Using Galaxy to Conduct Genomic Selection Analyses

1

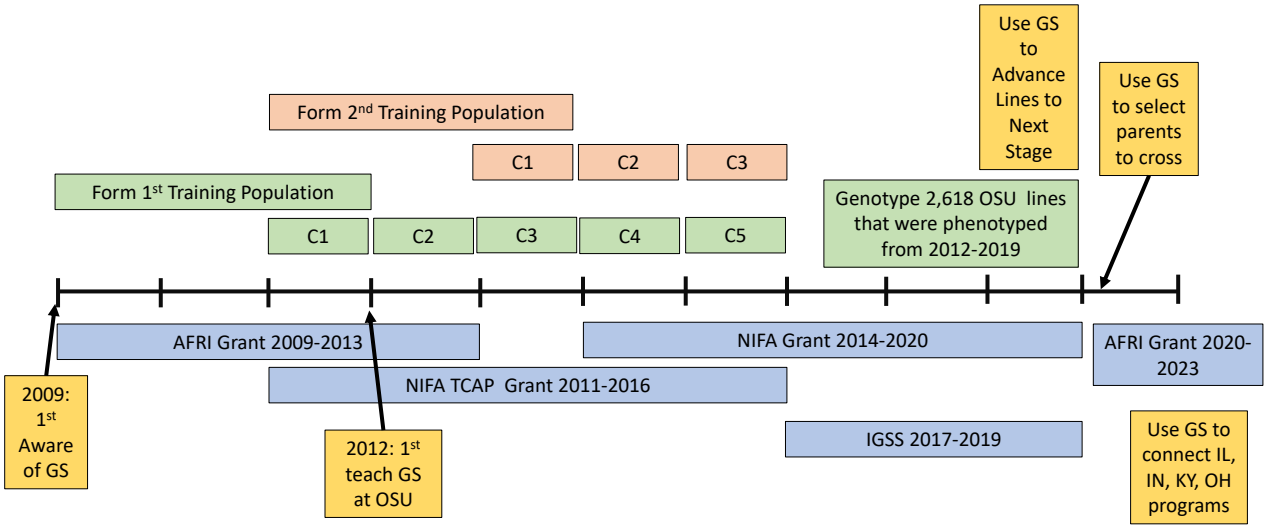
Myself
Professor: The Ohio State University, wheat
Assist, Assoc Prof: University of Arkansas, soybean
Consultant with IGSS

OSU, Michigan State University
41 Years in Agriculture
35 Years in Plant Breeding
Worked with crops in 3 states, 13 Countries
Wheat, Soybean
Taught classes/workshops in 9 Countries
Plant breeding, statistics, quantitative genetics, molecular breeding

- Quality
- Other Diseases
- Agronomic Traits
- Biomass
- Diversity

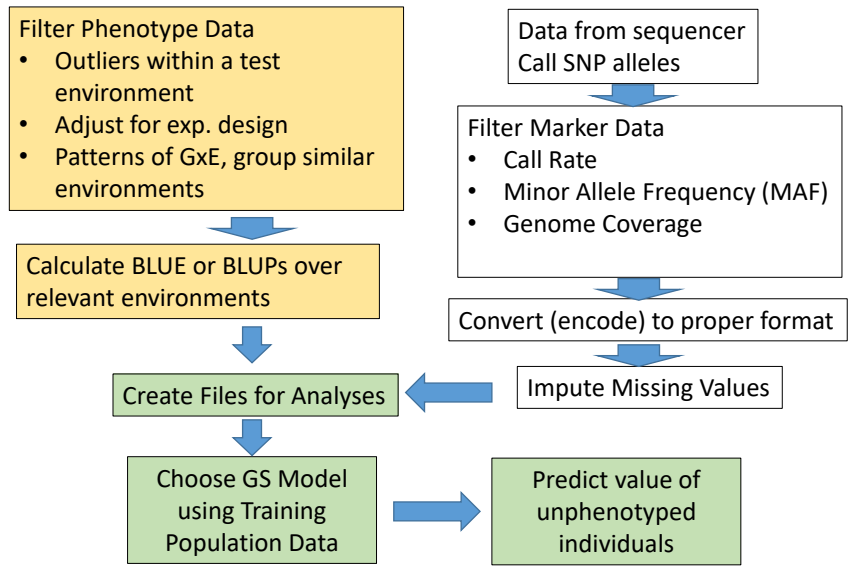
2

Genomic Selection in OSU Wheat Breeding: *Change is a Process*



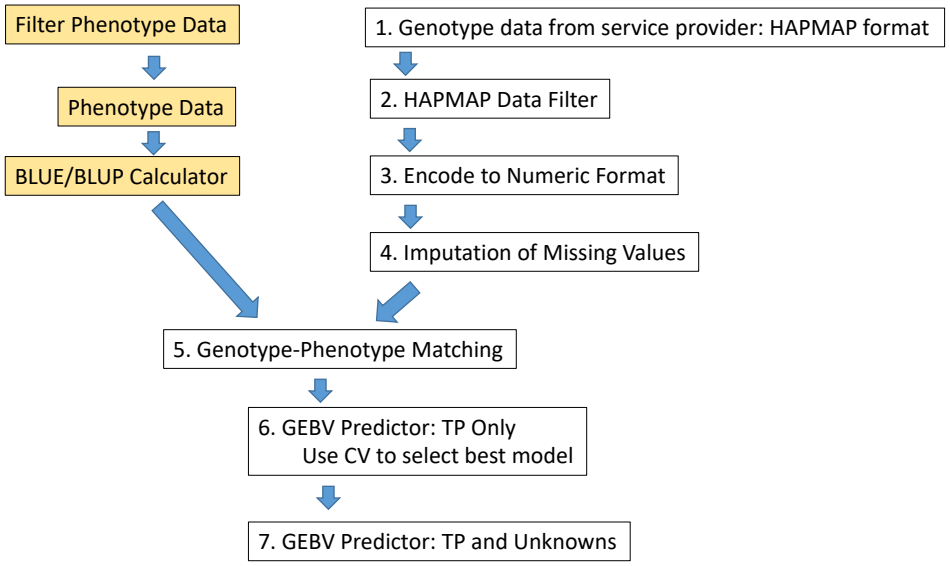
3

GS Data and Analysis Steps



4

GOBii, Galaxy for Genomic Selection



5

Numeric Data Formats for GS Analyses

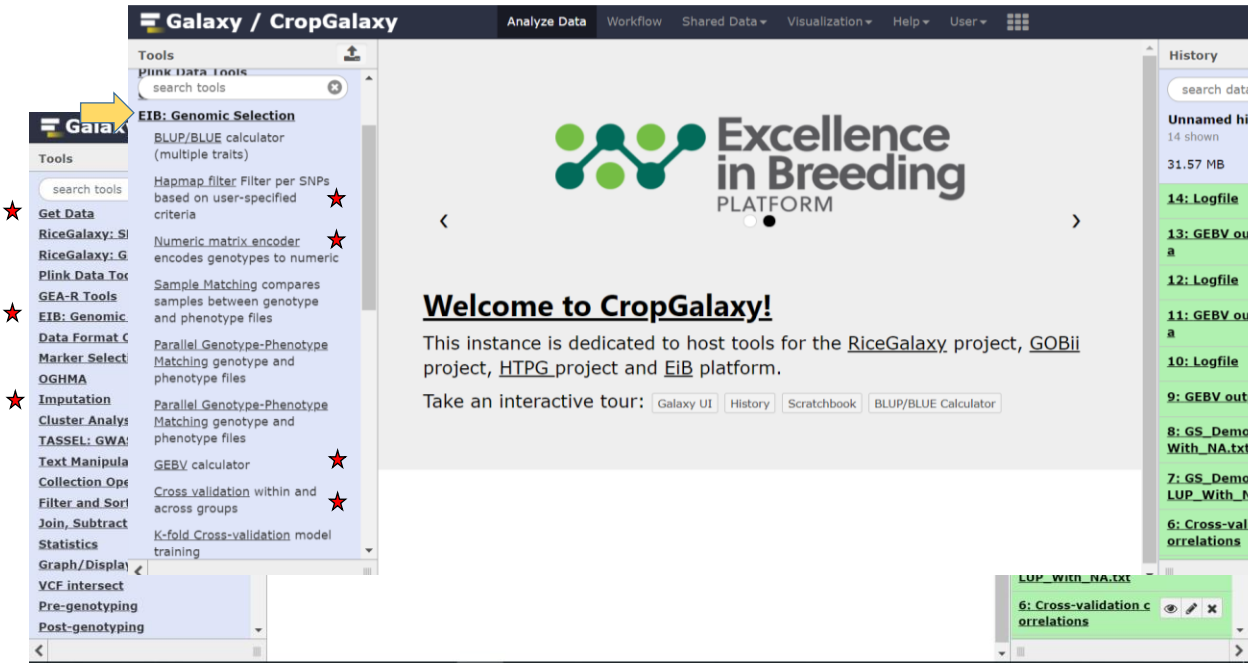
Source	Homozygous Reference Allele	Heterozygote	Homozygous SNP Allele
HAPMAP	AA	AC	CC
DArTseq 1-Row	0	2	1
Other 1-row	-1	0	1
Other 1-row	0	1	2

