

Ultrafast Approximation for Phylogenetic Bootstrap

Bui Quang Minh, Minh Anh Thi Nguyen, and Arndt von Haeseler

Center for Integrative Bioinformatics Vienna (CIBIV), Austria

Software freely available at <http://www.cibiv.at/software/igtree>



Introduction

The nonparametric bootstrap of Felsenstein (1985) is the de facto standard for phylogenetic analysis to assess branch supports of phylogenetic trees. Yet bootstrap with maximum likelihood (ML) is hampered by the enormous computation time required.

Here, we present an ultrafast bootstrap approximation approach (UFBoot) with the following advantages:

1. Unbiased branch support values (95% means a probability of 0.95 to be correct)
2. Robustness against moderate model violations
3. Substantial speedup compared to RAxML rapid bootstrap

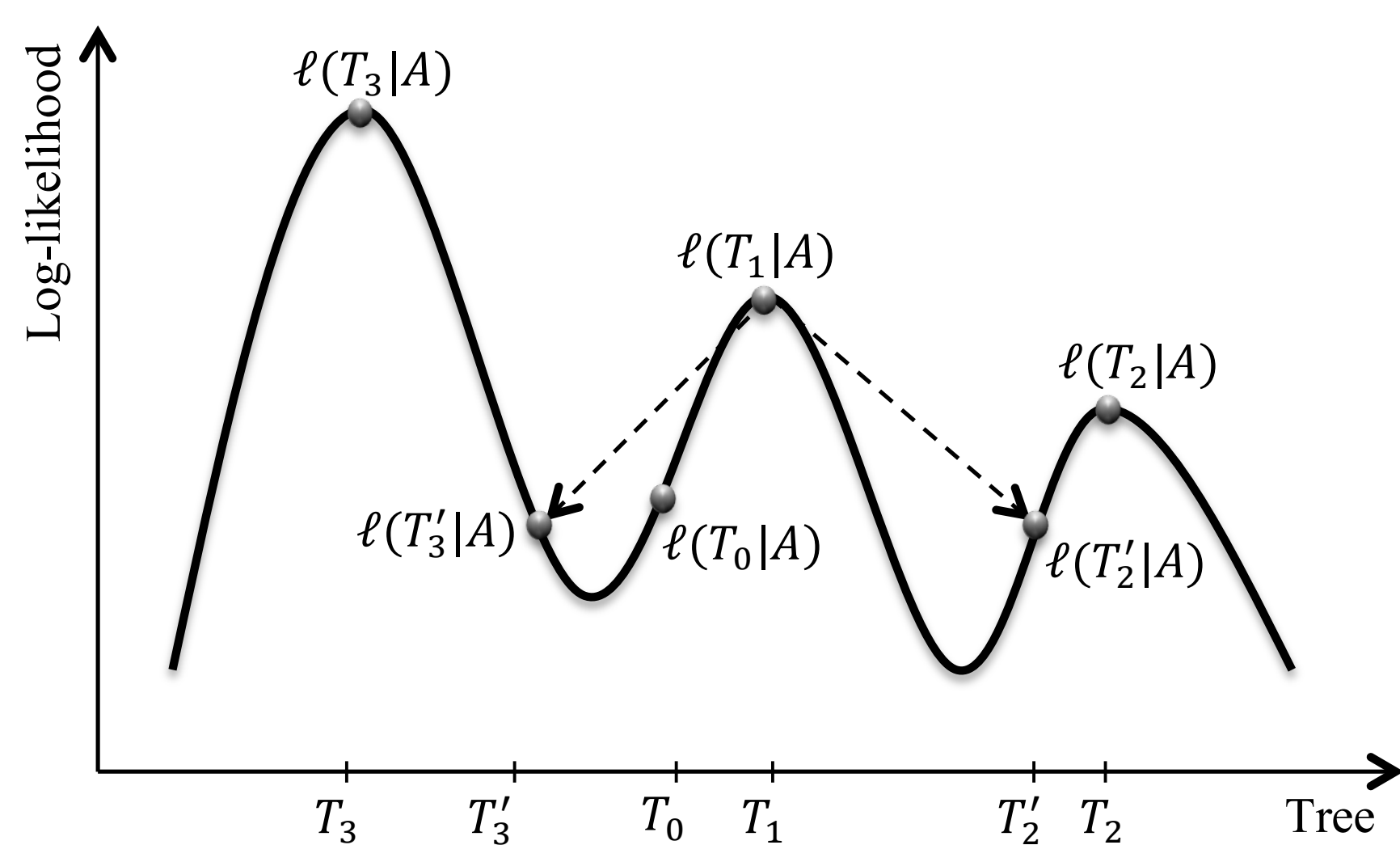
Minh B.Q., Nguyen M.A.T., and von Haeseler A. (2013) *Mol. Biol. Evol.*, 30:1188-1195. (Open Access Article)

UFBoot Method

1. Initialization step

- Generate B (≥ 1000) bootstrap replicates A_1^*, \dots, A_B^* from original alignment A .
- Initialize bootstrap trees $T_1^*, \dots, T_B^* := \emptyset$.

