

A tutorial to perform Association mapping analysis using TASSEL v 3.0 software

By

Dr. M. Awais Khan

University of Illinois, Urbana-Champaign

TASSEL software can be freely downloaded from www.maizegenetics.net website

In TASSEL software, two methods are implemented to perform association analysis

1. General Linear Model (GLM): Associations between markers and mean phenotypic values are identified using the population membership estimates as covariates to control for population structure. The GLM does not account for kinship as a potential cause of the genotype-phenotype relationship.

2. Mixed Linear Model (MLM): It takes account of population structure and kinship in the association analysis. It reduces Type I error due to relatedness and population structure.

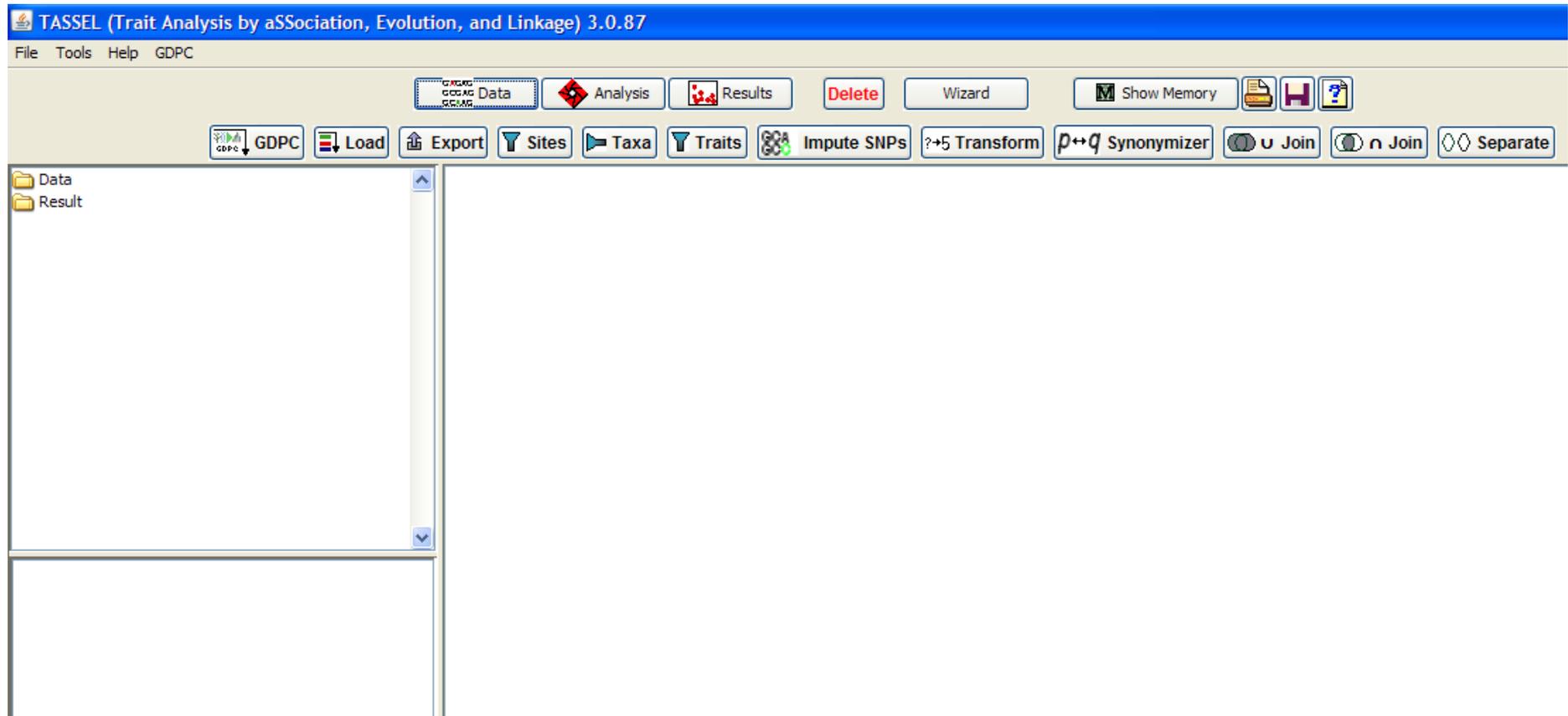
General Linear Model (GLM)

To perform GLM analysis, we need to load marker, trait, and population structure files into TASSEL

