Mapping quantitative trait loci with Mapmaker/QTL

There are three basic stages to doing a QTL analysis with Mapmaker/QTL:

1. Get the data into the format that Mapmaker needs.
2. Construct a genetic map for the marker data that you have (with Mapmaker).
3. Feed the marker and phenotype data into Mapmaker/QTL and run the analysis.

We'll go through each of these in turn.
The data format is relatively straightforward, as shown by this example from `sample.raw` that we'll be working with:

```
data type f2 intercross
333 12 1

*T175     HAHAHHA-HHHAHHAAHAAHHHAHAAAB-HAHHHAAHHHHHHHAHAHAAA-AHAH--HHA
          AAHHAA-AHHHAAAHAAAAHHAAHAAHAAAHHAHAHAHAH-HHAAAHHAHAAAAHAHAAH
          HHAAH-AAHHHHAAHHHHAAB-HAHAAHA-AAH-AAAHAAAHAHHHAH-AHHAH-HHAHH
          HHHAAAAAHAAAHHHAAHAHAAHHBAAHHHHHHHAHHAAAAHAHHHH--AAHABAA-AAA
          HAA-AHHHAAAHAAAHAAHAHHAAAAAAHHAAHHHAAHAAHAHAHHAAHHHHAHAHBHHA
          AAAAAAHAAHHHA--------------------
*T93      HAHBHHA-HHBABHHAHAAHHHAHAAAHAHABAHAAHHHHBHHBAAHAHAAAHHHAAAAA
          AAHHAHHABHHAAAHAAHAAH-ABAABA-HHAA--HAH-A-HH----HHH-H--H-HAAB
          -A-AA-HAH--HA--HHHB---A----H-A--HAHAHHAHHHHHHBAHBAHHAHAHAAHH
          HHHAAHAAHAHBBBHAAHAHAAHHBHBHHHHHHAAHHAAAHBAHAHHAAAHBABHHBAAA
          HAH-AHHHAAHHHHAHBAHAHHAHAHAAHHAAHHHAHAAAHHAAHHAABHAHAHAHBBHA
          HHAAAAHBAHHHAH-AAHAAHHHHAHAHAAHH-
     .
     .
     .
*weight        4.949       3.58         -         6.212       6.140
               5.330       5.761       5.470      7.897       6.790
               7.559        -          4.990      5.316       3.190
               6.160       8.150       5.370     16.330       8.990
               3.220       9.540        -         6.360       6.040
               5.480       4.710       6.310      4.390       1.829
               4.321       3.930       5.680      6.440       9.850
               7.660       9.260       1.780      1.780       3.620
               5.330       5.670       5.670       -          4.460
               5.940       3.880       5.740      3.770       4.810
               7.280      16.450       2.850      4.270        -
               4.430       2.320       9.760      3.550      10.540
              14.610       4.380       4.408       -         10.460
```

The first line in your data file is the cross type. The second is the number of individuals scored for phenotype and marker data, the number of markers scored, and the number of quantitative traits scored. Then comes the real data. The marker data has a name corresponding to each marker used. The genotypes are `A` if the individual is homozygous for the marker found in the A line, `B` if the individual is homozygous for the marker found in the B line, and `H` if the individual is

© 1999, 2001 Kent E. Holsinger

heterozygous. A - is used for missing data. After all of the marker data is entered the phenotype of each individual is entered.[1]

Once you have data in that format, you're ready to fire up Mapmaker and begin your analysis.[2] First, it's not a bad idea to turn on the recording function to keep a record of what we do:

```
1> photo "c:\temp\map.txt"
'photo' is on: file is 'C:\TEMP\MAP.TXT'
```

Next we prepare the data for analysis, and figure out which markers are in which linkage groups.

```
2> prepare sample.raw
preparing data from file 'SAMPLE.RAW'... ok
  F2 intercross data  (333 individuals, 12 loci)... ok
unable to run file 'SAMPLE.PRE'... skipping initialization
saving genotype data in file 'SAMPLE.DAT'... ok
saving map data in file 'SAMPLE.MAP'... ok
saving traits data in file 'SAMPLE.TRA'... ok

3> sequence 1 2 3 4 5 6 7 8 9 10 11 12
sequence #1= 1 2 3 4 5 6 7 8 9 10 11 12

4> group
Linkage Groups at min LOD 3.00, max Distance 50.0

group1= 1 2 3 5 7
-------
group2= 4 6 8 9 10 11 12
```

