



Bioinformatics workshop

3rd SoyaGen Annual Meeting

7 December 2018



UNIVERSITÉ
LAVAL

Genome Wide Association Study (GWAS)



Sidiki MALLE, François BELZILE

07/12/2018



GWAS: Overview

Step I II III IV V →

Line 1	T	G	G	T	A
Line 2	T	G	G	T	A
Line 3	T	G	G	T	A
Line 4	T	G	G	T	A
Line 5	T	G	G	T	A
Line 6	T	G	G	T	A
Line 7	T	G	G	T	A
Line 8	T	G	G	T	A
Line 9	T	G	G	T	A
Line 10	T	G	G	T	A
Line 11	T	G	G	T	A
Line 12	T	G	G	T	A
Line 13	T	G	G	T	A
Line 14	T	G	G	T	A
Line 15	T	G	G	T	A
Line 16	T	G	G	T	A
Line 17	T	G	G	T	A
Line 18	T	G	G	T	A
Line 19	T	G	G	T	A
Line 20	T	G	G	T	A
Line 21	T	G	G	T	A
Line 22	T	G	G	T	A
Line 23	T	G	G	T	A
Line 24	T	G	G	T	A
Line 25	T	G	G	T	A
Line 26	T	G	G	T	A
Line 27	T	G	G	T	A
Line 28	T	G	G	T	A
Line 29	T	G	G	T	A
Line 30	T	G	G	T	A
Line 31	T	G	G	T	A
Line 32	T	G	G	T	A
Line 33	T	G	G	T	A
Line 34	T	G	G	T	A
Line 35	T	G	G	T	A
Line 36	T	G	G	T	A
Line 37	T	G	G	T	A
Line 38	T	G	G	T	A
Line 39	T	G	G	T	A
Line 40	T	G	G	T	A
Line 41	T	G	G	T	A
Line 42	T	G	G	T	A
Line 43	T	G	G	T	A
Line 44	T	G	G	T	A
Line 45	T	G	G	T	A
Line 46	T	G	G	T	A
Line 47	T	G	G	T	A
Line 48	T	G	G	T	A
Line 49	T	G	G	T	A
Line 50	T	G	G	T	A

Genotypes
from
unrelated
individuals

Quality control (QC)

- Relatedness (K matrix)
- Population structure (Q or P matrix)
- Minor allele frequency (MAF)
- Heterozygosity

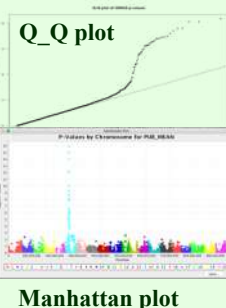


Phenotypes

Statistical analysis

- 1- Genotypes
- 2- Phenotypes
- 3- K matrix
- 4- P matrix

GWAS results visualization



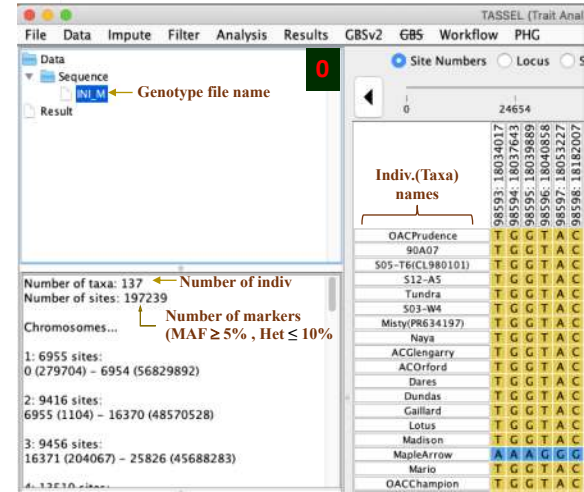
How to carry out a GWAS

TASSEL v 5.2.50

Step I SNP data filtering

**Covered in previous
presentations**

Sorting taxa



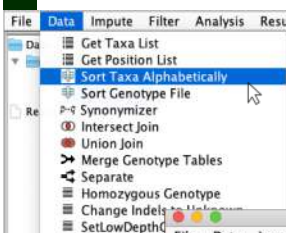
Genotype file name

Indiv.(Taxa) names

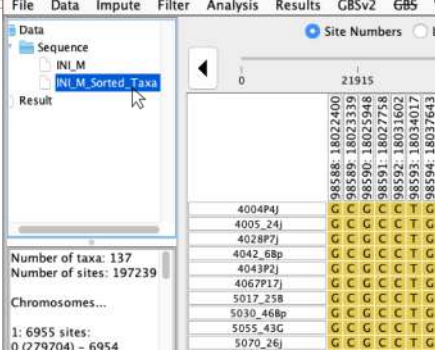
Number of taxa: 137
Number of sites: 197239