GS Uses numeric data
Cant Have Missing Values

Be VERY CAREFUL

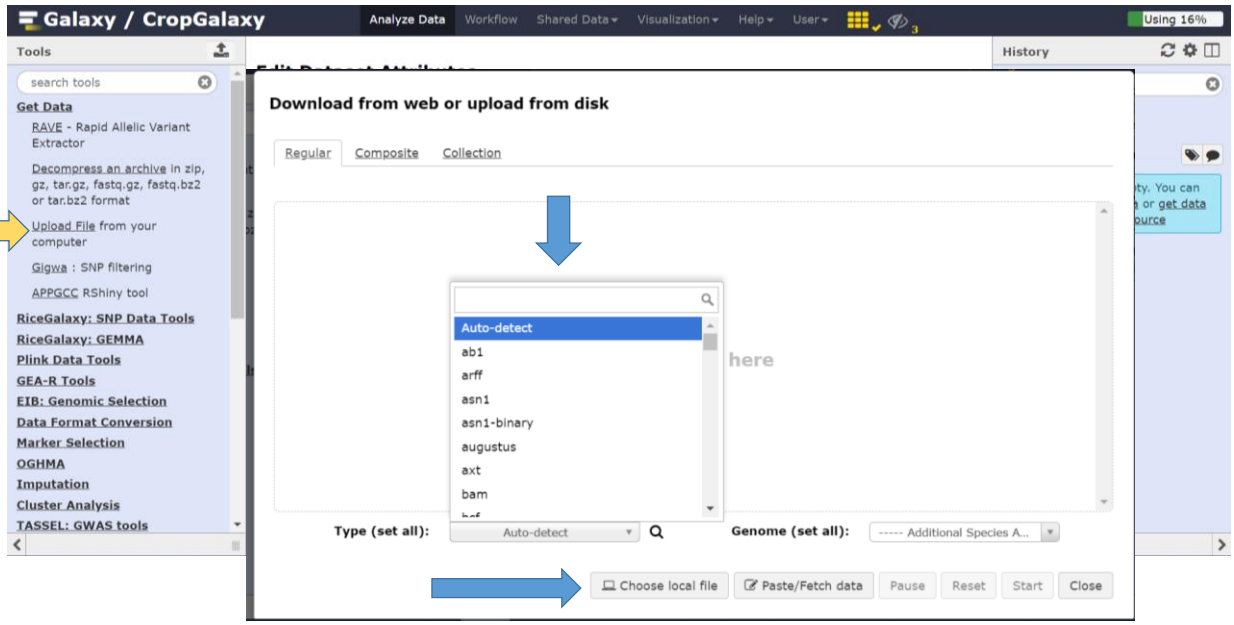
- Order of samples must be identical in genotype and phenotype files
- Order of markers must be identical in TP and PP data files
- Names and spellings must be identical in all files
- Designation of missing values need to be as specified in software

6


Galaxy / CropGalaxy
Analyze Data
Workflow
Shared Data
Visualization
Help
User
Tools
search tools
EIB: Genomic Selection
BLUP/BLUE calculator (multiple traits)
Hapmap filter Filter per SNPs based on user-specified criteria
Numeric matrix encoder encodes genotypes to numeric
Sample Matching compares samples between genotype and phenotype files
Parallel Genotype-Phenotype Matching genotype and phenotype files
Parallel Genotype-Phenotype Matching genotype and phenotype files
GEBV calculator
Cross validation within and across groups
K-fold Cross-validation model training
Get Data
OGHMA
Imputation
Cluster Analys
Text Manipula
Collection Ope
Join, Subtract
Statistics
VCF intersect
Pre-genotyping
Post-genotyping
Excellence in Breeding PLATFORM
Welcome to CropGalaxy!
This instance is dedicated to host tools for the RiceGalaxy project, GOBii project, HTPG project and EiB platform.
Take an interactive tour:
Galaxy UI
History
Scratchbook
BLUP/BLUE Calculator
History
Unnamed hi
14 shown
31.57 MB
14: Logfile
13: GEBV ou a
12: Logfile
11: GEBV ou a
10: Logfile
9: GEBV out
8: GS_Demo With_NA.txt
7: GS_Demo LUP_With_N
6: Cross-val orrelations
6: Cross-validation c orrelations

7

1. Get Data


Galaxy / CropGalaxy
Analyze Data
Workflow
Shared Data
Visualization
Help
User
Using 16%
Tools
search tools
Get Data
RAVE - Rapid Allelic Variant Extractor
Decompress an archive in zip, gz, tar.gz, fastq.gz, fastq.bz2 or tar.bz2 format
Upload File from your computer
Gigwa : SNP filtering
APPGCC RShiny tool
RiceGalaxy: SNP Data Tools
RiceGalaxy: GEMMA
Plink Data Tools
GEA-R Tools
EIB: Genomic Selection
Data Format Conversion
Marker Selection
OGHMA
Imputation
Cluster Analysis
TASSEL: GWAS tools
History
Download from web or upload from disk
Regular
Composite
Collection
Auto-detect
ab1
arff
asn1
asn1-binary
augustus
axt
bam
here
Type (set all):
Auto-detect
Genome (set all):
----- Additional Species A...
Choose local file
Paste/Fetch data
Pause
Reset
Start
Close

8

2. HAPMAP filter

Objective: Filter out markers that you want to exclude from the analysis

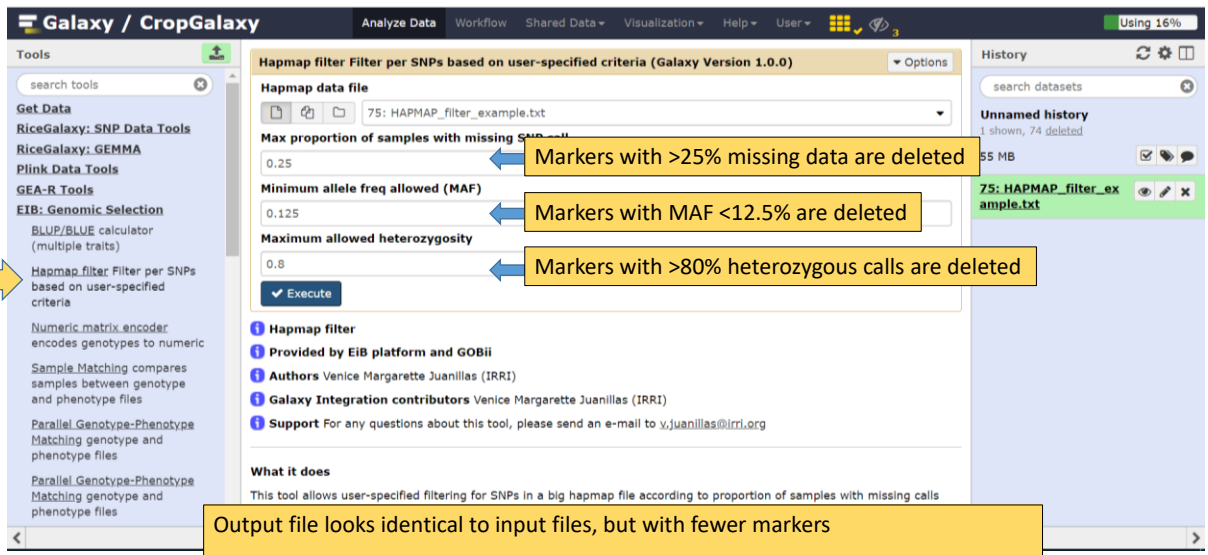
Required: a HAPMAP file of marker data

- Tab delimited
- First rows begin with #. These are meta data not used in any analysis
- Header row identify each column starts with "rs#"
- Marker data starts in first row after the header row: 1 row per marker
- First 11 columns contain data describing each marker
- SNP data begins in column 12: there is one column for each sample
- Example, 1st marker is 0100001151, the SNP is C or A, first four individuals are homozygous CC

