Bayzid, Md. Shamsuzzoha; Warnow, Tandy

Gene tree parsimony for incomplete gene trees. (English) [Zbl 1443.92117]


Summary: Species tree estimation from gene trees can be complicated by gene duplication and loss, and “gene tree parsimony” (GTP) is one approach for estimating species trees from multiple gene trees. In its standard formulation, the objective is to find a species tree that minimizes the total number of gene duplications and losses with respect to the input set of gene trees. Although much is known about GTP, little is known about how to treat inputs containing some incomplete gene trees (i.e., gene trees lacking one or more of the species). We present new theory for GTP considering whether the incompleteness is due to gene birth and death (i.e., true biological loss) or taxon sampling, and present dynamic programming algorithms that can be used for an exact but exponential time solution for small numbers of taxa, or as a heuristic for larger numbers of taxa. We also prove that the “standard” calculations for duplications and losses exactly solve GTP when incompleteness results from taxon sampling, although they can be incorrect when incompleteness results from biological loss. The software for the DP algorithm is freely available as open source code at https://github.com/shamsbayzid/DynaDup.

For the entire collection see [Zbl 1372.68022].

MSC:
92D10 Genetics and epigenetics
92D15 Problems related to evolution

Keywords:
gene duplication and loss; gene tree parsimony; deep coalescence

Software:
DynaDup; DupTree; PhyloNet; GitHub; Genetree; iGTP

Full Text: DOI

References:


