Project Proposal
W3500 Independent Biological Research

Kyelee Ruth Fitts

Fall 2017

GWAS (Genome-Wide Association Studies) have become increasingly important in the field of population genetics as a method of detecting polygenic adaptation, or adaption via traits that are affected by many different loci instead of just one as per traditional Mendelian genetics. GWAS have been used to detect polygenic adaptation in disease traits, for example detecting signals of selection in Type 2 diabetes, as well as a whole host of other quantitative traits that give valuable insight into how adapted genetic differences can be found, studied, and usefully applied [1] [4] . As this method of analysis becomes more and more crucial to analyzing such anthropometric traits, it is necessary to ensure that the underlying assumptions of GWAS and its methods remain as free from bias as possible to prevent false signals of selection.

One important model for studying polygenic traits is the additive model, where

$$y = \mu + A\alpha \tag{1}$$

This model refers to the trait value, y, of each SNP in an individual. μ is the average phenotype of the population. A refers to the parametrization of the genotypes and α is the average effect size of each allele on the phenotype. GWAS use robust statistical methods to obtain p-values that indicate whether a certain allele has a significant affect on the expression of a certain quantitative trait. However, GWAS assume that the average affect an allele has on a trait (α) is uncorrelated with the allele frequency over time of the trait– that is, uncorrelated with any dominance effects of the trait which could cause the allele frequency to shift over time or over different populations.

However, we know that under directional dominance, that assumption is not necessarily true. That is, if alleles that increase the effect size of a trait are dominant, then that allele will have a greater frequency and GWAS will overestimate the effect size– a bias that has the potential to result in false positive results.

To investigate this bias, I plan to work with Dr. Jeremy Berg in the Sella Lab, approaching this problem in two steps: first, to derive mathematically an expression that can quantify the bias due to directional dominance in GWAS given known expressions and concepts in population genetics. The second step would be to use the expression derived in step one to measure this bias using real data.

Height is an anthropometric trait for which many studies have found signals of selection using evidence from GWAS [5] . However, other studies have shown that height is also subject to directional dominance [3] – a combination that makes the trait well-suited for the purposes of my research. Data will come from the recent UK Biobank study, which has gathered genetic data on about 500,000 participants from the UK [2] .

Some progress has already been made on this project. In the spring of 2017 we found using the UK Biobank data further evidence of directional dominance in height. Over the summer, I worked with Dr. Berg to begin developing a mathematical expression for the bias. I hope this semester to make significant progress on what I believe is a fascinating project in mathematics and biology.

References

[1] Jeremy J. Berg. Polygenic Adaptation has Impacted Multiple Anthropometric Traits. *bioRxiv*, 2017.

[2] Clare Bycroft. Genome-wide genetic data on 500,000 UK Biobank participants. *bioRxiv*, 2017.

[3] Peter K Joshi. Directional dominance on stature and cognition in diverse human populations. *Nature*, 523:459–462, 2015.

[4] Fernando Racimo. Detecting polygenic adaptation in admixture graphs. *bioRxiv*, 2017.

[5] Michael C Turchin. Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nature Genetics*, 44:1015–1019, 2012.