#											dnarun_barcode			
34 other meta data (#) rows											project_	test_load_	test_load_	test_load_	test_load_
#											dnasample_ref_sample				
rs#	alleles	chrom	pos	strand	assembl	center	protLSID	assayLS	panelLS	QCcode	IRIS_313-1	IRIS_313-1	IRIS_313-1	IRIS_313-9
10100001151	C/A			+							CC	CC	CC	CC
10100001173	C/T			+							CC	CC	CC	CC
10100001178	G/T			+							GG	GG	GG	GG
10100001203	T/C			+							CC	CC	CC	CC
10100001248	G/A			+							GG	GG	GG	GG
10100001249	A/C			+							CC	CC	CC	CC
10100001266	G/A			+							AA	AA	AA	AA

9

2. Hapmap Filter



Output file looks identical to input files, but with fewer markers

Example: HAPMAP_filter_example file has 583 markers
Output file (Hapmap_data_filter_on_data_132) has 146 markers

10

3. Encode

Objective: Convert HAPMAP character data to numeric data

Required: a HAPMAP file of marker data

- Same as shown before
- Data should have been filtered before encoding
- **Homozygotes for minor allele converts to "0"**
- **Heterozygotes convert to "1"**
- **Heterozygotes for major allele converts to "2"**

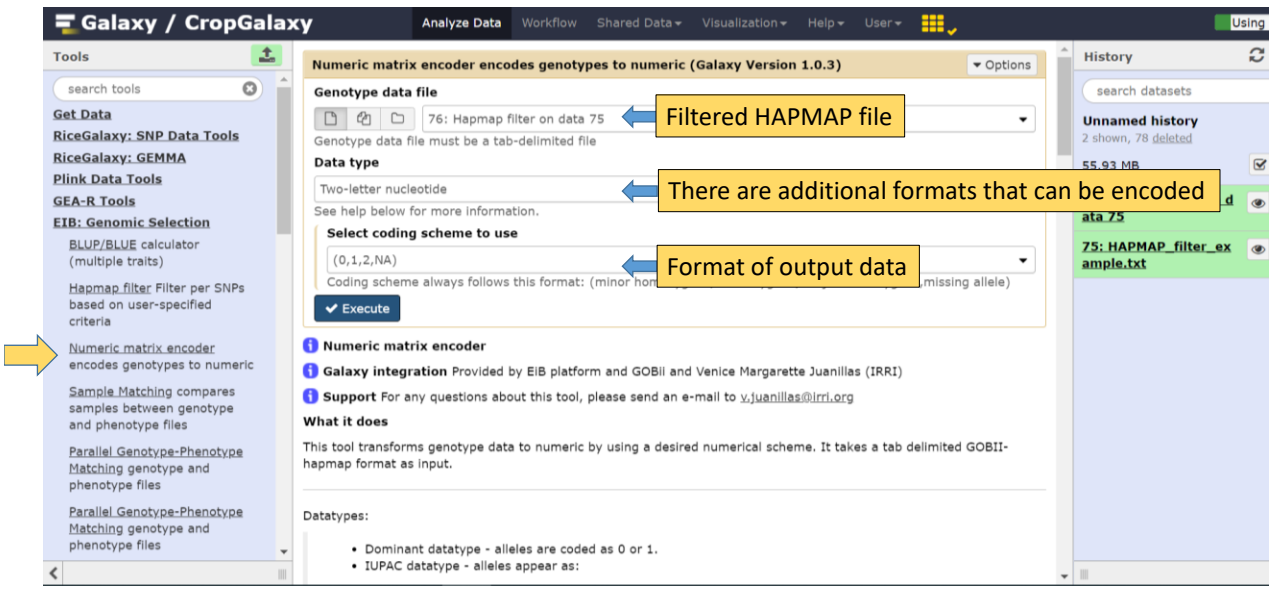
Note: Encoder can handle alternative HAPMAP formats

Example for one marker:
The frequency of the "C" allele is 0.8, thus it is the major allele
The frequency of the "A" allele is 0.2, thus it is the minor allele
Missing data is encoded as "NA"

Note: Encoder can also produce -1, 0, 1 scoring

IND1	IND2	IND3	IND4	IND5	IND6	IND7	IND8	IND9	IND10	IND11	
AA	CC	CC	CC	CC	CC	AC	CC	AC	CC	NN	HAPMAP
0	2	2	2	2	2	1	2	1	2	NA	Encoded

3. Numeric Encoder



Output file from Encoding:

Still in a HAPMAP format, just numeric scoring

#									dnarun_barcode								
34 addition # rows																	
#									dnasample_ref_sample								
rs#	alleles	chrom	pos	strand	sembl	center	rotLSI	ssayLSI	anelLSI	QCcode	IRIS_313	11806	IRIS_313	11644	IRIS_313	11515	IRIS_313	9944
10100002469	A/G			+							0	0	0	0	0	0	0	0
10100011130	A/G			+							0	NA	0	0	0	0	0	1
10100011334	T/C			+							0	0	0	0	2	0	2	2
10100021546	T/A			+							NA	1	0	1	2	1	2	2

13

4. Naïve Imputation

Objective: Replace missing genotypes (“NA”) with a numeric value

Missing marker values (NA) within a row are replaced with either the mean or the mode of the values in that row

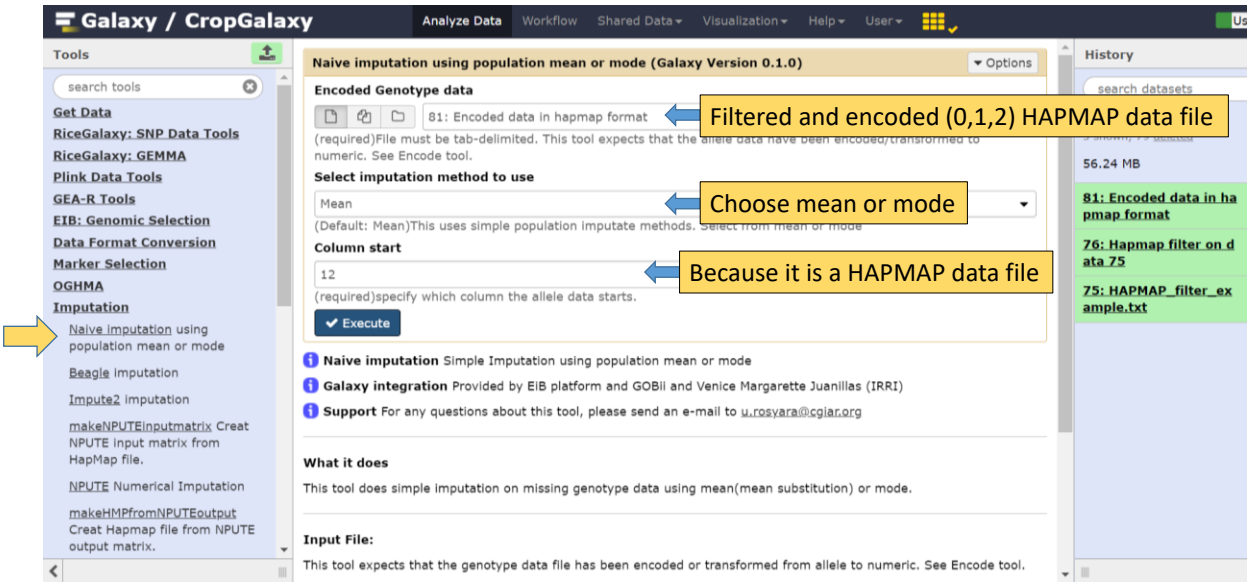
Required: a Filtered and Encoded HAPMAP file of marker data. Generated by #3

