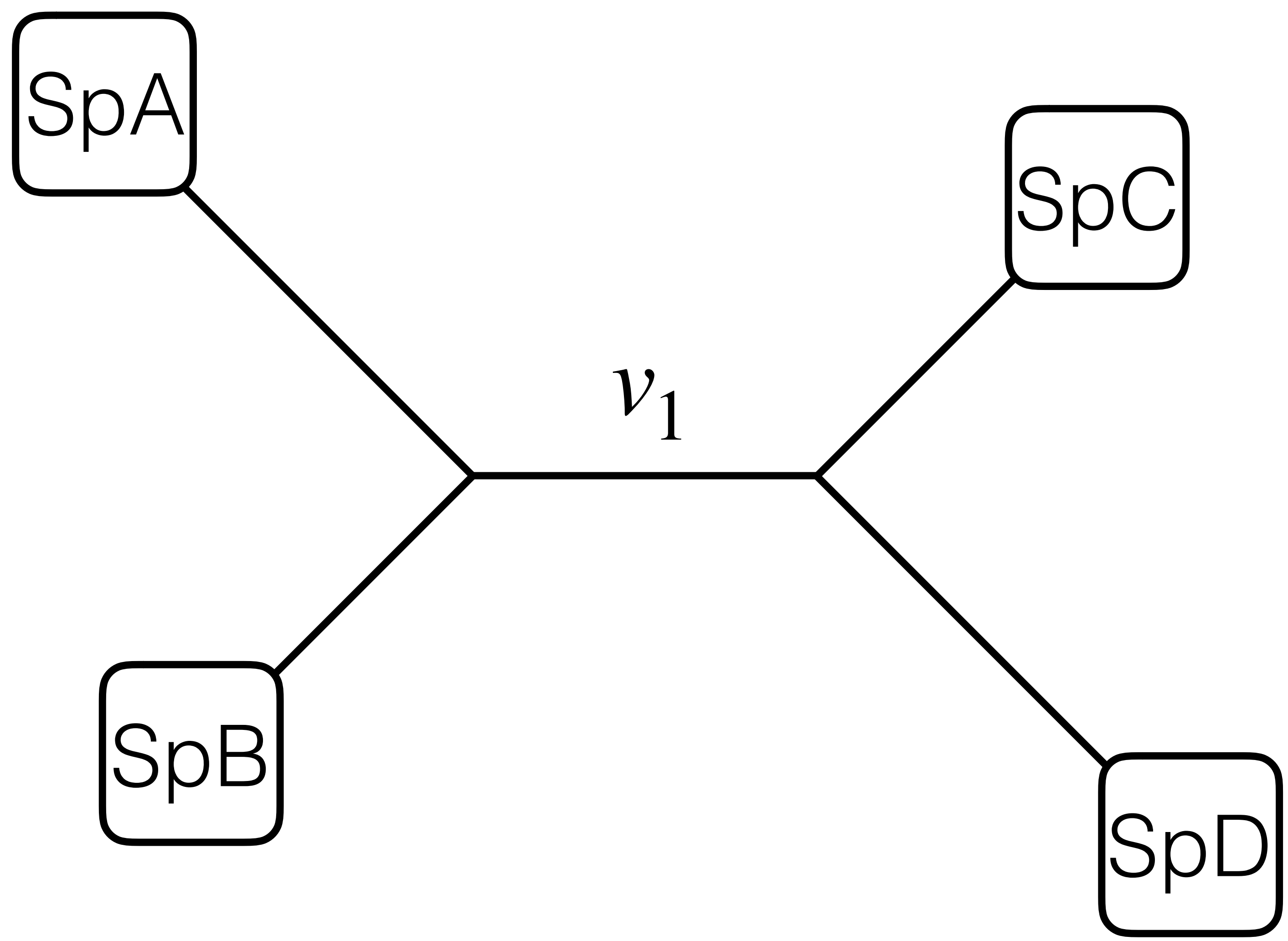
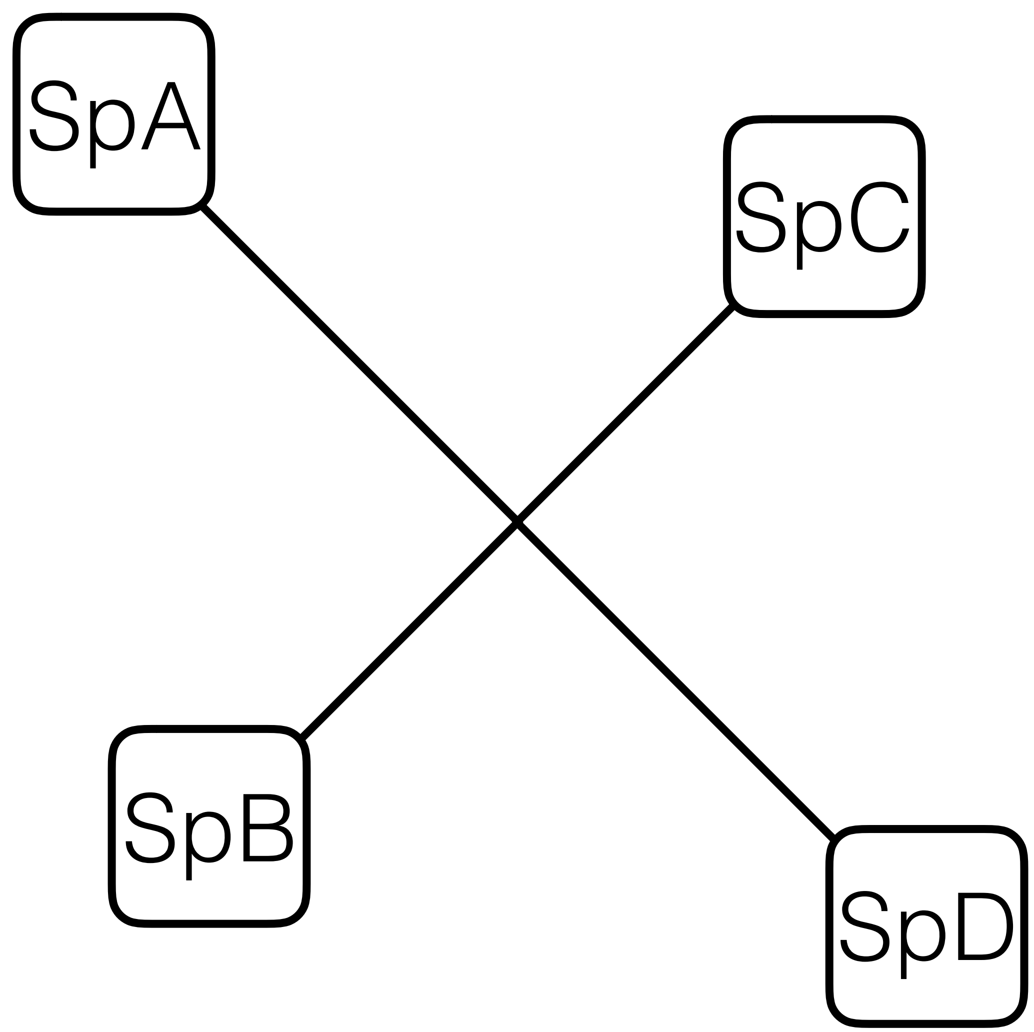
Testing Topological Hypotheses

