CSE 527
Lecture 10

Parsimony and
Phylogenetic Footprinting

Phylogenies
(aka Evolutionary Trees)

"Nothing in biology makes sense, except in the light of evolution"

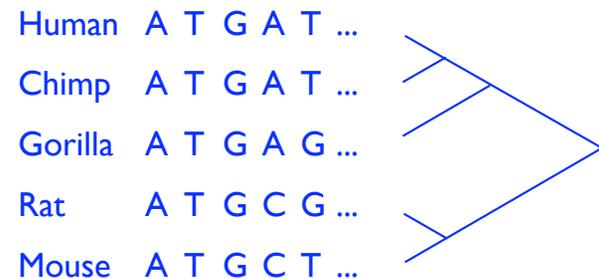-- Dobzhansky

---

- A Complex Question:

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

- A Simpler Question:

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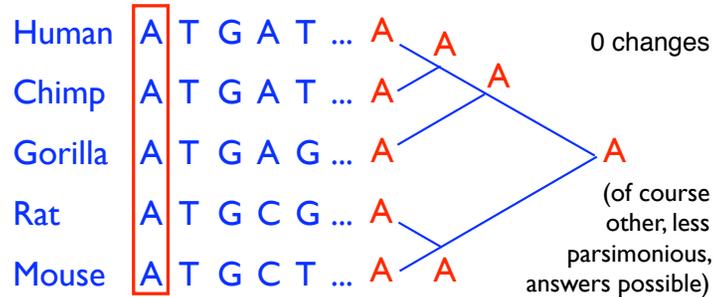

Parsimony

General idea ~ Occam's Razor:
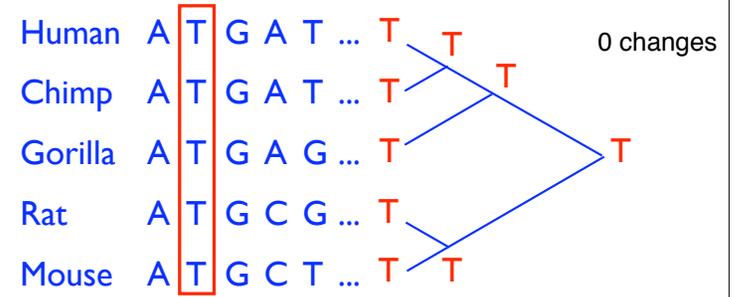Given data where change is rare, prefer an explanation that requires few events

Human	A T G A T ...
Chimp	A T G A T ...
Gorilla	A T G A G ...
Rat	A T G C G ...
Mouse	A T G C T ...

Parsimony

General idea ~ Occam's Razor:
Given data where change is rare, prefer an explanation that requires few events

Human   A T G A T ... A
Chimp   A T G A T ... A
Gorilla   A T G A G ... A
Rat   A T G C G ... A
Mouse   A T G C T ... A

0 changes

(of course other, less parsimonious, answers possible)

Parsimony

General idea ~ Occam's Razor:
Given data where change is rare, prefer an explanation that requires few events

Human   A T G A T ... T
Chimp   A T G A T ... T
Gorilla   A T G A G ... T
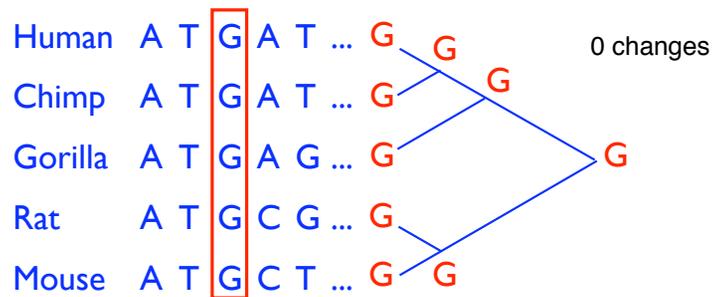Rat   A T G C G ... T
Mouse   A T G C T ... T

0 changes

Parsimony

General idea ~ Occam's Razor:
Given data where change is rare, prefer an explanation that requires few events

