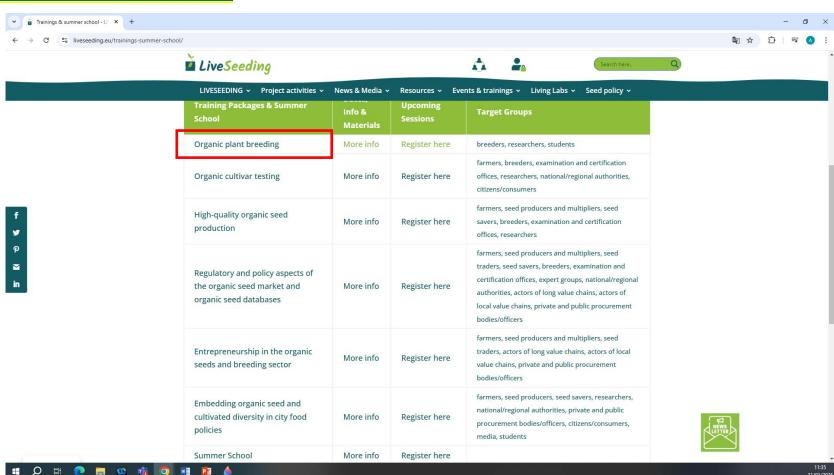
## TRAINING IN ORGANIC BREEDING!



#### **CONTEXT: Training in LIVESEEDING project**

https://liveseeding.eu/trainings-summer-school/





## Training in organic breeding organized in 5 Modules

- **1. Module 1** Plant Genetic Resources (PGRs): collection, conservation and exchange to support the increase of agrobiodiversity in farming systems
- 2. Module 2 Phenomics: approaches and tools for genetic resources and breeding material characterization FEBRUARY 3rd 2025, 9:00 to 17:30 CET
- 3. Module 3 Breeding methods fundamentals FEBRUARY 13th 2025, 9:00 to 18:00 CET
- 4. Module 4 Development and application of molecular methods in organic breeding MARCH 4th 2025, 9:00 to 18:00 CET
- **5. Module 5** Organic heterogeneous material (OHM) design and development MARCH 7th 2025, 9:00 to 18:00 CET



## Training in organic breeding organized in 5 Modules

- **1. Module 1** Plant Genetic Resources (PGRs): collection, conservation and exchange to support the increase of agrobiodiversity in farming systems
- 2. Module 2 Phenomics: approaches and tools for genetic resources and breeding material characterization FEBRUARY 3rd 2025, 9:00 to 17:30 CET
- 3. Module 3 Breeding methods fundamentals FEBRUARY 13th 2025, 9:00 to 18:00 CET
- **4. Module 4** Development and application of molecular methods in organic breeding MARCH 4th 2025, 9:00 to 18:00 CET
- **5. Module 5** Organic heterogeneous material (OHM) design and development MARCH 7th 2025, 9:00 to 18:00 CET



# Module 4 - Development and application of molecular methods in organic breeding

#### MARCH 4th 2025 9:00 to 18:00 CET

Unit 4.1: Genome Wide Association and its potential for understanding the genetic basis of complex traits

9:00-11:00 Partner in charge: UPV (Neus Ortega Albero)

11:00-11:30 Break

Unit 4.2: Genomic Selection and its potential in organic breeding

11:30-13:00 Partner in charge: FiBL (Michael Schneider)

13:00-14:30 Lunch Break

Unit 4.3: Application of molecular marker-based selection in practical breeding

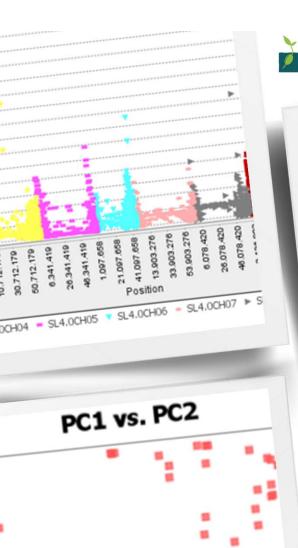
14:30-16:00 Partners in charge: KIS (Barbara Pipan) + FiBL (Lupins; Teresa Lazzaro)

16:00-16:30 Break

Unit 4.4: Overview on Experimental populations (e.g. NILs, RILs, introgression families, MAGIC) and their role in molecular breeding approaches

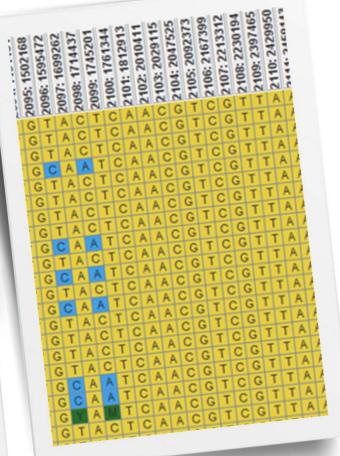
16:30-18:00 Partner in charge: UPV (Neus Ortega-Albero, Adrian Rodríguez-Burruezo) + KIS (Vladimir Meglic)





22 -20 -18 -16 -14 -12 -10

# LiveSeeding





Module 4: **Development and application of molecular methods in organic breeding** 

Unit 4.1: Genome Wide Association and its potential for understanding the genetic basis of complex traits

Author: Neus Ortega Albero (UPV)





Funded by the European Union, the Swiss State Secretariat for Education, Research and Innovation (SERI) and UK Research and Innovation (UKRI).

