

Analyzing GBS data using **TASSEL 5.0**

(Trait Analysis by aSSociation, Evolution, and Linkage)
Hands-on tutorial



Dev Paudel
Agronomy Department
Room 340 Cancer & Genetics Research Complex
University of Florida
Gainesville, FL

dpaudel@outlook.com

[@devRpaudel](https://twitter.com/devRpaudel)

11/23/2015

Outline

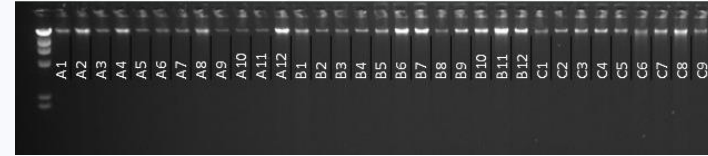
- Brief overview of genotyping by sequencing (GBS)
- File formats used in GBS
- TASSEL pipeline for calling SNPs
- Loading data in TASSEL GUI
- Creating linkage disequilibrium plots
- Creating Manhattan plots
- Creating QQ plots
- Creating cladograms

Genotyping by sequencing (GBS)

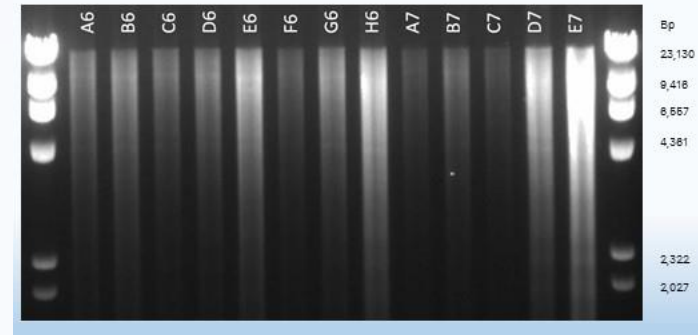
- Isolate DNA from samples/tissues
 - 33 genotypes
- Plate DNA & adapter pair
- Digest with Restriction Enzymes
 - Eg. PstI



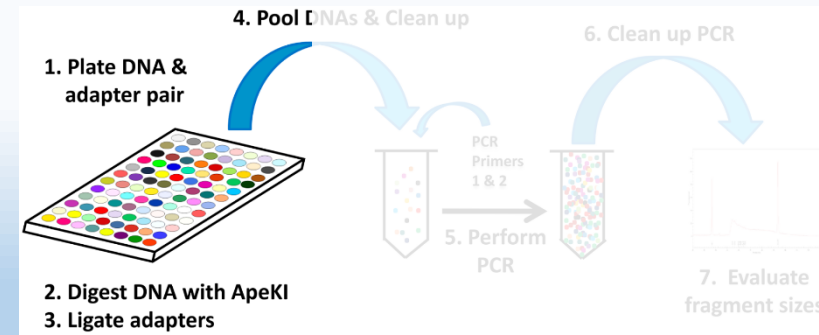
- Ligate adapters



DNA samples for GBS

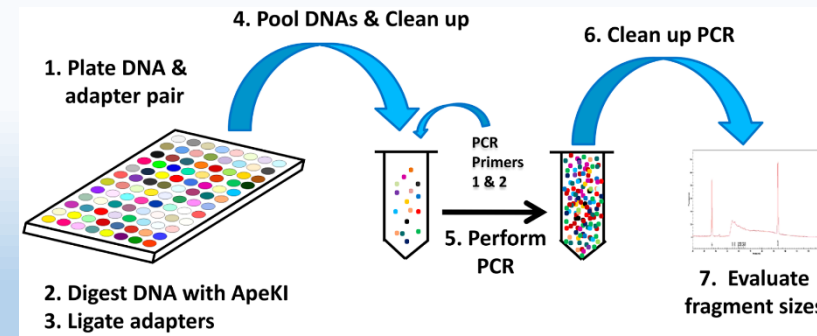
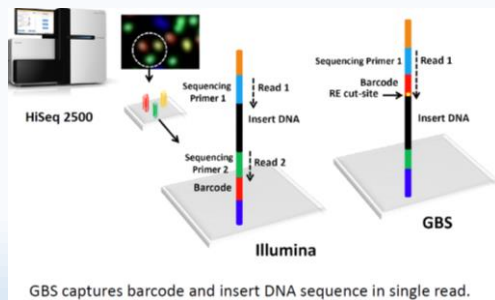


Restriction Digestion of DNA samples



Genotyping by sequencing (GBS)

- Isolate DNA from samples/tissues
 - 33 genotypes
- Digest with Restriction Enzymes
- Ligate with adapters
- Pool all DNA together
- PCR
- Sequence (Illumina)



Getting to know your GBS data

- 2 files
 - Key file
 - Sequence file

Key file / barcode file

Flowcell	Lane	Barcode	DNASample	LibraryPlate	Row	Col	LibraryPrepID	LibraryPlateID	Enz
C7U82ANXX		7 GAATGCAATA	34c1	NGflwgbs1	E	11	250515680	450023798	Pst
C7U82ANXX		7 ACGGTACT	34c10	NGflwgbs1	F	8	250515657	450023798	Pst
C7U82ANXX		7 CGCAACCACT	34c11	NGflwgbs1	F	9	250515665	450023798	Pst
C7U82ANXX		7 CAGTGCCATT	34c12	NGflwgbs1	F	10	250515673	450023798	Pst
C7U82ANXX		7 TAGCAG	34c13	NGflwgbs1	F	11	250515681	450023798	Pst
C7U82ANXX		7 GCTCCGA	34c14	NGflwgbs1	F	12	250515689	450023798	Pst
C7U82ANXX		7 CCGAACA	34c15	NGflwgbs1	G	1	250515602	450023798	Pst
C7U82ANXX		7 GGTGT	34c16	NGflwgbs1	G	2	250515610	450023798	Pst
C7U82ANXX		7 CGTGGACAGT	34c17	NGflwgbs1	G	4	250515626	450023798	Pst
C7U82ANXX		7 CTTGA	34c18	NGflwgbs1	G	5	250515634	450023798	Pst
C7U82ANXX		7 CTCTA	34c19	NGflwgbs1	G	6	250515642	450023798	Pst
C7U82ANXX		7 CAACCACACA	34c2	NGflwgbs1	E	12	250515688	450023798	Pst

Each barcode is associated with only 1 sample

Getting to know your GBS data

- 2 files
 - Key file
 - Sequence file
- Key file / barcode file
- Sequence file : FASTQ format (100 base pair long)

[illegible]

Each barcode is associated with only 1 sample

Getting to know your GBS data

- 2 files
 - Key file
 - Sequence file
- Key file / barcode file
- Sequence file : FASTQ format

Flowcell	Lane	Barcode	DNASample
C7U82ANXX	7	GAATGCAATA	34c1
C7U82ANXX	7	ACGGTACT	34c10
C7U82ANXX	7	CGCAACCAGT	34c11
C7U82ANXX	7	CAGTGCCATT	34c12
C7U82ANXX	7	TAGCAG	34c13
C7U82ANXX	7	CTGCAAG	34c14

@D00656:139:C7U82ANXX:7:1111:1302:2099 1:Y:0:7

ACGGTACTGCA GGGAAACTAATTCAGACACGGTAAACTCCACCGGTGCTCGCACCGGCAAATCTCGCTGGAAGCTCGAATGCGTTATGCCTGCCTGATCG
 +
 BBBB//<BF<FFFB/F<FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF<//BB<FBF<7FFFFFFFFF///<BF//7F//7B/////7B//77

Cut-site remnant

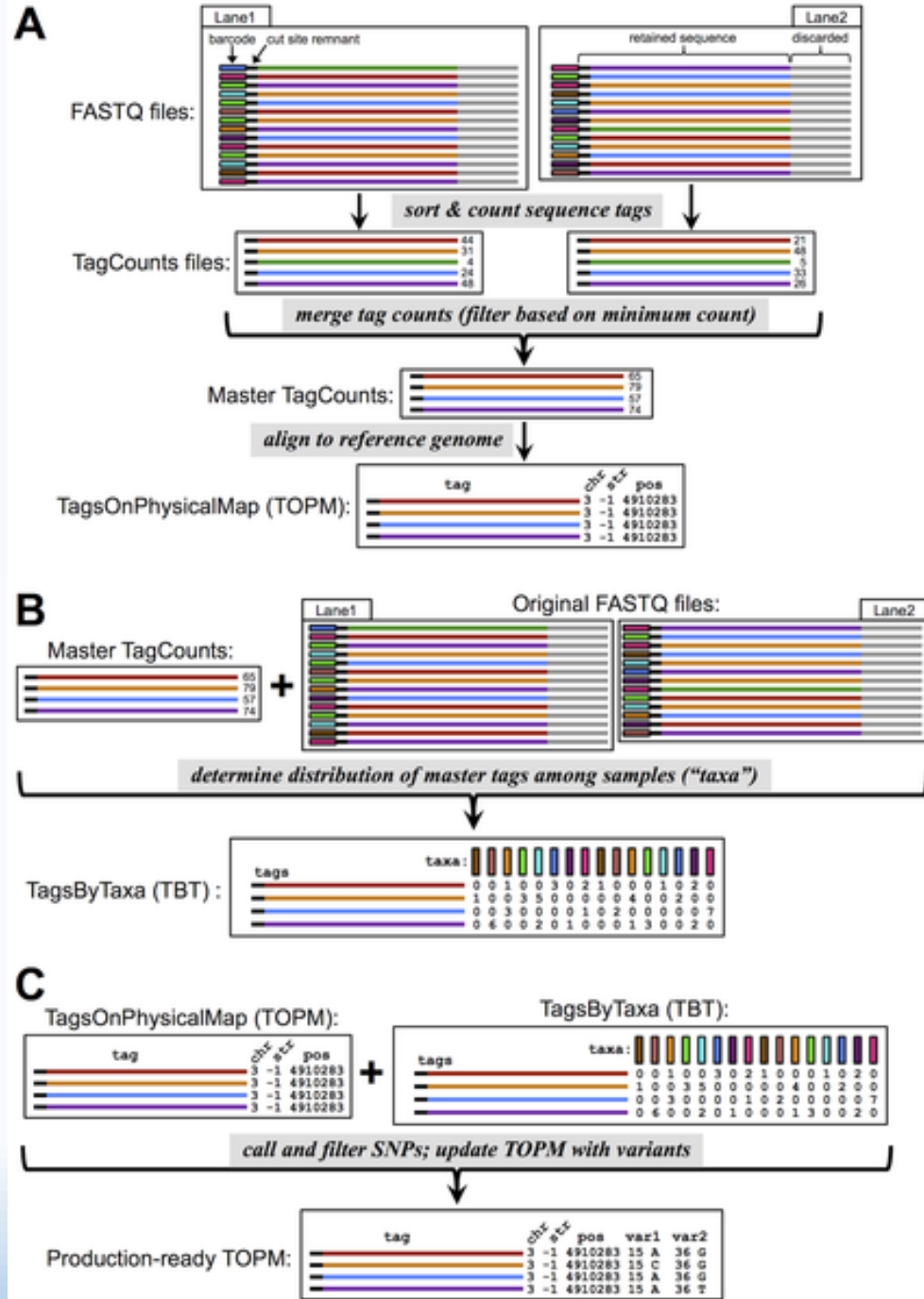
Barcode

5'...CTGCAAG...3'

3'...GACGTC...5'

Each barcode is associated with only 1 sample

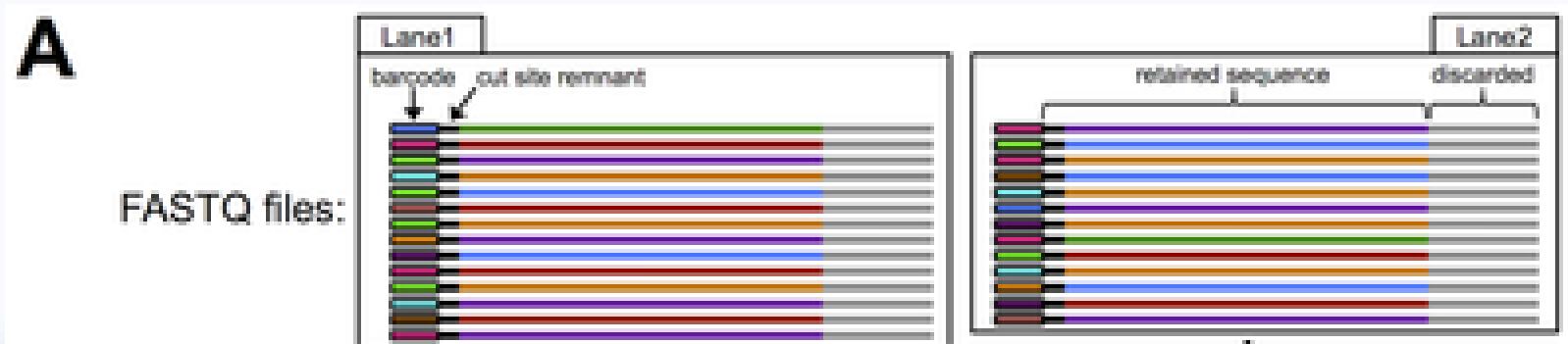
TASSEL pipeline



TASSEL pipeline

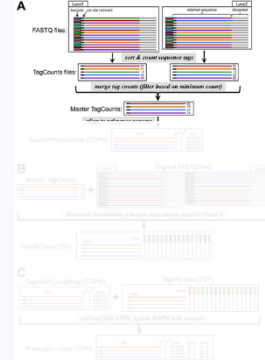
Filtering raw data:

- Remove reads that do not have barcodes
- Remove low quality reads
- Remove barcodes
- Remove nucleotides > 64bp
- Retained good barcoded sequences are called 'Tags'