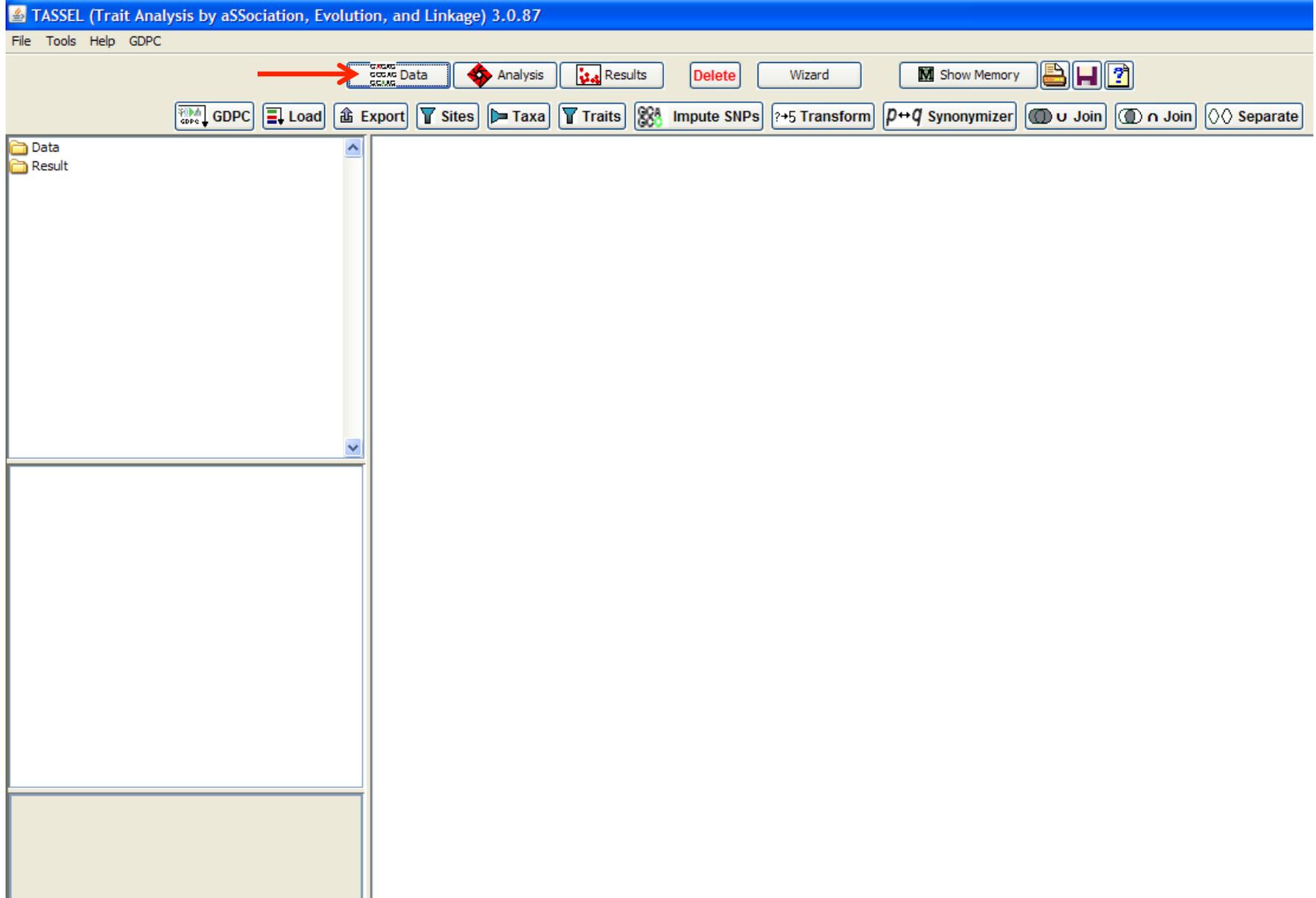
GLM analysis accounts only for population structure in the association analysis.

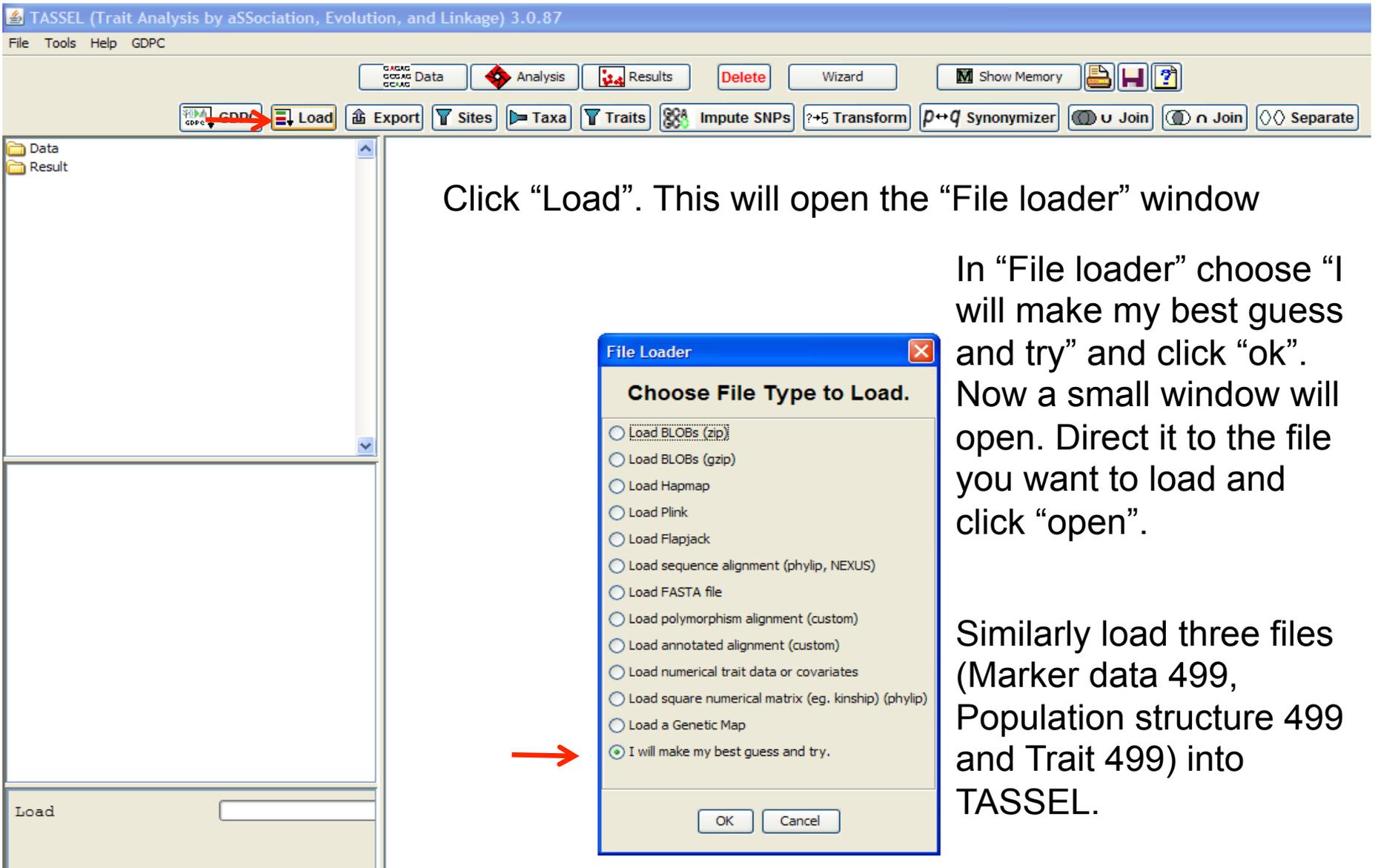
First, download TASSEL software from the www.maizegenetics.net website and install on your computer

