

# QTL Analysis with QGene Tutorial

Phillip McClean

**1. Getting the software.** The first step is to download and install the QGene software. It can be obtained from the following WWW site:

<http://qgene.org>

Then select “Download”. From that next page, you should note the system requirements. Check to see that you have the Java Development Kit (JDK) installed. If you don’t, you will need to install it. Then click on the “**Go to download page**” button at the bottom of the page. Here select “**QGene 4.3.3**” to download a .zip file, and click on “**QGene manual**” to download another .zip file.

**2. Installing the software and creating a shortcut.** Once you have installed QGene, you want to use it. I typically make a desktop shortcut for software and click on it to launch it. Alternatively you can go to the directory in which you installed QGene and click on “**qgene.exe**”.

**3. The data file.** Next you need to launch the data file that you will work with. You can go to the following link to read about the file format.

[http://www.qgene.org/qgene/current/docs/manual/Data\\_format.html](http://www.qgene.org/qgene/current/docs/manual/Data_format.html)

Use the extension **.qdf** as the file extension for QGene data files.

On the next page you will see an example of an input file. This consists of three sections: [Header], [Locus], and [Trait]. The [Header] section describes the experiment (Study Name), the population type (Mating String), the symbols for the genotype at a locus (Genotype Symbols; see manual link above for explanation of the symbol order), and the names of the two parents in the cross (Parent1, Parent2)

The [Locus] section provides the genotypic data for the markers that will be used of the analysis. The number of lines in this section equals the number of markers. For each, the first three items correspond to the locus name, the linkage group on which it resides, and the genetic position on the linkage group for the specific locus. Following the marker names in each line is the genotype of each individual in the population. The symbols used here should correspond to those found in the Genotype Symbols section of the header. For each marker line in this section, the genotype of each individual in the population must be in the same order.

The [Trait] section provides the phenotype for each individual in the population. The name of the trait starts each line. The second symbol in each line is either **n**, **o**, or **m**. This designates whether the data is nominal, ordinal, or metric. In most cases, you data will be metric. Missing data are represented by a period (.).

[NOTE. The object of this exercise is for you to perform a QTL analysis. Only those features of QGene related to identifying QTL will be described here.]

**4. Analysis of phenotypic data.** You will have downloaded several example datasets with which to work. To load this into QGene, go to **File/Load Data**, find your file in your directory system, and click on **Open**. This will launch a **Data Manager** panel inside the main QGene window. For example, open the data file **cbb-Lg7-only.qdf**.

You will see three tabs in the right panel of the window: **Traits**, **Chromosomes**, and **Marker Data**. The first step here is to investigate the phenotypic distribution for your population. In the right panel, you will see a mini version of this. To see the data in more detail, go to **Analysis/Trait analysis**.

