

Guide to *QTL Miner*

Rex Bernardo, University of Minnesota

Purpose

QTL Miner is software for **identifying markers useful for selection** on the basis of phenotypic, marker, and pedigree data available in a breeding program for self-pollinated crops. *QTL Miner* was developed as part of the USDA-funded Barley Coordinated Agricultural Project.

Installation

1. Download the *QTLMinerSetup.exe* file from <http://www.BarleyCAP.org>.
2. Run the program and follow the installation instructions. You may need to disable firewalls or antivirus programs before installation.
 - The program does not produce an icon on the desktop. Specify the folder where the program is to be installed.
3. The setup program installs a *QTLWindow.exe* file. Run *QTL Miner* by double-clicking this file.

Input Files

QTL Miner utilizes four input files. Each file is in tab-delimited (.txt) format. These input files are case-sensitive and should NOT contain any spaces.

I. Pedigree file

1. The first row is a header row. The list of inbreds begins in the second row.
2. The columns are separated by tabs and are organized as follows:
 - Name of the Inbred
 - Name of the first parent of the Inbred (i.e., Parent 1)
 - Name of the second parent of the Inbred (i.e., Parent 2)
 - Parental contribution of Parent 1 to the Inbred
 - Parental contribution of Parent 2 to the Inbred

Inbred	Parent1	Parent2	P1Contrib	P2Contrib
MOREX	CREE	BONANZA	0.5	0.5
M28	CREE	BONANZA	0.5	0.5
ROBUST	MOREX	MANKER	0.5	0.5
MN72-146	MANKER	M28	0.5	0.5
MN77-825	ROBUST	MN72-146	0.5	0.5
EXCEL	ROBUST	MN77-825	0.5	0.5
STANDER	MN80-224	EXCEL	0.5	0.5

3. The parental contributions can be based on either pedigree data (e.g., 0.50 for F₂-derived inbred, and 0.75 or 0.25 for BC₁-derived inbred) or marker data.
4. The sum of the parental contributions should not exceed 1.0 but can be less than 1.0.
5. If one of the parents is unknown, then it should be listed as "Unknown" and the corresponding parental contribution should be zero.

II. Phenotypic data file

1. The first row is a header row. The phenotypic records begin in the second row.
2. The columns are separated by tabs and are organized as follows:
 - Name of the Experiment
 - Name of the Inbred (format corresponds to inbred names in pedigree file)
 - Mean or value of the Inbred for the trait
 - Number of locations that correspond to the mean or value for the trait

Experiment	Inbred	Yield	Nloc
MN-2007	MOREX	4.45	18
MN-2007	ROBUST	4.61	18
MN-2007	STANDER	5.27	18
ND-2008	ROBUST	5.00	9
ND-2008	EXCEL	5.82	9
ND-2008	STANDER	5.79	9

III. Marker data file

1. The first row is a header row.
2. The names of the markers are given in the header row, beginning in the second column.
3. The columns are separated by tabs and are organized as follows:
 - Name of the Inbred
 - Genotype (coded as 1, 0, or -1) of the Inbred for the marker designated in the header
 - (Genotype for the second marker, third marker, ..., last marker)