Double click “TASSEL” to start the software



Click “Data” to start loading the data file into TASSEL





Click “Load”. This will open the “File loader” window

In “File loader” choose “I will make my best guess and try” and click “ok”. Now a small window will open. Direct it to the file you want to load and click “open”.

Similarly load three files (Marker data 499, Population structure 499 and Trait 499) into TASSEL.

Input files can be text delimited (.txt). For more information on the required input file layout, open the files provided with this tutorial

Right click the three files to highlight them and click “U Join” to join the three files

TASSEL (Trait Analysis by aSSociation, Evolution, and Linkage) 3.0.87

File Tools Help GDPC

Physical Positions
 Site Numbers
 Locus
 Site Name
 Alleles

0 68 136 204 272 340 408

	231: 231	232: 232	233: 233	234: 234	235: 235	236: 236	237: 237	238: 238	239: 239	240: 240	241: 241	242: 242	243: 243	244: 244
MadBSd032310	1:01	2:02	1:02	1:02	1:02	1:02	2:02	2:02	1:02	2:02	N	2:02	1:02	1:01
MadBSd032311	1:01	2:02	1:02	1:02	1:02	2:02	2:02	2:02	1:02	2:02	N	1:02	1:02	1:01
MadBSd032312	1:01	2:02	1:01	2:02	1:01	2:02	2:02	2:02	1:01	2:02	1:02	2:02	1:02	1:01
MadBSd032313	1:01	2:02	1:02	1:01	1:01	1:02	1:02	2:02	N	2:02	2:02	1:02	1:01	2:02
MadBSd032314	1:01	2:02	1:01	2:02	1:01	2:02	1:02	2:02	1:02	2:02	1:02	1:01	2:02	1:02
MadBSd032316	1:01	2:02	1:02	1:02	1:01	1:02	2:02	2:02	1:01	2:02	2:02	2:02	1:02	1:02
MadBSd032317	1:01	2:02	1:01	2:02	1:02	2:02	1:02	2:02	1:01	2:02	1:02	2:02	1:02	1:01
MadBSd032320	1:01	2:02	1:02	1:01	1:01	1:02	2:02	2:02	2:02	2:02	1:02	2:02	2:02	1:01
MadBSd032321	1:01	2:02	2:02	1:01	1:02	1:02	2:02	2:02	1:01	2:02	1:02	1:02	1:02	1:02
MadBSd032322	1:01	2:02	1:02	1:02	1:01	2:02	1:02	1:02	1:02	2:02	N	2:02	1:02	1:01
MadBSd032323	1:01	2:02	1:02	1:02	1:01	2:02	2:02	2:02	1:02	2:02	1:02	2:02	1:02	1:02
MadBSd032324	1:02	2:02	1:02	1:02	1:01	1:02	2:02	2:02	1:01	2:02	1:02	2:02	1:01	1:01
MadBSd032325	1:01	2:02	2:02	1:01	1:01	1:02	2:02	2:02	1:01	2:02	1:02	2:02	1:02	1:02
MadBSd032326	1:01	2:02	1:02	1:02	1:01	1:02	1:02	2:02	2:02	2:02	2:02	2:02	2:02	1:01
MadBSd032327	1:01	2:02	1:02	1:02	1:01	2:02	2:02	2:02	1:01	2:02	1:02	2:02	1:01	1:01
MadBSd032328	1:01	2:02	2:02	1:01	1:01	1:02	2:02	2:02	2:02	1:02	1:02	2:02	1:02	1:01
MadBSd032329	1:01	2:02	1:01	2:02	1:01	2:02	2:02	2:02	2:02	2:02	1:02	2:02	2:02	1:01
MadBSd032382	1:01	2:02	1:02	1:02	1:01	1:01	2:02	2:02	2:02	1:02	1:02	2:02	1:02	1:02
UgASd034311	1:01	2:02	1:02	1:02	1:01	1:02	2:02	2:02	2:02	1:02	1:02	2:02	2:02	1:01
UgASd034312	1:01	2:02	1:02	1:01	1:01	2:02	2:02	2:02	1:02	2:02	1:02	2:02	1:02	1:02
UgASd034313	1:02	2:02	1:02	1:02	1:01	1:02	2:02	2:02	1:02	2:02	1:02	1:02	1:02	1:01
UgASd034314	1:01	2:02	1:02	1:02	1:02	2:02	2:02	1:02	1:01	2:02	2:02	2:02	2:02	1:02
UgASd034315	1:01	2:02	1:01	2:02	1:01	2:02	2:02	2:02	1:02	1:02	N	1:02	2:02	1:01
UgASd034316	1:02	2:02	1:01	1:01	1:01	2:02	1:02	2:02	1:01	2:02	1:02	1:02	2:02	1:02
UgASd034317	1:01	2:02	1:01	2:02	1:01	2:02	2:02	2:02	1:02	2:02	1:02	2:02	1:02	1:01
UgASd034318	1:01	2:02	1:02	1:02	1:01	1:02	2:02	2:02	1:02	2:02	2:02	2:02	1:02	2:02
UgASd034320	1:01	2:02	1:01	2:02	1:01	1:01	2:02	2:02	1:02	2:02	1:01	2:02	1:02	1:02
UgASd034321	1:01	2:02	1:01	2:02	1:01	2:02	2:02	1:02	1:01	2:02	1:01	2:02	1:01	1:02

