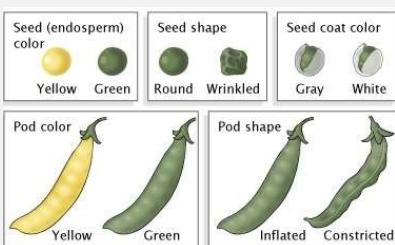


QTL MAPPING & ANALYSIS

Sujan Mamidi

Qualitative traits

1. Monogenic or oligogenic
2. Discrete phenotypic classes (nominal scale).
3. Typically, environmental effect on trait expression is absent or low
4. Discontinuous variation
5. Genes have large effect
6. Mapped as visible marker

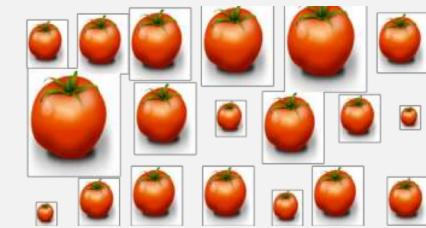


www.nature.com

Quantitative traits

1. Polygenic (quantitative trait loci)
2. Continuum of measures (interval scale).
3. Trait expression may show profound environmental effect
4. Continuous variation
5. Genes have smaller effects
6. Mapping requires QTL analysis

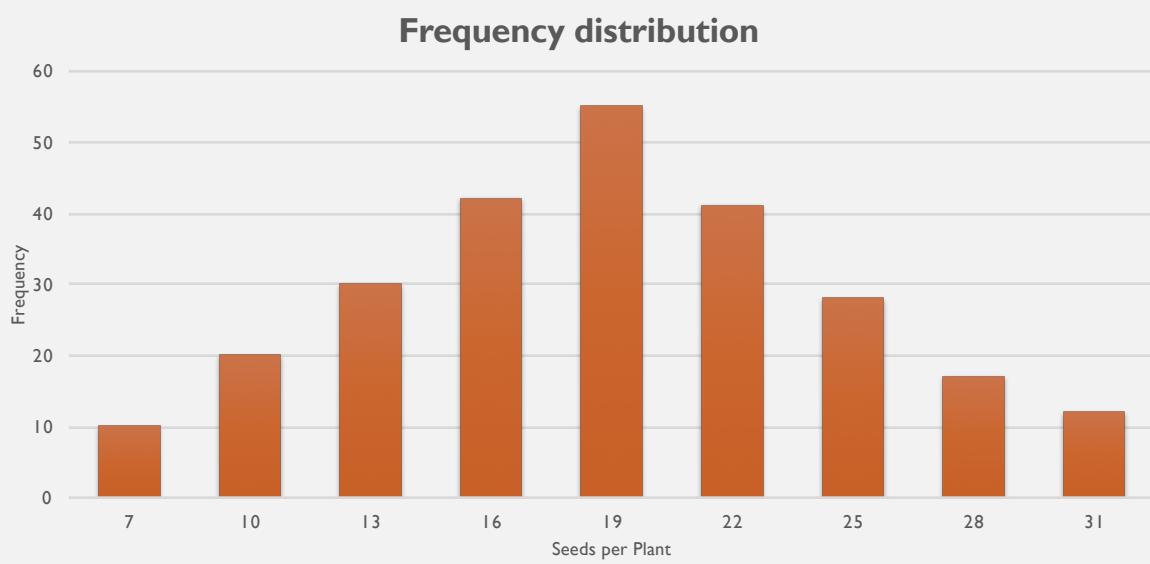
ExFruit shape: a quantitative trait



cubocube.com

QTL MAPPING

- Many agriculturally important traits such as yield, quality and some forms of disease resistance are controlled by many genes and are known as “**quantitative traits or polygenic or multifactorial or complex traits**”.
 - These traits show continuous variation in a population.
 - These traits **do not** fall into **discrete classes**.



Ex: Continuous distribution

Quantitative Trait Loci (QTL)

- Region of the genome that is associated with a quantitative trait.
- It can be a single gene or cluster of linked genes.

QTL characteristics

- These traits are controlled by multiple genes, each segregating according to Mendel's laws.
- These traits can be affected by environment to varying degrees.
- Individual gene effects is small
- The genes involved can be dominant, or co- dominant.
- The genes involved can be subject to epistasis or pleiotrophic effect.

QTL MAPPING

Identifying genomic regions associated with trait variation is known as **QTL mapping**. It involves testing DNA markers throughout the genome for the likelihood that they are associated with a QTL.

- QTL analysis is usually undertaken in segregating mapping populations.
- Genes and markers segregate via chromosome recombination during meiosis,

OBJECTIVES IN QTL MAPPING

The basic objective is to detect QTL, while minimizing the occurrence of false positives (Type I errors, that is declaring an association between a marker and QTL when in fact one does not exist).

- To identify the regions of the genome that affects the trait of interest.
- To analyze the effect of the QTL on the trait.
- How much of the variation for the trait is caused by a specific region?
- What is the gene action associated with the QTL – additive effect?
Dominant effect?
- Which allele is associated with the favorable effect?

PREREQUISITES FOR QTL MAPPING

- Availability of a good linkage map
- A **segregating population** derived from parents that differ for the trait(s) of interest
- Trait/Traits – Phenotypic values

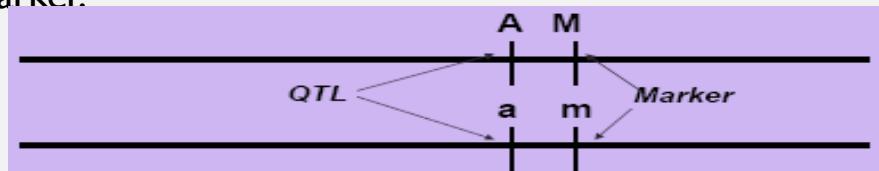
Methods to detect QTLs

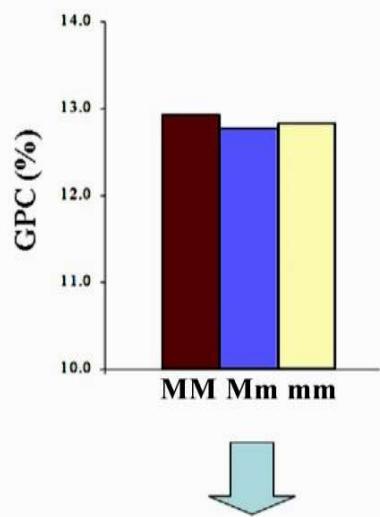
- Single-marker analysis (SMA)
- Simple interval mapping (SIM)
- Composite interval mapping (CIM)
- Multiple Interval Mapping (MIM)
- Bayesian Interval Mapping (BIM)
- And many more

Single-Marker Analysis (SMA)

Also known as single-point analysis. It is the simplest method for detecting QTLs associated with single markers.

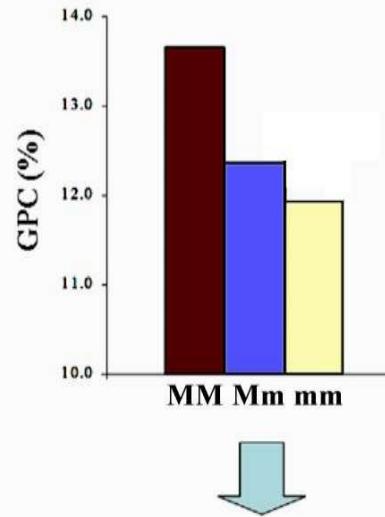
- This method **does not require a complete linkage map** and can be performed with basic statistical software programs.
- The statistical methods used for single-marker analysis **include t-tests, analysis of variance (ANOVA) and linear regression.**
- Linear regression is most commonly used because the coefficient of determination (R^2) from the marker explains the phenotypic variation arising from the QTL linked to the marker.





Non significant

No
QTL near such marker



F-test is significant

There is a QTL at or near
such marker

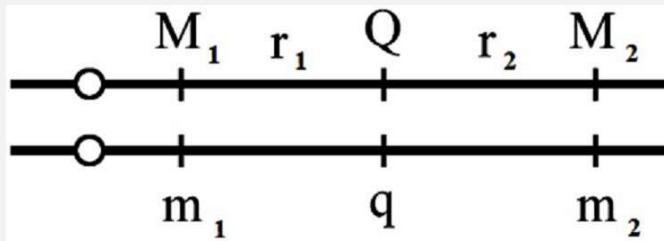
Limitations

- Likelihood of QTL detection significantly decreases as the distance between the marker and QTL increases
- It cannot determine whether the markers are associated with one or more markers QTLs
- Only good with a high dense map

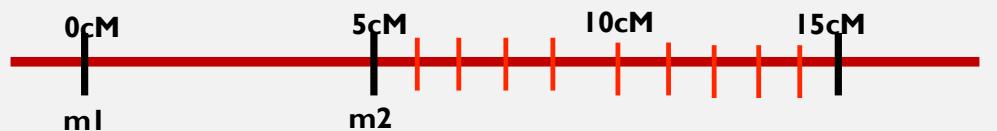


Simple Interval Mapping (SIM)

- It was first proposed by **Lander** and **Bolstein**.
- It takes full advantages of the linkage map.
- This method evaluates the target association between the trait values and the genotype of a hypothetical QTL (target QTL) at multiple analysis points between pair of adjacent marker loci (target interval).



- Presence of a putative QTL is estimated if the log of odds ratio exceeds a critical threshold.
- The principle behind interval mapping is to test a model for the presence of a QTL at many positions between two mapped



Statistical method used for SIM

Maximum Likelihood Approach

It is assumed that a QTL is located between two markers,

- Searches for the best approximation for quantitative trait distribution that are observed for each marker class.
- Models are evaluated by comparing the likelihood of the observed distributions with and without finding QTL effect

Logarithm of the odds ratio (LOD score):

- Linkage between markers is usually calculated using odds ratio.
- This ratio is more conveniently expressed as the logarithm of the ratio, and is called a logarithm of odds (LOD) value or LOD score.
- LOD values of >3 are typically used to construct linkage maps.