Interior Branch Test



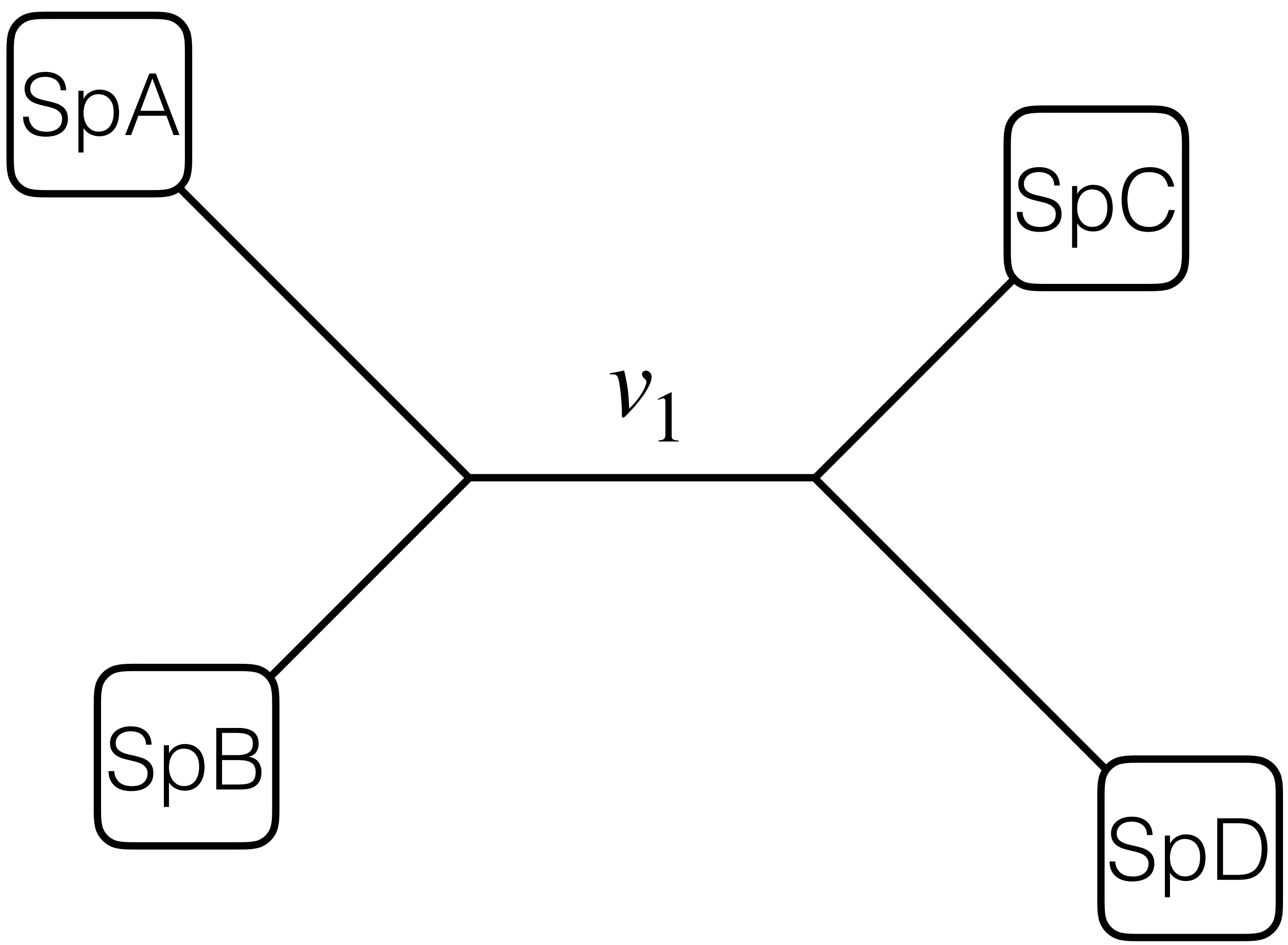
Interior Branch Test

H_0: The length of this branch is 0 (effectively, it does not exist).



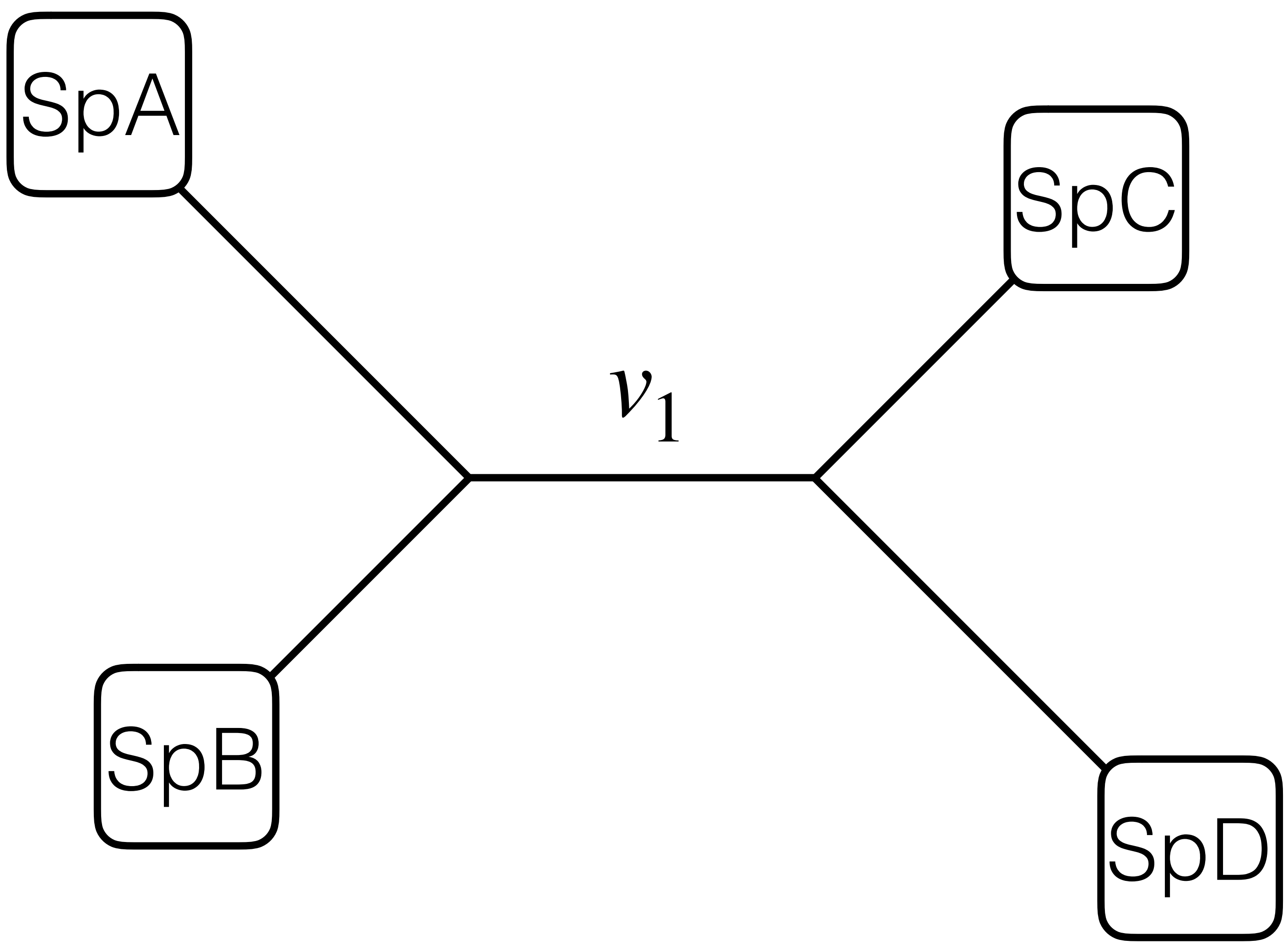
Interior Branch Test

H_1: The length of this branch is > 0.



Interior Branch Test

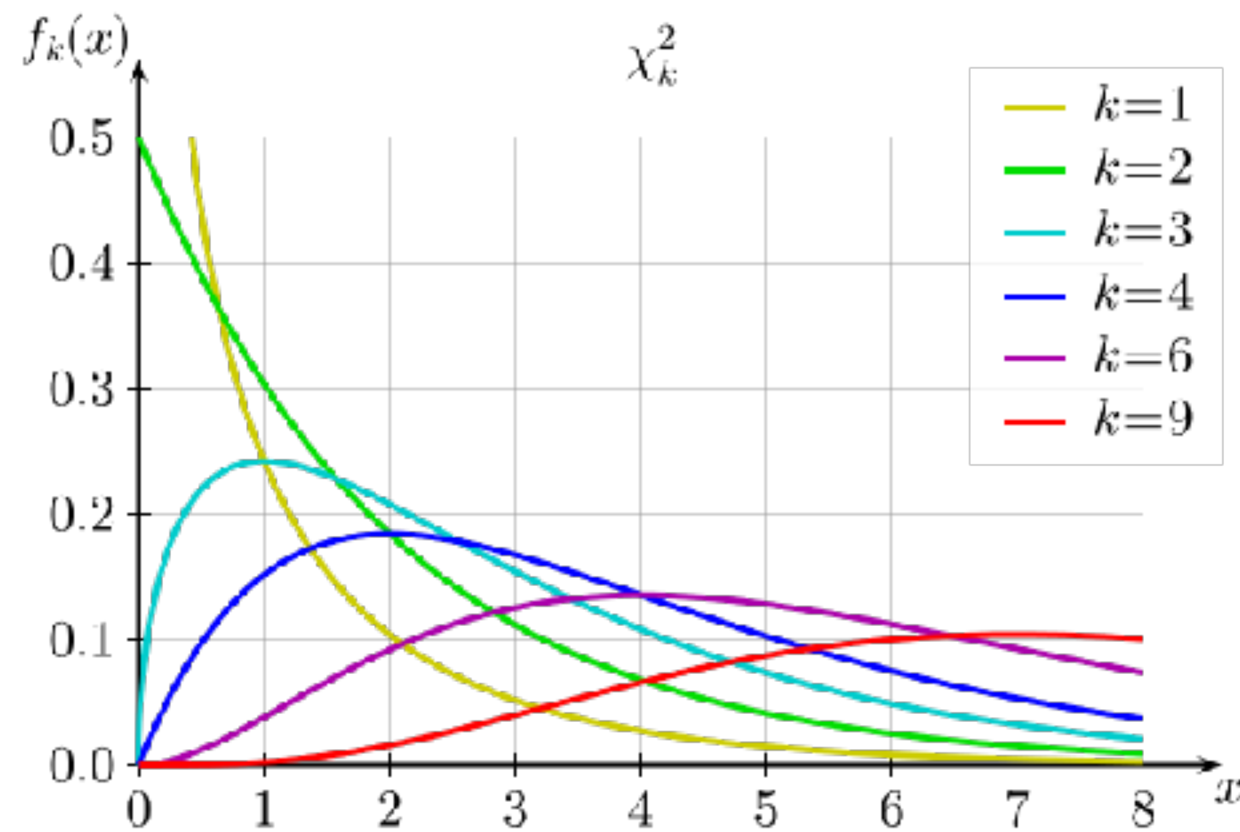
H_0 and H_1 are nested, so we can use an LRT.



