

Maximum Parsimony Inference of Phylogenetic Networks in the Presence of Polyploid Complexes

Supplementary Material

Zhi Yan, Zhen Cao, Yushu Liu, Huw A. Ogilvie, and Luay Nakhleh
Department of Computer Science
Rice University

- Gene tree estimation error on the simulated data sets is shown in Figure S1.
- Comparison results between MPAllopp with and without subgenome assignment in terms of inference accuracy and running time are shown in Figures S2 and S3
- Estimated gene trees for the *Pachycladon* data are shown in Figure S4.
- Reconstructed phylogenetic networks of New Zealand *Pachycladon* are shown in Figure S5.

S1 Estimating phylogenies where the subgenome assignment is unknown

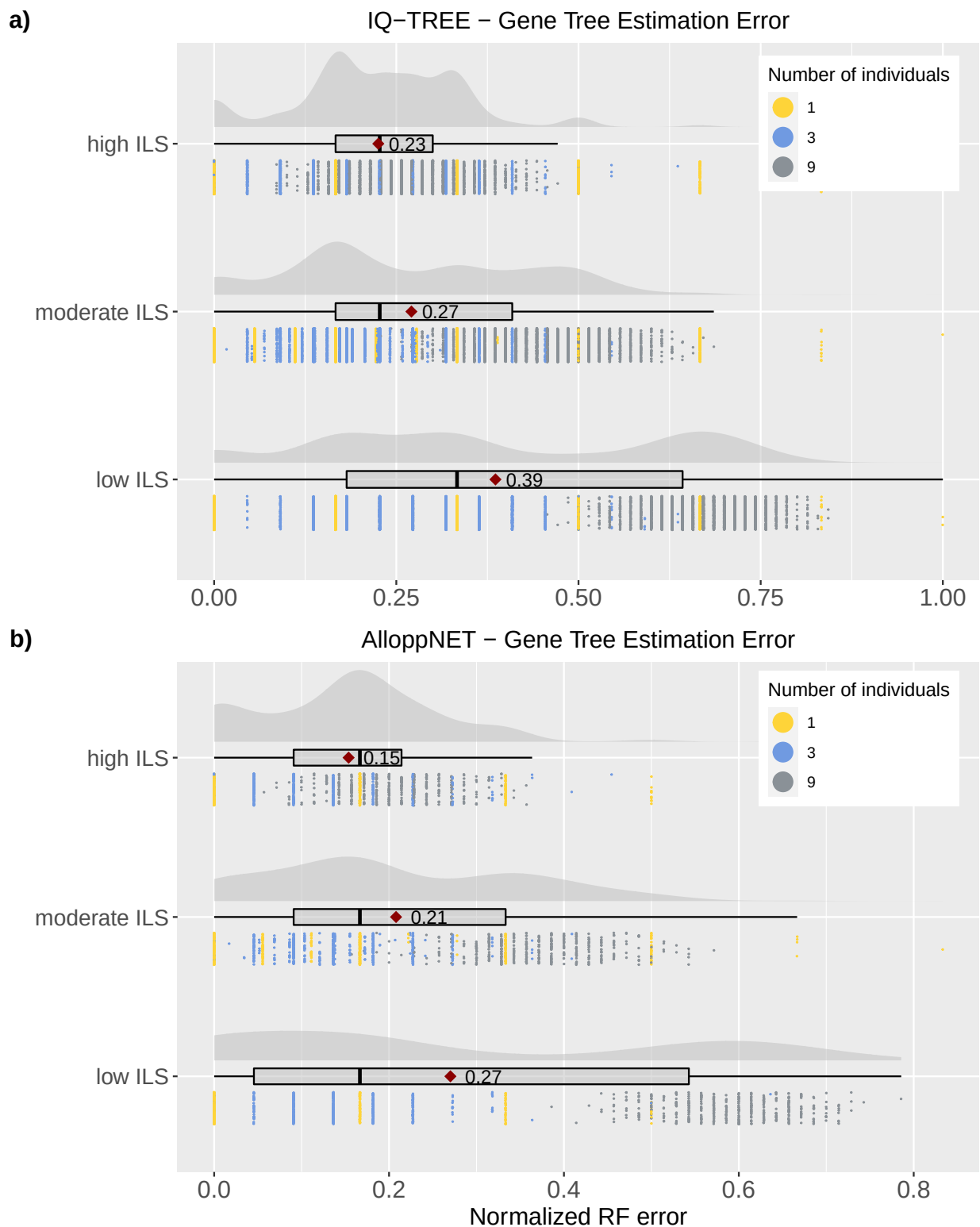


Figure S1: Gene tree estimation error. a) The rooted RF distance between the true gene trees and gene trees estimated by IQ-TREE on the 1/3/10/100 gene data sets. b) The rooted RF distance between true gene trees and maximum clade credibility gene trees derived from AlloppNET on the 1/3/10 gene data sets. Results are shown for gene trees excluding the outgroup. Red diamonds are means for ten replicates.