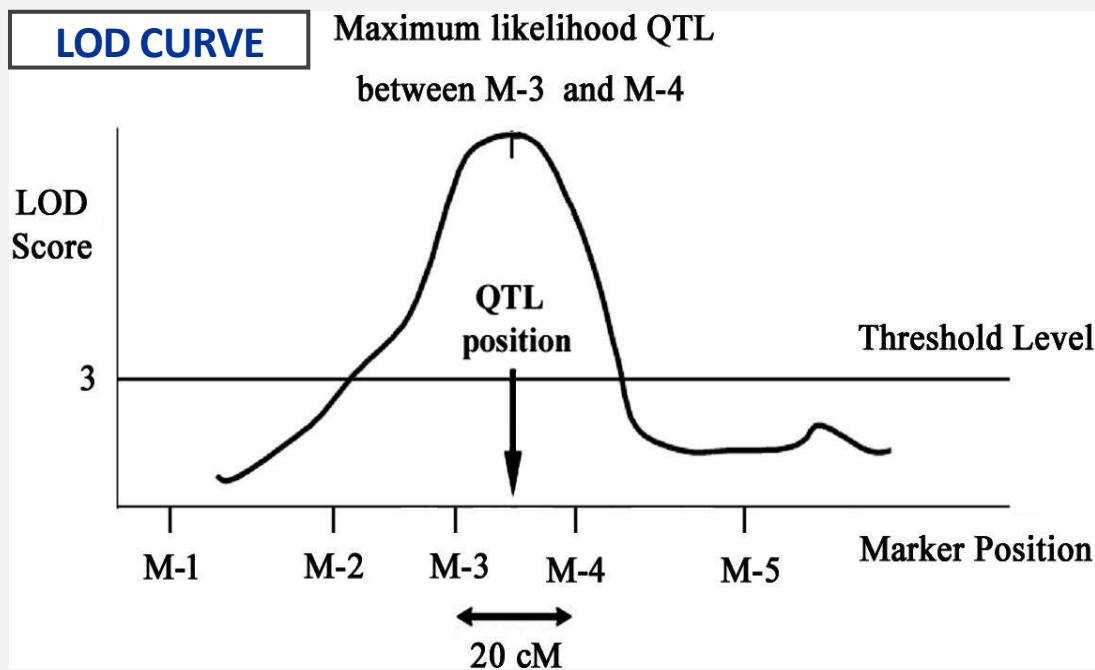
$$\text{Odds ratio} = \frac{\text{probability of the data occurring with a QTL}}{\text{probability of the data occurring with no QTL}}$$

- LOD of 2 means that it is 100 times more likely that a QTL exists in the interval than that there is no QTL.

- LOD of 3 between two markers indicates that linkage is 1000 times more likely (i.e. 1000:1) than no linkage.
- LOD values may be lowered in order to detect a greater level of linkage or to place additional markers within maps constructed at higher LOD values.

The LOD score is a measure of the strength of evidence for the presence of a QTL at a particular location.

Simple-Interval Mapping (SIM)



Limitations:

1. Location and effects of detected QTLs are confounded
2. QTL positions not precisely detected
3. Power to detect QTL is low when marker density is low
4. Multiple comparison increases false positives

Composite Interval Mapping (CIM)

- Developed by Jansen and Stam in 1994
- It combines interval mapping for a single QTL in a given interval with multiple regression analysis
 - Fits both the effects of a QTL as well as the effects of covariates (subset of selected genetic markers)
- Uses some markers as **cofactors**
- It is more precise and effective when linked QTLs are involved.
- It fits parameters for a target QTL in one interval while simultaneously fitting partial regression coefficients for "background markers" to account for variance caused by non-target QTL.

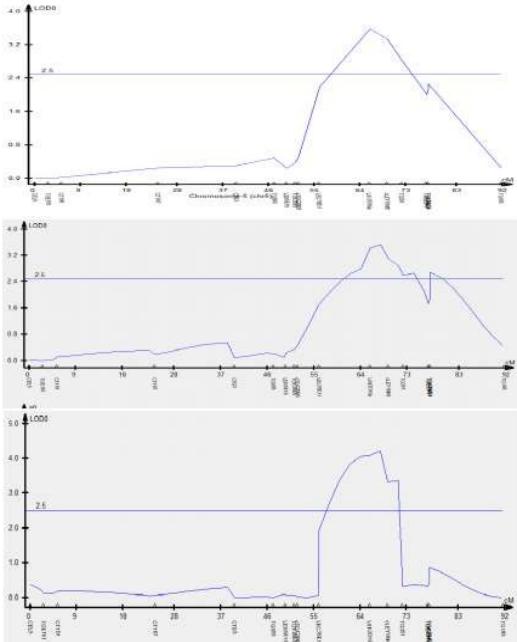
Advantages:

- By using linked markers as cofactors, the test is not affected by QTL outside the region, thereby increasing the precision of QTL mapping.
- By eliminating much of the genetic variance by other QTL, the residual variance is reduced, thereby increasing the power of detection of QTL.

Problems

- The effects of additional QTL will contribute to sampling variation.
- If two QTL are linked their combined effects will cause biased estimates.

Comparison of SMA, SIM and CIM for EB resistance in tomato



SMA

SIM

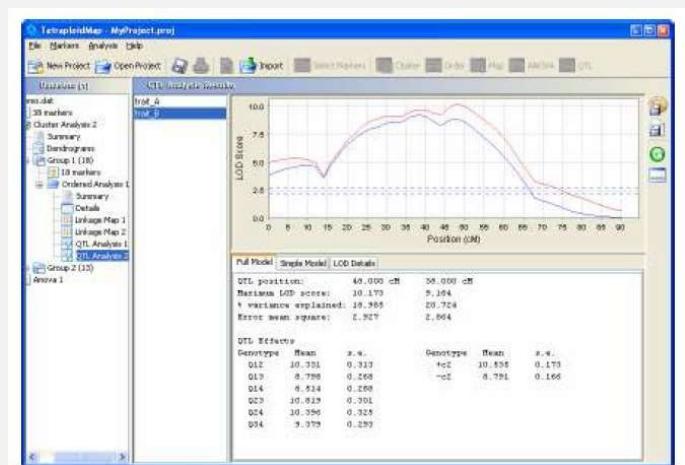
CIM



http://solcap.msu.edu/pdf%20files/5PAA_Douches_2_Mapping_Populations.pdf

Polyploids

- Generally, QTL mapping in allopolyploid genomes is same as diploids
- However, QTL mapping in autopolyploid genomes require different strategies



Tetraploid map

MERITS OF QTL MAPPING

- Identification of novel genes
- Where mutant approaches fail to detect genes with phenotypic functions , QTL mapping can help
- Good alternative when mutant screening is laborious and expensive
e.g circadium rhythm screens
- Can identify New functional alleles of known function genes
e.g.CRY2 gene

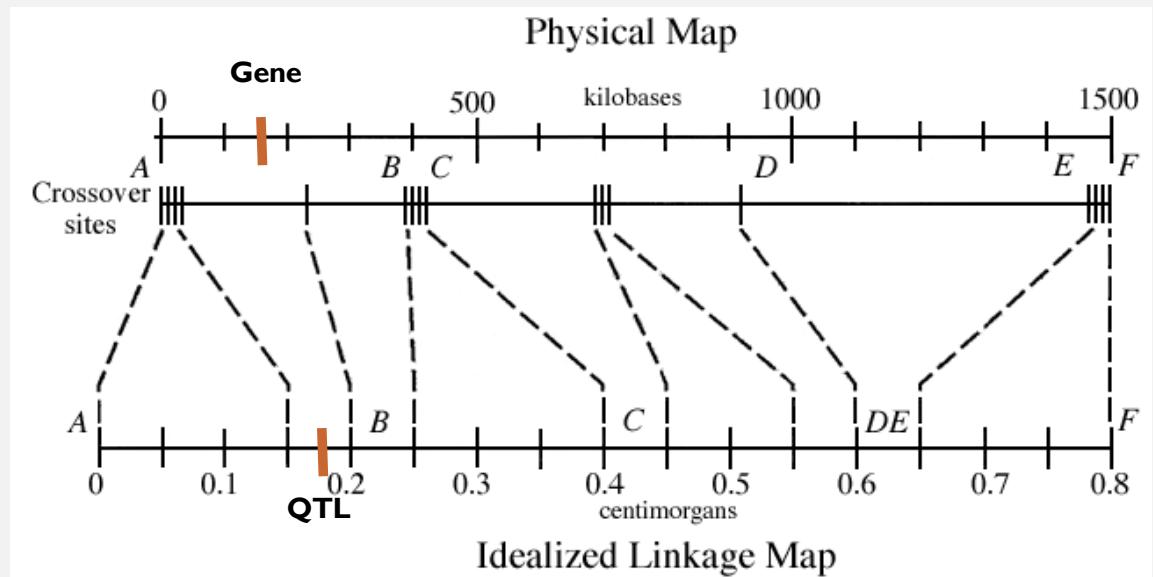
LIMITATIONS

- Mainly identifies loci with large effects.
- Small additive effects / epistatic loci are not detected and may require further analyses.
- QTL regions and effects specific for this population

PRACTICAL CONSIDERATIONS

- QTL regions different for different populations (for same trait) - Hard to compare unless we have common markers
- Presence of Physical map to compare helps identify physical positions and candidate genes

Genetic Map - Physical map



<http://www.informatics.jax.org/silver/figures/figure7-5.shtml>

Q gene input – Has 3 sections

[Header]

Study Name PI Height

Mating string r

Genotype symbols ABHxx-

Parent1 RT0034

Parent2 Cypress

Strings

- b=backcrossing,
- d=doubled-haploid creation,
- s=selfing, and
- i=random intercrossing
- r = ril
- BC₁F₁ - b
- F₂ - s
- F₃ -ss
- series of three backcrosses followed by a selfing - bbbs.
- backcross twice, self once, randomly intercross twice, and backcross again - bbsiib.