2. Exploration step

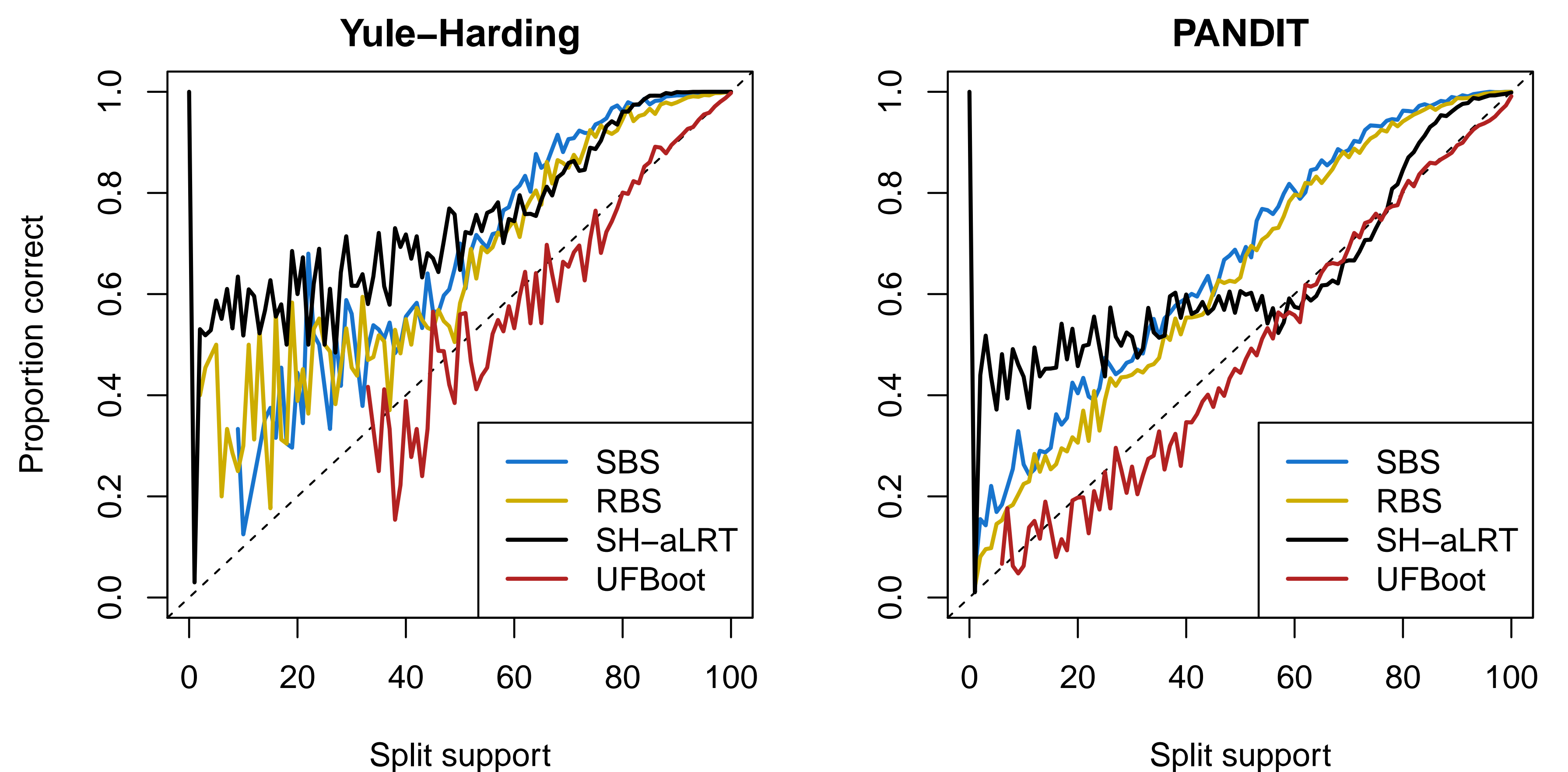


- Resample the tree space on A with important quartet puzzling and nearest neighbor interchange (IQPNNI) algorithm.
- Upon encountering a new tree T where $\ell(T|A) \geq \ell_{\min}$:
Compute approximate $\hat{\ell}(T|A_b^*)$ using resampling estimated log-likelihood (RELL). If $\hat{\ell}(T|A_b^*) > \hat{\ell}(T_b^*|A_b^*)$, update $T_b^* := T$.

