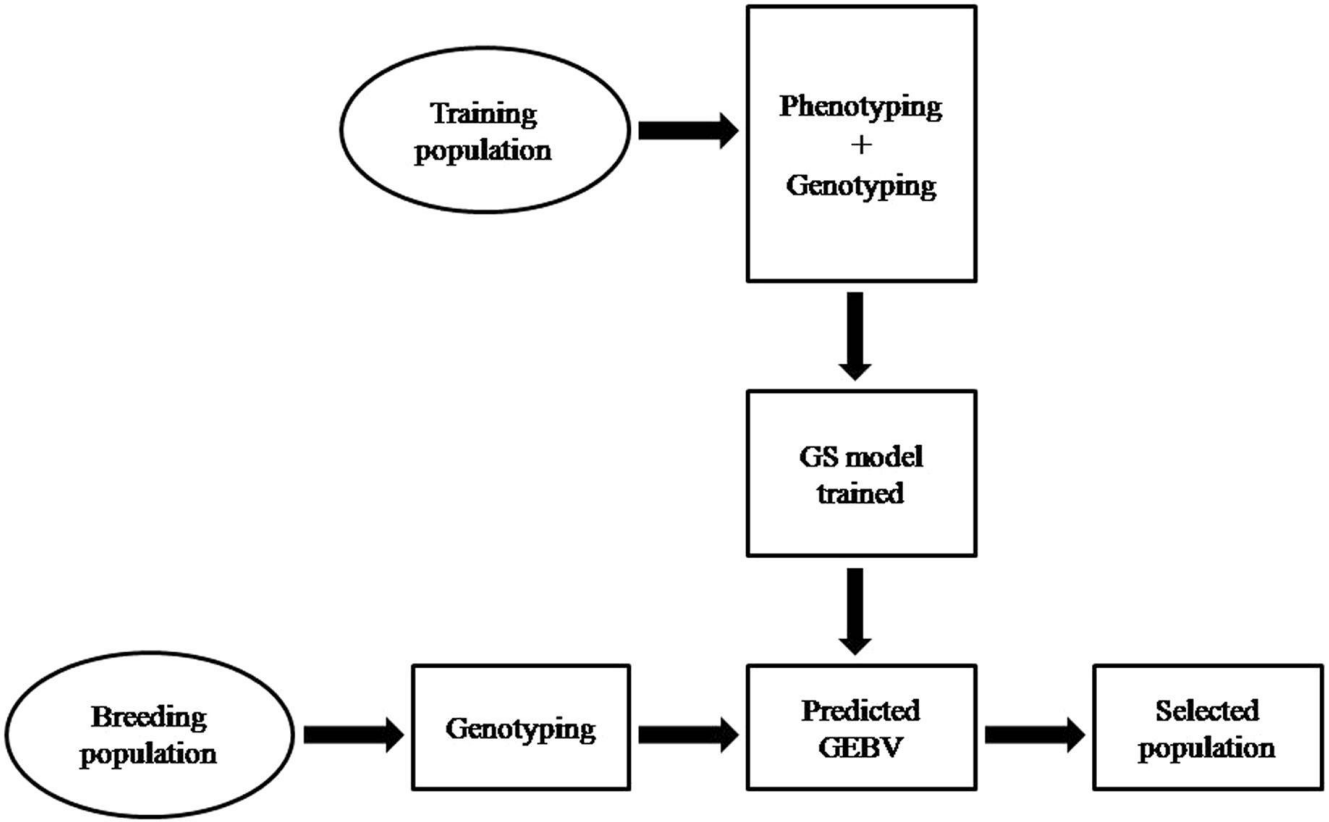
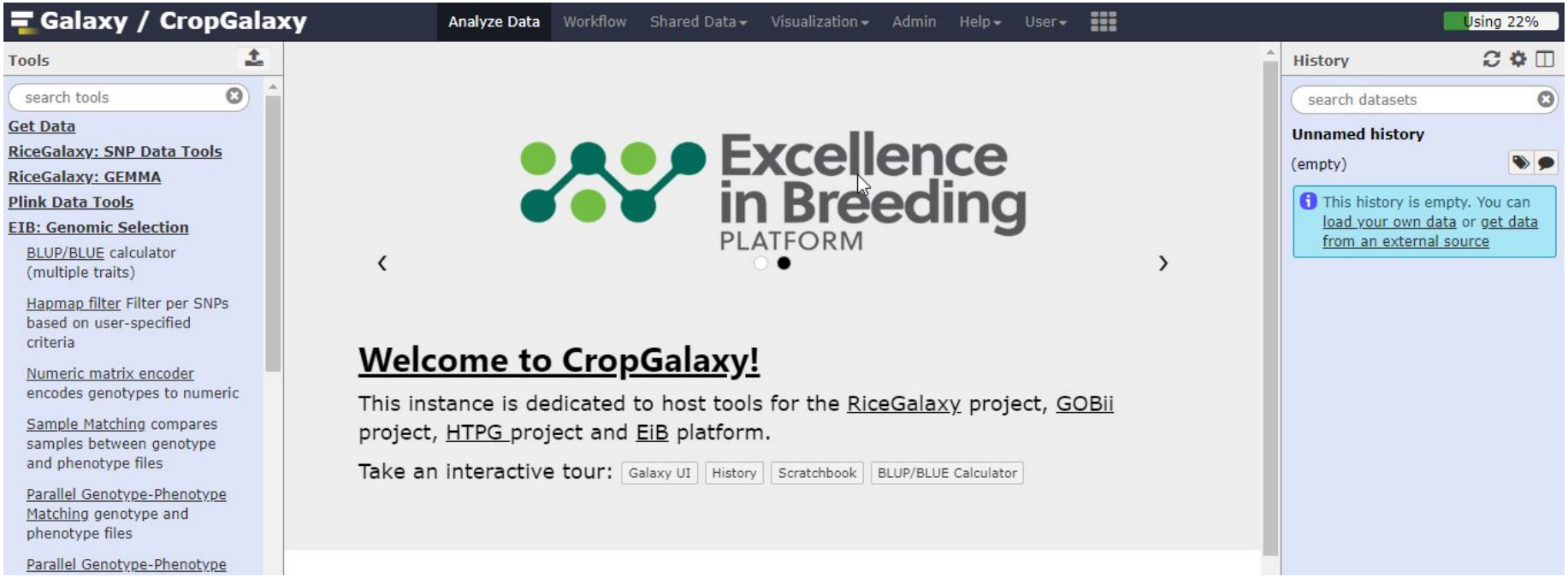
Galaxy for Crops: Genomic Selection

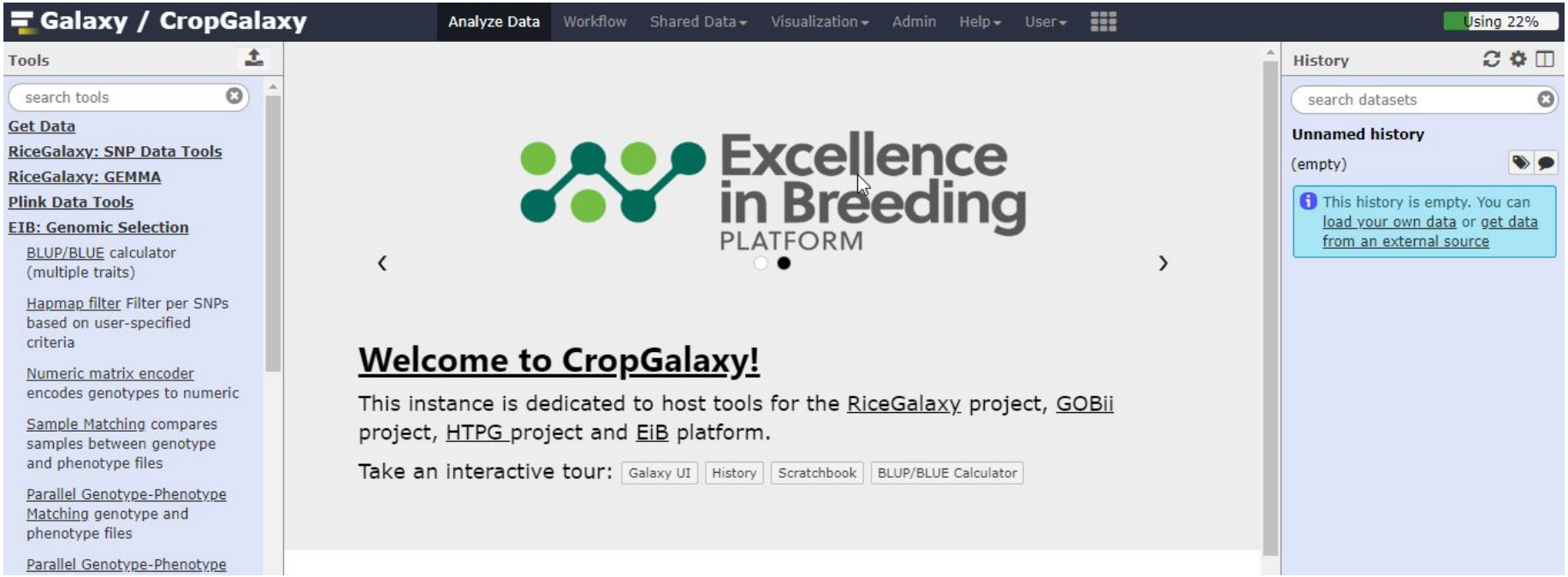
Venice Margarette Juanillas
Specialist- Information Systems



Source: *"Genomic Selection in the Era of Next Generation Sequencing for Complex Traits in Plant Breeding" (Bhat, J.A, et al, 2016)*



http://13.250.245.3/

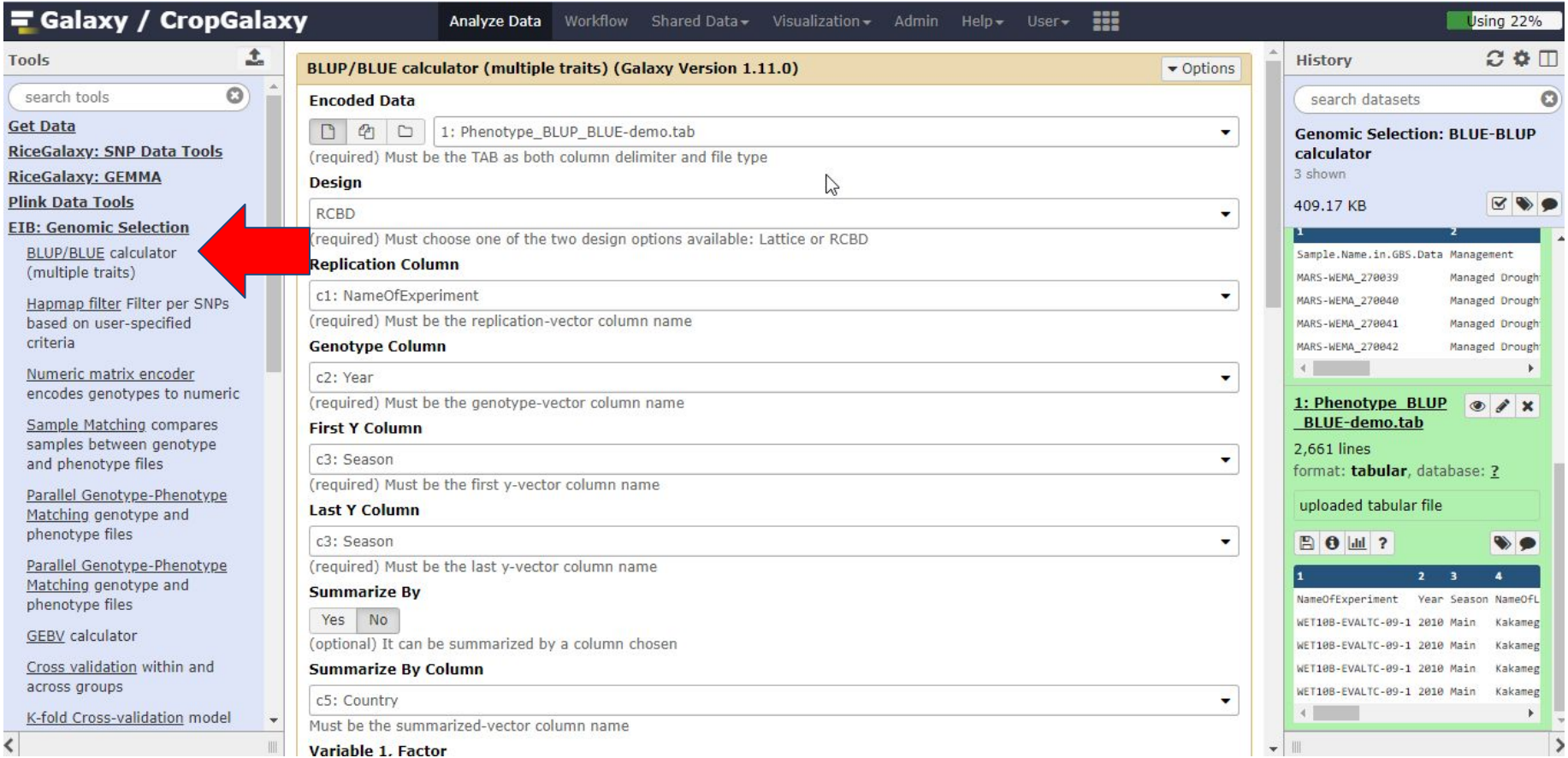


soon: **https://cropgalaxy.excellenceinbreeding.org**

IRRI

A. BLUE/BLUP Calculator

- Best Linear Unbiased Estimates (BLUE)
- Best Linear Unbiased Prediction (BLUP)
- Estimation of random effects