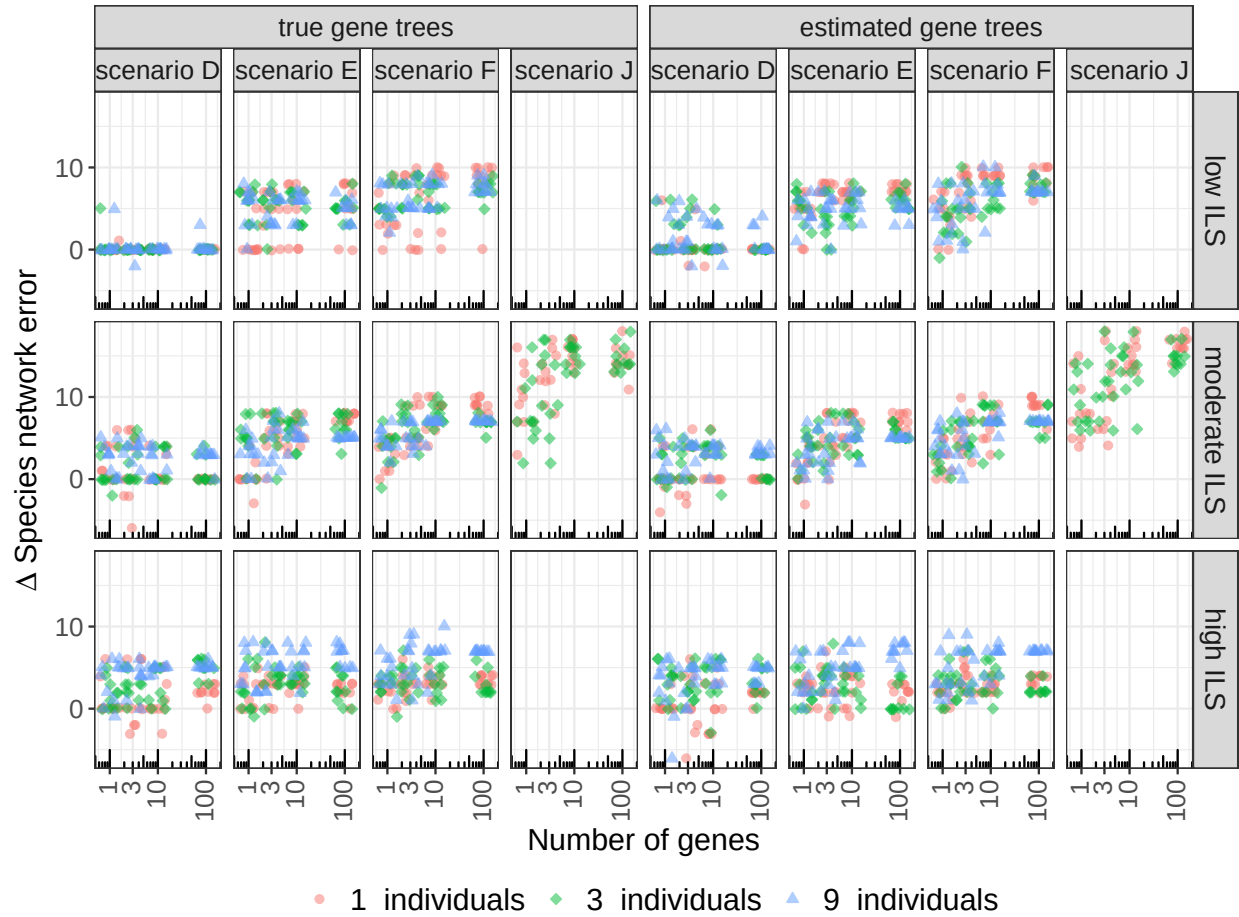


Figure S2: Scatter plots comparing the network estimation error difference between MPAllopp* and MPAllopp on simulated data over 10 replicates. The species network error was measured by the distance between true and estimated networks according to the metric of Nakhleh (2010). A positive value indicates that MPAllopp* yielded higher error than MPAllopp and vice versa for negative values. Here low, moderate and high ILS refer to mutation rates of 4×10^{-9} , 2×10^{-8} and 1×10^{-7} , respectively.