Genotype symbols

- Symbols used to represent the
- parent 1 homozygote **AA**,
- parent 2 homozygote **aa**,
- heterozygote **Aa**,
- the dominant marker phenotypes **a_** and **A_**,
- and missing data

Ex: Genotype symbols ABHxx-

2nd Section

[Locus]

Marker Chrom position(cM) Alleles

Ex:

AP2882 | 0 AAAAAA-BB-BBBAABAABAAAAAAA

RM10149 | 14.9 AAAAAAAABBBAABBAB-AAAAAA

- [Trait]

Name_of_trait Type Values

AMY_AR M 21.8 21.85 22.7 22.65 22.15 23.3 23.4 23.9 22.95

BLANKS_AR M 0 35 46.65 20 31.7 26.7 21.7 0 31.65 0 38.3

Type:

N - nominal

O - ordinal

M - metric.

Nominal – discrete classes

Ex: color of fruit – Yellow, orange, red, dark red

Ordinal – Classes based on order

Ex: Plant height – low, medium, tall

Disease – HS, S, N, R, HR

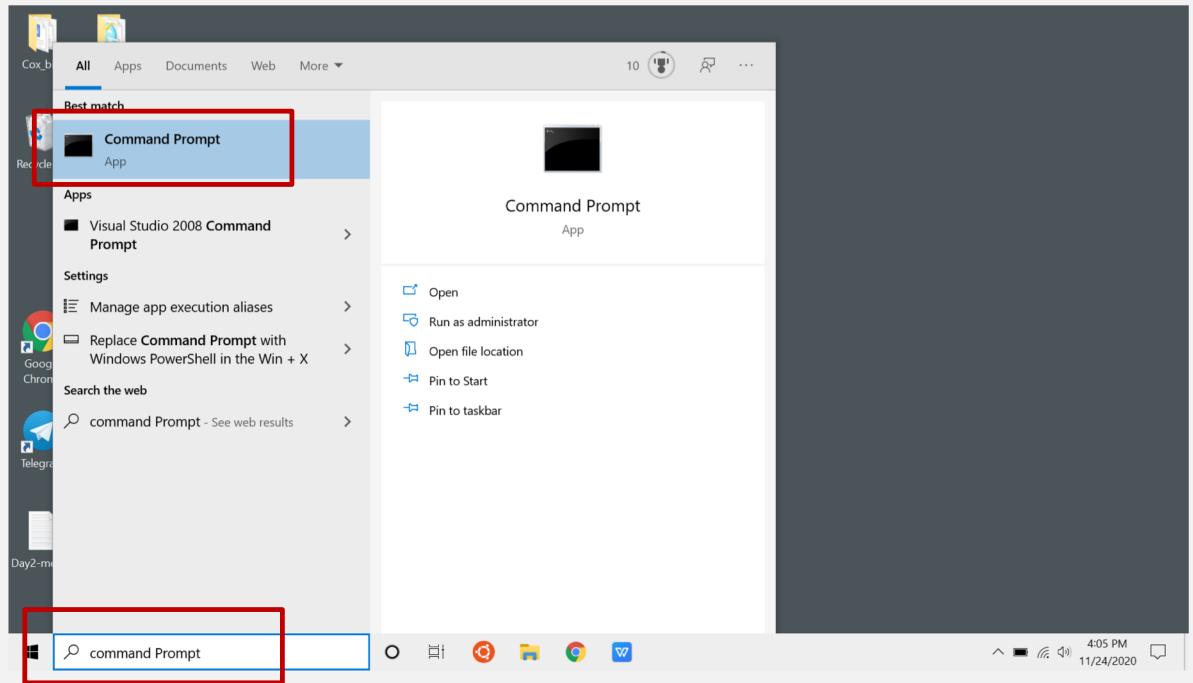
Metric – Numeric values

Ex: plant height - 5 5.1 5.4 5.7

Qgene

Data courtesy - Phillip McClean, NDSU

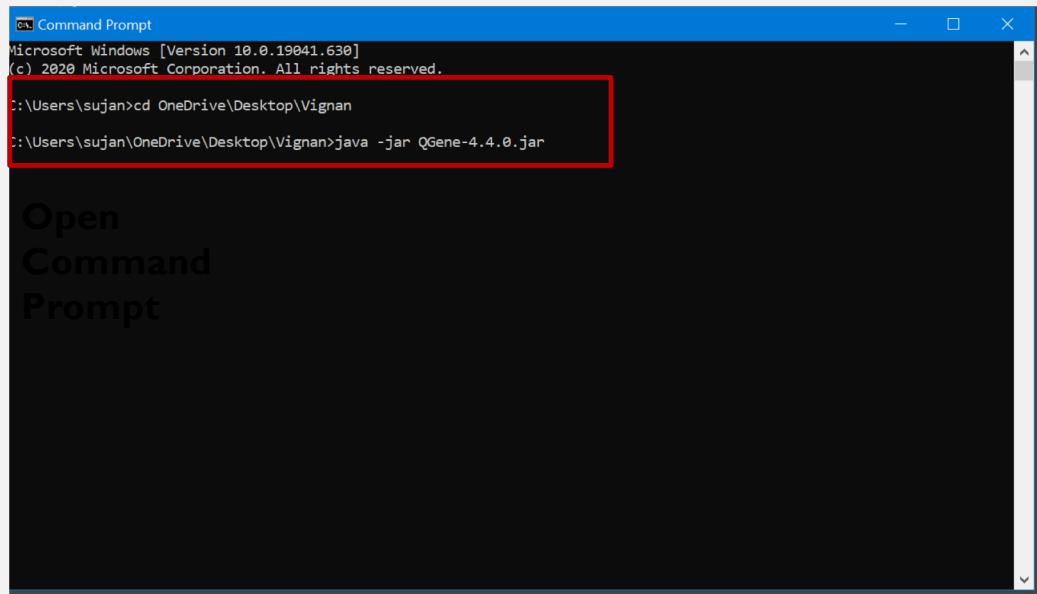
Open Command Prompt



type

cd path-foor-
qgene

java -jar
Qgene-4.4.0.jar

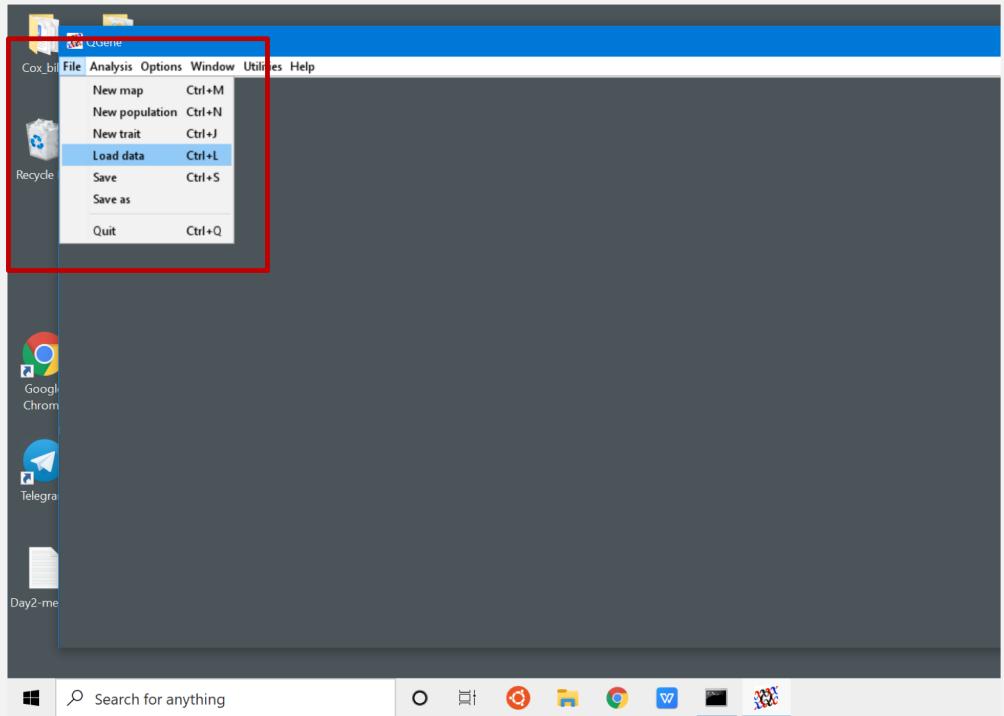


The screenshot shows a Windows Command Prompt window titled "Command Prompt". The window title bar includes the text "Command Prompt", the operating system version "Microsoft Windows [Version 10.0.19041.630]", and copyright information "(c) 2020 Microsoft Corporation. All rights reserved.". The main area of the window displays the command line history:

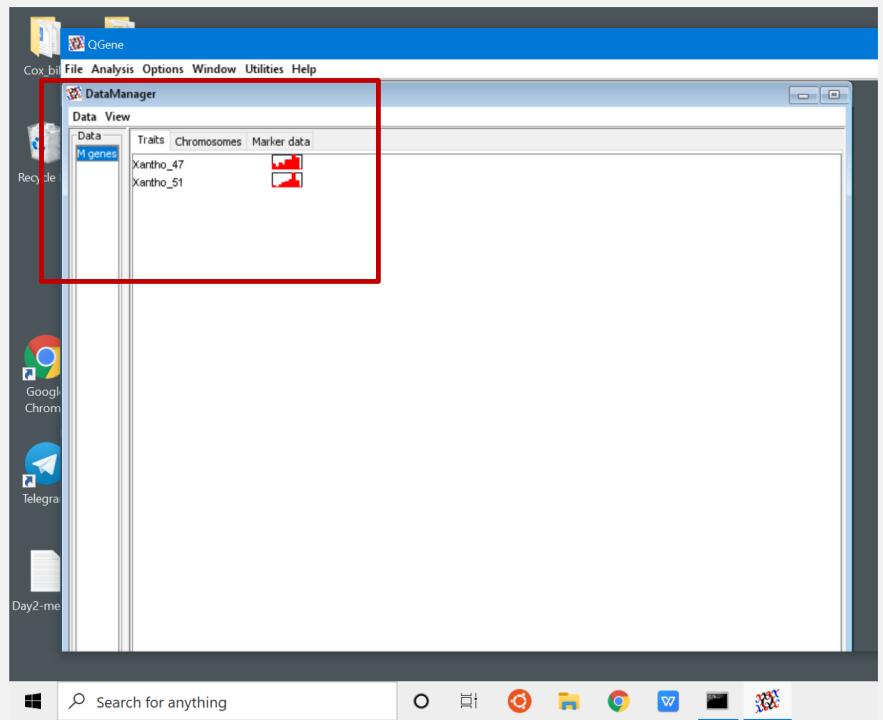
```
C:\Users\sujan>cd OneDrive\Desktop\Vignan
C:\Users\sujan\OneDrive\Desktop\Vignan>java -jar QGene-4.4.0.jar
```

A red rectangular box highlights the command "java -jar QGene-4.4.0.jar". To the left of the command prompt window, there is a vertical text overlay that reads "Open", "Command", and "Prompt" stacked vertically.

