# SPECIFIC AIMS

Multiparental populations of mice have revolutionized systems genetics studies by incorporating diverse collections of genetic variants into mapping cohorts, enabling deep phenotyping, and enhancing mapping resolution and biological reproducibility. In particular, these populations are being used extensively to model and understand infectious diseases. While a great deal of data are available from a variety of model infections, analytical methods that maximize statistical power to detect genetic loci (quantitative trait loci, QTL) that affect these complex traits are still lacking.

Two complementary mapping populations, the Collaborative Cross and the Diversity Outbred mice, arise from the same eight inbred founder lines and therefore share the same collection of genetic variants. However, the two mapping cohorts differ in their genetic architecture and possess complementary strengths. the Collaborative Cross is a collection of recombinant inbred lines which allow experimental replication and serial phenotyping of the same genotypes, while the Diversity Outbred is a collection of highly heterozygous and genetically unique mice with small recombination intervals.

Both CC and DO populations have been used to model tuberculosis pathogenesis. Our group has performed deep phenotyping of CC mice after infection with Mycobacterium tuberculosis (Mtb), and our collaborators have quantified the susceptibly of a large panel of DO animals. In both cases, Mtb infection produces a wide range of phenotypic outcomes, which in some cases can be associated with QTL. However, the lack of concordance between these studies suggests that many more associations might be observed with more sensitive methods.

We propose to develop an analytical framework that will leverage the strengths of both CC and DO datasets together. By developing a strategy to infer allelic series at a QTL in one populations, we will be able to probe the second population with enhanced statistical power. This strategy will increase the resolution of mapping in the CC panel, and allow susceptibility traits from the DO to be associated with more specific phenoytpes in the CC.

**Specific Aim 1.** *METHODS: Design and implement new biostatistical methods for parametric bayesian inference of allelic series at established QTL in multiparental populations.*

Ideally, this is only 2-3 sentences. “current methods to define alleles associated with a QTL are suboptimal because… We will overcome this using a Bayesian framework to determine the most likely allelic series.

We define allelic series to be both the number of alleles and the allocation of each allele to each founder. Standard QTL mapping methods permit each founder to have its own distinct allele. However, this is suboptimal because many QTL have fewer than eight alleles represented among the eight founders. Thus, a model that allows for eight allelic effects parameters is fitting too many parameters and, consequentially, is less powerful compared to a model that fits the true number of alleles after accounting for allelic series.

To overcome this problem, we propose to develop Bayesian statistical methods to infer allelic series at detected QTL. Knowing the allelic series offers multiple scientific benefits. In followup studies in cohorts with the same set of alleles (Collaborative Cross and Diversity Outbred mice, for example), our approach offers more statistical power to replicate QTL detection.

**Specific Aim 2.** *APPLICATION: Apply allelic series inferences to complex traits from Diversity Outbred and Collaborative Cross mice.*

Again 3-4 sentences. We will optimize on simulated data. Then we will do a CC ->DO translation. Then we will do a DO-> CC translation.

We first will perform simulation studies to characterize QTL detection power enhancements with allelic series-informed QTL mapping. We’ll use Diversity Outbred and Collaborative Cross mice’s genotypes to simulate traits. Because the traits are simulated, we know their allelic series. We then perform two methods of QTL mapping. In the first, we map traits with the standard eight-allele models. Second, we use allelic series to restrict the number of parameters in our models. We then map the same simulated traits, with the restricted models to quantify the enhancement in statistical power.

We then will apply our methods to gene expression traits in Diversity Outbred and Collaborative Cross mice. In so doing, we will draw on strengths of both the Collaborative Cross (genetic reproducibility) and Diversity Outbred (high heterozygosity levels and small mean haplotype length) mice. Specifically, we’ll leverage the small average haplotype size in the Diversity Outbred mice to identify a narrow genomic region that affects a trait. For each QTL in the Diversity Outbred, we’ll use our methods from Aim 1 to infer allelic series. Because the Collaborative Cross and the Diversity Outbred mice share the same set of founder alleles, we can transfer knowledge about QTL between the two cohorts. If we know the allelic series for a single QTL in the Diversity Outbred mice, then we automatically know the allelic series for the same QTL in the Collaborative Cross. We thus may use the allelic series that we inferred in analyses of Diversity Outbred mice to gain power when mapping QTL in Collaborative Cross mice.

A second area of application is in multivariate QTL mapping and tests of pleiotropy vs. separate QTL. We previously developed a multivariate mapping strategy that allows for eight alleles at every locus. Allelic series inferences for each of the univariate traits in a multivariate analysis would enable gains in statistical power to distinguish separate QTL from pleiotropy.

**Specific Aim 3.** *SOFTWARE: Create and share an open-source software package that implements our methods.*

We will create and share a user-friendly, open-source R software package and share it via Github. The R package will contain both functions that execute our new methods and code to reproduce all analyses from Specific Aim 2. Additionally, we will create a website that hosts and documents our software. We can achieve this feat with the pkgdown R package. To encourage durability of our software and methods, we’ll use Docker to execute our analyses and test our code. We will share the Docker images via Dockerhub.