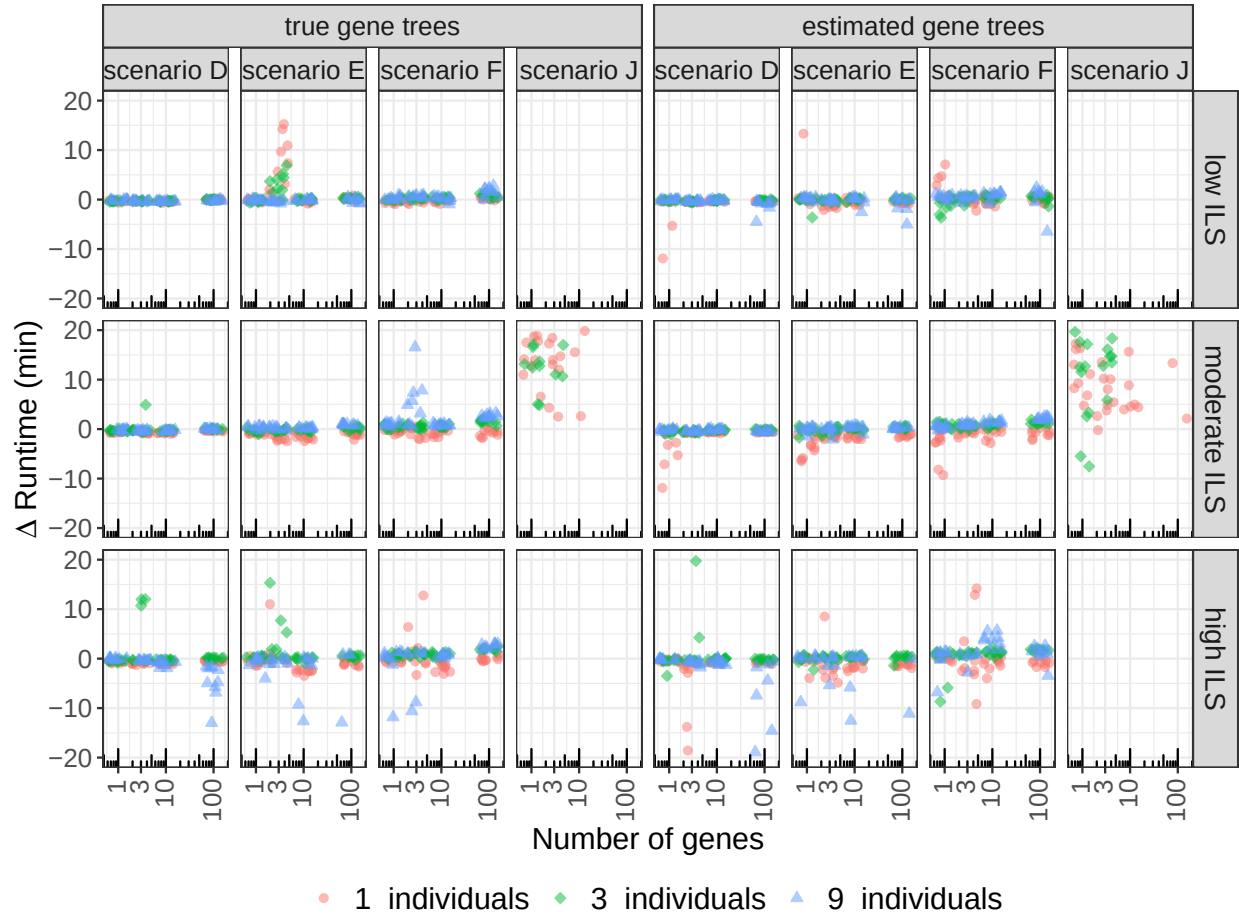


Figure S3: Scatter plots comparing running time (in CPU minutes) difference between MPAllopp* and MPAllopp on simulated data over 10 replicates. A positive value indicates that MPAllopp* required longer runtime than MPAllopp to complete the