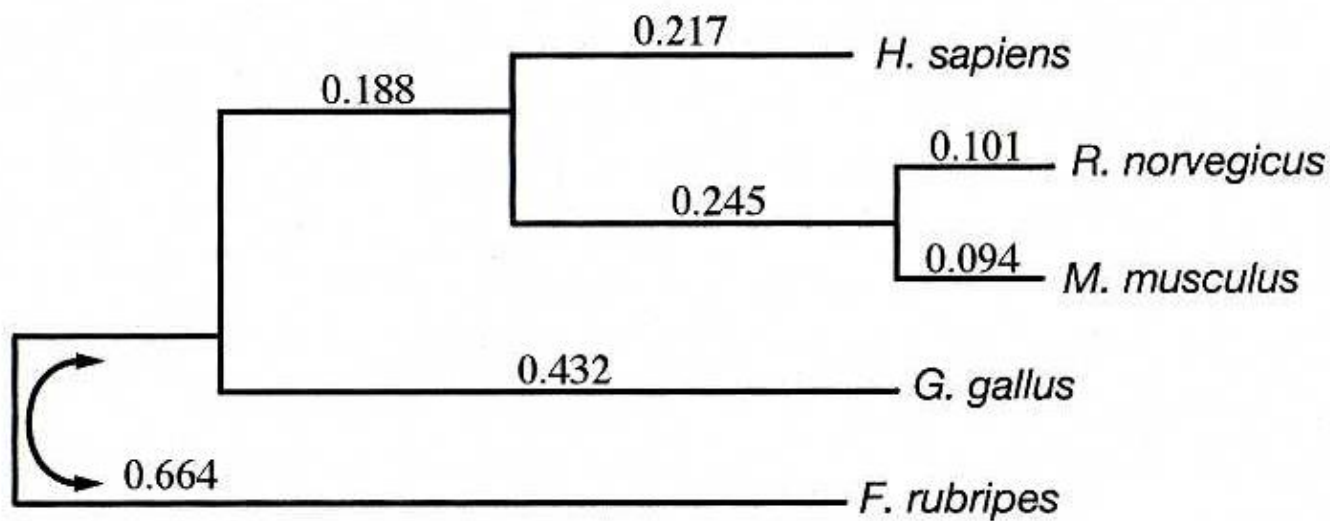
$$\mu = a_{cn}$$

$$v = a_{nc}$$

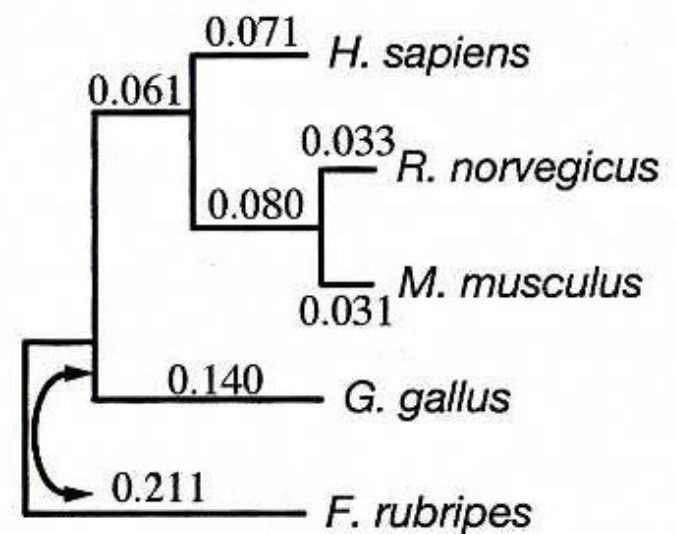
$\mathbf{x} =$

T	C	G	C	G	A	C	A	T	A	T	A	C	G	A	...
T	T	G	G	G	G	C	A	T	G	T	G	G	G	T	...
A	G	C	A	G	A	C	G	T	C	C	G	C	A	A	...

Nonconserved



Conserved



- branch lengths:
 - Expected # substitutions/site over corresponding evolutionary time period
 - for neutral state, should reflect underlying mutation rate
 - for conserved state: mutation rate \times scaling factor ρ
 - $\rho =$ frac of mutations that escape purifying selection
 - $\rho \approx .33$ (for vertebrates)