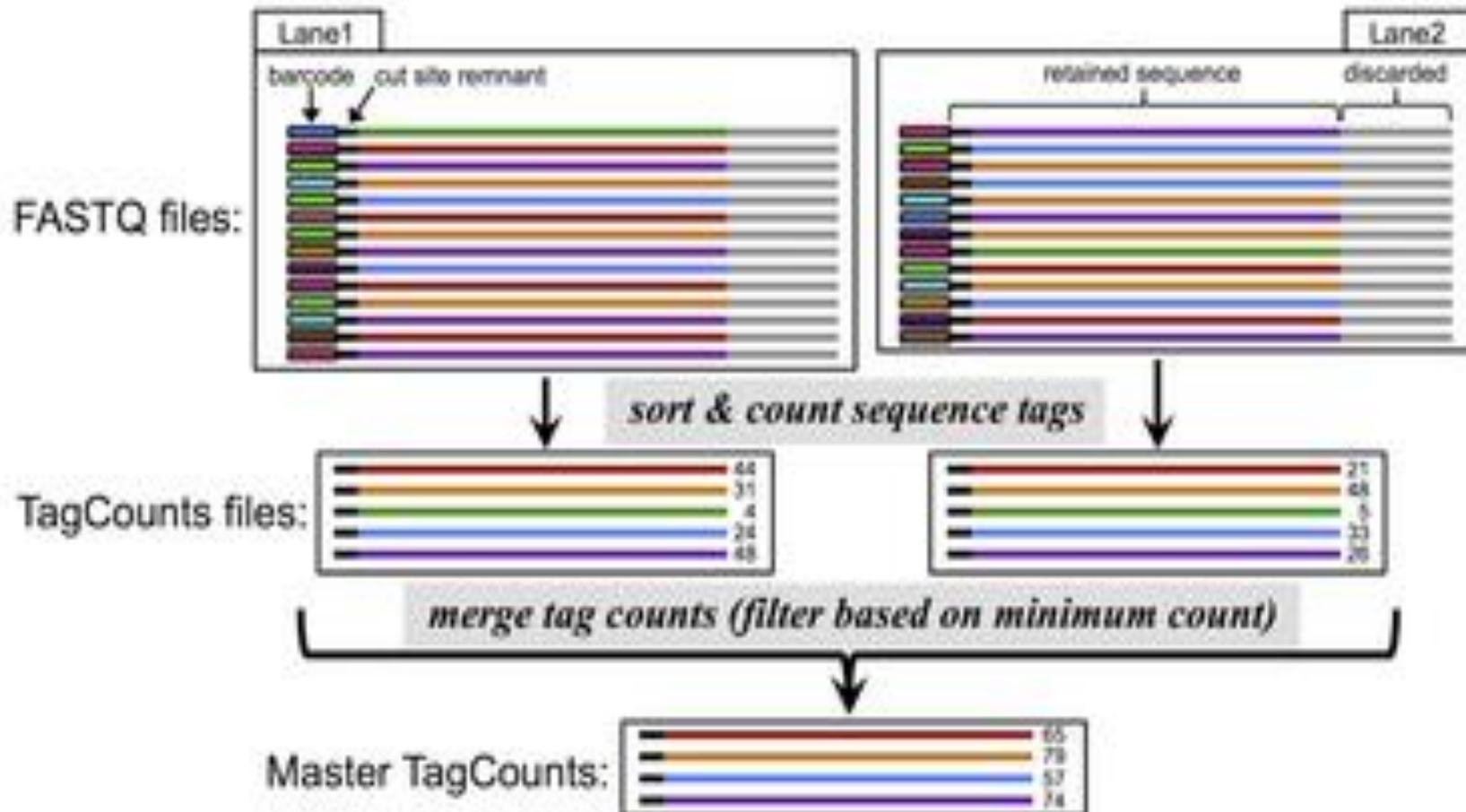


TASSEL pipeline

Extract unique 'tags'

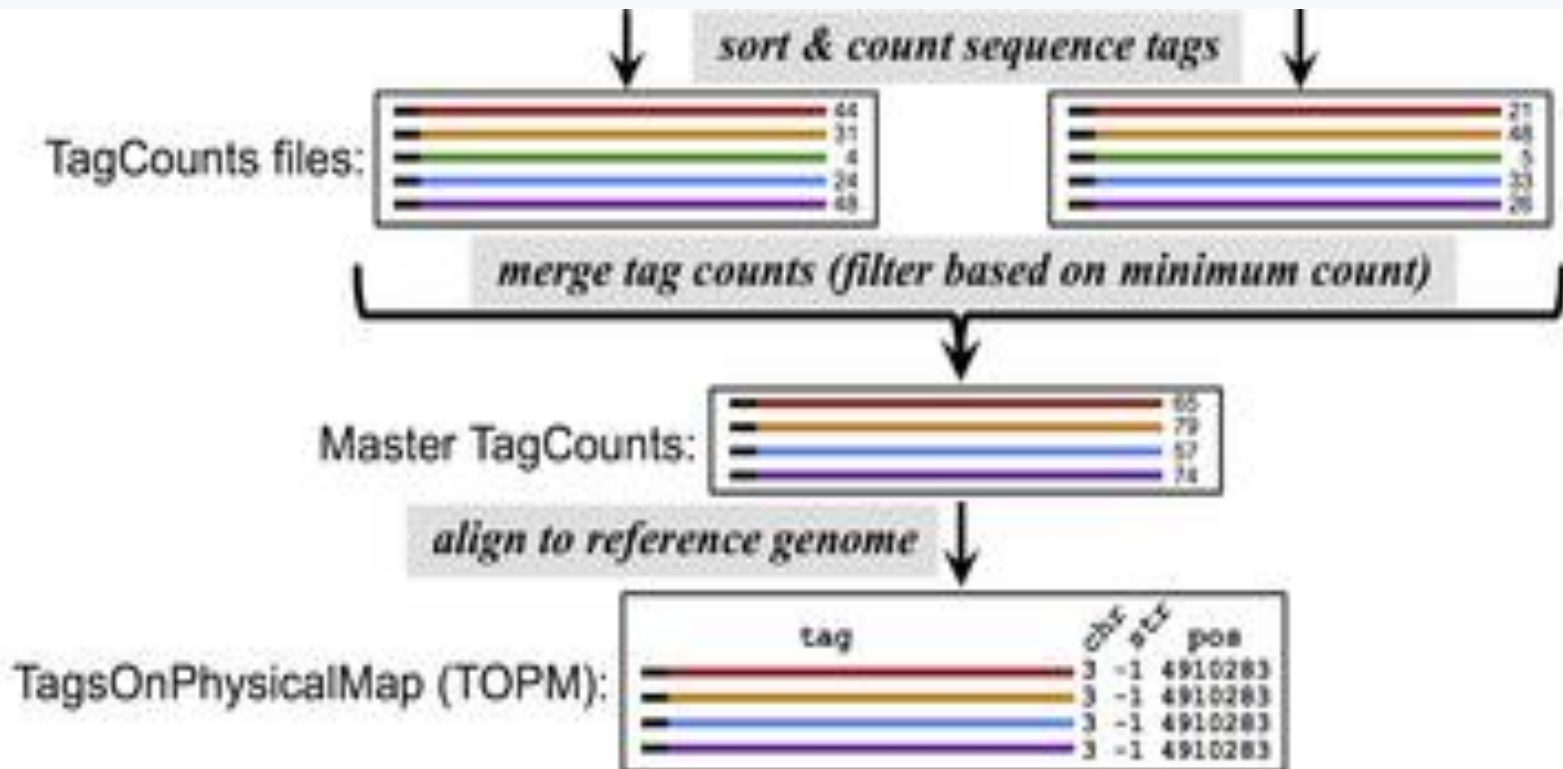
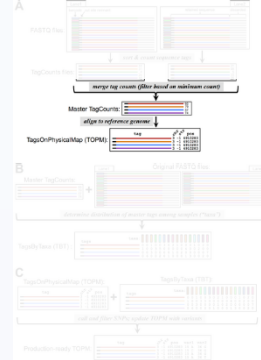


A



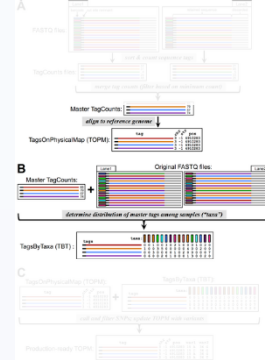
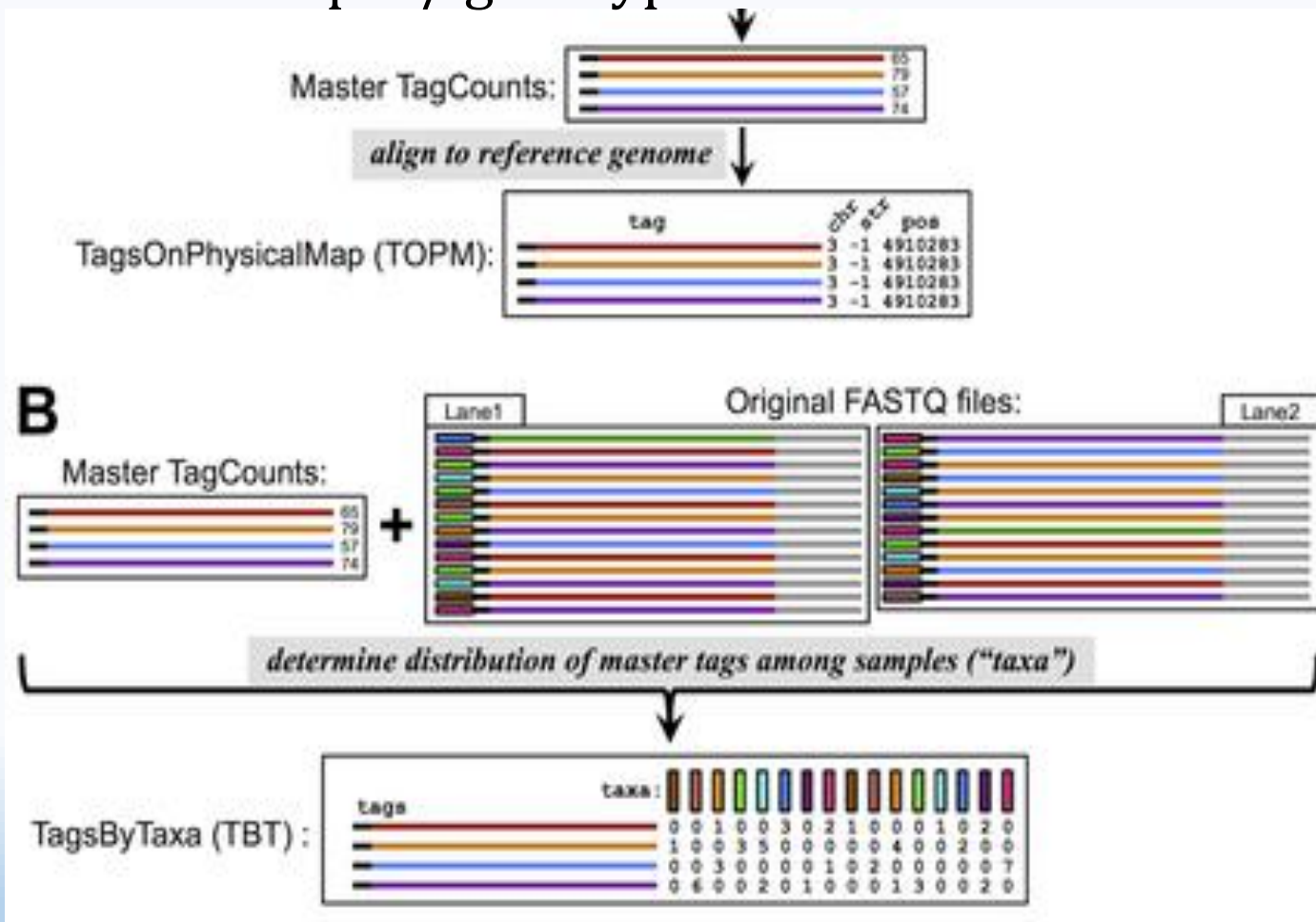
TASSEL pipeline

Align 'tags' to reference genome

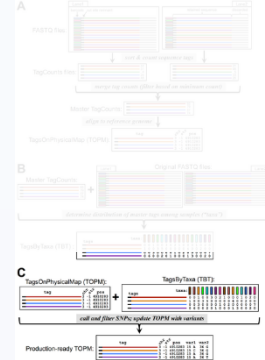
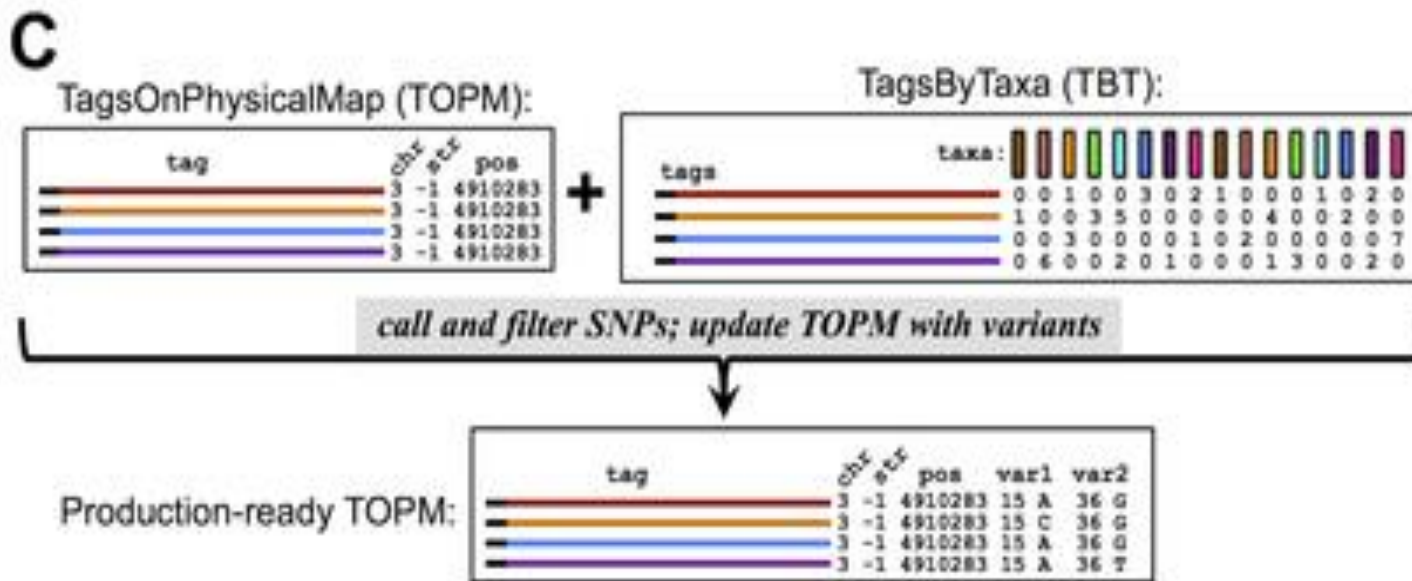