As with other QGene functions, when you select a function a new window appears. To maximize this, simply select the appropriate PC symbol in the upper right corner of this new window. You can explore the various drop down menus at a later time. For this example, you should look to see if your data is normally distributed. The common test of normality is the Kolmogorov-Smirnov test. You use the *P* value to determine if your data is normally distributed. If it is not, you can use the Transformation download menu to determine if transformed data is normally distributed.

```

[Header]
Study Name Lg7-only
Mating String r
Genotype Symbols AB...-
Parent1 Bat
Parent2 Jalo

[Locus]
m134 7 0.0 ABBAABABAAABBBAAAABBBABABAAABBABAAAAABAABABBBBAABBABAAABAABBBBAABBBBBBABBBA
CHI_INT3 7 10.8 BBBAABABBAABBBAAAABBBBABBBAAABBBAAAAABAABABBBBAABBBAAAAAABABAABBBBBBAABBBAA
m1719 7 19.4 AB-ABABBBBABBABAA-BBBABABBAABBBABABABAABABBABABBBAAAAABBAB-B--BBA-BBAABBBAA
DROF1b 7 24.1 AABAABABAABBBABAABBBBABBBABBBABABABABABB-AABBAAAAA---BA--ABBAB-AA-B-AA-
DBng199 7 26.7 ABBBABABAAABBBABAABBBBABABBAAABBABA-ABAABABBAABBAABAAAAABBABA--BBBAB--AAABAAA
m1615 7 29.3 -BBBABABAAABBABBAB-B---ABBAABBBABABABAABABBAABBAABAAAAABBABABBBBABBBAABBBAA
TA334Ind 7 29.3 ABBBABAB-AABBABBBBBAABBBABBBABABAABABBBAAABBBAAAAAABBABABBBBBAABBBAA
m501 7 33.1 ABBAABABAAABBBABABABBABABBAABBBBABABAABABBABABBBBAABBBABABBBBABBBAABBBAA
Pvcon3387 7 36.9 A-BBABAB-AAB-AB-ABABBABA-BAABBBBABABAABABBABAABBBAAAAAA-BBABABABBBAABBBAA
TA61Ind 7 37.6 ABBBABABAAABBBABABABBABBAABBBBABAB-ABABBABA--BAAAAABBBBABABBBBAABBBAA
m2129 7 38.3 ABBBABABAAABBBABAABBBBABABBBAABBBBABABAABABBABAABBBAAAAABBBBABABBBBAABBBAA
Pvcon4391 7 39.0 A-BBABABAABBBABAABAB-ABBAABBBBABABABAABBBBAABBBAAAAAABBBAABBBBAABBBAA
Pvcon8506 7 39.7 ABBBABABAAABBBAB-ABABBAA-BBAABBBBABABAABABBABAABBBAAAAAABBABABABBBAABBBAA
DBng060 7 41.9 AABBBBABAAABBBABAABABBAAABBBBBA-ABAABABBABABBBAAAAAABB--BABAB--AAABAAA
DPhs 7 43.7 AABBBBABAABBBABAABABBBAABBAABB-BBAABAABABBABABB-AAAABBBBABABABABAAABAAA
D1861 7 44.6 AABBB-ABAABBBABAAB-BBBBABBAABBBABBA-ABAABABBABA-BBAAAAAABB-ABABABABAAABA--
Pvcon732 7 46.0 ABBBBBABAABBBAB-ABABBBBABBAABBBAAABAABABBABABBBAAA-ABABBBBABABBBBAABBBAA
DP1090 7 51.8 AAABBBBA-AAABBBABAABBBBABBBBA-BABBA-ABAABABBA-ABBBAAAAAABAAB--BABABA-AAABAAA
m2531 7 55.9 AAABBBAAAAABBAAAAABBBBABBBABBABBAABAABABBBAAAAABBBAAAAAABAAAAAABAAAAAABAAA
m1378 7 59.7 AAABBBAAAAABBBAAAAAABBBBABBBBAABBBAAAAAABAABBBAAAAAABBBAAAAAABBA-BBAB-BAAAAAABAAA
m487 7 61.1 AAABBBABAABBBABAABABBABBABBABBBAAABAABBBAAAAAABBBAAAAAABBAABBAABAAAAAABAAA
DBng204 7 64.9 AAABBBABABBBBABAABABBABBBAAABABAA-ABA-BABBAAAAAABBAAB--AAAAAA-AAABBAAB
D0190 7 67.9 AAABB-ABABBBBABAAB-BABBABBABBABAA-BBBBABBBAAAAABBA-ABAAB-AAAAAB-AAA-AAB
m415 7 67.9 AAABBBBA-B--BBABA--ABABBABBABBA-AA-B-BABABAAAAABBA---BAABBA-AAAABAAAAABA
m2416 7 69.3 AAABBBABABBBAAAAAABAB-ABBB-BBABAABBBBABAB-AAAABBBAAAA-BABBBAAAAAABAAAAAAB
P_U3 7 73.4 AAA-BBAAB-BBBA-AAAA--BABBABBAAAAABBBAAAA-AAAABAAA--ABBAABA-AAAA-A-ABAAB
m1175 7 87.9 AAABB-ABAABBBAAAAAABABABAABAAAAAABAAAB-BBA--AA-AAAABBAABBAABBA---BAA-
m2459 7 98.4 AAABAAAAABBBBAAAAAABAAAAAABAAAAAABAAAAAABAAAAAABAAAAAABAAAAAABAAAAAABAAAA
m1065 7 107.7 AAABAAAAABBBBAAAAA-A-AAAB-BAABBAABAABBAABAAAAAABBAABAAAAAABAAAAAABAAA-AA-
m1357 7 111.3 AAABAAAAAABAAAAAABAAAAAABAAAAAABAAAAAABAAAAAABAAAAAABAAAAAABAAAAAABAAAA
m1853 7 116.0 AAABAA-BBAAA-A-A-A-A-AAABABAABBBAAAAAABAAAAA--AAAAAABAAAAAABBAABBAABBAAB
m1380 7 125.3 AAABAAAAAABAAAAAABAAAAAABAAAAAABAAAAAABAAAAAABAAAAAABAAAAAABAAAAAABAAAA
m290 7 128.5 --BAAAAAABAAAAA-A-A-AAABAAAAAABAAAAAABAAAAAABAAAAAABAAAAAABAAAAAABAAAA
m1947 7 141.6 AAABAAAAAABAAAAAABAAAAAABAAAAAABAAAAAABAAAAAABAAAAAABAAAAAABAAAAAABAAAA

