COVER SHEET

TITLE: Fine mapping for domestication traits using RC-NILs

AUTHOR'S NAME: Craig DeValk

MAJOR: Biochemistry

DEPARTMENT: Biochemistry

MENTOR: Dr. John Doebley

DEPARTMENT: Genetics

MENTOR(2): Ali York

DEPARTMENT(2): Genetics

YEAR: 2017

(The following statement must be included if you want your paper included in the library's electronic repository.)

The author hereby grants to University of Wisconsin-Madison the permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part in any medium now known or hereafter created.

ABSTRACT

Fine mapping for domestication traits using recombinant chromosome nearly isogenic lines

(Abstract content here not to exceed 150 words)

In past work the Doebley Lab used maize-teosinte hybrid recombinant inbred lines (RILs) to find various quantitative trait loci (QTL) for ear size in order to study the process of domestication. Of particular interest was a significant large effect QTL on chromosome 5 for ear diameter (Shannon, 2013). A different set of maize-teosinte lines than was used initially narrowed down that same QTL to a 2.654 Mbp region (Lemmon and Doebley, 2014). Recombinant chromosome nearly isogenic lines (RC-NILs) were created that were meant to segregate only for the region of interest. Both phenotypic and genotypic data was collected for the RC-NILs and results did not show a clear region of segregation between kernels per rank for maize and teosinte. Future work will utilize a subset of the RC-NILs that did not segregate on chromosome 7 and hopefully determine the gene responsible for the large effect QTL on chromosome 5.

Craig DeValk/Biochemistry
Author Name/Major

John Doebley
Mentor Name/Department

Craig DeValk
Author Signature

John Doebley
Mentor Signature

4-28-17
Date

Fine mapping for domestication traits using recombinant chromosome nearly isogenic lines

Craig DeValk, Alessandra York, and John Doebley

Abstract

There are many morphological differences between maize (Z*ea mays* ssp. *mays*) and its evolutionary ancestor teosinte (*Zea mays* ssp. *parviglumis*). Some of these traits have already been characterized at the genetic level which has provided valuable insights into the domestication process but at this point differences in ear structure have not been fully explained (Clark *et al*., 2006, Wang *et al.*, 2005 Wills *et al.*, 2013, Yang *et al.*, 2016). In past work the Doebley Lab used maize-teosinte hybrid recombinant inbred lines (RILs) to find various quantitative trait loci (QTL) for ear size. Of particular interest was a significant large effect QTL on chromosome 5 for ear diameter which was found using maize-teosinte BC_2S_3 lines (Shannon, 2013). This QTL also co-localized to a second large effect QTL for kernel row number. Further work was done with a different set of maize teosinte BC_6S_6 lines and end results narrowed down that same QTL to a 2.654 Mbp region (Lemmon and Doebley, 2014). Markers were used to find recombination points in the region of interest on the short arm of chromosome 5. With this information new recombinant chromosome nearly isogenic lines (RC-NILs) were created that were meant to be homozygous at every point except for the across the region of interest. The RC-NILs were made with two separate lines, b11b and b46, that were predicted to solely segregate in the region of the QTL chromosome 5. Both phenotypic and genotypic data was collected for the RC-NILs and results did not show a clear region of segregation between kernels per rank for maize and teosinte. It is believed that this is due to an unanticipated region on chromosome 7 that appeared to be segregating for all lines. Future work will utilize a subset of the RC-NILs that did not segregate on chromosome 7 to repeat the experiment and hopefully determine the gene or transcription factor responsible for the large effect QTL on chromosome 5.

Introduction

Evolution as a concept dates back to Charles Darwin and his work, *On the Origin of Species*. It was here that a respected scientist first published a theory to the general public that explained the diversity on Earth (Darwin, 1859). Since this time, knowledge of the topic has expanded exponentially and many new methods have been used to help improve its understanding. In particular, domestication has proved to be helpful in allowing researchers to learn about evolution. Domestication is the study of human directed evolution. In practice domestication is thought of as a type of evolution though, with the key difference being that domestication is the result of humans selecting for better traits consciously and therefore has the ability to cause more drastic changes quickly (Diamond, 2002). This can be found in many plants and animals from corn or rice, to dogs.