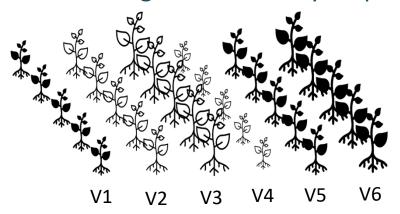




# **Relationship Gene - Trait**

# Genotype → Phenotype Meeting the causing genes.

- a) Develop markers to select interesting phenotypes easily (Unit 3)
- a) Establishing correlations between traits and finding if the genetic base is the same to select one complex trait through another easy-to-phenotype trait





# Qualitative vs. Quantitative traits

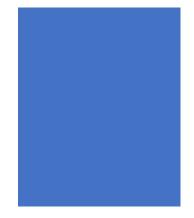
#### *Qualitative traits* → *categories*

Controlled by 1 or few genes



#### Quantitative traits → continuous expression

Controlled by many genes along the genome





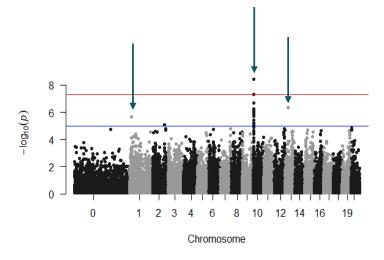
#### **GWAS**

#### Genome-Wide Association Analysis (GWAS)

**Quantitative Trait Loci (QTL)** → all the points (markers) in the genome associated statistically with the trait. The accumulative effect of these QTLs is the phenotype.

**Exploit existing natural variability** in a population accumulating recombination during decades.

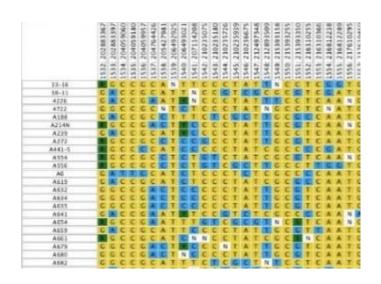
Used in a collection of varieties or a breeding population embedding variation for the studied trait.



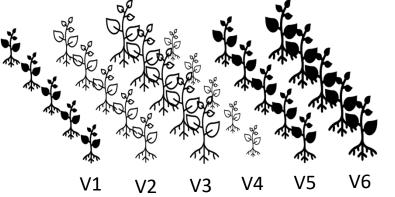


## **GWAS**

#### **Genotype**



#### **Phenotype**



#### **GWAS**



#### **GWAS**

#### **ADVANTAGES**

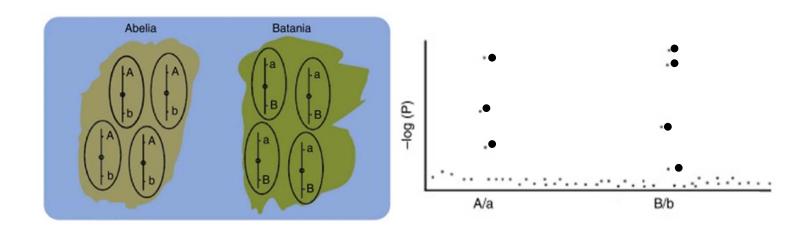
- No need to create a breeding population
- Analyze a wider range of diversity
- Useful for any type of trait
- Work on several traits in very variable populations
- High precisión in gene mapping
- Association for complex traits like adaptation to low input conditions

#### **DISADVANTAGES**

- Need variant availability
- The population of study must show variability for the trait
- Avoid population stratification



#### **Fake Associations**

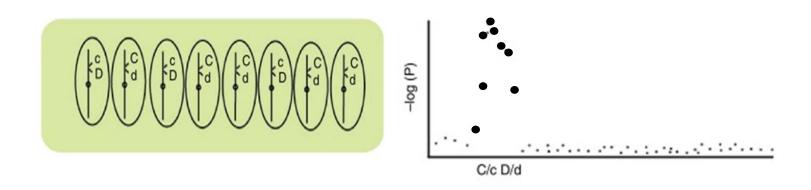


#### Fixed alleles in the population

When only two combination of haplotypes are fixed in the population, both genes will appear as associated with the trait, even only one is really associated.



