1117-92-59 **Elizabeth Allman**, **John Rhodes** and **Seth Sullivant*** (smsulli2@ncsu.edu).

Statistically-consistent k-mer methods for phylogenetic tree reconstruction.

Frequencies of k-mers in sequences are sometimes used as a basis for inferring phylogenetic trees without first obtaining a multiple sequence alignment. We show that a standard approach of using the squared-Euclidean distance between k-mer vectors to approximate a tree metric can be statistically inconsistent. To remedy this, we derive model-based distance corrections for orthologous sequences without gaps, which lead to consistent tree inference. The identifiability of model parameters from k-mer frequencies is also studied. Finally, we report simulations showing the corrected distance out-performs many other k-mer methods, even when sequences are generated with an insertion and deletion process. These results have implications for multiple sequence alignment as well, since k-mer methods are usually the first step in constructing a guide tree for such algorithms. (Received December 29, 2015)