Interior Branch Test

H$_0$ and H$_1$ are nested, so we can use an LRT.

$$LR = \frac{\mathscr{L}(v_1 = 0)}{\mathscr{L}(v_1 > 0)}$$



Because the value of v_1 is at a boundary of parameter space, we should use a 50:50 mixture of χ^2_0 and χ^2_1 .

Interior Branch Test

Challenges and Drawbacks

- Hypothesis not really specified *a priori* (only after inferring ML tree)
- Multiple tests if applied to all branches
- Not clear that H0 is actually correct/useful. Branch lengths may have expectations > 0 even for incorrect trees.

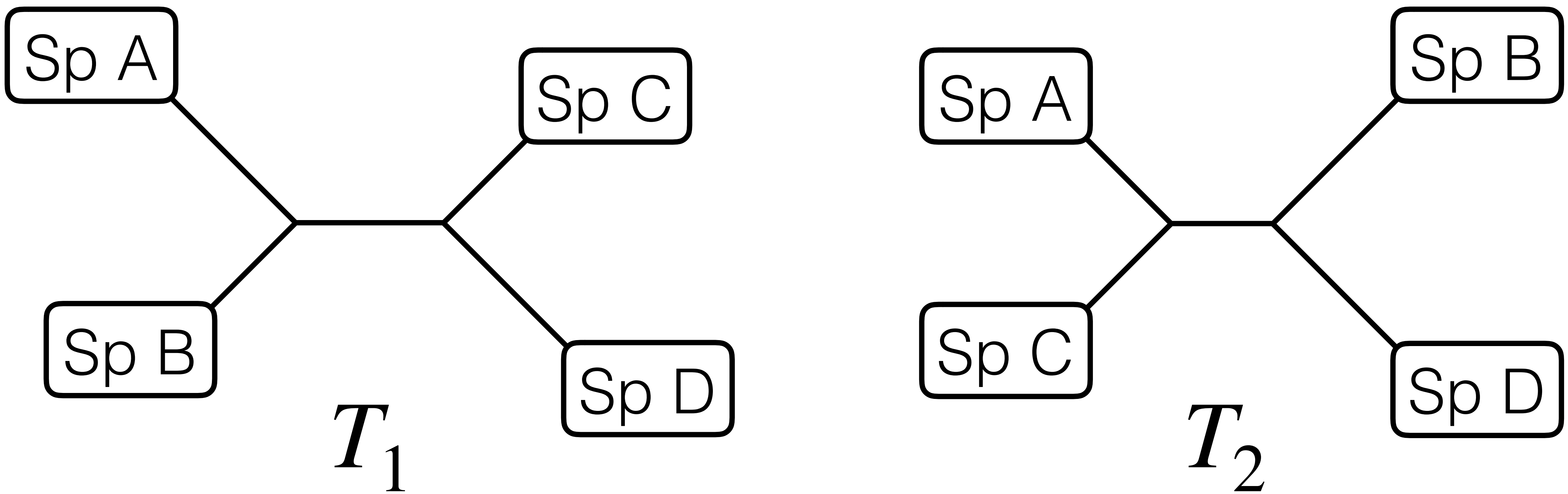
Interior Branch Test

Challenges and Drawbacks

- Hyp … after
  infe …
- Mu …
- No …
  Bra … ven
  for …

Not really used in empirical studies.

K-H Test



Test Statistic

$$\Delta = ln(\mathscr{L}_1) - ln(\mathscr{L}_2)$$

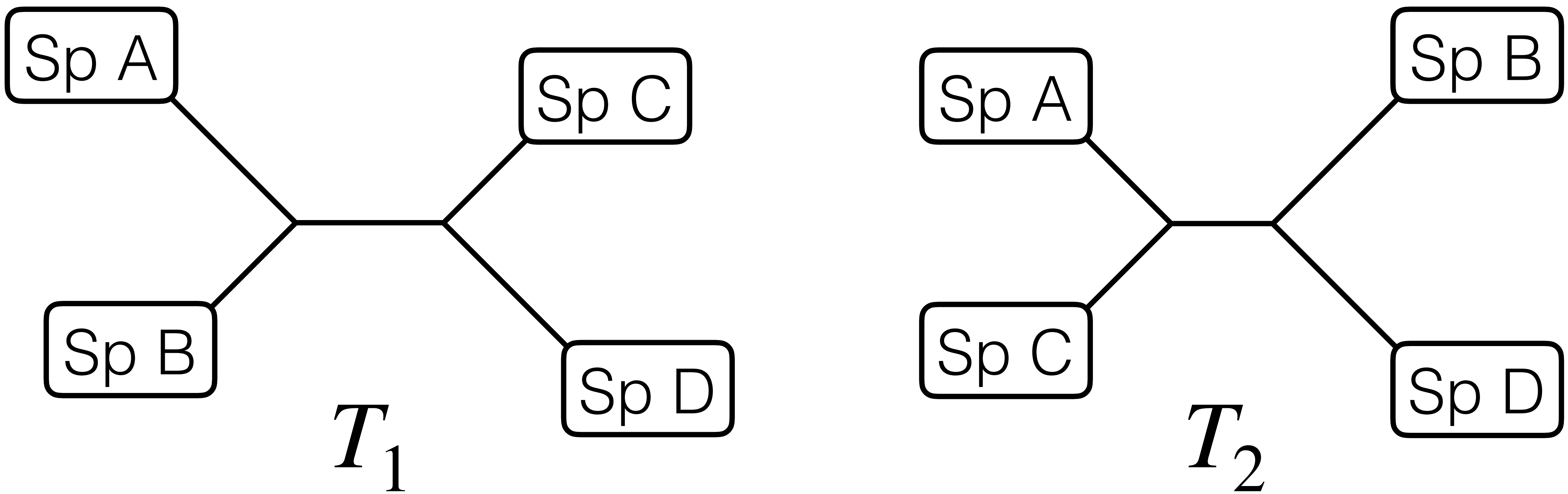
K-H Test

$$\Delta = ln(\mathscr{L}_1) - ln(\mathscr{L}_2)$$

The Intuitive Approach (*priNPfcd*)

1. Bootstrap data
2. Reoptimize and recalculate likelihood differences
3. After all bootstraps, take mean of differences and center distribution, so avg. difference is 0 (H_0).
4. Compare observed difference to centered, bootstrap distribution.