Chromosomes...
1: 6955 sites:
0 (279704) - 6954 (56829892)
2: 9416 sites:
6955 (1104) - 16370 (48570528)
3: 9456 sites:
16371 (204067) - 25826 (45688283)
4: 13510 sites:

1 Getting taxa sorted



Taxa sorted view



INI_M_Sorted_Taxa

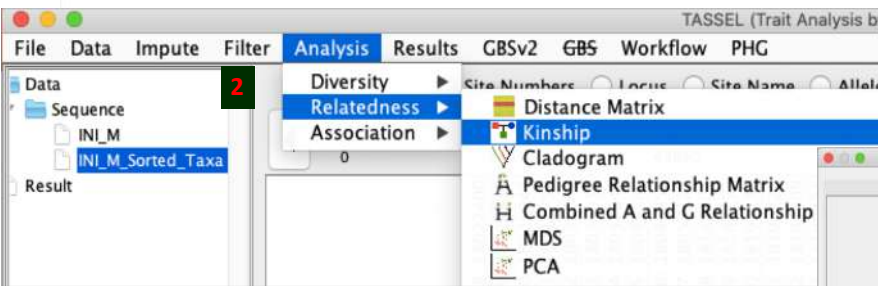
Number of taxa: 137
Number of sites: 197239

Chromosomes...
1: 6955 sites:
0 (279704) - 6954 (56829892)
2: 9416 sites:
6955 (1104) - 16370 (48570528)
3: 9456 sites:
16371 (204067) - 25826 (45688283)
4: 13510 sites:

www.soyagen.ca

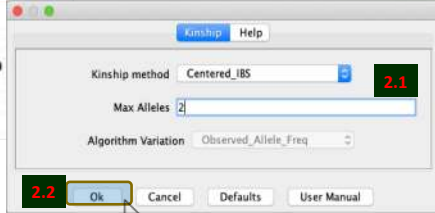
Kinship analysis (K matrix)

Step II



2

Kinship



2.1

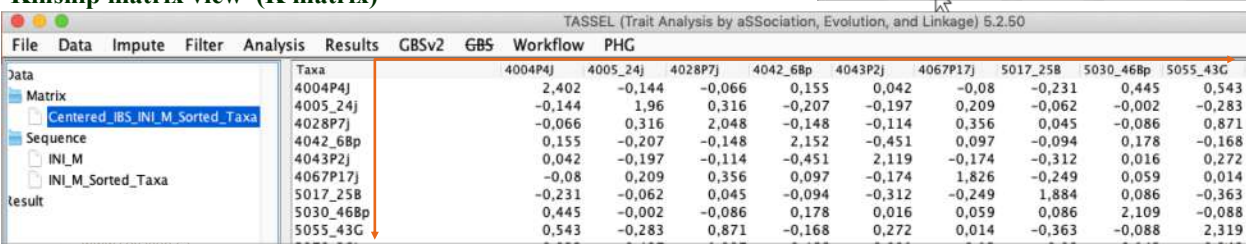
Kinship method: Centered_IBS

Max Alleles: 2

Algorithm Variation: Observed_Allele_Freq

2.2 Ok

Kinship matrix view (K matrix)



Centered_IBS_INI_M_Sorted_Taxa

Number of taxa: 137
Number of sites: 197239

Chromosomes...
1: 6955 sites:
0 (279704) - 6954 (56829892)
2: 9416 sites:
6955 (1104) - 16370 (48570528)
3: 9456 sites:
16371 (204067) - 25826 (45688283)
4: 13510 sites:

www.soyagen.ca

Pop'n structure analysis (P matrix) Step II

Pop. structure matrix view (P matrix)