TASSEL pipeline

- Match reads to lines and create a matrix of sequence reads by their lines of origin (tags by taxa or TBT)
- Taxa = Each sample / genotype is called a taxa



SNP calling



- Physical positions and taxa are known for the GBS tags
- Tags that map to the same location (according to the TOPM file) are considered a locus
- The sequence differences between those tags and their relative abundance (found in the TBT file) are used to identify SNPs at the locus > Tabulated in *hapmap* format

SNP calling – hapmap format













Locus

Taxa / genotype

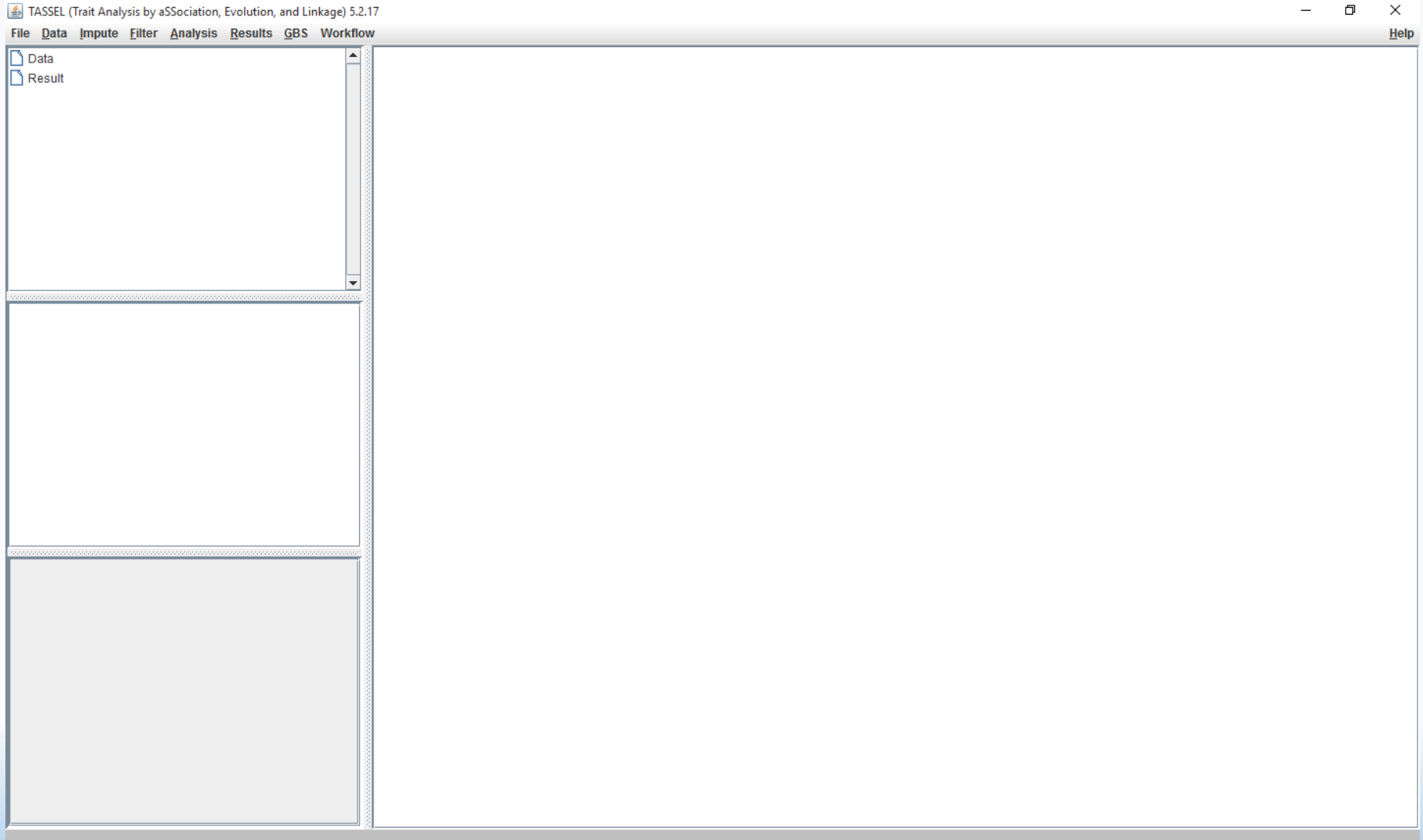
	0: 157104	1: 1947984	2: 2914066	3: 2914171	4: 2915078	5: 2915242	6: 2973508	7: 3205252	8: 3205262	9: 3206090	10: 3706018	11: 4175293	12: 4429897	13: 4429927	14: 4430055	15: 4490461	16: 4835434	17: 4835558	18: 4835658	19: 4836542	20: 4912526	21: 5353319	22: 5353655	23: 5562502	24: 8075572	25: 8366368	26: 8366411	27: 8367944	28: 8510027	29: 9023947	30: 9024005	31: 9029842	32: 9084948	33: 9084979	34: 9273299	35: 9300391	36: 9300541	37: 10065344	38: 10065856	39: 10066126	40: 10067544	41: 10067604	42: 10067746	43: 10067852	44: 10068726	45: 10069039	46: 10069145			
33-16	C	G	G	T	G	T	C	T	C	T	C	T	T	C	T	T	C	T	C	A	C	G	G	G	G	G	G	T	C	A	A	T	G	T	C	C	G	G	C	G	T	G	T	A	A	T	G	G	A	
38-11	C	C	G	T	G	T	C	T	C	C	T	C	C	T	A	A	C	T	C	A	C	G	G	G	G	G	A	G	A	A	A	N	G	C	C	C	G	A	C	A	C	G	T	A	A	T	G	G	A	
4226	C	G	G	T	G	T	C	T	C	T	C	C	N	N	C	A	A	C	T	C	A	C	G	G	G	G	G	T	C	N	N	N	G	T	C	T	G	G	C	G	T	G	T	A	A	T	N	G	G	C
A188	A	C	G	T	G	T	C	T	C	T	C	T	C	G	T	A	A	C	T	C	A	A	G	G	G	G	T	N	A	C	C	G	T	C	T	G	A	C	G	T	G	T	A	A	T	G	G	G	N	
A214N	C	C	T	A	G	A	C	T	C	T	T	T	C	C	T	T	A	C	T	C	A	C	G	G	C	G	T	C	T	A	T	G	T	C	T	G	A	C	G	T	A	T	A	T	T	G	C	A		
A239	A	C	T	T	A	A	C	T	C	T	C	T	C	C	T	A	N	C	T	C	A	C	G	G	G	G	T	N	A	A	T	G	C	C	C	G	A	C	G	T	G	T	A	A	T	G	G	G	A	
A272	A	C	T	T	A	A	G	T	C	T	C	T	C	C	T	A	A	C	T	C	A	A	G	G	G	G	T	C	A	A	T	G	T	G	T	A	G	C	G	T	G	T	A	A	T	G	G	G	A	
A441-5	C	C	G	T	G	T	C	T	C	T	T	T	C	G	T	A	A	C	T	C	A	C	G	G	G	G	T	C	T	C	C	G	T	C	T	G	G	C	A	C	G	T	A	A	T	G	G	A		
A554	C	G	T	T	A	T	C	T	C	T	T	C	C	C	T	A	A	C	T	C	A	C	G	G	G	G	T	C	T	A	T	G	T	C	T	G	A	C	A	C	G	T	A	A	T	G	G	A		
A556	C	C	G	T	G	T	N	N	N	T	C	T	C	C	T	T	A	C	T	C	A	C	G	C	G	G	G	C	T	A	T	T	C	C	T	G	A	C	A	C	G	T	A	A	T	G	G	G	A	
A6	A	C	T	T	A	A	C	T	C	T	T	T	C	C	T	A	A	C	T	C	A	C	G	G	C	G	T	C	T	A	T	G	C	C	C	G	G	C	G	T	G	T	A	A	T	G	G	G	A	
A619	C	G	G	T	G	T	C	A	G	T	C	C	C	C	T	A	A	C	T	C	A	A	G	G	G	G	T	C	A	C	C	G	T	C	C	G	G	C	G	T	A	C	A	A	A	A	C	G	A	
A632	C	C	T	A	G	A	C	T	C	T	C	C	C	C	T	A	A	C	T	C	A	C	G	G	C	G	T	C	T	A	T	G	T	C	T	G	A	C	G	T	A	T	A	T	T	G	C	A	C	
A634	C	C	T	A	G	A	C	T	C	T	T	T	C	C	T	T	A	C	T	C	A	C	G	G	C	G	T	C	T	A	N	G	T	C	T	G	A	C	G	T	A	T	A	T	T	G	C	A	C	
A635	C	C	T	A	G	A	C	T	C	T	T	T	C	C	T	T	A	C	T	C	A	C	G	G	C	G	T	C	T	A	N	G	T	C	T	G	A	C	G	T	A	T	A	T	T	G	C	A	C	
A641	A	C	T	T	A	T	C	T	C	T	T	T	C	C	T	T	A	C	T	C	A	C	G	G	C	N	T	C	T	A	T	G	T	C	T	G	A	C	G	T	A	T	A	T	T	G	C	A	C	
A654	N	G	T	T	A	T	C	T	C	T	T	T	C	C	T	A	A	C	T	C	A	C	G	G	N	A	G	A	A	A	N	G	C	C	T	A	G	C	G	T	G	T	A	A	T	G	G	A	C	
A659	A	G	T	T	A	T	C	T	C	T	T	T	C	C	T	A	A	C	T	C	A	C	G	G	C	G	G	C	A	N	C	G	T	C	C	G	A	C	G	T	G	T	A	A	T	G	G	A	N	
A661	C	G	T	T	A	N	C	T	C	T	C	C	C	C	T	A	A	C	T	C	A	C	G	G	G	G	T	C	T	C	C	G	T	C	C	G	G	C	N	T	G	T	A	A	T	G	G	A	N	
A679	C	C	T	A	A	T	G	T	C	N	T	T	C	C	T	T	A	C	T	C	A	C	G	C	C	N	G	A	T	C	C	G	C	C	C	G	A	C	G	T	A	C	A	A	A	A	C	A	C	
A680	C	C	T	A	A	T	G	T	C	T	T	T	C	C	T	T	A	C	T	C	A	C	G	C	G	G	T	N	T	N	C	G	C	C	C	G	A	C	G	T	A	C	A	A	A	A	C	A	C	
A682	C	C	T	A	G	T	G	T	C	T	C	C	C	C	T	A	A	C	T	C	A	C	G	G	G	G	T	C	T	A	T	G	C	C	C	G	A	C	G	T	G	T	A	A	T	G	G	A	C	
AB28A	A	G	T	A	G	N	C	N	N	T	C	T	T	G	C	A	A	C	T	C	A	C	C	G	C	G	T	C	T	C	C	G	T	C	C	G	A	C	A	C	G	T	A	A	T	G	G	G	A	
B10	A	C	G	T	G	T	C	T	C	T	T	T	C	C	T	T	A	C	T	C	A	C	G	C	C	A	G	A	A	A	T	G	C	C	C	G	A	C	G	T	A	C	A	A	A	A	C	A	C	
B103	A	C	G	T	G	T	C	T	C	T	T	T	C	C	N	A	A	C	T	C	A	A	G	G	G	G	G	C	T	A	T	G	T	G	C	G	A	C	G	T	G	T	A	A	T	G	G	A	C	
B104	C	C	T	T	A	T	C	T	C	C	T	T	C	C	T	A	A	C	T	C	A	C	G	C	C	A	G	A	T	A	T	G	T	C	T	G	G	C	G	T	A	C	A	A	A	A	C	A	C	
B105	C	C	T	T	A	T	C	T	C	C	T	T	C	C	T	A	A	C	T	C	A	C	G	C	C	G	T	C	T	A	T	G	T	C	T	G	A	C	G	T	A	T	A	T	T	G	C	A	C	
B109	C	C	G	T	G	T	C	T	C	T	T	T	C	C	T	T	N	C	T	C	N	C	G	C	C	A	G	A	T	C	C	G	N	C	C	G	A	C	G	T	A	C	A	A	A	A	C	A	C	
B115	C	G	G	T	G	T	C	A	G	T	C	T	C	C	T	A	N	C	T	C	A	C	G	G	C	G	G	A	T	A	T	G	C	C	C	G	A	C	G	T	G	T	A	A	T	G	G	G	A	
B14A	C	C	T	A	G	A	C	T	C	T	T	T	C	C	T	T	N	C	T	C	A	C	G	G	C	G	T	C	T	A	T	G	C	C	C	G	A	C	G	T	A	T	A	A	A	G	G	A	C	
B164	C	G	T	T	A	T	C	T	C	T	C	T	C	C	T	T	A	C	T	C	A	A	G	G	G	G	T	C	A	C	C	T	T	G	C	G	A	C	G	T	G	T	A	A	T	G	G	G	C	
B2	A	C	G	T	G	T	C	T	C	T	T	C	C	N	T	A	A	N	T	C	A	A	G	G	C	G	T	C	T	A	T	G	T	C	T	A	G	C	G	T	G	T	A	A	T	G	G	G	C	

Running TASSEL GUI

sseladmin-tassel-5-standalone-ebc440bcffb4 > tasseldmin-tassel-5-standalone-ebc440bcffb4 >

Name	Date modified	Type	Size
 example_pipelines	11/20/2015 1:01 PM	File folder	
 lib	11/20/2015 1:02 PM	File folder	
 TASSELTutorialData	11/20/2015 1:01 PM	File folder	
 cp.bat	11/20/2015 1:00 PM	Windows Batch File	1 KB
 run_anything.bat	11/20/2015 1:00 PM	Windows Batch File	1 KB
 run_anything.pl	11/20/2015 1:00 PM	PL File	2 KB
 run_pipeline.bat	11/20/2015 1:00 PM	Windows Batch File	1 KB
 run_pipeline.pl	11/20/2015 1:00 PM	PL File	2 KB
 start_tassel.bat	11/20/2015 1:01 PM	Windows Batch File	1 KB
 start_tassel.pl	11/20/2015 1:01 PM	PL File	2 KB
 sTASSEL.jar	11/20/2015 1:01 PM	Executable Jar File	2,921 KB
 Tassel5PipelineCLI.pdf	11/20/2015 1:00 PM	Adobe Acrobat D...	967 KB