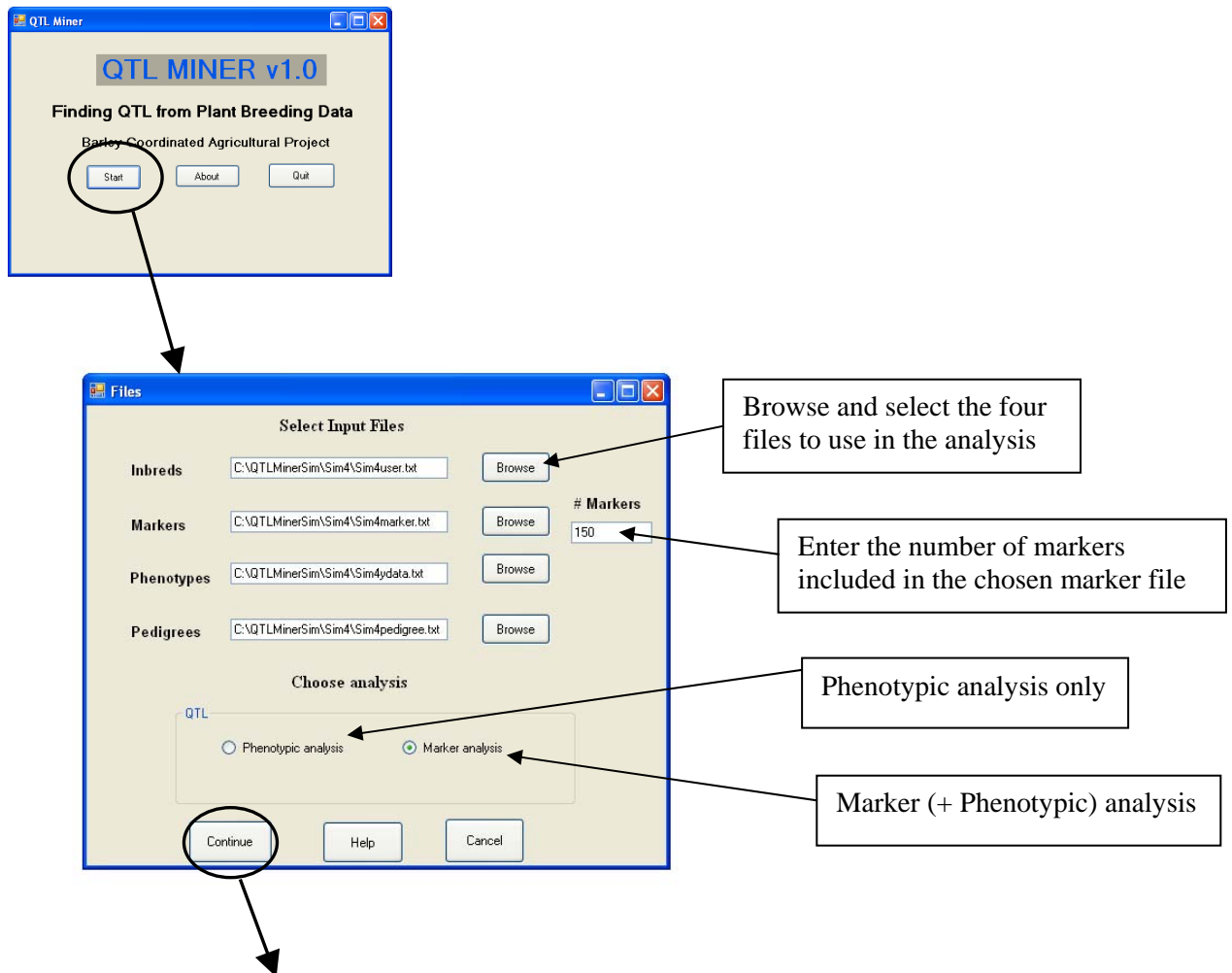
Inbred	1H-1973-796	1H-4226-570	3H-5646-568
EXCEL	-1	1	1
MOREX	-1	1	-1
ROBUST	-1	1	1
STANDER	-1	1	1
WA1614-95	-1	0	-1

IV. User inbreds file

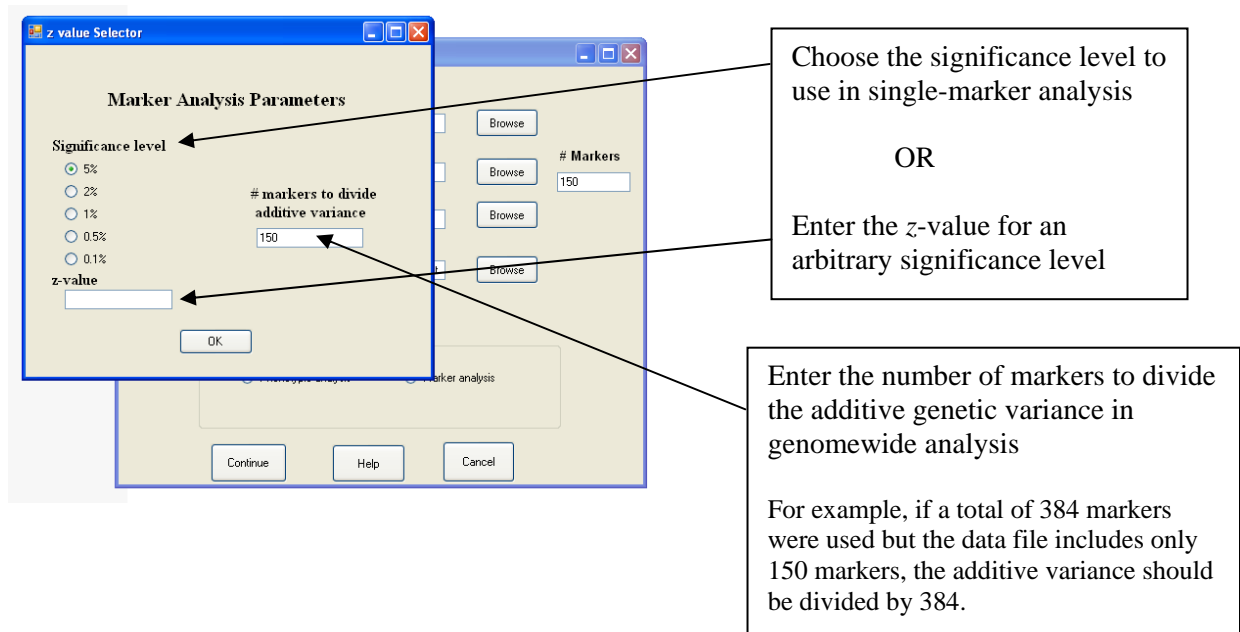
1. This file contains the names of the Inbreds to be included in a particular analysis.
2. No header row is needed.