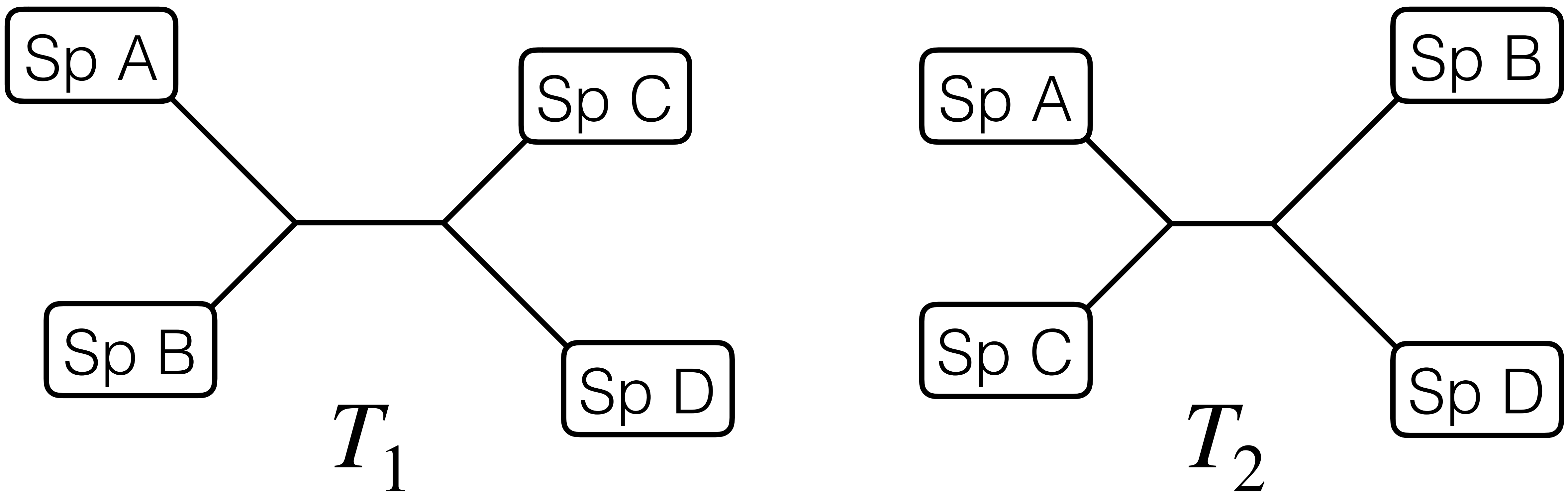
K-H Test



A Faster Approach (*priNPnca*)

This is the one described in Yang's book.