#### **Fake Associations**



#### **Ligated Genes or Linkage Disequilibrium**

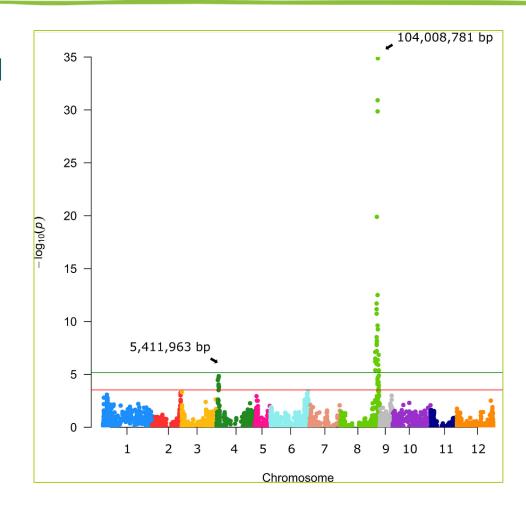
Two proximal genes always inherited together will be both associated to the trait, even only one of them is the cause of the phenotype.



## **GWAS** examples

Eggplant → pigmentation of the peel

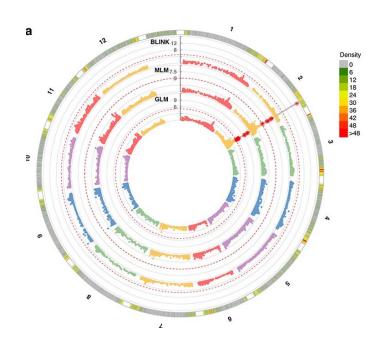


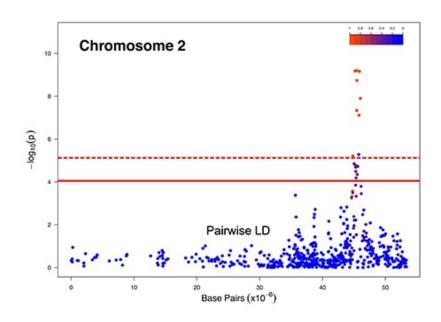




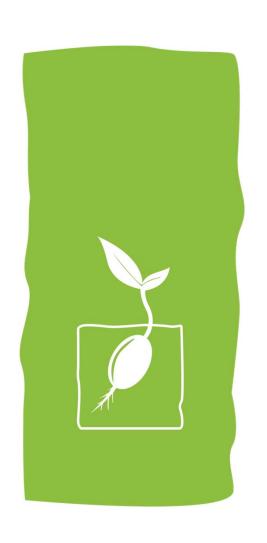
# **GWAS** examples

#### Tomato → Locule number



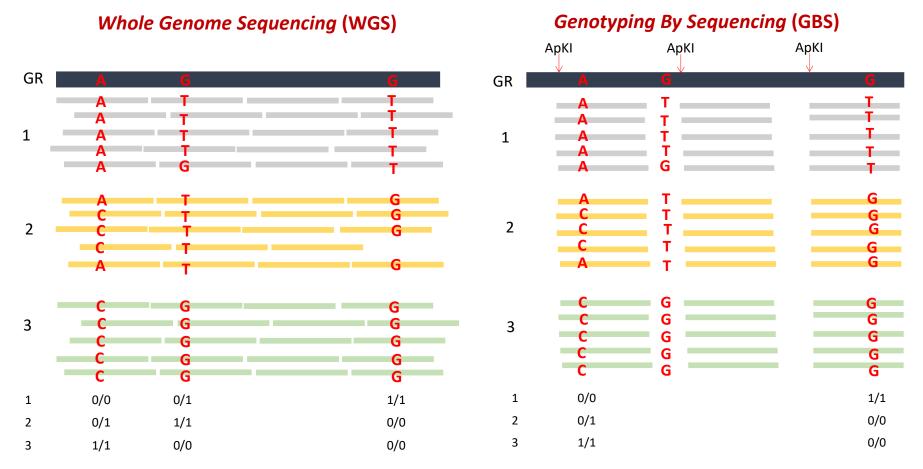






