

Abstract:

Phylogenetic networks are rooted, acyclic directed graph in which leaves are labelled with genes, genomes or species. They are used to model evolution with reticulate events. The reconstruction of phylogenetic networks is an important but challenging problem in phylogenetics and genome evolution, as the space of phylogenetic networks is vast and cannot be sampled well. One approach to the problem is to solve the minimum phylogenetic network problem, in which phylogenetic trees are first inferred, then the smallest phylogenetic network that displays all the trees is computed. A tree-child network is a phylogenetic network satisfying the condition that every non-leaf node has at least one child that is of indegree one. Here, we develop a new method that infers the minimum tree-child network by aligning lineage taxon strings in the phylogenetic trees, where we reduce the tree-child network inference problem to the shortest common supersequence problem. This algorithmic innovation enables us to get around the limitations of the existing programs for phylogenetic network inference. Our new program, named ALTS, is fast enough to infer a tree-child network with a large number of reticulations for a set of up to 100 phylogenetic trees with 50 taxa.