

Analysis of quantitative trait loci (QTL)

1. Background information

Over the last half century, a number of methods have been developed to map genes controlling quantitatively measured phenotypes segregating in experimental populations. In general these techniques work by finding correlations between the inheritance of particular genetic markers and variation in the phenotype for each individual in the population.

MAPMAKER/QTL extends these methods to 1) provide support for “interval mapping”, allowing one to fully exploit the information provided by a genetic linkage map, and 2) to calculate LOD scores for putative QTLs, providing a measure of the support for any particular hypothesis

2. Getting data into MAPMAKER/QTL

data type f2 intercross
74 38 1

*XNor	HHNA	VBHVBANHHHABA	HNABNANBNAHNAHNAHAA	HNBBVVHNBHAAAAHHBVBH	ABANHHABANHAABVHNNH																
*XGli1	HHNA	VBHVBANHHHABA	HNABNANBNAHNAHNAHAA	HNBBVVHNBHAAAAHHBVBH	ABANHHABANHAABVHNNH																
*Xmwg9201	HHNA	VBHVBANHHHABA	HNABNANBNAH-AHNAHAA	HNBCVBHNBHAAAAHHBVBH	ABAH-HABANHAABVHNNH																
*Xmwg9202	HHNA	VBHVBANHHHABA	HNABNANBNAH-AHNAHAA	HNBCVBHNBHAAAAHHBVBH	ABAH-HABANHAABVHNNH																
*Xmwg2021	HHNA	VBHVCABANHHABA	HNABNANBNAHNAHNAHAA	HNBBVVHNBHAAAAHHBVBH	ABANHHABANHAABVHNNH																
*Xmwg60	HHNA	VBHVBANHHHABA	HNABNANBNAHNAHNAHAA	CHVCVBHNBHAAAAHHBVBH	ABANHHABAN-HAABVHNNH																
*Xmwg2083	HHNA	VBHVBANHHHABA	HAABNAH-BABNAHNAHAA	HNBBVVHNBHAAAAHHBVBH	ABANHHABANHAABVHNNH																
*Xabc156	HHBA	VBHVBANHHHABA	HAABNANBNAHNAHNAHAA	HNH-VHNNHAAAAHDBVH	ANHHNBANHHHAABVHNNH																
*XChs3	BDBD	BDBDBDBDBDDDD	DDDBDBDBDBDDDDDDDD	BDBDBDBDDDDDBDBDBD	DDDBDBDBDDDDDBDBDBD																
*Xbcd98	BHBA	BHNBANHHHABA	HAABVBA-BHNNHAAAHND	BHNBVBHNDHNNHBAVHNA	ANHBVNAHNAABVHNNH																
*Xabg500	BD-H	BHNAHNBVCBA-	HANBBVHNBHNNHNAHAA	BHNBVBHNAHNNHBAVANN	ACCCVHNNHABVHNNHND-																
*XTri	BHHH	BHNAHNBVBAB	HANBBVHNBHNNHNAHAA	BHNBVBHNAHNNHBAVANN	ANHBVHNNHABVHNNHNAH																
*Xbcd762	B-VH	BA-A--VBVBAB	HANBBVH-NBNNHNBVB-	BHNNHVBNAHNNHBAVANN	ANHBVHNNHBA-HVBVHNAH																
*Xpsr1201	BHVB	BANANHHVBVBAB	HANBBVHNBHNNHNBHAA	BHNNHVBNAHNNHBAVANN	ANHBVHNAABVHVBVHNAH																
*Xmwg758	BHVB	BHNAHNBVBVBAB	HANBBVHNBHNNHNBHAA	BHNCVBHNAHNNHBAVANN	ABANHHABANHAABVHNNH																
*Xmwg10	BHVB	BHNAHNBVBVBAB	HAABVHNBHNNHNBHAA	BHNNHVBNAHNNHBAVANN	ANHBVHNAHNBVBVHNAH																