Since God told us what order the markers fall in, it's quite easy to construct our maps.

```
5> sequence 1 3 2 5 7
sequence #2= 1 3 2 5 7

6> map
===============================================================================
Map:
  Markers          Distance
    1  T175          4.2 cM
    3  C35          15.0 cM
    2  T93          11.9 cM
    5  C66          12.2 cM
    7  T50B       ----------
                    43.2 cM   5 markers   log-likelihood= -424.94
===============================================================================

7> sequence 4 11 8 12 9 6 10
sequence #3= 4 11 8 12 9 6 10
```

[1] It's obvious, isn't it, that the individuals have to correspond across markers and phenotypes?

[2] In the simplified example I'm describing to you, I'll assume that we already know (perhaps God told us) what the order of markers is. All we need to do is to get them on to chromosomes and mapped. Mapmaker has functions that allow you to figure out the order of markers, too, but we won't explore them here.

© 1999, 2001 Kent E. Holsinger

```
8> map
================================================================================
Map:
  Markers          Distance
    4  T24          14.8 cM
   11  C15           6.4 cM
    8  T125         18.9 cM
   12  T71          24.0 cM
    9  T83          18.1 cM
    6  T209         28.6 cM
   10  T17        ----------
                   110.8 cM   7 markers   log-likelihood= -688.99
================================================================================
```

Of course, we haven't made them into chromosomes yet, but fortunately that's pretty easy, too.

```
9> make chromosome one two
chromosomes defined: one two

10> sequence 1
sequence #4= 1

11> anchor 1
error: there is no chromosome named '1'

12> anchor one
1    - anchor locus on one
chromosome one anchor(s): T175

13> sequence 4
sequence #5= 4

14> anchor two
4    - anchor locus on two
chromosome two anchor(s): T24

15> sequence all
sequence #6= all

16> assign
1    - anchor locus on one...cannot re-assign
4    - anchor locus on two...cannot re-assign
2    - assigned to one at LOD 36.2
3    - assigned to one at LOD 90.6
5    - assigned to one at LOD 53.4
7    - assigned to one at LOD 54.2
8    - assigned to two at LOD 39.1
11   - assigned to two at LOD 80.9
12   - assigned to two at LOD 39.8
6    - assigned to two at LOD  9.1
9    - assigned to two at LOD 36.7
10   - assigned to two at LOD 22.5
```

© 1999, 2001 Kent E. Holsinger

```
17> sequence 1 3 2 5 7
sequence #7= 1 3 2 5 7

18> frame one
setting framework for chromosome one...
================================================================================
one framework:

  Markers          Distance
    1  T175          4.2 cM
    3  C35          15.0 cM
    2  T93          11.9 cM
    5  C66          12.2 cM
    7  T50B       ----------
                    43.2 cM   5 markers   log-likelihood= -424.94
================================================================================

19> sequence 4 11 8 12 9 6 10
sequence #8= 4 11 8 12 9 6 10

20> frame two
setting framework for chromosome two...
================================================================================
two framework:

  Markers          Distance
    4  T24          14.8 cM
   11  C15           6.4 cM
    8  T125         18.9 cM
   12  T71          24.0 cM
    9  T83          18.1 cM
    6  T209         28.6 cM
   10  T17        ----------
                   110.8 cM   7 markers   log-likelihood= -688.99
================================================================================
```

At this point we're done with Mapmaker, so it's time to exit and fire up Mapmaker/QTL.

```
21> q
save data before quitting? [yes] yes
saving map data in file 'SAMPLE.MAP'... ok
saving two-point data in file 'SAMPLE.2PT'... ok

        ...goodbye...
```

Again, it's a good idea to capture what we're doing as a record for future reference.

```
1> photo "c:\temp\qtl.txt"
'photo' is on: file is 'C:\TEMP\QTL.TXT'
```

First we load the data we have just prepared in Mapmaker, and examine the distribution of our trait

