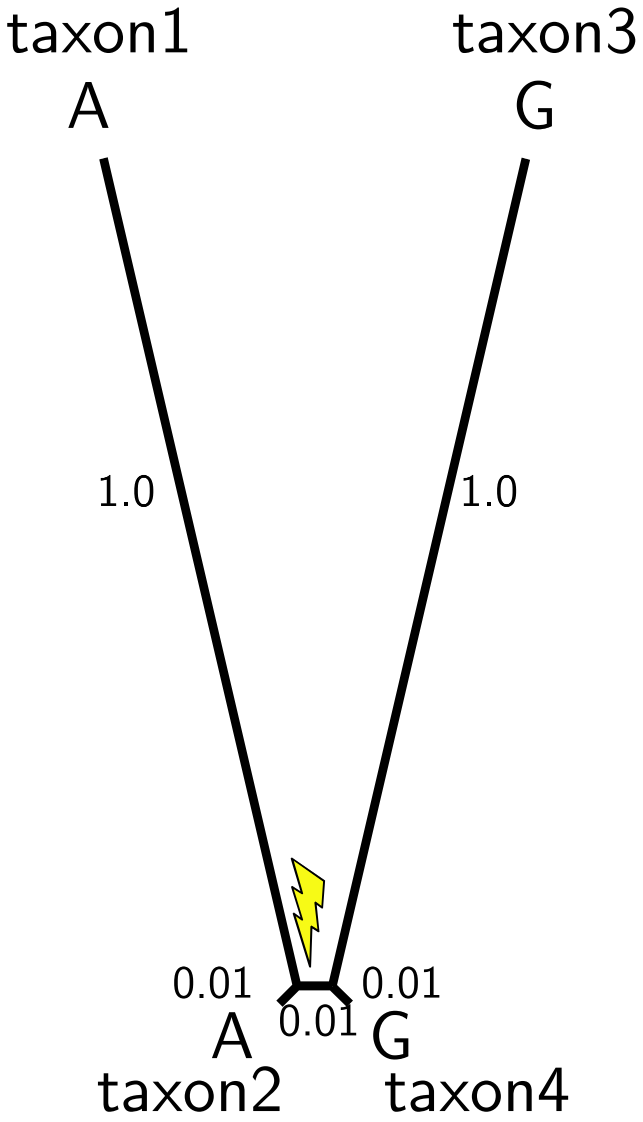
Thanks to Paul Lewis and Joe Felsenstein for the use of slides

Review

- Hennigian logic reconstructs the tree if we know **polarity** of characters and there is **no homoplasy**
- UPGMA infers a tree from a distance matrix:
  - groups based on **similarity**
  - fails to give the correct tree if rates of character evolution vary much
- Modern distance-based approaches:
  - find trees and branch lengths: patristic distances ≈ distances from character data.
  - do **not** use all of the information in the data.
- Parsimony:
  - prefer the tree that **requires** the fewest character state changes. Minimize the number of times you invoke homoplasy to explain the data.
  - can work well if if homoplasy is not rare
  - fails if homoplasy very common **or is concentrated on certain parts of the tree**