*Xabc160	BHVB	BHNAHNBVBVBAB	HAABVHNBHNNHNBHAA	BHNNHVBNAHNNHBAVANN	ANHBVHNAHNBVBVHNAH																
*Xabc152	BDBB	BDDDDDBBDBDB	-DDBBDBDBB-BDDDBD	BBD-DBDDDBDDDBDD	DDBB-DDDD-DBBDDDD																
*XksuG30	BDBB	BDDDDDBBDBDB	DDDBBDBDBB-BDDDBD	BDDDBBDDDBDDDBDD	DDBB-DDDDDBBDDDD																
*XGlul	BHVB	BHNAHNNHVBHAB	HAABVHNBHVBDBHNBHAA	BHNNHVBHNNHNBHNAH	HVBANHAANHHNBVHNNH																
*XLec	BVBH	BHNAHNNHVBHAB	HAABHNNHVBHNBHABNA	BHNNHVBHNNHNBHNAH	HVBANHAANHHNBVHNNH																
*Xmwg504	BVBH	BHAANHHVBHAB	HAABHNNHVBHNBHABH-	BHNNHVBHNNHNBHNAH	HVBANHAANHHNBVHNNH																
*Xmwg733	BVBA	BHAANHHVBHAB	HAABHNNHVBHNBHNNH	BHNVHNBHNAHNNHNAH	HVBA-HNAANHABV-HNBH																
*Xpsr375	B-BA	BHA-H-HVBH--	CAA--HNBVBE-VBNAH	BV--HVBHNAHNBHNAACH	HVBA-HNNANHABV-C-B-																
*XksuE8	BVBA	BHAANHHVBHNNH	HAANHHNBVBHNBHABH	BHNNHVBH--HNBHNAH	HVBANH-NA--ABHNNH																
*Xmwg947	ABBA	BHAANHHVBHNNH	HAANHHNBVBHNBHABH	BV-HVBHNAHNBHNAACH	HVBA-HNNANHABVHNNH																
*Xabc257	ABBA	BHAANHHVBHNNH	HAANHHNBVBHNBHABH	BHNVHNBHNAHNNHNAH	HVBANHNAH-ABHNNH																
*Xmwg706	ABNA	BHAANHHVBH-NH	AAANHHNBVBHNBHABA-	BHNCVBHNBANHNBHANNH	HVBANHNBAB-ABHNNH																
*Xmwg676	ABNA	BHAANHHVBHNNH	AAANHHNBVBHNBHABA-	BHNVHNBHNAHNBHANNH	BVBAHNBANHABVHNNH																
*Xmwg984	ABNA	BHAANHHVBHNNH	AAANHHNBVBHNBHABA-	HDBAVHNBANHNBHANNH	BVBA-HBAANHNNHNBH																
*Xmwg710	ABNA	BHAANHHNBHABA	AAANHHNBVBHNBHANNH	HBAABVHNBANHNBHANNH	BVBAANHBAANHNNHNBH																
*Xabc261	ABNA	BHAANHHNBHABA	AAANHHNBVBHNBHANNH	HBAABVHNBANHNBHANNH	BVBA-HBAANHNNHNBH																
*Xbcd508	ABNA	BHAANHHNBHABA	AAANHHNBVBHNBHANNH	HBAABVHNBANHNBHANNH	BVBAANHBAANHNNHNBH																
*Xbcd1261	AB-H	BHAANHHNBH-V-	AAANHHNA-BH-BANHNB-	HBA-VHNBANHNBHANNH	BVBAANHBAH-NHNCNB-																
*XAg7	ABNH	BHAABVHNBHNC--	AAADANHHNBANHNBHABH	HBAANHNBANHNBHANNH	BVBAANHBAANHNNHNBH																
*XksuE11	DDDB	DDDBBDD-DBDD	DDDDDD-BDD-ABDDDBD	-BA-BB-AB-B-AA-HDD-	-----B--A--D-D-																
*Xcdo393	ANHB	HANHHNBHNBABA	HHNNHNNHNBABAAHNC	HVNNHNBANHNBHANNH	HVBANHNBANHNBHANNH																
*Xwg241	ANHB	HANHHNBHNBABA	HHNNHNNHNBABAAHNB	HVNNHNBANHNBHANNH	HVBANHNBANHNBHANNH																
*EPS	130	130	130	61	104	-	130	85	58	60	91	74	79	98	99	71	130	89	109	86	99
	130	90	79	120	77	60	94	-	130	63	115	130	130	61	60	114	51	80	129	-	77
	82	124	98	64	69	90	124	91	68	126	123	86	-	84	-	72	127	100	63	105	114
	66	69	90	97	124	64	-	128	66	128	78										