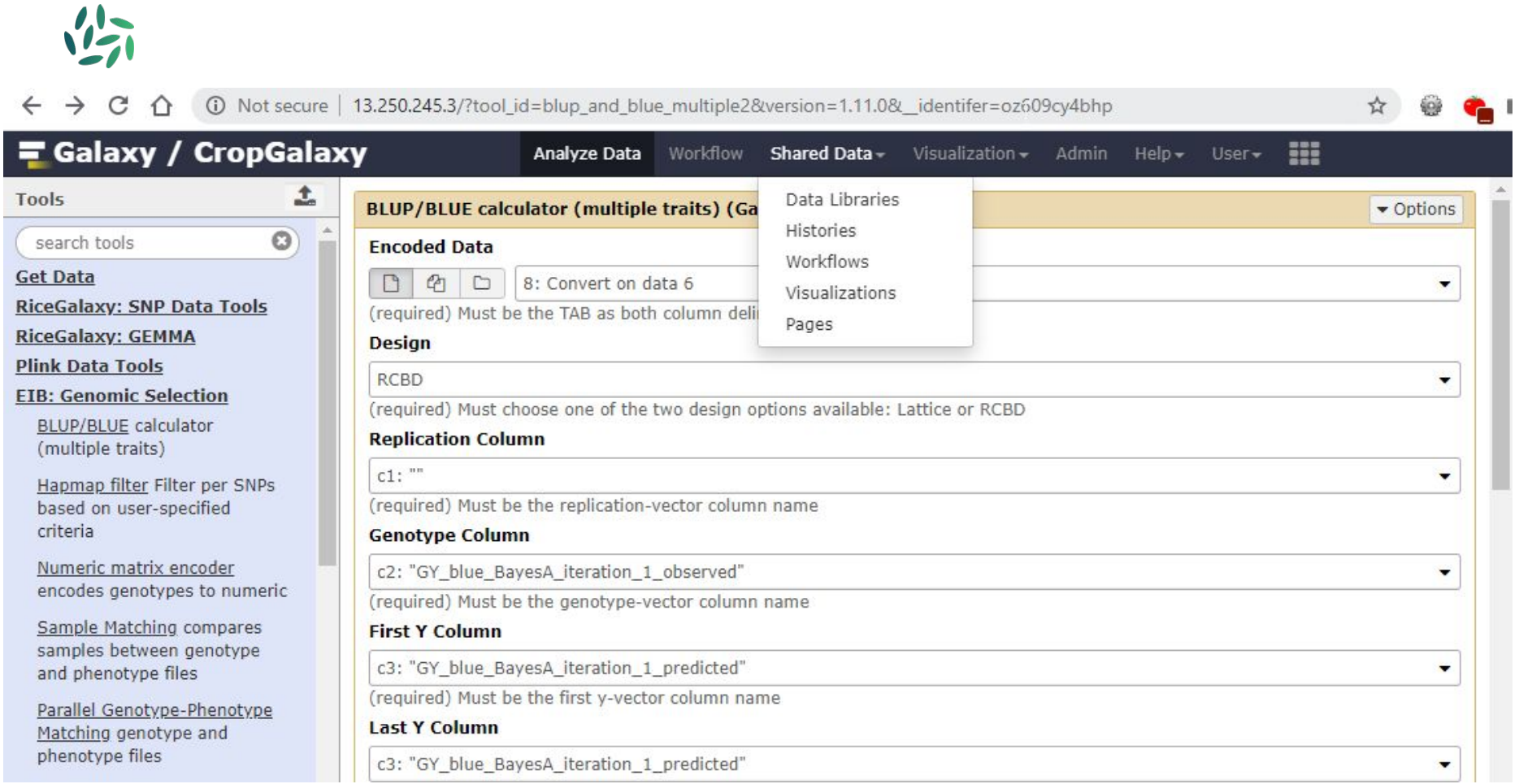


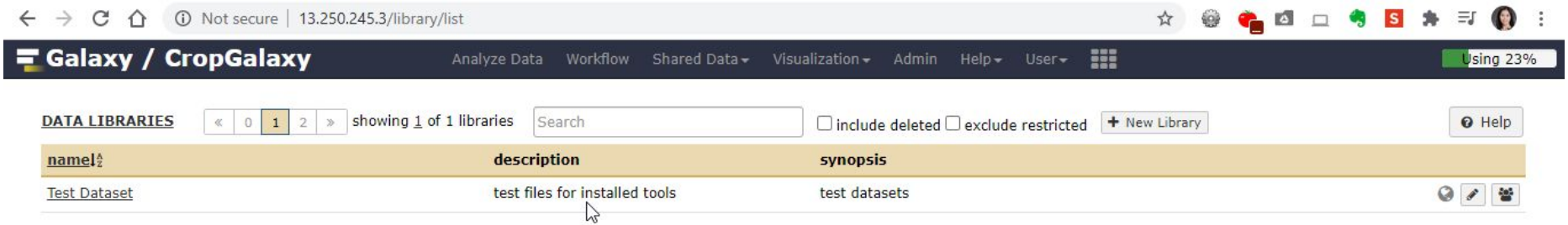
BLUE/BLUP calculator

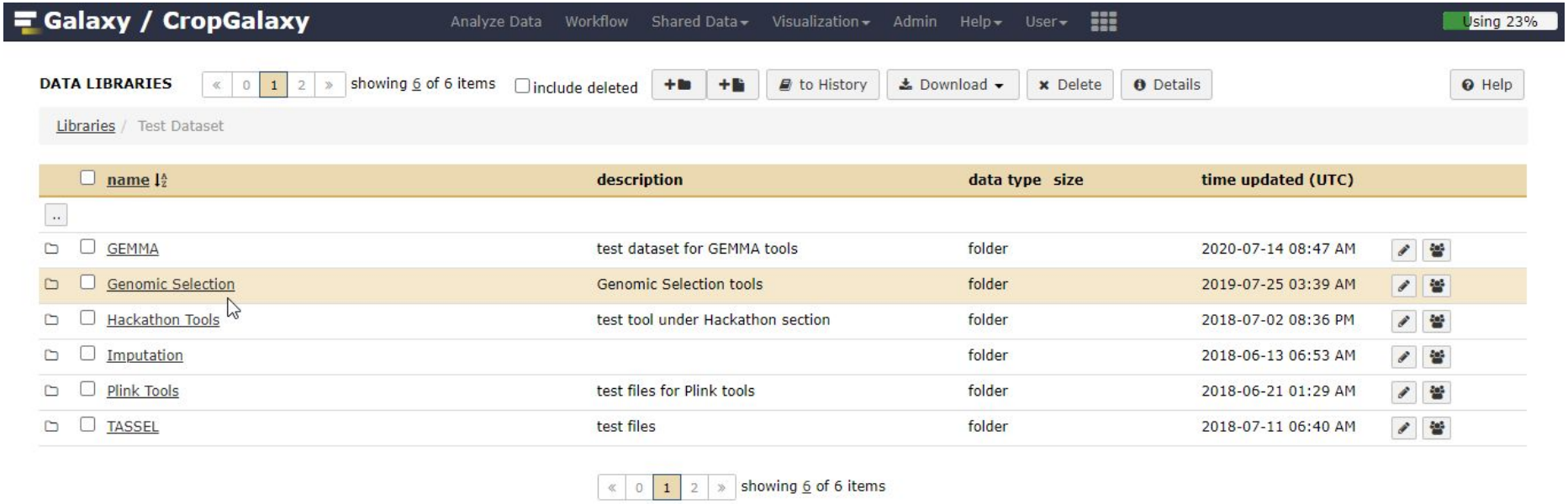
IRRI

Hands-on:

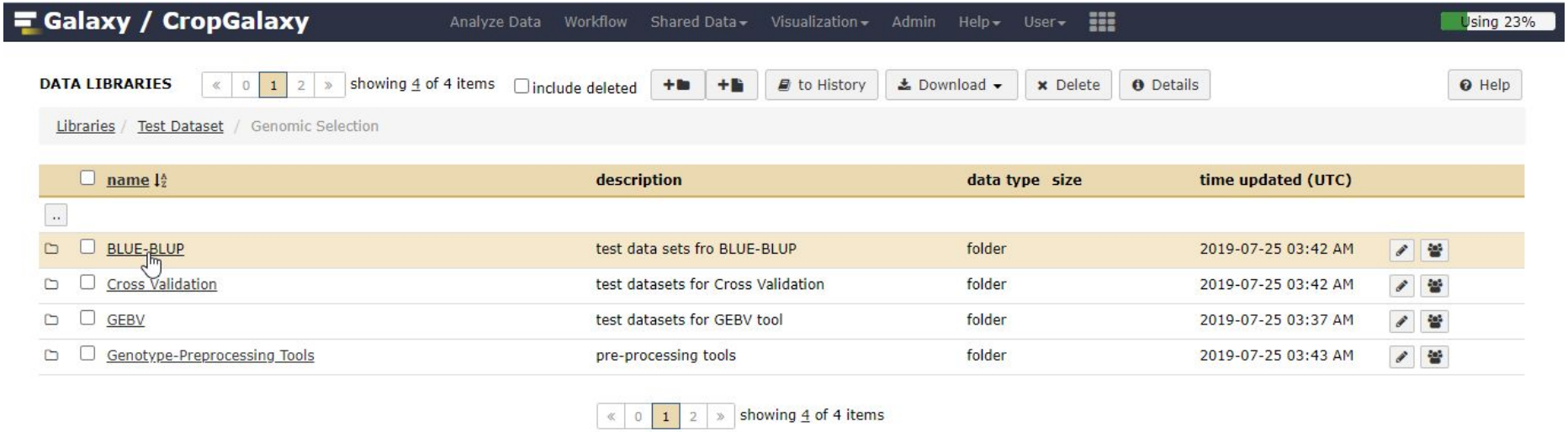
1. Dataset:
    a. Shared Data-> Data Libraries-> Test Dataset-> Genomic Selection-> BLUE/BLUP
    b. Select “Phenotype_BLUP_BLUE-demo.tab”
    c. Import to History.
    d. Click Green pop-up to start analyzing it.

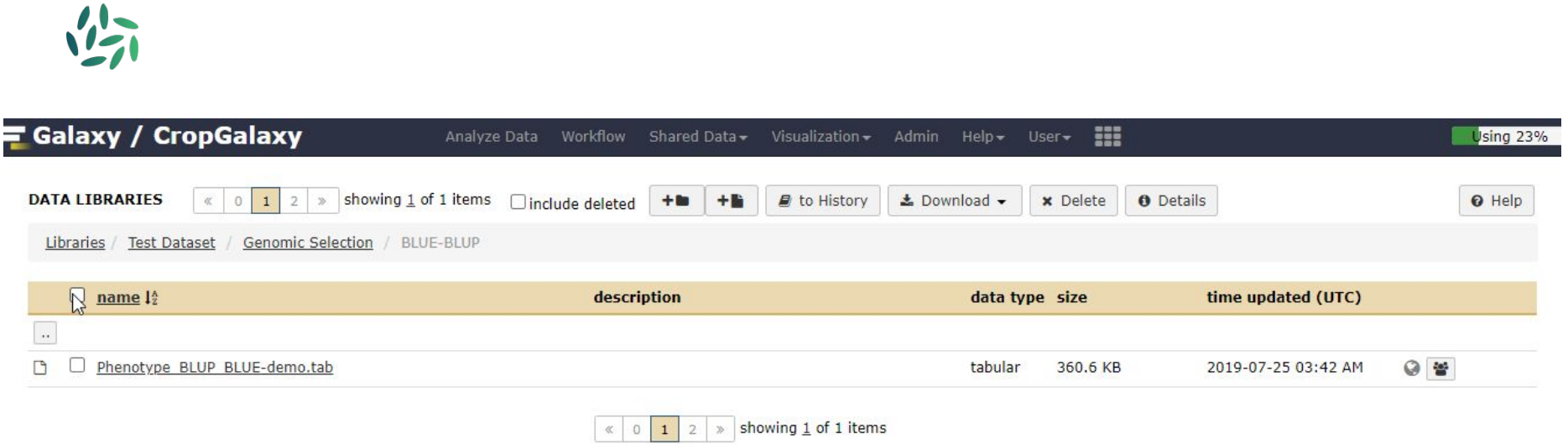

Not secure | 13.250.245.3/?tool_id=blup_and_blue_multiple2&version=1.11.0&__identifer=oz609cy4bhp
Galaxy / CropGalaxy
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Data Libraries
Histories
Workflows
Visualizations
Pages
Tools
search tools
Get Data
RiceGalaxy: SNP Data Tools
RiceGalaxy: GEMMA
Plink Data 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
BLUP/BLUE calculator (multiple traits) (Ga
Options
Encoded Data
8: Convert on data 6
(required) Must be the TAB as both column deli
Design
RCBD
(required) Must choose one of the two design options available: Lattice or RCBD
Replication Column
c1: ""
(required) Must be the replication-vector column name
Genotype Column
c2: "GY_blue_BayesA_iteration_1_observed"
(required) Must be the genotype-vector column name
First Y Column
c3: "GY_blue_BayesA_iteration_1_predicted"
(required) Must be the first y-vector column name
Last Y Column
c3: "GY_blue_BayesA_iteration_1_predicted"

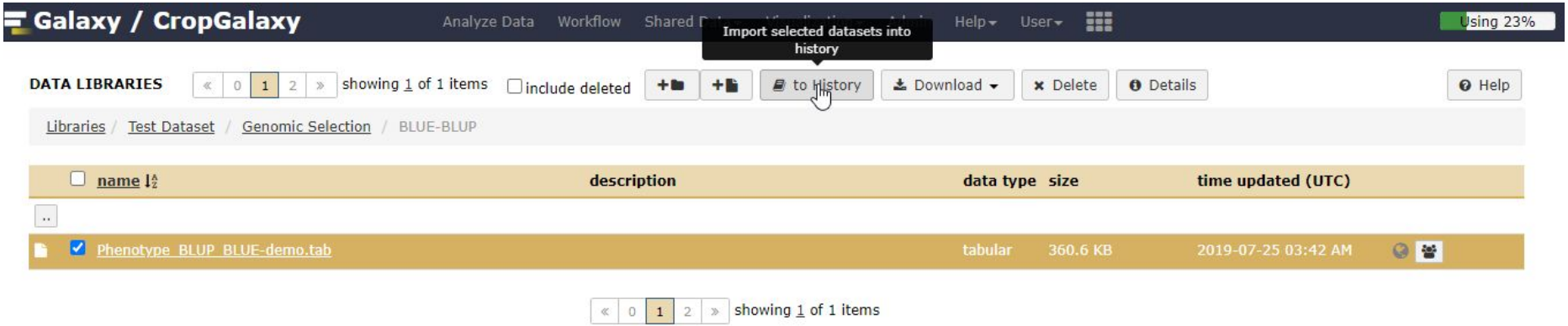

Not secure | 13.250.245.3/library/list
Galaxy / CropGalaxy
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Using 23%
DATA LIBRARIES
« 0 1 2 »
showing 1 of 1 libraries
Search
include deleted
exclude restricted
+ New Library
Help
name
description
synopsis
Test Dataset
test files for installed tools
test datasets

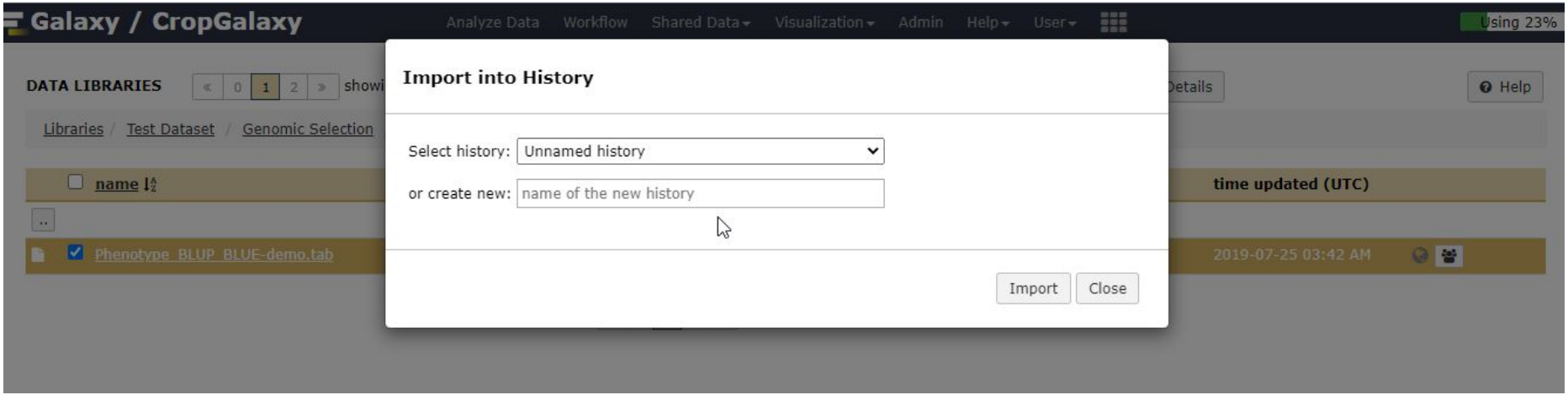

Galaxy / CropGalaxy
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Using 23%
DATA LIBRARIES
« 0 1 2 » showing 6 of 6 items
include deleted
to History
Download
Delete
Details
Help
Libraries / Test Dataset
name
description
data type
size
time updated (UTC)
..
GEMMA
test dataset for GEMMA tools
folder
2020-07-14 08:47 AM
Genomic Selection
Genomic Selection tools
folder
2019-07-25 03:39 AM
Hackathon Tools
test tool under Hackathon section
folder
2018-07-02 08:36 PM
Imputation
folder
2018-06-13 06:53 AM
Plink Tools
test files for Plink tools
folder
2018-06-21 01:29 AM
TASSEL
test files
folder
2018-07-11 06:40 AM
« 0 1 2 » showing 6 of 6 items