inference and vice versa for negative values. Here low, moderate and high ILS refer to mutation rates of 4×10^{-9} , 2×10^{-8} and 1×10^{-7} , respectively.

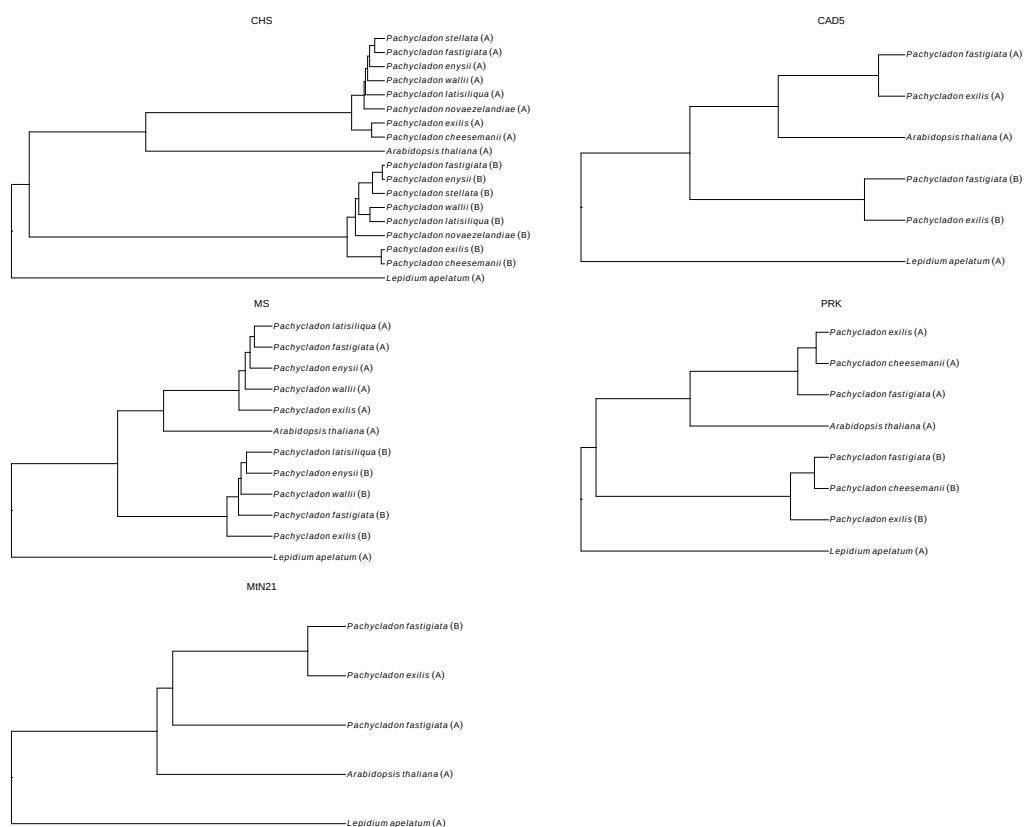
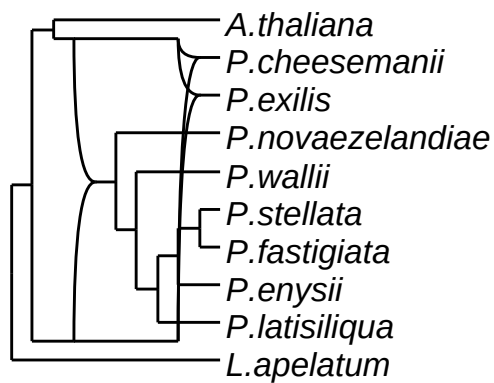
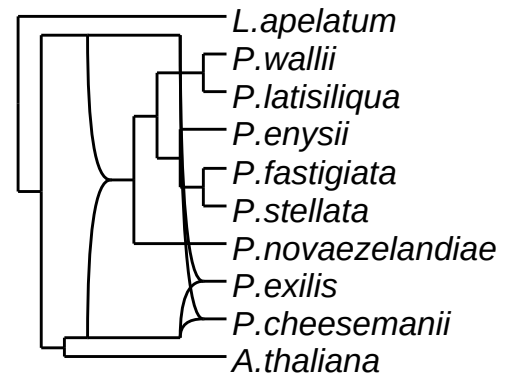


