Jeremiah Yarmie - 7712088
MBIO 7160 - Special Problems in Microbiology: Advanced Biostatistics Final Project Proposal

Question:
Why may the phylogenetic approach used in the R package *treeWAS*, which is based on phylogenetics and hypothesis testing, be more appropriate for conducting bacterial genome-wide association studies (GWAS) when compared to other approaches that incorporate linear mixed models, clustering, and multivariate analysis/principal component analysis?

The authors of the tool claim that *treeWAS* outperforms these other approaches in both terms of not only statistical power, but also in the appropriateness of the approach when considering bacterial population structure and events of homologous recombination.

To explore this claim, *treeWAS* will be run on two data sets that have been previously analyzed and published using these alternative GWAS approaches: the first is a genome collection of *Helicobacter pylori*, a highly recombinant bacterial pathogen, and the second is a genome collection of *Mycobacterium tuberculosis*, a clonal bacterial pathogen that undergoes almost no recombination.

Both data sets containing associated phenotypic and clinical metadata required for GWAS, and the results of each analysis will be compared to the published findings of each study.

References:
Collins C, Didelot X (2018) A phylogenetic method to perform genome-wide association studies in microbes that accounts for population structure and recombination. PLoS Comput Biol 14(2): e1005958
Accessed from: https://journals.plos.org/ploscompbiol/article/file?id=10.1371/journal.pcbi.1005958&type=printable

Berthenet, E., Yahara, K., Thorell, K., Pascoe, B., Meric, G., Mikhail, J. M., … Sheppard, S. K. (2018). A GWAS on Helicobacter pylori strains points to genetic variants associated with gastric cancer risk. *BMC biology*, 16(1), 84. doi:10.1186/s12915-018-0550-3
Accessed from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6090961/pdf/12915_2018_Article_550.pdf

Farhat, M.R., Freschi, L., Calderon, R. *et al.* GWAS for quantitative resistance phenotypes in *Mycobacterium tuberculosis* reveals resistance genes and regulatory regions. *Nat Commun* **10,** 2128 (2019) doi:10.1038/s41467-019-10110-6
Accessed from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6513847/pdf/41467_2019_Article_10110.pdf