Chapter 8

QUARTET SUPERTREES

Raul Piaggio-Talice, J. Gordon Burleigh, and Oliver Eulenstein

Abstract: We introduce two supertree methods that produce unrooted supertrees from unrooted input trees. The methods assemble supertrees from a weighted quartet (four-taxon) tree representation of the input trees. The first method, QLI, extends Willson's local inconsistency quartet method to construct supertrees. This method, which was designed originally to produce a tree from a taxon-character matrix, is not well suited for building accurate supertrees when there is little taxonomic overlap among the input trees. The second method, QIL1, builds additionally on Willson's quartet-rectifying process and infers missing phylogenetic information from the input trees. We examined the effectiveness of the quartet-supertree methods using simulated and empirical data sets. These studies suggest that QIL1 is relatively accurate when compared with the matrix representation with parsimony (MRP) supertree method.

Keywords: phylogeny reconstruction; quartet method; supertree

1. Introduction

Almost all published supertrees (e.g., Bininda-Emonds et al., 1999; Liu et al., 2001; Daubin et al., 2002; Kennedy and Page, 2002; Salamin et al., 2002; Mahon, 2004) have relied on a single supertree construction method, matrix representation with parsimony (MRP; Baum, 1992, Ragan, 1992). We describe two new methods that assemble supertrees from a weighted quartet (four-taxon) tree representation of the input trees. The performance of both methods is compared with MRP in a simulation and an empirical study.

Although the optimality criteria and the approaches for combining quartet trees vary among different methods, quartet methods generally combine quartet trees obtained from a character data set. This allows the construction of a quartet tree for all or almost all possible four-taxon subsets of the taxa using optimization criteria such as maximum likelihood or maximum parsimony. By contrast, supertree methods combine input trees that often have little taxonomic overlap. Thus, quartet supertrees would have to be built from sets of quartet trees in which a large number of the possible four-taxon subsets of the joint taxa of the input trees often are not represented. Furthermore, when quartet trees are derived from input trees rather than from a character data set, a new mechanism to weight the quartet trees has to be devised based on their observed frequency in the input trees. We note that RadCon (Thorley and Page, 2000) implements quartet MRP, which uses quartet trees in a variant of the MRP supertree method, but no supertree method has used quartet methods explicitly to build supertrees to our knowledge.

In our quartet-supertree methods, we adopt Willson’s (1999) local-inconsistency quartet method. In our first approach, we apply Willson’s method directly to the weighted quartet trees obtained from the input trees. We refer to this approach as the Quartet Local Inconsistency (QLI) method. Because of missing quartet trees, however, the supertrees constructed by the QLI method tend to be highly inaccurate. Thus, in a second method, we implement an intermediate step to infer missing quartet trees based on Willson’s (2001) rectifying process for quartet trees obtained from character matrices. We refer to this extended approach as the Quartet Inference and Local Inconsistency (QILI) method.

To assess the accuracy of the QLI and QILI methods in contrast to the MRP method, a simulation and an empirical study were performed. The simulation study showed that the QILI method always performed similarly or better than QLI. It also showed that the inference step improved the performance of QILI over QLI greatly when there was lower taxon overlap among the input trees such that the number of missing quartets was higher. Both the empirical and simulation studies showed that the accuracy of the QILI method came close to that of the MRP method. Although the accuracy of the QILI method declined more steeply than that of the MRP method when overlap among the input trees was reduced, the QILI method still performed well when applied to empirical data.

We first provide necessary definitions and notations in Section 2. In Section 3, we survey Willson’s approaches and introduce the QLI and QILI supertree methods derived from them. The simulation and empirical studies are described in Section 4. Section 5 presents the results from these studies.