© 1999, 2001 Kent E. Holsinger

```
2> load data sample
data files 'SAMPLE.DATA' and 'SAMPLE.TRAITS' are loaded.
(333 intercross progeny, 13 loci, 1 trait)
Unable to load any saved QTL map data.

3> trait 1
The current trait is now: 1 (weight)

4> show trait

Trait 1 (weight):

-----------------------------------------------------------------------------
distribution:                      quartile    fraction within n deviations:
mean   sigma    skewness  kurtosis ratio       1/4    1/2  1     2     3
6.11   3.51     1.59      3.53     0.83        0.23   0.42 0.75  0.95  0.99
-----------------------------------------------------------------------------

 -0.91
  0.85
  2.60  ************************
  4.35  ******************************************************
  6.11  ***********************************************************
  7.86  **************************************
  9.61  *******************
 11.37  ************
 13.12  *******
 14.87  ***
```

Remember that a fundamental assumption of QTL analysis is that trait values are normally distributed. The skewness and kurtosis of a normal distribution are zero, and just looking at the crude histogram above suggests that values are skewed right. Often in these circumstance, a log transform of the data will correct the problem, and it's quite easy to do:

```
5> make trait logwt=log(weight)

New trait number 2 (logwt) had been added to the data set.

-----------------------------------------------------------------------------
distribution:                      quartile    fraction within n deviations:
mean   sigma    skewness  kurtosis ratio       1/4    1/2  1     2     3
0.72   0.24     -0.06     -0.07    0.97        0.21   0.41 0.67  0.94  1.00
-----------------------------------------------------------------------------

  0.25  *****
  0.36  *******
  0.48  ***************************
  0.60  *********************************
  0.72  ***************************************************
  0.84  ***********************************************************
  0.96  **********************************
  1.08  **********************
  1.20  ***********
```

© 1999, 2001 Kent E. Holsinger

```
  1.32  ****
```

Just to make sure everything is in order, we take one last look at the linkage map for our marker loci, and select trait 2 for analysis.

```
6> seq [all]
The sequence is now '[all]'

7> show linkage maps

linkage maps:
==============================================================
 1-3           4.2 cM    4.0 %
 3-2          15.0 cM   13.0 %
 2-5          11.9 cM   10.6 %
 5-7          12.2 cM   10.8 %
==============================================================
 4-11         14.8 cM   12.8 %
 11-8          6.4 cM    6.0 %
 8-12         18.9 cM   15.7 %
 12-9         24.0 cM   19.1 %
 9-6          18.1 cM   15.2 %
 6-10         28.6 cM   21.8 %
==============================================================


8> trait 2
The current trait is now: 2 (logwt)
```

Now the fun begins. First, we simply scan the entire genome (at 2cM intervals) and see if we can find where the QTL loci are likely to be.

```
9> scan

QTL maps for trait 2 (logwt):
Sequence: [all]
LOD threshold: 2.00  Scale: 0.25 per '*'

No fixed-QTLs.
Scanned QTL genetics are free.

POS     WEIGHT  DOM     %VAR  LOG-LIKE  *
---------------------------------------- 1-3 4.2 cM
0.0    -0.033  -0.072   4.7%    3.083   *****
2.0    -0.053  -0.048   4.5%    2.814   ****
4.0    -0.067  -0.022   3.8%    2.474   **
---------------------------------------- 3-2 15.0 cM
0.0    -0.068  -0.021   3.8%    2.442   **
2.0    -0.076  -0.023   4.8%    2.873   ****
4.0    -0.080  -0.029   5.8%    3.291   ******
6.0    -0.081  -0.036   6.6%    3.667   *******
8.0    -0.081  -0.041   7.2%    3.983   ********
10.0   -0.079  -0.046   7.4%    4.228   *********
12.0   -0.077  -0.048   7.4%    4.395   **********
14.0   -0.075  -0.048   7.1%    4.483   **********
```