```

```

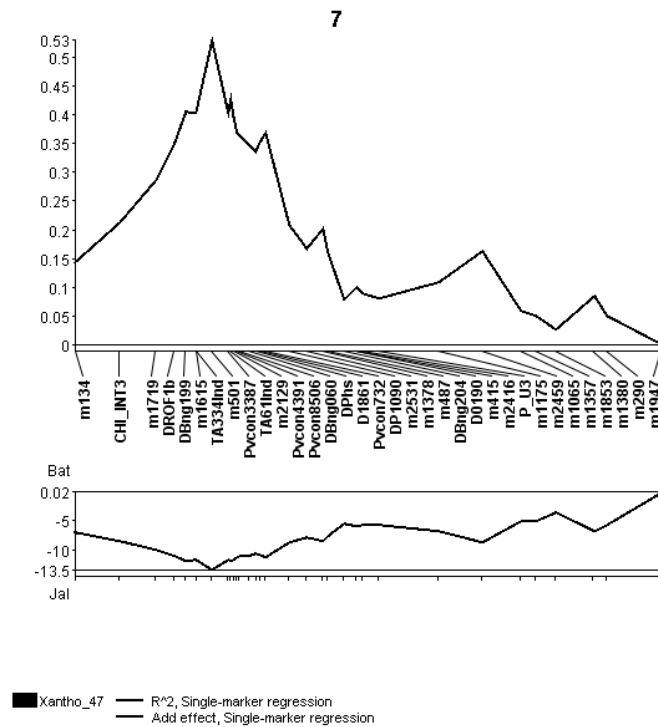
[Trait]
Xantho_47 M 53 53 72 79 43 68 24 85 46 32 62 83 78 68 53 74 67 74 36 71 65 67 76 22 84 74 36 58
79 57 72 53 82 18 53 53 73 60 53 74 23 69 76 58 82 67 57 75 80 34 25 46 65 69 71 40 71 35 79 26
78 76 85 42 66 74 46 39 45 65 85 . 29 46
Xantho_51 M 77 58 81 86 65 76 31 82 71 49 70 90 82 74 64 76 76 86 59 90 83 70 79 36 79 86 48 76
80 73 78 77 82 60 78 64 77 71 56 77 63 70 81 53 83 66 66 85 83 74 55 63 72 74 79 70 82 46 89 76
86 82 81 61 81 72 71 66 66 78 81 . 57 56

```

**5. Beginning a QTL analysis.** Now let's do a QTL analysis. To begin the analysis, you go to the **Analysis/QTL mapping** drop down menu. As with other QGene functions, this action will launch the "QTL mapping" window. The first step is to select a chromosome. In this example, you can just simply click on **7** in the **Chromosomes** section of the window. This will generate the linkage map from the marker data found in the data file.