# PRACTICAL EXAMPLE

#### WGS vs. GBS



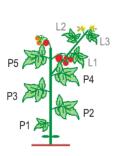


#### GWAS analysis in a collection of Solanum lycopersicum

#### Growth habit in tomato

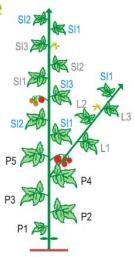
Determinate

1



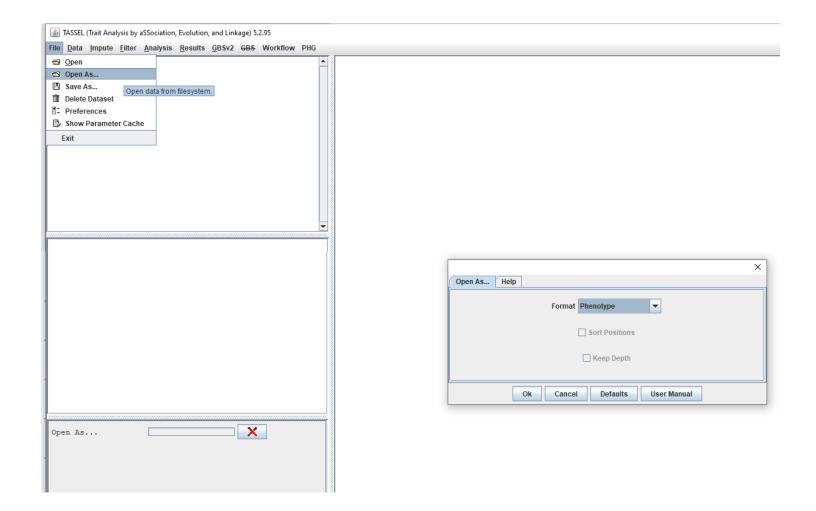
Semi - Determinate

Indeterminate

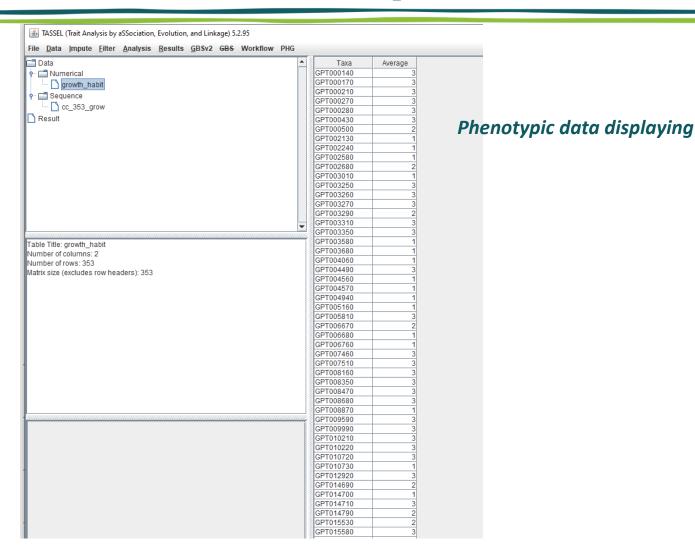




1. Open phenotypic data file

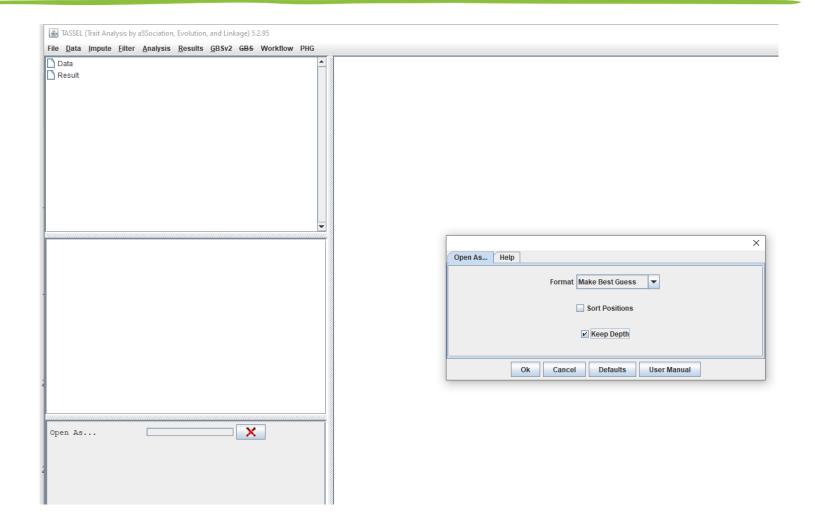




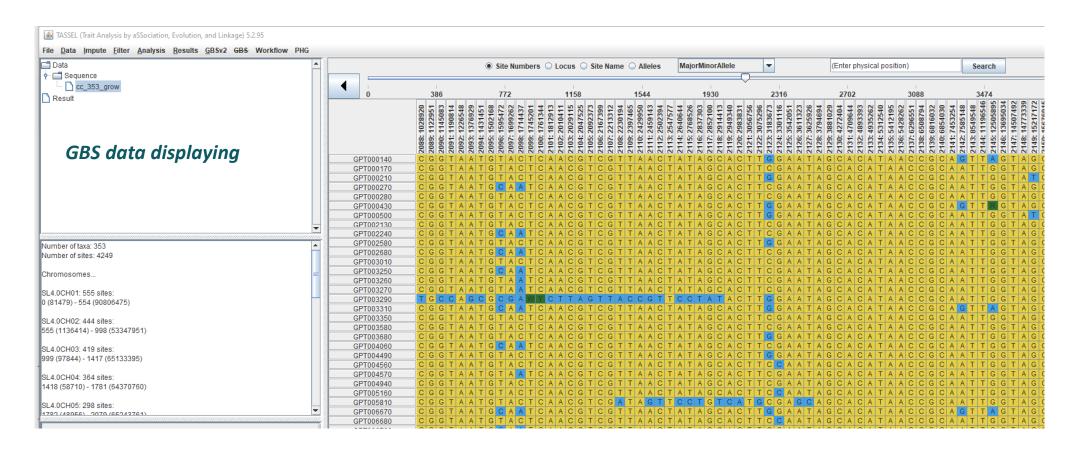