Taxa	PC1	PC2	PC3	PC4	PC5
4004P4J	-13,753	25,04	-104,11	39,693	-33,721
4005_24J	20,365	50,452	59,068	23,05	-11,298
4028P7J	33,033	103,32	53,705	4,267	19,848
4042_68p	73,517	-31,378	-27,872	-55,461	-31,098
4043P2J	-68,359	-0,936	-28,27	25,06	39,909
4067P17J	26,961	74,277	5,985	-31,216	25,955
5017_25B	-6,55	2,357	61,803	27,702	-63,865
5030_468n	14,236	21,56	-57,581	-10,284	-72,141

End of step II

Step I

Line 1 T G G T A
Line 2 T G G T A
Line 3 T G G T A
Line 4 T G G T A
Line 5 T G G T A
Line n T G G T A

Genotypes from unrelated individuals

Quality control (QC)

- Relatedness (K matrix)
- Population structure (Q or P matrix)
- Minor allele frequency (MAF)
- Heterozygosity

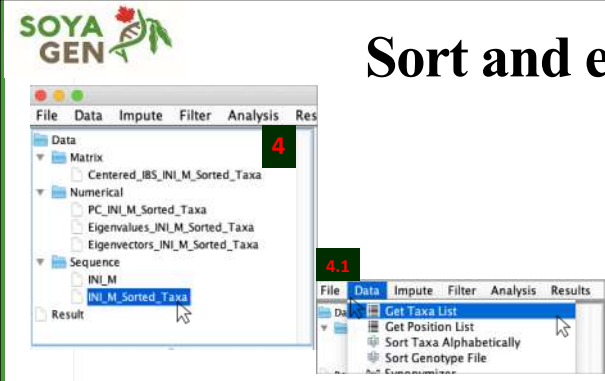
Step II

Number of taxa: 137
Number of sites: 197239 (MAF ≥ 5% , Het ≤ 10%)
Chromosomes...

NB: MAF and heterozygosity were done before

Sort and export taxa list

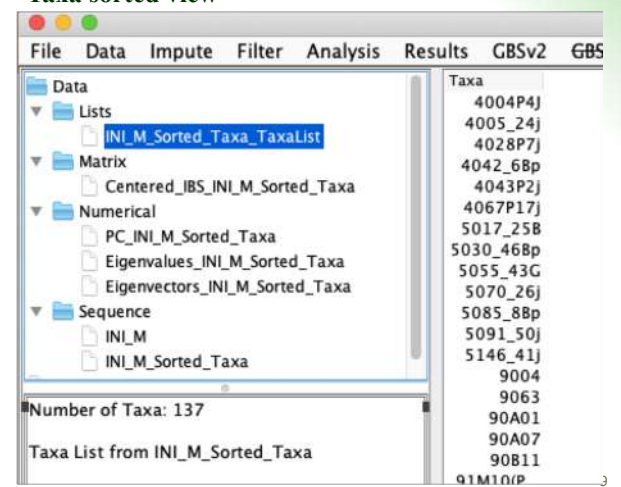
Step III



4

4.1

Taxa sorted view



File Data Impute Filter Analysis Results GBSv2 GBS

Data

- Lists
 - INI_M_Sorted_Taxa_TaxaList
- Matrix
 - Centered_IBS_INI_M_Sorted_Taxa
- Numerical
 - PC_INI_M_Sorted_Taxa
 - Eigenvalues_INI_M_Sorted_Taxa
 - Eigenvectors_INI_M_Sorted_Taxa
- Sequence
 - INI_M
 - INI_M_Sorted_Taxa

Number of Taxa: 137

Taxa List from INI_M_Sorted_Taxa

Taxa

4004P4J
4005_24J
4028P7J
4042_68p
4043P2J
4067P17J
5017_25B
5030_468p
5055_43G
5070_26J
5085_88p
5091_50J
5146_41J
9004
9063
90A01
90A07
90B11
91M10/P

www.soyagen.ca

Sort and export taxa list

Step III

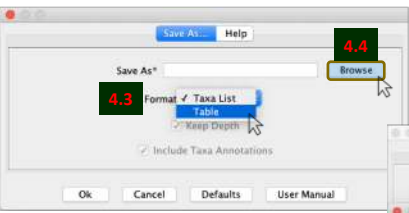


4.2

File Data Impute Filter An

- Open
- Open As...
- Save As...
- Delete Dataset
- Preferences
- Show Parameter Cache