The next step is to perform a QTL analysis. To do this, you must first select a trait to be analyzed in the **Traits** section of the display. For the purposes of this tutorial, we are only selecting the Xantho\_47 trait. (This trait is resistance score in a common bean population in response to infection by race 47 of *Xanthomonas*.) Then you need to select a specific type of analysis. You will find many of the popular methods listed in the right hand section of the display. For this example, let's compare the results from single-marker regression, simple interval mapping and composite interval mapping.

**6. Single-marker regression.** Single marker regression looks at each individual marker and essentially performs a one-way analysis of variance. Multiple statistics can be collected. Here, if you select "Single-marker regression" in the right side panel, you have multiple statistics that can be reported. For this analysis, let's choose R<sup>2</sup> and Add effect (additive effect). You will see the following figure when you perform these analyses.



**(Display Options.** I personally can better visualize the results if the background is white and the drawing is in black. To get this type of display, in the new Analysis/QTL mapping window, choose **Options/Reverse Video**. To get a black line, double click on the trait in the Traits section of the window and select a new color. Typically I choose black. To change the line format, double click on the line format box to the right of the test statistic you wish to display and select a new format. I prefer an unbroken line.)

The upper portion shows the  $R^2$  for each marker on the Y-axis. These values are then connected by a line. The lower portion shows the additive effect for each marker. A positive value means the presence of the allele from a particular parent increases the value of the phenotype while a negative value means the phenotype from a specific parent decreases the phenotypic value among the lines in the population.

Although this provides a graphical representation, the results of single-marker regression are typically presented in a table. The QGene results can also be saved in text format and then copied into and displayed in software like MS Excel. To do this, you select **File/Export text to clipboard** from the “QTL mapping window”. Next open MS Excel and paste the data into a worksheet. Table 1 is an example of the data. Table 1 shows the results of the analysis in table format. (**Note:** The  $-\log p(F)$  is added. This is a way of reporting the probability of the F test. For reference the  $-\log p(F)$  for 0.05, 0.01, and 0.001 are 1.30, 2.00, and 3.00 respectively.)

**Table 1.** Test statistics for single-marker regression.

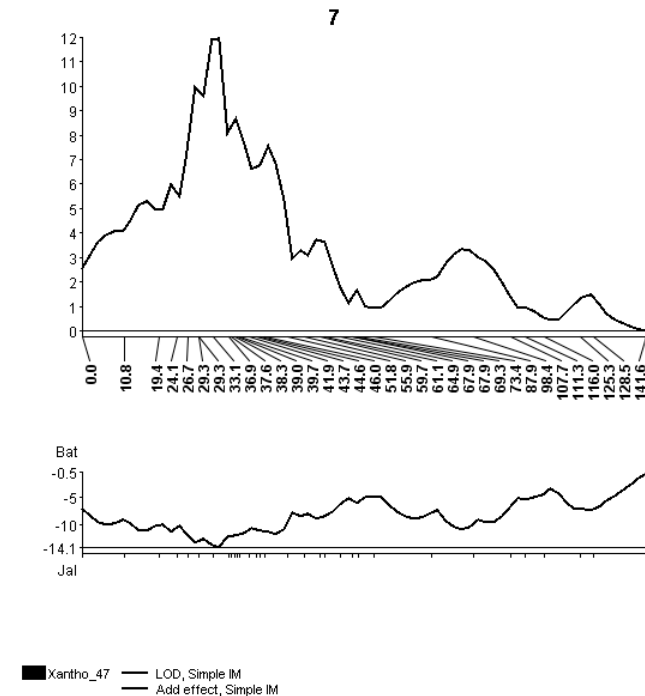
Chromosome	Locus name	Position	Xantho_47	Xantho_47	Xantho_47
			SMR (-log p(F))	SMR (Add effect)	SMR (R <sup>2</sup> )
7	m134	0	3.012	-6.955	0.143
7	CHI_INT3	10.8	4.398	-8.493	0.213
7	m1719	19.4	5.959	-9.937	0.286
7	DROF1b	24.1	7.467	-10.989	0.351
7	DBng199	26.7	8.811	-11.807	0.404
7	TA334Ind	29.3	8.774	-11.678	0.402
7	m501	33.1	12.441	-13.418	0.527
7	Pvcon3387	36.9	8.789	-11.685	0.403
7	TA61Ind	37.6	9.391	-12.009	0.425
7	m2129	38.3	8.642	-11.596	0.397
7	Pvcon4391	39	7.875	-11.149	0.367
7	Pvcon8506	39.7	7.755	-11.077	0.362
7	DBng060	41.9	7.372	-10.989	0.347
7	DPhs	43.7	7.073	-10.708	0.334
7	D1861	44.6	7.476	-10.904	0.351
7	Pvcon732	46	7.849	-11.148	0.366