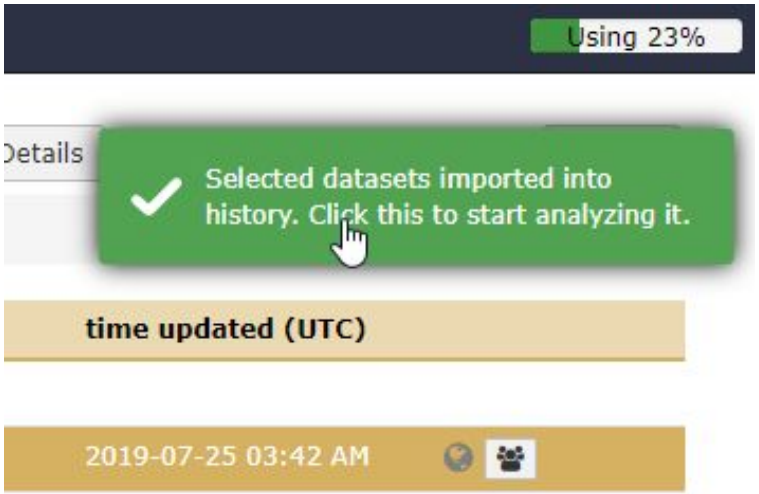
IRRI

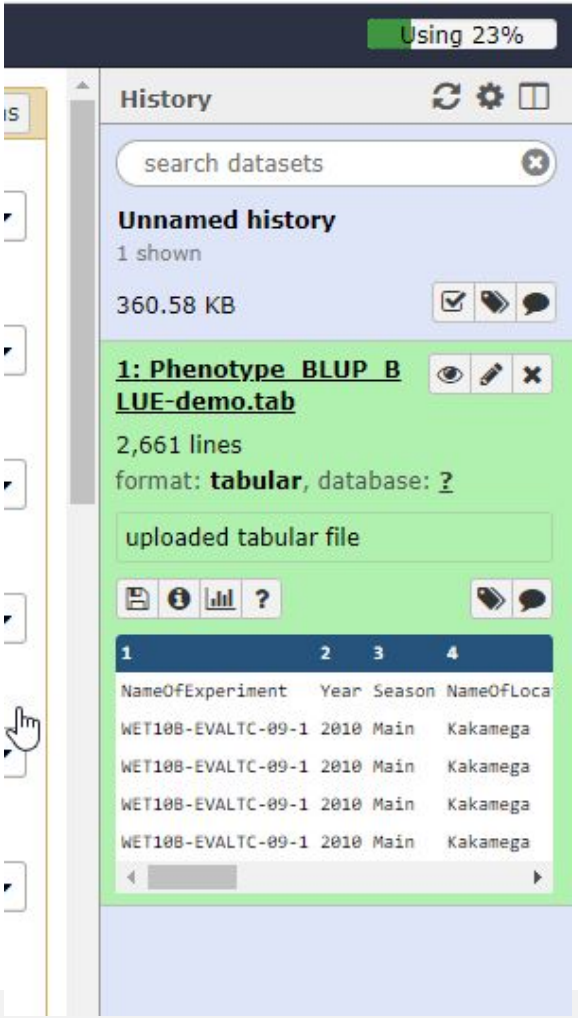

Galaxy / CropGalaxy
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Using 23%
DATA LIBRARIES
« 0 1 2 » showing 4 of 4 items
include deleted
to History
Download
Delete
Details
Help
Libraries / Test Dataset / Genomic Selection
name
description
data type
size
time updated (UTC)
..
BLUE-BLUP
test data sets fro BLUE-BLUP
folder
2019-07-25 03:42 AM
Cross Validation
test datasets for Cross Validation
folder
2019-07-25 03:42 AM
GEBV
test datasets for GEBV tool
folder
2019-07-25 03:37 AM
Genotype-Preprocessing Tools
pre-processing tools
folder
2019-07-25 03:43 AM
« 0 1 2 » showing 4 of 4 items


Galaxy / CropGalaxy
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Using 23%
DATA LIBRARIES
« 0 1 2 » showing 1 of 1 items
include deleted
to History
Download
Delete
Details
Help
Libraries / Test Dataset / Genomic Selection / BLUE-BLUP
name
description
data type
size
time updated (UTC)
..
Phenotype_BLUP_BLUE-demo.tab
tabular
360.6 KB
2019-07-25 03:42 AM
« 0 1 2 » showing 1 of 1 items


Galaxy / CropGalaxy
Analyze Data
Workflow
Help
User
Using 23%
Import selected datasets into history
DATA LIBRARIES
showing 1 of 1 items
include deleted
to History
Download
Delete
Details
Help
Libraries / Test Dataset / Genomic Selection / BLUE-BLUP
name
description
data type
size
time updated (UTC)
Phenotype_BLUP_BLUE-demo.tab
tabular
360.6 KB
2019-07-25 03:42 AM
showing 1 of 1 items


Galaxy / CropGalaxy
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Using 23%
DATA LIBRARIES
Libraries / Test Dataset / Genomic Selection
Help
name
time updated (UTC)
Phenotype_BLUP_BLUE-demo.tab
2019-07-25 03:42 AM
Import into History
Select history:
Unnamed history
or create new:
name of the new history
Import
Close

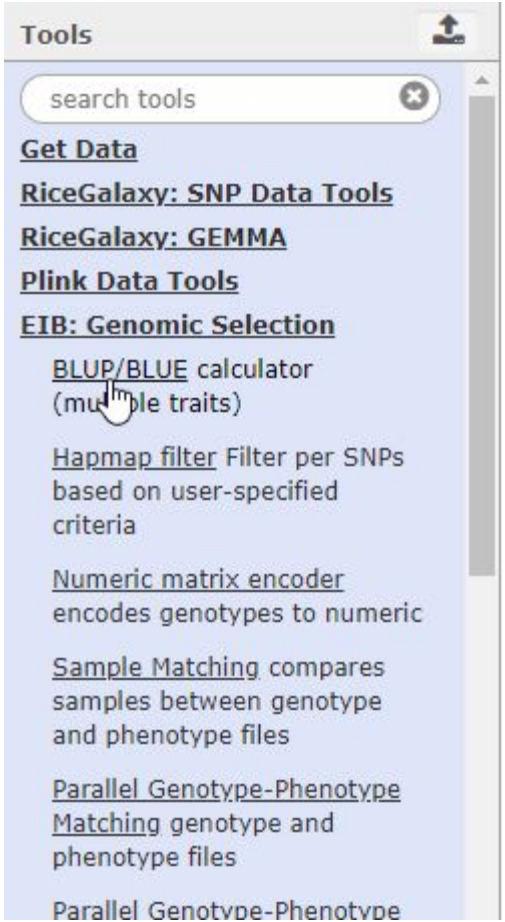

Using 23%
Details
Selected datasets imported into history. Click this to start analyzing it.
time updated (UTC)
2019-07-25 03:42 AM


Using 23%
History
search datasets
Unnamed history
1 shown
360.58 KB
1: Phenotype_BLUP_BLUE-demo.tab
2,661 lines
format: tabular, database: ?
uploaded tabular file
1 2 3 4
NameOfExperiment Year Season NameOfLoca
WET10B-EVALTC-09-1 2010 Main Kakamega
WET10B-EVALTC-09-1 2010 Main Kakamega
WET10B-EVALTC-09-1 2010 Main Kakamega
WET10B-EVALTC-09-1 2010 Main Kakamega

Hands-on:

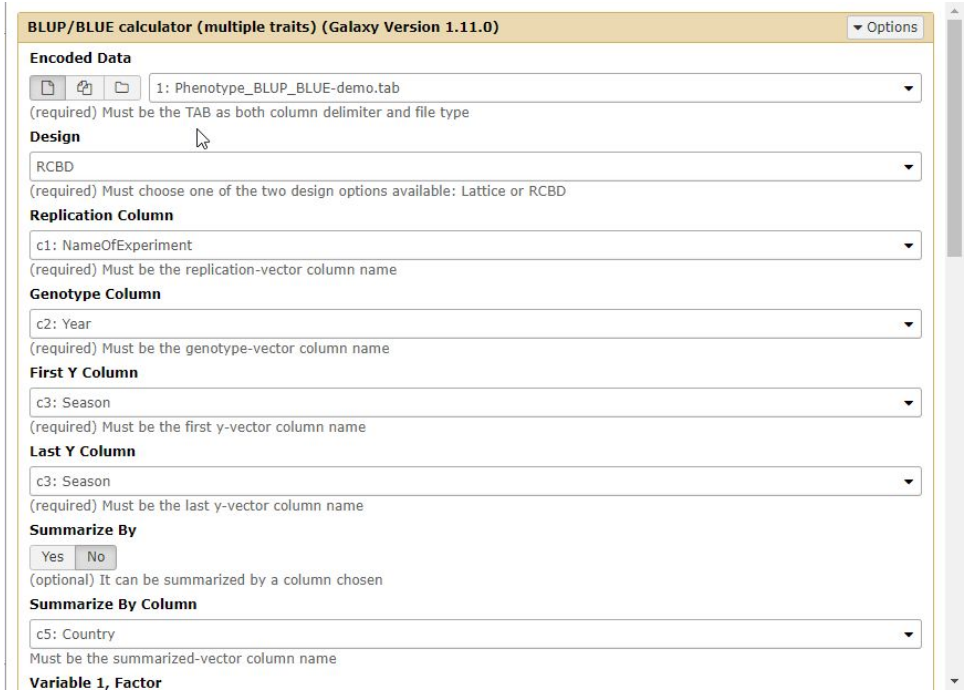
2. Tool:

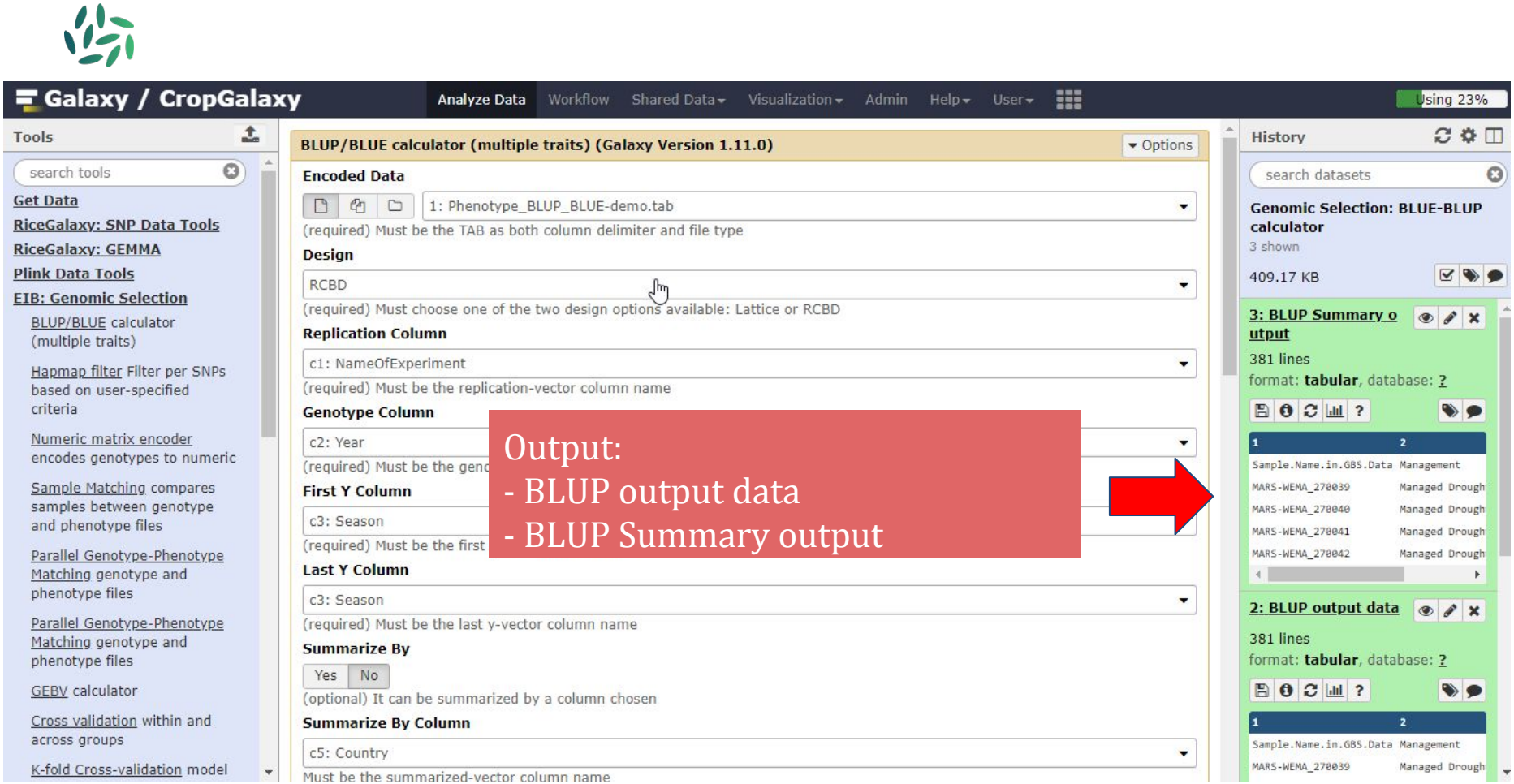
a. Go to Tools-> EIB: Genomic Selection->BLUE/BLUP calculator

