

Quantitative Trait Loci Mapping Part III: Linkage Analysis of our Quantitative Trait using WebQTL

Based on a module developed by William Grisham and Natalie A. Schottler at the University of California, Los Angeles (mdcune.psych.ucla.edu). Information on GeneNetwork was obtained from their website. Creative Commons Copyright 2009

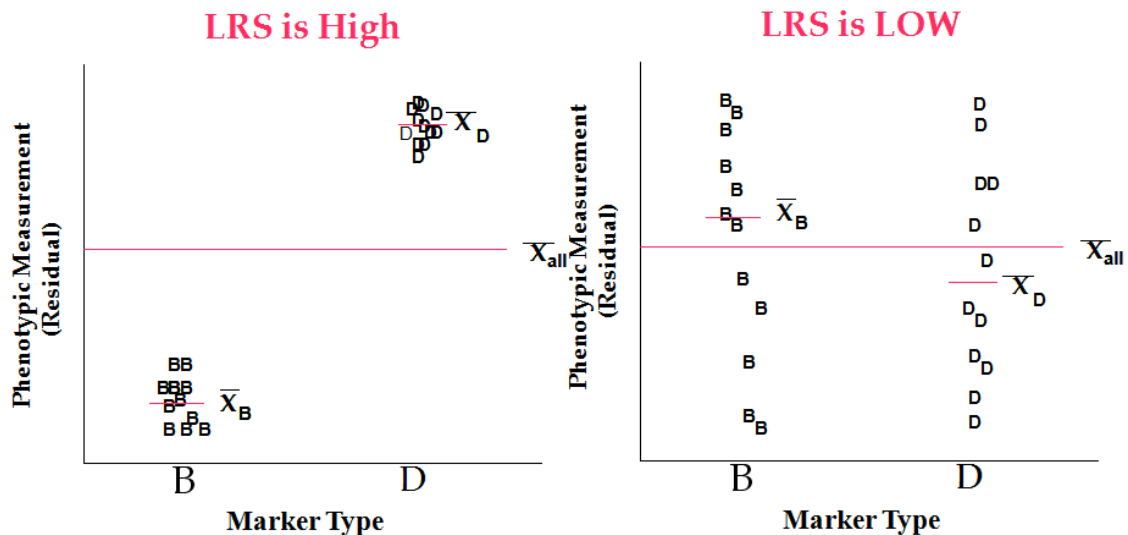
- *Pierce, Page 665-666 has a nice basic overview of QTL Mapping*

We will be using a genetic data set that is stored at the GeneNetwork website to map our QTLs. GeneNetwork consists of linked resources and analysis tools for systems genetics. It has been designed for the analysis of networks of genes, transcripts, and classic phenotype data sets. GeneNetwork combines more than 25 years of legacy data generated by hundreds of scientists with genome sequence data and massive transcriptome data sets. WebQTL is a major module of GeneNetwork that is optimized for fast on-line quantitative trait locus (QTL) analysis of traits that are controlled by combinations of gene variants and environmental factors. WebQTL exploits reference populations of humans (CEPH), mice (BXD, AXB, LXS, etc.), rats (HXB), *Drosophila*, barley, and *Arabidopsis*. Each of these genetic reference populations is accompanied by dense genetic maps used to locate modifier genes that cause downstream differences in expression and phenotypes, including disease susceptibility. We will be using their mouse BXD genetic data set to identify genetic variation associated with our trait.

Significant linkage between our trait variation and genetic variation between the B and D strains is determined through calculation of a Likelihood Ratio Statistic (discussed below) by means of a permutation test. Briefly, a distribution is determined by randomly scrambling and reassigning the residual phenotype to the various markers and then calculating the highest LRS over 1000 replications. A distribution is constructed from these random draws and then the genome-wide alpha level is determined from this distribution. A large number of LRSs are calculated (one for each marker, which at this writing is about 3500 informative markers). Using permutation tests, you may not all get the exact same values for the significant and suggested criteria lines—it depends on the

“draw” that each each of you gets in the random reassignment done by the algorithm. These differences probably won’t matter much unless the calculated LRS value is teetering right on the edge of “suggested,” in which case some of you will generate a much longer/shorter list of markers reaching the criterion of “suggested,” than others.