Number of sequences: 499
 Number of sites: 476
 Data type: Text
 Loci: Unknown

TASSEL (Trait Analysis by aSSociation, Evolution, and Linkage) 3.0.87

File Tools Help GDPC

GDPC Data Analysis Results Delete Wizard Show Memory

GDPC Load Export Sites Taxa Traits Impute SNPs Transform Synonymizer Join Join Separate

Data

- Sequence
 - Marker data 499
 - Marker data 499 + Population structure 499 +
- Polymorphisms
- Numerical
 - Population structure 499
 - Trait 499
- Matrix
- Tree
- Fusions
- Synonymizer
- Result

Taxa	Cluster 1	Height	Haplotype
CutSd035071	0.982	79	ABBCCAAA...
CutSd035072	0.186	76.67	ABBACABAB...
CutSd035073	0.91	72.33	ABBAAAABBB...
CutSd035074	0.984	97.5	ABBAAAAAB...
CutSd035075	0.977	72	ABBCCCBAB...
CutSd035076	5.4E-2	79.33	CBBACCBAB...
CutSd035077	0.412	85	BABBACBBB...
CutSd035078	0.929	70	ABBCCCBBB...
CutSd035079	0.987	74	ABBACABAB...
CutSd035081	0.988	70.33	CBBCCCBAB...
CutSd035085	0.145	89	CBBCCCBAB...
CutSd035087	0.731	90.33	ABBACCBAB...
CutSd035088	0.839	93	BBBACCACB...
CutSd035089	5.6E-2	72.33	ABBCAACB...
CutSd035090	0.963	71.33	ABBAACAN...
CutSd035291	0.965	92.33	AABAACBCB...
CutSd035292	0.983	85.67	CBBAACAAB...
CutSd035294	0.973	76	ABBACABBB...
CutSd035295	0.964	83.33	ABBCCCAAB...
CutSd035296	0.746	76	CBBAACBBB...
CutSd035297	0.819	61	CBBACAAB...
CutSd035298	0.165	72	ABBACAAB...
CutSd035299	0.872	78.33	CBBCAABBB...
CutSd035302	0.968	68	ABBACABAB...
CutSd035303	0.924	98.33	ABBACABAB...
CutSd035304	0.153	88.33	ABBACCBCB...
CutSd035305	0.921	83.33	CBBCCCAAB...
CutSd035306	0.722	73.67	CBBABCAB...
CutSd035307	0.978	64.67	ABBACABAB...
CutSd035308	0.932	66	ABBACABAB...
CutSd035311	0.974	82	ABBCCCBAB...
CutSd035312	0.986	75.67	ABBACBCB...
CutSd035313	0.624	73	CBBACAAB...
CutSd035316	0.982	69.33	ABBAACBCB...
CutSd035317	0.869	69.67	BBBACABAB...
CutSd035318	0.903	77	ABBACCABB...

Number of sequences: 499
 Number of sites: 476
 Data type: Text
 Loci: Unknown
 Table Title: Phenotypes and Genotypes
 Number of columns: 4
 Number of rows: 499
 Number of elements: 1996
 Union Join

Now click “New Created File” and check to see if the files joined correctly by making sure that the genotypes (Taxa) in the new file correspond with the respective data of the original file.

Afterwards, click the “Analysis” tab to begin association mapping analysis



Data



Analysis



Results

Delete

Wizard

Show Memory



Diversity



Link. Diseq.



Cladogram



Kinship



GLM



MLM



GS


 Data

- Sequence
 - Marker data 499
 - Marker data 499 + Population structure 499 +**
 - Polymorphisms
- Numerical
 - Population structure 499
 - Trait 499
- Matrix
- Tree
- Fusions
- Synonymizer
- Result

Number of sequences: 499
 Number of sites: 476
 Data type: Text
 Loci: Unknown
 Table Title: Phenotypes and Genotypes
 Number of columns: 4
 Number of rows: 499
 Number of elements: 1996
 Union Join

Taxa	Cluster1	Height	Haplotype
CutSd035071	0.982	79	ABBCCAAA...
CutSd035072	0.186	76.67	ABBACABAB...
CutSd035073	0.91	72.33	ABBAABBB...
CutSd035074	0.984	97.5	ABBAAAAAB...
CutSd035075	0.977	72	ABBCCCBAB...
CutSd035076	5.4E-2	79.33	CBBACCBAB...
CutSd035077	0.412	85	BABBACBBB...
CutSd035078	0.929	70	ABBCCCBBB...
CutSd035079	0.987	74	ABBACABAB...
CutSd035081	0.988	70.33	CBBCCCBAB...
CutSd035085	0.145	89	CBBCCCBAB...
CutSd035087	0.731	90.33	ABBACCBAB...
CutSd035088	0.839	93	BBBACCACB...
CutSd035089	5.6E-2	72.33	ABBBCAACB...
CutSd035090	0.963	71.33	ABBAACAN...
CutSd035291	0.965	92.33	AABAACBCB...
CutSd035292	0.983	85.67	CBBAACAAB...
CutSd035294	0.973	76	ABBACBBB...
CutSd035295	0.964	83.33	ABBCCCAAB...
CutSd035296	0.746	76	CBBAACBBB...
CutSd035297	0.819	61	CBBCACAAB...
CutSd035298	0.165	72	ABBACCAAB...
CutSd035299	0.872	78.33	CBBCAABBB...
CutSd035302	0.968	68	ABBACABAB...
CutSd035303	0.924	98.33	ABBACABAB...
CutSd035304	0.153	88.33	ABBACCBAB...
CutSd035305	0.921	83.33	CBBCCCAAB...
CutSd035306	0.722	73.67	CBBABCBAB...
CutSd035307	0.978	64.67	ABBACABAB...
CutSd035308	0.932	66	ABBACABAB...
CutSd035311	0.974	82	ABBCCCBAB...
CutSd035312	0.986	75.67	ABBACABCB...
CutSd035313	0.624	73	CBBAACAAB...
CutSd035316	0.982	69.33	ABBAACBCB...
CutSd035317	0.869	69.67	BBBACABAB...


 With the joint file selected, click the “GLM” tab to perform association mapping analysis using GLM

TASSEL (Trait Analysis by aSSociation, Evolution, and Linkage) 3.0.87

File Tools Help GDCP

Analysis Results Delete Wizard Show Memory

Diversity Link. Diseq. Cladogram Kinship **GLM** MLM GS

Data
 Sequence
 Marker data 499
 Marker data 499 + Population structure 499 +
 Polymorphisms
 Numerical
 Population structure 499
 Trait 499
 Matrix
 Tree
 Fusions
 Synonymizer
 Result

Taxa	Cluster 1	Height	Haplotype
CutSd035071	0.982	79	ABBCCAAA...
CutSd035072	0.186	76.67	ABBACABAB...
CutSd035073	0.91	72.33	ABBAAABBB...
CutSd035074	0.984	97.5	ABBAAAAAB...
CutSd035075	0.977	72	ABBCCCBAB...
CutSd035076	5.4E-2	79.33	CBBACCBAB...
CutSd035077	0.412	85	BABBACBBB...
CutSd035078	0.929	70	ABBCCCBBB...
CutSd035079	0.987	74	ABBACABAB...
CutSd035081	0.988	70.33	CBCCCCBAB...
CutSd035085	0.145	89	CBCCCCBAB...
CutSd035087	0.731	90.33	ABBACCBAB...
CutSd035088	0.839	93	BBBACCACB...
CutSd035089	5.6E-2	72.33	ABBBCAACB...
CutSd035090	0.963	71.33	ABBAACAN...
CutSd035291	0.965	92.33	AABAACBCB...
CutSd035292	0.983	85.67	CBBAACAAB...
CutSd035294	0.973	76	ABBACABBB...
CutSd035295	0.964	83.33	ABBCCAAB...
CutSd035296	0.746	76	CBBAACBBB...
CutSd035297	0.819	61	CBBCACAAB...
CutSd035298	0.165	72	ABBACCAAB...
CutSd035299	0.872	78.33	CBBCAABBB...
CutSd035302	0.968	68	ABBACABAB...
CutSd035303	0.924	98.33	ABBACABAB...
CutSd035304	0.153	88.33	ABBACCBAB...
CutSd035305	0.921	83.33	CBCCCCAAB...
CutSd035306	0.722	73.67	CBBABCBAB...
CutSd035307	0.978	64.67	ABBACABAB...
CutSd035308	0.932	66	ABBACABAB...
CutSd035311	0.974	82	ABBCCCBAB...
CutSd035312	0.986	75.67	ABBACABCB...
CutSd035313	0.624	73	CBBAACAAB...
CutSd035316	0.982	69.33	ABBAACBCB...
CutSd035317	0.869	69.67	BBBACABAB...
CutSd035318	0.903	77	ABBACABB...
CutSd035319	0.242	68.33	CBBAACBNB...

Number of sequences: 499
 Number of sites: 476
 Data type: Text
 Loci: Unknown
 Table Title: Phenotypes and Genotypes
 Number of columns: 4
 Number of rows: 499
 Number of elements: 1996
 Union Join

GLM

In the "GLM Options" window, specify the number of permutations as 1000 and click "OK"

GLM Options

Use permutation test for markers

Number of Permutations: 1000

OK

Data
 Analysis
 Results
 Delete
 Wizard
 Show Memory

Diversity
 Link. Diseq.
 Cladogram
 Kinship
 GLM
 MLM
 GS

Data
 Sequence
 ● Marker data 499
 ● Marker data 499 + Population structure 499 +
 ● Polymorphisms
 Numerical
 ● Population structure 499
 ● Trait 499
 ● Matrix
 ● Tree
 ● Fusions
 ● Synonymizer
 Result

Number of sequences: 499
 Number of sites: 476
 Data type: Text
 Loci: Unknown
 Table Title: Phenotypes and Genotypes
 Number of columns: 4
 Number of rows: 499
 Number of elements: 1996
 Union Join

Taxa	Cluster1	Height	Haplotype
CutSd035071	0.982	79	ABBBCAAA...
CutSd035072			
CutSd035073			
CutSd035074			
CutSd035075			
CutSd035076			
CutSd035077			
CutSd035078			
CutSd035079			
CutSd035080			
CutSd035081			
CutSd035082			
CutSd035083			
CutSd035084			
CutSd035085			
CutSd035086			
CutSd035087			
CutSd035088			
CutSd035089			
CutSd035090			
CutSd035091			
CutSd035092			
CutSd035093			
CutSd035094			
CutSd035095			
CutSd035096			
CutSd035097			
CutSd035098			
CutSd035099			
CutSd035100			
CutSd035101			
CutSd035102			
CutSd035103			
CutSd035104			
CutSd035105			
CutSd035106			
CutSd035107			
CutSd035108			
CutSd035109			
CutSd035110			
CutSd035111			
CutSd035112			
CutSd035113			
CutSd035114			
CutSd035115			
CutSd035116			
CutSd035117			
CutSd035118			
CutSd035119			
CutSd035120			
CutSd035121	0.985	169.67	ABRACRRB...

