
DANIEL BROWN, University of Waterloo

Fast algorithms for phylogenetic reconstruction of aligned sequences

We present LSHtree and LSHplace, two fast methods for phylogenetics problems.

In LSHtree, we build phylogenies by starting with each taxon in its own tree, and merging trees (not necessarily at their roots, which is the practice in algorithms like Neighbour-Joining) until a single tree remains. The algorithm is sped up by using locality-sensitive hashing to find pairs of nearby sequences in the tree. We also approximately reconstruct the sequences at ancestral positions of the tree, to enable joins of trees where the proper edge is far from the leaves of the tree. The algorithm is extremely fast in theory, and can be shown to yield high-quality trees in Markov models of evolution. It also gives excellent trees in practice, and scales very well for large data sets.

We also present LSHplace, which adapts the idea of LSHtree to phylogenetic placement, giving a very fast method for putting new taxa onto an existing phylogeny. We discuss the implications of this work for applications such as metagenomics.

Both algorithms are joint work with PhD student Jakub Trzuskowski.