Loading data into TASSEL GUI



Loading data into TASSEL GUI

mdp_genotype_subset.hmp.txt

mdp_genotype_hmp_subset.txt.hmp.txt - Notepad

File	Edit	Format	View	Help													
rs#	alleles	chrom	pos	strand	assembly#	center	protLSID	assayLSID	panelLSID	QCcode	33-16	38-11	4226				
4722	A188	A214N	A239	A272	A441-5	A554	A556	A6	A619	A632	A634	A635	A641	A654	A659	A661	A679
A680	A682	AB28A	B10	B103	B104	B105	B109	B115	B14A	B164	B2PZB00859.1	C/A	1	157104	+	NA	
NA	NA	NA	NA	NA	C	C	C	C	A	C	A	A	C	C	C	A	C
C	C	A	N	A	C	C	C	C	A	A	A	C	C	C	C	C	C
APZA01271.1	C/G	1	1947984	+	NA	NA	NA	NA	NA	NA	NA	C	G	C	G	C	C
C	C	G	C	C	G	C	C	C	C	G	G	G	C	C	C	G	C
C	C	C	G	C	G	CPZA03613.2	T/G	1	2914066	+	NA	NA	NA	NA	NA	NA	NA
G	G	G	G	T	T	T	G	T	G	T	G	T	T	T	T	T	T
T	T	T	T	G	G	T	T	G	G	T	T	GPZA03613.1	T/A	1	2914171	+	
NA	NA	NA	NA	NA	NA	T	T	T	T	T	A	T	T	T	T	T	T
A	A	A	T	T	T	T	A	A	A	A	T	T	T	T	T	T	A

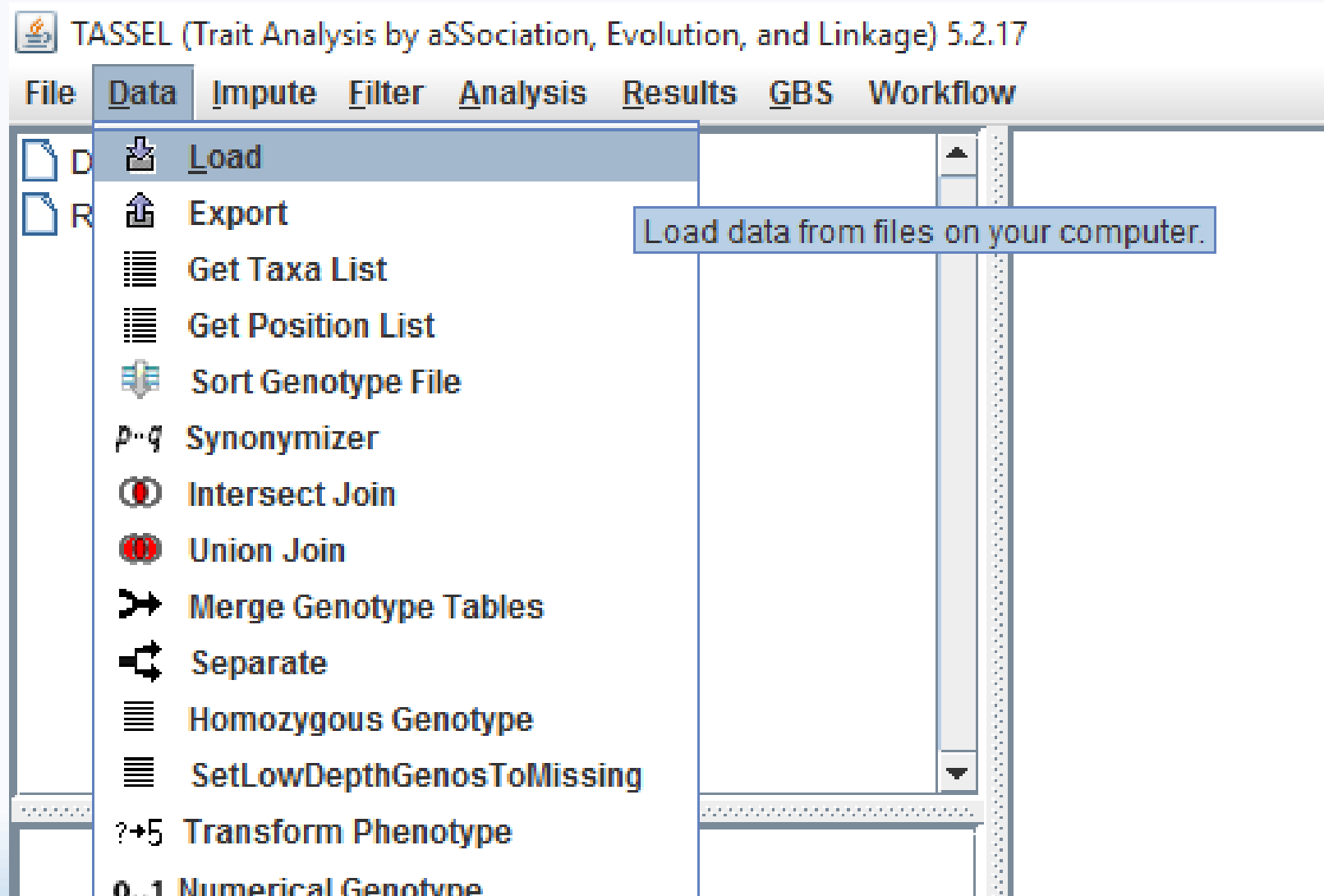
mdp_traits_subset.txt

mdp_traits_subset.txt - Notepad

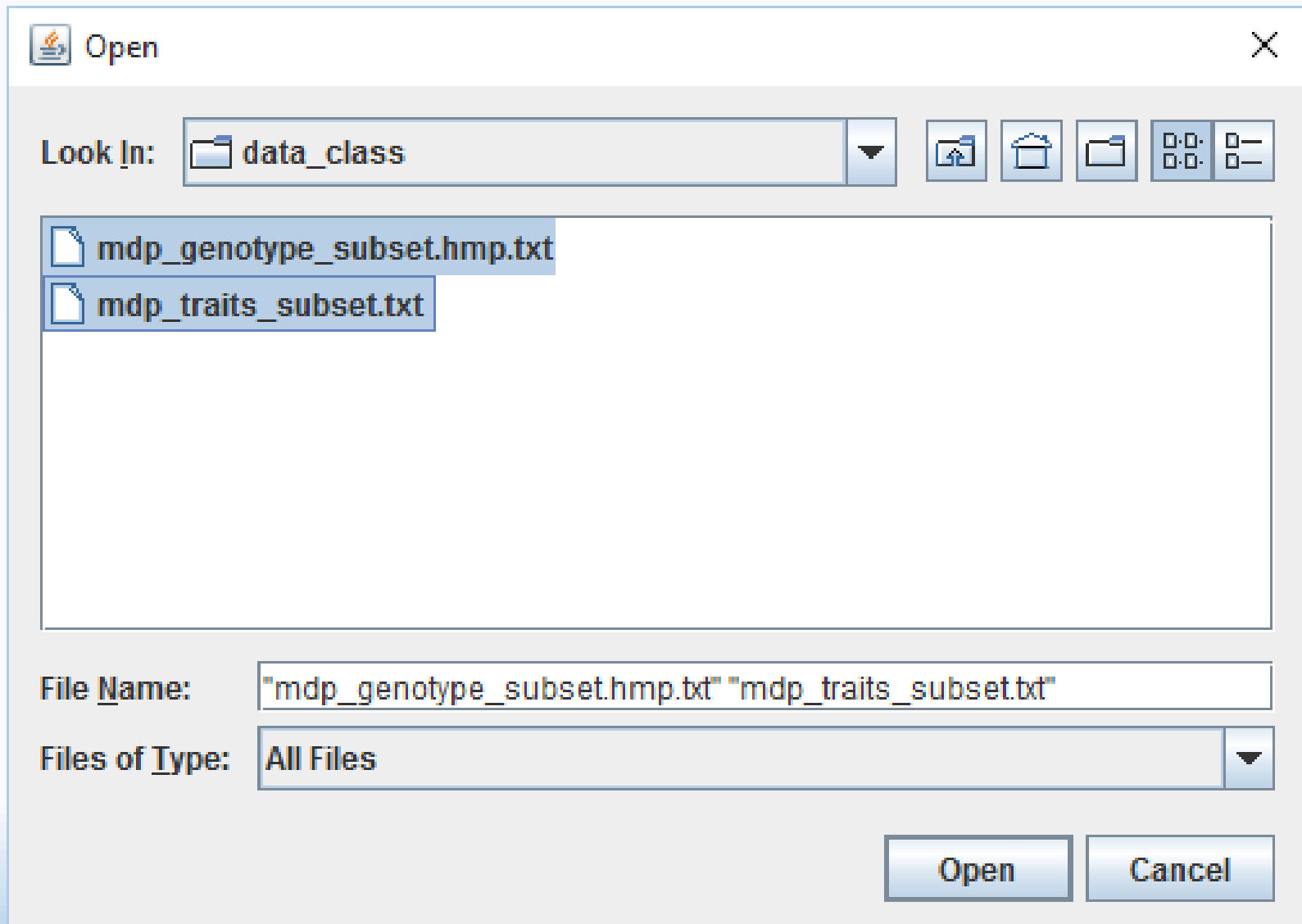
kTrait>	EarHT	dpoll	EarDia
811	59.5	-999	-999
33-16	64.75	64.5	-999
38-11	92.25	68.5	37.897
4226	65.5	59.5	32.21933
4722	81.13	71.5	32.421
A188	27.5	62	31.419
A214N	65	69	32.006
A239	47.88	61	36.064
A272	35.63	70	-999

Complete data set available with TASSEL 5.0
TASSELTutorialData>data

Loading data into TASSEL GUI



Loading data into TASSEL GUI



Loading data into TASSEL GUI

TASSEL (Trait Analysis by aSSociation, Evolution, and Linkage) 5.2.17

File Data Impute Filter Analysis Results GBS Workflow

Data

- Numerical
 - mdp_traits_subset
- Sequence
 - mdp_genotype_subset
- Result

Table Title: mdp_traits_subset
Number of columns: 4
Number of rows: 34
Number of elements: 136

Taxa	EarHT	dpoll	EarDia
811	59.5	<input type="checkbox"/>	<input type="checkbox"/>
33-16	64.75	64.5	<input type="checkbox"/>
38-11	92.25	68.5	37.897
4226	65.5	59.5	32.219
4722	81.13	71.5	32.421
A188	27.5	62	31.419
A214N	65	69	32.006
A239	47.88	61	36.064
A272	35.63	70	<input type="checkbox"/>
A441-5	53.5	67.5	35.008
A554	38.5	66	33.418
A556	28	65	31.929
A6	109.5	80.5	31.517
A619	36	61	40.63
A632	60	61	35.953
A634	54	59	35.601
A635	37	64	35.3
A641	54.5	66	33.727
A654	39	64	<input type="checkbox"/>
A659	46.5	58.5	38.846
A661	51.5	59	39.323
A679	65	66	42.471
A680	68	65.5	41.152
A682	47	57.5	35.928
AB28A	73.5	78	32.504
B10	74	69	36.561
B103	37	57.5	<input type="checkbox"/>
B104	56.25	64.5	44.773
B105	67.5	68	39.857
B109	67	64	38.951
B115	68	65.5	37.06
B14A	57	63.5	38.067
B164	66	58	35.562
B2	39.5	70	<input type="checkbox"/>

Loading data into TASSEL GUI

TASSEL (Trait Analysis by aSSociation, Evolution, and Linkage) 5.2.17

File Data Impute Filter Analysis Results GBS Workflow Help

☒ Site Numbers ☐ Locus ☐ Site Name ☐ Alleles MajorMinorAllele (Enter physical position) Search

0 309 618 927 1236 1545 1854 2163 2472 2781 3090

1541: 207114208 1542: 210235075 1543: 210235180 1544: 210235726 1545: 210235039 1546: 210236675 1547: 212497348 1548: 212893509 1549: 215393158 1550: 215393255

33-16 T C C C C T A T T N C
38-11 N C C C G T C G C G C
4226 N C C C C T A T T T C
4722 T C C C C T A T N G C
A188 T C C T C G C T T G C
A214N C C C C C T A T T G C
A239 C C C C C T A T T G C
A272 C C G C C T A T T G C
A441-5 C C C C C T A T C G C
A554 T G T C T A T C G C
A556 T G T C G C T G C
A6 T C C C T C T C G C
A619 T C C C C T A T C G C
A632 C C C C T A T T G C
A634 C C C C T A T T G C
A635 C C C C T A T T G C
A641 T C C G T C T C G C
A654 T G T C G C T N C
A659 C C C C T A T T G C
A661 N N C C T A T C G C
A679 C C C N T A T T G C
A680 C C C C T A T T G C
A682 T C T C G C T N T C
AB28A N C C G T C G C G C
B10 T C C C T A T T G C
B103 T C C C T A T T G C
B104 C C C C T A T T G C
B105 T G T C G C T T G C
B109 C C C C T A T T G C
B115 T C C C T A T T T C
B14A C C C C T A T T G C
B164 T G T C G C T T N C
B2 T C C C C T A T T G C

Number of taxa: 33
Number of sites: 3093

Chromosomes...

1: 540 sites:
0 (157104) - 539 (299170077)

2: 393 sites:
540 (736367) - 932 (234574991)

3: 355 sites:
933 (1240310) - 1287 (229544509)

4: 319 sites:
1288 (139753) - 1606 (245131801)

5: 357 sites:
1607 (656148) - 1963 (216431558)

6: 213 sites:
1964 (2379148) - 2176 (167883450)

7: 246 sites:
2177 (729478) - 2422 (170346253)

8: 256 sites:
2423 (169137) - 2678 (172323795)

9: 213 sites:
2679 (3873116) - 2891 (151289948)

class net.maizegenetics.dna.snp.CoreGenotypeTable

Filtering for sites on chr 1 only

TASSEL (Trait Analysis by aSSociation, Evolution, and Linkage) 5.2.17

File Data Impute **Filter** Analysis Results GBS Workflow

Data

- Numerical
 - mdp_traits
- Sequence
 - mdp_genc
- Result

Filter

- Sites**
- Site Names
- Taxa Names
- Taxa
- Traits
- Filter Genotype Table Sites

Filter and Process Alignment

☒ Site Numbers ☐ Locus ☐


	0	309	618
		1541: 207114208	
33-16		T	
38-11		N	
4226		N	
4722		T	
A188		T	
A214N		C	
A239		C	
A272		C	
A441-5		C	
A554		T	
A556		T	
A6		T	

Number of taxa: 33
Number of sites: 3093

Chromosomes...