2. Open GBS data file

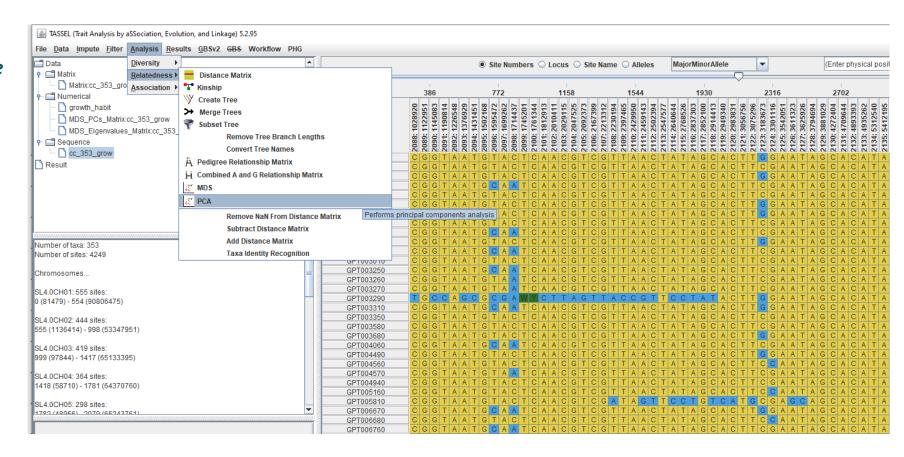






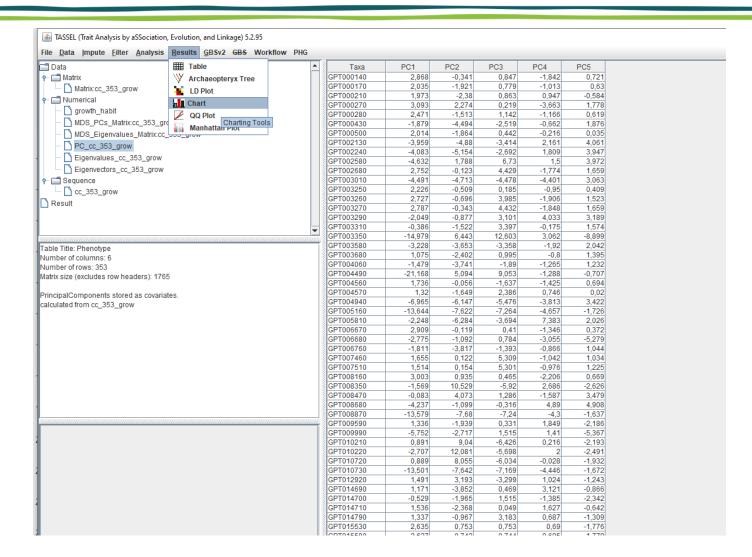


3. Checking population structure with PCA



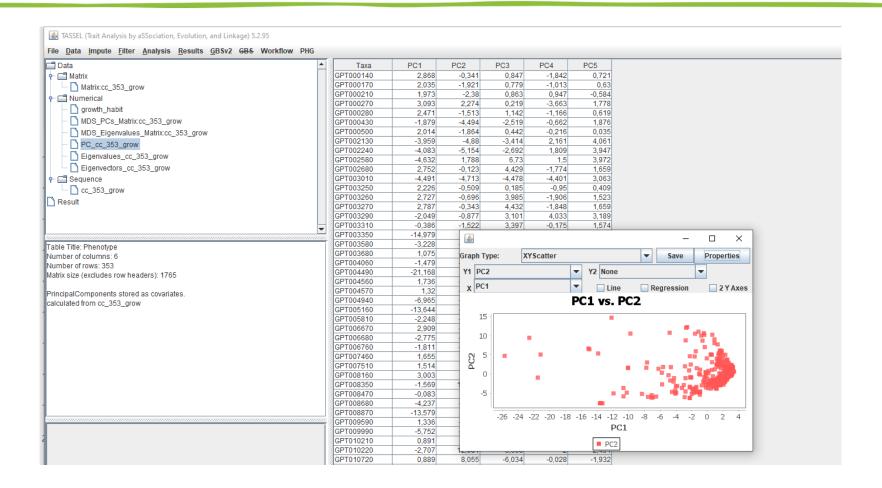


3. Checking population structure with PCA



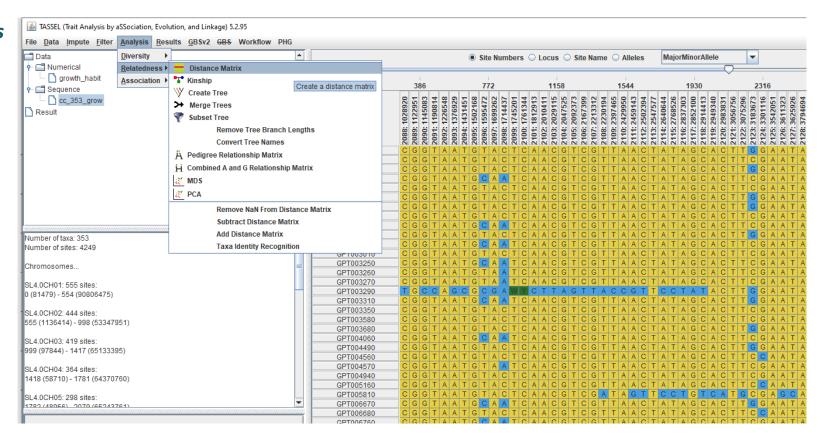


#### **Displaying PCA**



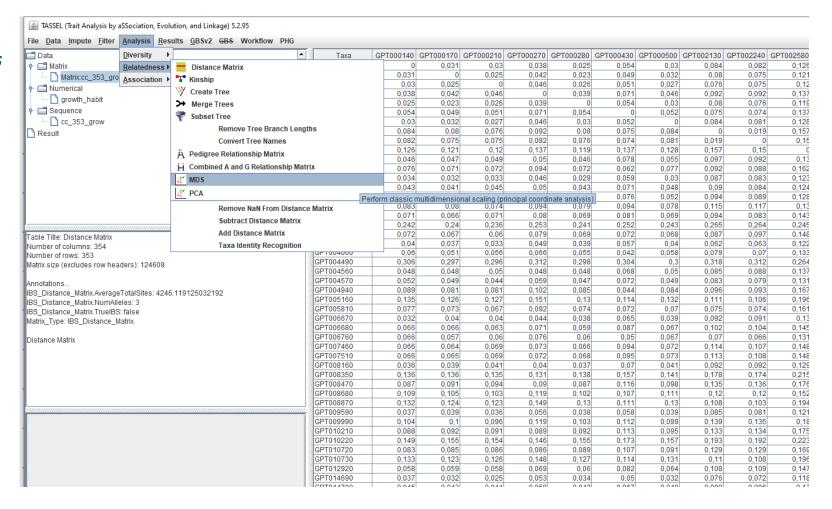


4. Calculating relationships within the population with the Distance Matrix



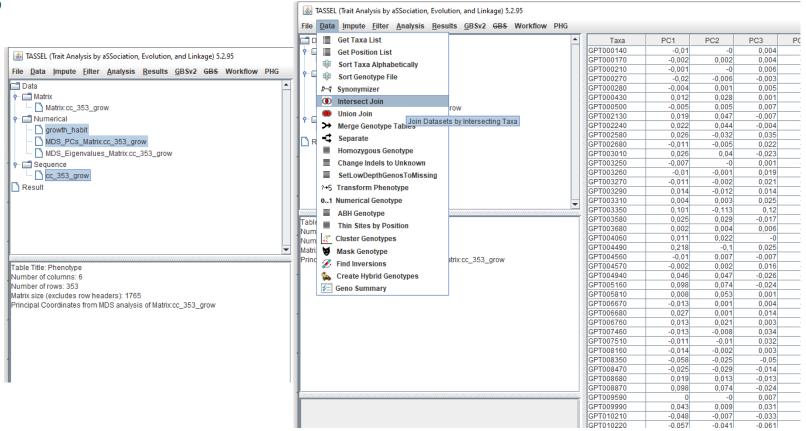