Exit



4.3

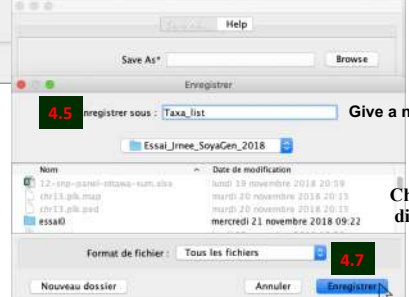
4.4

Save As*

Format: Taxa List

Include Taxa Annotations

Ok Cancel Defaults User Manual



4.5

Enregistrer

Enregistrer sous: Taxa_list

Essai_Insee_SoyaGen_2018

Nom

Nom	Date de modification
12-rnp-panel-ottawa-rum.xlsx	lundi 19 novembre 2018 20:09
chr13.gls.msap	mardi 20 novembre 2018 20:15
chr13.gls.ped	mardi 20 novembre 2018 20:15
essai0	mercredi 21 novembre 2018 09:22

Format de fichier: Tous les fichiers

Nouveau dossier Annuler Enregistrer

Give a name

Choose your working directory

saved

www.soyagen.ca

SOYA GEN Step III

Sort individuals in phenotype file

Open phenotype file in Excel (v 16.19 was used)

Click «données»

5.2

5.3

select Trait

5.4

5.1

Select individuals and phenotype values columns

Indiv. Sorted view

NB
One phenotypic value for each individual

11

SOYA GEN Step III

Check names in genotype and phenotype files in Excel (Must be identical)

Phenotype file

Open genotype file

6.1

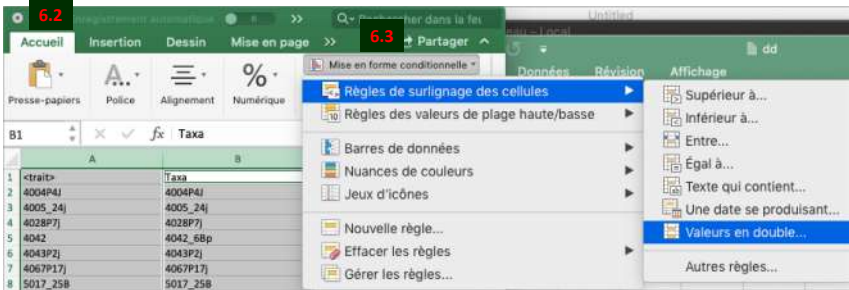
Copy and paste

Phenotype file

12

Check names in genotype and phenotype files in Excel

6.2 Sélectionner les colonnes A et B



6.3 Menu Accueil > Mise en forme conditionnelle > Règles de surlignage des cellules > Valeurs en double...

6.4 Nouvelle règle de mise en forme

Style : Classique

Appliquer une mise en forme uniquement aux valeurs uniques...

valeurs en double dans la plage sélectionnée

Mettre en forme avec : Remplissage rouge...

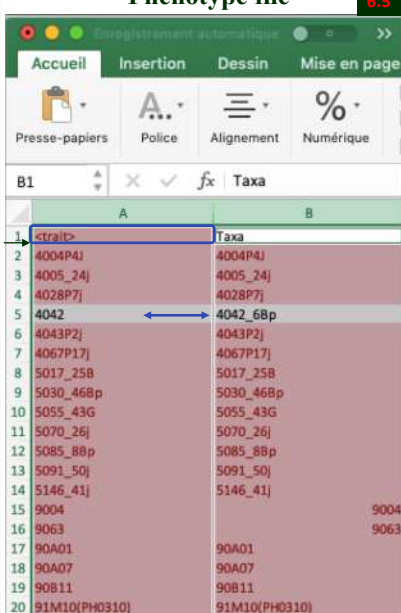
6.4 OK

Select both columns

www.soyagen.ca

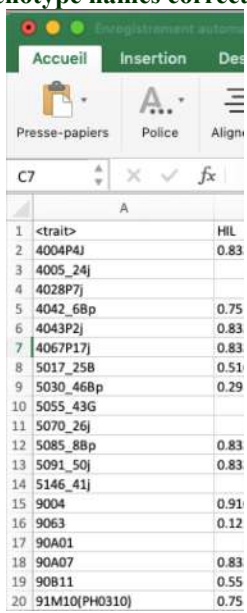
Check names in genotype and phenotype files in Excel