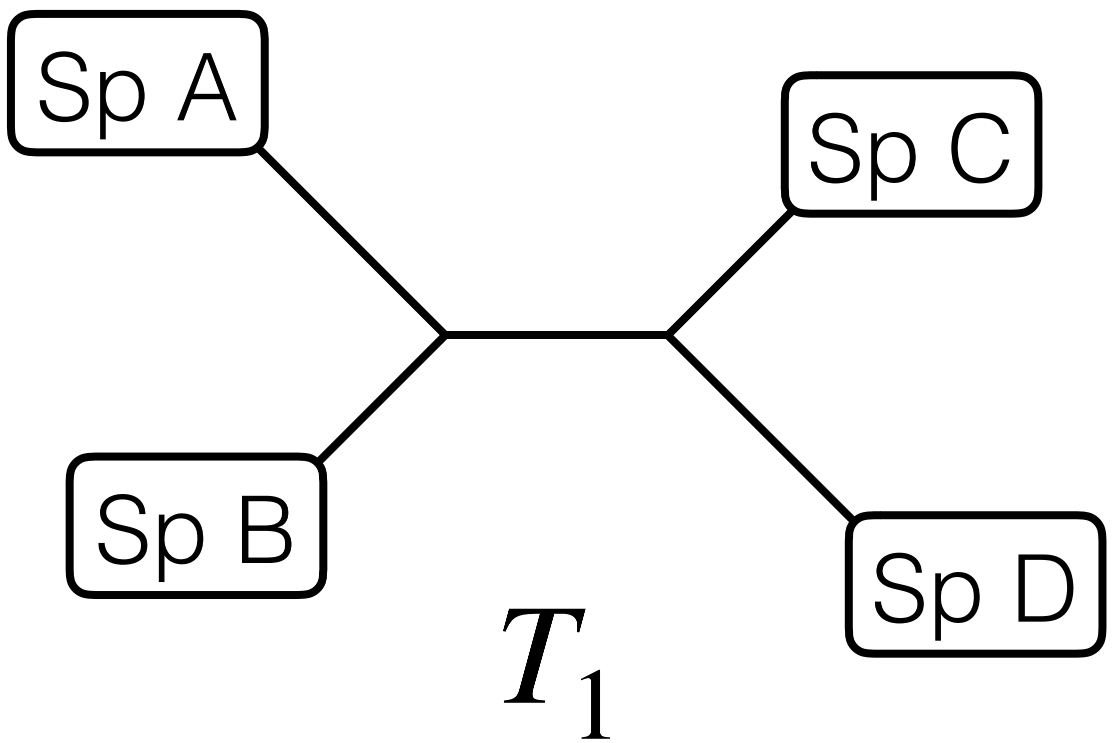
K-H Test

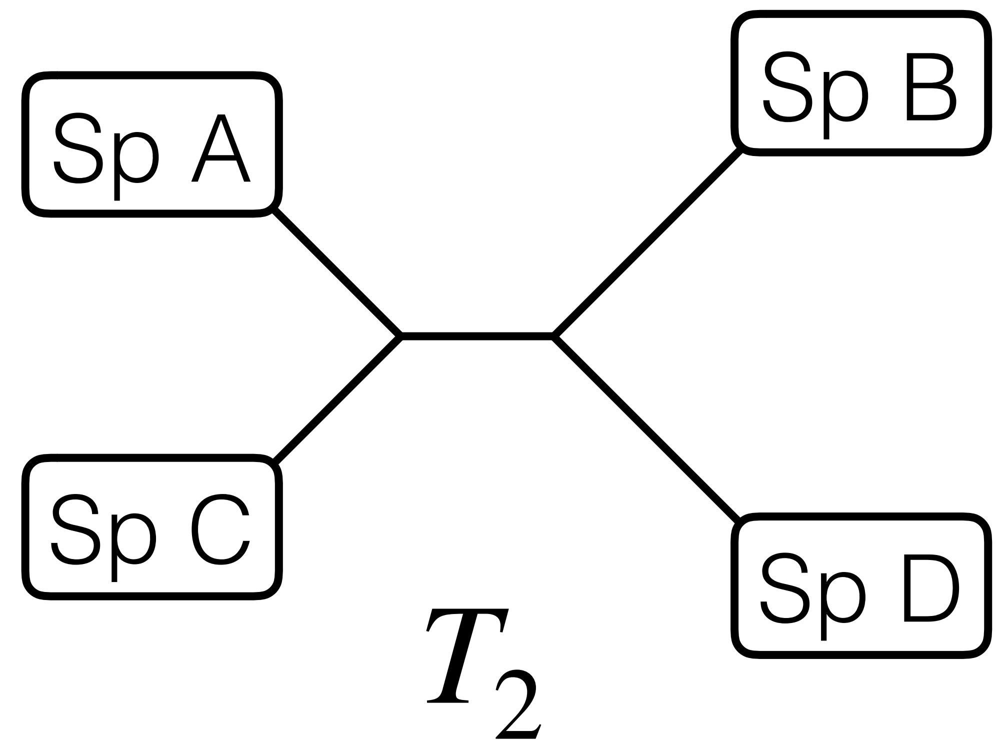


$$ln(\mathscr{L}_1) = \sum_{h=1}^{n} lnP(\mathbf{x}_h | \hat{\theta}_1)$$

Site Likelihoods

K-H Test



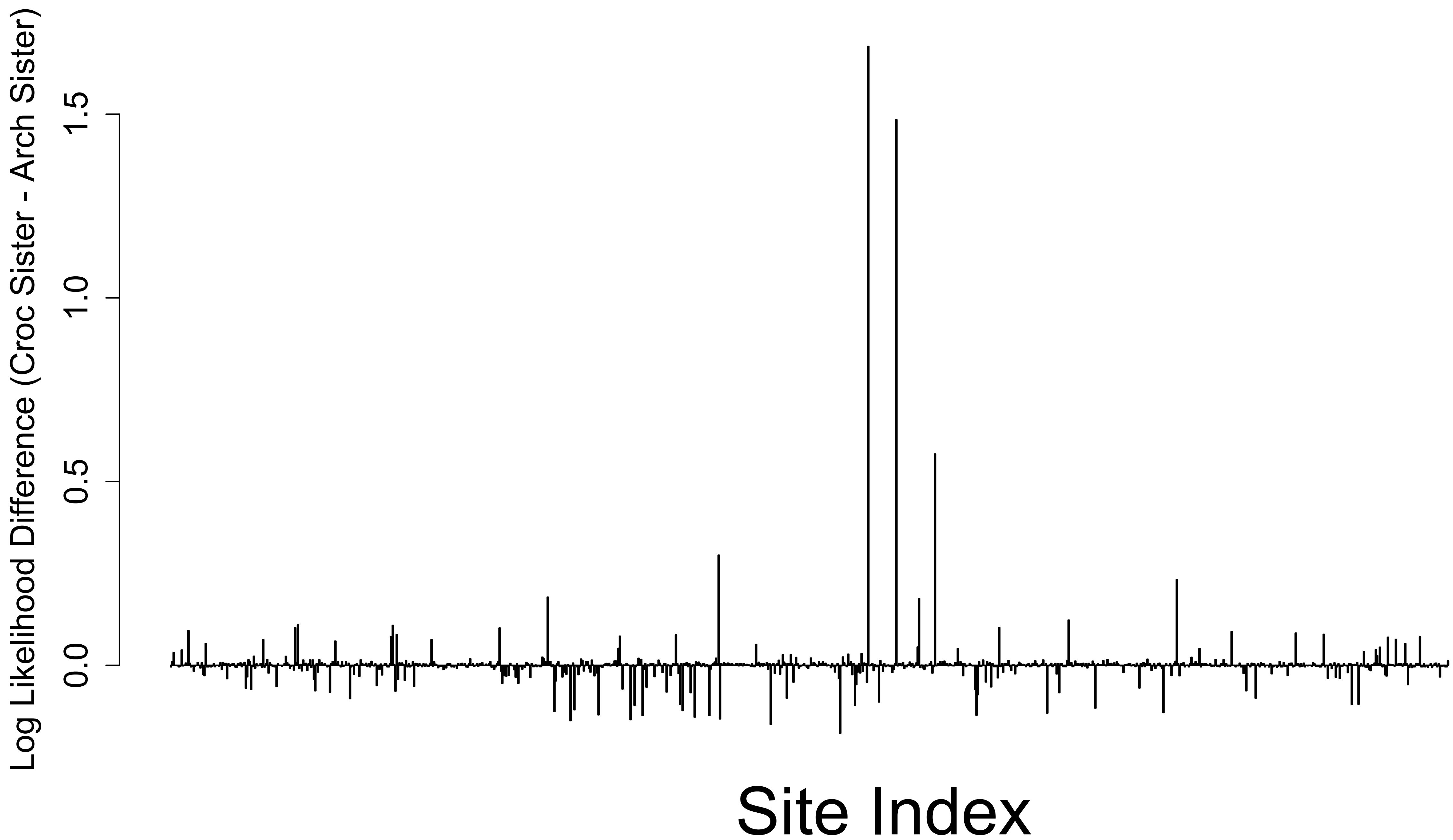


$$d_h = lnP(\mathbf{x}_h \mid \hat{\theta}_1) - lnP(\mathbf{x}_h \mid \hat{\theta}_2)$$

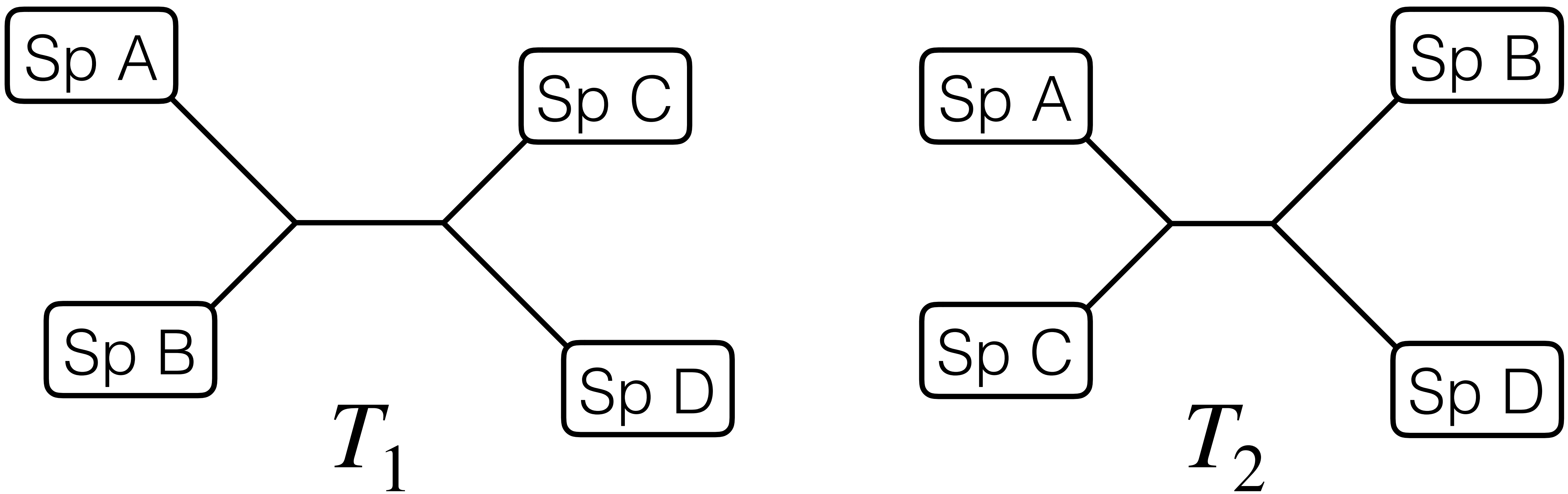
These can be used to estimate expected variance of Δ (with some assumptions)

$$\bar{d} = \frac{\Delta}{n}$$

Site Likelihood Example



K-H Test

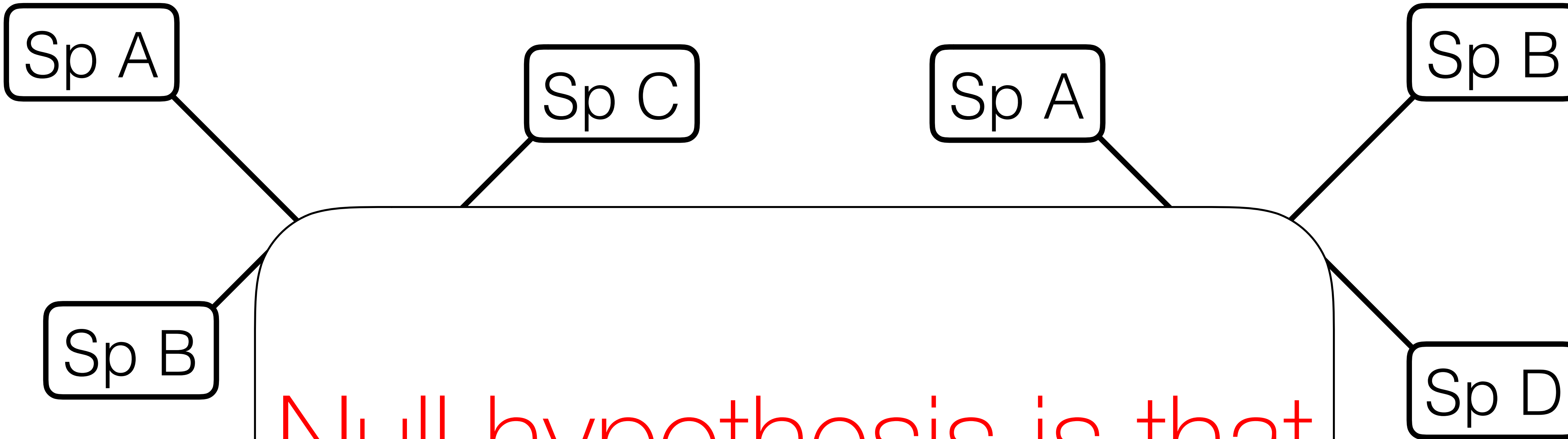


$$var(\Delta) = \frac{n}{n-1}\sum_{h=1}^{n}(d_h - \bar{d})^2$$

Reject T2 if Δ is $> [var(\Delta)]^{1/2}$

priNPnca from Goldman et al. (2000)