Tab delimited

#									dnarun_barcode								
34 addition # rows																	
#									dnasample_ref_sample								
rs#	alleles	chrom	pos	strand	sembl	center	rotLSI	ssayLSI	anelLSI	QCcode	IRIS_313	11806	IRIS_313	11644	IRIS_313	11515	IRIS_313	9944
10100002469	A/G			+							0	0	0	0	0	0	0	0
10100011130	A/G			+							0	NA	0	0	0	0	0	1
10100011334	T/C			+							0	0	0	0	2	0	2	2
10100021546	T/A			+							NA	1	0	1	2	1	2	2

14

4. Imputation



15

Output file from Naïve imputation

#										dnarun_barcode						
34 addition # rows																
#										dnasample_ref_sample						
rs#	alleles	chrom	pos	strand	sembl	center	rotLSI	sayLSI	nelLSI	QCode	IRIS_313	11806	IRIS_313	11644	IRIS_313	11515
10100002469	A/G	NA	NA	+	NA	NA	NA	NA	NA	NA	0	0	0	0	0	0
10100002706	C/T	NA	NA	+	NA	NA	NA	NA	NA	NA	0	0	0	0	0	0
10100011130	A/G	NA	NA	+	NA	NA	NA	NA	NA	NA	0	1.7	0	0	0	0
10100011334	T/C	NA	NA	+	NA	NA	NA	NA	NA	NA	0	0	0	0	2	0
10100011408	G/C	NA	NA	+	NA	NA	NA	NA	NA	NA	0	0	0	0	0	0
10100021546	T/A	NA	NA	+	NA	NA	NA	NA	NA	NA	1	1	0	1	2	1
10100028751	T/G	NA	NA	+	NA	NA	NA	NA	NA	NA	0	0	0	0	0	1.6
10100028865	T/C	NA	NA	+	NA	NA	NA	NA	NA	NA	0	0	0	0	0	0
10100028959	T/C	NA	NA	+	NA	NA	NA	NA	NA	NA	0	0	0	0	0	0
10100029432	T/C	NA	NA	+	NA	NA	NA	NA	NA	NA	0	0	0	0	0	1.7

Mean of the numeric values in that row

16

5. Genotype-Phenotype Matching

Objective: Ensure the phenotype and genotype files contain data on the same lines and in the same order

The number of samples in each file type must be identical
The names of the samples must be identical
The order of the samples must be identical

Pheno File

GID	Trait1
AAA	100
BBB	97
CCC	68
DDD	58
EEE	92
FFF	102
GGG	85
HHH	76
III	48
JJJ	96
etc	105

Genotype File

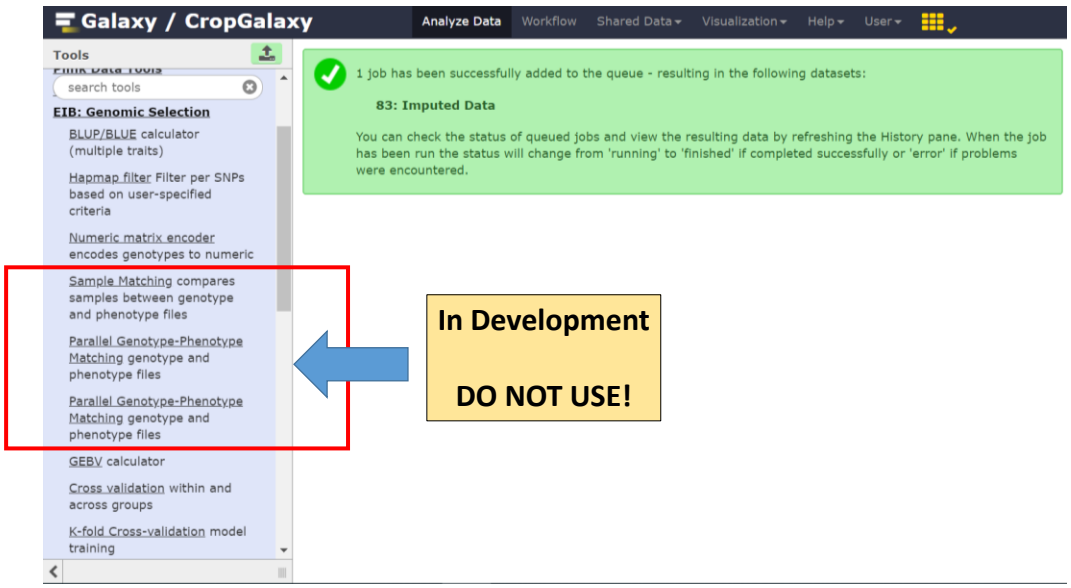
GID	M1	M2	M3	M4
AAA	0	2	2	0
BBB	2	2	2	0
CCC	0	0	1	0
DDD	2	2	0	0
EEE	2	0	1	2
FFF	2	0	2	2
GGG	1	0	0	2
HHH	0	0	0	0
III	0	0	0	0
JJJ	0	2	0	0
etc	2	2	2	0

Genotype File in format for the analysis

M1	M2	M3	M4
0	2	2	0
2	2	2	0
0	0	1	0
2	2	0	0
2	0	1	2
2	0	2	2
1	0	0	2
0	0	0	0
0	0	0	0
0	2	0	0
2	2	2	0

17

5. Matching Genotype and Phenotype Files



18

6. GEBV Calculator: GS Accuracy by Cross Validation

Objective: Run various prediction models **on the training population data only** to assess which model is best

Required Files:

Filtered, Encoded, imputed genotype data file with column of sample names removed

Phenotype file, Tab delimited

Pheno File

GID	Trait1
AAA	100
BBB	97
CCC	68
DDD	58
EEE	92
FFF	102
GGG	85
HHH	76
III	48
JJJ	96
etc	105

Genotype File

M1	M2	M3	M4
0	2	2	0
2	2	2	0
0	0	1	0
2	2	0	0
2	0	1	2
2	0	2	2
1	0	0	2
0	0	0	0
0	0	0	0
0	2	0	0
2	2	2	0

19

Sample Data Files

Pheno_for_Demo 190 samples

Sample.Name.in.GBS.Data	Season	GY_blue	GY_blup	GY_PEV	GY_pevReliability	GY_means
MARS-WEMA_270039	Main	6.715	6.975	0.05	0.176	6.715
MARS-WEMA_270040	Main	5.959	6.841	0.05	0.176	5.959
MARS-WEMA_270041	Main	6.925	7.012	0.05	0.176	6.925
MARS-WEMA_270042	Main	6.925	7.127	0.05	0.176	7.576
MARS-WEMA_270043	Main	8.101	7.22	0.05	0.176	8.101
MARS-WEMA_270044	Main	7.335	7.084	0.05	0.176	7.335
MARS-WEMA_270045	Main	8.015	7.205	0.05	0.176	8.015
MARS-WEMA_270046	Main	7.234	7.066	0.05	0.176	7.234
MARS-WEMA_270047	Main	6.87	7.002	0.05	0.176	6.87
MARS-WEMA_270048	Main	7.43	7.101	0.05	0.176	7.43
MARS-WEMA_270049	Main	7.178	7.056	0.05	0.176	7.178
MARS-WEMA_270050	Main	6.555	6.946	0.05	0.176	6.555
MARS-WEMA_270051	Main	6.491	6.935	0.05	0.176	6.491
MARS-WEMA_270052	Main	6.811	6.992	0.05	0.176	6.811
MARS-WEMA_270053	Main	7.82	7.17	0.05	0.176	7.82
MARS-WEMA_270054	Main	6.384	6.916	0.05	0.176	6.384
MARS-WEMA_270055	Main	6.384	6.993	0.05	0.176	6.817
MARS-WEMA_270056	Main	7.374	7.091	0.05	0.176	7.374
MARS-WEMA_270057	Main	7.186	7.058	0.05	0.176	7.186