The Likelihood Ratio Statistic or LRS. As stated above, the LRS is performed at each interval marker. A great discussion of the LRS can be found in Beatty and Laughlin (2006). The images below display what the data would look like when the LRS is high and low:



Here the marker type refers to whether the individuals had the form of the DNA (same allele) that was characteristic of the B or D strain at that particular point on that particular chromosome. The individual Bs and Ds represent data points from individual RISs. Basically, when the LRS is high, all the phenotype (or its residual) of individual RISs with a given marker at that point in the chromosome will be clustered around their group mean, which will be distant from the cluster of phenotype of individuals with the other marker type. Both clusters will be distant from the mean of all (the grand mean). If the LRS is high, then that marker is likely at or near a gene(s) that have an impact on the phenotype.

In contrast, if the LRS is low, the phenotype of individual RISs will be scattered so that the variation from their individual group means is about the same as the variation away from the grand mean. This would indicate that this portion of the chromosome is not likely to have genes that had a big impact on the phenotype of interest. (The LRS can

be converted to a LOD score, which is another method of determining linkage that we have discussed.)

What constitutes a significant peak? Although at first blush it should seem obvious that there is a peak on the interval map that you will generate, when one zooms in on a single chromosome, one can see that it is not so obvious and needs some sort of operational definition. We define peaks as being a drop of 4.6 in the LRS from the maximum, which is approximately 1 log unit or a factor of 10. You will zoom into the graph further by selecting a range of megabases in which encompasses the peaks that are highest. In your zoomed in view, you can see the SNP track (single nucleotide polymorphism track). In this analysis we are interested in identifying the genetic variation that contributes to our olfactory bulb volume phenotype. To find genes that can differentially impact the phenotype, we will look where the SNP density is the greatest. We are doing this on the assumption that genes with higher number of SNPs are going to produce more variance in the phenotype than genes with few or no SNPs. Genes that are low in SNPs are going to be fairly consistent across recombinant inbred strains and so would be unlikely to contribute to differential phenotypes. So, when you select a region at the end of this section that will lead you into the UCSC Genome Browser choose a region where the density of SNPs is high.

Olfactory Bulb Analysis: Using GeneNetwork/WebQTL

Take Home Points Part III (Testable Material)

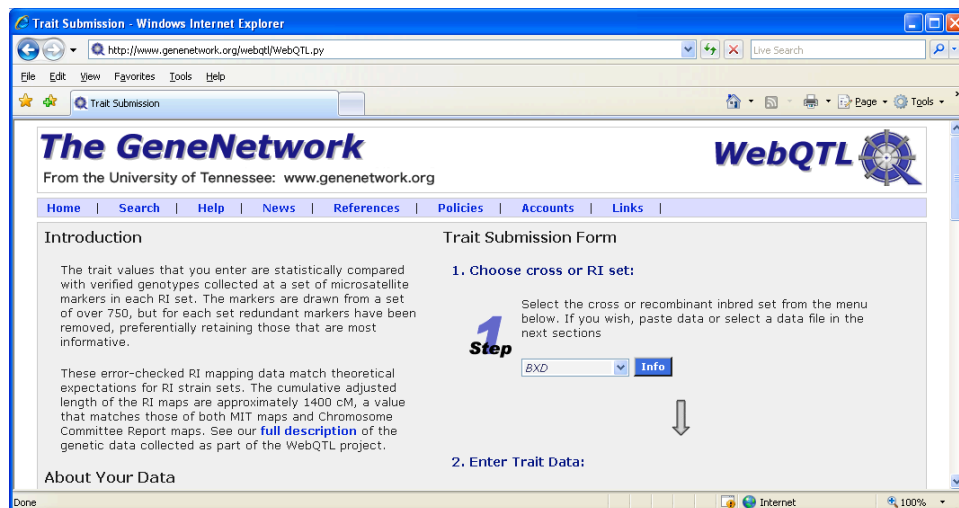
- What is the difference between a high LRS value for a marker and a low LRS value?
- What kind of variation do you find at SNPs?
- At the end, why did we look at regions that have high SNP density when picking a genome region to further analyze?