Human   A T G A T ... G
Chimp   A T G A T ... G
Gorilla   A T G A G ... G
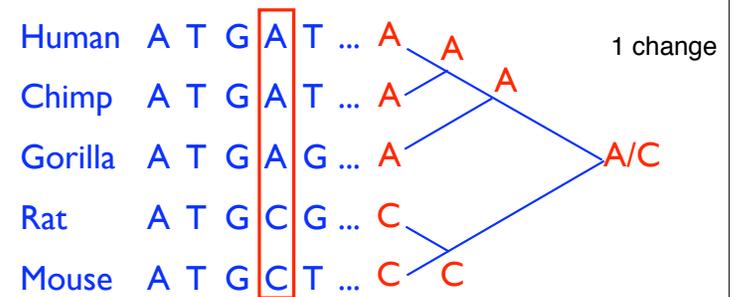Rat   A T G C G ... G
Mouse   A T G C T ... G

0 changes

Parsimony

General idea ~ Occam's Razor:
Given data where change is rare, prefer an explanation that requires few events

Human   A T G A T ... A
Chimp   A T G A T ... A
Gorilla   A T G A G ... A
Rat   A T G C G ... C
Mouse   A T G C T ... C

1 change

A/C
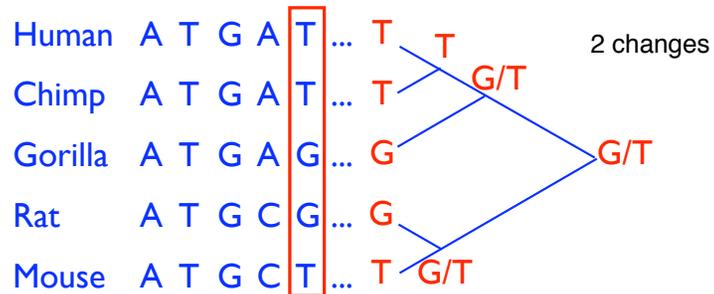
Parsimony

General idea ~ Occam's Razor:
Given data where change is rare, prefer an explanation that requires few events

Human   A T G A T ... T
Chimp   A T G A T ... T
Gorilla A T G A G ... G
Rat   A T G C G ... G
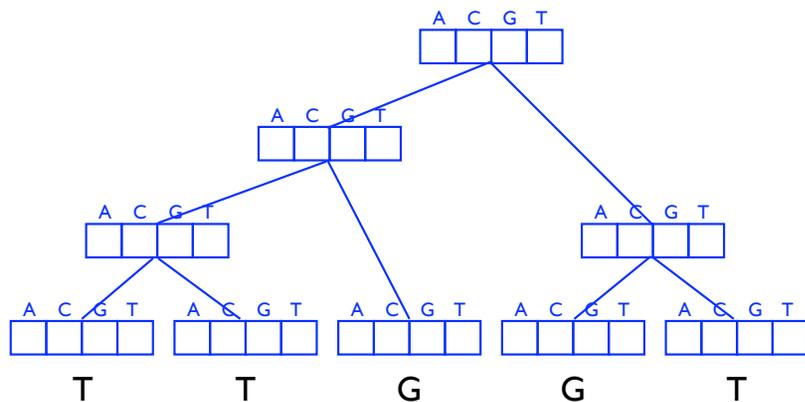Mouse   A T G C T ... T

2 changes

Counting Events Parsimoniously

- Lesson of example – no unique reconstruction

- But there is a unique minimum number, of course

- How to find it?

- Early solutions 1965-75

Sankoff & Rousseau, '75

$P_u(s)$ = best parsimony score of subtree rooted at node u, assuming u is labeled by character s

ACGT

T   T   G   G   T

Sankoff-Rousseau Recurrence

$P_u(s)$ = best parsimony score of subtree rooted at node u, assuming u is labeled by character s

For Leaf u:

$$P_u(s) = \begin{cases} 0 & \text{if } u \text{ is a leaf labeled } s \\ \infty & \text{if } u \text{ is a leaf not labeled } s \end{cases}$$
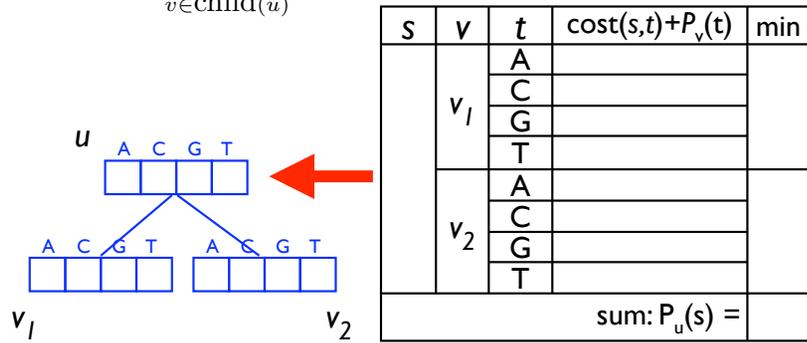
For Internal node u:

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

Time: O(alphabet^2 x tree size)

Sankoff & Rousseau, '75

$P_u(s)$ = best parsimony score of subtree rooted at node u, assuming u is labeled by character s

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

s	v	t	cost(s,t)+P$_v$(t)	min
		A		
	v_1	C		
		G		
		T		
		A		
	v_2	C		
		G		
		T		
			sum: P$_u$(s) =	

u A C G T

A C G T   A C G T

v_1 v_2

Sankoff & Rousseau, '75

$P_u(s)$ = best parsimony score of subtree rooted at node u, assuming u is labeled by character s

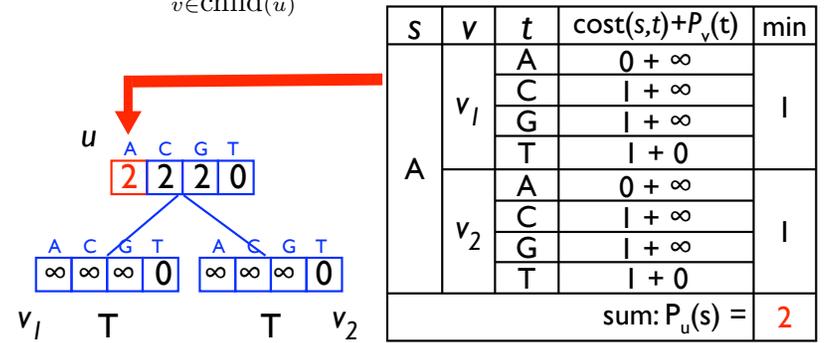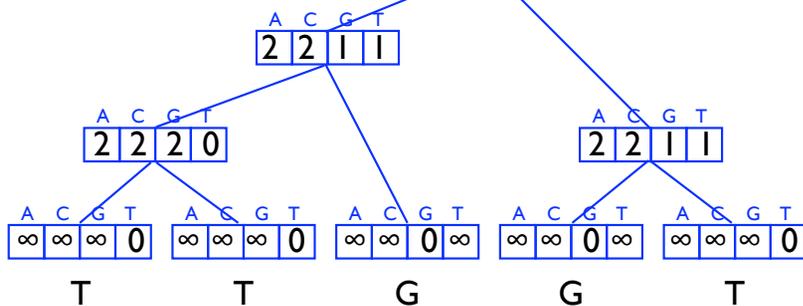$$P_u(s) = \sum_{v \in \text{child}(u)} \min_{t \in \{A,C,G,T\}} \text{cost}(s,t) + P_v(t)$$

s	v	t	cost(s,t)+P$_v$(t)	min
		A	0 + ∞	
	v_1	C	1 + ∞	
		G	1 + ∞	1
		T	1 + 0	
A		A	0 + ∞	
	v_2	C	1 + ∞	
		G	1 + ∞	1
		T	1 + 0	
			sum: P$_u$(s) =	2

u A C G T
   2 2 2 0

A C G T   A C G T
∞ ∞ ∞ 0   ∞ ∞ ∞ 0

v_1 T   T v_2

Sankoff & Rousseau, '75

$P_u(s)$ = best parsimony score of subtree rooted at node u, assuming u is labeled by character s

Min = 2 (G or T) ⟶ A C G T
                     4 4 2 2

A C G T
2 2 1 1

A C G T   A C G T
2 2 2 0   2 2 1 1

A C G T   A C G T   A C G T   A C G T   A C G T
∞ ∞ ∞ 0   ∞ ∞ ∞ 0   ∞ ∞ 0 ∞   ∞ ∞ 0 ∞   ∞ ∞ ∞ 0

   T   T   G   G   T

Parsimony – Generalities

- Parsimony is not necessarily the best way to evaluate a phylogeny (maximum likelihood generally preferred)

- But it is a natural approach, & fast.

- Finding the best tree: a much harder problem

- Much is known about these problems; *Inferring Phylogenies* by Joe Felsenstein is a great resource.

Phylogenetic Footprinting

See link to Tompa's slides on course web page
http://www.cs.washington.edu/homes/tompa/papers/ortho.ppt