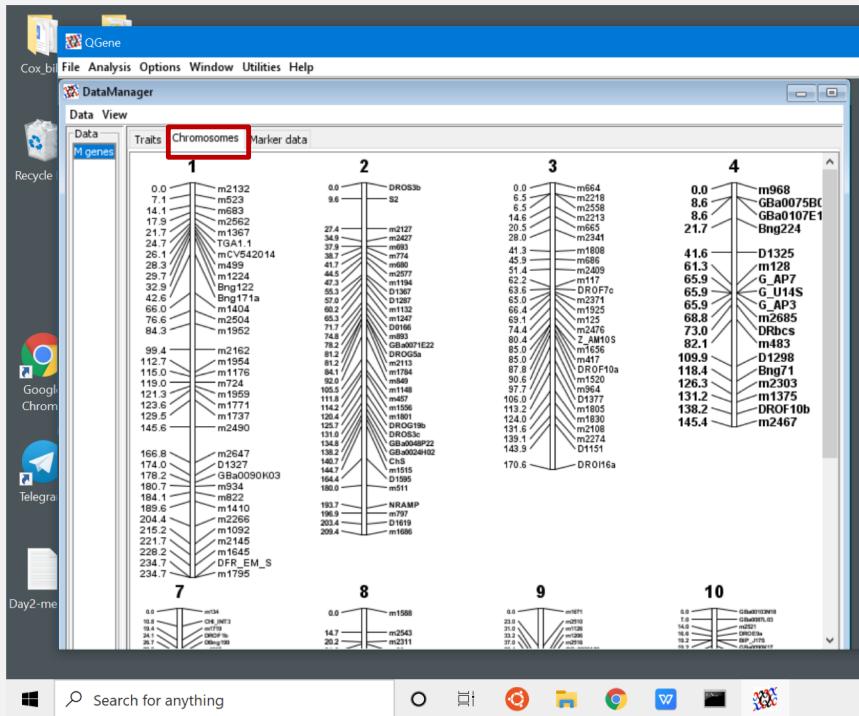
File/Load data



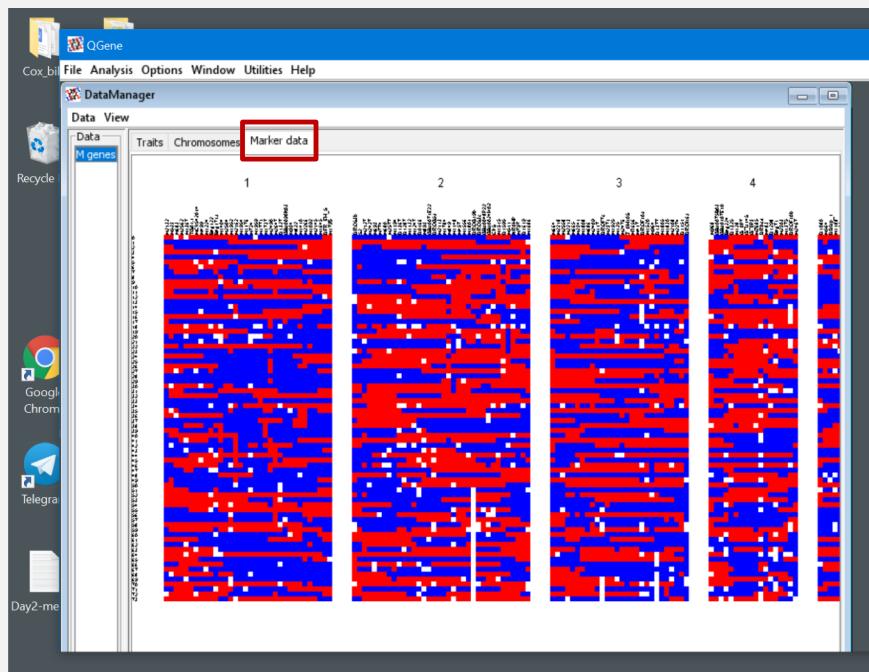
Data loaded
- 2traits



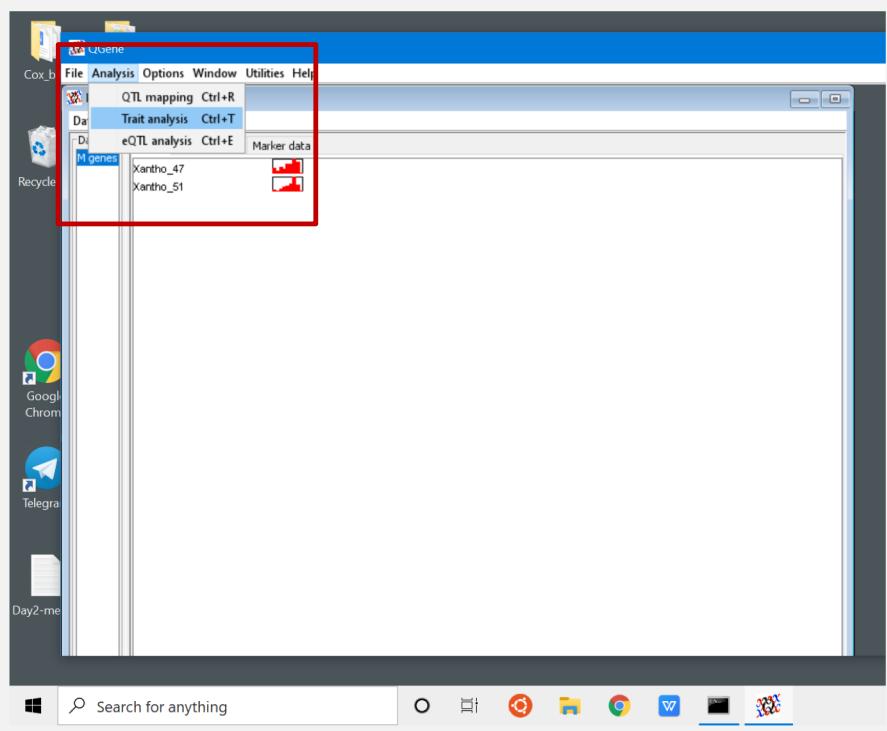
Data loaded - Chromosomes view



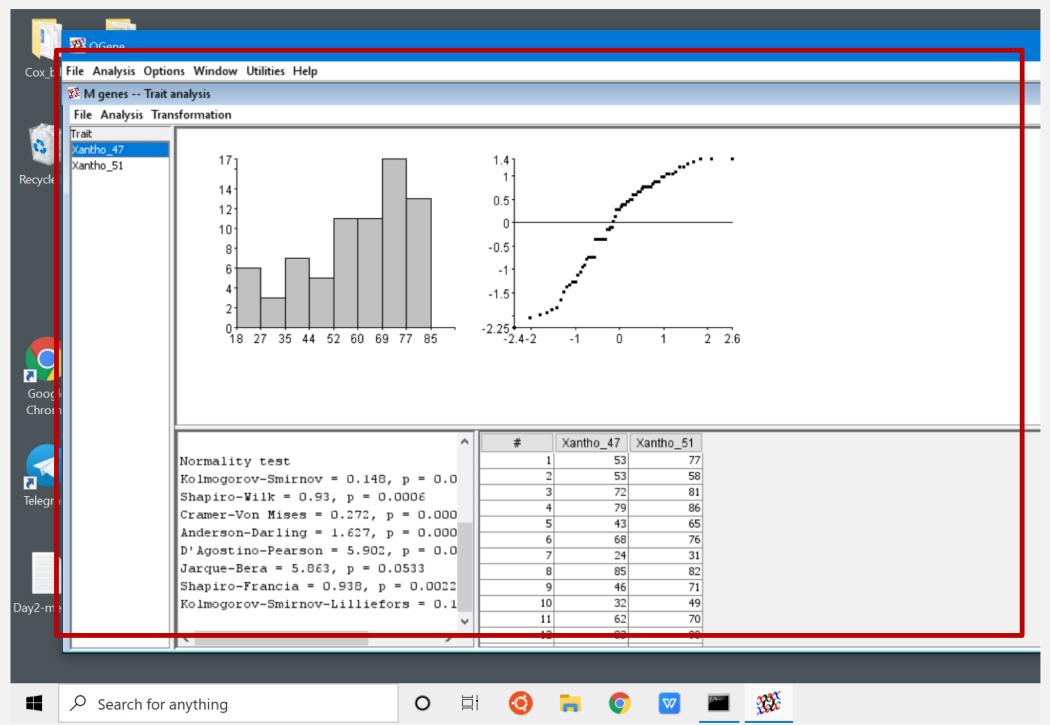
Data Loaded - Marker Data



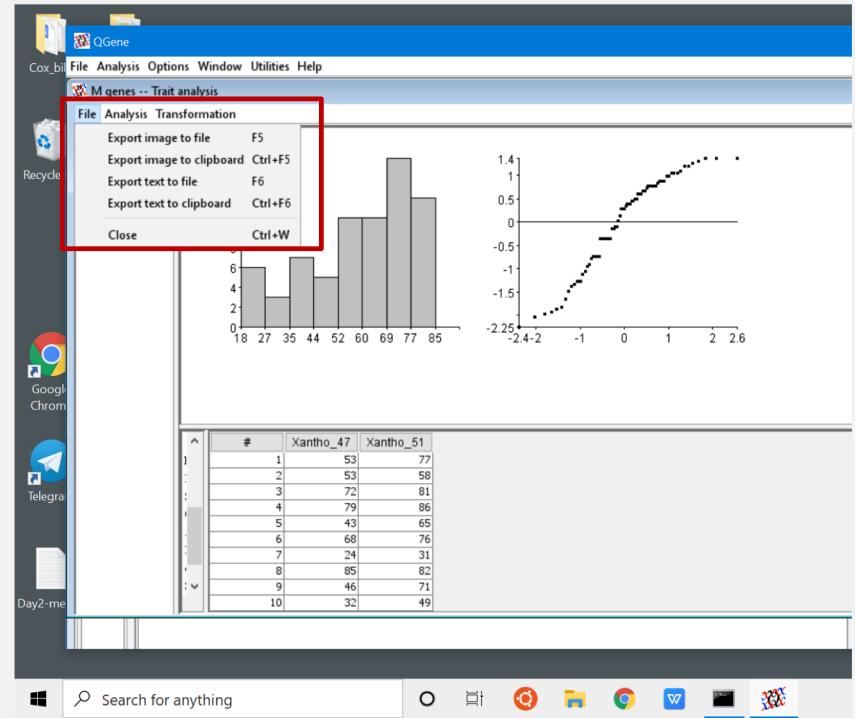
Analysis/Trait analysis



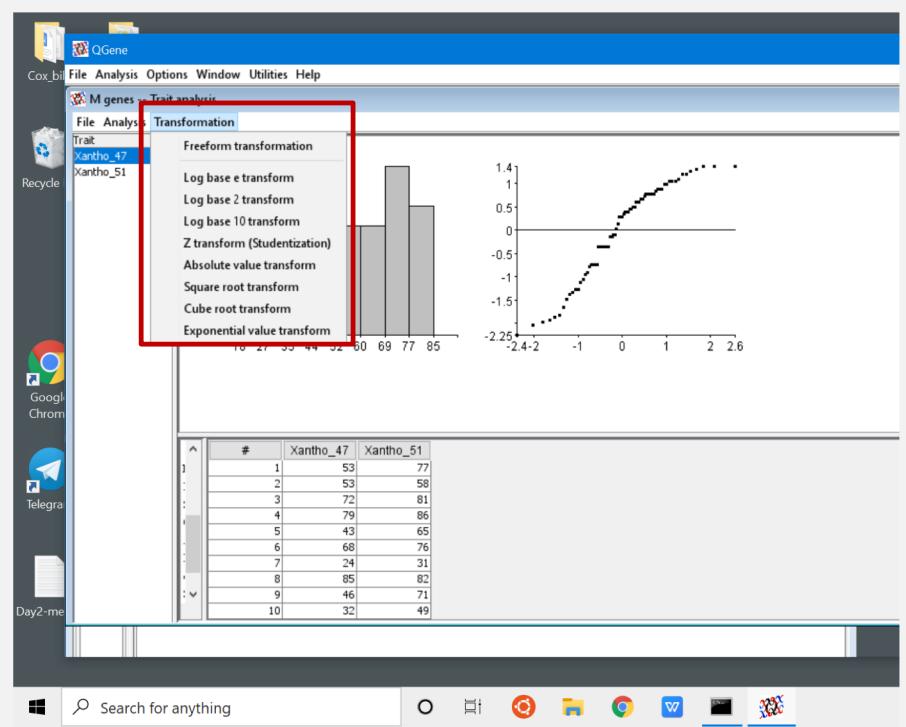
Trait : histogram Normality tests



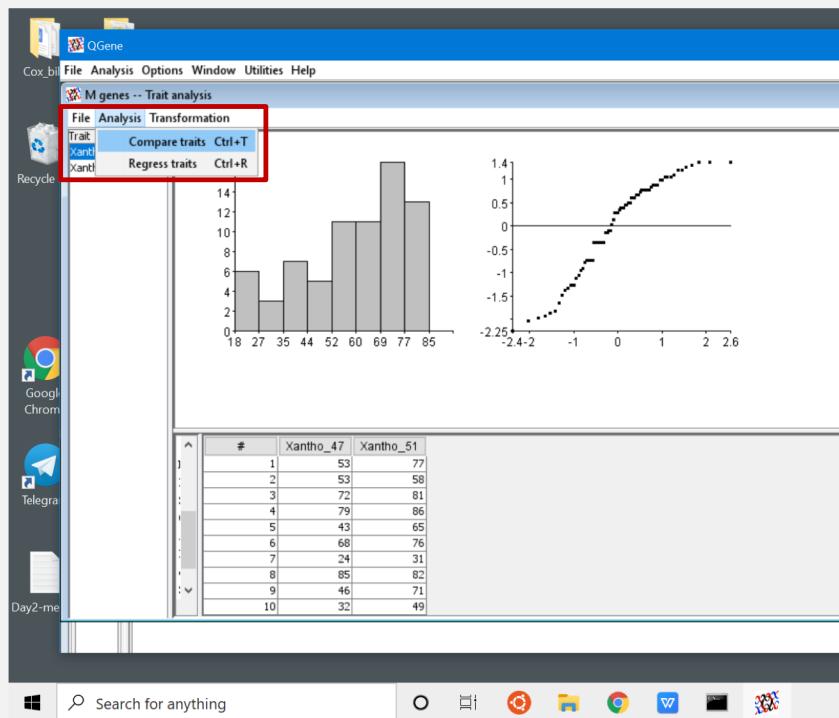
File - options to save data or copy test to clipboard