6.5 Phenotype file



headers

6.6 Phenotype names corrected



Correct name according to genotype file

6.6

- ✓ Delete the column of genotype name
- ✓ Check
 - Phenotype file header
 - Number of individuals must be the same in all files
- ✓ Saved phenotype file (.txt)

Step III

Load phenotype checked and saved files

Format de fichier : Tous les fichiers

Annuler Ouvrir

Loaded phenotype view in TASSEL

www.soyagen.ca

SOYA GEN

End of step III

Step I

Genotypes from unrelated individuals

II

Quality control (QC)

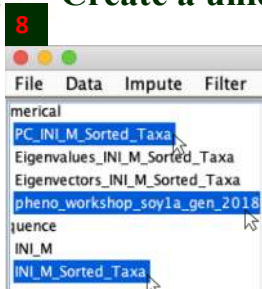
- Relatedness (K matrix)
- Population structure (Q or P matrix)
- Minor allele frequency (MAF)
- Heterozygosity

III

Phenotypes

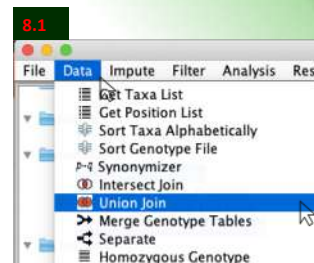
www.soyagen.ca

Create a unique joint file with: phenotype + PCs + genotype Step IV



NB: to select the three files

- Hold « cmd (Mac) » or « ctrl (Windows) »
- select PC
- select Phenotype
- select Genotype



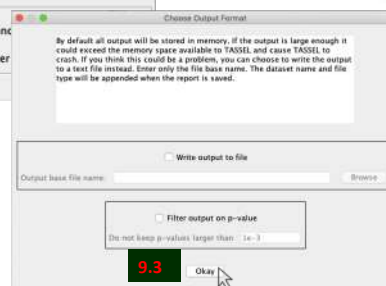
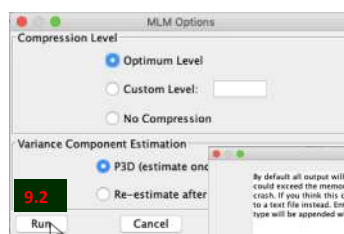
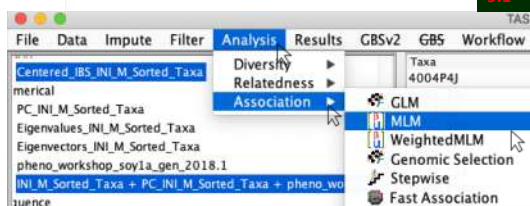
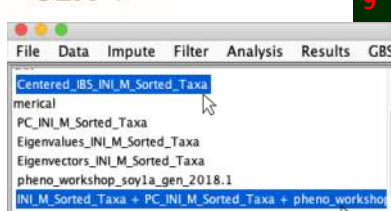
Joint file view

Taxa	PC1	PC2	PC3	PC4	PC5	HIL	Genotype
4004P4J	-13,753	25,04	-104,11	39,693	-33,721	0,833	G:C:C:G:G:A:T:C:T:A...
4005_24J	20,365	50,452	59,068	23,05	-11,298	1	G:C:C:G:G:A:T:C:T:A...
4028P7J	33,033	103,32	53,705	4,267	19,848	1	G:C:C:G:G:A:T:C:T:A...
4042_68P	73,517	-31,378	-27,872	-55,461	-31,098	0,75	A:A:G:T:A:G:A:T:C:T...
4043P2J	-68,359	-0,936	-28,27	25,06	39,909	0,833	A:A:G:T:A:G:A:T:C:T...
4067P17J	26,961	74,277	5,985	-31,216	25,955	0,833	G:C:C:G:G:A:T:C:T:A...
5017_25P	-6,55	2,357	61,803	27,702	-63,865	0,517	A:A:G:T:A:G:A:T:C:T...

www.soyagen.ca

17

Launching GWAS using MLM Step IV



MLM analysis in progress



www.soyagen.ca

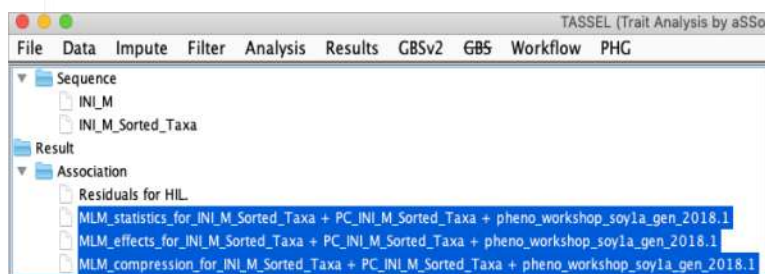
18



End of step IV

Output of the MLM analysis

For more info : <http://www.maizegenetics.net/tassel>.