© 1999, 2001 Kent E. Holsinger

```
--------------------------------------- 2-5 11.9 cM
0.0    -0.073  -0.048   6.8%    4.500   ***********
2.0    -0.079  -0.048   7.7%    4.755   ************
4.0    -0.084  -0.046   8.2%    4.912   ************
6.0    -0.088  -0.043   8.4%    4.969   ************
8.0    -0.088  -0.041   8.2%    4.920   ************
10.0   -0.087  -0.038   7.7%    4.757   ************
--------------------------------------- 5-7 12.2 cM
0.0    -0.083  -0.034   6.8%    4.501   ***********
2.0    -0.084  -0.037   7.2%    4.427   **********
4.0    -0.084  -0.038   7.3%    4.236   *********
6.0    -0.081  -0.038   6.9%    3.931   ********
8.0    -0.077  -0.036   6.2%    3.526   *******
10.0   -0.071  -0.032   5.2%    3.046   *****
12.0   -0.063  -0.026   4.0%    2.535   ***
--------------------------------------- *
 *
--------------------------------------- 4-11 14.8 cM
0.0    -0.102  -0.007   9.0%    5.645   ***************
2.0    -0.110  -0.008  10.4%    6.159   *****************
4.0    -0.116  -0.008  11.4%    6.584   *******************
6.0    -0.119  -0.007  12.1%    6.897   ********************
8.0    -0.120  -0.006  12.3%    7.083   *********************
10.0   -0.120  -0.005  12.1%    7.135   *********************
12.0   -0.117  -0.006  11.4%    7.054   *********************
14.0   -0.111  -0.009  10.4%    6.853   ********************
--------------------------------------- 11-8 6.4 cM
0.0    -0.109  -0.010   9.9%    6.752   ********************
2.0    -0.118  -0.012  11.4%    7.418   **********************
4.0    -0.122  -0.014  12.0%    7.802   ************************
6.0    -0.122  -0.016  11.8%    7.932   ************************
--------------------------------------- 8-12 18.9 cM
0.0    -0.121  -0.016  11.7%    7.931   ************************
2.0    -0.130  -0.014  13.6%    8.409   **************************
4.0    -0.136  -0.011  15.1%    8.753   ****************************
6.0    -0.140  -0.009  16.0%    8.926   ****************************
8.0    -0.140  -0.009  16.3%    8.914   ****************************
10.0   -0.138  -0.010  16.0%    8.723   ***************************
12.0   -0.134  -0.013  15.2%    8.369   **************************
14.0   -0.128  -0.016  13.9%    7.880   ************************
16.0   -0.119  -0.020  12.2%    7.292   **********************
18.0   -0.109  -0.022  10.3%    6.647   *******************
--------------------------------------- 12-9 24.0 cM
0.0    -0.104  -0.022   9.5%    6.357   ******************
2.0    -0.106  -0.022   9.8%    6.123   *****************
4.0    -0.107  -0.021  10.0%    5.825   ****************
6.0    -0.107  -0.020   9.9%    5.461   **************
8.0    -0.105  -0.019   9.7%    5.032   *************
10.0   -0.102  -0.018   9.1%    4.543   ***********
12.0   -0.097  -0.017   8.3%    4.004   *********
14.0   -0.090  -0.015   7.2%    3.434   ******
16.0   -0.082  -0.013   5.9%    2.856   ****
18.0   -0.072  -0.011   4.6%    2.301   **
```

© 1999, 2001 Kent E. Holsinger

```
20.0   -0.062  -0.008   3.4%    1.798   *
22.0   -0.052  -0.006   2.3%    1.365   *
24.0   -0.042  -0.003   1.6%    1.010   *
--------------------------------------- 9-6 18.1 cM
0.0    -0.042  -0.003   1.5%    1.003   *
2.0    -0.045  -0.010   1.8%    1.078   *
4.0    -0.049  -0.018   2.2%    1.171   *
6.0    -0.052  -0.027   2.6%    1.275   *
8.0    -0.054  -0.034   3.0%    1.375   *
10.0   -0.055  -0.040   3.2%    1.457   *
12.0   -0.055  -0.044   3.3%    1.505   *
14.0   -0.053  -0.046   3.1%    1.515   *
16.0   -0.051  -0.045   2.8%    1.487   *
18.0   -0.048  -0.043   2.4%    1.428   *
--------------------------------------- 6-10 28.6 cM
0.0    -0.047  -0.043   2.4%    1.423   *
2.0    -0.048  -0.046   2.7%    1.378   *
4.0    -0.047  -0.048   2.9%    1.310   *
6.0    -0.046  -0.049   3.0%    1.217   *
8.0    -0.045  -0.048   2.9%    1.099   *
10.0   -0.043  -0.046   2.7%    0.960   *
12.0   -0.040  -0.041   2.4%    0.807   *
14.0   -0.036  -0.034   1.9%    0.651   *
16.0   -0.032  -0.025   1.4%    0.506   *
18.0   -0.028  -0.015   0.9%    0.383   *
20.0   -0.025  -0.006   0.6%    0.291   *
22.0   -0.022   0.002   0.4%    0.231   *
24.0   -0.020   0.009   0.4%    0.199   *
26.0   -0.018   0.015   0.3%    0.190   *
28.0   -0.017   0.020   0.3%    0.199   *
--------------------------------------- *

Results have been stored as scan number 1.
```