By default all output will be stored in memory. If the output is large enough it could exceed the memory space available to TASSEL and cause TASSEL to crash. If you think this could be a problem, you can choose to write the output to a text file instead. Enter only the file base name. The dataset name and file type will be appended when the report is saved.

Write output to file

Output base file name:

Filter output on p-value

Do not keep p-values larger than

GLM

File Tools Help GDFC

Data Analysis Results Delete Wizard Show Memory

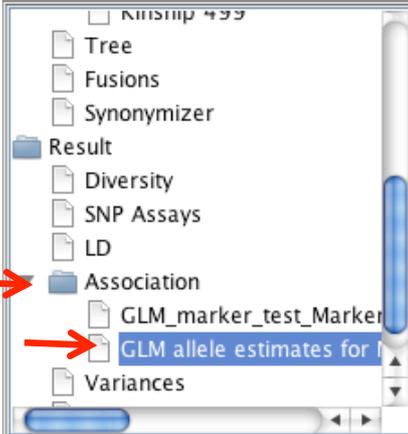
Diversity Link. Diseq. Cladogram Kinship GLM MLM GS

Trait	Marker	Locus	Locus_pos	marker_F	marker_p	markerR2	markerDF	markerMS	errorDF	errorMS	modelDF	modelMS
Freshweight	M76	Unknown	75 22.28985	5.4124E-...	8.259E-2	2	3.5224E6	492	1.5803E5	3	2.5164E6	
Freshweight	M223	Unknown	222 17.26153	3.8658E-5	3.4935E-2	1	2.829E6	473	1.6389E5	2	1.7289E6	

Table Title: Marker Test
 Number of columns: 13
 Number of rows: 2
 Number of elements: 26
 Tests of Marker-Phenotype Association
 Data set: Marker data 499 + Population
 model: trait = mean + Cluster1 + ...

Click the “Association” tab under the “Result” table. There are two output files from GLM analysis “GLM_marker_test...” and “GLM allele estimate...”

The “GLM_marker_test...” file identifies two markers (M76 and M223) as associated with the trait “Freshweight” at the significance threshold chosen (1-e3) chosen in the previous slide.



Trait	Marker	Obs	Locus	Locus_pos	Allele	Estimate
Freshweight	M76	164	Unknown	75	2:02	-2.6362E2
Freshweight	M76	234	Unknown	75	1:02	-1.3974E1
Freshweight	M76	98	Unknown	75	1:01	0
Freshweight	M223	467	Unknown	222	2:02	566.4604
Freshweight	M223	9	Unknown	222	1:02	0



The “GLM allele estimates...” file provides effect estimates for each genotypic class (homozygous or heterozygous) for the markers associated with freshweight.

title: Allele Estimates
 number of columns: 7
 number of rows: 5
 number of elements: 35
 allele effect estimates
 fixed effect linear model
 data set: Marker data 499 + Population s
 trait = mean + Cluster1 + marker

Mixed Linear Model (MLM)

To perform MLM analysis, a kinship matrix is required in addition to the files required for GLM

MLM analysis includes both population structure and kinship in the association analysis. It reduces Type I error due to relatedness and population structure.

Click “Data” to load the kinship file

The screenshot shows the TASSEL (Trait Analysis by aSSociation, Evolution, and Linkage) 3.0.87 interface. The title bar is blue with the text 'TASSEL (Trait Analysis by aSSociation, Evolution, and Linkage) 3.0.87'. Below the title bar is a menu bar with 'File', 'Tools', 'Help', and 'GDPC'. The main toolbar contains several buttons: 'Data', 'Analysis', 'Results', 'Delete', 'Wizard', 'Show Memory', and a help icon. A red box highlights the 'Data', 'Analysis', and 'Results' buttons, with a red arrow pointing to the 'Data' button. Below this, another red box highlights the 'GDPC' and 'Load' buttons, with a red arrow pointing to the 'Load' button. The left sidebar shows a tree view with 'Data' selected, containing sub-items like 'Sequence', 'Numerical', 'Matrix', 'Tree', 'Fusions', 'Synonymizer', 'Result', 'Diversity', 'SNP Assays', and 'LD'. The main panel area is mostly empty, with the text 'Node on data tree' at the top and 'Holds the basic data structures' at the bottom. A 'Main Panel' label is located at the bottom right of the interface.

Click “Load” it will open the “File loader” window

The image shows the TASSEL (Trait Analysis by aSSociation, Evolution, and Linkage) 3.0.87 software interface. The main window has a menu bar (File, Tools, Help, GDPC) and a toolbar with buttons for Data, Analysis, Results, Delete, Wizard, Show Memory, and others. A red box highlights the Data, Analysis, and Results buttons. Below the toolbar, another red box highlights the GDPC, Load, Export, Sites, Taxa, Traits, Impute SNPs, Transform, Synonymizer, Join, and Separate buttons. The left sidebar shows a tree view of data structures, including Sequence, Polymorphisms, Numerical, Matrix, Tree, Fusions, Synonymizer, and Result. The main area displays the text "Node on data tree" and "Holds the basic data structures".