5. Incorporating the relationships to the analysis with the MDS



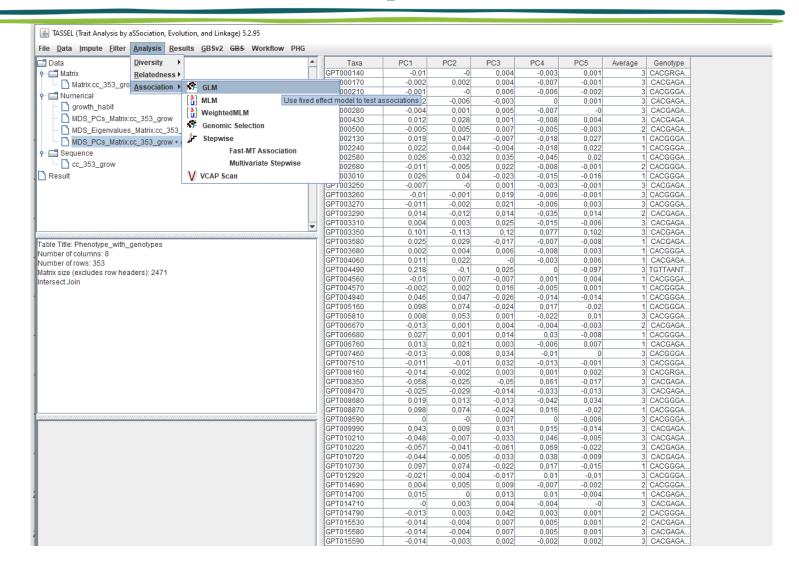


6. Select available data to the analysis (phenotype, genotype and MDS) and Intersect them



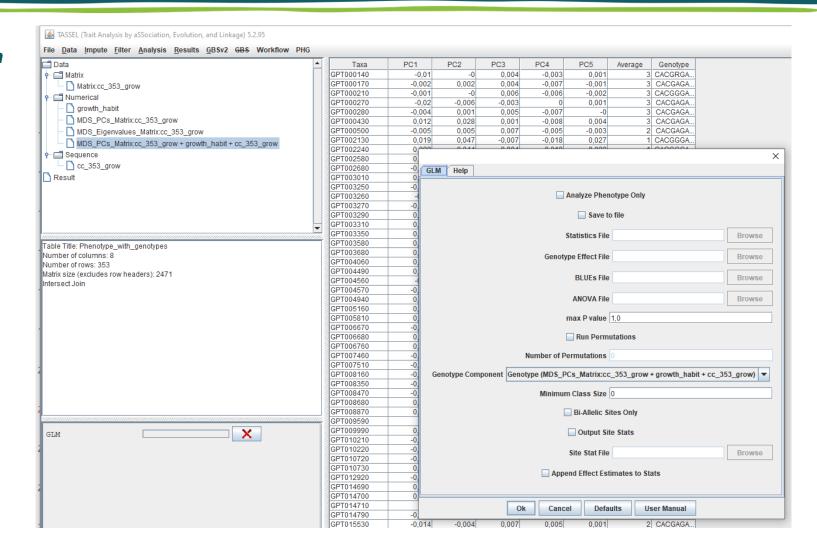


7. Association analysis with GLM





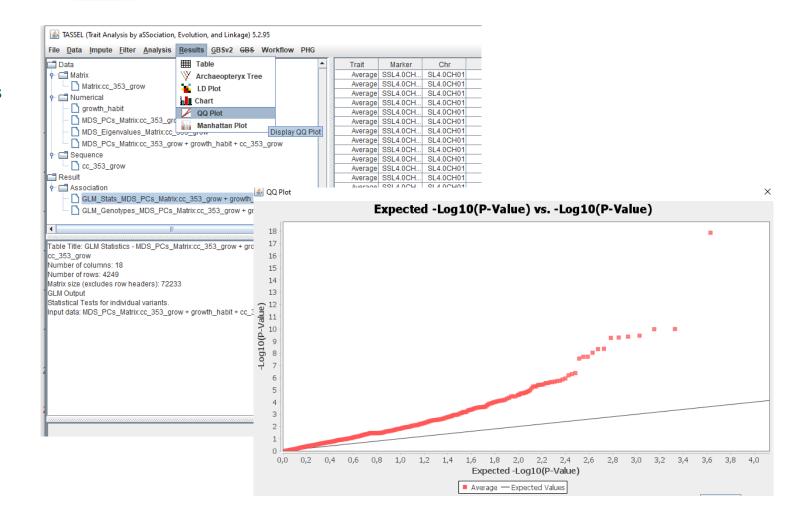
#### 7. Association analysis with GLM





#### 8. Represent QQPlot

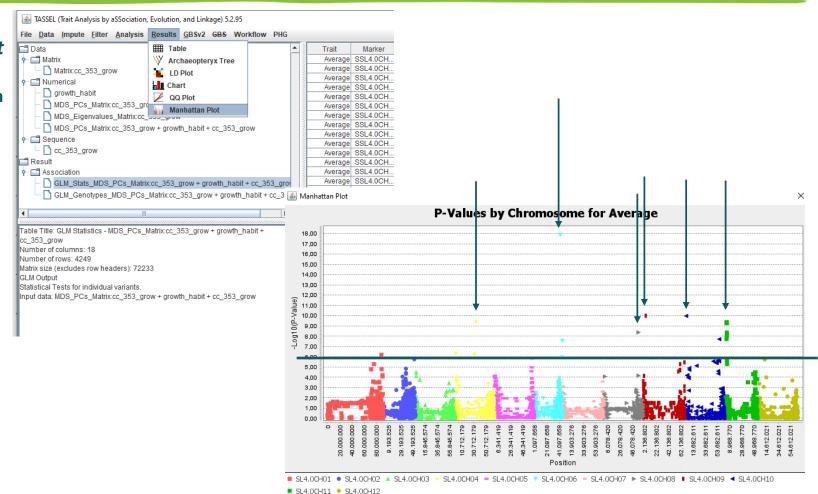
The distribution of the markers is according to the expected if any QTL is associated with the trait





#### 9. Represent Manhattan plot

The position of the markers with their association to the trait





## **Next steps**

- a) Select the most significant QTLs and establish a region upstream and downstream
- b) Use available genomic information of the species to check annotated genes in this region
- c) Analyze the function of these genes to know if they could be involved in the expression of the trait
- d) Search information in other species and in other databases