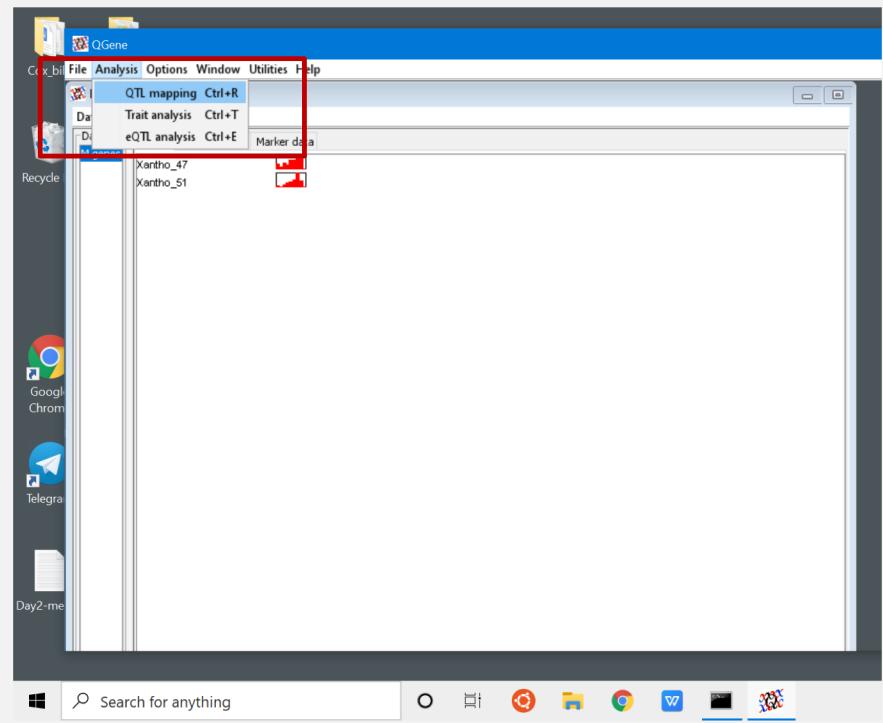
Transformation” - To convert data



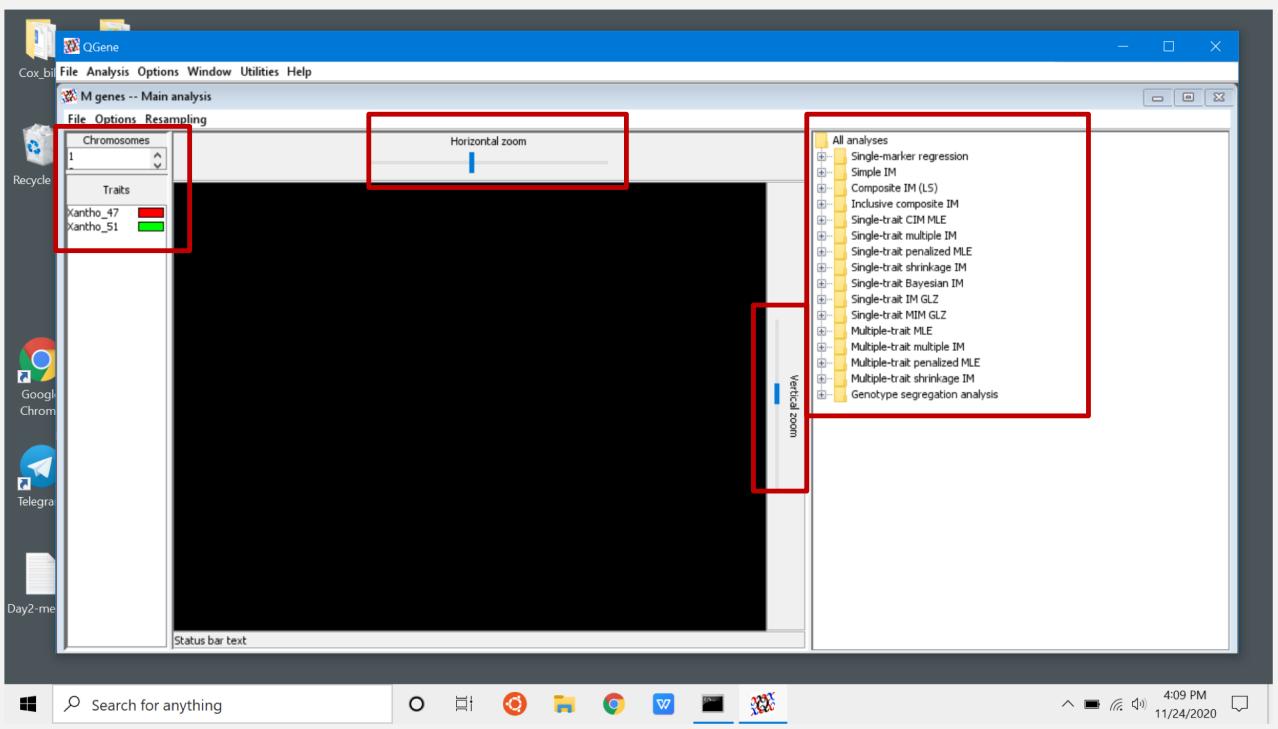
Analysis/Compare traits - to estimate correlation among different traits



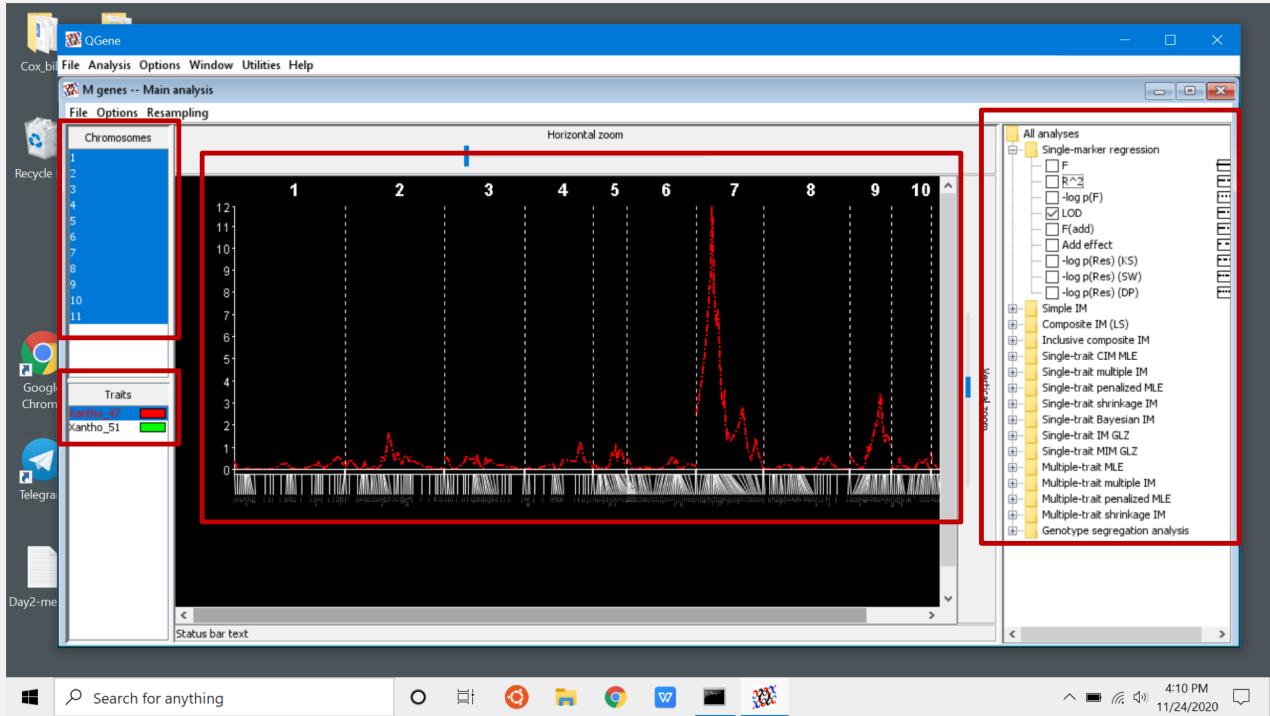
Analysis / QTL Mapping



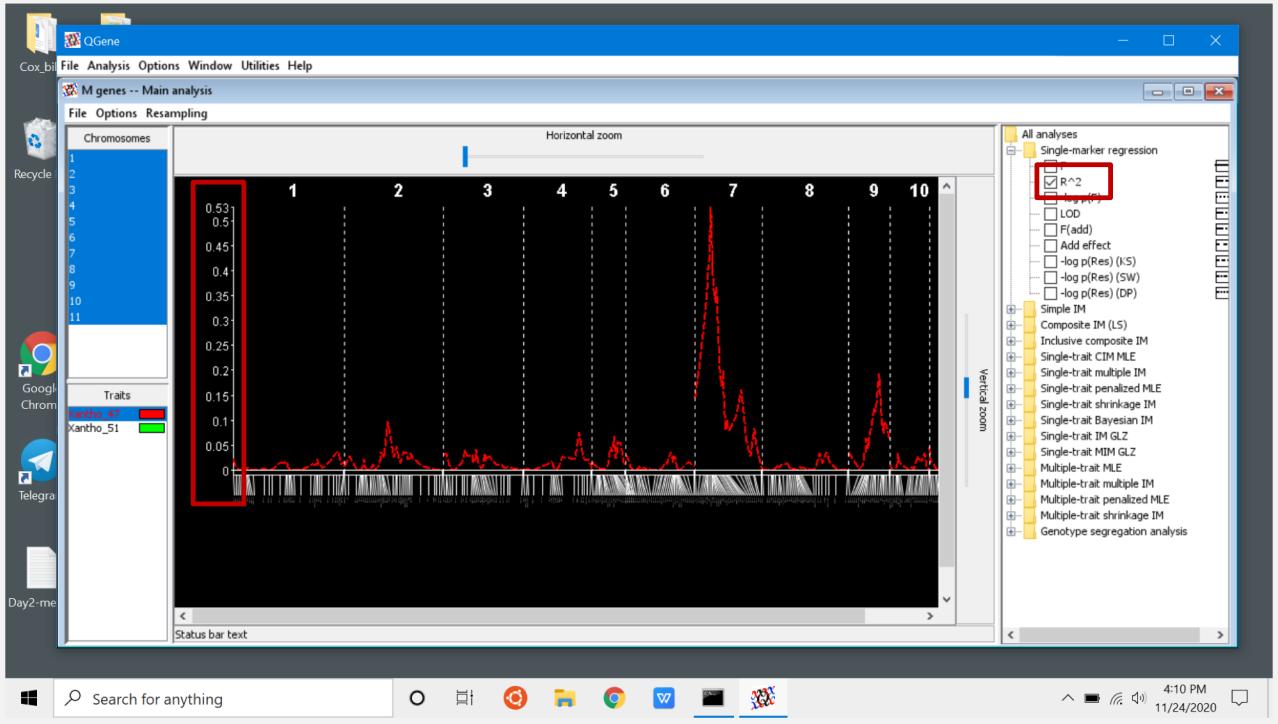
Various Windows - All adjustable in size