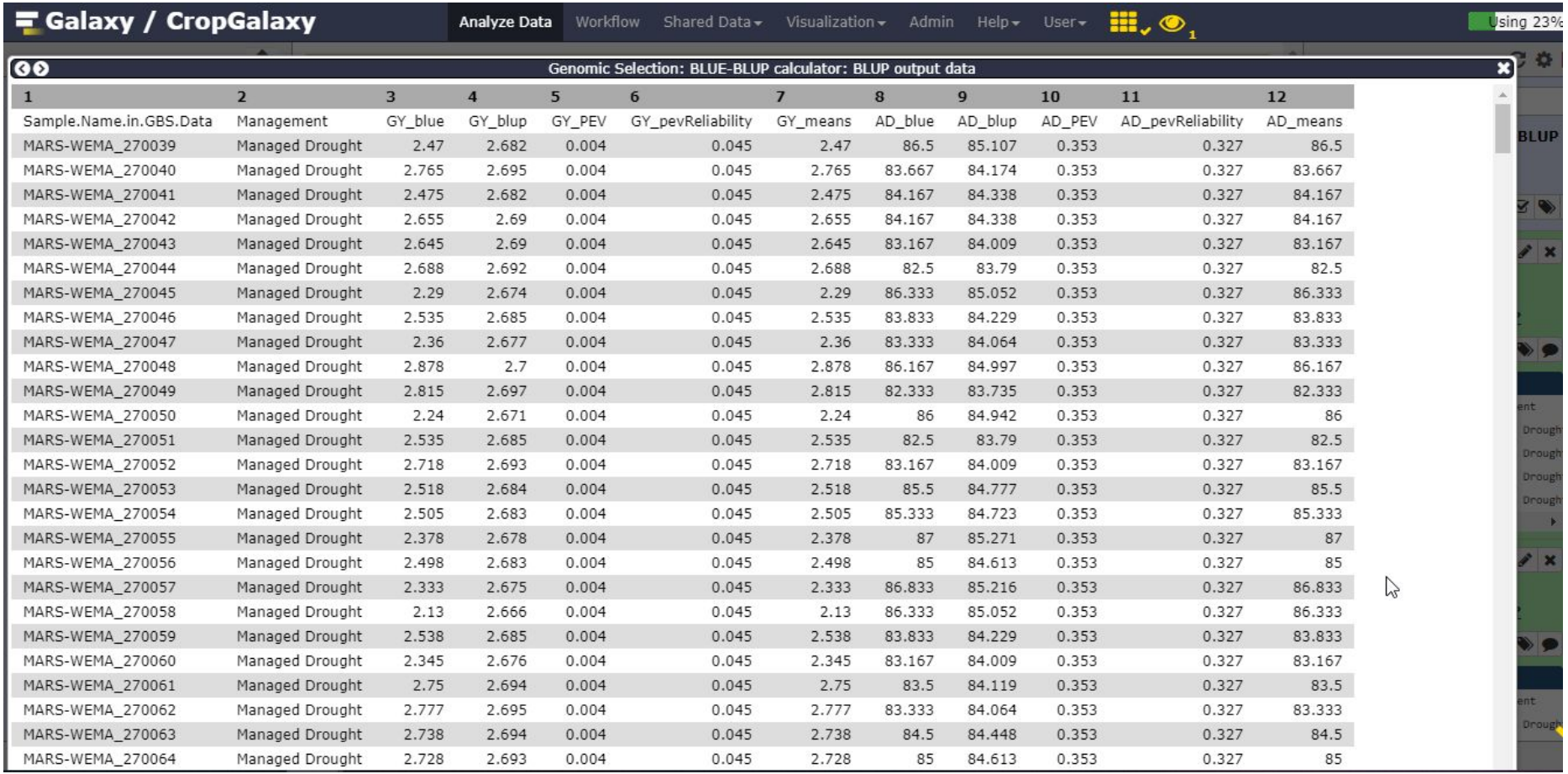


Hands-on:
3. BLUE/BLUP Calculator

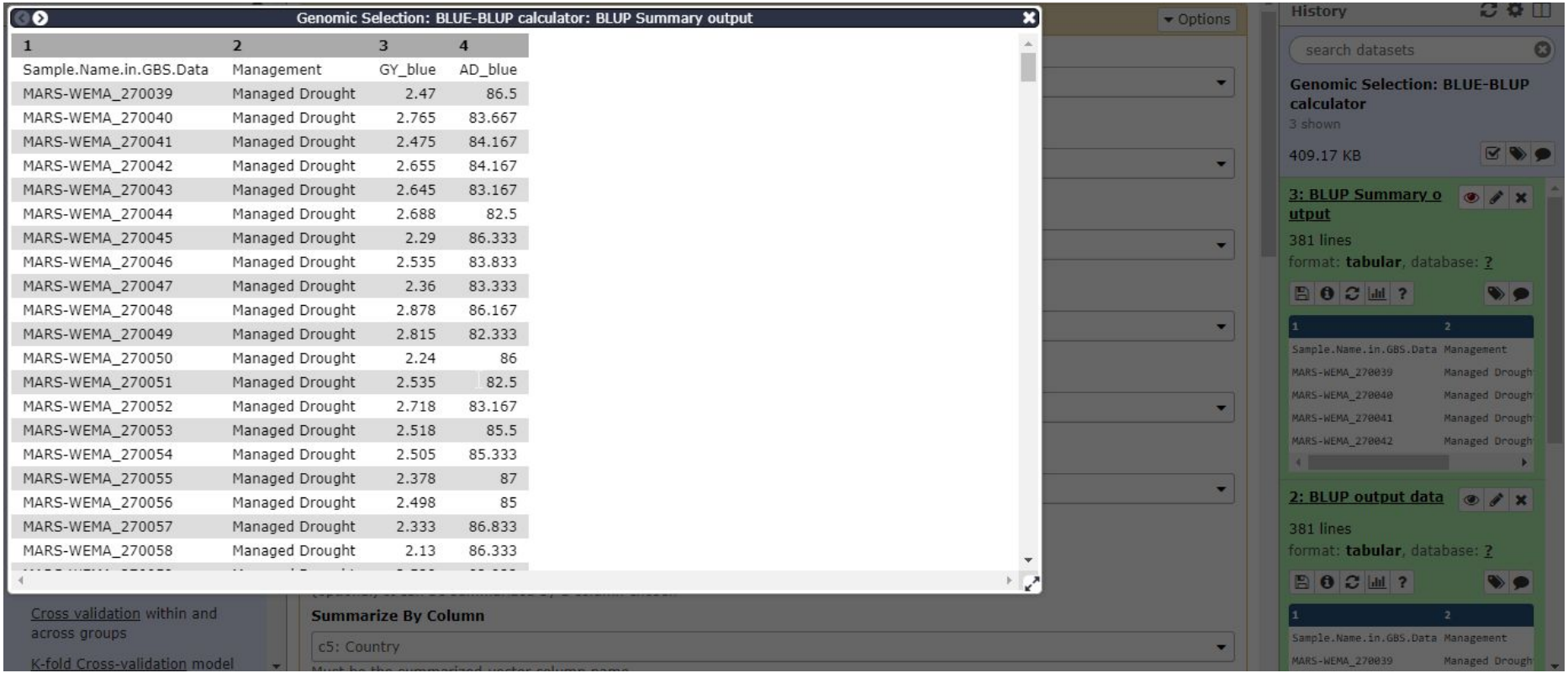
Encoded Data:
Phenotype_BLUE_BLUP-demo.tab




Galaxy / CropGalaxy
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Using 23%
Tools
search tools
Get Data
RiceGalaxy: SNP Data Tools
RiceGalaxy: GEMMA
Plink Data 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
BLUP/BLUE calculator (multiple traits) (Galaxy Version 1.11.0)
Options
Encoded Data
1: Phenotype_BLUP_BLUE-demo.tab
(required) Must be the TAB as both column delimiter and file type
Design
RCBD
(required) Must choose one of the two design options available: Lattice or RCBD
Replication Column
c1: NameOfExperiment
(required) Must be the replication-vector column name
Genotype Column
c2: Year
First Y Column
c3: Season
Last Y Column
c3: Season
(required) Must be the last y-vector column name
Summarize By
Yes
No
(optional) It can be summarized by a column chosen
Summarize By Column
c5: Country
Must be the summarized-vector column name
Output:
- BLUP output data
- BLUP Summary output
History
search datasets
Genomic Selection: BLUE-BLUP calculator
3 shown
409.17 KB
3: BLUP Summary output
381 lines
format: tabular, database: ?
1 2
Sample.Name.in.GBS.Data Management
MARS-WEMA_270039 Managed Drough
MARS-WEMA_270040 Managed Drough
MARS-WEMA_270041 Managed Drough
MARS-WEMA_270042 Managed Drough
2: BLUP output data
381 lines
format: tabular, database: ?
1 2
Sample.Name.in.GBS.Data Management
MARS-WEMA_270039 Managed Drough

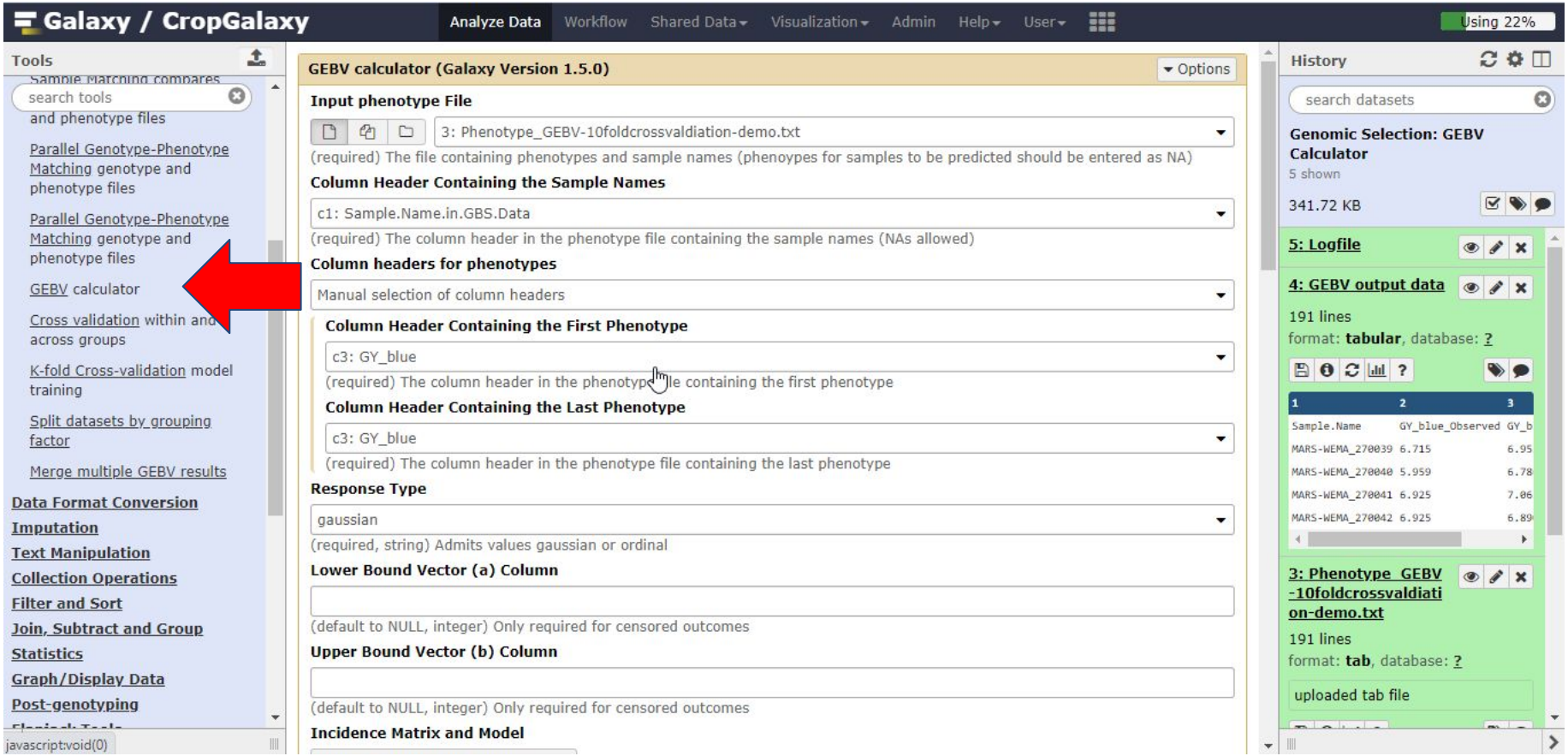


1	2	3	4	5	6	7	8	9	10	11	12
Sample.Name.in.GBS.Data	Management	GY_blue	GY_blup	GY_PEV	GY_pevReliability	GY_means	AD_blue	AD_blup	AD_PEV	AD_pevReliability	AD_means
MARS-WEMA_270039	Managed Drought	2.47	2.682	0.004	0.045	2.47	86.5	85.107	0.353	0.327	86.5
MARS-WEMA_270040	Managed Drought	2.765	2.695	0.004	0.045	2.765	83.667	84.174	0.353	0.327	83.667
MARS-WEMA_270041	Managed Drought	2.475	2.682	0.004	0.045	2.475	84.167	84.338	0.353	0.327	84.167
MARS-WEMA_270042	Managed Drought	2.655	2.69	0.004	0.045	2.655	84.167	84.338	0.353	0.327	84.167
MARS-WEMA_270043	Managed Drought	2.645	2.69	0.004	0.045	2.645	83.167	84.009	0.353	0.327	83.167
MARS-WEMA_270044	Managed Drought	2.688	2.692	0.004	0.045	2.688	82.5	83.79	0.353	0.327	82.5
MARS-WEMA_270045	Managed Drought	2.29	2.674	0.004	0.045	2.29	86.333	85.052	0.353	0.327	86.333
MARS-WEMA_270046	Managed Drought	2.535	2.685	0.004	0.045	2.535	83.833	84.229	0.353	0.327	83.833
MARS-WEMA_270047	Managed Drought	2.36	2.677	0.004	0.045	2.36	83.333	84.064	0.353	0.327	83.333
MARS-WEMA_270048	Managed Drought	2.878	2.7	0.004	0.045	2.878	86.167	84.997	0.353	0.327	86.167
MARS-WEMA_270049	Managed Drought	2.815	2.697	0.004	0.045	2.815	82.333	83.735	0.353	0.327	82.333
MARS-WEMA_270050	Managed Drought	2.24	2.671	0.004	0.045	2.24	86	84.942	0.353	0.327	86
MARS-WEMA_270051	Managed Drought	2.535	2.685	0.004	0.045	2.535	82.5	83.79	0.353	0.327	82.5
MARS-WEMA_270052	Managed Drought	2.718	2.693	0.004	0.045	2.718	83.167	84.009	0.353	0.327	83.167
MARS-WEMA_270053	Managed Drought	2.518	2.684	0.004	0.045	2.518	85.5	84.777	0.353	0.327	85.5
MARS-WEMA_270054	Managed Drought	2.505	2.683	0.004	0.045	2.505	85.333	84.723	0.353	0.327	85.333
MARS-WEMA_270055	Managed Drought	2.378	2.678	0.004	0.045	2.378	87	85.271	0.353	0.327	87
MARS-WEMA_270056	Managed Drought	2.498	2.683	0.004	0.045	2.498	85	84.613	0.353	0.327	85
MARS-WEMA_270057	Managed Drought	2.333	2.675	0.004	0.045	2.333	86.833	85.216	0.353	0.327	86.833
MARS-WEMA_270058	Managed Drought	2.13	2.666	0.004	0.045	2.13	86.333	85.052	0.353	0.327	86.333
MARS-WEMA_270059	Managed Drought	2.538	2.685	0.004	0.045	2.538	83.833	84.229	0.353	0.327	83.833
MARS-WEMA_270060	Managed Drought	2.345	2.676	0.004	0.045	2.345	83.167	84.009	0.353	0.327	83.167
MARS-WEMA_270061	Managed Drought	2.75	2.694	0.004	0.045	2.75	83.5	84.119	0.353	0.327	83.5
MARS-WEMA_270062	Managed Drought	2.777	2.695	0.004	0.045	2.777	83.333	84.064	0.353	0.327	83.333
MARS-WEMA_270063	Managed Drought	2.738	2.694	0.004	0.045	2.738	84.5	84.448	0.353	0.327	84.5
MARS-WEMA_270064	Managed Drought	2.728	2.693	0.004	0.045	2.728	85	84.613	0.353	0.327	85


Genomic Selection: BLUE-BLUP calculator: BLUP Summary output
1 2 3 4
Sample.Name.in.GBS.Data Management GY_blue AD_blue
MARS-WEMA_270039 Managed Drought 2.47 86.5
MARS-WEMA_270040 Managed Drought 2.765 83.667
MARS-WEMA_270041 Managed Drought 2.475 84.167
MARS-WEMA_270042 Managed Drought 2.655 84.167
MARS-WEMA_270043 Managed Drought 2.645 83.167
MARS-WEMA_270044 Managed Drought 2.688 82.5
MARS-WEMA_270045 Managed Drought 2.29 86.333
MARS-WEMA_270046 Managed Drought 2.535 83.833
MARS-WEMA_270047 Managed Drought 2.36 83.333
MARS-WEMA_270048 Managed Drought 2.878 86.167
MARS-WEMA_270049 Managed Drought 2.815 82.333
MARS-WEMA_270050 Managed Drought 2.24 86
MARS-WEMA_270051 Managed Drought 2.535 82.5
MARS-WEMA_270052 Managed Drought 2.718 83.167
MARS-WEMA_270053 Managed Drought 2.518 85.5
MARS-WEMA_270054 Managed Drought 2.505 85.333
MARS-WEMA_270055 Managed Drought 2.378 87
MARS-WEMA_270056 Managed Drought 2.498 85
MARS-WEMA_270057 Managed Drought 2.333 86.833
MARS-WEMA_270058 Managed Drought 2.13 86.333
Options
Cross validation within and across groups
K-fold Cross-validation model
Summarize By Column
c5: Country
History
search datasets
Genomic Selection: BLUE-BLUP calculator
3 shown
409.17 KB
3: BLUP Summary output
381 lines
format: tabular, database: ?
2: BLUP output data
381 lines
format: tabular, database: ?