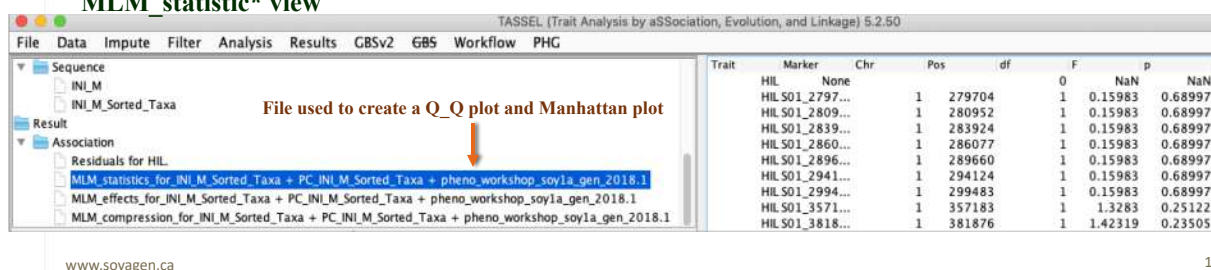
Three files are created with prefixes

-MLM_statistics*

-MLM_effects*

-MLM_compression*

MLM_statistic* view



www.soyagen.ca

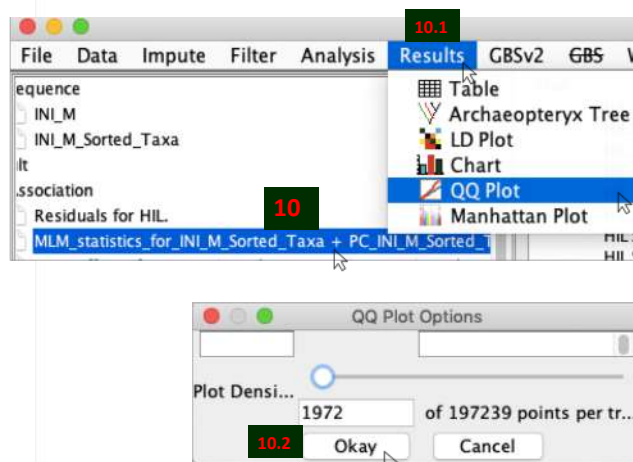
19



Visualization of GWAS results

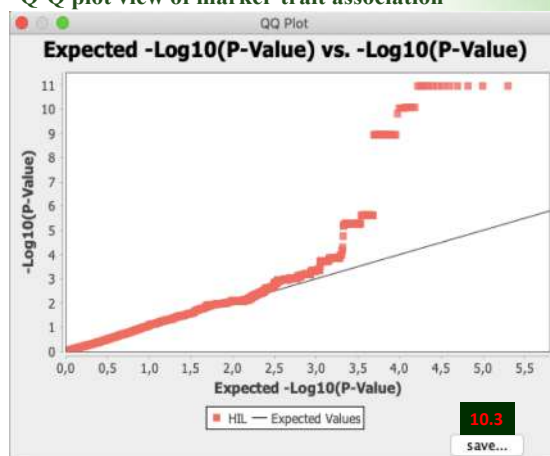
Step V

Association distribution, quantile-quantile (Q-Q) plots of observed P values versus expected P values