A "File Loader" dialog box is open, titled "File Loader". It contains the text "Choose File Type to Load." and a list of radio button options:

- Load BLOBs (zip)
- Load BLOBs (gzip)
- Load Hapmap
- Load Plink
- Load Flapjack
- Load sequence alignment (phylip, NEXUS)
- Load FASTA file
- Load polymorphism alignment (custom)
- Load annotated alignment (custom)
- Load numerical trait data or covariates
- Load square numerical matrix (eg. kinship) (phylip)
- Load a Genetic Map
- I will make my best guess and try.

At the bottom of the dialog box are "OK" and "Cancel" buttons. A red arrow points to the "I will make my best guess and try." option.

In “File Loader” choose “I will make my best guess and try” and click “OK”. Now a small window will open. Direct it to the file “kinship 499” and click “open”

Click the “Analysis” tab, highlight both the “kinship 499” and “three files combined previously” by right clicking. Click the “MLM” tab to perform mixed linear model analysis for association mapping.

TASSEL (Trait Analysis by aSSociation, Evolution, and Linkage) 3.0.87

File Tools Help GDPC

Analysis Results Delete Wizard Show Memory

Diversity Link. Diseq. Cladogram Kinship GLM MLM GS

Data

- Sequence
 - Marker data 499
 - Marker data 499 + Population structure 499 +
- Polymorphisms
- Numerical
 - Population structure 499
 - Trait 499
- Matrix
 - Kinship 499
- Tree
- Fusions
- Synonymizer
- Result
 - Diversity
 - SNP Assays

Table Title: Alignment Distance Matrix
 Number of columns: 500
 Number of rows: 499
 Number of elements: 249500

Taxa	CutSd035...							
CutSd035071	2	0	6.95E-2	0	0	4.1E-2	1.5E-2	0
CutSd035072	0	2	6.27E-2	9.42E-2	8.0E-2	0	0	0
CutSd035073	6.95E-2	6.27E-2	2	8.58E-2	2.05E-2	0	0	0.1739
CutSd035074	0	9.42E-2	8.58E-2	2	0	0	4.26E-2	9.2E-3
CutSd035075	0	8.0E-2	2.05E-2	0	2	0	0	0
CutSd035076	4.1E-2	0	0	0	0	2	0	1.6E-2
CutSd035077	1.5E-2	0	0	4.26E-2	0	0	2	0
CutSd035078	0	0	0.1739	9.2E-3	0	1.6E-2	0	2
CutSd035079	0	9.2E-2	0	0	9.2E-3	0	0	0.1108
CutSd035081	0	0	0	3.17E-2	0	9.03E-2	0	0
CutSd035085	7.49E-2	0	0	1.36E-2	2.6E-2	0	5.31E-2	0
CutSd035087	7.91E-2	0	0	1.61E-2	0	0	0	0
CutSd035088	0	0	0	9.7E-3	3.0E-4	1.67E-2	7.58E-2	2.49E-2
CutSd035089	1.51E-2	0.1408	0	0	1.79E-2	0	1.2E-2	0
CutSd035090	0.1306	0	7.86E-2	0	0	0	0	0
CutSd035291	6.06E-2	0	0	0.1028	6.34E-2	6.0E-3	0.1162	0
CutSd035292	2.4E-2	8.52E-2	0	1.44E-2	0	0	0	0
CutSd035294	0	0	0	0	1.12E-2	0	1.14E-2	0
CutSd035295	0	2.11E-2	4.15E-2	5.8E-2	0	3.89E-2	0	0
CutSd035296	0	0	0	6.12E-2	7.36E-2	4.21E-2	0.1147	1.98E-2
CutSd035297	0	7.4E-2	0	1.74E-2	0	0	0	0
CutSd035298	0	1.46E-2	0	0	2.08E-2	5.76E-2	0	1.86E-2
CutSd035299	0	0.1497	0.1514	6.43E-2	0	0	4.48E-2	0.1784
CutSd035302	3.61E-2	7.02E-2	0	0	0	0	2.43E-2	0.1006
CutSd035303	0	2.43E-2	0	2.3E-2	0	3.9E-3	0	3.34E-2

TASSEL (Trait Analysis by aSSociation, Evolution, and Linkage) 3.0.87

File Tools Help GDPC

Data Analysis Results Delete Wizard Show Memory

Diversity Link. Diseq. Cladogram Kinship GLM MLM GS

Data

- Sequence
 - Marker data 499
 - Marker data 499 + Population structure 499 +
- Polymorphisms
- Numerical
 - Population structure 499
 - Trait 499
- Matrix
 - Kinship 499
- Tree
- Fusions
- Synonymizer
- Result
 - Diversity
 - SNP Assays

Table Title: Alignment Distance Matrix
 Number of columns: 500
 Number of rows: 499
 Number of elements: 249500