B. Genomic Breeding Values (GEBV) calculator

- Samples have been genotyped but no phenotypic data
- Predict GEBV based on phenotypes and marker effects identified in the training dataset
- Input:
  - Genotype matrix
  - Phenotyping File
- Output:
  - Table of observed and predicted values for each trait

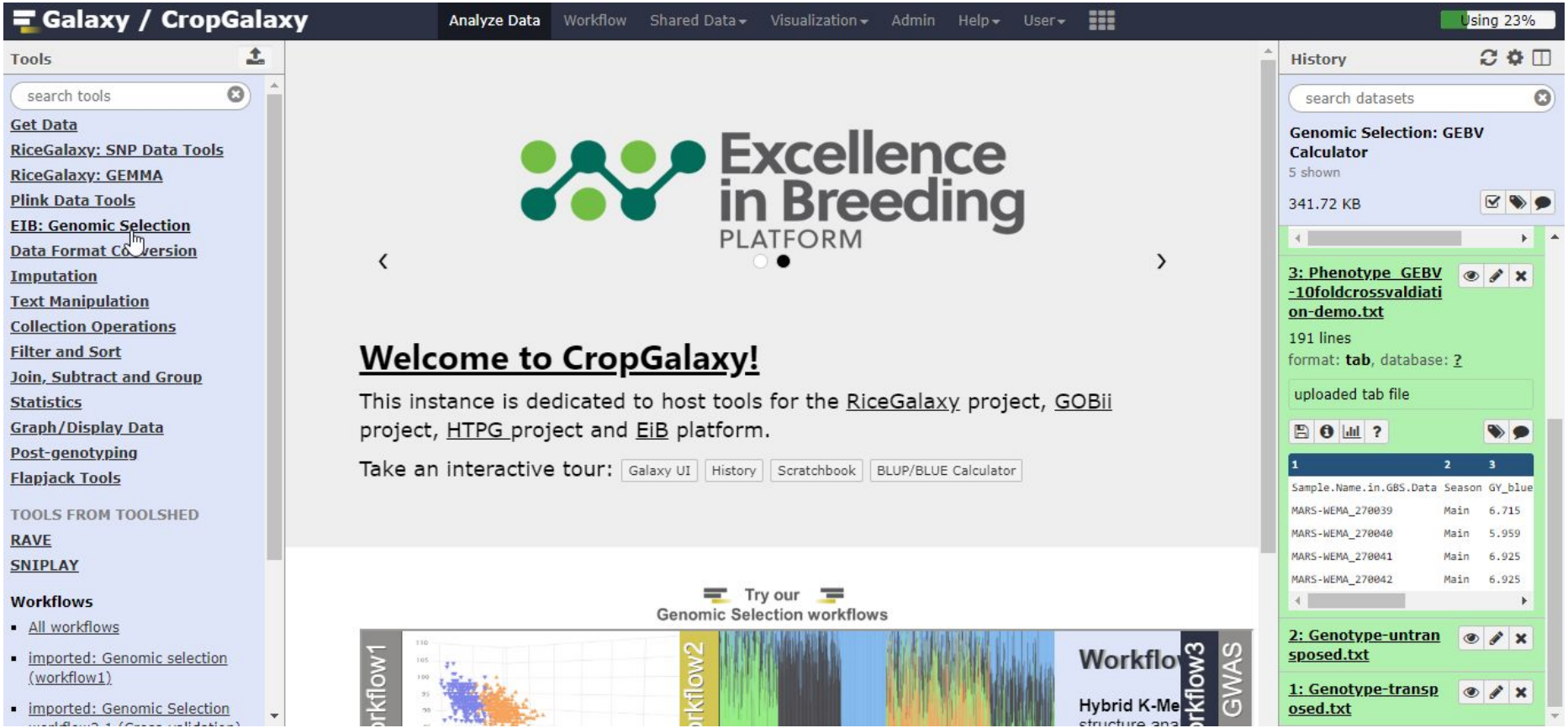


GEBV calculator

IRRI

Hands-on:

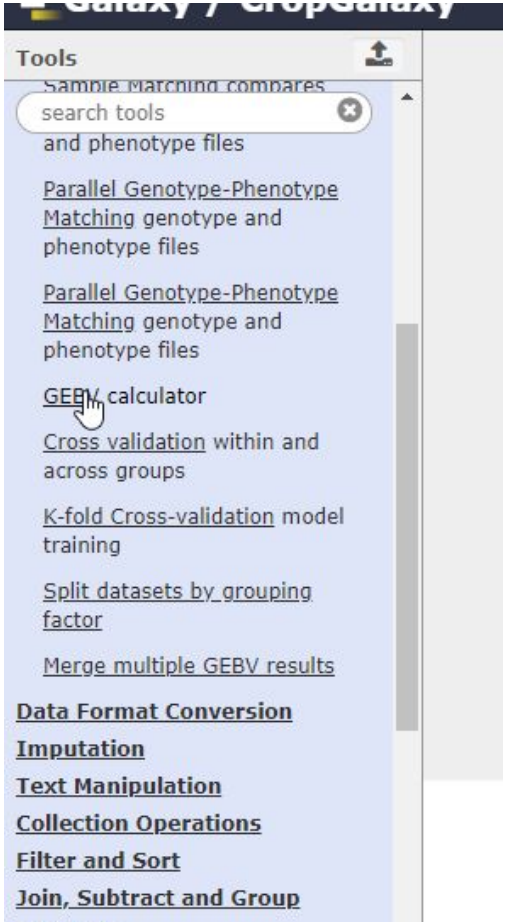
1. Dataset:
    a. Shared Data-> Data Libraries-> Test Dataset-> Genomic Selection-> GEBV
    b. Select all datasets.
    c. Import to History.
    d. Click Green pop-up to start analyzing it.


Galaxy / CropGalaxy
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Using 23%
Tools
search tools
Get Data
RiceGalaxy: SNP Data Tools
RiceGalaxy: GEMMA
Plink Data Tools
EIB: Genomic Selection
Data Format Conversion
Imputation
Text Manipulation
Collection Operations
Filter and Sort
Join, Subtract and Group
Statistics
Graph/Display Data
Post-genotyping
Flapjack Tools
TOOLS FROM TOOLSHED
RAVE
SNIPLAY
Workflows
All workflows
imported: Genomic selection (workflow1)
imported: Genomic Selection
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
Try our
Genomic Selection workflows
History
search datasets
Genomic Selection: GEBV Calculator
5 shown
341.72 KB
3: Phenotype_GEBV-10foldcrossvaldiation-demo.txt
191 lines
format: tab, database: ?
uploaded tab file
Sample.Name.in.GBS.Data Season GY_blue
MARS-WEMA_270039 Main 6.715
MARS-WEMA_270040 Main 5.959
MARS-WEMA_270041 Main 6.925
MARS-WEMA_270042 Main 6.925
2: Genotype-untransposed.txt
1: Genotype-transposed.txt

Hands-on:

2. Tool:

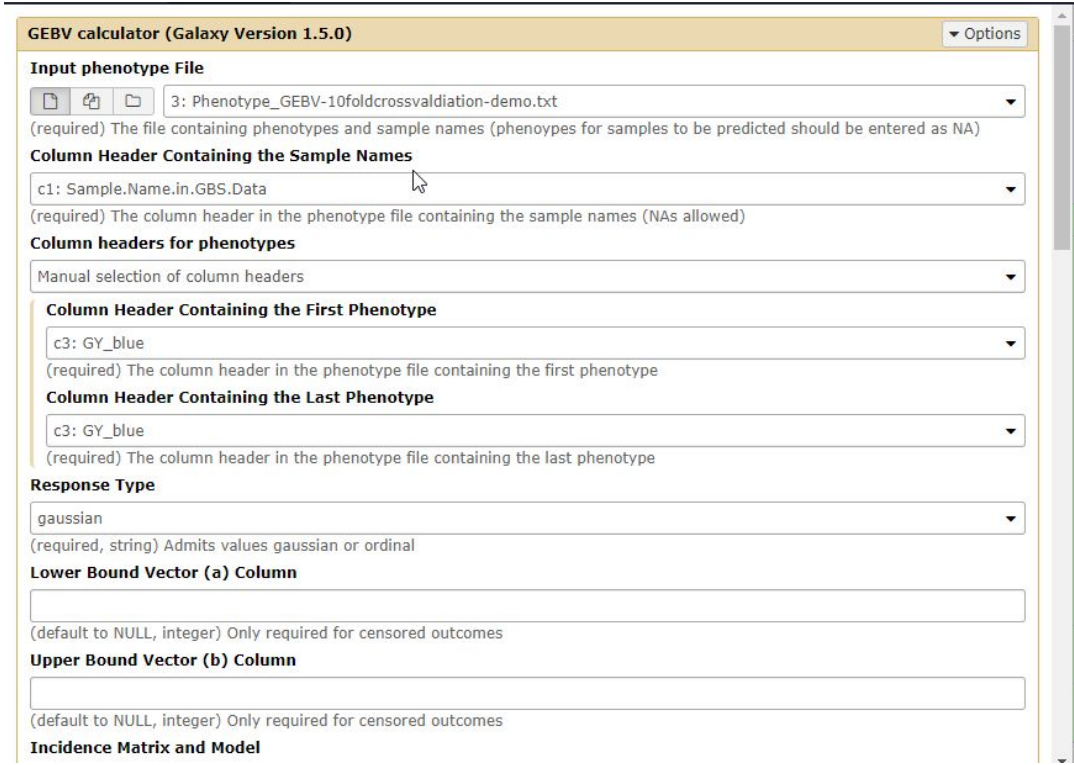
a. Go to Tools-> EIB: Genomic Selection->GEBV calculator

