|  |  |  |
| --- | --- | --- |
| **Meeting Summary** | **Project Name:** | Quantitative Trait Locus (QTL) Mapping of Peas |
| **Date:** | Wednesday, October 12, 2016 |
| **Time:** | 3:00-4:45pm |
| **Attendees:** | Client: Jamin Smitchger  Consultants: Andrea Mack, Paul Harmon and Nnamdi Ezike | |
| **Meeting Summary** | | |
| **About the Project:**  Quantitative trait locus (QTL) analysis is a statistical method that links two types of information—phenotypic data (trait measurements) and genotypic data (usually molecular markers)—in an attempt to link variation in genetic bases associated with of variation in expression of complex traits such that an inheritance pattern in progeny can predict where genes are located.   * The research focus on collection of data on Field Pea in Montana. Data were collected over a 5-year **I have this written as 6 years, please verify** period in Bozeman and 2-year in. There are 255 progeny + 2 parents, resulting in 50,000 data points of phenotypic data on 16 quantitative traits thought to be associated with lodging such as height, stem width, maturity time. There are also 1.2 million data points of genotypic data most of which Jamin thought to be junk. Genetic data may be junk because it is too similar to the reference or because it is associated with a phenotype that is not expressed. * Jamin is interested in the lodging of peas, which is a measure stalk strength.   . **With each parent contributing to different DNA sequence to its progeny**   * + **Genotypic data indicates what allele of a gene has been inherited from which parent.**   **Statistical Analysis Needed**   * Part 1: Assess correlations between phenotypic data   + Jamin needs basic help with correct interpretation of correlation and would like the consultants to assist in a more visually understandable correlation plot than the 17x17 he has already made   + Jamin would like a more indepth explanation of the difference between R2 and R   + Jamin was unsure of how to account for correlations in phenotypic traits in a QTL analysis   + QTL is not generally done on two traits that are highly correlated to reduce the number of tests done and increase power to detect genetic variation   + Thresholds for *highly correlated* will need to be determined, one source suggests the general cut off to be 0.5 in human genetic data * Part 2: QTL mapping to find evidence of associations between phenotypic and genetic, for each trait that is not highly correlated with others   + Standard quality protocol     - Assess plant quality - - within each plant there should be an approximately uniform distribution of base pairs     - Assess Hardy-Weinberg equilibrium -- random mating, uniform distribution of base pairs in similar locations across the entire population     - Allele frequency > 5%   + QTL   + Choosing appropriate experimental error rate (because roughly 500 tests are performed)     - False discovery rate assessment     - Type I error rate can also be assessed through simulation studies, but this is less common   + Interpretation * Part 3: Based on the phenotypic traits which exhibit variation in expression is most highly associated with genetic variation, fit a mixed model to answer which loci is the best predictor of lodging (gblups) * Note that Part 3 was not explicitly discussed in the meeting, but is a common direction to go following a QTL analysis, further discussion with Jamin is need to decide if this is what his end goal is.   **Study Considerations**   * DNA sequence determines characteristics * Meiosis reduces number of chromosomes in the parent cell and produces gamete cells combined from 2 individuals. Pea has 7 different chromosomes * Genes located close together on the same chromosome are more likely to be inherited together * Inheritance pattern of genes are further analyzed in hundreds of progeny * Peas are self pollinating (homozygous), which can be thought of as getting each chromosome twice from the parent * Genes assessed in this study are random, opposed to targeted * Software can result in bad calls, which is where a base pair cannot be distinguished, resulting in missing data (Map Disto records as "H") * **This part is unclear to me but is important:** Data (which -- genetic?, I thought this was recorded as base pairs?) are recorded in centimorgans (cMs) which represents the distance loci, adjacent groups of base pairs, are from the end of the chromosome * cM's are the measurements used to assess linkage, or how likely traits are to be inherited together, for example, genes that are associated with hair and eye color are often inherited together, meaning that they have high linkage * Genetic markers are not independent of each other – develop/utilize methodology to handle this lack on independence * The consultants will consider analyzing QTL Mapping in different programs from ‘*Map Disto*”, including R package “RQTL” or QTL Cartographer, and bioconductor.org * Software associated with bioconductor.org will most likely be used to replicate Jamin's QTL mapping   **Progress Made on Analysis**   * Client used correlation matrix to assess the relationship of the explanatory variables for the phenotypic data (using R package “Psych”) * Client has already logged genotypic data collected into a program called “Map Disto” – a software for genotyping dataset embedded in Microsoft Excel   **Challenges faced by Client**   * Still collecting data – expected to collect all required data in another month * Organizing data gathered * Constrained with what statistical methods to use – QTL relies on a series of F tests **Why is this a constraint?**   **Questions for Jamin**  -> What is considered *high* correlation in traits for peas?  -> Are you doing a QTL for all phenotypic traits (that aren't highly correlated) or only for lodging?  -> Please explain the study design and environmental variation.  Jamin plans to graduate in Spring 2017. | | |