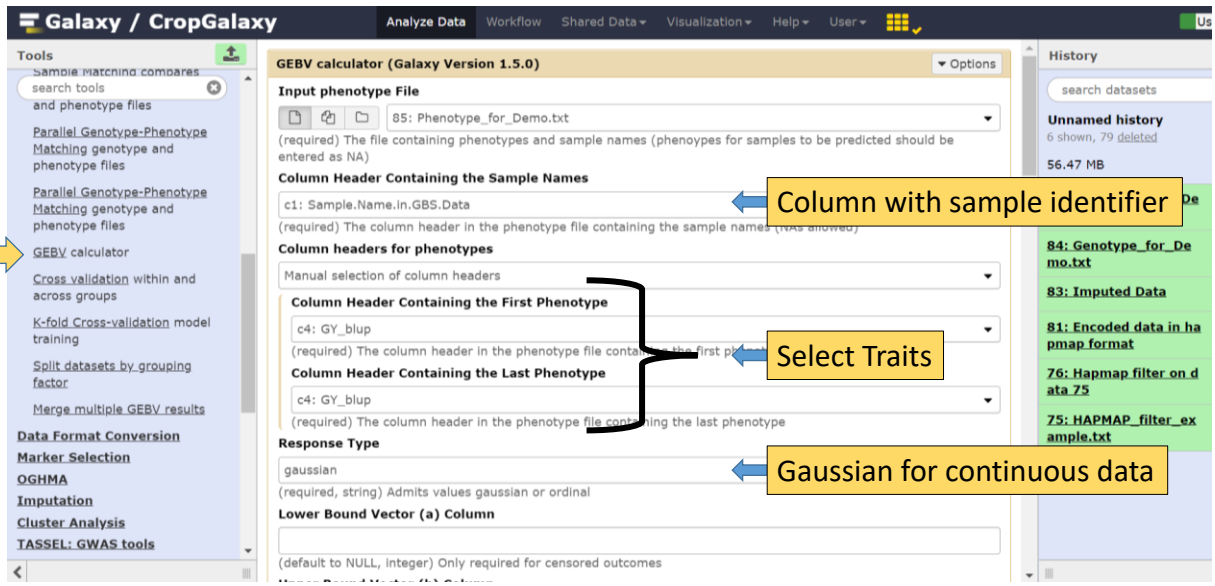
Geno_for_demo 190 markers (columns)

rs145689	rs150837	rs150103	rs150977	rs148542	rs148897	rs148817	rs145887	rs151228	rs145371
1	0	1	0	0	0	0	2	1	2
1	0	1	0	1	0	0	1	1	1
0	1	1	0	2	0	1	0	1	0
1	0	1	1	1	0	0	1	2	1
0	1	1	0	2	1	0	1	0	0
0	1	2	1	1	1	0	1	2	2
2	1	1	1	0	2	1	2	2	1
1	0	1	2	0	2	2	1	1	1
0	1	1	1	0	1	1	1	1	1
1	1	1	0	0	0	1	1	1	1
0	2	1	1	0	2	1	0	2	1
0	1	0	1	1	1	1	2	1	1
0	0	0	1	0	0	1	1	1	1
1	1	0	1	2	2	1	0	1	1
0	1	1	2	0	2	1	2	0	1
2	0	1	2	1	0	2	0	2	2
1	1	1	1	0	1	0	2	1	1
2	1	0	1	0	1	1	1	0	2
1	1	2	1	0	1	1	2	1	2

No Identifier Column

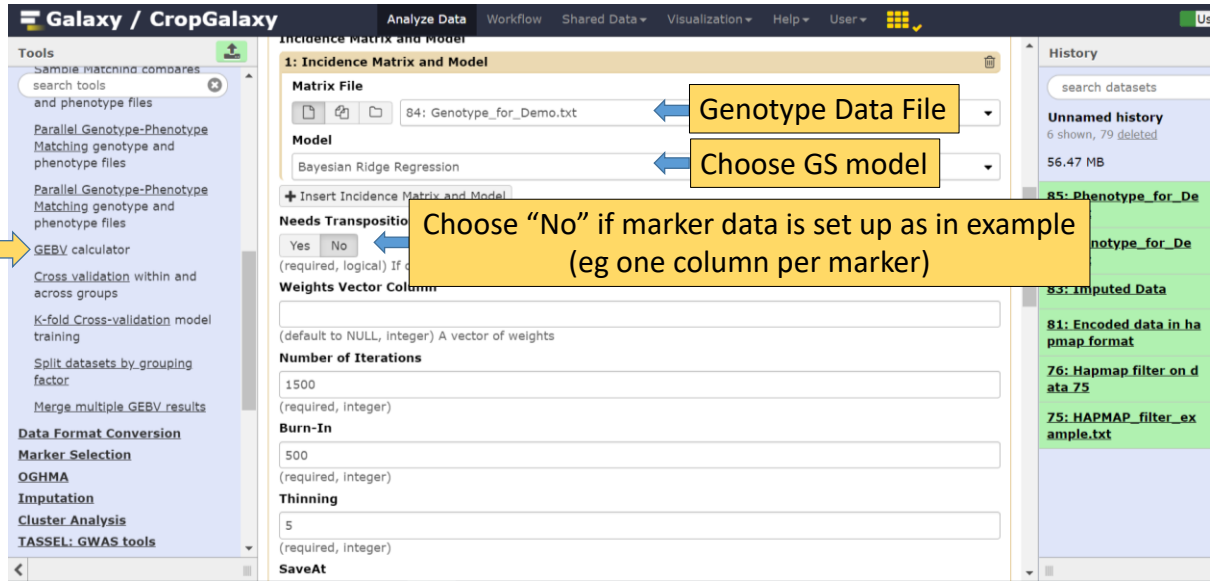
20

6. GS Accuracy via CV



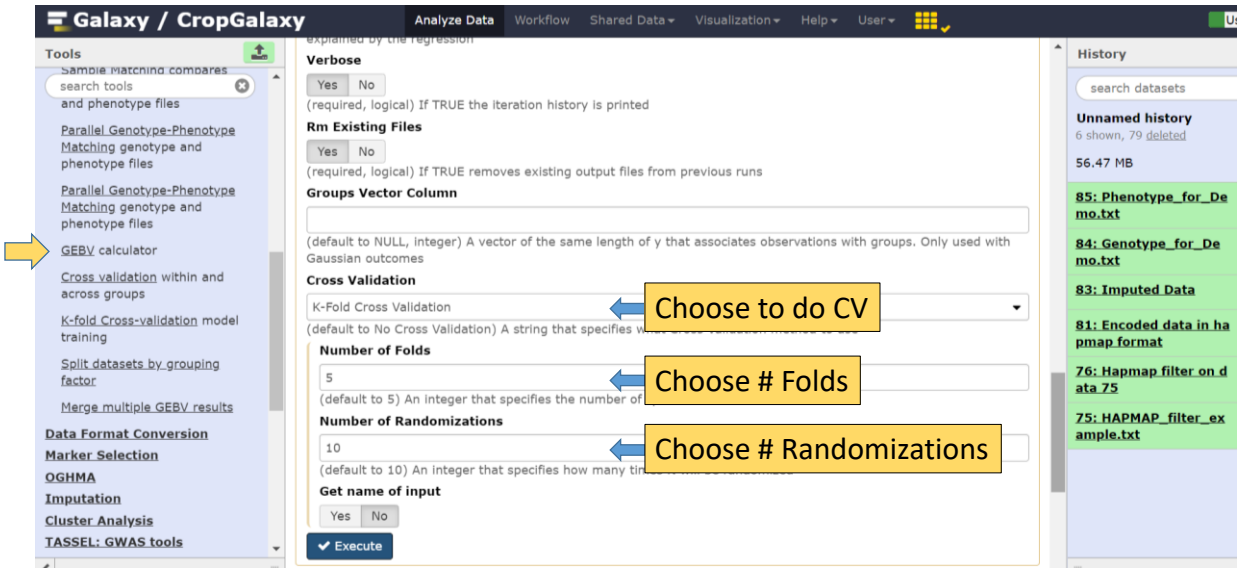
21

6. GS Accuracy via CV (cont)



22

6. GS Accuracy via CV (cont)



Output

GS Accuracy

Randomization Number	GY_blup correlation
1	0.125
2	0.189
3	0.244
4	0.19
5	0.223
6	0.203
7	0.157
8	0.157
9	0.189
10	0.169

