

MAPPING QUANTITATIVE TRAIT LOCI WITH QTL CARTOGRAPHER

There are two stages to making a QTL map for a particular trait (once you've scored tens or hundreds of marker loci in tens or hundreds of F_2 or backcross progeny):

1. Construct a genetic map of your markers.
2. Feed the genetic map, marker data, and phenotype data into QTL Cartographer and run the analysis.

Although constructing the genetic map of your markers is really important step, we're not going to talk about it further. If you'd like to see how to do it, take a look at <http://darwin.eeb.uconn.edu/notes/qtl-mapmaker.pdf>. It is, after all, simply an elaboration of classical Mendelian genetics. We're going to focus on the second step.

The data

You'll find two sample files that are distributed with QTL Cartographer (<http://darwin.eeb.uconn.edu/e> and <http://darwin.eeb.uconn.edu/eeb348/realdatac.inp>) on the course web site. They are text files that contain the basic data we'll be using in the analysis. `realdatam.inp` contains information about the genetic map of our marker loci. `realdatac.inp` contains information about the marker genotypes and associated phenotypes of the individuals in our crosses.

After some introductory information about the markers and the genetic map, `realdatam.inp` provides the following description of our genetic map:

```
# 13234789    bychromosome    -filetype map.inp
-type intervals  the interval distance following the marker is given.
-function      1  where 1 => haldane mapping function,
                2  => kosambi and
                3  => fixed (no mapping function).
-Units        cM  where cM means centiMorgans
```

```

-chromosomes 1 the haploid number of chromosomes
-maximum      9 markers on any chromosome.
-named        yes markers will have names.

-start
-Chromosome c10
D10MIT31  9.1
D10MIT42  4.2
IGF1      1.3
D10MIT9   1.3
D10MIT10  1.6
D10MIT41  3.3
D10MIT12  2.2
D10NDS2   8.3
D10MIT14  0.0
-stop

-end

```

There is one chromosome represented in our marker data (labeled c10). The last table gives the name of each marker, in linkage map order, and the distance to the next marker. The crossing data (`realdatac.inp`) is quite a bit more complicated than this.

```

# 1472195375 -filetype cross.inp
#
# Data from: Horvat and Medrano, 1995. Genetics 139:1737-1748
# We are grateful to Juan Medrano for kindly allowing us to
# distribute this data set with QTL Cartographer.
#
# This is a data set in the standard QTL Cartographer input format.
# It is defined in the manual and in the file cross.inp.
#
-SampleSize      190 is the sample size
-Cross RF2 is the type of cross
-traits 1 is the number of traits
-otraits 0 is the number of other traits
-case yes
-TranslationTable
  AA 2 AA
  Aa 1 Aa
  aa 0 aa
  A- 12 A-
  a- 10 a-
  -- -1 --
-missingtrait . A missing trait is encoded by a solitary period.
-start markers
D10MIT31 Aa Aa AA Aa aa AA Aa Aa aa Aa Aa Aa AA aa Aa aa Aa Aa
AA Aa AA Aa Aa AA Aa Aa AA Aa Aa Aa AA Aa Aa AA Aa Aa Aa