Open a word document for your group. Name it **WebQTL analysis** and save it to the desktop. This will be the document that you will use to compile your data from this section. Lightning bolts will highlight each place where data must be recorded. Email it to your group members at the end of this class.

1. Open the GeneNetwork Main website

- <http://www.genenetwork.org/>. It is also linked on our course website under today's discussion material.

When the page loads, select **Home** then **Enter Trait Data** from its drop down menu bar. You will be directed to a new page that will allow you to submit your trait data.



2. Submit trait data

Step 1: Choose 'BxD' as the recombinant inbred set.

Step 2: Enter the trait data by pasting your average residual values into the box below the heading 'By Pasting or Typing Multiple Values.' You will get these average residual values from "Olfactory Bulb Residuals.xls" file on our website. Copy the whole column

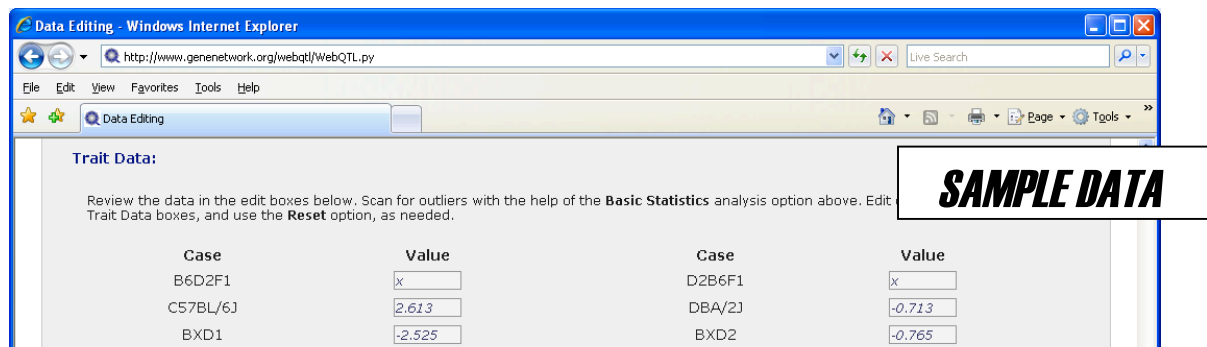
that is highlighted yellow, including the 'x's. Do not copy the header of the column. Simply click the box on the WebQTL page, then select **Edit** then **Paste** from the menu bar. Your residual values should appear in a column in the box.

Step 3: Do not select anything here.

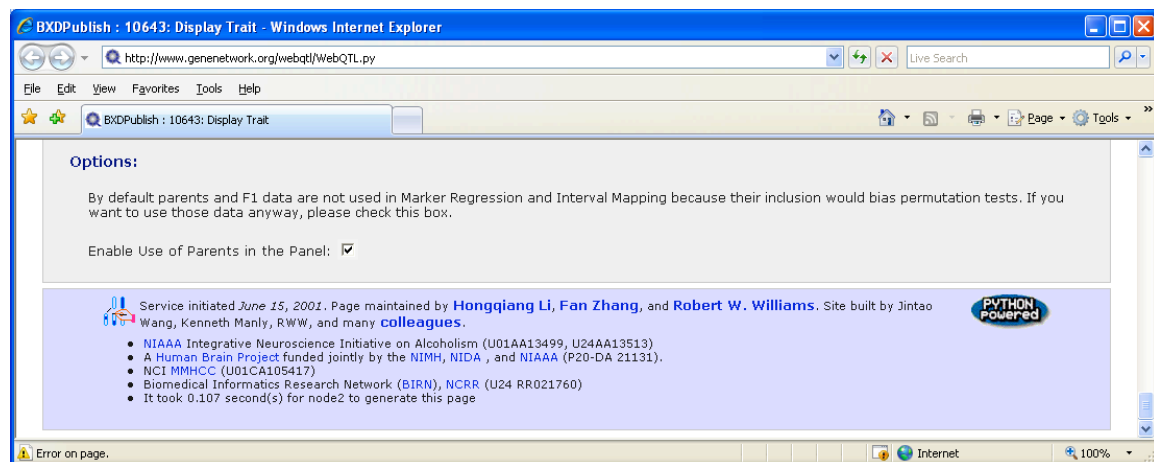
Select **Next** to submit your data.