The Doebley Lab uses maize and its ancestor, teosinte, to study the process of domestication. Maize is now known to have been first domesticated around 10,000 years ago in the Balsas River Valley region of Mexico from a local form of teosinte, but this theory was at one time controversial (Doebley and Stec, 1991). George Beadle, the same man who proposed the now famous one gene, one enzyme hypothesis, was also one of the first people to suggest that maize evolved from teosinte. However, for much of his life people did not believe this because the morphological differences between the two species were considered too numerous to have evolved in only a few thousand years. It was only after he retired and began to devote significant time towards this idea did he prove that indeed teosinte was the ancestor of maize (Beadle, 1972). This led to the investigation of factors controlling the domestication of maize, many of which appeared to be traits of complex inheritance in five main regions of the maize genome (Beadle, 1939).

Complex or quantitative traits are those in which many genes or environmental factors are responsible for controlling a particular phenotype. The best way to analyze complex traits historically is through quantitative trait loci (QTL) mapping, and more recently genome-wide association studies (GWAS). A QTL is a section of a chromosome that is determined to be the cause of trait variation or in this case domestication trait variation (Grisel, 2000). Take for example the gene *teosinte branched 1 (tb1)*. This gene, which was first identified by the Doebley Lab using QTL mapping, controls for the branching in maize and teosinte by repressing

outgrowth via repression of the cell cycle. It also can control for other traits throughout the organism such as inflorescence sex and lateral branch length (Doebley *et al.*, 1995). Through many QTL mapping studies, researchers have discovered that transcription factors such as *tb1* are pivotal in crop domestication (Doebley *et al.*, 2006). This makes sense because transcription factors have a more versatile role altering downstream gene expression.

While many important maize domestication genes have been identified, none have been found for the change in ear size. During domestication, ancestral people would have wanted to select for larger ears due to ease of harvestability and increase in yield. Therefore, this has been of great interest to those studying maize domestication. To look for QTL for changes in ear size, phenotypes are collected for traits such as the diameter of the ear, the length of the ear, how many kernel rows, or how ranks of kernels an ear has. Doebley and Stec first identified an ear size QTL on chromosome five in 1991 and reaffirmed it in other populations (Doebley and Stec, 1993). Many years later, Shannon found the same results when mapping for ear diameter in her domestication QTL mapping study, with the most significant QTL have a LOD score of 144.43 in a 6.85 Mbp region on chromosome five (Shannon, 2013) (Figure 1).

Attempts to fine-map the QTL Shannon used failed, which was likely because the lines were too heterogeneous. This caused segregation for multiple small effect QTL distributed throughout the genome. Additionally there appears to be multiple linked QTL in this region with one of these having a greater effect than the others. However, the QTL for ear size was later confirmed by Lemmon and Doebley to a more narrow interval of 2.654 Mbp using a set of more isogenic lines (BC_6S_6) (Lemmon and Doebley, 2014) (Figure 2). Within this region, QTLs for ear diameter, kernel row number and kernels per rank all overlap. It is with these lines that the fine-mapping study was repeated due to the ability to segregate only for the region where the small QTL for ear size is located because it alleviates the issues encountered before when fine-mapping.

Methods

Development of Lines: A set of recombinant chromosome nearly isogenic lines (RC-NILs) was developed by the lab to create breakpoints within the region of interest to fine-map the QTL. They started with a set of W22 maize lines used by Lemmon and Doebley in 2014 that were believed to be homozygous throughout except for the short arm of chromosome five. Lines were then chosen that only segregated for the 2.654 Mbp region where the three ear size QTL were. These lines were crossed to generate recombination and in 2015, 3200 F2 plants from this line were grown and then screened. Plants that were heterozygous for a recombination point in the region of interest were selfed to create the F3 generation. The following generation was screened for individuals that were homozygous recombinants and from this a set of 163 RC-NILs was obtained. This whole process is highlighted in Figure 3.