Taxa	CutSd035...							
CutSd035071	2	0	6.95E-2	0	0	4.1E-2	1.5E-2	0
CutSd035072	0	2	6.27E-2	9.42E-2	8.0E-2	0	0	0
CutSd035073								0.1739
CutSd035074								E-2
CutSd035075								9.2E-3
CutSd035076								0
CutSd035077								1.6E-2
CutSd035078								0
CutSd035079								2
CutSd035081								0.1108
CutSd035085								0
CutSd035087								E-2
CutSd035088								0
CutSd035089	1.51E-2	0.1408	0	0	1.79E-2	0	1.2E-2	0
CutSd035090	0.1306	0	7.86E-2	0	0	0	0	0
CutSd035291	6.06E-2							3
CutSd035292	2.4E-2							0
CutSd035294	0							1.14E-2
CutSd035295	0							0
CutSd035296	0							E-2
CutSd035297	0							0.1147
CutSd035298	0							1.98E-2
CutSd035299	0							0
CutSd035302	3.61E-2							E-2
CutSd035303	0							0
CutSd035304	0							4.48E-2
CutSd035305	6.24E-2							0.1784
CutSd035306	0							2.43E-2
CutSd035307	3.61E-2							0.1006
CutSd035308	0							3
CutSd035311	6.64E-2							0
CutSd035312	5.79E-2							8.99E-2
CutSd035313	0.1032	9.93E-2	0	0	0	7.45E-2	0	0
CutSd035316	1.3E-2	3.24E-2	1.28E-2	0	0	1.02E-2	0	0
CutSd035317	0	0.115	2.2E-3	1.87E-2	8.29E-2	2.55E-2	3.09E-2	3.3E-3
CutSd035318	0	0	5.88E-2	0.1271	0	0	0	0

In the "MLM Options" window, select "Optimum Level" for Compression Level and "P3D" for Variance Component Estimation. Then click "Run".

MLM Options

Compression Level

- Optimum Level
- Custom Level:
- No Compression

Variance Component Estimation

- P3D (estimate once)
- Re-estimate after each marker

Run Cancel Help Me Choose

Table Title: MLM_statistics_for
 Number of columns: 12
 Number of rows: 4
 Number of elements: 48
 MLM statistics for compressed
 Dataset: Marker data 499 + Po
 Compression = trueCompressi
 P3D = true. Variance compone
 Model: trait = mean + Cluster1

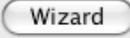
Trait	Marker	Locus	Site	df	F	p	errordf	markerR2	Genetic Var	Residual Var	-2LnLikeli...	
Freshweight	None				0	NaN	NaN	496	NaN	4.9087E4	7.5164E4	7.3548E3
Freshweight	M76	Unknown	75	2	15.89462	2.046E-7	496	6.4046E-2	4.9087E4	7.5164E4	7.3548E3	
Freshweight	M161	Unknown	160	2	7.16575	8.5606E-4	495	2.886E-2	4.9087E4	7.5164E4	7.3548E3	
Freshweight	M223	Unknown	222	1	12.68088	4.068E-4	476	2.666E-2	4.9087E4	7.5164E4	7.3548E3	

Click the “Association” tab under the “Result” tab. There are three output files from MLM analysis: “MLM_statistics...”, “MLM_effects...”, and “MLM_compression...”.

Click “MLM_statistics..” This file identifies three markers (M76, M161, M223) as significantly associated with freshweight at the significance threshold selected in the previous slide.

Note that marker “M161” was not identified using GLM analysis.

File Tools Help GDPC

 Data
  Analysis
  Results
  Delete
  Wizard
  Show Memory
 


 Diversity
  Link. Diseq.
  Cladogram
  Kinship
  GLM
  MLM
  GS

Trait	Marker	Locus	Site	Allele	Effect	Obs
Freshweight	M76	Unknown	75	2:02	-2.3856E2	164
Freshweight	M76	Unknown	75	1:02	-2.1142E1	234
Freshweight	M76	Unknown	75	1:01	0	98
Freshweight	M161	Unknown	160	2:02	127.26722	174
Freshweight	M161	Unknown	160	1:02	182.81396	208
Freshweight	M161	Unknown	160	1:01	0	113
Freshweight	M223	Unknown	222	2:02	520.04945	467
Freshweight	M223	Unknown	222	1:02	0	9

Table Title: MLM_effects_for_M
 Number of columns: 7
 Number of rows: 8
 Number of elements: 56
 MLM SNP effect estimates
 Dataset: Marker data 499 + Pop
 Compression = trueCompression
 P3D = true. Variance componer
 Model: trait = mean + Cluster1

Under the “Association” tab under the “Result” tab, click the “MLM_effects...” file. This includes the effect estimates for each genotypic class (homozygous or heterozygous) for each of the markers associated with freshweight.

Conclusion

This tutorial demonstrates that association mapping analysis can help identify the molecular markers significantly linked to traits of interest.

Implementation of GLM and MLM models in TASSEL software allows one to account for effects due to both population genetic structure and relatedness.

References and Readings

Bradbury, P. J., Z. Zhang, D. E. Kroon, T. M. Casstevens, Y. Ramdoss and E. S. Buckler. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 23:2633–2635. Available online at: <http://dx.doi.org/10.1093/bioinformatics/btm308> (verified 7 Feb 2012).

Myles, S., J. Peiffer, P. J. Brown, E. Ersoz, Z. Zhang, D. E. Costich, and E. S. Buckler. 2009. Association mapping: Critical considerations shift from genotyping to experimental design. *Plant Cell* 21:2194-2202. Available online at: <http://dx.doi.org/10.1105/tpc.109.068437> (verified 7 Feb 2012).

Zhu, C., M. Gore, E. S. Buckler, and J. Yu. 2008. Status and prospects of association mapping in plants. *Plant Genome* 1:5–20. Available online at: <http://dx.doi.org/10.3835/plantgenome2008.02.0089> (verified 7 Feb 2012).

Book

Oraguzie, N. C., E.H.A. Rikkerink, S. E. Gardine, and H. N. de Silva (eds.) *Association mapping in plants*. Springer, NY.