K-H Test

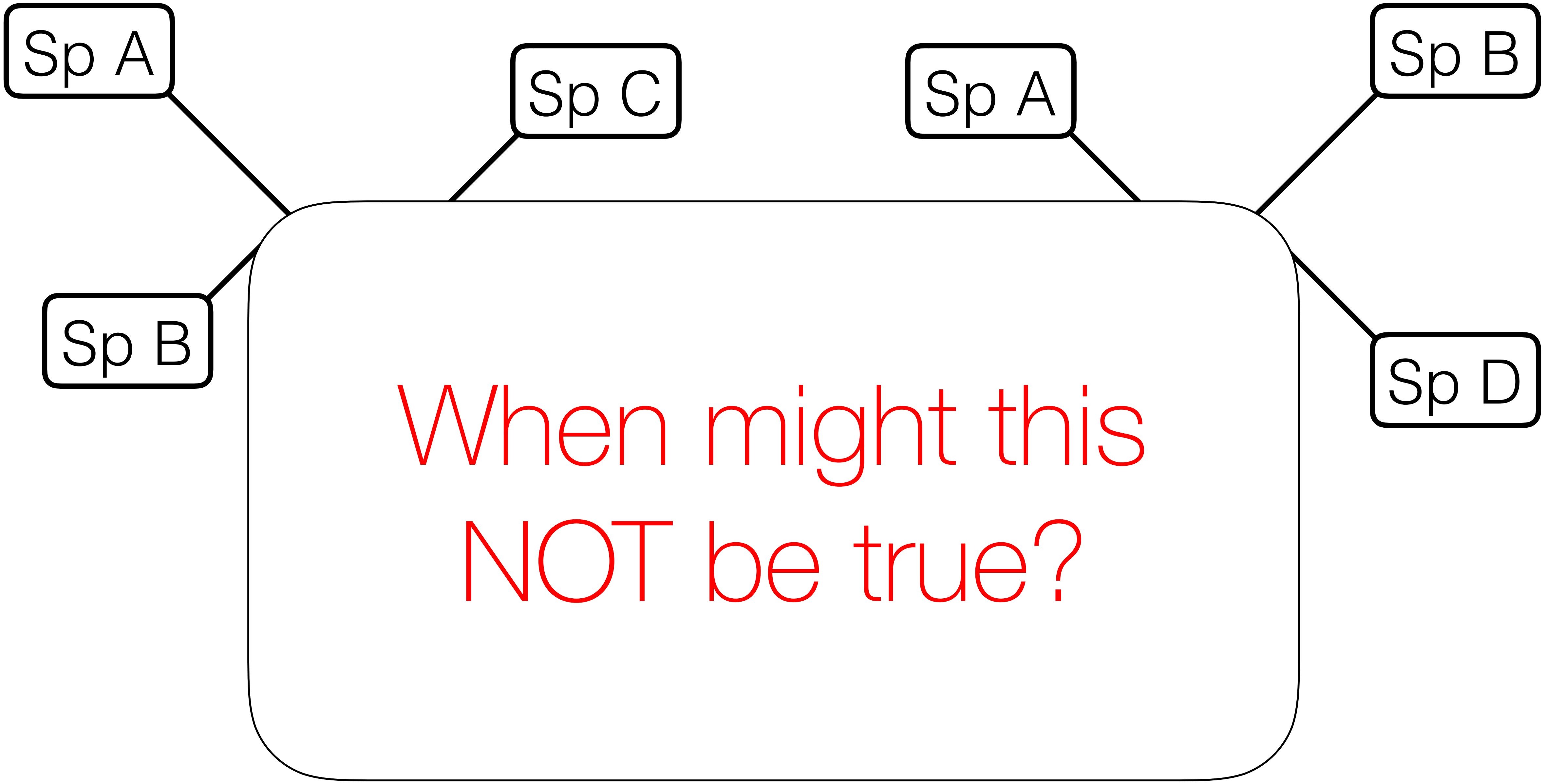


Null hypothesis is that $\Delta = 0$.

Reject T2 if Δ is $> [var(\Delta)]^{1/2}$

priNPnca from Goldman et al. (2000)

K-H Test

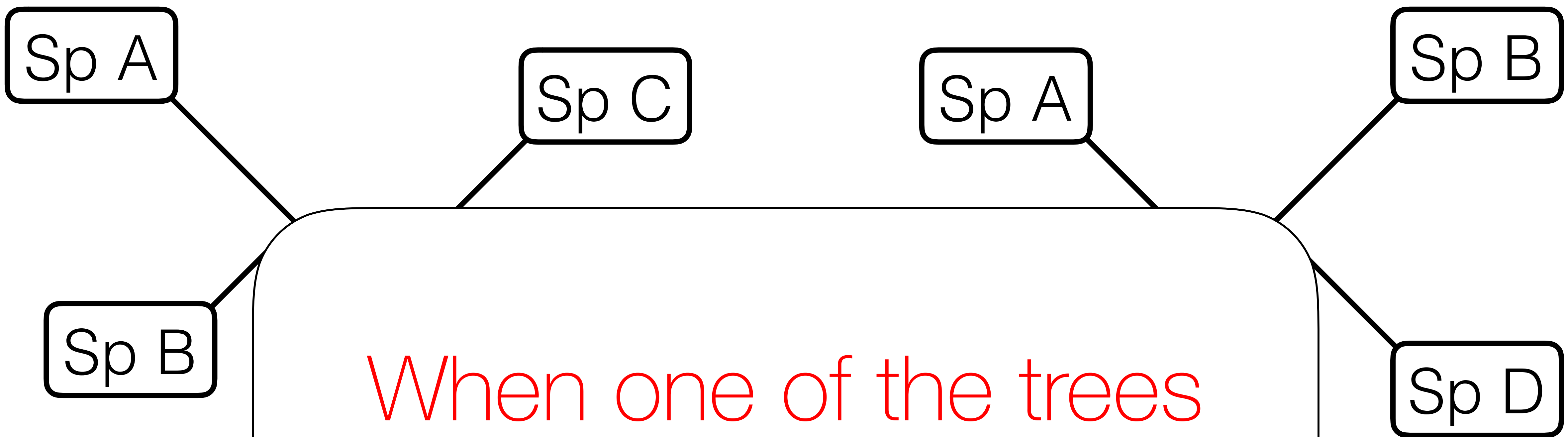


When might this NOT be true?

Reject T2 if Δ is $> [var(\Delta)]^{1/2}$

priNPnca from Goldman et al. (2000)

K-H Test

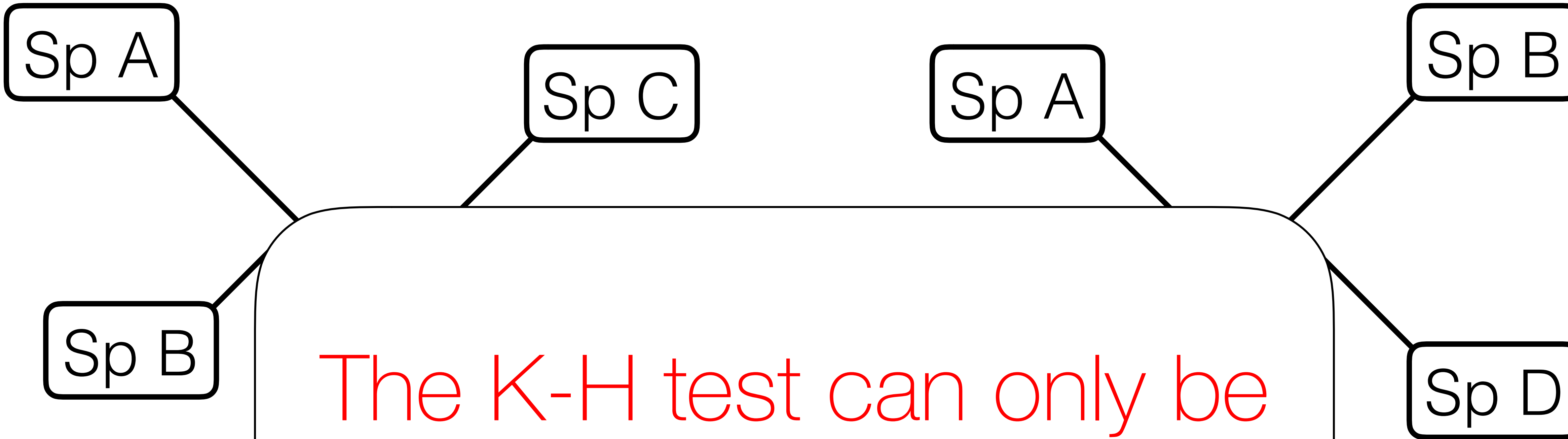


When one of the trees being compared is already known to be the ML tree!

Reject T2 if Δ is $> [var(\Delta)]^{1/2}$

priNPnca from Goldman et al. (2000)

K-H Test

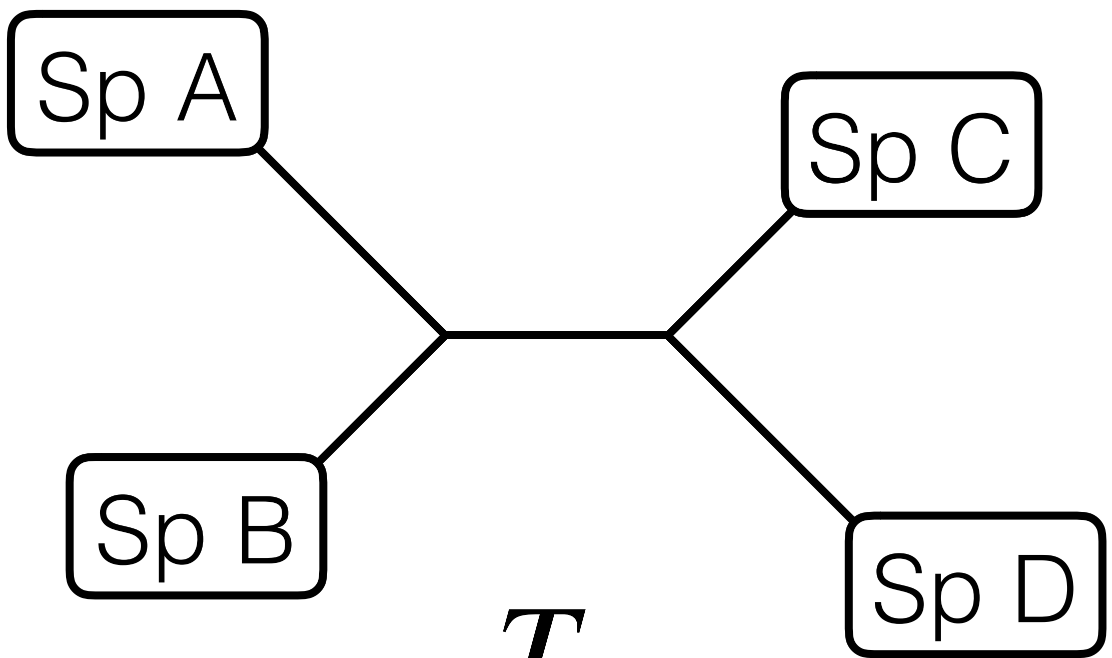


The K-H test can only be applied when topologies are selected **a priori**.