Phenotypic Analysis and Fine-Mapping: To find where each line's breakpoint was located, DNA was extracted from leaf tissue for each of the 163 recombinant lines (Doebley and Stec, 1991). The DNA was then sent in for sequencing using the Genotype-by-Sequencing method (Elshire et al., 2011).

All 163 RC-NILs were grown in a randomized complete block design which contained six blocks. The block could affect the variation in the phenotype and the way the locations of each plant in a block were assigned was random. This randomness is accounted for in the model below. Phenotypes for kernels per rank were measured by counting the number of kernels vertically per ear. Additionally kernel row number was measured by counting the number of kernels around the ear. Ear diameter was measured in centimeters using a caliper towards the middle of each ear. This data was analyzed with a mixed linear model in the software SAS (Version 13.1) according to the following parameters to generate least squared means (LSMs).

$$Y_{ij} = \mu + a_i + b_j + e_{ij}$$

The data was fit using least square means calculations. Y_{ij} is the phenotypic trait value (in this case that is for kernels per rank), μ is the mean, a_i is the line effect, b_j is the block effect, and e_{ij} is the experimental error. The reason for a mixed linear model is because of the experiments design, which has both random and fixed effects. The line effect is the fixed effect because it

was expected that each line would have an individual-specific effect on the kernels per rank of each ear.

Results

Phenotypic Analysis: This analysis, shown in Figure 4, compares the phenotypic values for kernels per rank to their genetic counterparts for each species in a more visual way. On the right hand portion of the figure all of the RC-NILs are organized from smallest to largest. On the left side, presences of teosinte (green), maize (yellow), or heterozygous (grey) markers are displayed. Ideally the phenotypes would have segregated into two distinct classes when sorted from smallest to largest with the smaller phenotype class corresponding to teosinte and the larger class corresponding to maize. This, when compared to genotypes that should have been more uniform in where their breakpoints occurred, would preferably have pointed to a narrow region where the gene or transcription factor in question was. This region would have identified the causual interval for the QTL. Results for ear diameter and kernel row number are not shown but had a similar lack of uniformity as well as the same continuous distribution in the phenotype.

QTL mapping: Since the causal interval did not fall within the original region of interest, QTL mapping was performed to see where the causal region appears to be located. Using R/qtl all plots were generated showing all significant QTL for each of the three populations (Figures 5A and 5B). Populations b11b and b46 were used to derive the RC-NILs. Much of the analysis was done independently with these populations because each population was fixed for different genotypes. Therefore it is possible that doing a combined analysis would lead to unexpected and biased results.

With this being said, an additional analysis was done using the combined data of both populations as this could show any larger trends. Table 1 is provided to show this data in text form. From here it is easy to see the physical location for various QTL as well as the lod scores. All of the QTL found are on chromosomes 5 and 7 with lod scores going up to a maximum of 8.164. The QTL found on chromosome 7 indicate a problem with the analysis since everything other than the region of interest on chromosome 5 was supposed to be controlled for and non-

segregating. QTL maps were also found for ear diameter and kernel row number with both also containing an unanticipated QTL on chromosome 7.

Discussion

Some interesting results were obtained during this study. The first of which is the presence of a shadow peak in all of the QTL maps in Figure 5. On chromosome 5 there are always two large peaks with one of the two peaks not regarded as significant by R/qtl. The reason for the presence of the second peak is due to the way in which the lines were developed. Looking at the marker analysis in Figure 4 it is possible to see the inverse relationship between presence of maize and presence of teosinte in the region of interest. Because QTL studies are simply statistical analysis meant to find significant correlations in data when a pattern like this appears, a shadow peak will appear.