```

```

aa AA Aa Aa aa aa Aa Aa Aa AA AA aa aa AA Aa Aa aa Aa aa AA
AA Aa Aa aa Aa AA AA AA Aa Aa Aa Aa Aa AA aa Aa AA Aa Aa Aa
Aa AA AA aa AA AA aa aa AA aa AA Aa AA aa AA aa Aa Aa Aa AA
Aa aa Aa aa Aa Aa AA Aa Aa Aa Aa Aa Aa aa Aa Aa Aa AA Aa Aa
AA Aa AA Aa Aa Aa aa aa AA Aa Aa AA aa Aa AA aa AA AA Aa AA
Aa AA aa AA Aa Aa Aa Aa Aa Aa AA Aa Aa AA Aa Aa AA AA AA Aa
AA Aa aa aa aa AA AA Aa Aa AA Aa Aa Aa AA aa Aa Aa AA AA AA
aa Aa Aa Aa Aa AA Aa AA AA aa aa
D10MIT42 Aa aa Aa Aa aa Aa Aa Aa aa Aa Aa Aa aa AA aa Aa Aa aa Aa
AA Aa AA AA Aa Aa Aa Aa AA Aa Aa Aa Aa AA Aa Aa AA Aa Aa Aa
Aa AA Aa Aa Aa aa Aa AA Aa AA AA aa aa Aa Aa Aa Aa Aa AA
AA Aa Aa aa Aa AA AA AA Aa Aa AA Aa Aa AA aa Aa AA Aa Aa Aa
.
.
Aa Aa aa AA Aa aa Aa aa Aa Aa AA AA AA Aa Aa Aa Aa Aa Aa Aa
Aa Aa Aa Aa aa AA AA Aa Aa AA Aa AA AA Aa AA Aa Aa AA AA Aa
Aa AA Aa aa aa AA Aa AA AA Aa aa
-stop markers
-start traits
GAIN29 12.100000 15.600000 14.000000 14.600000 13.500000 13.200000 17.300000 13.000000 16.000000
11.600000 12.700000 13.000000 14.800000 11.400000 16.600000 19.700000 16.300000 16.200000 22.200000
28.400000 20.100000 10.900000 22.500000 15.100000 14.800000 19.000000 13.600000 21.700000 15.400000
13.800000 15.100000 19.600000 18.800000 16.500000 15.100000 23.300000 13.300000 13.600000 16.800000
16.900000 17.100000 16.700000 18.000000 17.500000 13.800000 20.500000 18.000000 15.000000 14.300000
.
.
18.400000 17.800000 14.600000 12.000000 10.300000 11.200000 16.000000 19.200000 20.800000 13.300000
11.800000
-stop traits

```

After some basic information about the structure of the data the line `-start markers` denotes the beginning of the marker information. The genotype of the 190 individuals scored is entered following the label form the marker. When the marker data is finished (`-stop markers`), we start the trait information (`-start traits`). After a short label describing the trait, the information for each individual is entered.

Rmap, Rcross, and Qstats

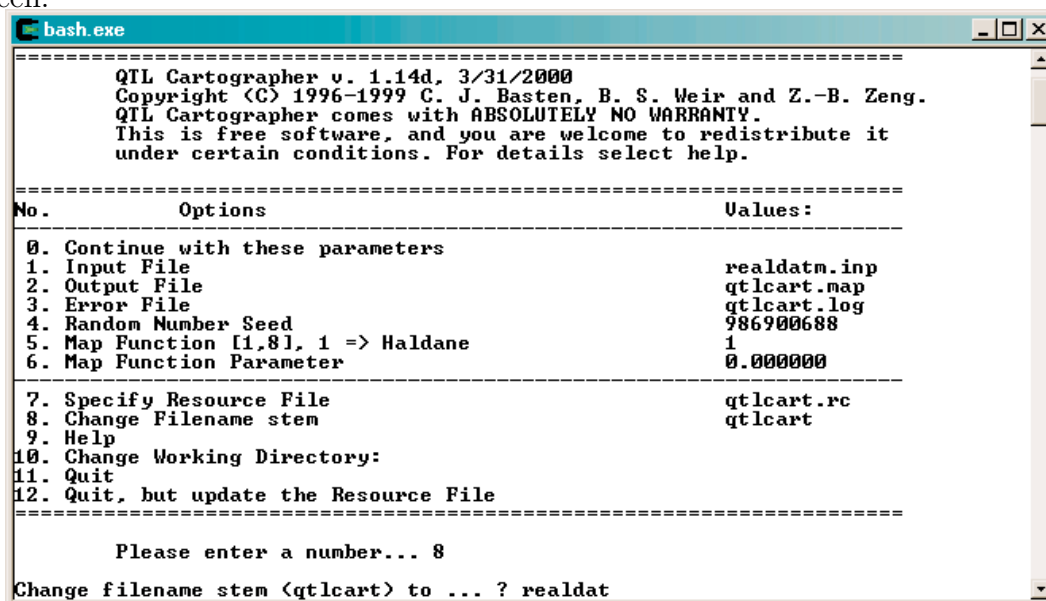
QTL Cartographer was written on Unix workstations, and its design reflects the Unix philosophy. Instead of writing one, big, monolithic program that does everything. Write a series of small tools that work well together, and allow users to work with them individually as they see fit. That's great for complex analyses, but it makes learning the package somewhat more difficult. For our purposes, we'll just run through a series of programs sequentially to produce a simple result.