Randomization Number	RR Blups	Bayes A	Bayes B	Bayes C	Bayes LASSO
1	0.125	0.194	0.118	0.108	0.173
2	0.189	0.188	0.109	0.123	0.174
3	0.244	0.200	0.077	0.127	0.076
4	0.190	0.191	0.144	0.145	0.137
5	0.223	0.093	0.078	0.149	0.155
6	0.203	0.214	0.137	0.212	0.176
7	0.157	0.087	0.131	0.130	0.081
8	0.157	0.132	0.142	0.189	0.184
9	0.189	0.084	0.116	0.129	0.174
10	0.169	0.208	0.145	0.168	0.210
Average	0.185	0.159	0.120	0.148	0.154
Stdev	0.035	0.054	0.025	0.033	0.044

RR Blup gives highest Average Accuracy. Select it to predict value of unphenotyped individuals

7. Predict value of unphenotyped individuals

Objective: Run best model on data from training population and genotype data for unphenotyped individuals

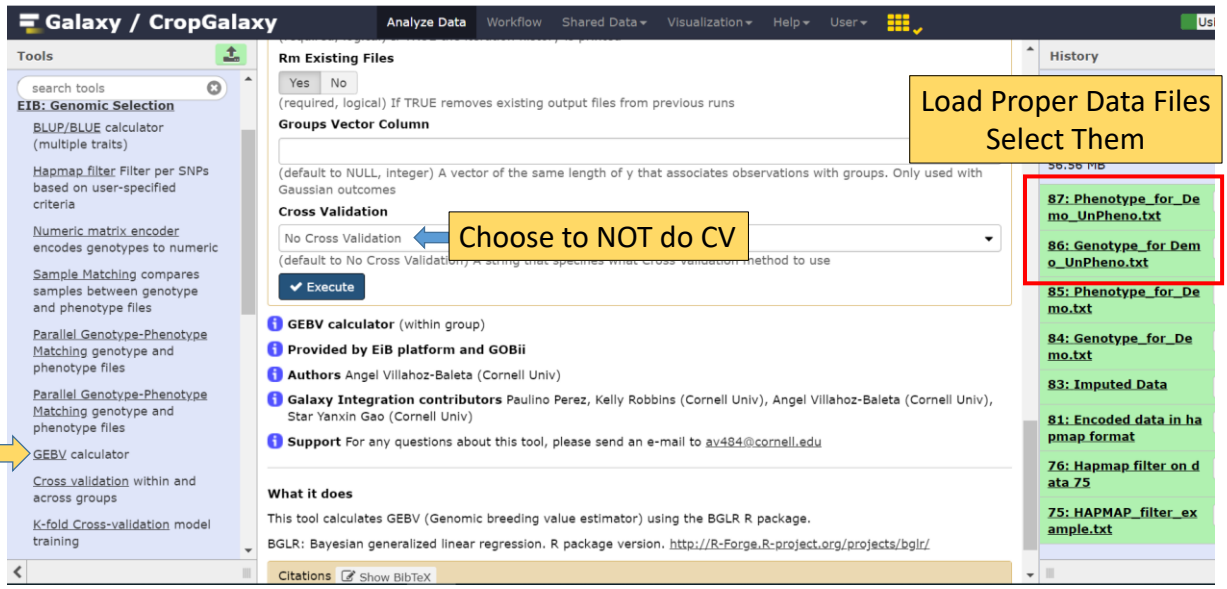
Sample.Name.in.GBS.Data	Season	GY_blue	GY_blup	GY_PEV	_pevReliabi	GY_means	rs145689	rs150837	rs150103	rs150977	rs148542	rs148897	rs148817	rs145887	rs151228	rs145371
MARS-WEMA_270039	Main	6.715	6.975	0.05	0.176	6.715	1	0	1	0	0	0	0	2	1	2
MARS-WEMA_270040	Main	5.959	6.841	0.05	0.176	5.959	1	0	1	0	1	0	0	1	1	1
MARS-WEMA_270041	Main	6.925	7.012	0.05	0.176	6.925	0	1	1	0	2	0	1	0	1	0
MARS-WEMA_270042	Main	6.925	7.127	0.05	0.176	7.576	1	0	1	1	1	0	0	1	2	1
MARS-WEMA_270043	Main	8.101	7.22	0.05	0.176	8.101	0	1	1	0	2	1	0	1	0	0
MARS-WEMA_270044	Main	7.335	7.084	0.05	0.176	7.335	0	1	2	1	1	1	0	1	2	2
MARS-WEMA_270045	Main	8.015	7.205	0.05	0.176	8.015	2	1	1	1	0	2	1	2	2	1
MARS-WEMA_270046	Main	7.234	7.066	0.05	0.176	7.234	1	0	1	2	0	2	2	1	1	1
MARS-WEMA_270047	Main	6.87	7.002	0.05					1	1	0	1	1	1	1	1
MARS-WEMA_270048	Main	7.43	7.101	0.05					1	0	0	0	1	1	1	1
MARS-WEMA_270049	Main	7.178	7.056	0.05					1	1	0	2	1	0	2	1
MARS-WEMA_270050	Main	6.555	6.946	0.05					0	1	1	1	1	2	1	1
MARS-WEMA_270051	Main	6.491	6.935	0.05	0.176	6.491	0	0	0	1	0	0	1	1	1	1
MARS-WEMA_270052	Main	6.811	6.992	0.05	0.176	6.811	1	1	0	1	2	2	1	0	1	1
MARS-WEMA_270053	Main	7.82	7.17	0.05	0.176	7.82	0	1	1	2	0	2	1	2	0	1
MARS-WEMA_270054	Main	6.384	6.916	0.05	0.176	6.384	2	0	1	2	1	0	2	0	2	2
MARS-WEMA_270055	Main	6.384	6.993	0.05	0.176	6.817	1	1	1	1	0	1	0	2	1	1
MARS-WEMA_270056	Main	7.374	7.091	0.05	0.176	7.374	2	1	0	1	0	1	1	1	0	2
MARS-WEMA_270057	Main	7.186	7.058	0.05	0.176	7.186	1	1	2	1	0	1	1	2	1	2
UnPheno 1	NA	NA	NA	NA	NA	NA	0	0	1	0	1	0	2	2	1	1
UnPheno 2	NA	NA	NA	NA	NA	NA	2	0	0	1	1	1	2	2	0	1
UnPheno 3	NA	NA	NA	NA	NA	NA	0	1	0	0	0	1	1	2	0	2
UnPheno 4	NA	NA	NA	NA	NA	NA	1	0	0	1	2	0	1	2	0	1
UnPheno 5	NA	NA	NA	NA	NA	NA	1	1	2	1	1	1	0	2	2	1

Training Population Data

Add phenotypic and genotypic data from UnPhenotyped Individuals whose value you want to predict