A second area of interest is the results depicted in Figure 4. It is readily apparent that the phenotypic values did not segment into two distinct classes. Instead, a continuous distribution of values was found. This means narrowing down the region to a smaller QTL is very difficult. An ideal set of data is shown in Figure 6. Notice how in this analysis larger phenotypes entirely correspond to maize genotypes while smaller ones correspond to teosinte in the highlighted black box. The proposed reason for the differences between Figure 4 and Figure 6 is the unanticipated presence heterozygosity on chromosome 7 (Figure 7). This is of importance because, as shown in Figure 1, traits controlling for ear size, similar to kernels per rank, have been found to have QTL here. An added QTL adds complexity to the analysis due to possible interactions among multiple varying genes. Furthermore this inhibits the ability of the data to segregate as nicely as was hoped.

One interesting thing to note about the data collected in this study is that the results for kernels per rank looked the most promising of all three traits collected. One explanation for this different pattern could be that the analyses done by Alessandra York for ear diameter and kernels per row are looking at a different gene, or genes, than what is controlling for kernels per rank. These genes would both still be in the same region initially identified by the lab in 2013 but simply be different genetic structures. One hypothesis for this is, since all three genes control for

related functions and are in a similar area, that the different genes are a result of gene duplication and subsequent modification by evolution (Lynch and Conery, 2000). Duplication is a very common route by which new genes are formed in Eukaryotic cells and this would fall in line with the early results found by Shannon in 2013. If it is in fact determined that multiple genes are controlling for the different traits further analysis could be done to determine if they are the result of homology or some other evolutionary function.

Future Directions

The next steps in this analysis will be to find a way around this issue of chromosome 7. To do this a small subset of the current RC-NILs will be selected. This subset will be ones that did not segregate on chromosome 7 and therefore should not result in the presence of unwanted QTL being found. This will eliminate interactions and will hopefully allow for the phenotypes to segregate into two classes which will allow for narrowing down of the region of interest. This experiment will be repeated again with similar methods on these lines. It is possible after all of this analysis that the gene for kernels per rank may fall in a different range than the genes for ear diameter and kernels per row. This would lead to further research to elucidate the differences between the multiple genes and possibly determine if there are any interactions among them.

References

Beadle G. 1939. Teosinte and the Origin of Maize. *Journal of Heredity*. 30: 245-247.

Beadle G. 1972. The Mystery of Maize. *Fields Museum of Natural History Bulletin*. 43: 2-11.

Darwin C. 1859. *On the Origin of Species*. John Murray.

Diamond J. 2002. Evolution, consequences and future of plant and animal domestication. *Nature*. 418: 700-707.

Doebley J, Stec A. 1991. Genetic Analysis of the Morphological Differences between Maize and Teosinte. *Genetics*. 129: 285-295.

Doebley J, Stec A. 1993. Inheritance of the Morphological Differences Between Maize and Teosinte: Comparison of Results for Two F_2 Populations. *Genetics*. 134: 559-570.

Doebley J, Gaut S. Smith D. 2006. The Molecular Genetics of Crop Domestication. *Cell*. 127: 1309-1321.

Doebley J, Stec A, Gustas A. 1995. Teosinte branced1 and the origin of maize: evidence for epistasis and the evolution of dominance. *Genetics*. 141: 333-346

Grisel J. 2000. Quantitative Trait Locus Analysis. *Alcohol Research and Health*. 24: 169-174.

Lemmon ZH, Doebley JF. 2014. Genetic Dissection of a Genomic Region with Pleiotropic Effects on Domestication Traits in Maize Reveals Multiple Linked QTL. *Genetics*. 198:345-353.

Lynch M, Conery J. 2000. The Evolutionary Fate and Consequences of Duplicate Genes. *Science*. 290: 1151-1155.

SAS. *Computer software*. Vers. 13.1.