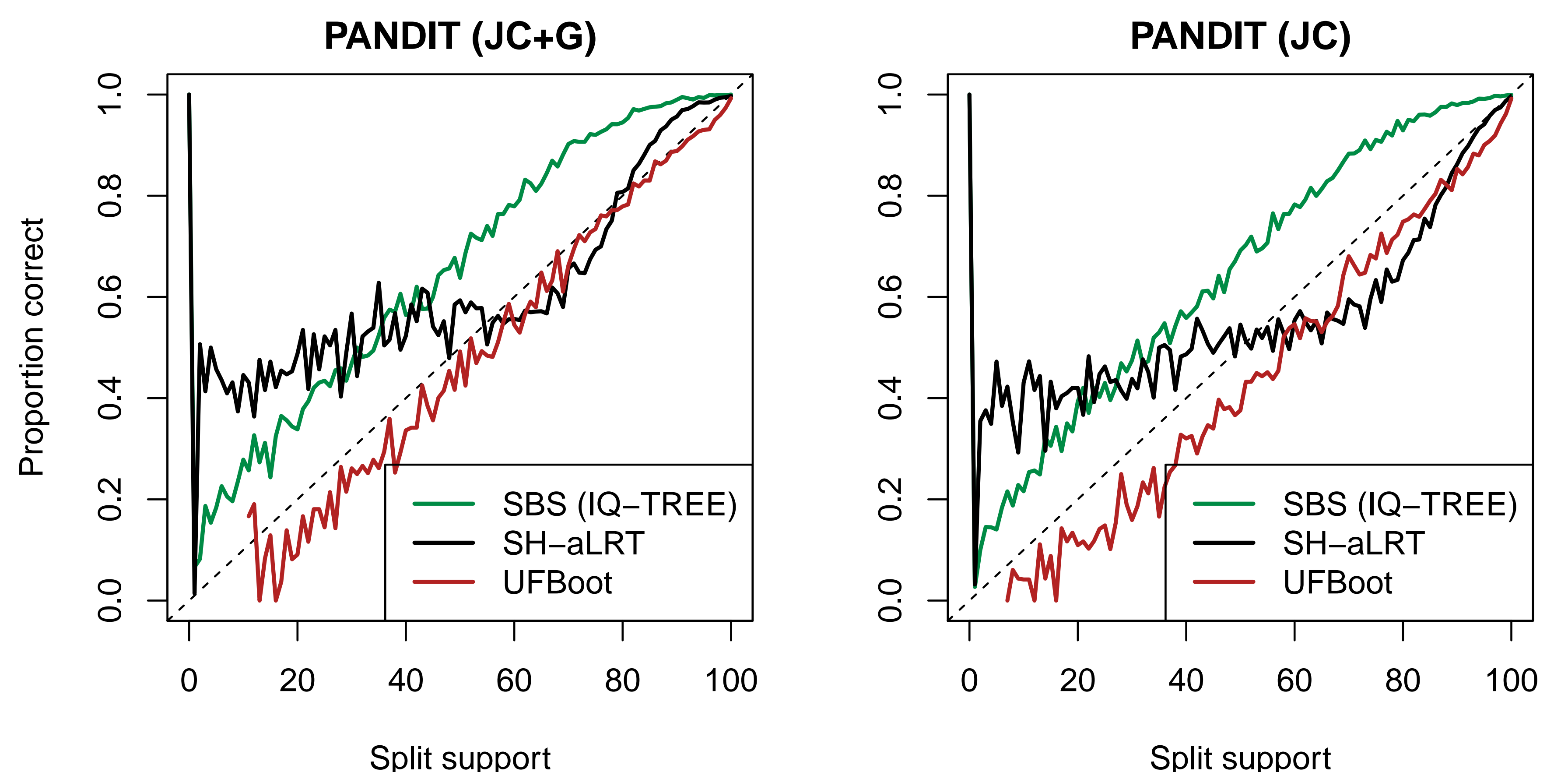
3. Summarization step. Construct a consensus tree from $\{T_1^*, \dots, T_B^*\}$ and place support values on the ML tree.