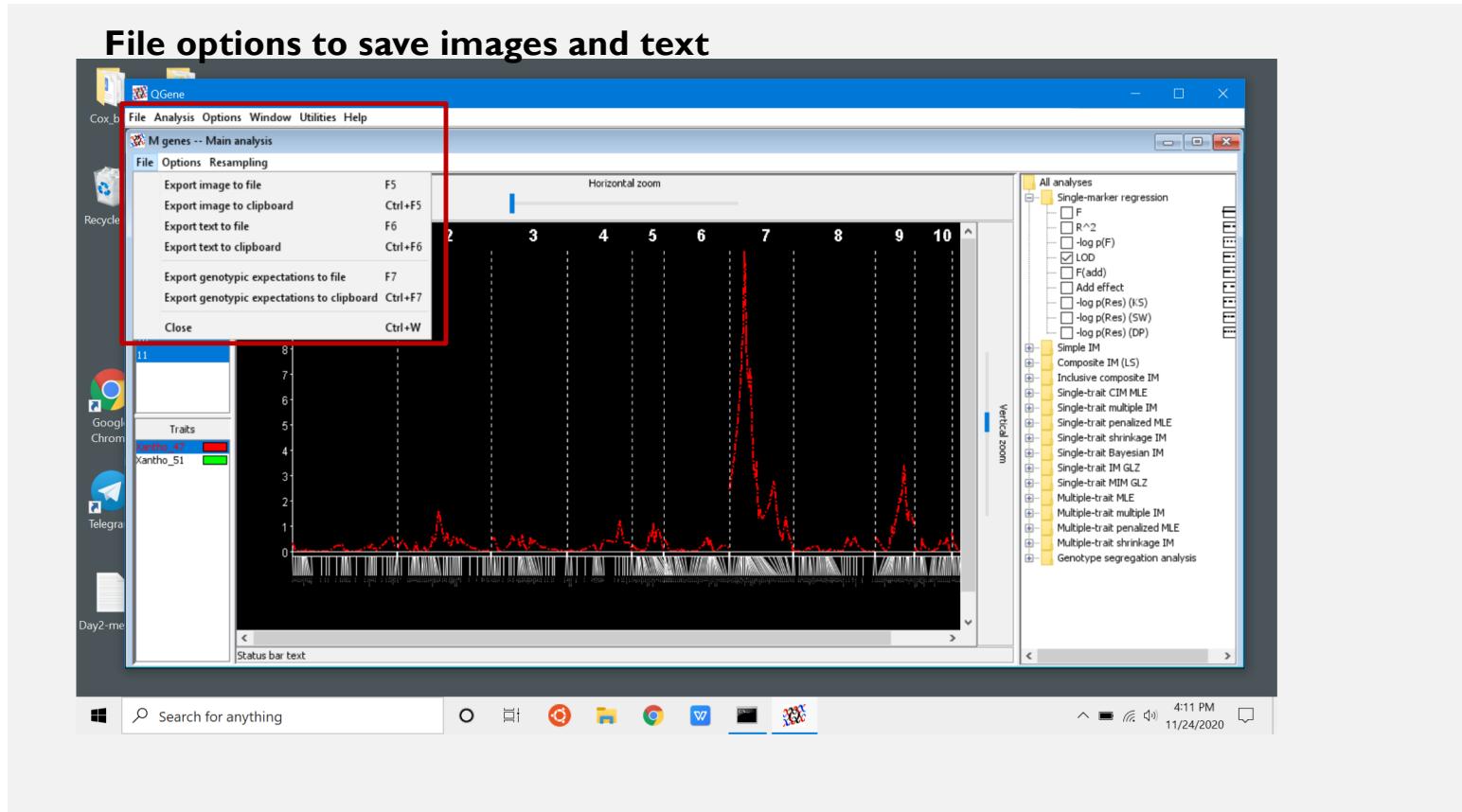
Single Marker regression (SMA) - LOD



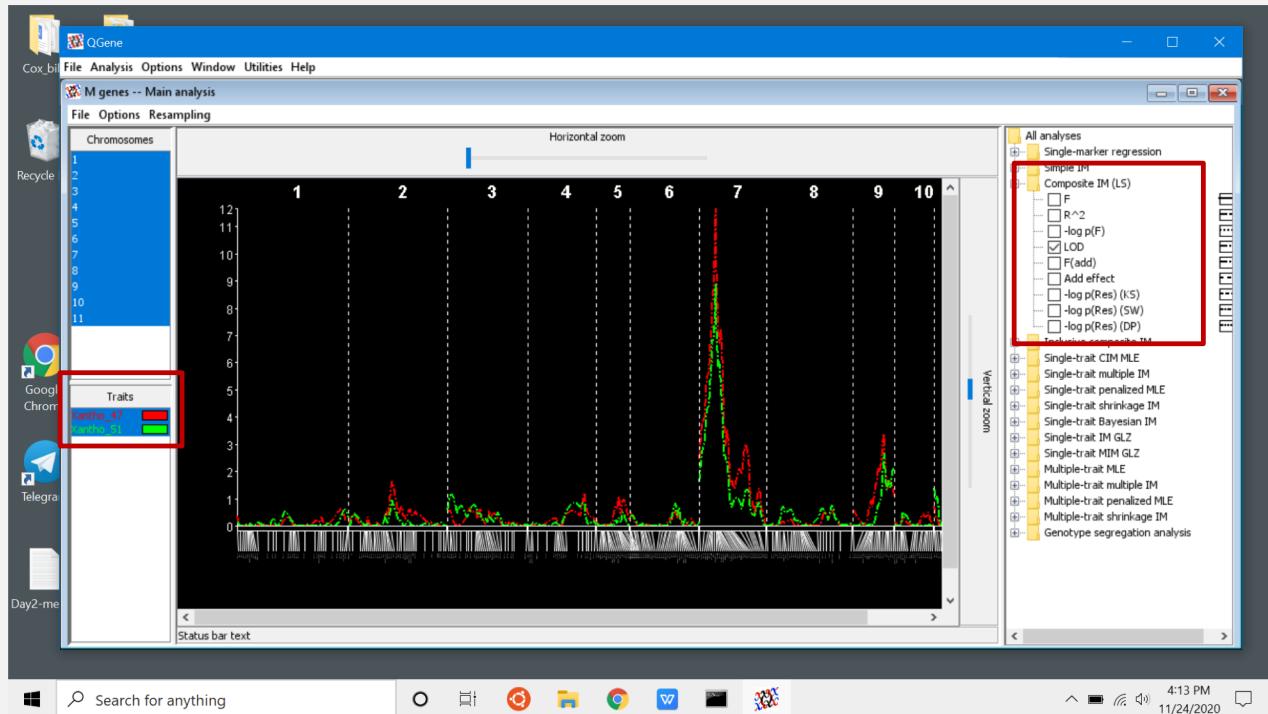
Single Marker Regression - R^2



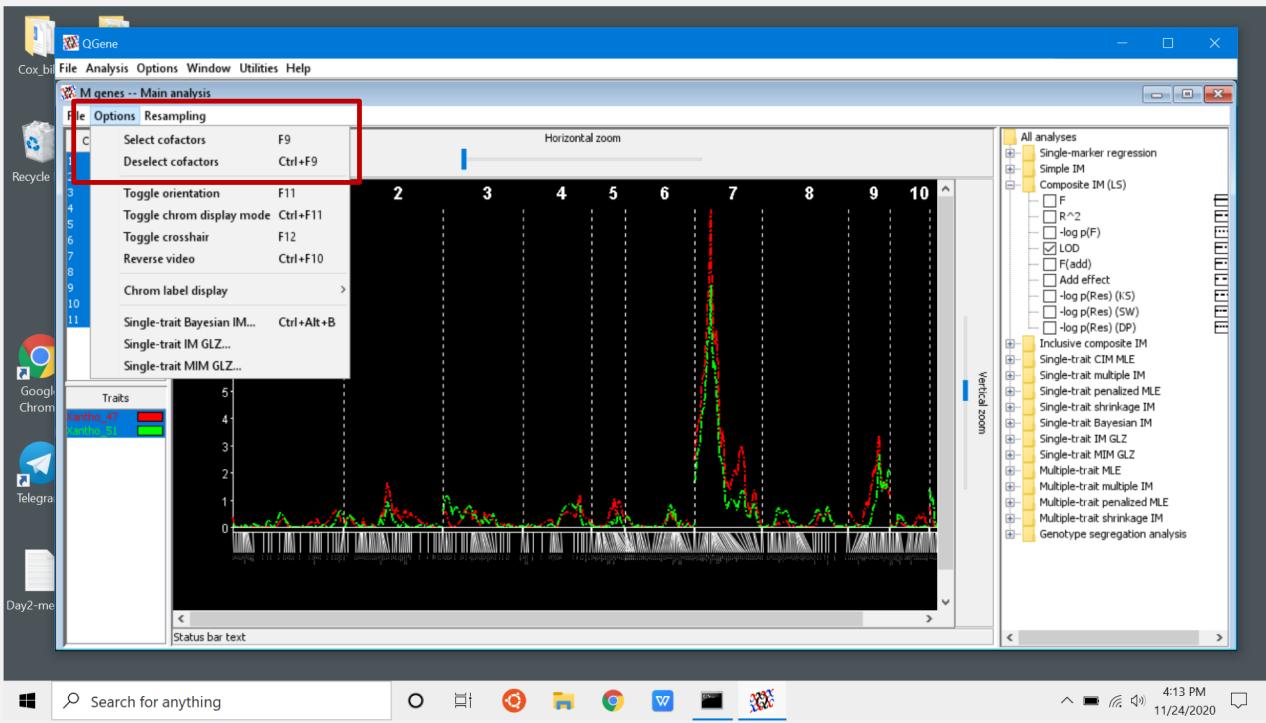
File options to save images and text



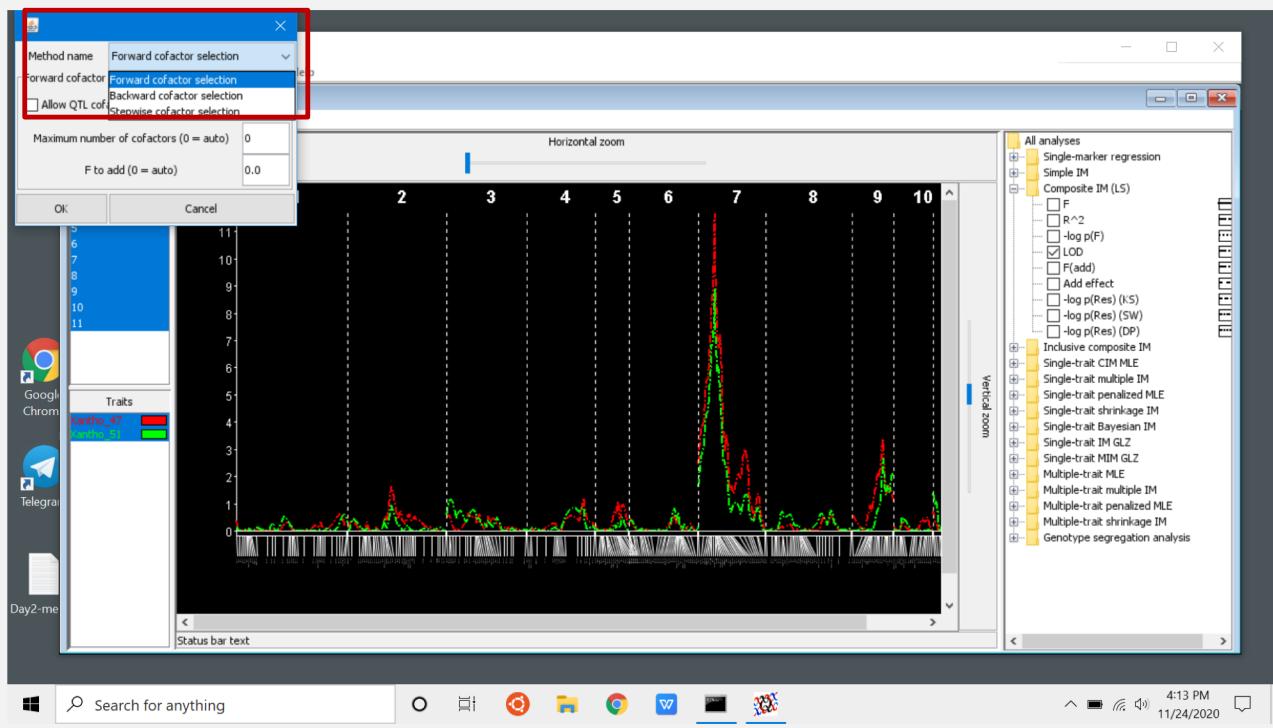
Composite Interval Mapping - LOD



