

1112-92-23

Elizabeth Allman, John Rhodes and Seth Sullivant*, smsulli2@ncsu.edu. *Statistically consistent k-mer methods for phylogenetic tree reconstruction.*

Algorithms based on k-mer distances are used to reconstruct phylogenetic trees without first constructing a multiple sequence alignment. We show that the standard method for reconstructing trees based on k-mer distances is statistically inconsistent (that is, they reconstruct the wrong tree even with increasing amounts of data) and we also derive statistically consistent model-based distance corrections in the case of sequences without gaps. We report on numerous simulations which show that the new formulas significantly out-perform older (statistically inconsistent) methods, even in sequences with gaps. These results also have implications for multiple sequences alignment, since many widely used multiple sequence alignment programs use the statistically inconsistent methods to construct a guide tree for multiple sequence alignment. (Received May 31, 2015)