Results

1. UFBoot is unbiased while other methods are conservative:

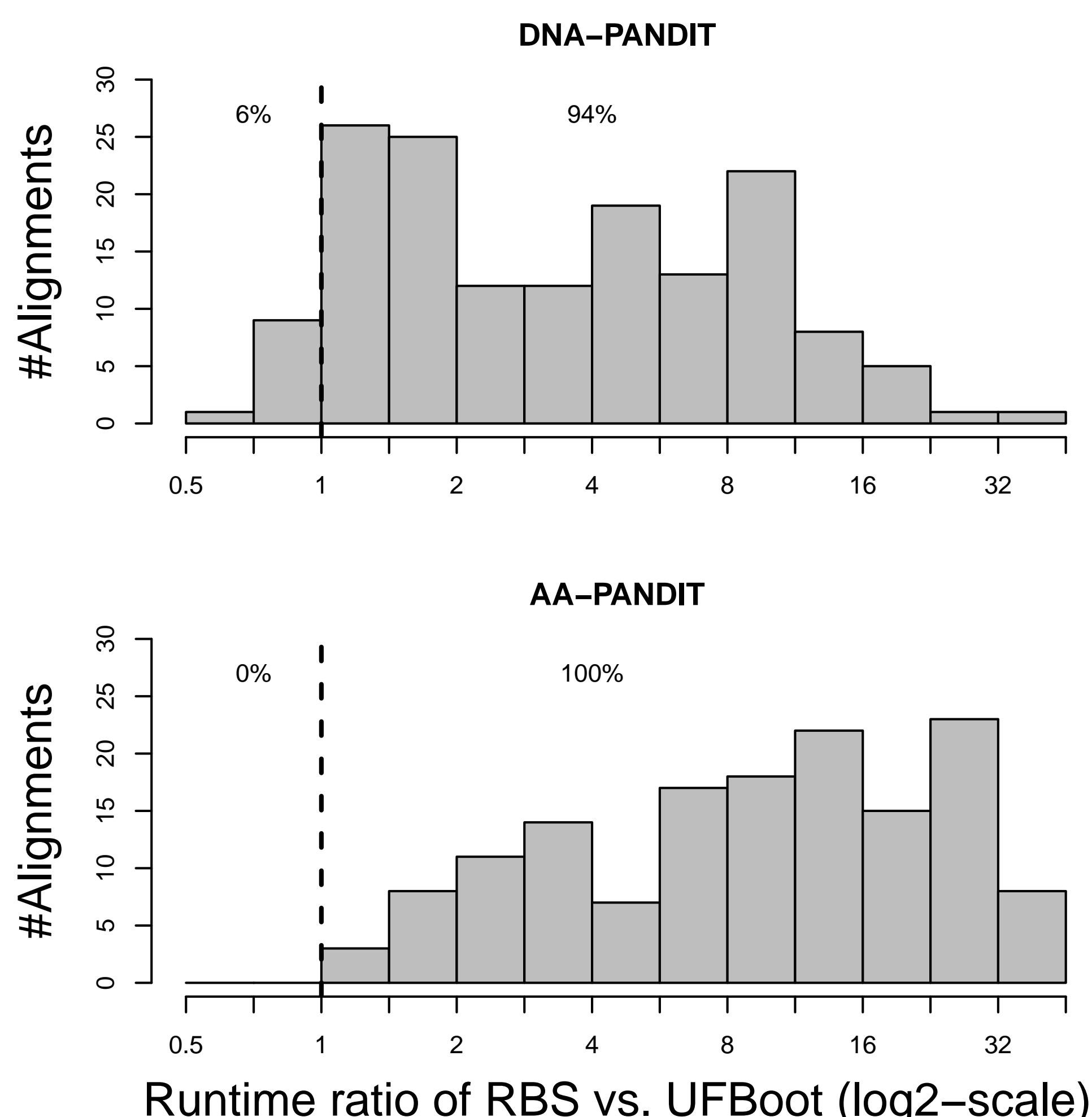


2. UFBoot is robust against moderate model violations:



SBS - standard bootstrap, RBS - RAxML rapid bootstrap, SH-aLRT - Shimodaira-Hasegawa-like approximate likelihood ratio test (PhyML). These accuracy plots follow Hillis and Bull (1993).

3. UFBoot is substantially faster than RBS:



Other key features of IQ-TREE:

- Parallel threading for multi-core CPUs (via OpenMP)
- Substantially faster ML tree reconstruction than IQPNNI
- Numerous substitution models for binary, DNA, protein alignments with automatic selection of best-fit models
- Partition models for phylogenomic data
- Various tree topology tests
- Test of the model homogeneity assumption along the tree