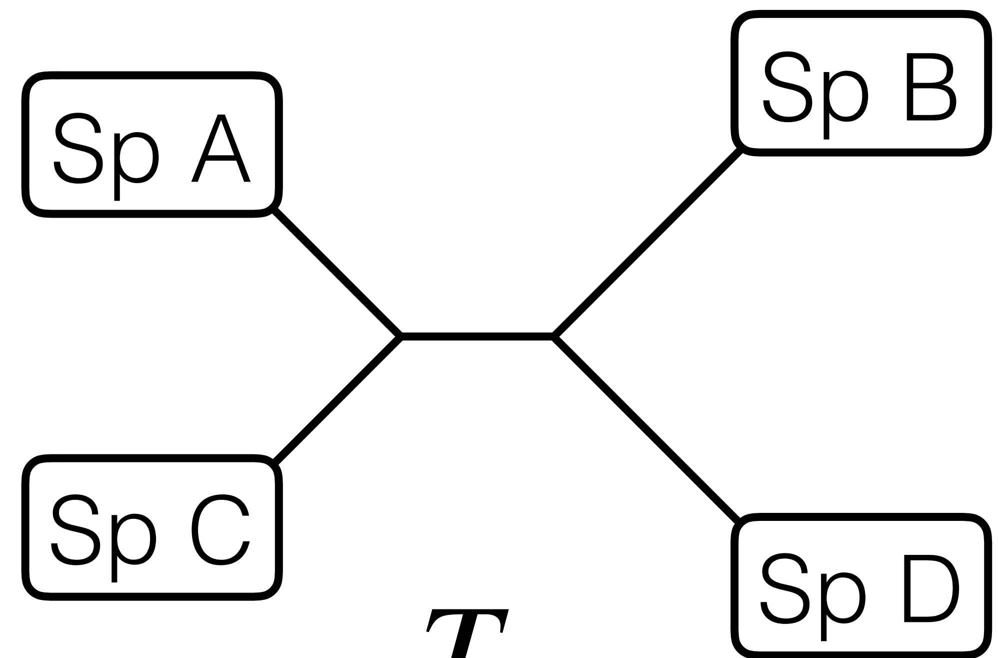
Reject T2 if Δ is $> [var(\Delta)]^{1/2}$

priNPnca from Goldman et al. (2000)

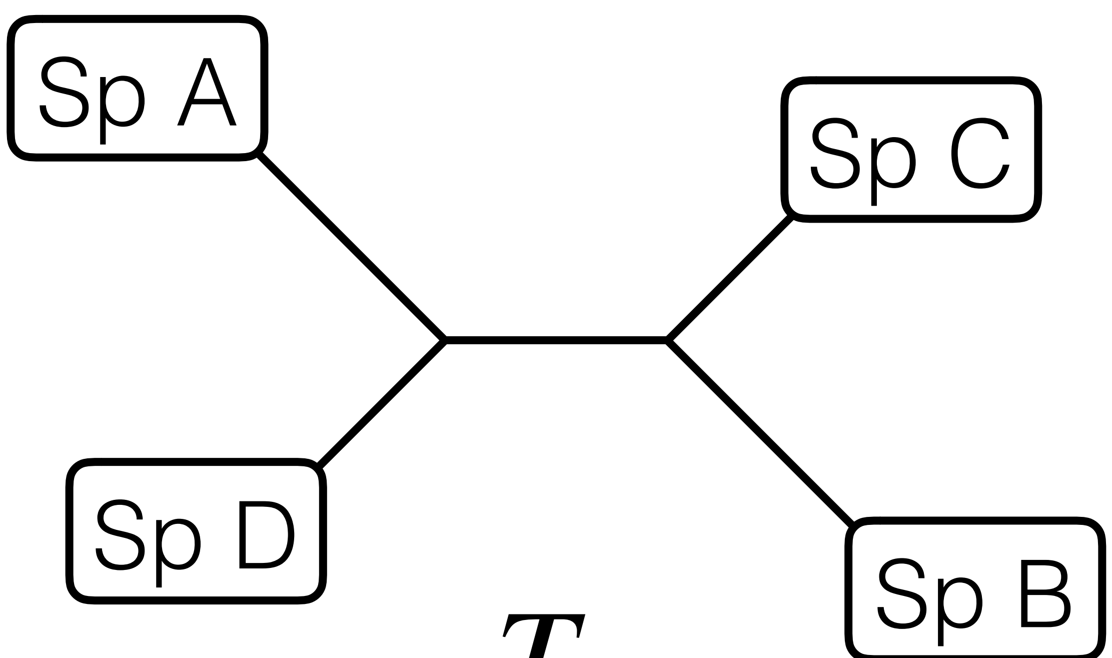
S-H Test



T_1



T_2



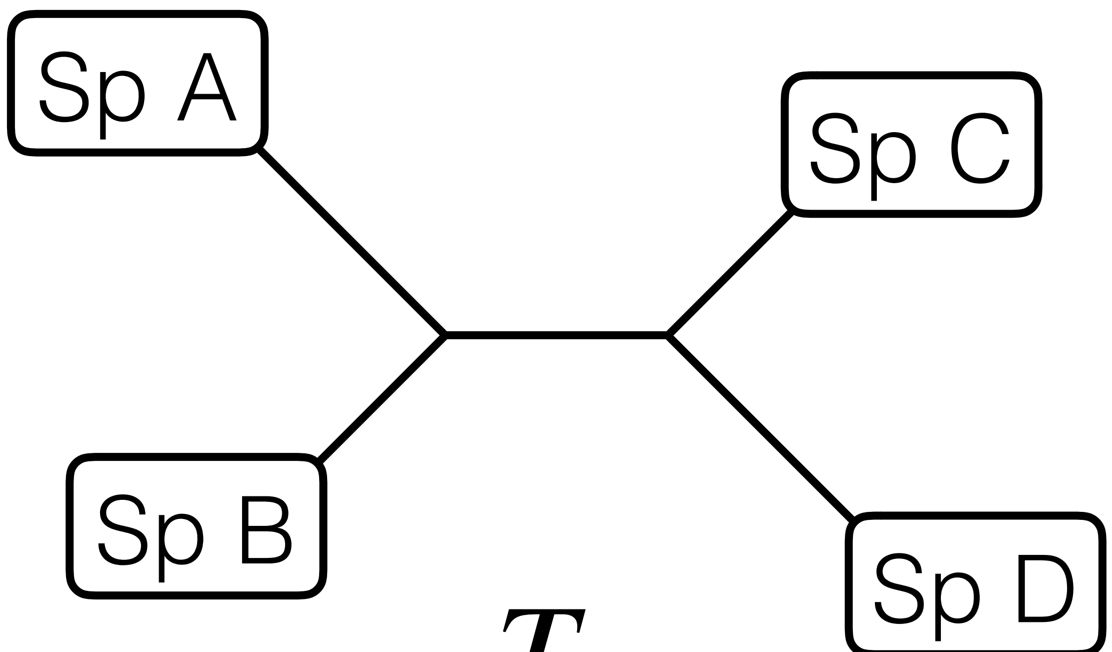
T_3

Start by considering all possible trees.