Figures

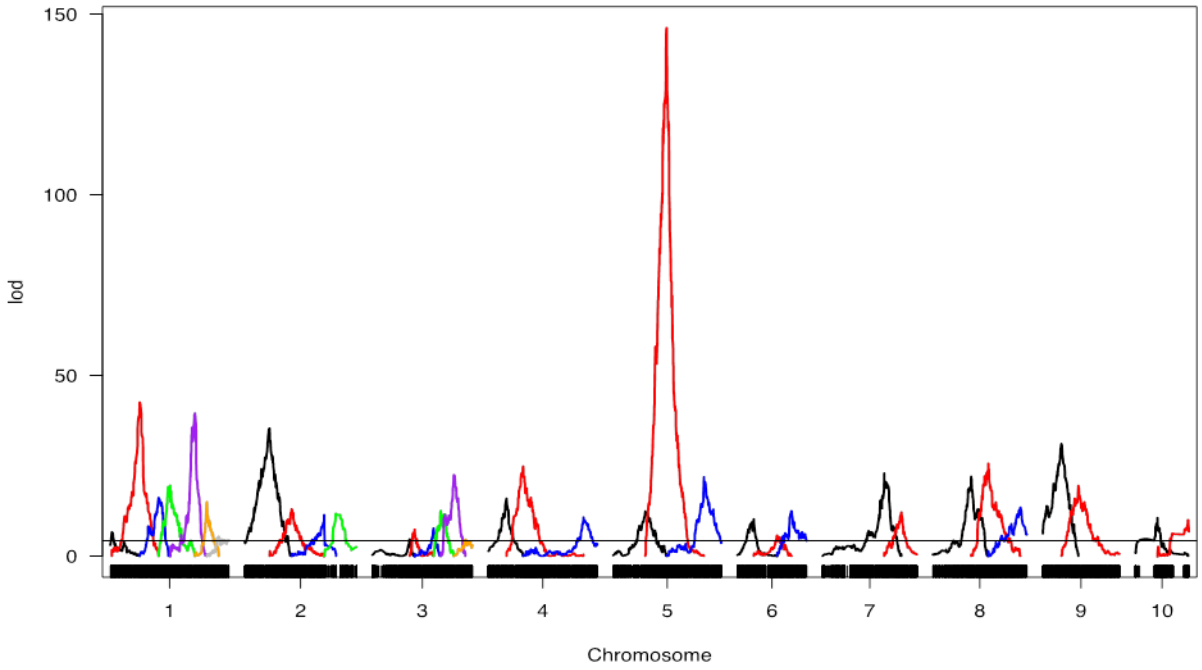

Figure 1: Shannon's QTL mapping for ear diameter in the BC253 population. The large spike on chromosome 5 has a lod score of 144.43. Lod scores are a measure of statistical significance with larger lod scores indicating more significance. This QTL was the largest one found in