Rmap takes the more or less human-readable map data and massages it into a form that LRmapqt1, SRmapqt1, and Zmapqt1 find easier to work with. Rcross does the same thing

with the crossing data. Running them is quite simple.¹ Simply type

```
..\bin\Rmap -i realdatm.inp
```

at the command prompt to start Rmap. When you do you'll see something like the following screen:



Hit "8" to change the filename stem, and enter realdat when prompted. Then hit "0" to do the analysis. You'll find the results in realdat.map. Then type

```
..\bin\Rcros -i realdatc.inp
```

, hit "0" to do the analysis,² and you'll find the results in realdat.cro.

Qstats provides summary information about the data that we'll be analyzing. You run it just the way you've run the other programs, and the results will be written to realdat.qst. The first section of tells us about the phenotype distribution in our data.

```
-----
This is for -trait 1 called GAIN29
-----
```

```
Sample Size..... 189
M(1)..... 16.2333
```

¹I'm going to assume that you've put the data files in a working directory under QTLCartWin.

²When you ran Rmap you set up a configuration file that remembered your preference for the filename stem.

```

M(2)..... 275.5310
M(3)..... 4886.7546
M(4)..... 90456.0754
Mean Trait Value..... 16.2333
Variance..... 12.0737
Standard Deviation..... 3.4747
Coefficient of Variation... 0.2140
Average Deviation..... 2.7746
Skw..LW(24)..... 24.4358
.....Sqrt(6/n)..... 0.1782
Kur..LW(29)..... 479.5400
.....Sqrt(24/n)..... 0.3563
k3...LW(24)..... 0.5825
k4...LW(28)..... 0.2896
S (5%: 5.99, 1%: 9.21)..... 11.3470
-----
-----

```

The most important number here is on the last line. It indicates that our phenotypic data are not well-described by a normal distribution. It's not too hard to see why if you look at the histogram that appears just below this table in the file. The data are bunched to the left, with only a few high values. In a real analysis we'd want to transform the data so that it fits a normal distribution better, but we won't do that today.

The last table in the output provides two tests of whether our markers are behaving as Mendel would have predicted.

Summary of marker segregation

```

-----
-----
Chrom  Mark   Name           type   n(m)           Chi2           LR
-----
-begin segregation
  1     1   D10MIT31       co     190            5.2421         5.6179
  1     2   D10MIT42       co     190           10.6737        11.9570
  1     3   IGF1           co     190            9.6947        10.7630
  1     4   D10MIT9        co     190            8.9474         9.8102
  1     5   D10MIT10       co     190            7.9263         8.6400
  1     6   D10MIT41       co     190            5.1474         5.5322
  1     7   D10MIT12       co     190            5.9053         6.3923
  1     8   D10NDS2        co     190            6.8316         7.4279

```

```

1      9      D10MIT14                co      190                5.6526                5.8318
-----
-----
-end segregation

```

Both can be compared to a χ^2 with one degree of freedom (3.84). You'll notice that our markers aren't behaving especially well. Unfortunately, there's not much we can do about that, but it's interesting to see it.

SRmapqtl, Zmapqtl, and Eqtl

SRmapqtl uses stepwise-linear regression to identify the markers that are associated with trait differences. Option "6" tells the program which type of stepwise selection to use: forward selection without elimination (0), backward elimination (1), or forward selection combined with backward elimination (2). The first two methods only attempt to rank the markers in the order of their influence. The last attempts to identify those with a significant influence on expression of the trait. As explained in the QTL Cartographer manual, it is probably the best choice when you intend to use marker information as a statistical control for differences in the genetic background. It's the one we'll use for our analyses.

```

-----
-----
Chromosome      Marker      Rank      F-Stat      DOF
-----
1              7          1         49.89893     186
1              9          2         4.54364     184
-----
-----

```

As you can see, only two of our markers appear to be statistically associated with variation in our trait.