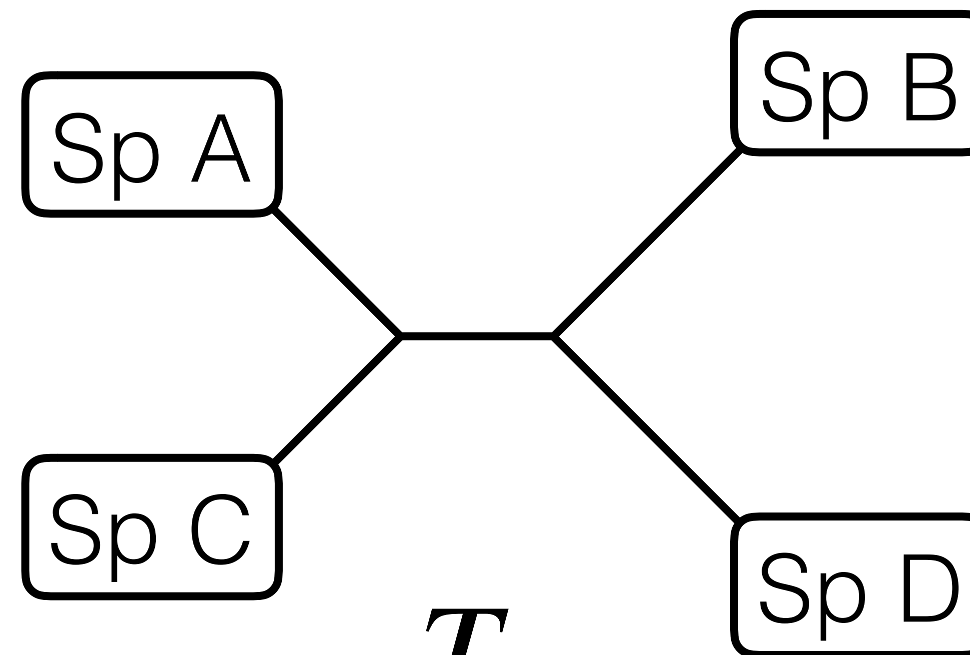
Or a set that MUST contain the true tree.

This set must be selected a priori.

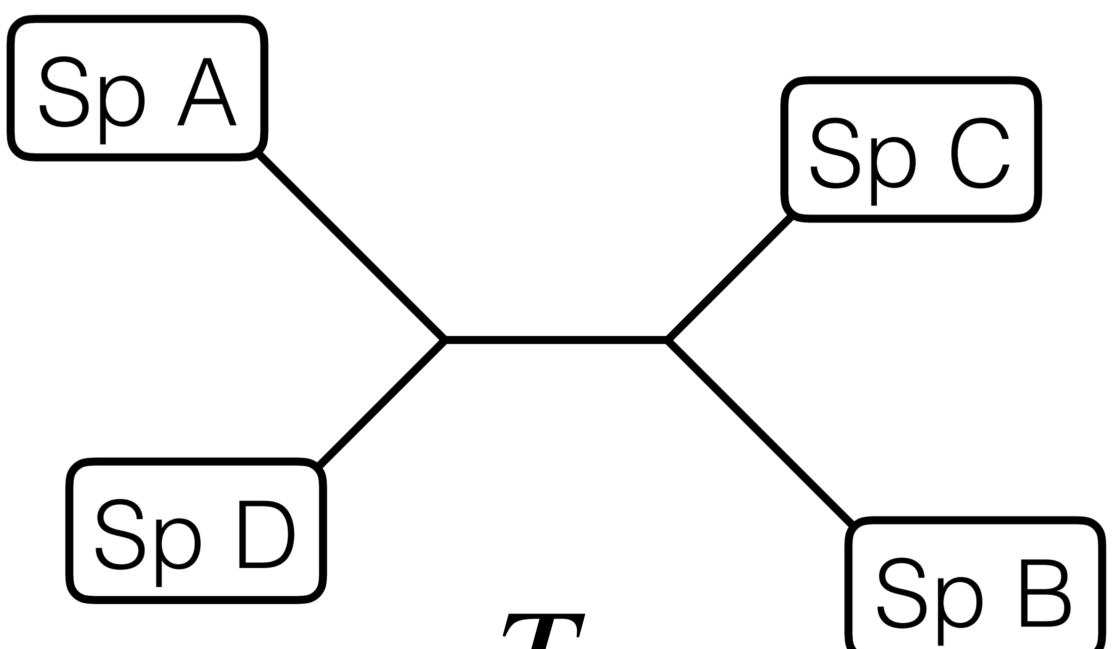
S-H Test



T_1



T_2

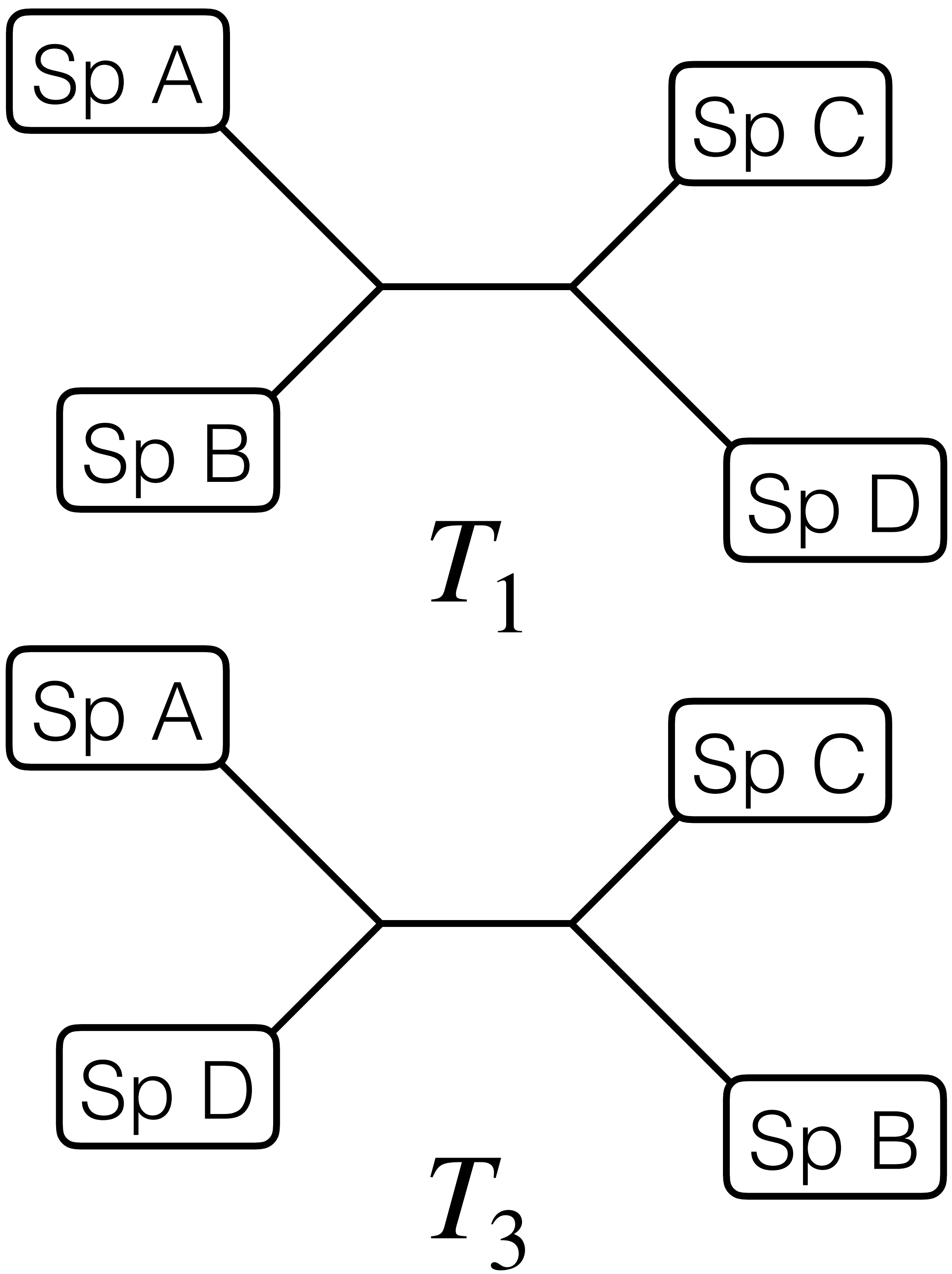


T_3

H_0: All *T* are equally good explanations of the data.

H_A: Some *T* are not equally good explanations.

S-H Test



H_0: All *T* are equally good explanations of the data.

H_A: Some *T* are not equally good explanations.

S-H Test

The Intuitive Explanation (*posNPfcd)*

1. Calculate L_{ML} - L_x.
2. Bootstrap
3. Reoptimize for each bootstrap
4. Center the bootstrapped likelihoods for each topology across replicates.
5. For each replicate, recalculate L_{ML} - L_x from the recentered values.
6. For each topology, compare observed difference to the reentered, bootstrap distribution. This test is one-sided and gives a P-value for each topology.