1: 540 sites:

Filtering for sites on chr 1 only

 Filter Alignment ×

Filter Alignment

Minimum Count:	<input type="text" value="20"/>	out of 33 sequences
Minimum Frequency:	<input type="text" value="0.0"/>	
Maximum Frequency:	<input type="text" value="1.0"/>	
Position Type:	<input type="text" value="Position index"/>	
Start Position:	<input type="text" value="0"/>	
End Position:	<input type="text" value="3092"/>	of 3092 sites

☐ Remove minor SNP states

☐ Generate haplotypes via sliding window

Haplotype Length

Step Length

Filtering for sites on chr 1 only

Filter Alignment

Select Chromosomes to Filter

<input checked="" type="checkbox"/> 1	<input type="checkbox"/> 2
<input type="checkbox"/> 3	<input type="checkbox"/> 4
<input type="checkbox"/> 5	<input type="checkbox"/> 6
<input type="checkbox"/> 7	<input type="checkbox"/> 8
<input type="checkbox"/> 9	<input type="checkbox"/> 10
<input type="checkbox"/> Select/Deselect All	

Select Cancel

Filtering for sites on chr 1 only

TASSEL (Trait Analysis by aSSociation, Evolution, and Linkage) 5.2.17

FileDataImputeFilterAnalysisResultsGBSWorkflow

Data

Numerical

Sequence

Result

mdp_traits_subset

mdp_genotype_subset

mdp_genotype_subset_chr1_157104-299170077

Number of taxa: 33

Number of sites: 540

Chromosomes...

1: 540 sites:

0 (157104) - 539 (299170077)

FilterAlignment...

Site Filter

Base Type: CoreGenotypeTable

Nucleotide Codes

(Derived from IUPAC)...

A A:A

C C:C

G G:G

T T:T

R A:C

Physical Positions

Site Numbers

Locus

Site Name

Alleles

157104

37533725

74910346

112286967

	244: 177870650	245: 179852076	246: 180171012	247: 180171048	248: 182386577	249: 182391102	250: 183647544	251: 183831590	252: 184088942	253: 188047564	254: 188047739	255: 188047772	256: 188047850	257: 188048080	258: 189467910	259: 193887303	260: 193887369	261: 195425179	262: 195589422
33-16	C	G	A	A	G	G	A	T	G	C	A	C	C	G	T	T	C	G	A
38-11	C	C	A	A	G	G	R	T	G	A	A	C	C	G	T	T	C	G	A
4226	C	N	N	A	A	G	G	C	G	C	A	C	C	G	T	T	A	G	G
4722	C	N	G	A	G	G	A	T	G	M	G	C	Y	G	T	T	C	G	A
A188	C	G	A	A	N	G	G	T	G	C	G	C	C	G	T	T	C	G	G
A214N	C	C	A	A	G	S	G	T	G	C	A	C	C	G	T	T	C	G	G
A239	C	C	G	A	G	G	A	C	G	C	A	C	C	G	T	T	A	G	A
A272	C	C	G	A	G	G	R	N	G	C	A	C	C	G	T	T	A	G	A
A441-5	C	C	G	A	A	G	G	T	G	C	G	C	T	G	T	T	A	A	A
A554	C	C	G	A	G	G	G	T	G	C	G	C	T	G	T	T	C	G	G
A556	C	C	G	A	N	G	G	T	G	C	A	C	C	G	C	T	A	G	G
A6	C	G	G	A	G	G	R	T	G	A	A	C	C	G	T	T	C	G	A
A619	C	C	G	A	G	G	G	T	G	C	G	C	C	G	T	T	A	G	A
A632	C	C	A	A	G	G	G	T	G	C	N	C	C	G	T	T	C	G	G
A634	C	C	A	A	G	N	G	T	G	C	A	C	C	G	T	T	C	G	G
A635	C	C	A	A	G	G	G	T	G	C	N	C	C	G	T	T	C	G	G
A641	C	C	N	A	G	G	T	G	C	A	C	C	G	T	T	C	G	G	G
A654	C	G	A	A	G	G	G	C	G	C	G	C	N	G	C	T	A	R	A
A659	C	C	G	A	G	G	G	T	G	C	G	C	T	G	T	T	C	G	G
A661	N	C	G	T	G	G	G	T	G	C	A	C	C	G	C	T	A	G	A
A679	C	C	G	A	A	G	R	C	A	C	A	C	C	G	T	T	C	A	A
A680	C	C	G	A	A	G	R	C	R	C	A	C	C	G	T	T	C	A	A
A682	C	C	A	A	G	G	G	T	G	C	A	C	C	G	T	T	C	G	A
AB28A	C	C	G	A	G	G	G	T	G	A	A	C	C	G	T	T	C	G	A
R10	C	C	A	A	G	S	G	T	G	C	A	C	C	G	T	T	C	G	A

Calculating linkage disequilibrium

TASSEL (Trait Analysis by aSSociation, Evolution, and Linkage) 5.2.17

FileDataImputeFilterAnalysisResultsGBSWorkflow

Data

Numerical

Sequence

Result

mdp_traits_subse

mdp_genotype_su

mdp_genotype_su

Diversity

Linkage Disequilibrium

Distance Matrix

Cladogram

Kinship

Pedigree Relationship Matrix

Combined A and G Relationship Matrix

PCA

MDS

GLM

MLM

WeightedMLM

Physical Positions

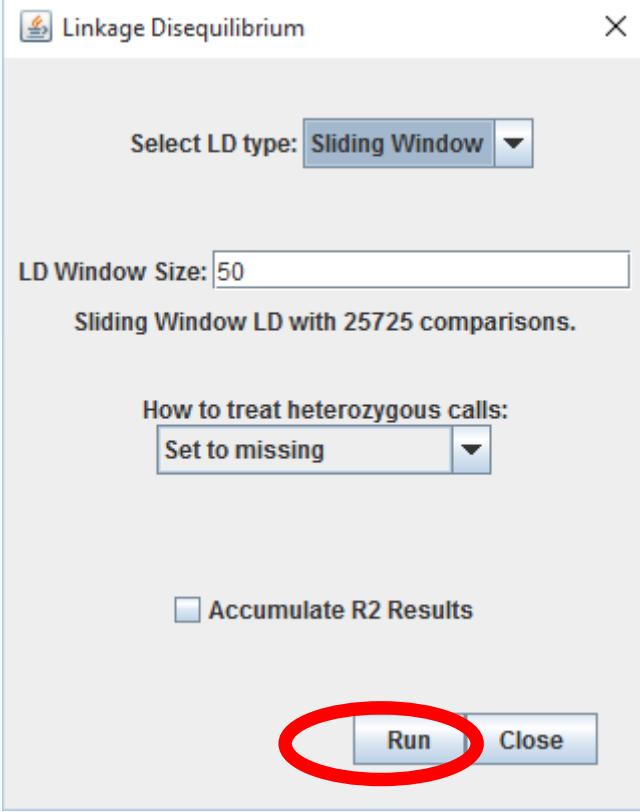
Site Number

Linkage Disequilibrium

37533725

	244: 177870650	245: 179852076	246: 180171012	247: 180171048	248: 182386577
33-16	C	G	A	A	G
38-11	C	C	A	A	G
4226	C	N	N	A	A
4722	C	N	G	A	G
A188	C	G	A	A	N
A214N	C	C	A	A	G

Calculating linkage disequilibrium



A screenshot of a software dialog box titled "Linkage Disequilibrium". The dialog box has a standard Windows-style title bar with a close button (X) in the top right corner. The main area is light gray and contains several controls:

- Select LD type:** A dropdown menu with "Sliding Window" selected.
- LD Window Size:** A text input field containing the number "50".
- Sliding Window LD with 25725 comparisons.** A line of text providing context for the window size.
- How to treat heterozygous calls:** A dropdown menu with "Set to missing" selected.
- Accumulate R2 Results:** An unchecked checkbox.
- Buttons:** Two buttons at the bottom right, "Run" and "Close". The "Run" button is circled in red.

Calculating linkage disequilibrium

TASSEL (Trait Analysis by aSSociation, Evolution, and Linkage) 5.2.17

File Data Impute Filter Analysis Results GBS Workflow

Data

- Numerical
 - mdp_traits_subset
- Sequence
 - mdp_genotype_subset
 - mdp_genotype_subset_chr1_157104-299170077
- Result
 - LD
 - LD:mdp...**

Locus1	Position1	Site1
1	1947984	1
1	2914066	2
1	2914066	2
1	2914171	3
1	2914171	3
1	2914171	3
1	2915078	4
1	2915078	4
1	2915078	4
1	2915078	4
1	2915242	5
1	2915242	5
1	2915242	5
1	2915242	5
1	2915242	5
1	2973508	6
1	2973508	6
1	2973508	6
1	2973508	6
1	2973508	6
1	2973508	6
1	3205252	7

Table Title: Linkage Disequilibrium
Number of columns: 17
Number of rows: 25725
Number of elements: 437325
LD Analysis

Creating LD plots

TASSEL (Trait Analysis by aSSociation, Evolution, and Linkage) 5.2.17

FileDataImputeFilterAnalysisResultsGBSWorkflow

Data

Numerical

Sequence

Result

LD

mdp_traits_subset

mdp_genotype_subset

mdp_genotype_subset_chr1

LD:mdp...

Table

Archaeopteryx Tree

LD Plot

Chart

QQ Plot

Manhattan Plot

Display Linkage Disequilibrium

Locus1	Position1	
1	1947984	
1	2914066	
1	2914066	
1	2914171	
1	2914171	
1	2914171	
1	2915078	
1	2915078	
1	2915078	
1	2915078	
1	2915242	
1	2915242	
1	2915242	
1	2915242	
1	2915242	
1	2973508	
1	2973508	
1	2973508	
1	2973508	
1	2973508	
1	2973508	