MOREX
ROBUST
EXCEL
STANDER
MN-BRITE

Running OTL Miner



(See next page)



QTL Miner Analysis and Results

1. The output file is *(TraitName)_Output.txt*. The names of the input files and the numbers of inbreds, records or observations, and experiments are given.
2. Output from phenotypic analysis:
 - Mean of each experiment
 - Estimates of additive variance, residual variance, and entry-mean heritability
 - Grand mean
 - Breeding values of inbreds. The breeding value is a prediction of the genetic merit of each inbred and accounts for the heritability (h^2) of the trait. For example, if $h^2 = 0$, then the breeding values are all equal to the grand mean. If the inbreds are all unrelated, the data are balanced, and $h^2 = 1.0$, then the breeding value is equal to the observed mean performance of the inbred.
3. Output from genomewide analysis:
 - Marker effects are random and are obtained by best linear unbiased prediction
 - All markers are included in the analysis
 - No significant tests are conducted
 - The breeder will improve the trait by selecting for all markers
 - See the following reference for details:

Bernardo, R., and J. Yu. 2007. Prospects for genomewide selection for quantitative traits in maize. *Crop Sci.* 47:1082-1090

4. Output from single-marker analysis

- Marker effects are fixed and are obtained by mixed-model analysis
- Relatedness among inbreds is accounted for
- Single-marker significance tests are conducted with the specified comparison-wise significance level
- See the following reference for details:

Arbelbide, M., and R. Bernardo. 2006. Mixed-model QTL mapping for kernel hardness and dough strength in bread wheat. *Theor. Appl. Genet.* 112: 885-890.

QTL Miner Simulator

QTLMinerSimulator.exe is software for creating virtual barley inbreds and test files for use in *QTL Miner*.

1. Data sets are generated for a given number of inbreds, number of QTL, number of markers, and trait h^2 .
2. The data sets can then be analyzed in *QTL Miner* to study the usefulness of the procedures in *QTL Miner* for finding markers useful for selection.
3. To run ***QTLMinerSimulator***:
 - Create the following folder: c:\QTLMinerSim\
 - Download the ***QTLMinerSimulator.exe*** file from <http://www.BarleyCAP.org> and copy this file to the c:\QTLMinerSim\ folder
 - Double click ***QTLMinerSimulator.exe*** to run the program
 - The software will ask whether the QTL should be placed randomly in the barley genome or whether they should be in fixed locations in the genome. If the QTL locations are fixed, the program will ask for one input CSV file with the cM positions for each of the QTL. A sample CSV file for fixed locations of seven QTL is as follows:

```
70
219
376
522
680
842
991
```

4. The QTL effects follow a geometric series, where QTL1 has the largest effect, QTL2 has the second largest effect, and the last-numbered QTL has the smallest effect. The inbreds are assumed evaluated in eight experiments.
5. In addition to producing the four test files for *QTL Miner*, ***QTLMinerSimulator*** produces an output file with genotypic values of the inbreds (c:\QTLMinerSim\genval.csv) and an output file

with locations of the markers and QTL (c:\QTLMinerSim\geneloc.csv). In the latter file, the markers are the loci with negative numbers while the QTL are the loci with positive numbers.

6. Missing data can be simulated by manually deleting rows at random in the phenotypic data file.

Issues and Tips for Using *QTL Miner*

1. *QTL Miner* is designed to run on Windows XP.
2. *QTL Miner* is designed to use input files generated by THT in Barley CAP. Tab-delimited TXT files can be generated on a spreadsheet, but *Microsoft Excel* is known to insert extra spaces when files are converted to a tab-delimited format. Files therefore need to be searched for extra spaces and these spaces need to be deleted. One way of doing this is through *Search and Replace* in *Microsoft Word*: search for a <space> and replace it with <delete> (i.e., press the Delete key).
3. Also, make sure there are no blanks after the last column or row. Suppose your spreadsheet has N_R rows and N_C columns. In *Microsoft Excel*, delete column ($N_C + 1$) and delete row ($N_R + 1$) before you export the spreadsheet as a TXT file.
4. The limits are unknown regarding the size of the data sets that can be analyzed in *QTL Miner*, and this is something that users would need to learn by trial and error. Try experimenting with fewer numbers of markers as needed (e.g., 96 then working with higher increments).