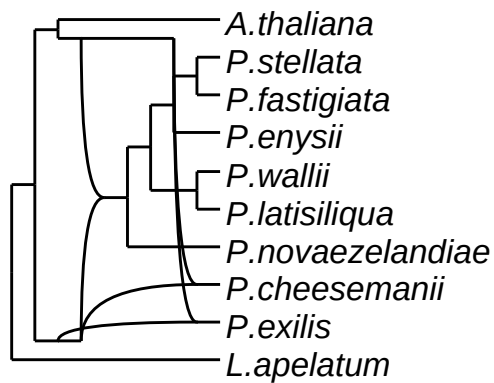
Figure S4: **Maximum clade credibility gene trees of the *Pachycladon* data set.** Gene names are abbreviated as follows: *CHS*, chalcone synthase; *CAD5*, cinnamyl alcohol dehydrogenase 5; *MS*, malate synthase; *PRK*, phosphoribulokinase; *MIN21*, MtN21 nodulation gene.



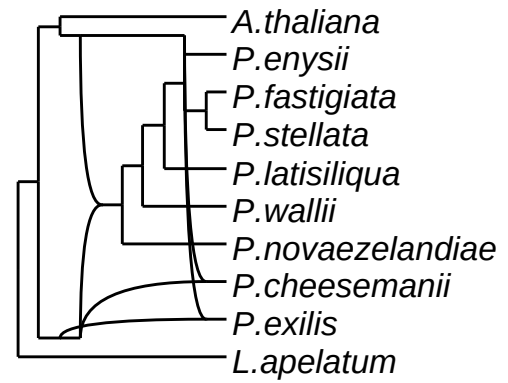
(a)



(b)



(c)



(d)

Figure S5: **Results on *Pachycladon*.** Species network reconstructed from the 10-taxon *Pachycladon* data set with five genes from Joly et al. (2009).

References

- Joly, S., P. B. Heenan, and P. J. Lockhart. 2009. A pleistocene inter-tribal allopolyploidization event precedes the species radiation of pachycladon (brassicaceae) in new zealand. *Molecular phylogenetics and evolution* 51:365–372.
- Nakhleh, L. 2010. A metric on the space of reduced phylogenetic networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)* 7:218–222.