The data set we will analyze here is an RFLP map of chromosome 1 of *Triticum monococcum* based on 74 F₂ plants and the average days from sowing to flowering of each F₃ family derived from these F₂ plants. The name of the file is **QTL.txt** and is a plain ASCII file.

Note that the first row is identical to previous examples in MAPMAKER, but that the third number of the second row has a number 1 instead of a 0, indicating the presence of one QTL. Then all the RFLP markers are identical to previous examples and the last row includes the quantitative data. This data includes the name of the trait after a * and then the quantitative values for the 74 individuals separated by spaces.

The first step in running MAPMAKER/QTL is to run the regular MAPMAKER program to prepare the data for the MAPMAKER/QTL program.

After starting mapmaker type:

1> System

This command allows you to go to the DOS system and change the directory to the place where your data files are located. Use the letter of the hard drive first

i.e. CD C:\MAP\DATA

Then type EXIT to go back to MAPMAKER.

2> Prepare qtl.txt

To prepare the data for MAPMAKER/QTL you need to

- Define each chromosome you wish to search for QTLs

3> Make chromosome one

Where “one” is an arbitrary name assigned to our chromosome.

- Assign the markers in each chromosome’s map to the chromosome

4> Sequence 1

Indicates MAPMAKER that we are now considering the first marker

5> Anchor one

Anchors the RFLP marker 1 to the chromosome named “one”.

6> Sequence all: Selects all the markers in the matrix

7> Assign 0.5

Assign all the markers linked to RFLP marker 1 to the same chromosome. The default LOD is 3.0 but we are using a lower one 0.5 to be sure that all the markers will be included (we know by nullitetrasonic analysis that all these markers are on this particular chromosome).

- Set the framework of each chromosome to specify the correct order of these markers

8> sequence 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38

Indicates MAPMAKER the correct sequence of the markers

9> frame one

The frame command declares the framework order of the specified chromosome to be that specified by the previous sequence command. It is like running the map command but the difference is that this order will be saved.

To facilitate the preparation of these files for MAPMAKER /QTL it is possible to create a batch file that will tell MAPMAKER to perform all these commands. An example for this particular chromosome is indicated below and saved as QTLIN.txt.

```
prepare qtl.txt
make chromosome one
sequence 1
anchor one
sequence all
assign 0.5
sequence 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38
frame one
```

If you have prepared a file like this you just type: **RUN QTLIN.txt**

10> quit

Save data before quitting? **YES**

Now all the files are prepared to run MAPMAKER/QTL. To run the QTL program type QTL in the MAPMAKER directory or double click the icon for MAPMAKER/QTL.

3. Preliminary examination of the trait data

After starting mapmaker type:

1> System

To go to the DOS system and change the directory to the place where your data files are located using the CD command. Then type EXIT to go back to MAPMAKER.

2> load data QTL

This command will load all the files prepared by MAPMAKER that have the starting name QTL.

3> photo qtl.out

This command generates a copy of the session to the file qtl.out

The algorithms on which MAPMAKER/QTL is based assume that the values of the quantitative trait vary across the population following a normal distribution. Thus, the first step in most QTL mapping projects is to examine the trait data and determine how well they fit this assumption. If they do not closely follow a normal distribution, it is often possible to transform the data (i.e. log transformation) to fit the normal distribution.

We will assume that the distribution of our trait is not significantly different from a normal distribution.

4> trait 1

Using the trait command we first inform MAPMAKER/QTL to perform all further analyses on this trait. This is particularly important when various QTL are included in the same file.