7	DP1090	51.8	4.28	-8.63	0.207
7	m2531	55.9	3.472	-7.777	0.167
7	m1378	59.7	4.173	-8.498	0.202
7	m487	61.1	3.338	-7.56	0.16
7	DBng204	64.9	1.788	-5.352	0.079
7	m415	67.9	2.194	-5.889	0.1
7	m2416	69.3	1.962	-5.522	0.088
7	P_U3	73.4	1.794	-5.524	0.079
7	m1175	87.9	2.352	-6.731	0.108
7	m2459	98.4	3.367	-8.569	0.161
7	m1065	107.7	1.413	-5.011	0.059
7	m1357	111.3	1.246	-4.888	0.05
7	m1853	116	0.765	-3.469	0.026
7	m1380	125.3	1.885	-6.76	0.084
7	m290	128.5	1.254	-5.625	0.051
7	m1947	141.6	0.002	0.014	0

While remembering that this data set contains markers along a single chromosome in the correct map order, you should notice that almost all of the markers are significant, many at a  $P$  value  $< 0.001$ ). Therefore, we need to determine if all of these markers have independent effects, whether they are all part of one or several QTL, and what is the markers that border the one (or many) QTL. To address these, further analyses, simple interval mapping and composite interval mapping will be performed.

**7. Simple interval mapping.** Again, go to the right panel of the “QTL mapping” window, deselect “Single-marker regression, and then click on the “+” symbol next to **Simple IM**. This is the simple interval mapping function. When this is expanded, you will see different types of test statistics for this QTL mapping approach.

To perform the actual QTL mapping, you then select a specific statistic. Select the **LOD** statistic, and the data will be analyzed to discover QTL regions (if any) from this data set. This is the display you should see.



As with all analysis in QGene, you can save the data and paste it into MS Excel for viewing. Table 2 provides that data for the simple interval mapping that was just completed. Whereas the test statistics were reported for each marker in the simple regression analysis table, with interval mapping those statistics are reported every 2 cM. Why 2 cM? That is the step-wise interval for the analysis. The names of only those markers that map at a specific interval position are reported.

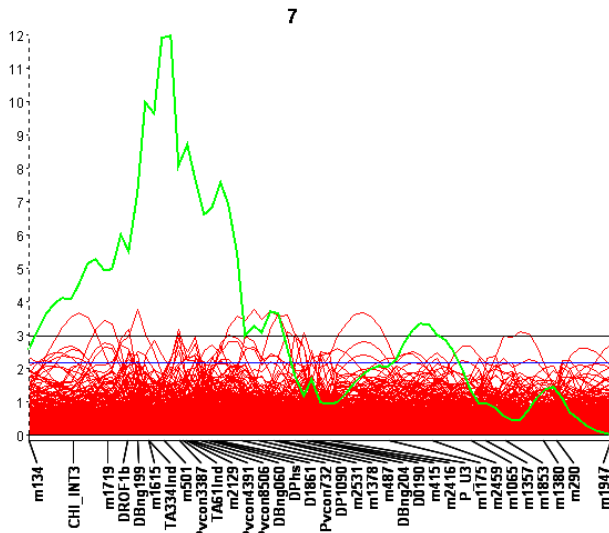
Here you can see two peaks, a major peak on the left end and minor peak on the right end. The important question, to address is whether these significant peaks. The way to calculate the significant LOD value is to perform a permutation test. To perform the test, select **Resampling/Permutation** from the menu part of the “QTL mapping”