#### GWAS analysis in a collection of Solanum lycopersicum

#### Fruit shape in tomato



Flattened

1



Slightly flattened

2



Rounded

3



High-rounded

4



Heart-shaped

5

LiveSeeding



Cylindrical

6



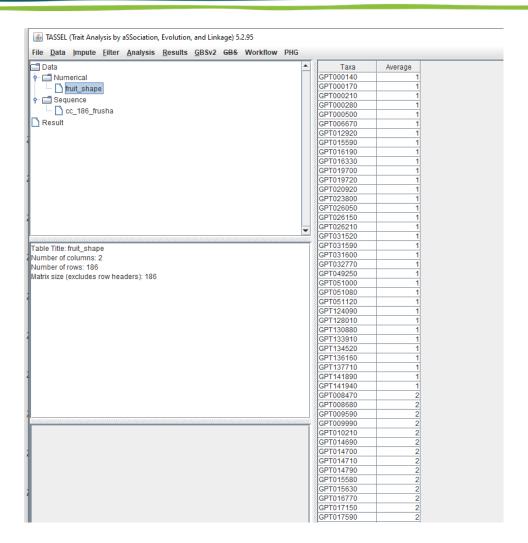
**Pyriform** 

7



8

1. Open phenotypic data file



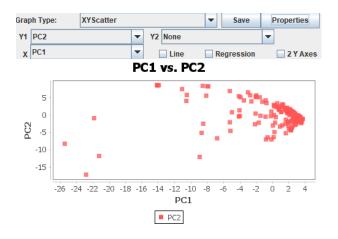


#### 2. Open GBS data file



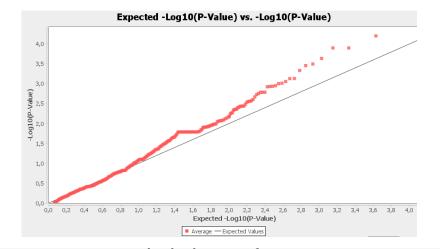


#### 3. PCA

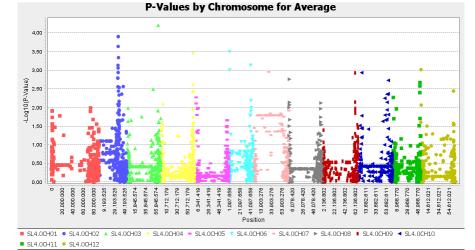


- 4. Distance Matrix
- 5. MDS
- 6. Intersect analysis information

- 7. Association analysis
- 8. QQPlot



9. Manhattan plot





## **Fast Quiz**

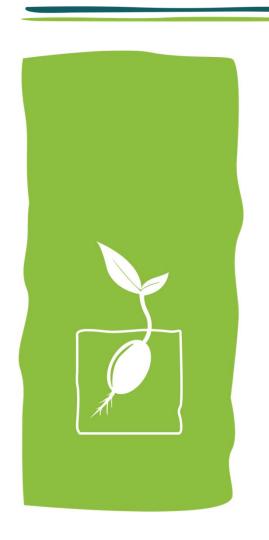


- ☐ How can the results be affected by the existence of two subpopulations?
- ☐ Which could be the reason why in this case we do not obtain any associated QTL in the second case? Be aware of the data.

Send your answers to the chat



# **Take Home Messages**



- ✓ GWAS is used to associate traits with genomic regions
- ✓ We need a non-structured and variable study population
- ✓ Specialized programs are required to intersect phenotypic and genotypic data
- ✓ Significant markers need to be analyzed to find the candidate genes responsible for the trait

# **Proposed Homework**



Try to perform your own analysis with Tassel5 and the data in additional resources (genotyping\_data.vcf, phenotyping\_data.txt).

Send your results with screenshots of the QQplots and Manhattan plots and explain the obtained results.

Send your results to <a href="mailto:neuoral@doctor.upv.es">neuoral@doctor.upv.es</a>, adrodbur@upvnet.upv.es, petra.jelincic@ips-konzalting.hr



#### **Additional Resources**

# Tassel5 □ https://www.maizegenetics.net/tassel □ https://www.youtube.com/watch?v=73hD5n-OOJc&list=PLdf-U83sN48OPSQ-DqL3AgqKOvu8nUoMT Novel High throughput phenotyping platforms: □ https://www.youtube.com/watch?v=zw02B\_oWL9k Data for homework and presentation: □ https://github.com/jmartinezpoq/LiveSeeding\_Module4/tree/master/Unit1\_PracticalExercices



#### **Additional Resources**

#### **How to Install Tassel5**





TASSEL Version 5.0 (Getting Started!) (Build: February 22, 2025 Requires: Java 1.8)

Tassel 5 Mac OS
Tassel 5 Windows 64 Bit
Tassel 5 Windows 32 Bit
Tassel 5 UNIX

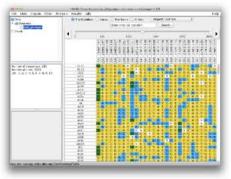
TASSEL Version 5.0 Standalone

(GBS Pipeline V2 - Preferred Version)

Using Git - Recommended!

Download ("Tags" are Versions)

#### Alignment Viewer





















































































Funded by the European Union, the Swiss State Secretariat for Education, Research and Innovation (SERI) and UK Research and Innovation (UKRI). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or REA, nor SERI or UKRI.





