CSE 527
Lecture 16

Parsimony and
Phylogenetic Footprinting

Phylogenies or Evolutionary Trees

"Nothing In Biology Makes Sense, Except in Light of Evolution "

Dobzhansky

A Complex Question:

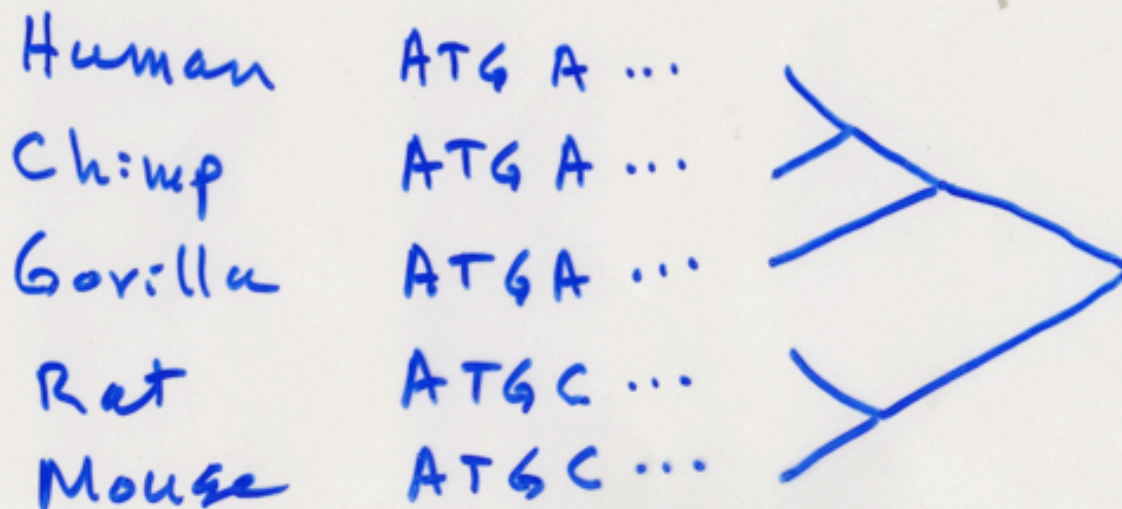Given data (& sequences, anatomy...)
infer phylogeny

A Simpler Question:

Given data and a phylogeny
evaluate "how much change"
is needed to fit data to tree