**Table 2.** Test statistics for simple interval mapping.

Chromosome	Locus name	Position	Xantho_47	Xantho_47	Xantho_47
			SIM (-log p(F))	SIM (Add effect)	SIM (LOD)
7	m134	0	3.012	-6.955	2.445
7		2	3.588	-8.381	3.002
7		4	4.049	-9.365	3.45
7		6	4.366	-9.81	3.761
7		8	4.515	-9.667	3.908
7		10	4.473	-8.939	3.867
7		12	4.922	-9.644	4.309
7		14	5.513	-10.879	4.895
7		16	5.662	-11.062	5.042
7		18	5.333	-10.15	4.716
7		20	5.351	-9.815	4.734
7		22	6.344	-11.259	5.722
7		24	5.865	-10.003	5.244
7		26	7.557	-11.594	6.937
7		28	10.098	-13.064	9.499
7		30	9.79	-12.506	9.187
7		32	11.932	-13.626	11.358
7		34	11.996	-14.022	11.423
7		36	8.33	-12.167	7.714
7		38	8.909	-11.909	8.297
7		40	7.944	-11.408	7.326
7		42	6.929	-10.65	6.307
7		44	7.098	-10.95	6.476
7	Pvcon732	46	7.849	-11.148	7.23
7		48	7.194	-11.726	6.573
7		50	5.785	-10.712	5.165
7		52	3.397	-7.679	2.816
7		54	3.71	-8.473	3.12
7		56	3.518	-7.869	2.934
7		58	4.137	-8.9	3.537
7		60	4.058	-8.469	3.459
7		62	3.055	-7.48	2.487
7		64	2.22	-6.284	1.696
7		66	1.569	-5.198	1.1
7		68	2.098	-5.914	1.582
7		70	1.385	-4.8	0.937
7		72	1.338	-4.989	0.896
7		74	1.355	-4.943	0.911
7		76	1.672	-6.347	1.193
7		78	1.972	-7.576	1.466
7		80	2.228	-8.452	1.703
7		82	2.412	-8.827	1.876
7		84	2.503	-8.634	1.962
7		86	2.494	-7.914	1.953
7		88	2.676	-7.352	2.125
7		90	3.195	-9.13	2.622
7		92	3.588	-10.357	3.001
7		94	3.785	-10.775	3.193
7		96	3.741	-10.27	3.151
7		98	3.452	-8.921	2.87
7		100	3.284	-9.368	2.708
7		102	2.949	-9.45	2.386
7		104	2.417	-8.471	1.88
7		106	1.788	-6.666	1.297
7		108	1.326	-5.073	0.885
7		110	1.316	-5.348	0.877
7		112	1.188	-4.987	0.766
7		114	0.91	-4.404	0.534
7	m1853	116	0.765	-3.469	0.42
7		118	0.775	-4.193	0.427

7	120	1.151	-5.875	0.735
7	122	1.513	-6.982	1.05
7	124	1.785	-7.164	1.295
7	126	1.881	-7.218	1.383
7	128	1.499	-6.638	1.038
7	130	1.038	-5.494	0.64
7	132	0.785	-4.734	0.436
7	134	0.553	-3.658	0.263
7	136	0.357	-2.496	0.134
7	138	0.2	-1.423	0.052
7	140	0.076	-0.537	0.009

window. This will open a dialog box. The box is already set to perform the test based on the QTL Mapping Method (here simple interval mapping) and report the test statistic (LOD) that was used in your original analysis. You can use the drop down methods here if you wish to change these. Here, you should not change these. Finally, you can change the number of permutations in the upper right corner. Now perform the test, and you should see something like this.



There are two points to notice when the analysis is complete. First, all of the red lines are the results of the many permutation tests. Secondly, two lines have been added. The top line refers to the significant LOD score for the  $\alpha_{0.01}$  experiment-wide error, and the lower line refers to the LOD score for the  $\alpha_{0.05}$  experiment-wide error. The actual values are reported in the upper panel of the “Resampling” window. From my run, those values were  $\alpha_{0.01}=2.955$ , and  $\alpha_{0.05}=2.153$ .