Eyeballing this, you can see that there appear to be peaks 6.0cm past marker 2 and 6.0cm past marker 8. Fortunately, there's an easier way to pick out those peaks than just scanning, namely

```
10> show peaks

LOD score peaks for scan 1.1 of trait 2 (logwt).
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= 3-2 + 4.0
                     Right Boundary= 5-7 + 10.0

INTERVAL   LENGTH  QTL-POS  GENETICS    WEIGHT  DOMINANCE
2-5        11.9    6.0      free       -0.0875 -0.0434
```

© 1999, 2001 Kent E. Holsinger

```
chi^2= 22.884 (2  D.F.)        log-likelihood= 4.97
mean= 0.805   sigma^2= 0.052   variance-explained= 8.4 %
==============================================================
QTL-Map for peak 2:
Confidence Interval:  Left Boundary= 11-8 + 2.0
                     Right Boundary= 8-12 + 16.0

INTERVAL   LENGTH  QTL-POS  GENETICS    WEIGHT  DOMINANCE
8-12       18.9    6.0      free       -0.1396 -0.0093

chi^2= 41.106 (2  D.F.)        log-likelihood= 8.93
mean= 0.866   sigma^2= 0.047   variance-explained= 16.0%
==============================================================
```

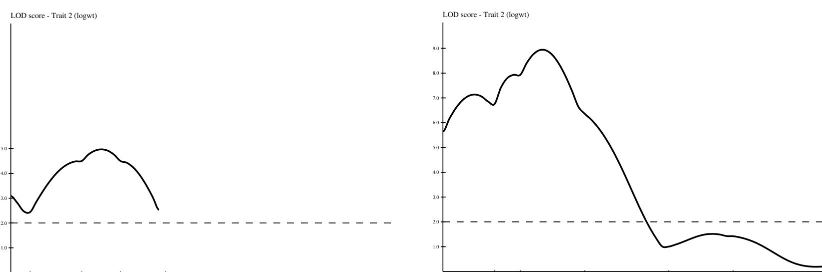
The trait value for a given QTL genotype is equal mean for AA homozygotes at this locus (0.805) plus the `WEIGHT` coefficient times the number of B alleles in the genotype plus the `DOMINANCE` coefficient, if the individual is heterozygous. Thus, individuals heterozygous for the QTL at 2+6 have a log-weight 0.1309 units less than AA individuals, and BB individuals have a log-weight 0.1750 units less than AA individuals. Notice that this QTL also explains about 8.4% of the total phenotypic variance.

It's also possible to produce nice-looking postscript plots of the QTL positions.

```
11> draw scan

scan 1.1 saved in PostScript file 'scan1_1.ps'
```

The file will be save in `c:\mapmaker`, and when you print it, you'll get figures that look like these:



Since we have identified two potential QTL's, it is useful to look at their joint effects.

```
17> sequence [2+6][8+6]
The sequence is now '[2+6][8+6]'

18> map

==============================================================
QTL map for trait 2 (logwt):

INTERVALS  LENGTH  QTL-POS  WEIGHT  DOMINANCE
2-5        11.9    6.0     -0.1083 -0.0301
8-12       18.9    6.0     -0.1496 -0.0151
```

© 1999, 2001 Kent E. Holsinger

```
chi^2= 74.385 (4  D.F.)        log-likelihood= 16.15
mean= 0.971   sigma^2= 0.041   variance-explained= 27.0%
==============================================================

19> quit
save data before quitting? [yes] yes
Now saving SAMPLE.QTLS...
Now saving SAMPLE.TRAITS...

        ...goodbye...
```

Notice that the WEIGHT and DOMINANCE coefficients change in magnitude, albeit not drastically, and that the variance explained is slightly greater than the sum of the individual variances.

© 1999, 2001 Kent E. Holsinger