Long Branch Attraction

Results of one simulation of 1000 sites

True tree

0.01
0.01
0.01

987 -> 0 542 -> 1+

0.01 0.01

2 true pars. inf.

True tree

Accurate topology

LBA topology

"Felsenstein Zone" Trees

Fig. 3, p. 531, from Swofford et al. 2001. Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. Systematic Biology 50: 525-539.

Long Branch Repulsion?

Fig. 2, p. 529, from Swofford et al. 2001. Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. Systematic Biology 50: 525-539.
Failing to estimate edge lengths correctly leads to LBA.

Correctly estimated long edges mean convergence explanation is more reasonable.

Underestimated edges make convergence explanation is less reasonable.

A

C

C

JC+G

True
(JC+G)

A

A

C

C

JC
Likelihood ratio test of the molecular clock


\[
\begin{align*}
\text{Codium} & \rightarrow -5226.835 \\
\text{Osmunda} & \rightarrow \text{Zamia} \rightarrow \text{Gnetum} \rightarrow \text{Zea} \rightarrow \text{Oryza} \rightarrow \text{Acer} \\
\end{align*}
\]

KH test

- Original dataset: \(\delta = 3.90\)
- Bootstrap 1: \(\delta = 3.23\)
- Bootstrap 2: \(\delta = 4.35\)
- Bootstrap 3: \(\delta = -0.42\)
- Bootstrap 100: \(\delta = 1.59\)

KH test (RELL method)

\[
\begin{align*}
\delta_0 &= 0.168 \\
\delta_1 &= -0.063 \\
\end{align*}
\]

HIV-1 subtypes

2,000 nucleotide sites from gag and pol genes.
Substitution model: GTR+\(\Gamma\) (using empirical base frequencies)

\[
\begin{align*}
\ln L_1 &= -5073.75 \\
\ln L_{ML} &= -5069.85 \\
\end{align*}
\]

Example: HIV-1 subtypes

\[ \delta_0 = \ln L_1 - \ln L_2 = (-5069.85) - (-5073.75) = 3.9 \]

3.9 does not lie in the rejection region, so we cannot reject the null hypothesis of equal support

\[ \ln L_2 = -5073.75 \]

\[ \ln L_1 = -5069.85 \]

KH Test (Normal Approximation)

```
begin paup;
exe hiv1.nex;
gettrees file=treefile.tre;
lscore 1 2 / khtest=bootstrap[none|normal] RELL=yes;
end;
```

```
lscore 1 2 / shtest RELL=yes bootreps=1000;
```

```
lscore 1 2 / autest RELL=yes bootreps=1000;
```