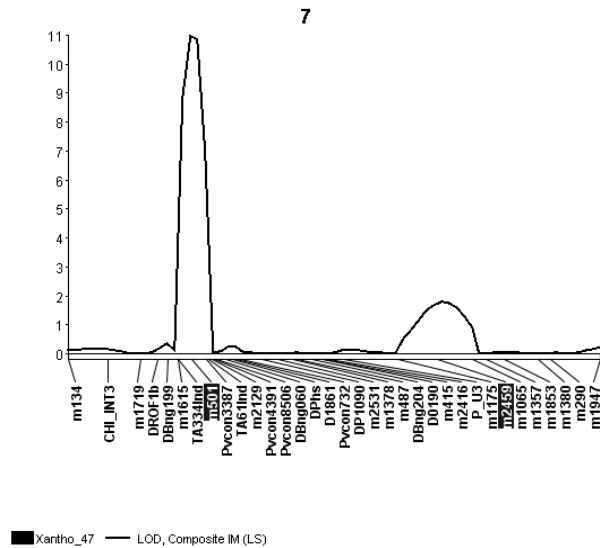
**8. Composite interval mapping (CIM).** Lander and Bostein (Genetics, 1989. 121:185) were the first to introduce interval mapping. But shortly after that publication, statistical issues were recognized with the technique. This is best illustrated by a quote from Zeng (PNAS, 1993: 90:10972).

“The test statistic of an interval can be affected by QTLs located at other regions of the chromosome. Even though there is no QTL within an interval, the test statistic on the interval can still be very significant if there is a QTL at some nearby point on the chromosome. Moreover, if there is more than one QTL on a chromosome, the test statistic at a testing position will be affected by all those QTLs, and the estimated positions and effects of “QTLs” identified by this method are likely to be biased.”

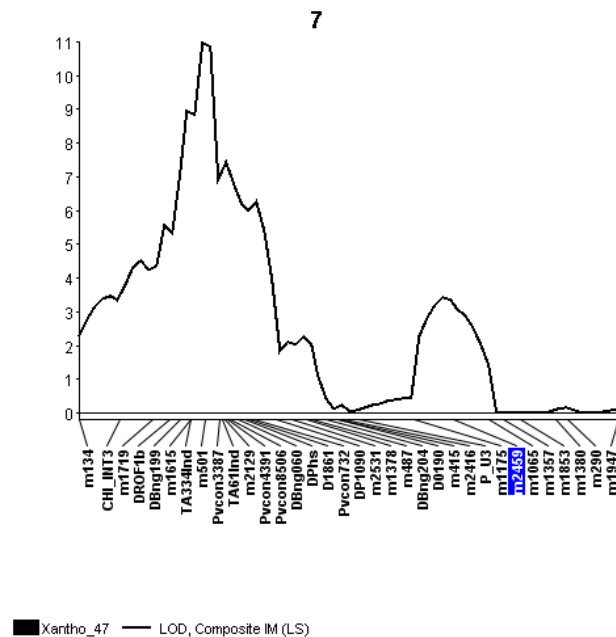
Composite interval mapping, first described by Zeng (Genetics, 1994. 136:1457), is a refinement of the simple interval mapping (SIM) method. With SIM, a QTL effect is searched in the interval between two markers. But QTL linked to markers that fall outside the interval may affect the position and level of effect of any QTL within the interval. With CIM, markers outside the interval are considered cofactors. By removing the effects of these cofactors, the location and effect of a QTL within the interval can be better estimated. The major challenge is to select the cofactors. QGene allows you to select cofactors one of two ways. First, you can use the default parameters and let QGene select the cofactor. This is often the case in published papers. Alternatively, you can

select those factors and determine if they improve your QTL discovery. So how do you do these alternative steps in QGene.

The first step is make sure you have your chromosome and trait selected. Then choose CIM and the appropriate test statistic. This will provide the results for the interval mapping portion of CIM. To determine the effect of neighboring loci, you must first set the parameters that control the selection of cofactors. This is done by selecting **Options/Select cofactors** in the header of the QTL mapping window. Without a full understanding of the approach to cofactor selection, it is always safe to use the program defaults. Use Forward cofactor selection and uncheck the “All QTL cofactors instead of markers” box. If you set the other parameters to 0, the program will select the cofactors. Here is the image that was generated.