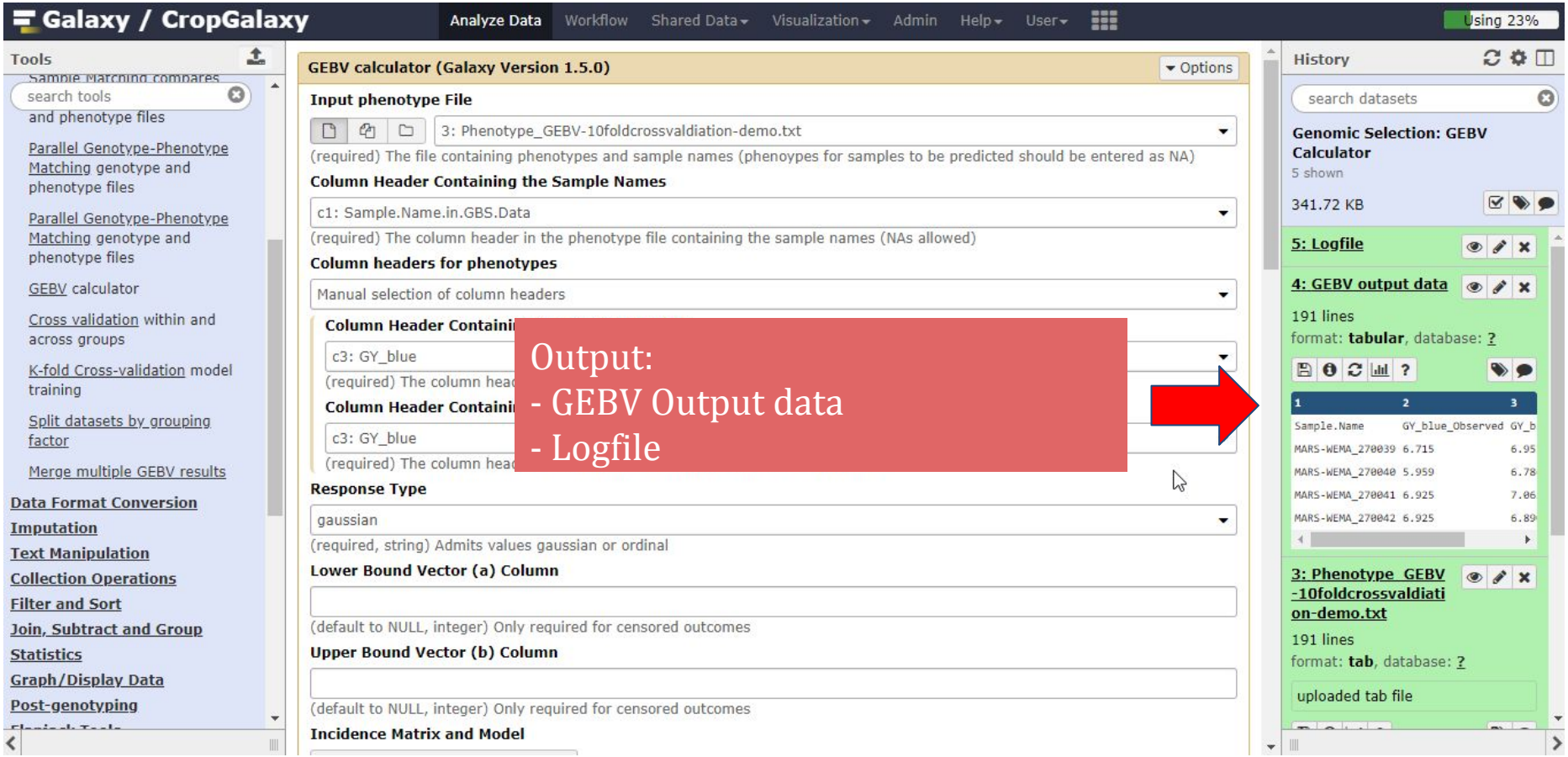


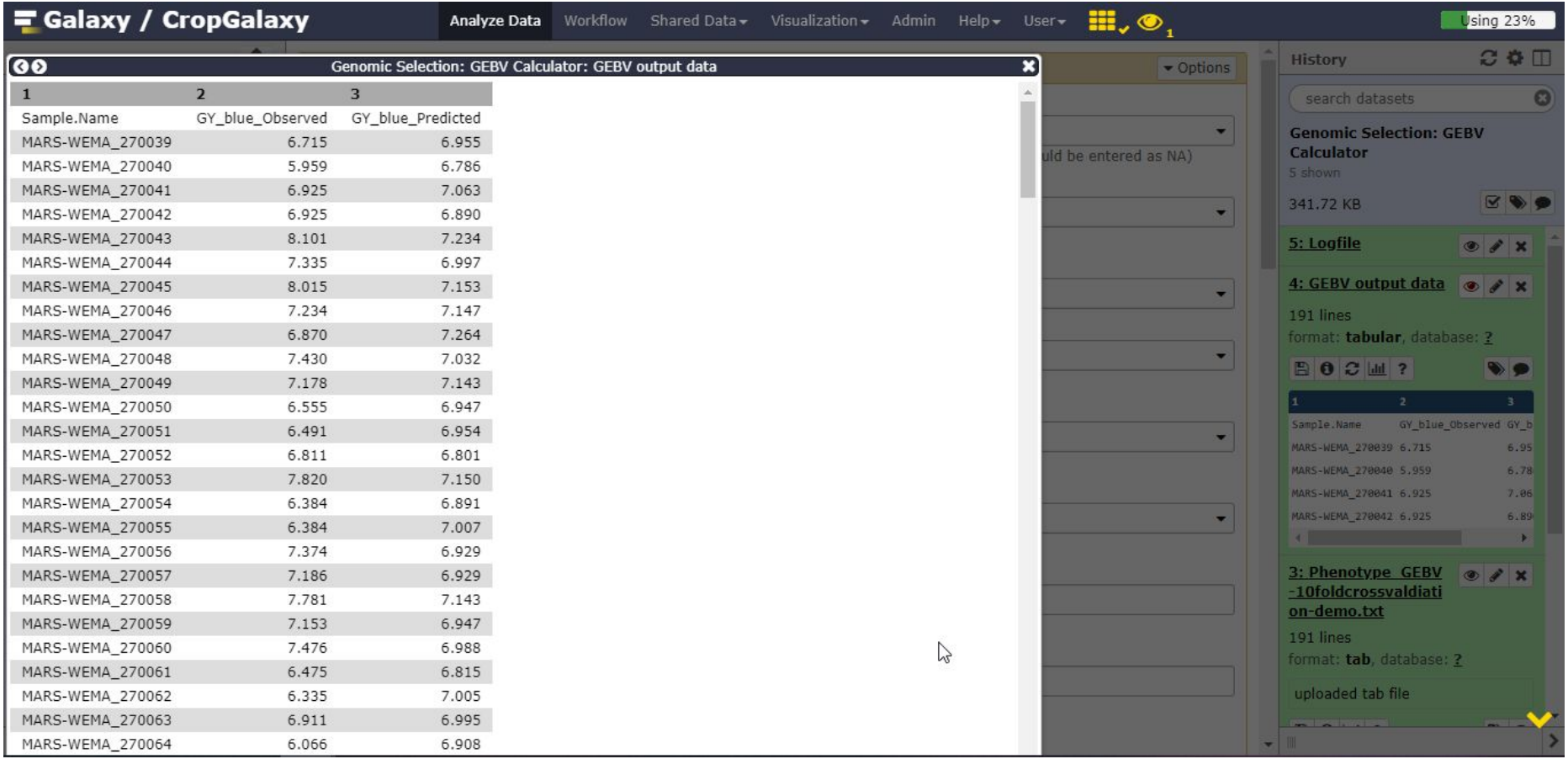
Hands-on:
3. GEBV Calculator

Input phenotype data:
Phenotype_GEBV-10fold crossvaldiation-demo.txt

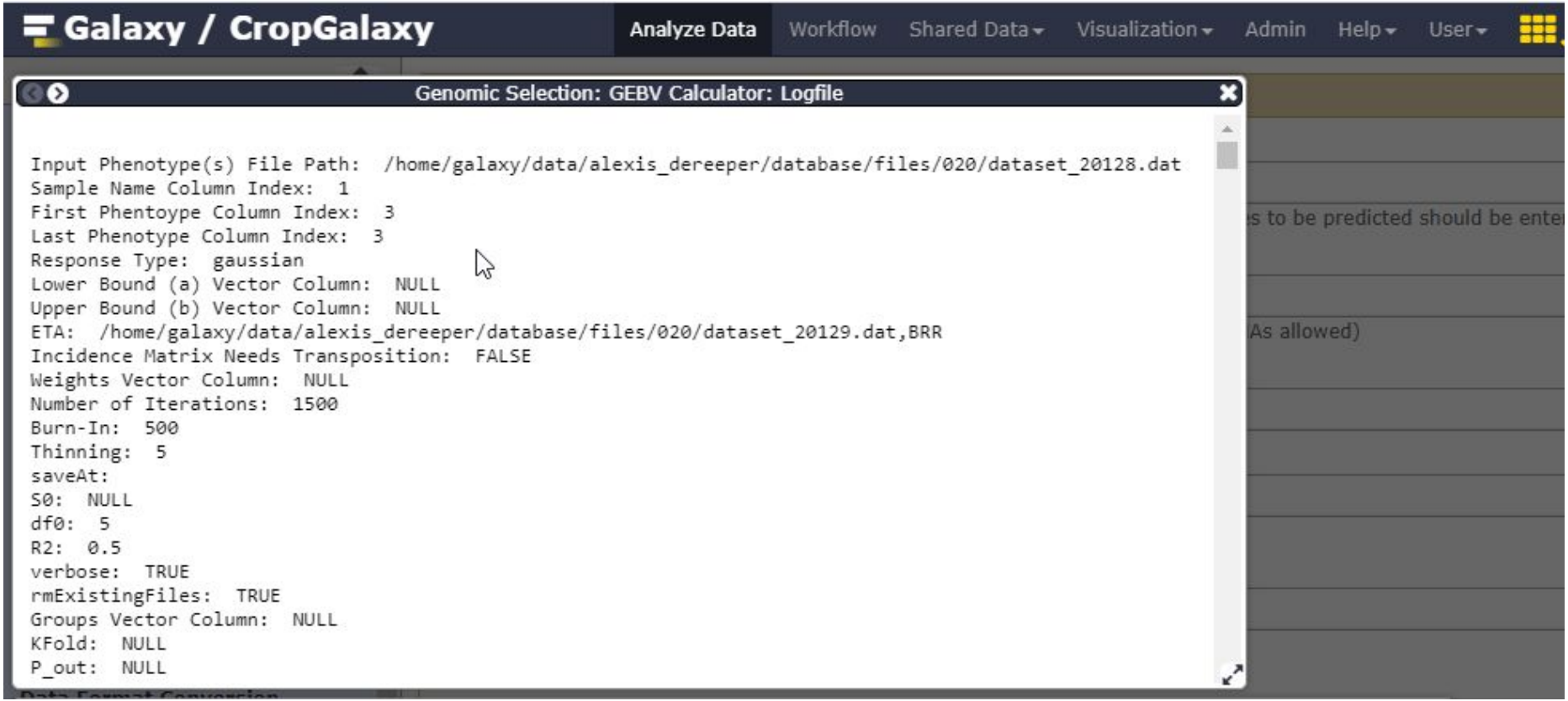
Incidence Matrix:
Genotype-transposed.txt




Galaxy / CropGalaxy
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Using 23%
Tools
search tools
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
Split datasets by grouping factor
Merge multiple GEBV results
Data Format Conversion
Imputation
Text Manipulation
Collection Operations
Filter and Sort
Join, Subtract and Group
Statistics
Graph/Display Data
Post-genotyping
GEBV calculator (Galaxy Version 1.5.0)
Options
Input phenotype File
3: Phenotype_GEBV-10foldcrossvaldiation-demo.txt
(required) The file containing phenotypes and sample names (phenoypes for samples to be predicted should be entered as NA)
Column Header Containing the Sample Names
c1: Sample.Name.in.GBS.Data
(required) The column header in the phenotype file containing the sample names (NAs allowed)
Column headers for phenotypes
Manual selection of column headers
c3: GY_blue
c3: GY_blue
Response Type
gaussian
(required, string) Admits values gaussian or ordinal
Lower Bound Vector (a) Column
(default to NULL, integer) Only required for censored outcomes
Upper Bound Vector (b) Column
(default to NULL, integer) Only required for censored outcomes
Incidence Matrix and Model
Output:
- GEBV Output data
- Logfile
History
search datasets
Genomic Selection: GEBV Calculator
5 shown
341.72 KB
5: Logfile
4: GEBV output data
191 lines
format: tabular, database: ?
1 2 3
Sample.Name GY_blue_Observed GY_b
MARS-WEMA_270039 6.715 6.95
MARS-WEMA_270040 5.959 6.78
MARS-WEMA_270041 6.925 7.06
MARS-WEMA_270042 6.925 6.89
3: Phenotype_GEBV-10foldcrossvaldiation-demo.txt
191 lines
format: tab, database: ?
uploaded tab file

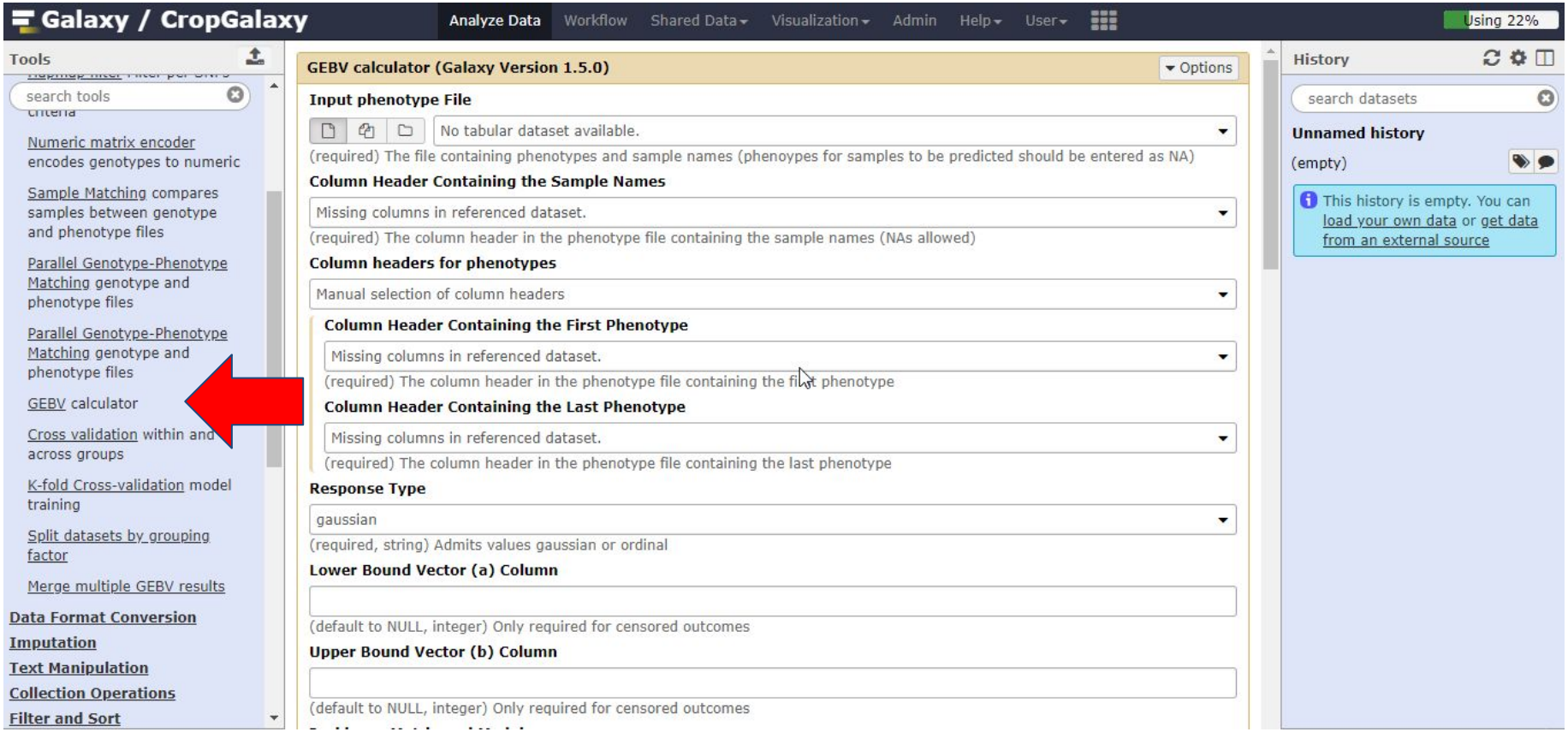


1	2	3
Sample.Name	GY_blue_Observed	GY_blue_Predicted
MARS-WEMA_270039	6.715	6.955
MARS-WEMA_270040	5.959	6.786
MARS-WEMA_270041	6.925	7.063
MARS-WEMA_270042	6.925	6.890
MARS-WEMA_270043	8.101	7.234
MARS-WEMA_270044	7.335	6.997
MARS-WEMA_270045	8.015	7.153
MARS-WEMA_270046	7.234	7.147
MARS-WEMA_270047	6.870	7.264
MARS-WEMA_270048	7.430	7.032
MARS-WEMA_270049	7.178	7.143
MARS-WEMA_270050	6.555	6.947
MARS-WEMA_270051	6.491	6.954
MARS-WEMA_270052	6.811	6.801
MARS-WEMA_270053	7.820	7.150
MARS-WEMA_270054	6.384	6.891
MARS-WEMA_270055	6.384	7.007
MARS-WEMA_270056	7.374	6.929
MARS-WEMA_270057	7.186	6.929
MARS-WEMA_270058	7.781	7.143
MARS-WEMA_270059	7.153	6.947
MARS-WEMA_270060	7.476	6.988
MARS-WEMA_270061	6.475	6.815
MARS-WEMA_270062	6.335	7.005
MARS-WEMA_270063	6.911	6.995
MARS-WEMA_270064	6.066	6.908


Galaxy / CropGalaxy
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Genomic Selection: GEBV Calculator: Logfile
Input Phenotype(s) File Path: /home/galaxy/data/alexis_dereeper/database/files/020/dataset_20128.dat
Sample Name Column Index: 1
First Phentoype Column Index: 3
Last Phenotype Column Index: 3
Response Type: gaussian
Lower Bound (a) Vector Column: NULL
Upper Bound (b) Vector Column: NULL
ETA: /home/galaxy/data/alexis_dereeper/database/files/020/dataset_20129.dat,BRR
Incidence Matrix Needs Transposition: FALSE
Weights Vector Column: NULL
Number of Iterations: 1500
Burn-In: 500
Thinning: 5
saveAt:
S0: NULL
df0: 5
R2: 0.5
verbose: TRUE
rmExistingFiles: TRUE
Groups Vector Column: NULL
KFold: NULL
P_out: NULL

C. Cross Validation (within and across groups)

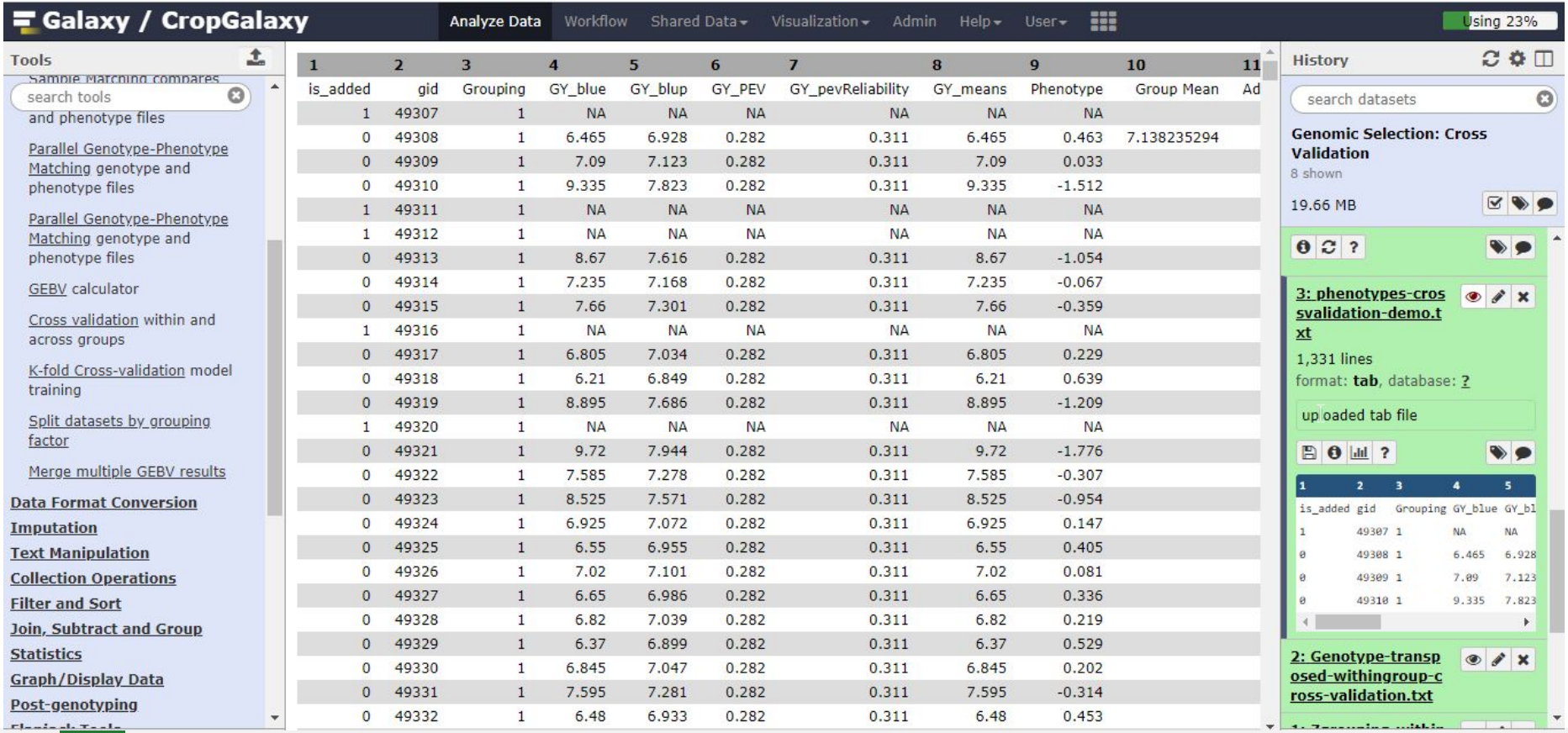
- applied to estimate how the model will perform in general
-