25

7. Predict value of unphenotyped individuals



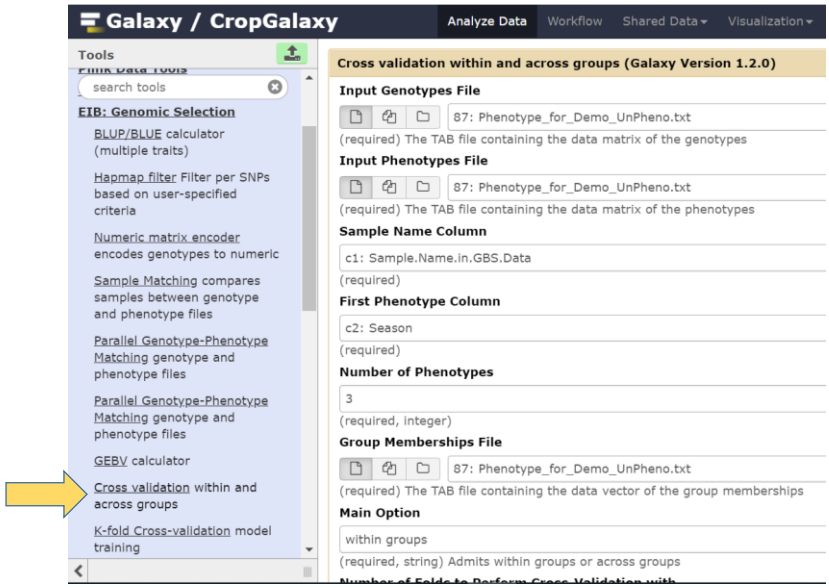
26

Output

Sample.Name	GY_blue_Observed	GY_blue_Predicted
MARS-WEMA_270039	6.72	6.95
MARS-WEMA_270040	5.96	6.77
MARS-WEMA_270041	6.93	7.06
MARS-WEMA_270042	6.93	6.91
…	…	…
MARS-WEMA_293000	7.28	7.05
MARS-WEMA_293001	6.33	6.83
MARS-WEMA_293002	7.28	7.11
MARS-WEMA_293003	6.60	6.41
MARS-WEMA_293004	6.54	7.24
UnPheno 1	NA	7.05
UnPheno 2	NA	6.83
UnPheno 3	NA	7.11
UnPheno 4	NA	6.41
UnPheno 5	NA	7.24

27

Predictions with and across groups of data



28

Example: Same set of lines, phenotyped in three environments

GID	GY_ENV_1	GY_ENV_2	GY_ENV_3
49319	7.7	8.5	3.5
49446	NA	8.5	7.2
49359	7.0	8.7	4.2
49322	7.3	NA	3.8
49413	7.4	8.8	4.6
49339	6.8	8.9	3.4
49315	7.3	8.9	3.9
49397	7.4	9.0	4.2
49444	7.8	9.0	4.9
49340	6.9	9.0	NA

GID	GROUP	GY
49319	1	7.7
49446	1	NA
49359	1	7.0
49322	1	7.3
49413	1	7.4
49339	1	6.8
49315	1	7.3
49397	1	7.4
49444	1	7.8
49340	1	6.9
49319	2	8.5
49446	2	8.5
49359	2	8.7
49322	2	NA
49413	2	8.8
49339	2	8.9
49315	2	8.9
49397	2	9.0
49444	2	9.0
49340	2	9.0
49319	3	3.5
49446	3	7.2
49359	3	4.2
49322	3	3.8
49413	3	4.6
49339	3	3.4
49315	3	3.9
49397	3	4.2
49444	3	4.9
49340	3	NA

CV within and across groups

CV within groups

Use G1 data to predict value of G1 lines
Use G2 data to predict value of G2 lines
Use G3 data to predict value of G3 lines

CV across groups

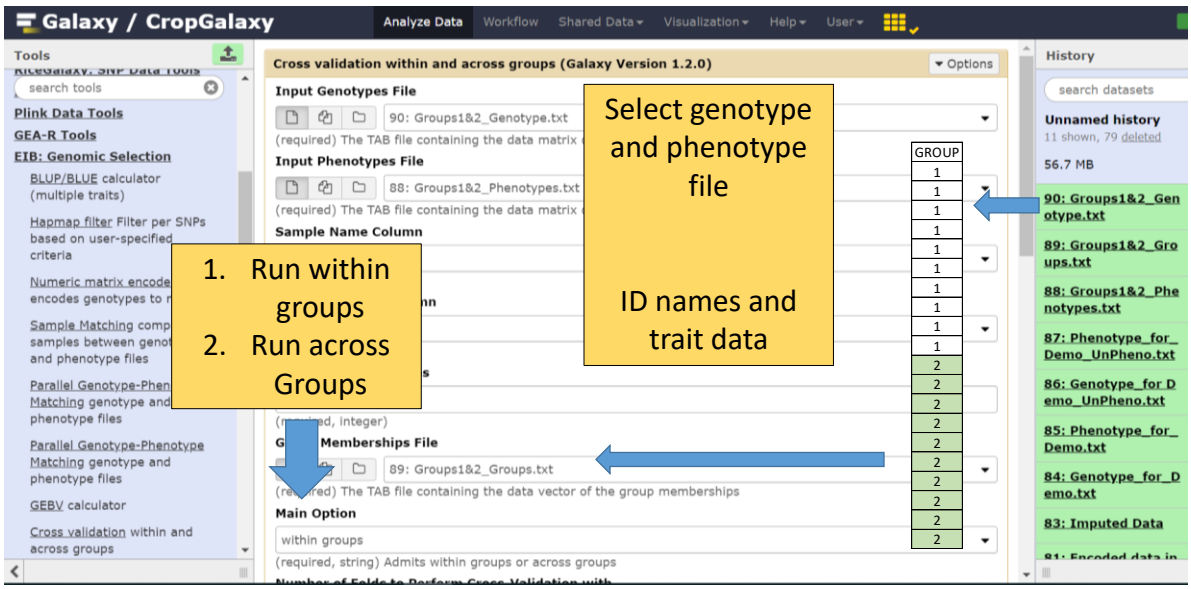
Use G2&3 data to predict value of G1 lines
Use G1&3 data to predict value of G2 lines
Use G1&2 data to predict value of G3 lines

29

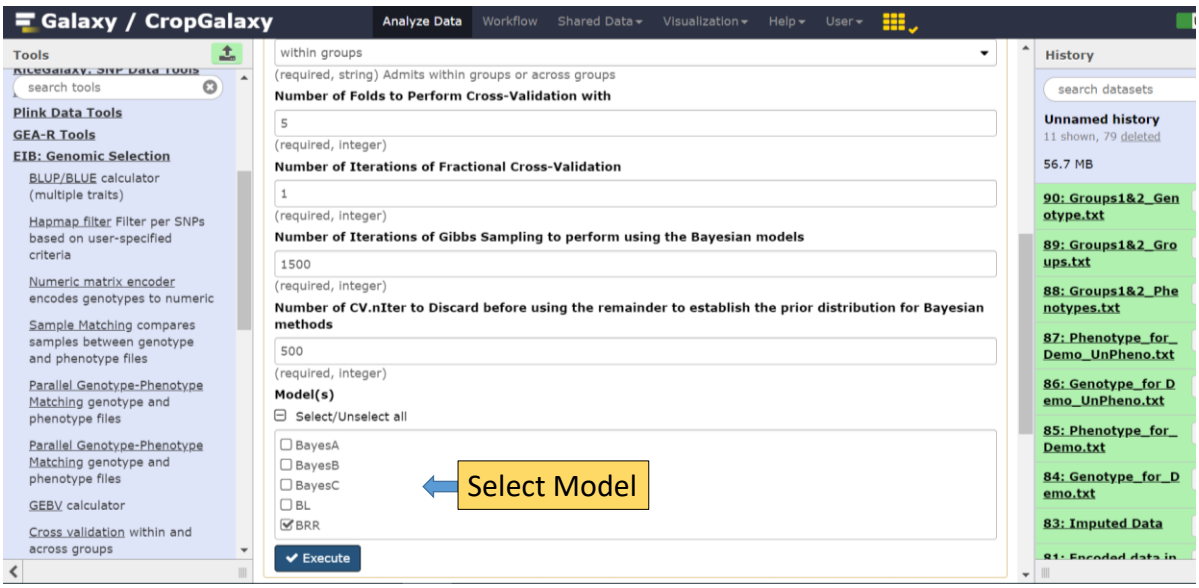
Example: Three sets of lines, phenotyped in same environment

GID	GROUP	GY
A1	1	7.7
A2	1	NA
A3	1	7.0
A4	1	7.3
A5	1	7.4
A6	1	6.8
A7	1	7.3
A8	1	7.4
A9	1	7.8
A10	1	6.9
B1	2	8.5
B2	2	8.5
B3	2	8.7
B4	2	NA
B5	2	8.8
B6	2	8.9
B7	2	8.9
B8	2	9.0
B9	2	9.0
B10	2	9.0
C1	3	3.5
C2	3	7.2
C3	3	4.2
C4	3	3.8
C5	3	4.6
C6	3	3.4
C7	3	3.9
C8	3	4.2
C9	3	4.9
C10	3	NA

