November 21, 2020

Dear selection committee:

I am writing to apply to enroll in the Advanced Gene Mapping course for January 2021. I have studied statistical genetics and related subjects since 2007. I began with work in human genome-wide association studies at the University of Washington in 2007. For two years, I served as a member of both the data analysis team and the data quality control team. My colleagues developed an analytic pipeline for quality control measures for genotypic data in human GWAS. We also implemented these methods in a freely available R package. My specific roles included analyzing allele probe intensity data in efforts to infer sample mislabeling errors.

In September 2011, I enrolled in the statistics Ph.D. program at the University of Wisconsin-Madison. In addition to pursuing a rigorous course of study in theoretical statistics, I collaborated in a breast cancer GWAS analysis. We sought to characterize sample size requirements in post-GWAS studies. Our ultimate goal was to characterize breast cancer-related perturbations in biomolecular networks.

Following that project, I began work on my Ph.D. thesis research in complex trait genetics. I studied multivariate methods for mapping quantitative trait loci in multiparental populations of model organisms, like the Diversity Outbred and Collaborative Cross mice. These cohorts incorporate DNA from more than two founder lines. Proper analysis of data from these mice requires novel statistical methods and tools. I developed and characterized a test of pleiotropy against separate QTL for coincident univariate traits. When two traits univariately map to a single genomic region, one may wish to distinguish between the competing possibilities that 1) a single QTL affects both traits and 2) two distinct QTL, that happen to be near each other, each affect one trait. We implemented our methods in an R software package, `qtl2pleio` and reported our findings in journal articles.

Following completion of my statistics Ph.D., I accepted a postdoctoral research position at the University of Massachusetts Medical School. I started in September 2019. My current research investigates QTL mapping-related questions in genetically diverse mice and requires further development of applied statistical methods. One question involves the extent to which we may borrow information from a study in Collaborative Cross mice to inform QTL mapping in a similar study of Diversity Outbred mice. I'm currently studying strategies aimed at enhancing statistical power to detect tuberculosis-related QTL in Diversity Outbred mice by leveraging results from tuberculosis-related studies in Collaborative Cross mice. My leading approach, so far, involves using allelic series at Collaborative Cross mice QTL to simplify statistical models for mapping QTL in Diversity Outbred mice. Traditionally, one would fit a model like:

$$\text{Trait} = \sum_{i=1}^{8} p_i b_i + \text{error}$$

with p_i being the probability that a specified mouse has founder i allele at the genomic position of interest and b_i is the effect of founder allele i on the trait value.

In my strategy to enhance statistical power to detect QTL, I first use the Diversity Outbred mice to infer the allelic series (number of functional alleles, from one to eight, and assignment of founder alleles to functional alleles) at a genomic position, then use that allelic series to inform QTL scans of the same region in Collaborative Cross mice. In a setting where the allelic series inference reveals presence of two distinct functional alleles, with one functional allele belonging to founder 1 and the other functional allele belonging to the other seven lines, I fit this linear model in the Collaborative Cross mice:

$$\text{Trait} = p_A b_A + p_B b_B + \text{error}$$

where $p_A = p_1$ from the eight-allele model and $p_B = 1 - p_1$.

I anticipate that this strategy both will enhance detection of QTL affecting similar traits and will refine credible intervals for position of known QTL.

In future work, I'm eager to explore mediation analysis methods for molecular phenotypes in genetically diverse mice.

Thank you for considering my application.

Sincerely,

Fred Boehm