3. Edit trait data

When the **Trait Data and Analysis Form** page loads, scroll down the page to ensure that your data was correctly entered. Be sure that the strains labeled 'B6D2F1,' 'D2B6F1,' and 'BxD41' have corresponding values of 'x.' Also be sure that the first real value in the list corresponds to 'C57BL/6J' and the last real value corresponds to 'BxD42.' There will be a bunch of other strains with x's below our last listed values—these correspond to recombinant inbred strains that we did not measure.

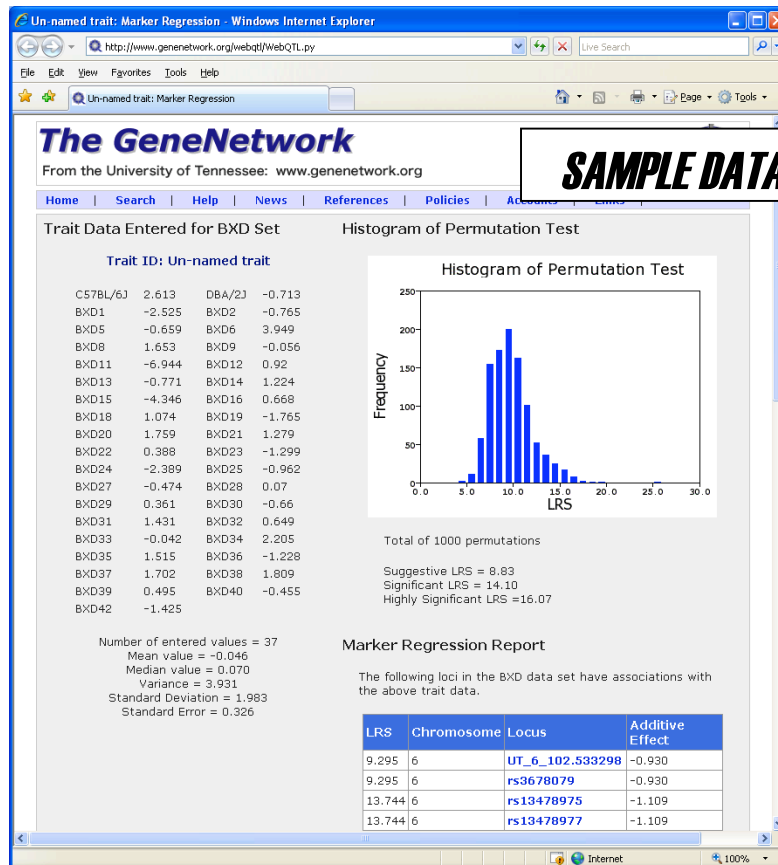


Under 'Options' below the trait data at the **bottom of the page**, choose 'Enable the Use of Parents.'



4. Observe the results of marker regression

Select the **Marker Regression** button near the upper right side of the **Trait Data and Analysis Form** page. A new window will appear, giving you the **Histogram of Permutation Test** and **Marker Regression Report**. Do not close the Trait Data and Analysis Form.