30


Galaxy / CropGalaxy
Analyze Data
Workflow
Shared Data
Visualization
Help
User
Tools
search tools
Plink Data Tools
GEA-R Tools
EIB: Genomic Selection
BLUP/BLUE calculator (multiple traits)
Hapmap filter Filter per SNPs based on user-specified criteria
Numeric matrix encoder encodes genotypes to numeric
Sample Matching compares samples between genotype and phenotype files
Parallel Genotype-Phenotype Matching genotype and phenotype files
Parallel Genotype-Phenotype Matching genotype and phenotype files
GEBV calculator
Cross validation within and across groups
Cross validation within and across groups (Galaxy Version 1.2.0)
Options
Input Genotypes File
90: Groups1&2_Genotype.txt
(required) The TAB file containing the data matrix
Input Phenotypes File
88: Groups1&2_Phenotypes.txt
(required) The TAB file containing the data matrix
Sample Name Column
Memberships File
89: Groups1&2_Groups.txt
(required) The TAB file containing the data vector of the group memberships
Main Option
within groups
(required, string) Admits within groups or across groups
GROUP
1
1
1
1
1
1
1
1
1
1
2
2
2
2
2
2
2
2
2
2
Select genotype and phenotype file
ID names and trait data
1. Run within groups
2. Run across Groups
History
search datasets
Unnamed history
11 shown, 79 deleted
56.7 MB
90: Groups1&2_Genotype.txt
89: Groups1&2_Groups.txt
88: Groups1&2_Phenotypes.txt
87: Phenotype_for_Demo_UnPheno.txt
86: Genotype_for Demo_UnPheno.txt
85: Phenotype_for_Demo.txt
84: Genotype_for_Demo.txt
83: Imputed Data

31

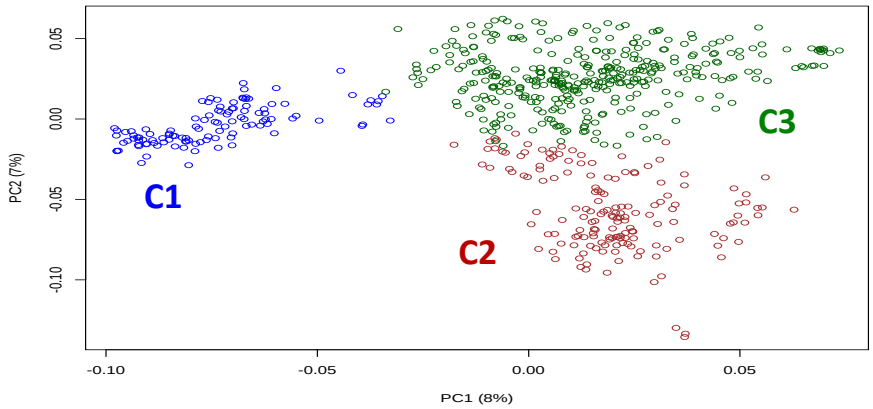

Galaxy / CropGalaxy
Analyze Data
Workflow
Shared Data
Visualization
Help
User
Tools
search tools
Plink Data Tools
GEA-R Tools
EIB: Genomic Selection
BLUP/BLUE calculator (multiple traits)
Hapmap filter Filter per SNPs based on user-specified criteria
Numeric matrix encoder encodes genotypes to numeric
Sample Matching compares samples between genotype and phenotype files
Parallel Genotype-Phenotype Matching genotype and phenotype files
Parallel Genotype-Phenotype Matching genotype and phenotype files
GEBV calculator
Cross validation within and across groups
within groups
(required, string) Admits within groups or across groups
Number of Folds to Perform Cross-Validation with
5
(required, integer)
Number of Iterations of Fractional Cross-Validation
1
(required, integer)
Number of Iterations of Gibbs Sampling to perform using the Bayesian models
1500
(required, integer)
Number of CV.nIter to Discard before using the remainder to establish the prior distribution for Bayesian methods
500
(required, integer)
Model(s)
Select/Unselect all
BayesA
BayesB
BayesC
BL
BRR
Select Model
Execute
History
search datasets
Unnamed history
11 shown, 79 deleted
56.7 MB
90: Groups1&2_Genotype.txt
89: Groups1&2_Groups.txt
88: Groups1&2_Phenotypes.txt
87: Phenotype_for_Demo_UnPheno.txt
86: Genotype_for Demo_UnPheno.txt
85: Phenotype_for_Demo.txt
84: Genotype_for_Demo.txt
83: Imputed Data

32

Your Practicum Files and Objectives

576 Winter Wheat Lines
Genotyped with 8,703 GBS SNPs
All Pheno. for 5 Fusarium Head Blight Traits

346 Winter Wheat Lines
1,451 GBS SNPs
307 Pheno. 39 need to be predicted



33

- Incidence (INC) : Percentage of spikes showing symptoms
- Severity (SEV): % of florets showing symptoms within an infected spike
- *Index (IND):* % of symptomatic florets in the plot (IND=INC*SEV/100)
- Fusarium Damaged Kernels (FDK): % of kernels showing symptoms
- Deoxynivalenol (DON): concentration of DON in grain

SEV
25% 50% 65%

INC 5% 95%
SEV 5% 80%
IND 0.3% 75%

34

GS_Demo_FHB_**BLUP** (307 rows)
GS_Demo_FHB_**BLUP_With_NA** (346 rows)

NAME	GROUP	IND	FDK	ISK	DON	PC1
IL95-4162	3	-0.16	-0.26	-0.23	-0.09	-0.40
IL02-18228	3	-0.65	-0.86	-0.91	-0.74	-1.76
PEMBROKE	3	-0.05	0.30	0.28	0.11	0.30
MO080104	3	-0.56	-0.58	-0.62	-0.66	-1.39
IL79-002DH	3	-0.40	-0.52	-0.51	-0.62	-1.21
TRUMAN	2	-0.50	-0.48	-0.54	-0.55	-1.18
0128A1--2	3	-0.15	0.55	0.32	-0.15	0.23
JENSEN	1	0.74	1.03	0.91	0.70	1.92
OH10-143-23	3	0.08	0.94	0.39	0.34	1.11

2 Data Files to Upload
- 2 for Phenotypes
- 2 for Genotypes
- 1 for grouping

GS_Demo_FHB_**GENO** (307 rows)
GS_Demo_FHB_**GENO_With_NA** (346 rows)

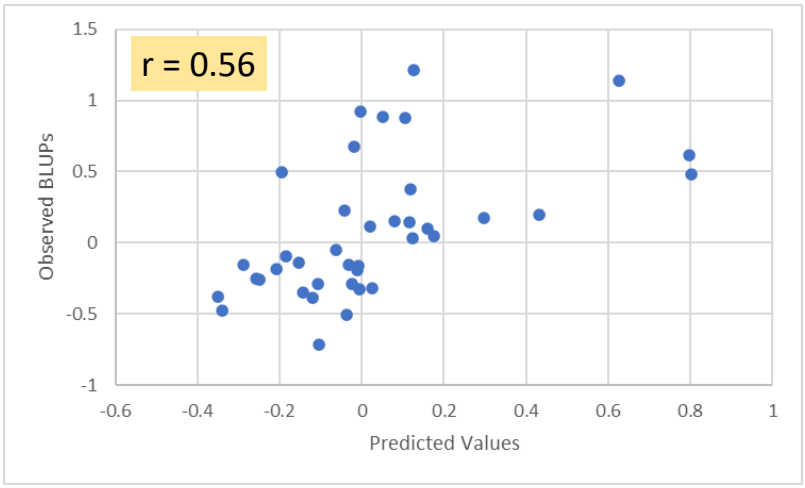
TPM59	TPM383	TPM1663	TPM2559	TPM3009	TPM3622
0	0	0	0	0	1
0	1	0	1	2	1
0	0	0	0	0	0
0	0	1	0	0	0
0	0	0	0	2	0
0	1	0	1	0	1
1	0	1	1	2	1
0	0	0	0	0	1
0	0	0	1	2	1

GS_Demo_FHB_**Group**
(346 rows)

GROUP
3
3
3
3
3
2
3
1
3

35

Results for IND



36