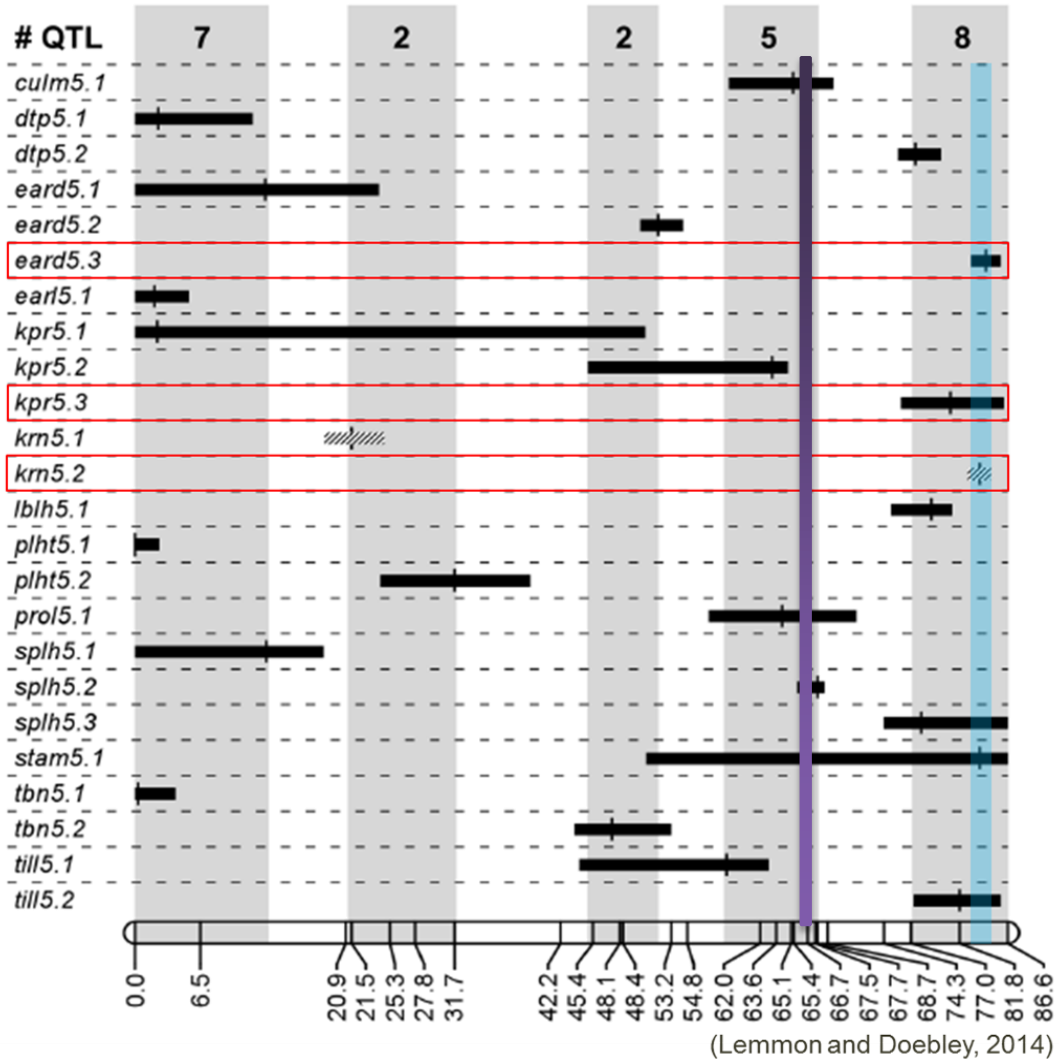

Figure 2: A map of a portion of chromosome 5 labeled in centimorgans (cM) at the bottom. The darker shaded regions labeled with numbers at the top are areas of high gene frequency and the purple region is the centromere. Shannon's data gave the QTL for *eard5.3* and this region overlapped heavily with QTL for kernel row number (*krn5.2*) and kernels per rank (*kpr5.3*). This narrowed down the QTL for kernel row number to a smaller 4.81 cM region shown in light blue (Edited from Lemmon and Doebley, 2014).