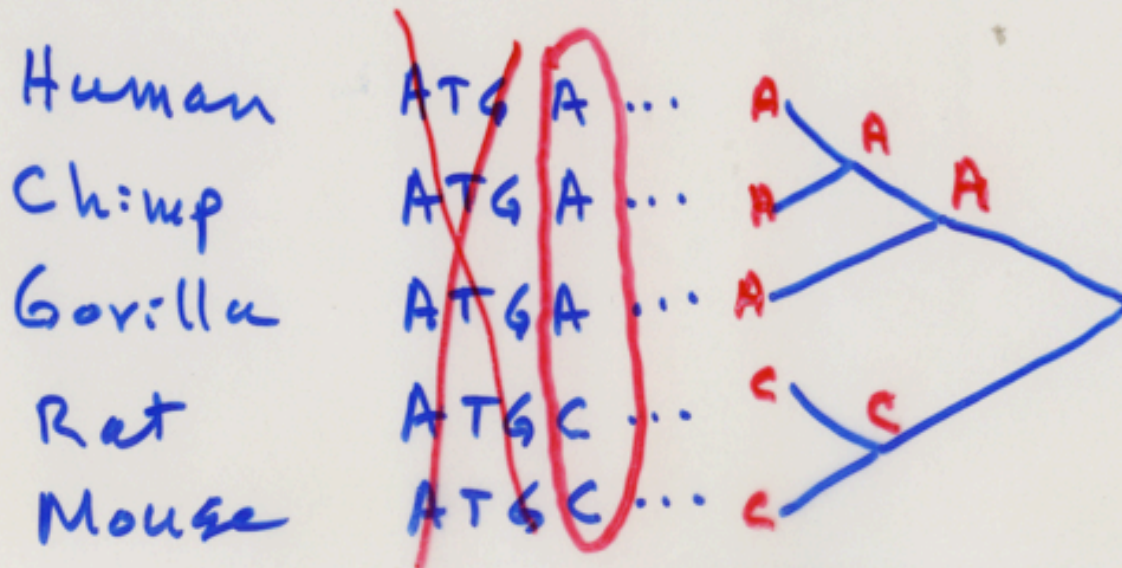# Parsimony

General idea ~ Occam's Razor

Given data where change is rare
prefer an explanation that
requires few events

Human	ATG A ...	
Chimp	ATG A ...	
Gorilla	ATGA ...	
Rat	ATG C ...	
Mouse	ATG C ...	

Parsimony

General idea ~ Occam's Razor

Given data where change is rare
prefer an explanation that
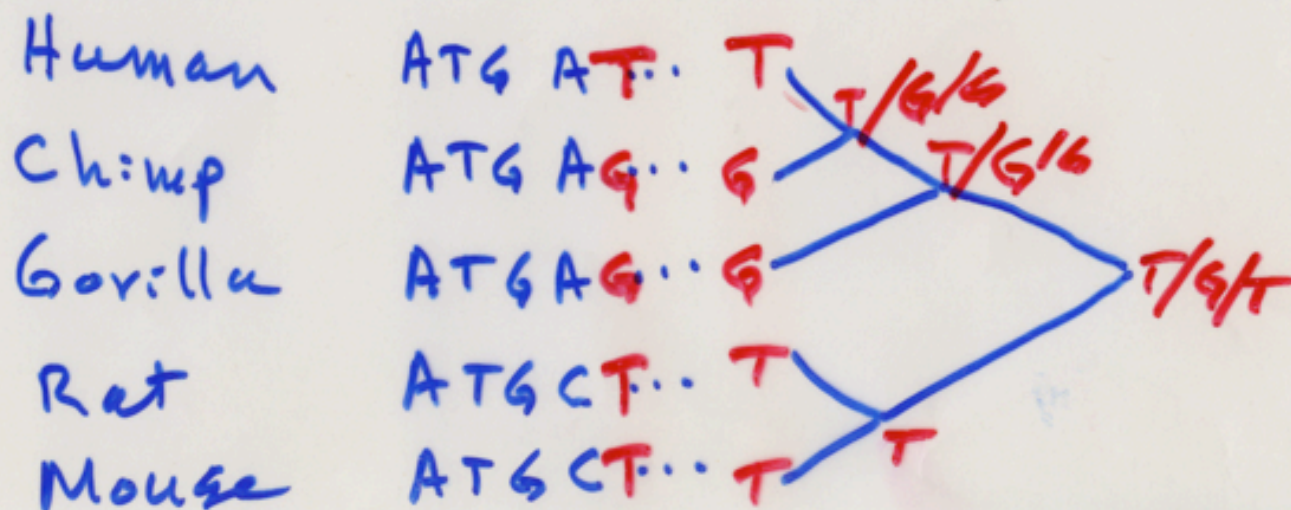requires few events

Parsimony

General idea ~ Occam's Razor

Given data where change is rare
prefer an explanation that
requires few events

Human	ATG AT.. T	
Chimp	ATG AG.. G	T/G/G
Gorilla	ATGAG.. G	T/G/G
Rat	ATG CT.. T	T/G/T
Mouse	ATG CT.. T	

Counting Events Parsimoniously

Bottom line - no unique reconstruction.
Is, of course, unique minimum number.
How to find it?