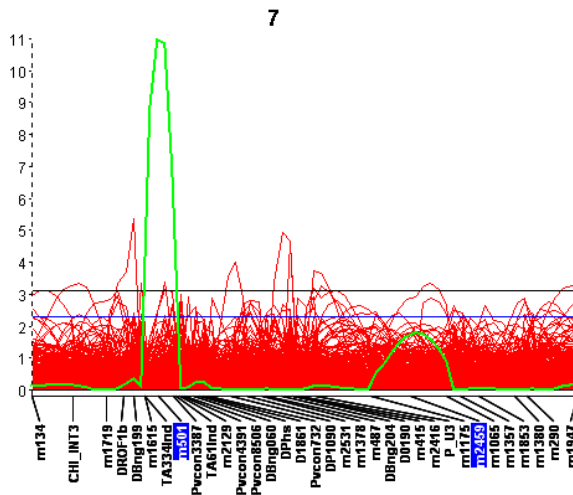


Now we want to use QGene to demonstrate the value of cofactor selection in CIM. The markers used as cofactors are highlighted. Now click on m501. This removes it as a cofactor. Notice the difference in the figure below.



The important observation is that the position of the QTL is less clear. And while the effect (LOD value) of the first QTL is essentially unchanged, the magnitude (breadth of the effect ) of QTL is increased. This clearly demonstrates that CIM is preferred because it can better locate the QTL and more accurately measure the effect.

Now that you have completed the demonstration of the advantage of CIM, let's determine the LOD score for which a QTL will be considered significant. Again, this is done by performing permutation test. Use the method described above and the same default parameters.



Notice now that there is one major peak on the left portion and a second near the right of the linkage group. Also notice that the  $\alpha_{0.01}=3.127$ , and  $\alpha_{0.05}=2.285$ . With respect to the experiment-wide error rate, the peak on the right side is not significant. Therefore you can conclude that there is only QTL on this chromosome, and that it maps between m1615 and m2129. Also notice how these results are different from the SIM analysis.

**9. Conclusion.** Using another feature of QGene, we can also determine the distance between these two markers. Again, go back to the “Data Manager” window. Now select the **Chromosomes** tab. Here your linkage group will be displayed. Now find m1615 and you can see that it is located a map position 29.3. Similarly m2129 is located at position 38.3. Therefore these two markers are 9.0 cM apart. Your conclusion is that within this 9.0 cM interval you should be able to locate a gene which affects resistance to Xanthomonas race 47. To discover that gene will require a map-based cloning experiment.

**10. Expanding the analysis to a whole genome.** This example was designed to show a QTL analysis of a single linkage group. This group was selected because it was known that a major QTL was found on this linkage. More often, your will data will be from all linkage groups, rather than a single. To demonstrate whole genome QTL analysis, close out the current data set (by closing QGene), load the **cbb-all-LG.qdf** data file, and opening the “QTL mapping” window. First, you should notice that 11 chromosomes are listed. Now select all of the chromosomes, pick the Xantho\_47 trait, perform a CIM analysis, and determine the  $\alpha_{0.01}$  and  $\alpha_{0.05}$  significance levels.

A final point to mention is in regards to the permutation test to set the significant value for different  $\alpha$  levels. As implemented in QGene, this is a genome-wide test. If a region of the linkage map is skewed, then the significant value will be conservative (too high) within that region. Although QGene cannot compensate for that you should be aware of it. One work around is to create a .qdf input file for just that chromosome. For example, the whole genome analysis also detected a QTL for CBB tolerance on LG 9. Following a CIM analysis, a permutation test was performed only on chromosome 9, and the  $\alpha_{0.01}=3.465$ , and  $\alpha_{0.05}=2.215$ . This compares to the  $\alpha_{0.01}=3.465$ , and  $\alpha_{0.05}=2.215$  values for the whole genome permutation test. You might wish to consider mapping just individual chromosomes to determine your significance value. You can perform the linkage group 9 analysis using the data file. **cbb-LG9-only.qdf**.