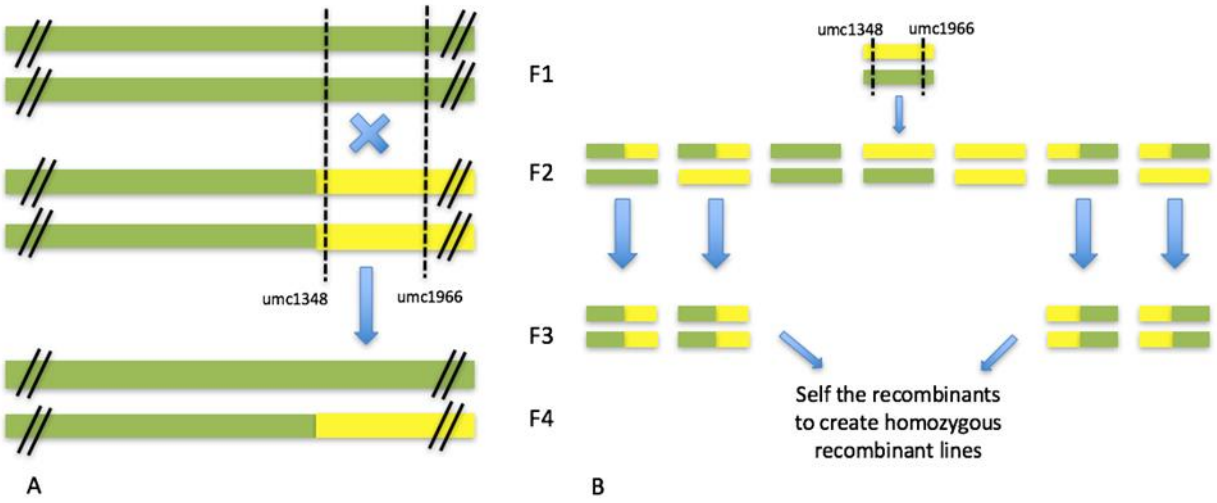

Figure 3A and 3B: In A, both lines are homozygous for teosinte throughout the short arm of chromosome five but begin to differ for maize or teosinte in the region enclosed by the markers umc1348 and umc1966. When crossed they can generate a heterozygous chromosome in this region if a crossover event occurs. In B, a heterozygous chromosome has already been generated and as it is selfed for three generations to create the RC-NILs.

Figure 4: Graphical representation of the phenotypic and genotypic data for the b46 population. The left region boundary is the marker umc1348 and the right boundary is the marker umc1946 as referenced in Figure 3. On the right green spaces are markers that were genotyped homozygous teosinte, yellow spaces are homozygous maize, grey spaces are heterozygous for maize and teosinte, and blank spaces are where genotyping failed. All of the lines were sorted by their kernel per rank value from smallest to largest as shown on the right. Notice the lack of segregation.

