

Study Questions for “Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates” by Peter et al.

1. Considering the assumptions of linear models, why was it important to sample wild yeast strains rather than using lab strains?
2. The phenotype measurements did not follow a Normal distribution. How did the authors overcome this so that they could use a linear modeling framework for association tests?
3. The GWAS was performed using a linear mixed effects model. How is this similar to and different from the standard linear models we discussed in class? (you may want to look at reference 19)
4. Yeast and humans both have a missing heritability problem. What can this study teach us about human GWAS and missing heritability? Think about characteristics of *S. cerevisiae* genetic variation described in the paper and consider if each one is shared with humans or not.
5. The authors say that the result in Figure 6c may reflect differences in power between CNVs and SNVs. What three things determine power? Which of them do you think is different for CNVs versus SNVs?
6. **Challenge question:** How does the ADMIXTURE method used to generate Supplementary Figure 8 work? Think about the input data and what an ancestry component is. Are humans mosaics?
7. **Challenge question:** The authors identify 4,940 core open reading frames (ORFs) and 2,856 variable ORFs in the *S. cerevisiae* pangenome. Do humans have a pangenome with core and variable genes?