

***Bayesian Model-Based Alignment and Clustering of Terminal Restriction Fragment Length Polymorphism (T-RFLP) Profiles of 16S rRNA***

Zaid Abdo. Department of Mathematics, University of Idaho

Clustering is a popular method utilized in analyzing T-RFLP data that is used in studying some human microbial ecosystems (Zhou et al. 2007 for example). Clustering is used to identify microbial community kinds that share similar structure. Clustering T-RFLP data takes place in two steps: 1) similar operational taxonomic units (OTUs) between different samples are identified based on the similarity of associated fragment lengths (a step referred to as alignment or binning) and 2) different sample profiles are clustered based on the abundance of the identified OTU's in each sample. Both steps jointly affect the accuracy of the final clustering.

In this poster I present a new, model-based-Bayesian-simultaneous clustering and alignment approach to inferring the relationships between sampled T-RFLP profiles. I compare this method to a two stage method that first aligns using a naive Bayes approach and then clusters using T-BAPS (Tang et al. 2007) and to a distance-based clustering method. I utilize simulated data and well characterized real data in these comparisons.

Preliminary results indicate that model-based approaches are superior to distance-based ones in recovering the true clustering structure of the data. This is due to the improved efficiency of the model-based approaches in utilizing the information available in the data.

Supported by NIH Grant Number P20 RR016448 (COBRE Program of the National Center for Research Resources).