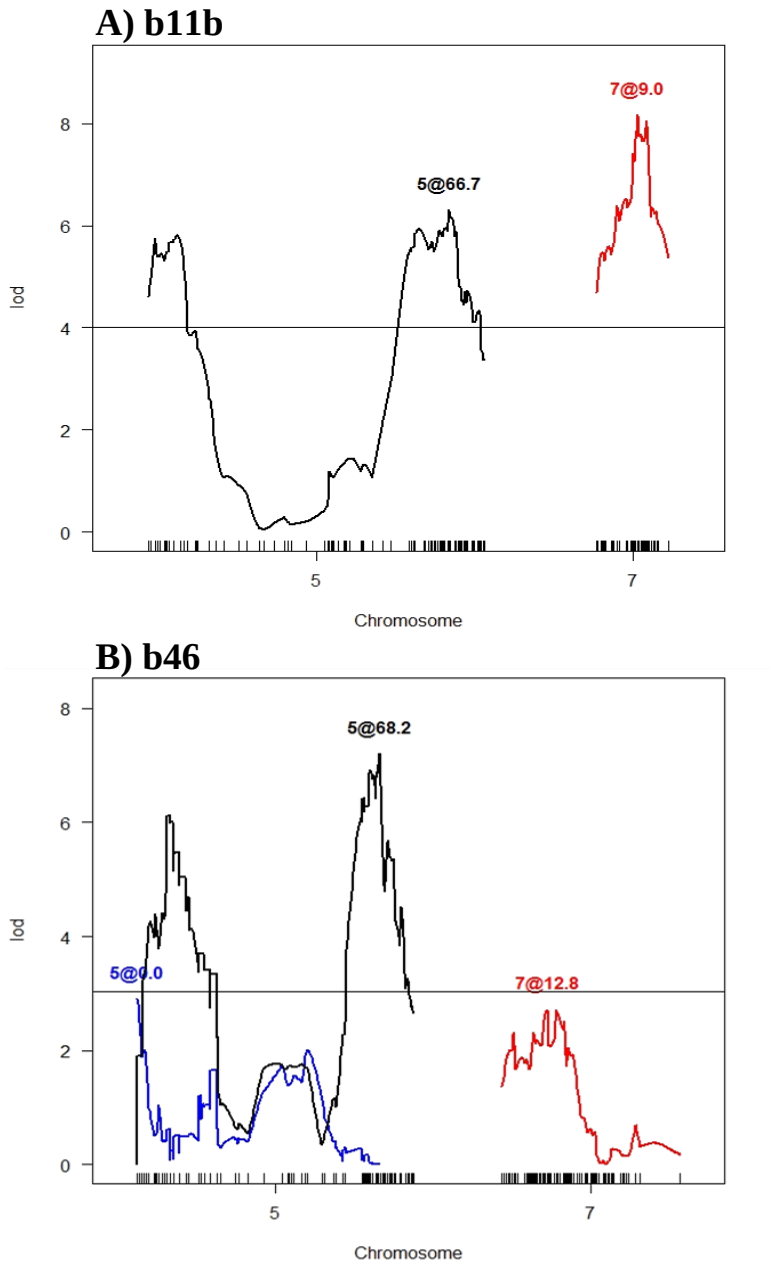

Figure 5A and 5B: QTL maps for analysis A), b11b B) b46. The minimum lod score threshold is shown by the horizontal line in each graph. Additionally all QTL found are labeled with the chromosome they were found on and their position in cM. Different colors on each respective graph are used to differentiate among QTL. Notice the presence of QTL on chromosome 7.

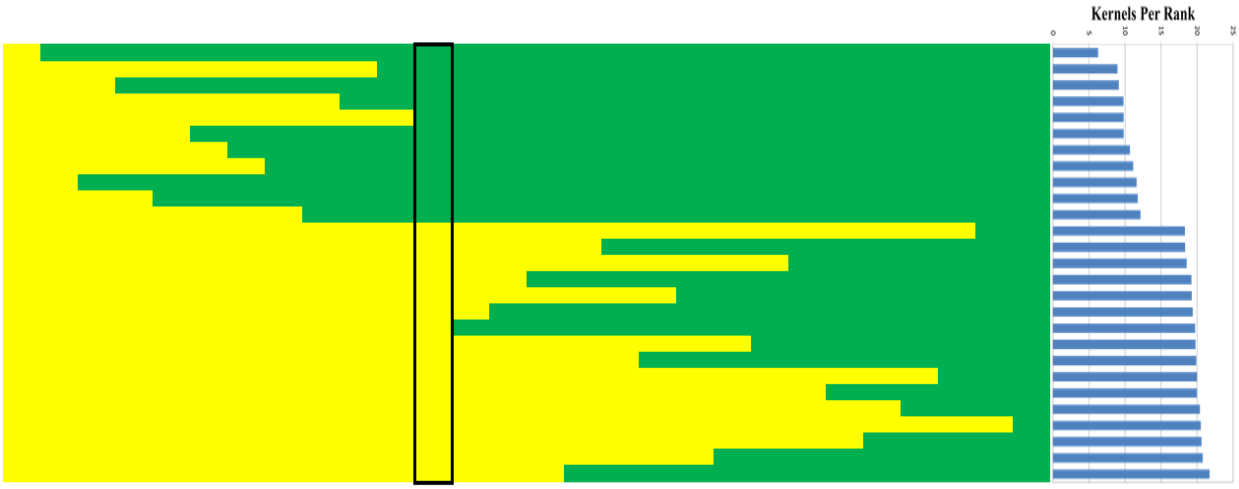

Figure 6: An ideal result for the analysis done in Figure 4. Again green represents markers genotyped to be teosinte and yellow is maize. Notice the two distinct phenotypic classes on the right. In the black box all of the genotypes for the smaller class are teosinte and for the larger glass all of the genotypes are maize. This would indicate the portion of the genome responsible for the QTL of interest is immediately upstream of the marker in the box.

Figure 7: Graphical representation of segregation in chromosome 7. Again maize is yellow and green is teosinte. Notice the breakpoints which create unwanted complexity in the analysis.

Table 1: All significant QTL found in analysis A) b11b and B) b46. The range of the QTL is given by its physical position in base pairs. The peak lod score, which can also be found in Figure 5, indicates to some extent the certainty of the result. Percent variance that the QTL is predicted to control tiller number for is shown. "***" indicates that the QTL interacts with other QTL at the level of $p < 0.0001$.

A) b11b

Chr.	Position (Mbp)	LOD Score	% Variance
5***	163.420 - 176.910	5.956	18.9
7***	148.326 - 154.608	7.855	26.1

B) b46

Chr.	Position (Mbp)	LOD Score	% Variance
5***	147.541 - 172.951	6.873	33.493