Table Title: Linkage Disequilibrium

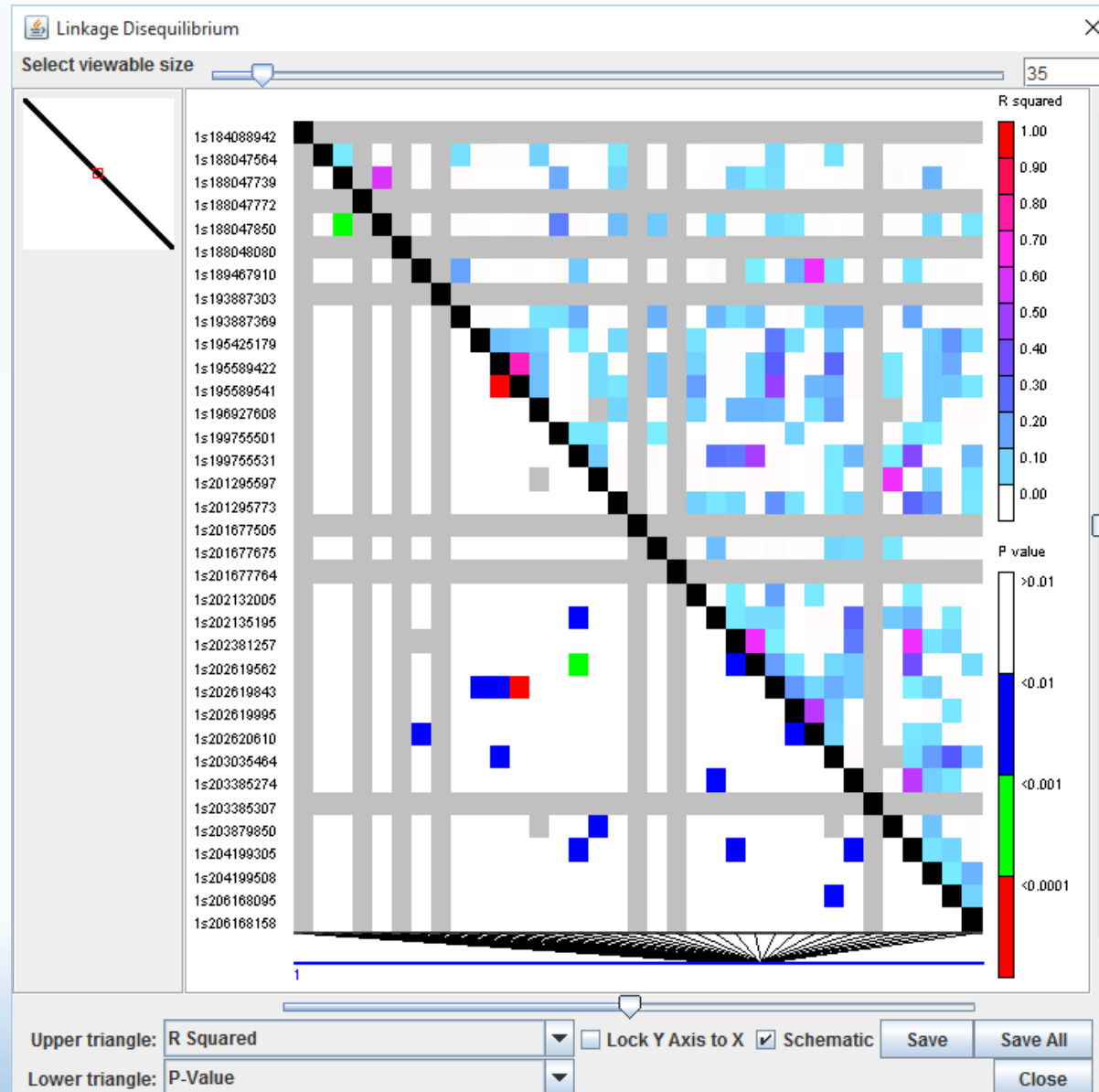
Number of columns: 17

Number of rows: 25725

Number of elements: 437325

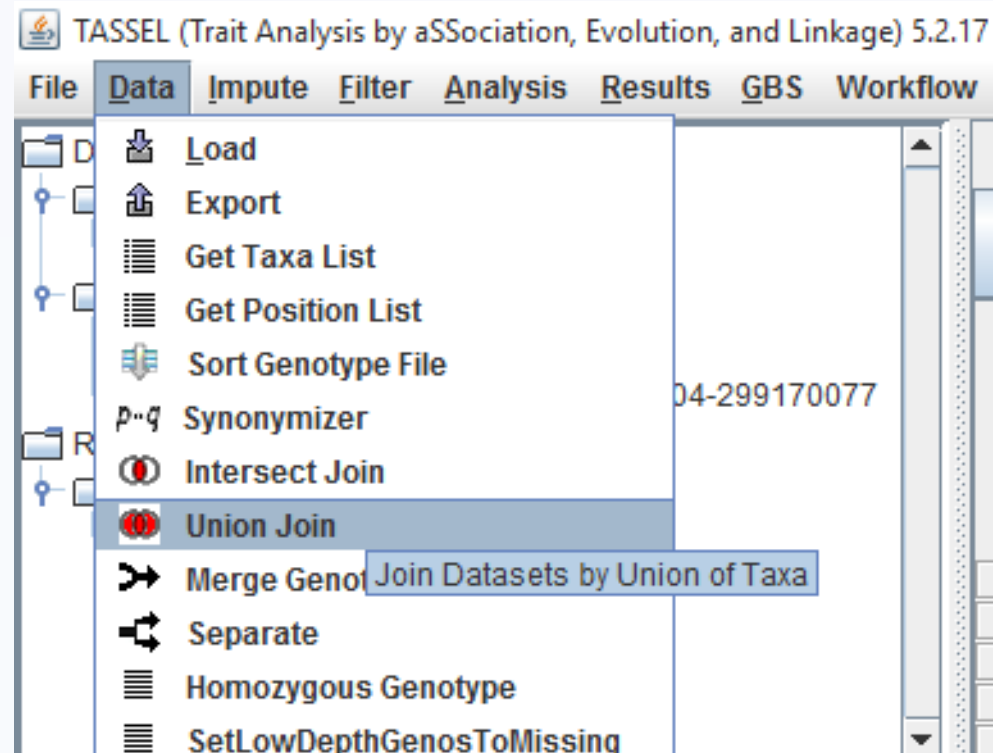
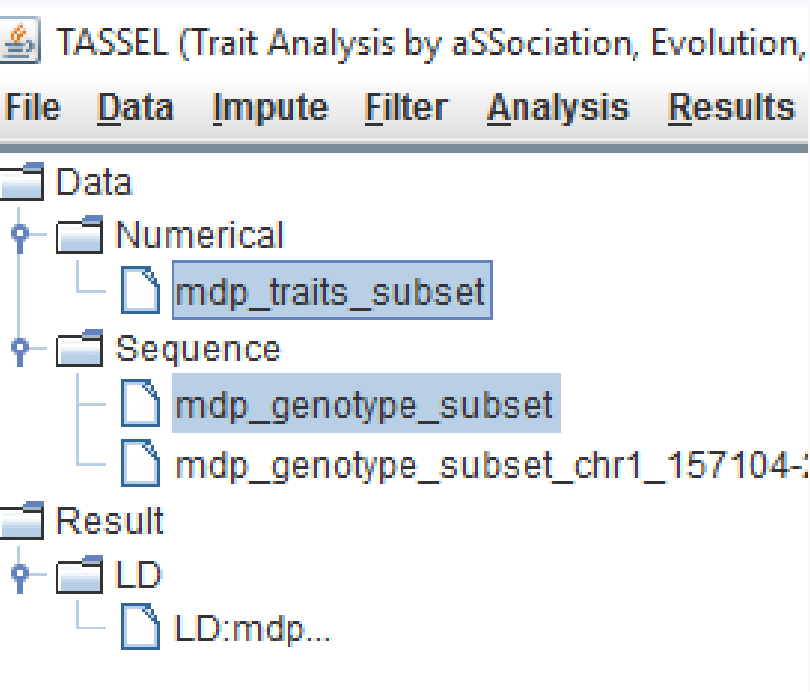
LD Analysis

Creating LD plots



Creating Manhattan plots

Merge genotype and phenotype data



Creating Manhattan plots

TASSEL (Trait Analysis by aSSociation, Evolution, and Linkage) 5.2.17

File Data Impute Filter Analysis Results GBS Workflow

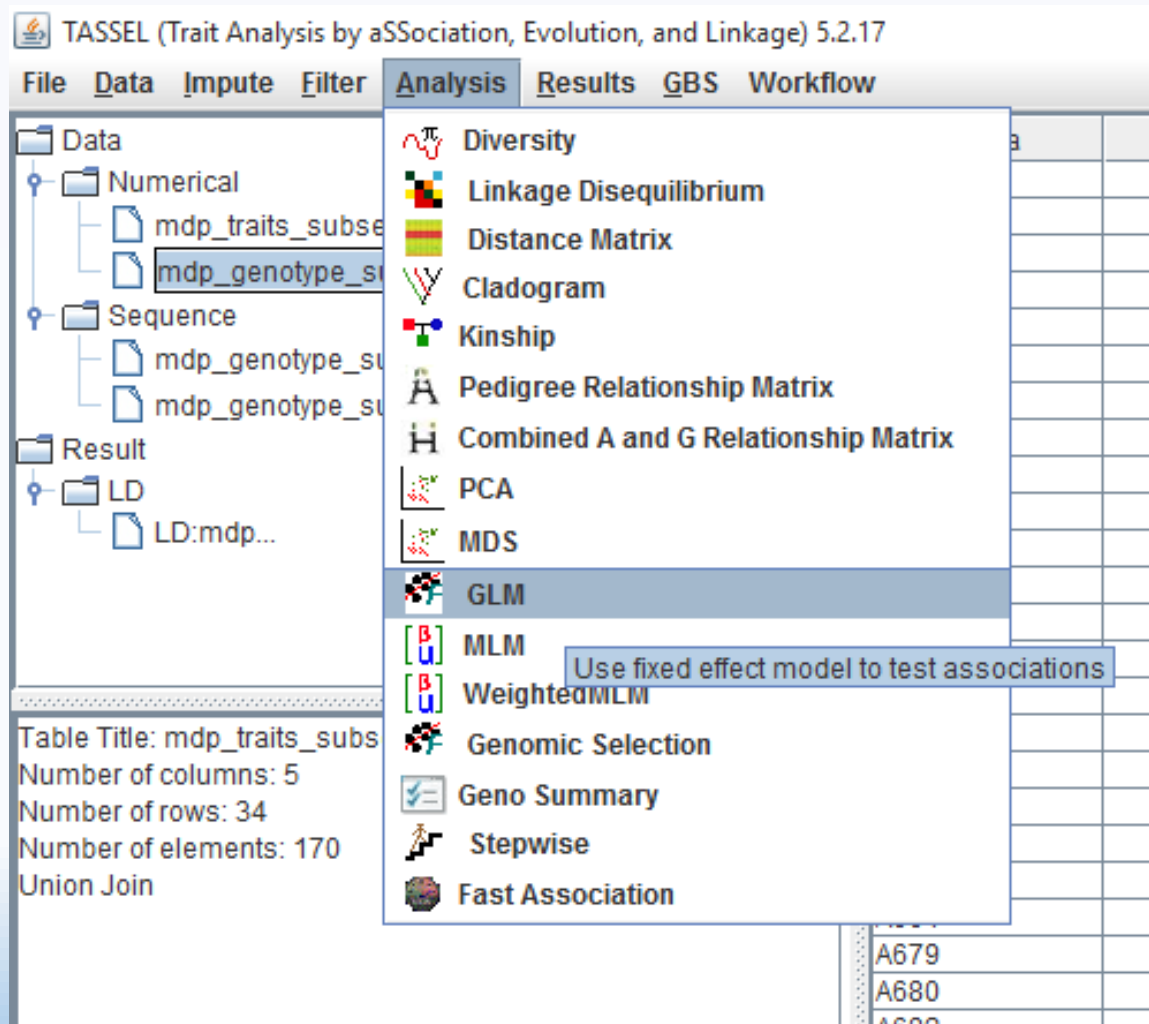
Data
Numerical
mdp_traits_subset
mdp_genotype_subset + mdp_traits_subset
Sequences
mdp_genotype_subset
mdp_genotype_subset_chr1_157104-299170077
Result
LD
LD:mdp...

Table Title: mdp_traits_subset_with_genotypes
Number of columns: 5
Number of rows: 34
Number of elements: 170
Union Join

Taxa	EarHT	dpoll	EarDia	Genotype
811	59.5	<input type="checkbox"/>	<input type="checkbox"/>	CCGTGTG...
33-16	64.75	64.5	<input type="checkbox"/>	CCGTGTC...
38-11	92.25	68.5	37.897	CCGTGTC...
4226	65.5	59.5	32.219	CCGTGTS...
4722	81.13	71.5	32.421	none
A188	27.5	62	31.419	ACGTGTC...
A214N	65	69	32.006	CCTAGAC...
A239	47.88	61	36.064	ACTTAAC...
A272	35.63	70	<input type="checkbox"/>	ACTTAAG...
A441-5	53.5	67.5	35.008	CCGTGTC...
A554	38.5	66	33.418	CGTTATCT...
A556	28	65	31.929	CCGTGTN...
A6	109.5	80.5	31.517	ACTTAAC...
A619	36	61	40.63	CCGTGTC...
A632	60	61	35.953	CCTAGAC...
A634	54	59	35.601	CCTAGAC...
A635	37	64	35.3	CCTAGAC...
A641	54.5	66	33.727	ACTTATCT...
A654	39	64	<input type="checkbox"/>	NGTTATCT...
A659	46.5	58.5	38.846	AGTTATCT...
A661	51.5	59	39.323	CGTTANC...
A679	65	66	42.471	CCTAATGT...
A680	68	65.5	41.152	CCTAATGT...

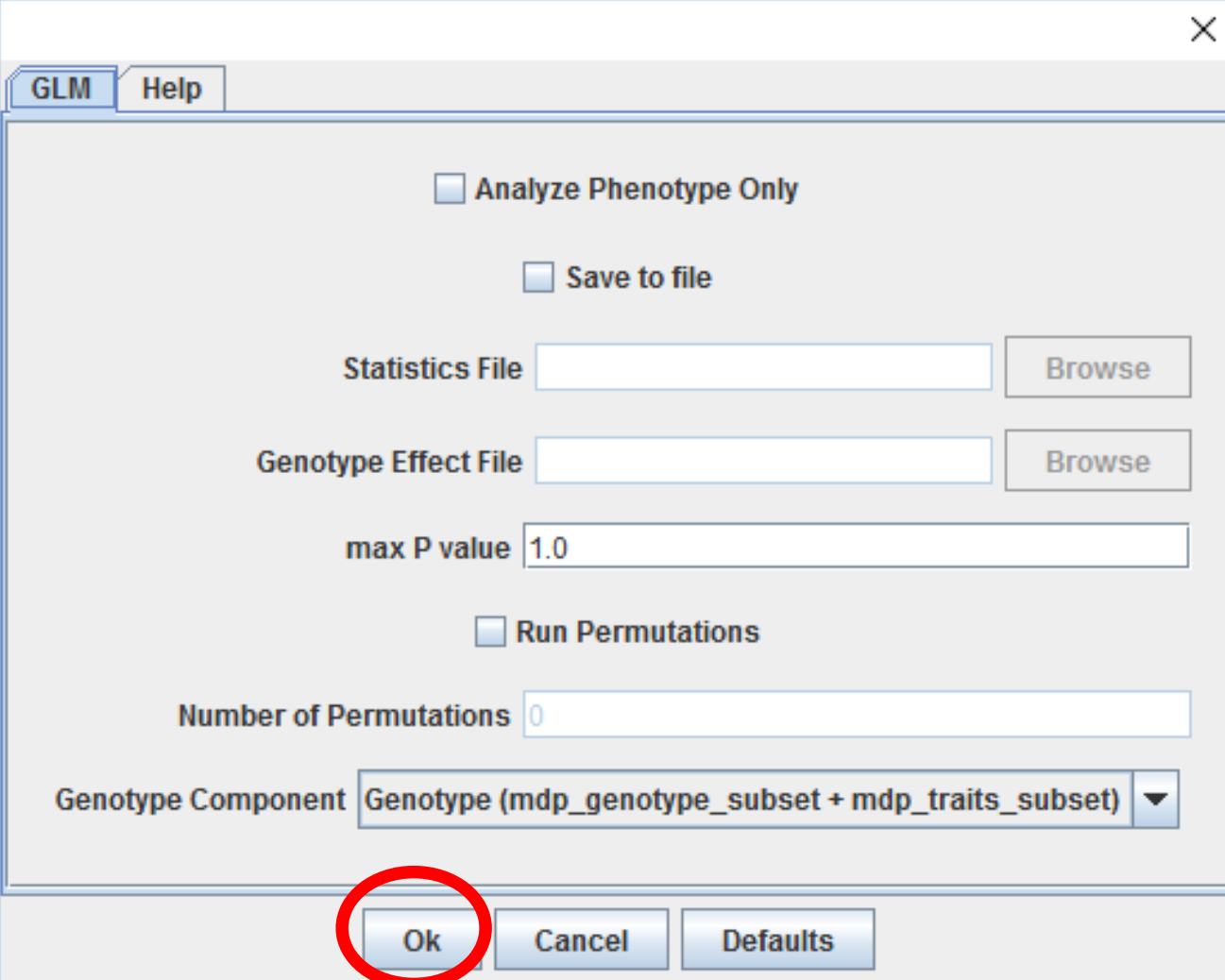
Creating Manhattan plots

Use fixed effect model to test association



Creating Manhattan plots

Use fixed effect model to test association



A screenshot of a software dialog box titled "GLM" with a "Help" button. The dialog contains several options and input fields for creating Manhattan plots. At the bottom, the "Ok" button is circled in red.