Early solutions from 1965-75
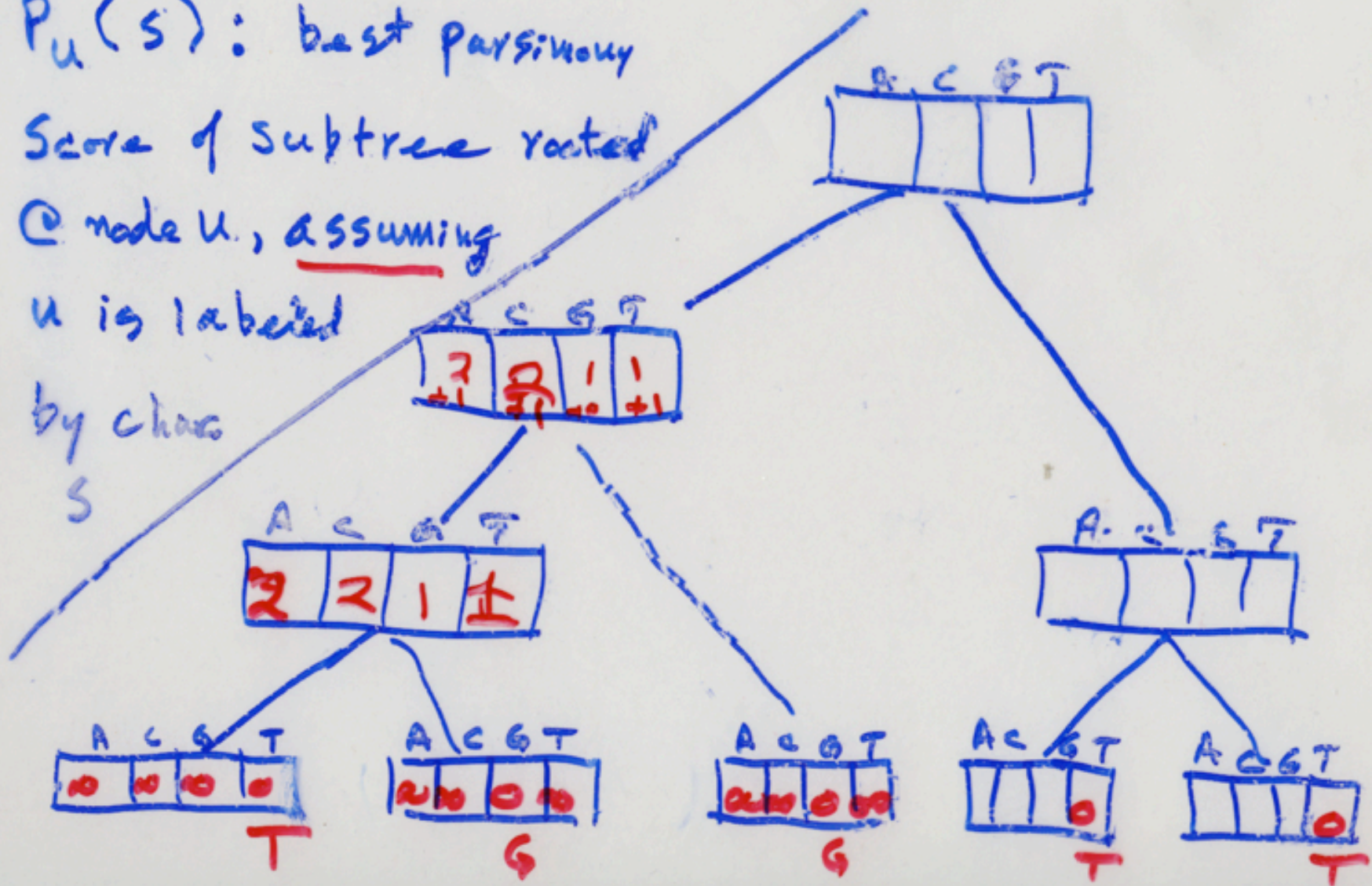
Sankoff & Rousseau '75

$P_u(s)$: best parsimony
Score of subtree rooted
@ node u, assuming
u is labeled
by char
s

Recurrence

$$P_u(s) = \sum_{\substack{v \text{ child} \\ \text{of } u}} \min_{t \in \{A,C,G,T\}} \text{Cost}(s,t) + P_v(t)$$

$$P_{\text{leaf } u}(s) = \begin{cases} 0 & \text{if leaf } u \text{ labeled } s \\ \infty & \text{otherwise} \end{cases}$$

Time: linear in alphabet × tree size

Parsimony

- Parsimony is not necessarily the best way to evaluate a phylogeny, but is a natural approach, & fast.

- Finding the best tree is a much harder problem

- Much is known about these problems; soon-to-be-published book by Joe Felsenstein will be great resource.

Phylogenetic Footprinting

See link to Tompa's slides on course web page