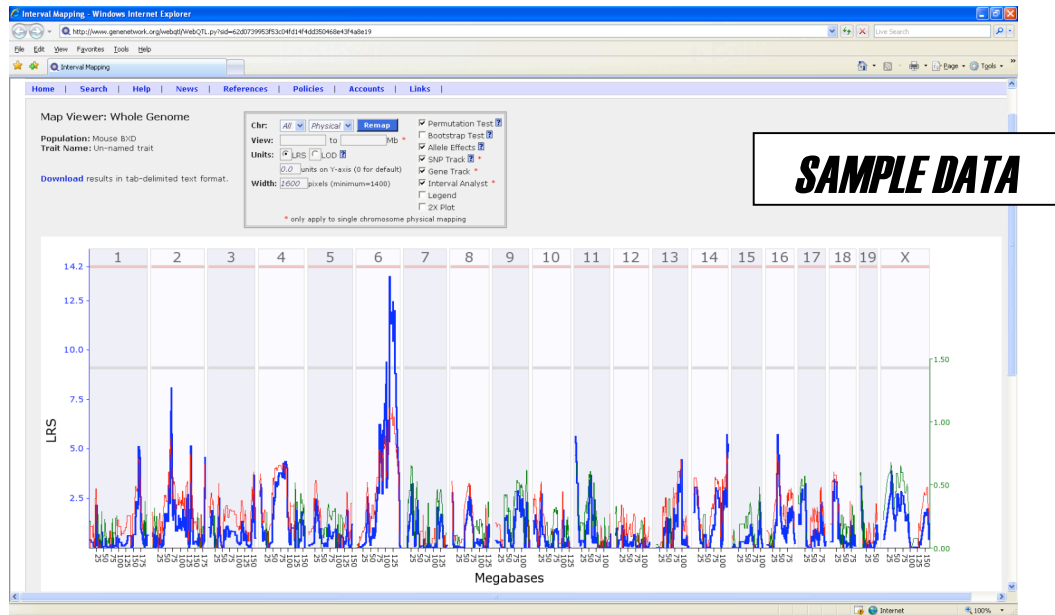
The **Marker Regression report** shows you the identity of the marker SNPs that showed significant linkage to our trait.



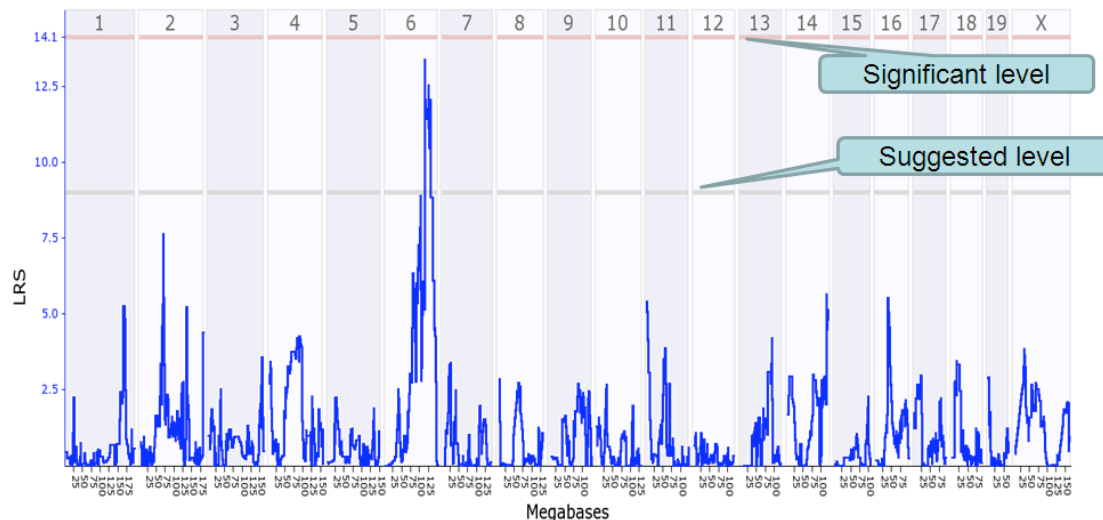
Record how many SNPs showed significant linkage on your word document. Choose one that had the highest LRS value and click on it. You will be taken to another page that is specific to that marker. At the top of the page will be the **Location of the SNP**. **On your word document, enter the name of the SNP and the Location.** Then click on the **NCBI link** below the location and copy and paste the actual SNP sequence into your **QTL word document**.