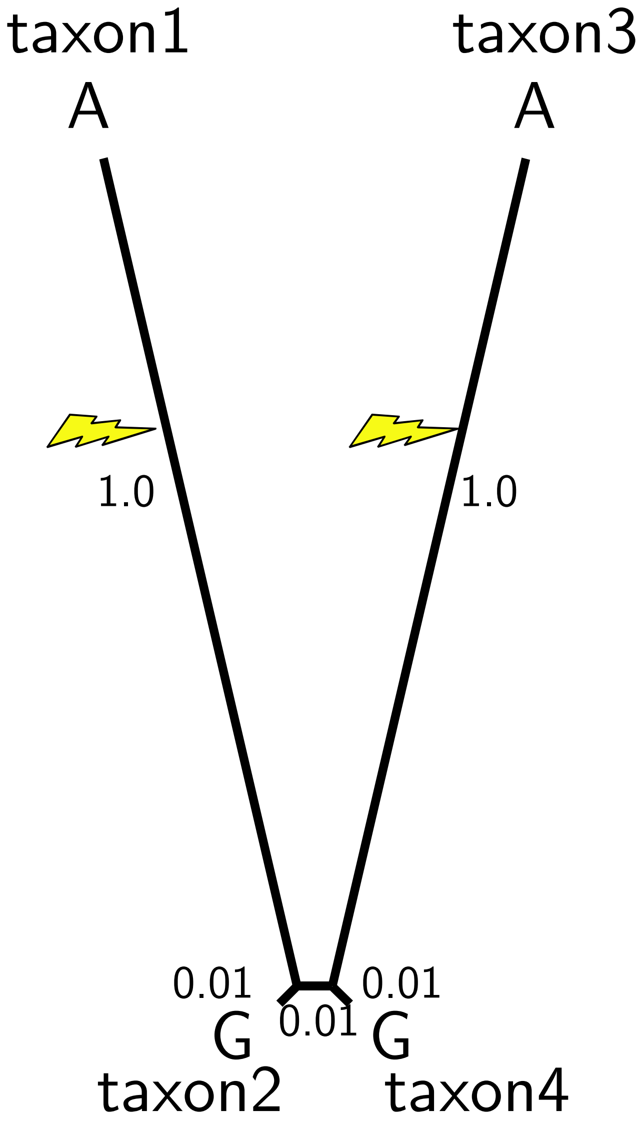
Long branch attraction



Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology* **27**: 401-410.

The probability of a parsimony informative site due to inheritance is very low, (roughly 0.0003).

Long branch attraction



Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology* **27**: 401-410.

The probability of a parsimony informative site due to inheritance is very low, (roughly 0.0003).

The probability of a misleading parsimony informative site due to parallelism is much higher (roughly 0.008).

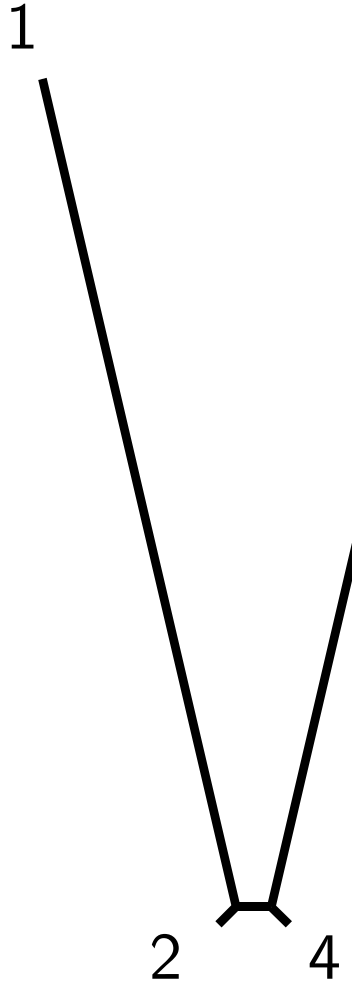
Long branch attraction data

Under such a tree misleading characters are more common that characters that favor the true tree.

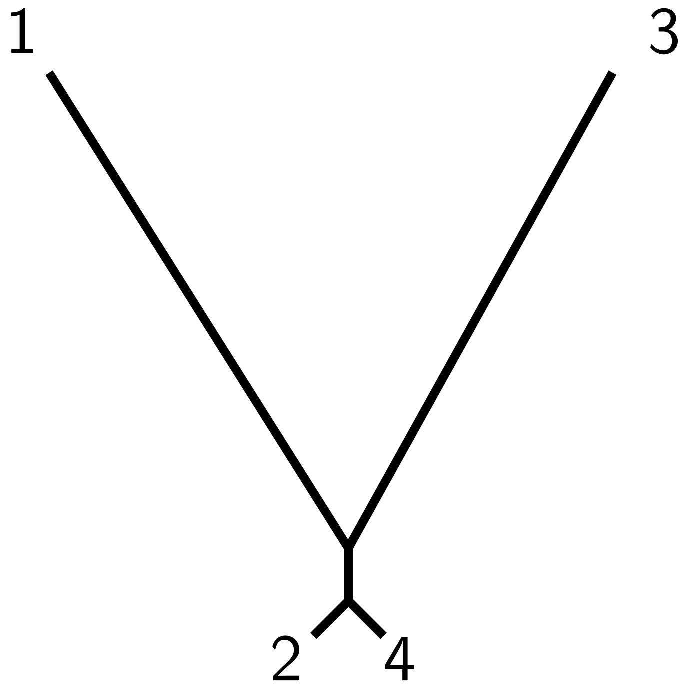
	Rare					Common			
taxon1	A	A	C	C		A	A	C	C
taxon2	A	A	C	C		G	C	T	G
taxon3	G	C	T	G		A	A	C	C
taxon4	G	C	T	G		G	C	T	G

Long branch attraction

Parsimony is almost guaranteed to get this tree wrong.



True



Inferred

Likelihood

X is the data.

T is the tree.

ν is a vector of branch lengths.

$\Pr(X|T, \nu)$ is the *likelihood*; this is sometimes denoted $L(T, \nu)$.

Maximum likelihood: find the T and ν that gives the highest likelihood.