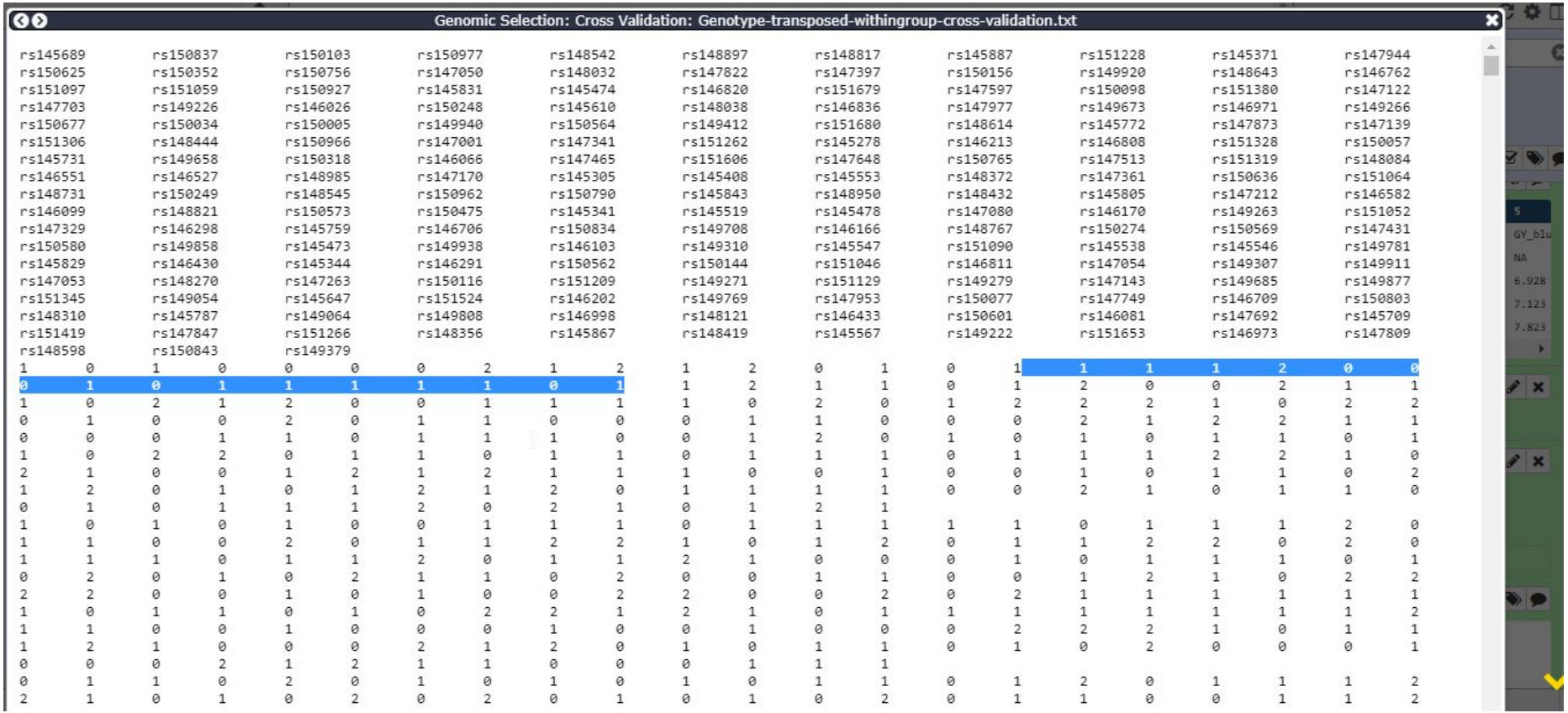


Cross Validation

Hands-on:

1. Dataset:
    a. Shared Data-> Data Libraries-> Test Dataset-> Genomic Selection-> Cross Validation
    b. Select all datasets.
    c. Import to History.
    d. Click Green pop-up to start analyzing it.


Galaxy / CropGalaxy
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Using 23%
Tools
search tools
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
Split datasets by grouping factor
Merge multiple GEBV results
Data Format Conversion
Imputation
Text Manipulation
Collection Operations
Filter and Sort
Join, Subtract and Group
Statistics
Graph/Display Data
Post-genotyping
1 2 3 4 5 6 7 8 9 10 11
is_added gid Grouping GY_blue GY_blup GY_PEV GY_pevReliability GY_means Phenotype Group Mean Ad
1 49307 1 NA NA NA NA NA NA
0 49308 1 6.465 6.928 0.282 0.311 6.465 0.463 7.138235294
0 49309 1 7.09 7.123 0.282 0.311 7.09 0.033
0 49310 1 9.335 7.823 0.282 0.311 9.335 -1.512
1 49311 1 NA NA NA NA NA NA
1 49312 1 NA NA NA NA NA NA
0 49313 1 8.67 7.616 0.282 0.311 8.67 -1.054
0 49314 1 7.235 7.168 0.282 0.311 7.235 -0.067
0 49315 1 7.66 7.301 0.282 0.311 7.66 -0.359
1 49316 1 NA NA NA NA NA NA
0 49317 1 6.805 7.034 0.282 0.311 6.805 0.229
0 49318 1 6.21 6.849 0.282 0.311 6.21 0.639
0 49319 1 8.895 7.686 0.282 0.311 8.895 -1.209
1 49320 1 NA NA NA NA NA NA
0 49321 1 9.72 7.944 0.282 0.311 9.72 -1.776
0 49322 1 7.585 7.278 0.282 0.311 7.585 -0.307
0 49323 1 8.525 7.571 0.282 0.311 8.525 -0.954
0 49324 1 6.925 7.072 0.282 0.311 6.925 0.147
0 49325 1 6.55 6.955 0.282 0.311 6.55 0.405
0 49326 1 7.02 7.101 0.282 0.311 7.02 0.081
0 49327 1 6.65 6.986 0.282 0.311 6.65 0.336
0 49328 1 6.82 7.039 0.282 0.311 6.82 0.219
0 49329 1 6.37 6.899 0.282 0.311 6.37 0.529
0 49330 1 6.845 7.047 0.282 0.311 6.845 0.202
0 49331 1 7.595 7.281 0.282 0.311 7.595 -0.314
0 49332 1 6.48 6.933 0.282 0.311 6.48 0.453
History
search datasets
Genomic Selection: Cross Validation
8 shown
19.66 MB
3: phenotypes-crossvalidation-demo.txt
1,331 lines
format: tab, database: ?
uploaded tab file
1 2 3 4 5
is_added gid Grouping GY_blue GY_bl
1 49307 1 NA NA
0 49308 1 6.465 6.928
0 49309 1 7.09 7.123
0 49310 1 9.335 7.823
2: Genotype-transposed-withingroup-cross-validation.txt

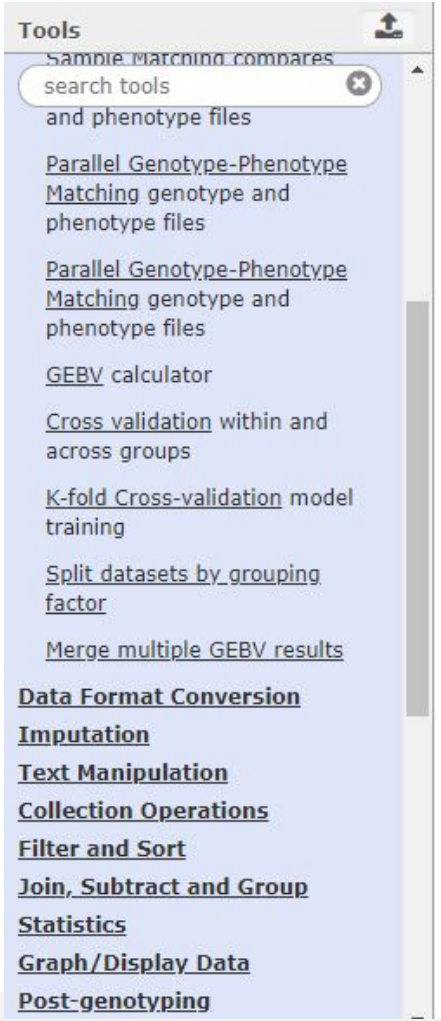

Genomic Selection: Cross Validation: Genotype-transposed-withingroup-cross-validation.txt
rs145689 rs150837 rs150103 rs150977 rs148542 rs148897 rs148817 rs145887 rs151228 rs145371 rs147944
rs150625 rs150352 rs150756 rs147050 rs148032 rs147822 rs147397 rs150156 rs149920 rs148643 rs146762
rs151097 rs151059 rs150927 rs145831 rs145474 rs146820 rs151679 rs147597 rs150098 rs151380 rs147122
rs147703 rs149226 rs146026 rs150248 rs145610 rs148038 rs146836 rs147977 rs149673 rs146971 rs149266
rs150677 rs150034 rs150005 rs149940 rs150564 rs149412 rs151680 rs148614 rs145772 rs147873 rs147139
rs151306 rs148444 rs150966 rs147001 rs147341 rs151262 rs145278 rs146213 rs146808 rs151328 rs150057
rs145731 rs149658 rs150318 rs146066 rs147465 rs151606 rs147648 rs150765 rs147513 rs151319 rs148084
rs146551 rs146527 rs148985 rs147170 rs145305 rs145408 rs145553 rs148372 rs147361 rs150636 rs151064
rs148731 rs150249 rs148545 rs150962 rs150790 rs145843 rs148950 rs148432 rs145805 rs147212 rs146582
rs146099 rs148821 rs150573 rs150475 rs145341 rs145519 rs145478 rs147080 rs146170 rs149263 rs151052
rs147329 rs146298 rs145759 rs146706 rs150834 rs149708 rs146166 rs148767 rs150274 rs150569 rs147431
rs150580 rs149858 rs145473 rs149938 rs146103 rs149310 rs145547 rs151090 rs145538 rs145546 rs149781
rs145829 rs146430 rs145344 rs146291 rs150562 rs150144 rs151046 rs146811 rs147054 rs149307 rs149911
rs147053 rs148270 rs147263 rs150116 rs151209 rs149271 rs151129 rs149279 rs147143 rs149685 rs149877
rs151345 rs149054 rs145647 rs151524 rs146202 rs149769 rs147953 rs150077 rs147749 rs146709 rs150803
rs148310 rs145787 rs149064 rs149808 rs146998 rs148121 rs146433 rs150601 rs146081 rs147692 rs145709
rs151419 rs147847 rs151266 rs148356 rs145867 rs148419 rs145567 rs149222 rs151653 rs146973 rs147809
rs148598 rs150843 rs149379
1 0 1 0 0 0 0 2 1 2 1 2 0 1 0 1 1 1 1 2 0 0
0 1 0 1 1 1 1 1 0 1 1 2 1 1 0 1 2 0 0 2 1 1
1 0 2 1 2 0 0 1 1 1 1 0 2 0 1 2 2 2 1 0 2 2
0 1 0 0 2 0 1 1 0 0 0 1 1 0 0 0 2 1 2 2 1 1
0 0 0 1 1 0 1 1 1 0 0 1 2 0 1 0 1 0 1 1 0 1
1 0 2 2 0 1 1 0 1 1 0 1 1 1 0 1 1 1 2 2 1 0
2 1 0 0 1 2 1 2 1 1 1 0 0 1 0 0 1 0 1 1 0 2
1 2 0 1 0 1 2 1 2 0 1 1 1 1 0 0 2 1 0 1 1 0
0 1 0 1 1 1 2 0 2 1 0 1 2 1
1 0 1 0 1 0 0 1 1 1 0 1 1 1 1 1 0 1 1 1 2 0
1 1 0 0 2 0 1 1 2 2 1 0 1 2 0 1 1 2 2 0 2 0
1 1 1 0 1 1 2 0 1 1 2 1 0 0 0 1 0 1 1 1 0 1
0 2 0 1 0 2 1 1 0 2 0 0 1 1 0 0 1 2 1 0 2 2
2 2 0 0 1 0 1 0 0 2 2 0 0 2 0 2 1 1 1 1 1 1
1 0 1 1 0 1 0 2 2 1 2 1 0 1 1 1 1 1 1 1 1 2
1 1 0 0 1 0 0 0 1 0 0 1 0 0 0 2 2 2 1 0 1 1
1 2 1 0 0 0 2 1 2 0 1 0 1 1 0 1 0 2 0 0 0 1
0 0 0 2 1 2 1 1 0 0 0 1 1 1
0 1 1 0 2 0 1 0 1 0 1 0 1 1 0 1 2 0 1 1 1 2
2 1 0 1 0 2 0 2 0 1 0 1 0 2 0 1 1 0 0 1 1 2

Genomic Selection: Cross Validation: 7grouping-withingroup-demo.txt
1
Grouping
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1

Hands-on:

2. Tool:

a. Go to Tools-> EIB: Genomic Selection->Cross Validation



Hands-on:

3. Cross Validation

Input genotype file:

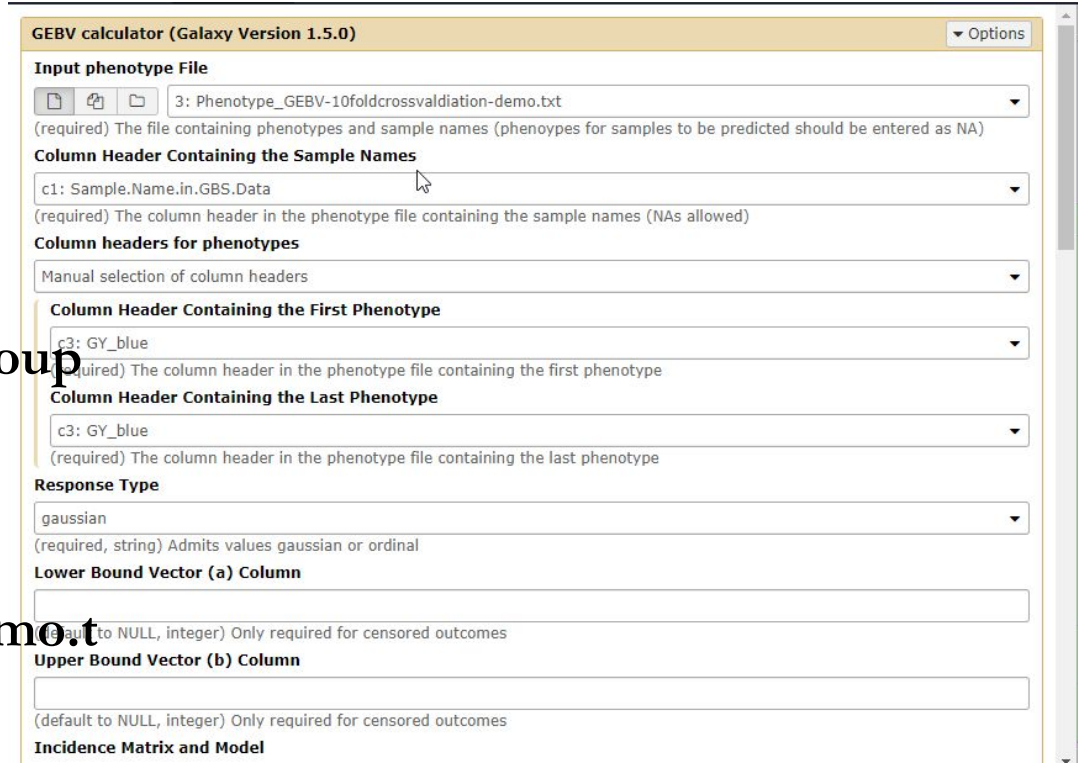
Genotype-transposed-withingroup-cross-validation.txt