☐ Analyze Phenotype Only

☐ Save to file

Statistics File

Genotype Effect File

max P value

☐ Run Permutations

Number of Permutations

Genotype Component ▼

Creating Manhattan plots

Use fixed effect model to test association

TASSEL (Trait Analysis by aSSociation, Evolution, and Linkage) 5.2.17

File Data Impute Filter Analysis Results GBS Workflow

Data

- Numerical
 - mdp_traits_subset
 - mdp_genotype_subset + mdp_traits_subset
- Sequence
 - mdp_genotype_subset
 - mdp_genotype_subset_chr1_157104-299170077

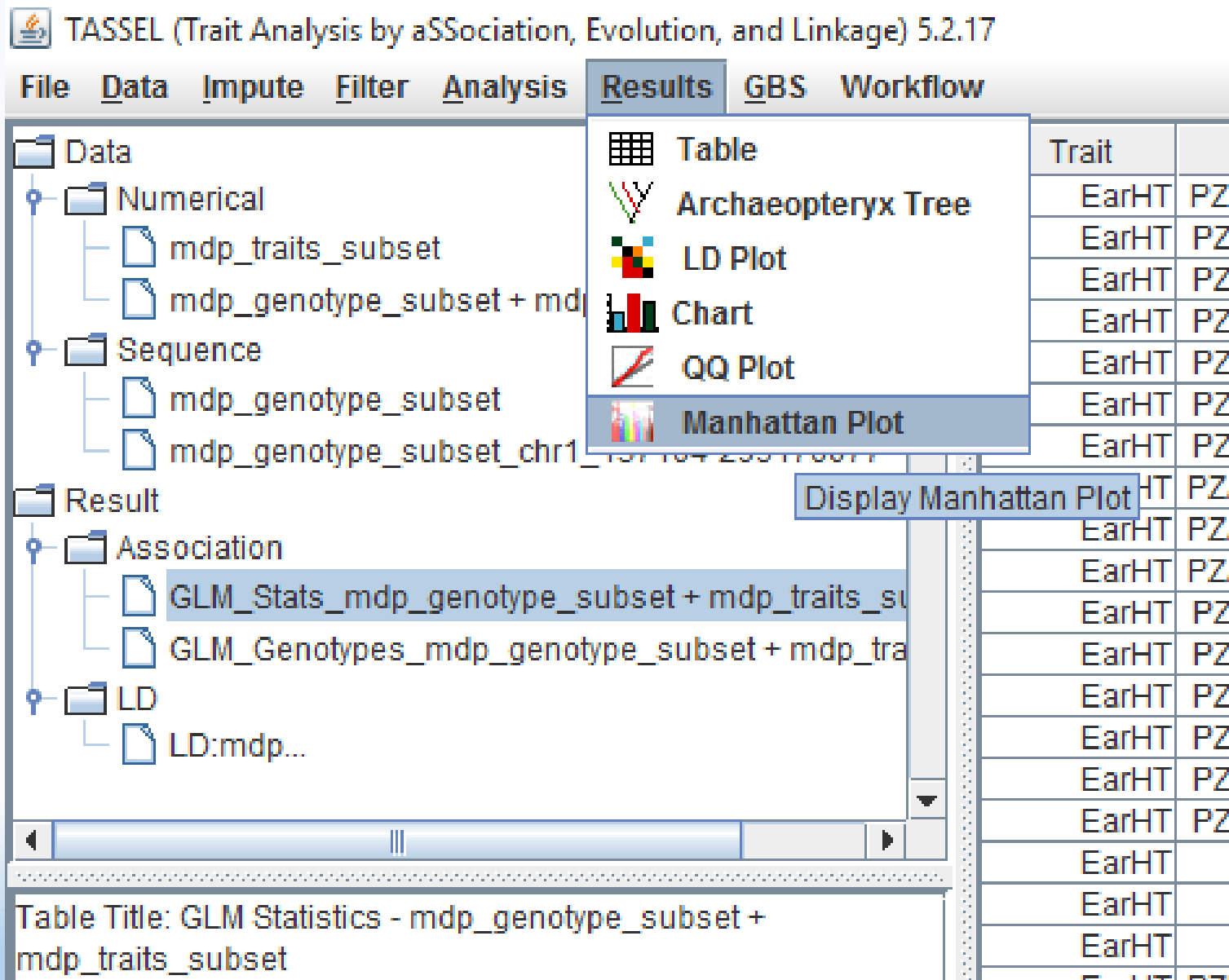
Result

- Association
 - GLM_Stats_mdp_genotype_subset + mdp_traits_subset**
 - GLM_Genotypes_mdp_genotype_subset + mdp_traits_subset
- LD
 - LD:mdp...

Trait	Marker	Chr	Pos	marker_F	p	ma
EarHT	PZB00859.1	1	157104	0.32788	0.57118	
EarHT	PZA01271.1	1	1947984	0.25598	0.61648	
EarHT	PZA03613.2	1	2914066	0.00553	0.94122	
EarHT	PZA03613.1	1	2914171	0.10799	0.74465	
EarHT	PZA03614.2	1	2915078	0.0222	0.88253	
EarHT	PZA03614.1	1	2915242	0.1038	0.74962	
EarHT	PZA00258.3	1	2973508	0.89718	0.41873	
EarHT	PZA02962....	1	3205252	0.17348	0.6801	
EarHT	PZA02962....	1	3205262	0.17348	0.6801	
EarHT	PZA00599....	1	3206090	0.64495	0.42824	
EarHT	PZA02129.1	1	3706018	0.38301	0.54052	
EarHT	PZA00393.1	1	4175293	1.04186	0.31529	
EarHT	PZA02869.8	1	4429897	5.41939	0.02685	
EarHT	PZA02869.4	1	4429927	0.25234	0.61923	
EarHT	PZA02869.2	1	4430055	2.69704	0.11134	
EarHT	PZA02032.1	1	4490461	0.84521	0.36501	
EarHT	zagl1.5	1	4835434			
EarHT	zagl1.2	1	4835558			
EarHT	zagl1.6	1	4835658			
EarHT	PZD00081.2	1	4836542			
EarHT	zagl1.1	1	4912526			
EarHT	PZB00919.1	1	5353319	7.12811	0.01197	
EarHT	PZB00919.2	1	5353655	0.86145	0.3605	
EarHT	PHM2244....	1	5562502	0.47616	0.49547	
EarHT	PZA03093....	1	8075572	2.8417	0.10222	
EarHT	PZA00181.5	1	8366368	0.13641	0.87306	

Table Title: GLM Statistics - mdp_genotype_subset + mdp_traits_subset
Number of columns: 17
Number of rows: 9279
Number of elements: 157743
GLM Output
Statistical Tests for individual variants.
Input data: mdp_genotype_subset + mdp_traits_subset

Creating Manhattan plots



Creating Manhattan plots

TASSEL (Trait Analysis by aSSociation, Evolution, and Linkage) 5.2.17

FileDataImputeFilterAnalysisResultsGBSWorkflow

Data

Numerical

mdp_traits_subset

mdp_genotype_subset + mdp_traits_subset

Sequence

mdp_genotype_subset

mdp_genotype_subset_chr1

Result

Association

GLM_Stats_mdp_genotype_subset + mdp_traits_subset

GLM_Genotypes_mdp_genotype_subset + mdp_traits_subset

LD

LD:mdp...

Table

Archaeopteryx Tree

LD Plot

Chart

QQ Plot

Manhattan Plot

Trait	
EarHT	PZ
EarHT	PZ
EarHT	PZ
EarHT	PZ
EarHT	PZ
EarHT	PZ
EarHT	PZ
EarHT	PZ
EarHT	PZ
EarHT	PZ
EarHT	PZ
EarHT	PZ
EarHT	PZ
EarHT	PZ
EarHT	PZ

Display Manhattan Plot

Table Title: GLM Statistics - mdp_genotype_subset + mdp_traits_subset

1	4835558			0	
6	4835658			0	
2				0	
1				0	
1				8695	
2				2704	
...				1562	
...				8653	
5				0965	
2	1	8366	79031	0.00232	
1	1	8367	39568	0.02589	
2	1	8510027	0.18461	0.67051	0.00612
6	1	9023947	0.00091	0.97609	0.00003
8	1	9024005	0.02717	0.87034	0.00104

Manhattan Plot Options

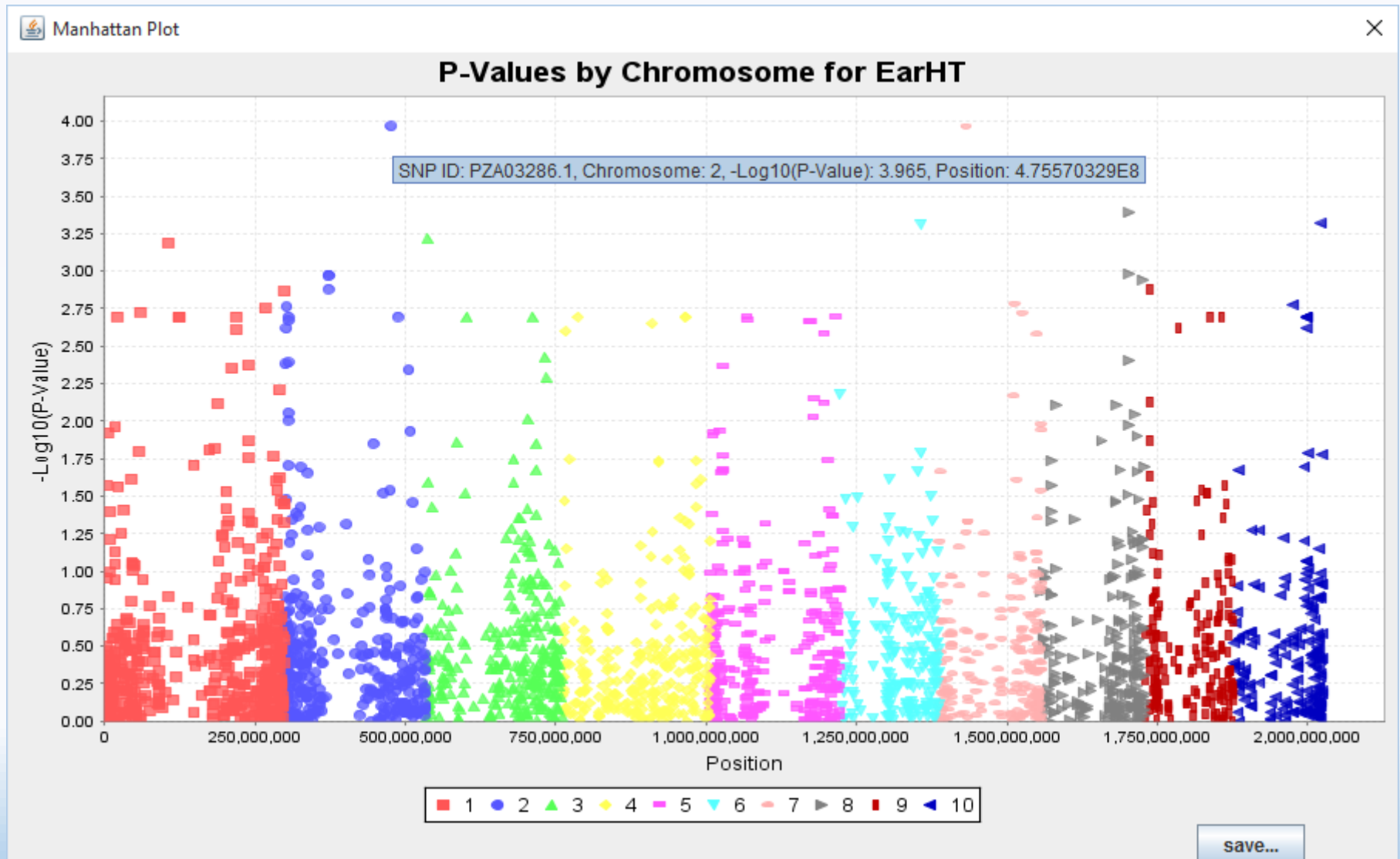
Select trait

EarHT

dpoll

EarDia

Creating Manhattan plots



Creating QQ plot

TASSEL (Trait Analysis by aSSociation, Evolution, and Linkage) 5.2.17

File Data Impute Filter Analysis Results GBS Workflow

mdp_traits_subset
mdp_genotype_subset + mdp_traits_subset
Sequence
mdp_genotype_subset
mdp_genotype_subset_chr1_157
Tree
Tree:mdp_genotype_subset
Tree:mdp_genotype_subset
result
Association
GLM_Stats_mdp_genotype
GLM_Genotypes_mdp_genotype

Table
Archaeopteryx Tree
LD Plot
Chart
QQ Plot
Manhattan Plot

Trait	Marker	Chr
EarHT	PZB00859.1	1
EarHT	PZA01271.1	1
EarHT	PZA03613.2	1
EarHT	PZA03613.1	1
EarHT	PZA03614.2	1
EarHT	PZA03614.1	1
EarHT	PZA00258.3	1
EarHT	PZA02962....	1
EarHT	PZA02962....	1
EarHT	PZA00599....	1

Display QQ Plot

QQ Plot Options

Available Traits

Traits to Plot

>> > < <<

EarHT
dpoll
EarDia

Plot Density:

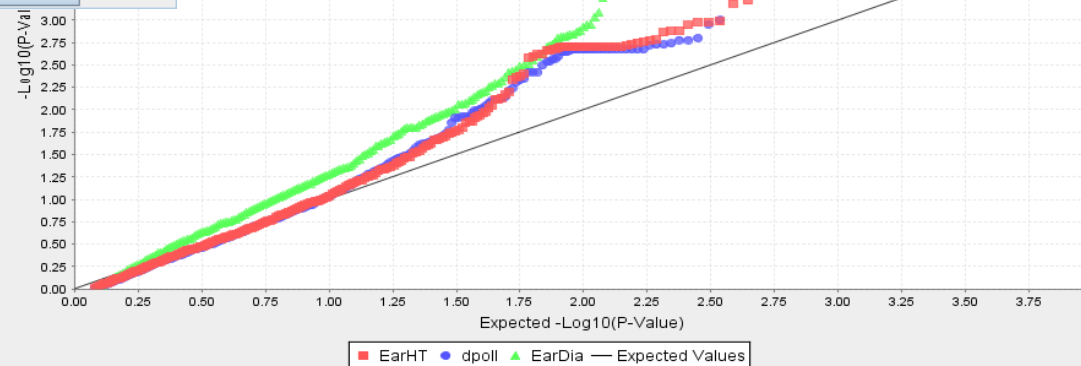
30

of 3093 points per trait.

