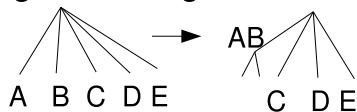


## Figures

Figure 1 - Overview of FastTree.

## Neighbor Joining with Profiles



Profile (A)    **ACGTACGTACGT**  
 Profile (B)    **A-CGACGTAC-T**  
 Profile (AB)   **A<sub>-ccg</sub>ACGTAC<sub>gt</sub>-T**  
 O(NLa) space instead of O(N<sup>2</sup>) space

Neighbor-joining Criterion:

find the join that minimizes  $d(A,B) - r(A) - r(B)$

Distances to joined nodes:

Neighbor joining:  $d(AB,C) = (d(A,C)+d(B,C))/2 - d(A,B)/2$

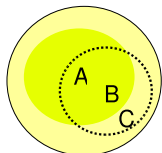
FastTree:  $= \Delta(AB,C) - u(C) - u(AB)$     ← “up-distance” u

Average out-distances:

Neighbor joining:  $r(A) = \sum d(A,X)/(n-2)$

FastTree:  $= \frac{n \cdot \Delta(A, \sum X/n) - \Delta(A,A) - \sum (u(A)+u(X))}{n-2}$

## Top-hits Heuristic



If B is close to A, then the best join for B is also close to A:

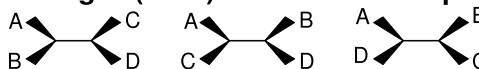
$\text{TopHits}(B) \subset \text{TopHits}(A)$  with a larger radius

When we do a join:

$\text{TopHits}(AB) \subset \text{TopHits}(A) \cup \text{TopHits}(B)$

## Nearest-neighbor interchanges (NNIs) & Local bootstrap

Consider three local topologies:

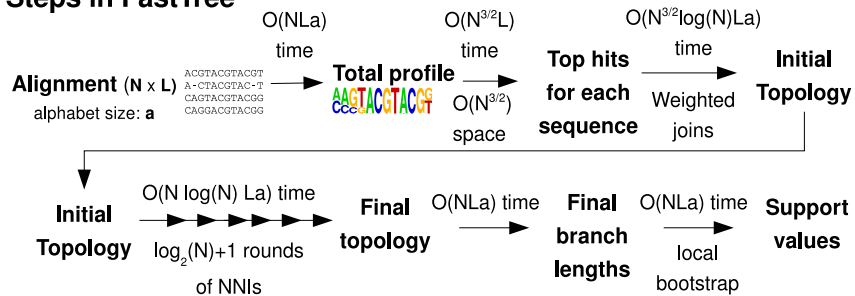


Minimum evolution criterion:

prefer AB|CD if  $d_{AB} + d_{CD} < \min(d_{AC} + d_{BD}, d_{AD} + d_{BC})$

where  $d_{AB} = \text{LogCorrection}(\Delta(A,B))$

## Steps in FastTree



**Figure 2 - Distribution of support values for simulated alignments of 250 protein sequences with gaps.**

We compare the distribution of FastTree's local bootstrap and the traditional (global) bootstrap for correctly- and incorrectly-inferred splits. The right-most bin contains the strongly-supported splits (0.95-1.0).