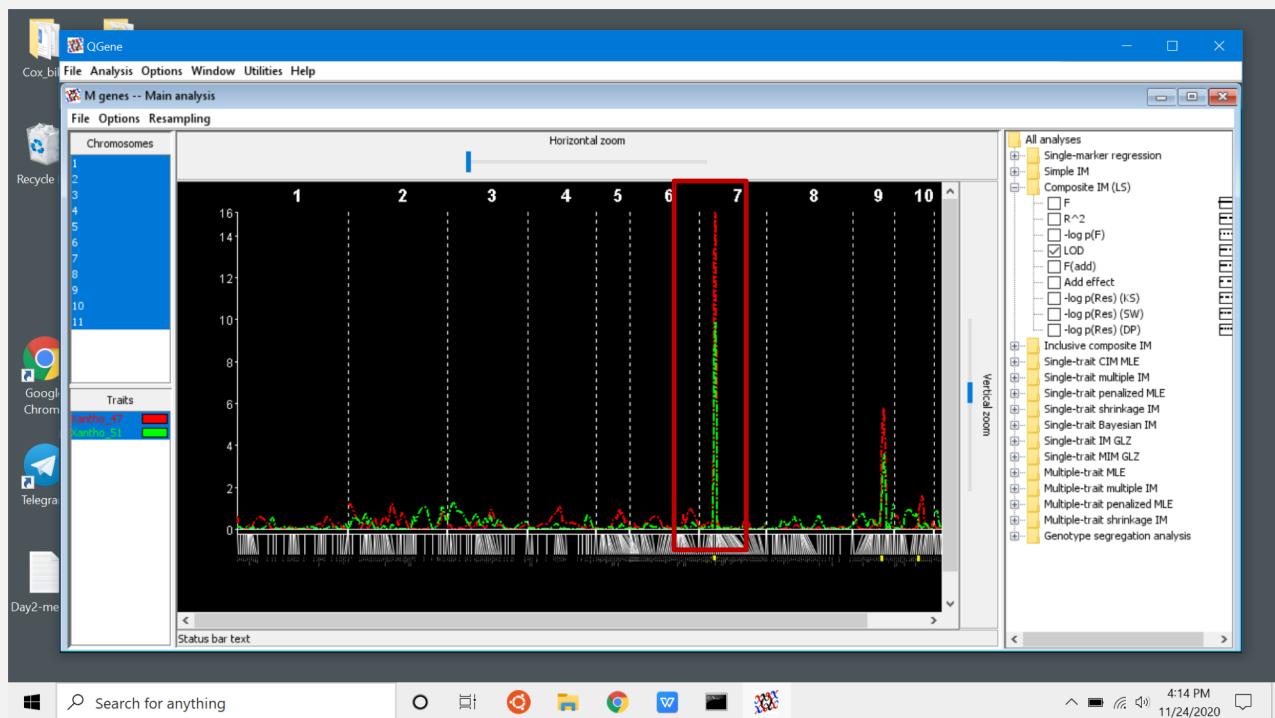
Options / select cofactors



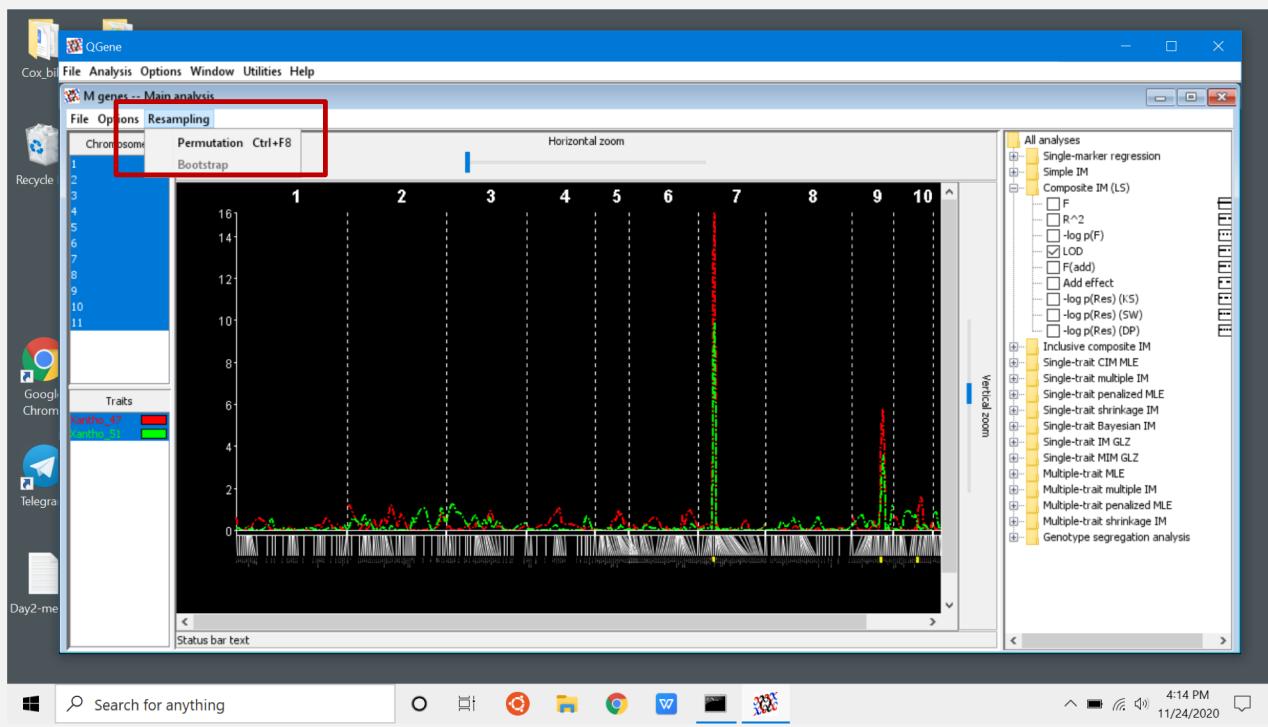
Select “Stepwise cofactor selection”

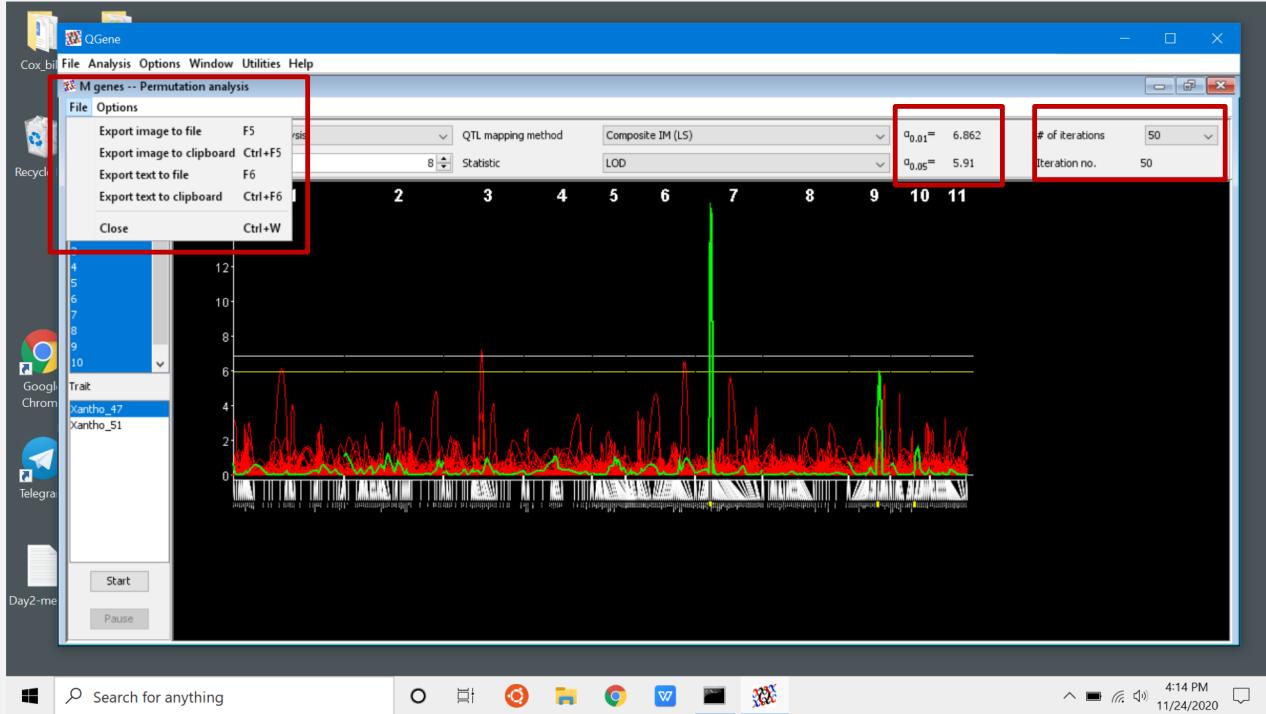


CIM - narrowed QTL region



Resampling - to determine cutoffs





Note the cut-offs

Final QTL image