Okay

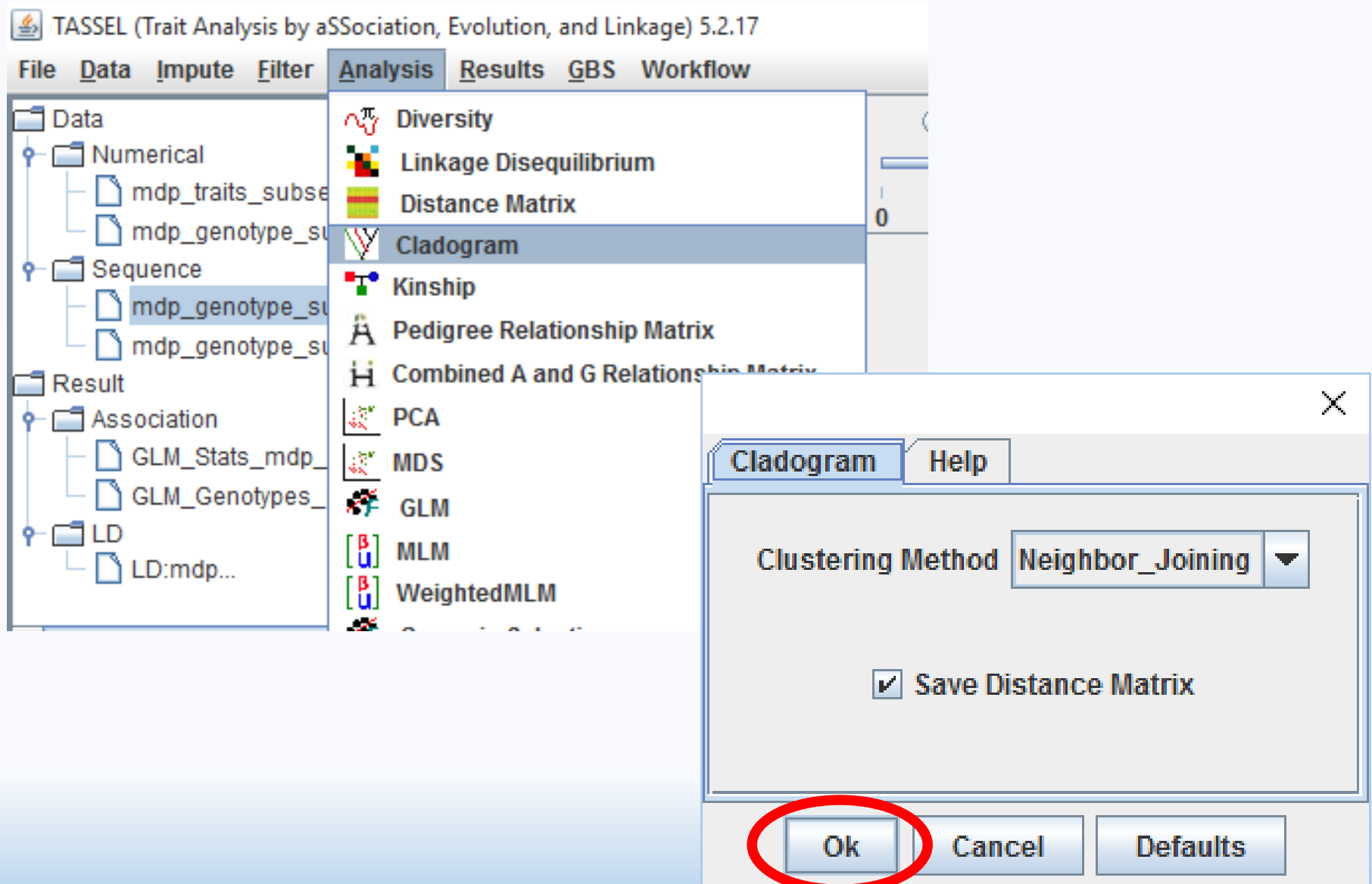
Cancel

Expected -Log10(P-Value) vs. -Log10(P-Value)

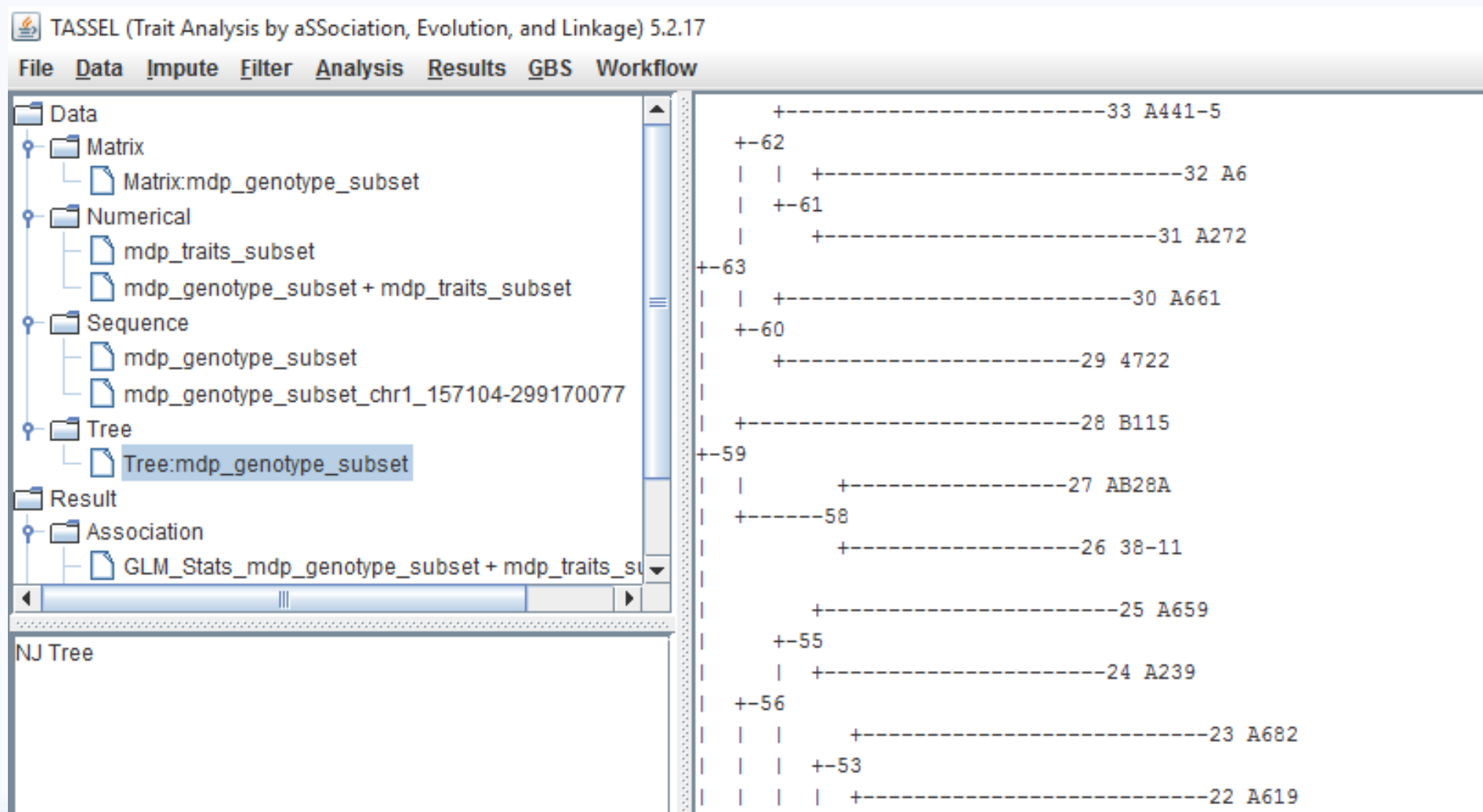


save...

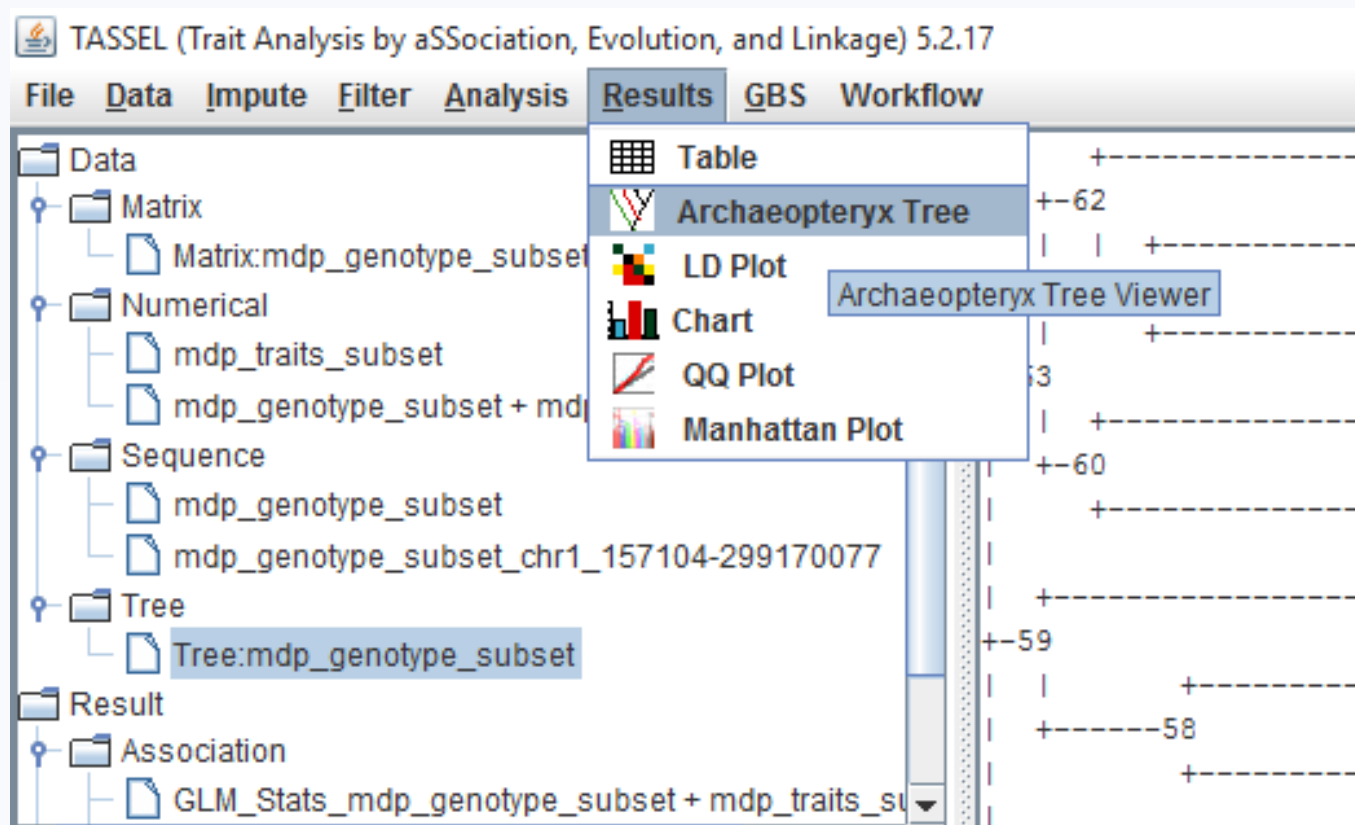
Creating cladograms



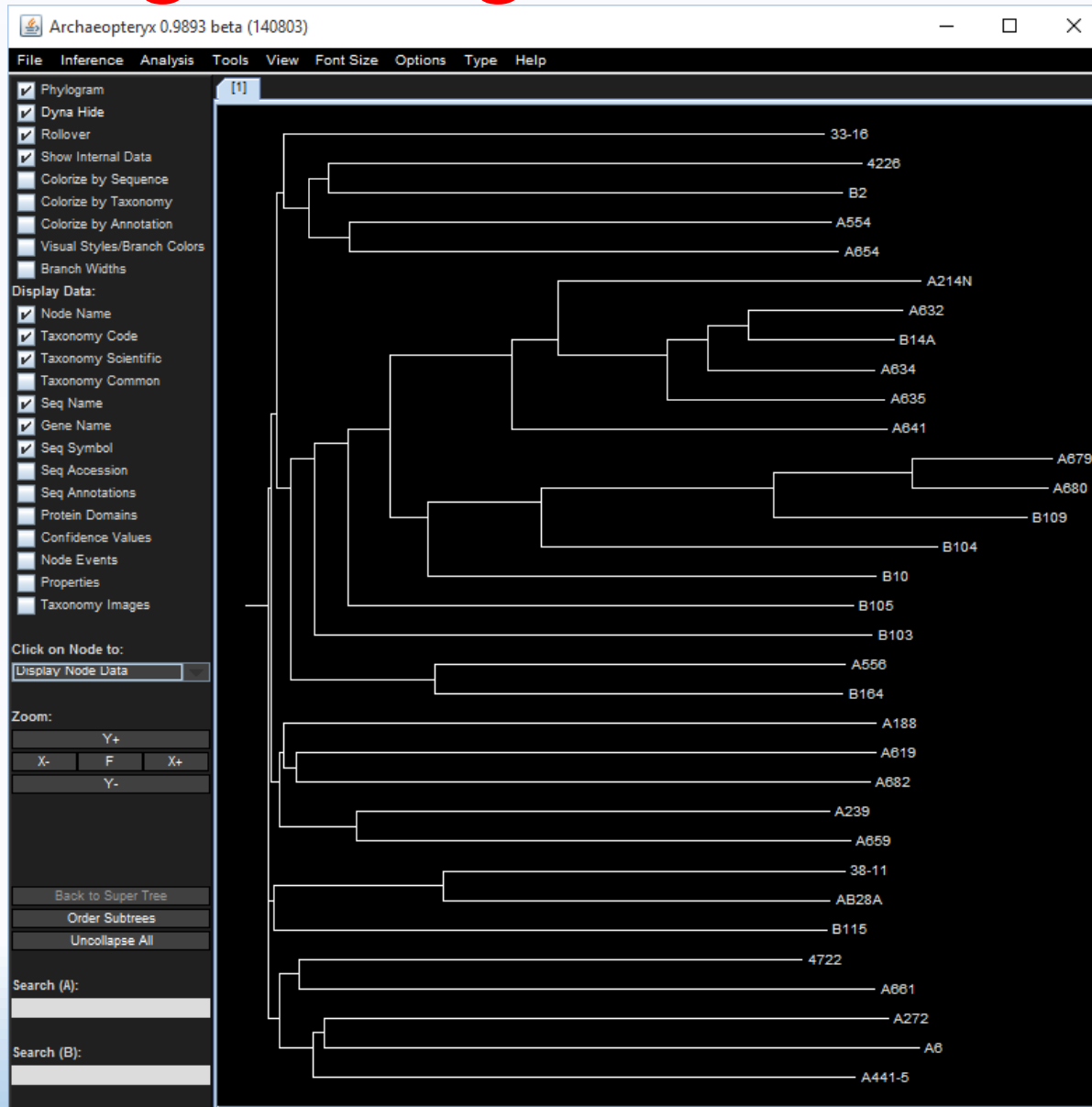
Creating cladograms



Creating cladograms



Creating cladograms



Thank you!