5> show trait

The command “show trait” displays a simple statistical analysis of the data and prints a rough histogram.

```
Trait 1 (EPS):
-----
distribution:          quartile |   fraction within n deviations:
mean  sigma  skewness  kurtosis  ratio  |  1/4   1/2   1   2   3
94.46 25.08  0.12    -1.37    1.54  |  0.19 0.29 0.49 1.00 1.00
-----

44.29 |
56.83 |***
69.38 |*****
81.92 |*****
94.46 |*****
107.00|*****
119.54|*****
132.08|*****
144.62|
157.16|
```

15. 4. Scanning the genome for putative QTLs

Perhaps the most useful capability of MAPMAKER/QTL is to compute “QTL likelihood plots” covering the entire genome. These graphs provide a visual representation of 1) regions in the genome which are likely to contain putative QTLs, 2) the strength of the data supporting the hypotheses that particular QTLs exist, and 3) the likely position of these putative QTLs.

The fundamental method MAPMAKER/QTL uses to generate these plots works as follows: The program iteratively ‘steps’ along the chromosome and at each point calculates a “maximum likelihood map”. That is, at each point, MAPMAKER/QTL asks the question, “assuming there is a QTL right here, what is the maximally likely manner in which it’s inheritance affects the trait? Moreover, what is the strength of the data supporting this hypothesis?”

These results are expressed as a number of real valued parameters including:

- The effect of the QTL on the trait, expressed in terms of additive (“weight”) and dominance (“Dom”) effects for F₂ data, or a “weight” term alone for backcross data. The “Weight” term indicates the effect of the B allele at this particular locus on the phenotype.
- The fraction of the total variation in the trait across the population explained by the QTL (this is equivalent to an R² value calculated by linear regression).

- A lod score, also called log-likelihood, indicating the strength of the data supporting the hypothesis of a QTL at this locus.

6> sequence [all]

Tell the program which regions of the genome we wish to search

7> scan

The “scan” command instructs MAPMAKER/QTL to search for QTLs throughout the genome. By default MAPMAKER/QTL looks for QTLs every 2.0 cM generating a large amount of output. Below is the end of that output for the flowering time QTL on *T. monococcum* chromosome 1.

```

QTL maps for trait 1 (EPS):
Sequence: [all]
LOD threshold: 2.00  Scale: 0.25 per '*'
... part of the output deleted
POS  WEIGHT  DOM  %VAR  LOG-LIKE
10.0  -5.640  -6.026  4.5%  0.642 |
-----| 30-31 3.5 cM
0.0   -5.707  -6.170  4.7%  0.704 |
2.0   -5.960  -3.896  3.9%  0.573 |
-----| 31-32 2.8 cM
0.0   -5.942  -1.886  3.3%  0.493 |
2.0   -8.319  -4.253  6.5%  0.970 |
-----| 32-33 1.4 cM
0.0   -8.935  -3.925  7.2%  1.106 |
-----| 33-34 1.8 cM
0.0   -8.010  -3.127  5.4%  0.819 |
-----| 34-35 9.0 cM
0.0   -9.867  -6.848  9.6%  1.472 |
2.0   -11.85  -7.987  14.6%  2.024 | *
4.0   -13.22  -7.838  18.1%  2.527 | ***
6.0   -14.06  -6.867  19.9%  2.942 | ****
8.0   -14.46  -5.676  20.6%  3.263 | *****
-----| 35-36 16.8 cM
0.0   -14.51  -4.988  20.5%  3.385 | *****
2.0   -16.53  -5.044  25.9%  3.924 | *****
4.0   -18.45  -4.409  31.2%  4.501 | *****
6.0   -20.08  -3.611  36.1%  5.099 | *****
8.0   -21.35  -3.053  40.4%  5.685 | *****
10.0  -22.20  -2.679  43.4%  6.222 | *****
12.0  -22.73  -2.341  45.3%  6.688 | *****
14.0  -22.99  -1.919  46.2%  7.070 | *****
16.0  -22.98  -1.279  46.0%  7.362 | *****
-----| 36-37 9.1 cM
0.0   -22.93  -0.995  45.7%  7.457 | *****
2.0   -24.15  -2.466  49.1%  7.968 | *****
4.0   -24.47  -2.436  48.9%  8.235 | *****
6.0   -24.55  -1.939  47.4%  8.340 | *****
8.0   -24.49  -1.148  44.7%  8.324 | *****
-----| 37-38 1.4 cM
0.0   -24.41  -0.624  43.0%  8.280 | *****
-----|

```

Weight: additive effect of allele B. DOM: Dominance effect in the Het.