Zmapqtl is (finally) where we actually map the position of any QTL's revealed by our analysis and display the genetic parameters associated with each QTL. We'll use model 6, which uses the analysis we just completed to provide statistical control for background genetic variation. There are 21 columns in the output, which are described on p. 63 of the QTL Cartographer manual. Fortunately, Eqtl scans the output of Zmapqtl and produces a more easily readable summary.

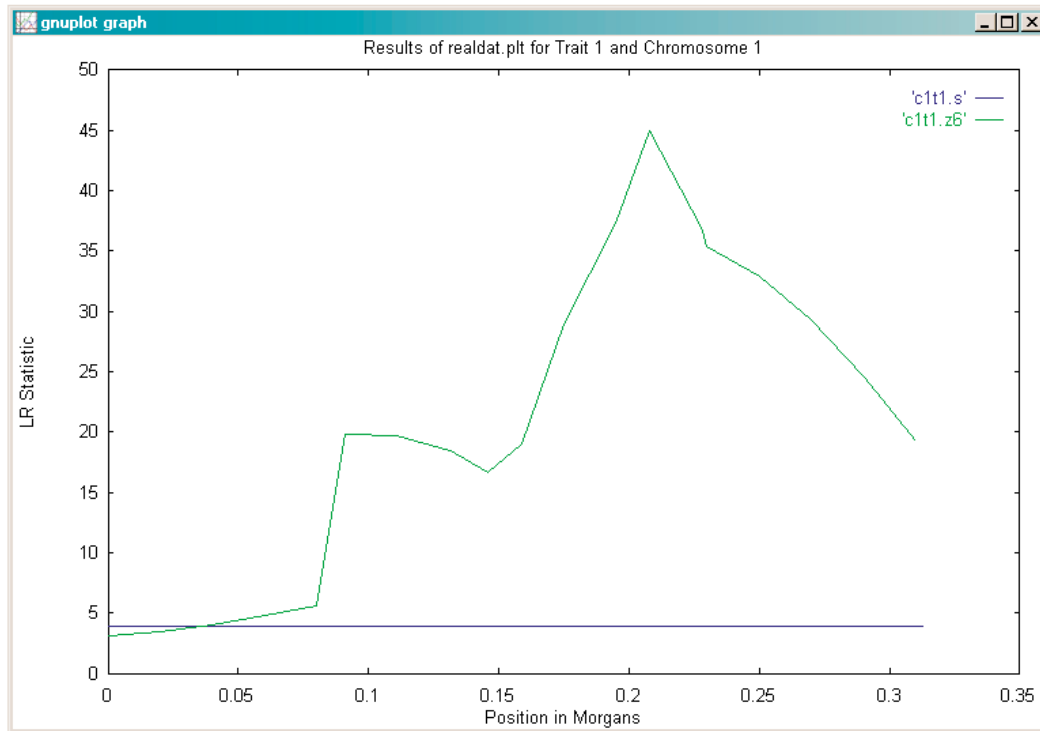
```

# for trait -named GAIN29 which is -number      1
#
#      #      ..Chrom..Markr.      .Position.      .Test Stat.      .Additive.      .Dominance.      .R2.      .TR2.      .S.
#      1      1      2      9.1100      19.7556      1.6445      -0.7939      0.1300      0.2145      2.3757
#      2      1      7      20.8100      44.8940      2.2573      -0.7565      0.2671      0.2724      1.0104

```

Preplot and GNUPlot

An even nicer way to look at the results of these analyses is to plot them. QTL Cartographer provides an easy way to do just that. Simply run preplot which will produce a GNUPlot control file. Then fire up GNUPlot, open the control file you just produced, and you should see something like this:



Since these data are from a published paper (*Genetics* 139:1737–1748; 1995), you might want to compare these results with those that they present.