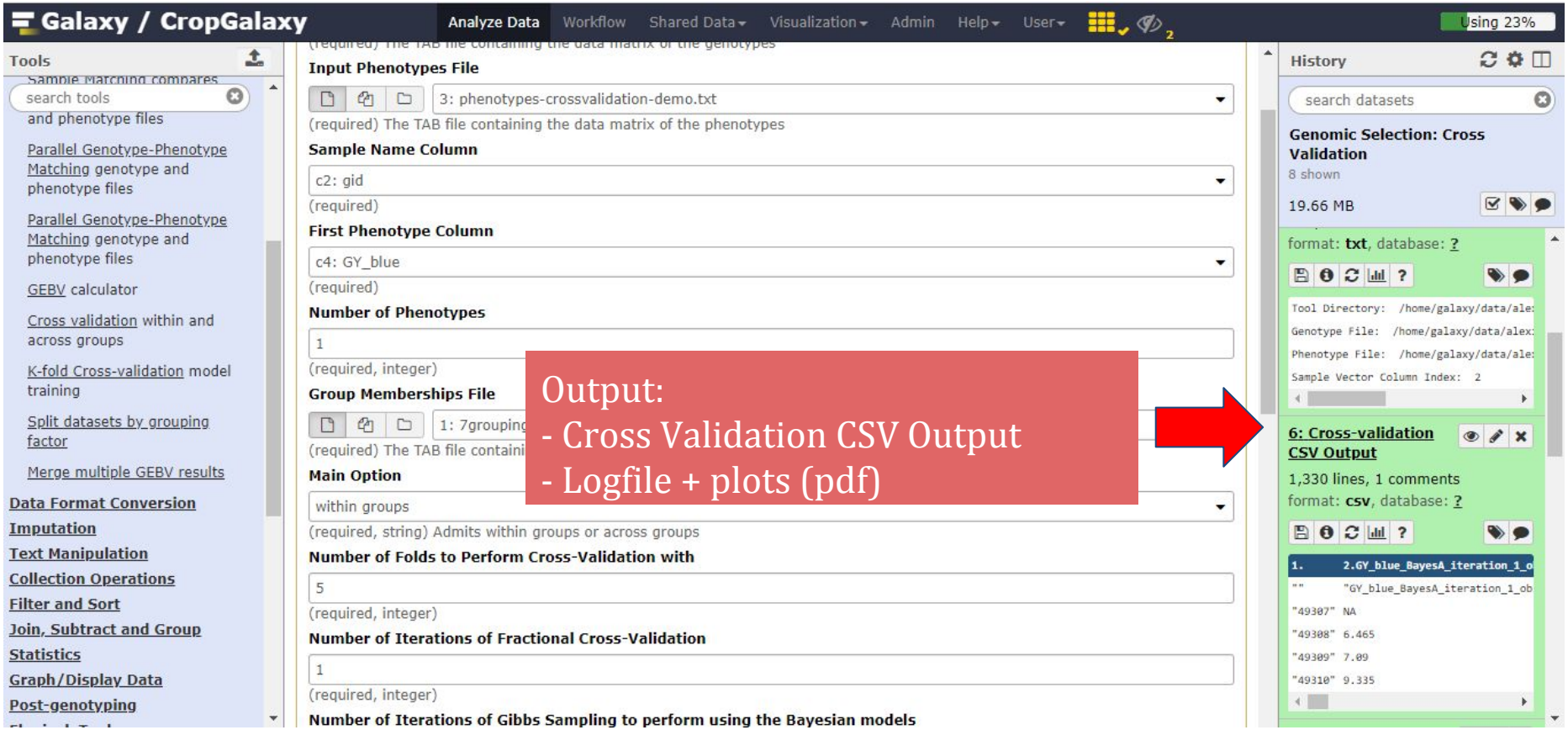
Input phenotype file:

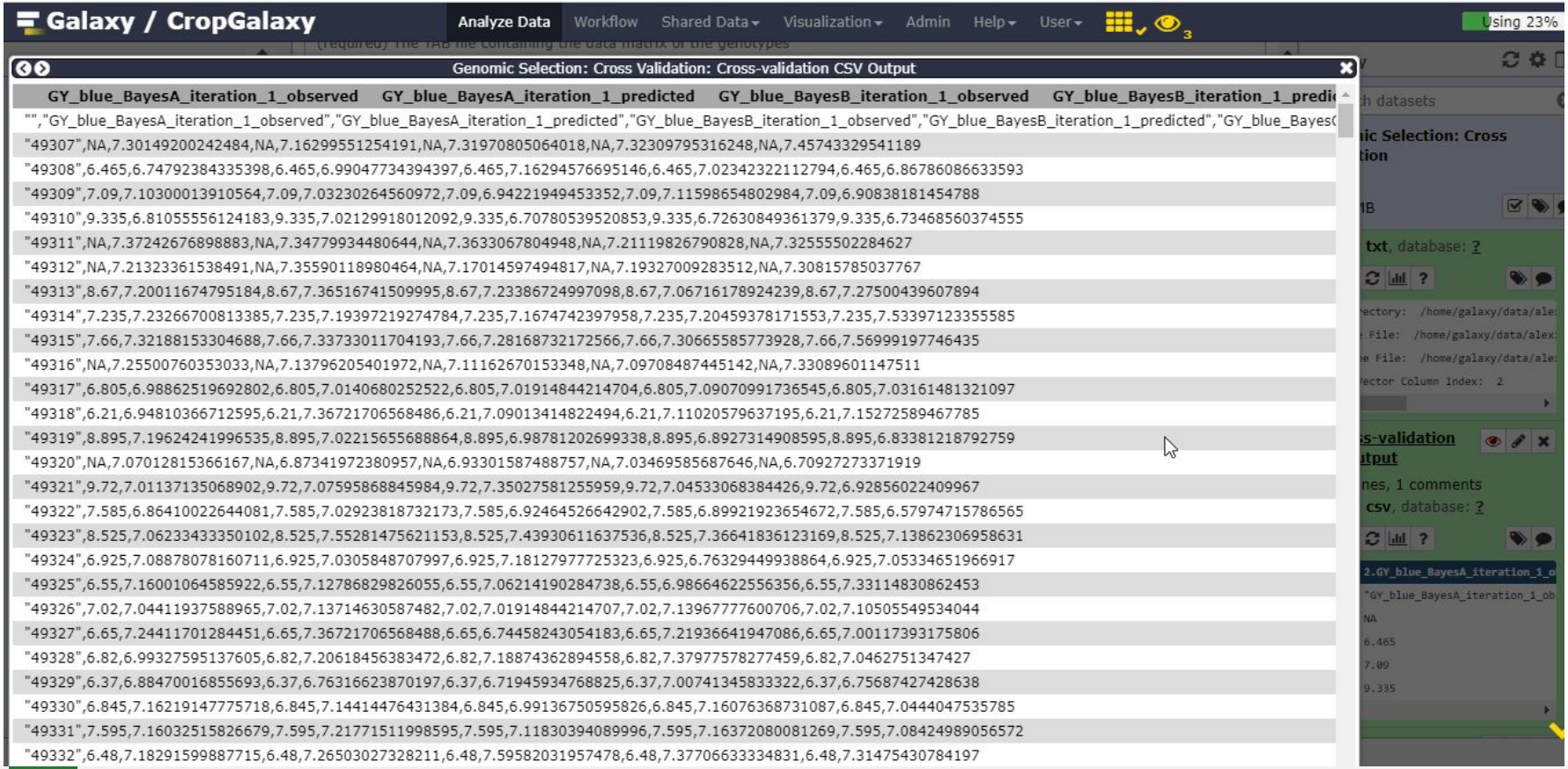
phenotypes-crossvalidation-demo.txt

Group Membership file:

7grouping-withingroup-demo.txt




Galaxy / CropGalaxy
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Using 23%
Tools
search tools
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
Split datasets by grouping factor
Merge multiple GEBV results
Data Format Conversion
Imputation
Text Manipulation
Collection Operations
Filter and Sort
Join, Subtract and Group
Statistics
Graph/Display Data
Post-genotyping
Input Phenotypes File
3: phenotypes-crossvalidation-demo.txt
(required) The TAB file containing the data matrix of the phenotypes
Sample Name Column
c2: gid
(required)
First Phenotype Column
c4: GY_blue
(required)
Number of Phenotypes
1
(required, integer)
Group Memberships File
(required) The TAB file containi
Main Option
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
Output:
- Cross Validation CSV Output
- Logfile + plots (pdf)
History
search datasets
Genomic Selection: Cross Validation
8 shown
19.66 MB
format: txt, database: ?
Tool Directory: /home/galaxy/data/ale
Genotype File: /home/galaxy/data/alex
Phenotype File: /home/galaxy/data/ale
Sample Vector Column Index: 2
6: Cross-validation CSV Output
1,330 lines, 1 comments
format: csv, database: ?
1. 2.GY_blue_BayesA_iteration_1_o
"" "GY_blue_BayesA_iteration_1_ob
"49307" NA
"49308" 6.465
"49309" 7.09
"49310" 9.335


Galaxy / CropGalaxy
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Using 23%
Genomic Selection: Cross Validation: Cross-validation CSV Output
GY_blue_BayesA_iteration_1_observed
GY_blue_BayesA_iteration_1_predicted
GY_blue_BayesB_iteration_1_observed
GY_blue_BayesB_iteration_1_predi
"","GY_blue_BayesA_iteration_1_observed","GY_blue_BayesA_iteration_1_predicted","GY_blue_BayesB_iteration_1_observed","GY_blue_BayesB_iteration_1_predicted","GY_blue_Bayes(
"49307",NA,7.30149200242484,NA,7.16299551254191,NA,7.31970805064018,NA,7.32309795316248,NA,7.45743329541189
"49308",6.465,6.74792384335398,6.465,6.99047734394397,6.465,7.16294576695146,6.465,7.02342322112794,6.465,6.86786086633593
"49309",7.09,7.10300013910564,7.09,7.03230264560972,7.09,6.94221949453352,7.09,7.11598654802984,7.09,6.90838181454788
"49310",9.335,6.81055556124183,9.335,7.02129918012092,9.335,6.70780539520853,9.335,6.72630849361379,9.335,6.73468560374555
"49311",NA,7.37242676898883,NA,7.34779934480644,NA,7.3633067804948,NA,7.21119826790828,NA,7.32555502284627
"49312",NA,7.21323361538491,NA,7.35590118980464,NA,7.17014597494817,NA,7.19327009283512,NA,7.30815785037767
"49313",8.67,7.20011674795184,8.67,7.36516741509995,8.67,7.23386724997098,8.67,7.06716178924239,8.67,7.27500439607894
"49314",7.235,7.23266700813385,7.235,7.19397219274784,7.235,7.1674742397958,7.235,7.20459378171553,7.235,7.53397123355585
"49315",7.66,7.32188153304688,7.66,7.33733011704193,7.66,7.28168732172566,7.66,7.30665585773928,7.66,7.56999197746435
"49316",NA,7.25500760353033,NA,7.13796205401972,NA,7.11162670153348,NA,7.09708487445142,NA,7.33089601147511
"49317",6.805,6.98862519692802,6.805,7.0140680252522,6.805,7.01914844214704,6.805,7.09070991736545,6.805,7.03161481321097
"49318",6.21,6.94810366712595,6.21,7.36721706568486,6.21,7.09013414822494,6.21,7.11020579637195,6.21,7.15272589467785
"49319",8.895,7.19624241996535,8.895,7.02215655688864,8.895,6.98781202699338,8.895,6.8927314908595,8.895,6.83381218792759
"49320",NA,7.07012815366167,NA,6.87341972380957,NA,6.93301587488757,NA,7.03469585687646,NA,6.70927273371919
"49321",9.72,7.01137135068902,9.72,7.07595868845984,9.72,7.35027581255959,9.72,7.04533068384426,9.72,6.92856022409967
"49322",7.585,6.86410022644081,7.585,7.02923818732173,7.585,6.92464526642902,7.585,6.89921923654672,7.585,6.57974715786565
"49323",8.525,7.06233433350102,8.525,7.55281475621153,8.525,7.43930611637536,8.525,7.36641836123169,8.525,7.13862306958631
"49324",6.925,7.08878078160711,6.925,7.0305848707997,6.925,7.18127977725323,6.925,6.76329449938864,6.925,7.05334651966917
"49325",6.55,7.16001064585922,6.55,7.12786829826055,6.55,7.06214190284738,6.55,6.98664622556356,6.55,7.33114830862453
"49326",7.02,7.04411937588965,7.02,7.13714630587482,7.02,7.01914844214707,7.02,7.13967777600706,7.02,7.10505549534044
"49327",6.65,7.24411701284451,6.65,7.36721706568488,6.65,6.74458243054183,6.65,7.21936641947086,6.65,7.00117393175806
"49328",6.82,6.99327595137605,6.82,7.20618456383472,6.82,7.18874362894558,6.82,7.37977578277459,6.82,7.0462751347427
"49329",6.37,6.88470016855693,6.37,6.76316623870197,6.37,6.71945934768825,6.37,7.00741345833322,6.37,6.75687427428638
"49330",6.845,7.16219147775718,6.845,7.14414476431384,6.845,6.99136750595826,6.845,7.16076368731087,6.845,7.0444047535785
"49331",7.595,7.16032515826679,7.595,7.21771511998595,7.595,7.11830394089996,7.595,7.16372080081269,7.595,7.08424989056572
"49332",6.48,7.18291599887715,6.48,7.26503027328211,6.48,7.59582031957478,6.48,7.37706633334831,6.48,7.31475430784197
h datasets
ic Selection: Cross
tion
txt, database: ?
ectory: /home/galaxy/data/ale
File: /home/galaxy/data/alex
e File: /home/galaxy/data/ale
ector Column Index: 2
s-validation
tput
nes, 1 comments
csv, database: ?
2.GY_blue_BayesA_iteration_1_o
"GY_blue_BayesA_iteration_1_ob
NA
6.465
7.09
9.335

D. OGHMA Tools

Thank you!

BILL & MELINDA
GATES foundation

Excellence
in Breeding
PLATFORM

GOBii

GS
GALAXY

CIMMYT
International Maize and Wheat
Improvement Center

ICRISAT
INTERNATIONAL CROPS RESEARCH
INSTITUTE FOR THE SEMI-ARID TROPICS

IRRI

Southern Cross
University

Thank you!

- Kelly Robbins
- Elizabeth Jones
- Star Xiao Gao
- Mike Olsen
- Susanne Dreisigacker
- JOsh Cobb
- Rajeev Varshney
- Victor Jun Ulat
- Angel Villahoz-Baleta
- Alexis Dereeper
- Umesh Rosyara
- Xuecai Zhang
- Juan Andres Burgueño
- Pancho Crossa
- Kate Dehner
- Manish Roorkiwal
- Hima Kudapa
- Santosh Deshpande
- Paulino Perez
- Fernando Toledo
- Nicolas Beaume
- Ramil Mauleon
- Kenneth McNally
- Dmytro Chebotarov
- Abhishek Rathore
- Jean-Luc Jannink
- Mark Sorrells
- Nickolai Alexandro