Press Okay

Q-Q plot view of marker-trait association



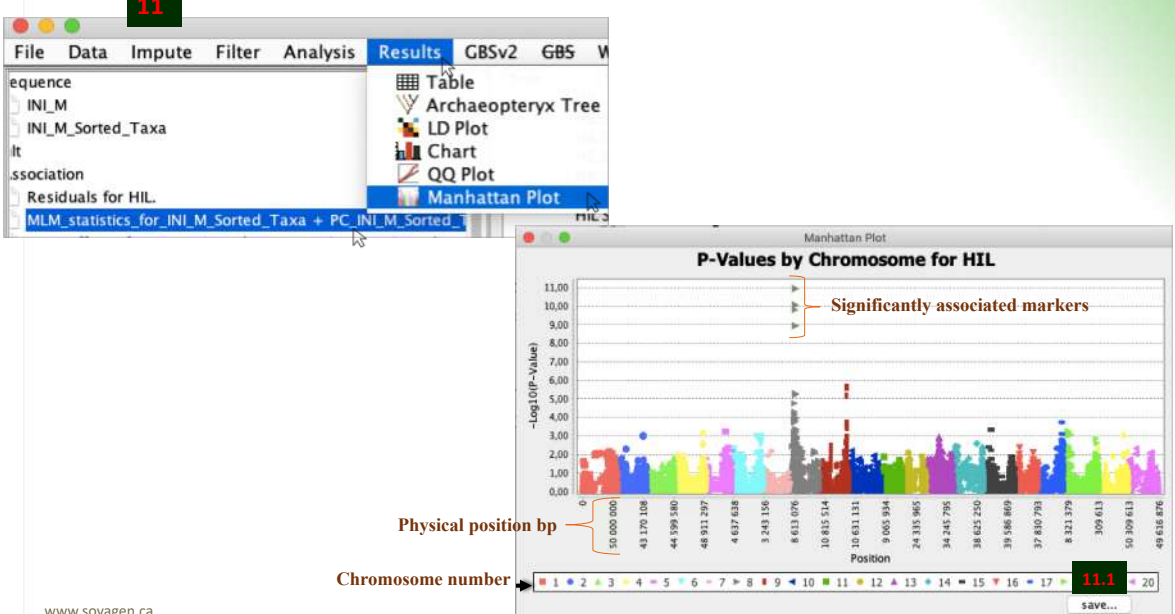
www.soyagen.ca

20

Visualization of GWAS results

Step V

11



Physical position bp

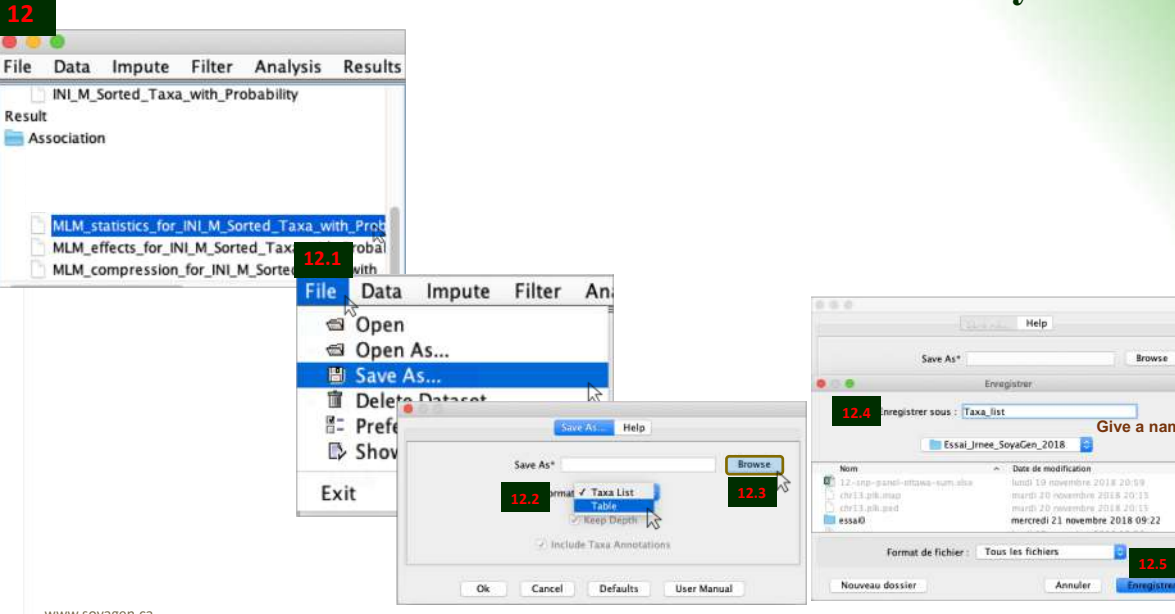
Chromosome number

11.1

21

Save GWAS Results files for further analysis

12



12.1

12.2

12.3

12.4

12.5

22