5. Observe the results of interval mapping

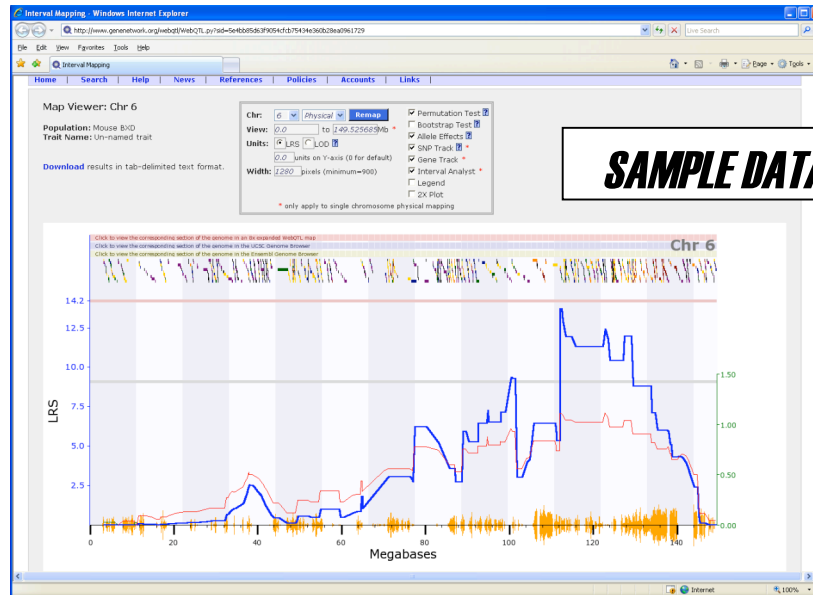
Select the **Interval Mapping** button near the upper left side of the **Trait Data and Analysis Form** page. A new window will appear with a **Map Viewer**.



De-select “allele effects” and “interval analyst” and Re-map in order to get a less “noisy” view of the LRS like the one below.

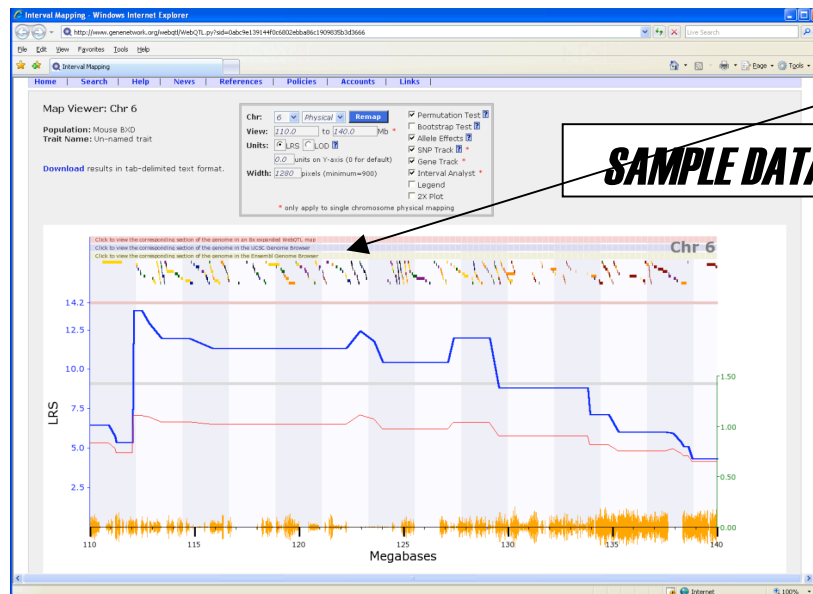


To zoom in on a chromosome of interest, simply click the chromosome number on the map. A new window will appear with a map of just that chromosome. In this example, chromosome 6 is selected.



The yellow lines that appear at the bottom of the graph indicate the SNP density.

Zoom into the graph further by selecting a range of megabases in which the peaks are highest, and enter this range in the boxes corresponding to 'View.' In this example only, 110-140 Mb is chosen (you may choose another range). Select the **Remap** button. A new window will appear with the zoom.



Take a screen shot of this linkage interval. This is the region of the mouse genome where genetic variation is linked to our variation in olfactory bulb volume. Paste the image into your word document.

Finally, the arrow in the above image points to continuous blue bar that runs across the linkage interval and is labeled “Click to view the corresponding section of the genome in the UCSC Genome Browser”. **Click a square within the significant linkage interval.** This will take you to the UCSC genome browser for this region. We will work on this region in the next class.



Make sure to record the position/search number that you find in the box on this page. It will have the form: chr6: xxx,xxx,xxx-xxx,xxx,xxx. This information will allow you to return to this exact region of the genome at the start of our next class.

To summarize what should be on your WebQTL document:

1. The number of linked SNP markers
2. The name, location, and variation of one highly linked SNP
3. A screen shot of your linkage interval
4. The positional information of your linked region.

Email this document to each group member. You will each have to turn this in at the end of the project.