MAPMAKER/QTL scan command generates one line of output for each point examined in the genome. To declare the presence of any QTL, one needs to select an appropriate LOD score threshold above which the QTL map data will be considered significant. Though many papers have been written about selection of the appropriate LOD score, a LOD of 3 is usually conservative enough to be sure that the QTLs found are not artifacts. A LOD of 3 indicates that the probability of having a QTL in that

particular region is 1000 higher than the probability of not having a QTL giving the actual data. This high threshold is selected in order to be sure that no false QTLs will be considered in all the datapoints analyzed in the genome. If lower numbers of markers are analyzed, the threshold can be lower.

We can quickly examine the output for these regions by looking at the row of asterisk (*) characters printed next to each data line. These visually indicate the shape and height of the "log-likelihood surface". By default, one star is printed if the log-likelihood exceeds 2.0, with each additional star indicating an increase of 0.25 over this level. These thresholds may be changed by the user.

For example, on this particular example we see that a putative QTL has been detected on the distal end, with the most likely position six centimorgans distal to RFLP marker 36. As you can see, however, the likelihood surface around this putative QTL is fairly flat, that is the QTL LOD do not change substantially at close positions in the chromosome. A confidence interval can be constructed between the LOD scores that are one unit below the QTL peak. All information about the QTL peaks can be obtained by the show peaks command.

8> show peaks

```

LOD score peaks for scan 1.1 of trait 1 (EPS).
Sequence: [all]
No fixed-QTLs.
Scanned QTL genetics are free.
Peak Threshold: 2.00  Falloff: -2.00
=====
QTL-Map for peak 1:
Confidence Interval:  Left Boundary= 35-36 + 12.0
                      Right Boundary= 37-38 (off end)

INTERVAL  LENGTH  QTL-POS  GENETICS  WEIGHT  DOMINANCE
36-37     9.1     6.0     free     -24.556 -1.9391

chi^2= 38.409 (2 D.F.)      log-likelihood= 8.34
mean= 120.696  sigma^2= 331.173  variance-explained= 47.4%

```

The mean flowering date of

AA individuals is 120.696 days (mean from peaks)

AB heterozygous is $120.7 - 24.6 - 1.9 = 94.2$ (mean + weight + dominance)

BB individuals is $120.7 + (-24.6 * 2) = 71.5$ days

Note that in this case the QTL alleles inherited from the B parents give rise to shorter flowering times.

Dominance is the difference between the measures of the observed $AB=H$ and the intermediate values $(AA+BB)/2$

$$120.7 + 71.5 = 192.2$$

$$192.2 / 2 = 96.1$$

$$94.2 - 96.1 = 1.9$$

Degree of dominance= $1.9/24.6=0.077$ Close to 0 so mainly additive

9> quit Save data before quitting? **YES**

Exercise

1) The heights of the 36 plants from file MAP7A.txt are indicated below. Heights are in the same order as plants in the matrix MAP7A.txt.

Heights: 90 100 101 93 94 88 101 102 99 103 98 90 87 77 99 91 105

81 79 97 103 90 105 78 91 80 92 99 88 97 90 102 89 88 92 93

2a) Use MAPMAKER /QTL to calculate the most likely location of a QTL for height. Indicate clearly what is the most likely location of the QTL

2b) What is the LOD score of this QTL?

2c) What is the meaning of this LOD score?

2d) What proportion of the variation in height in this population is explained by the locus at the peak of the QTL?