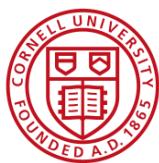




*Genomic Open-source Breeding
informatics initiative*

BILL & MELINDA
GATES foundation



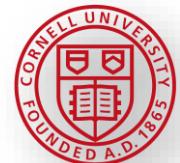
Genomic Open-source Breeding informatics initiative



To transform breeding by enabling the implementation of genomic and marker-assisted selection as part of routine breeding programs in developing countries

- ✓ Build scalable databases for genomics and marker data management
- ✓ Connect to breeding management systems
- ✓ Build and connect tools and analysis pipelines

GOBii A Global Community



IRRI



Excellence in Breeding Platform



Module 1

Breeding program management

Module 2

Trait discovery and breeding

Module 3

Genotyping and sequencing

Module 4

Phenotyping

Module 5

Bioinformatics, biometrics, and Data Management

Enable step changes in increasing genetic gains and scale of impact of breeding programs in the developing world

Enterprise Breeding System (EBS)



Core Breeding (B4R)

- Breeding Program Planning
- Inventory Management
- Find Germplasm
- Nursery Creation
- Trial Design
- Crossing
- Harvest Manager
- Services
- Shipment

Field mapping, Phenotyping, QC (KDX)

Sample Manager

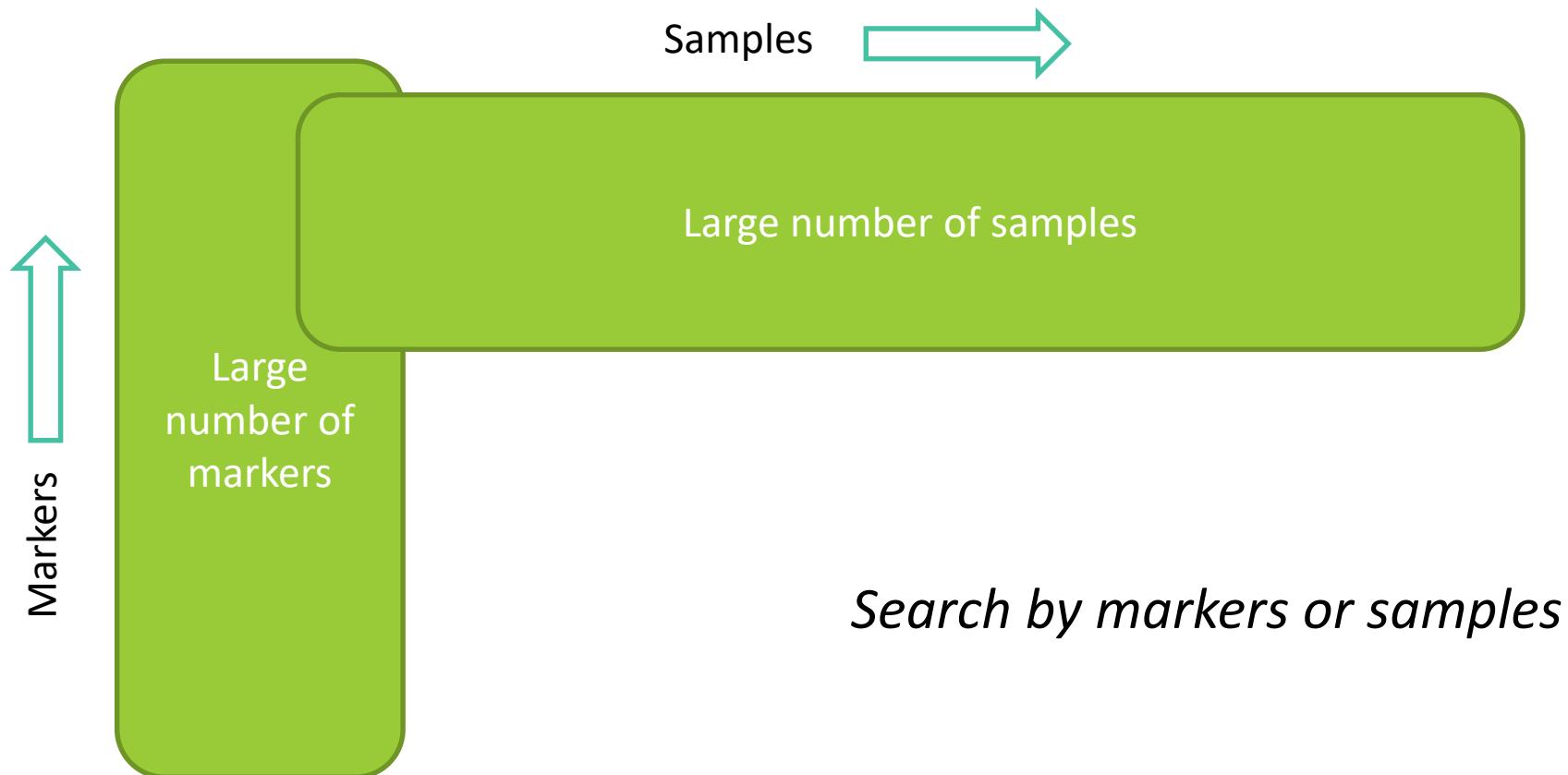
Genotype Management (GOBii)

- Loader
- Extractor
- QC (KDC)
- MAS (Ped Ver, MABC, FB) (Flapjack)
- GWAS (Galaxy)
- Genomic Selection (Galaxy)

Trial Analytics

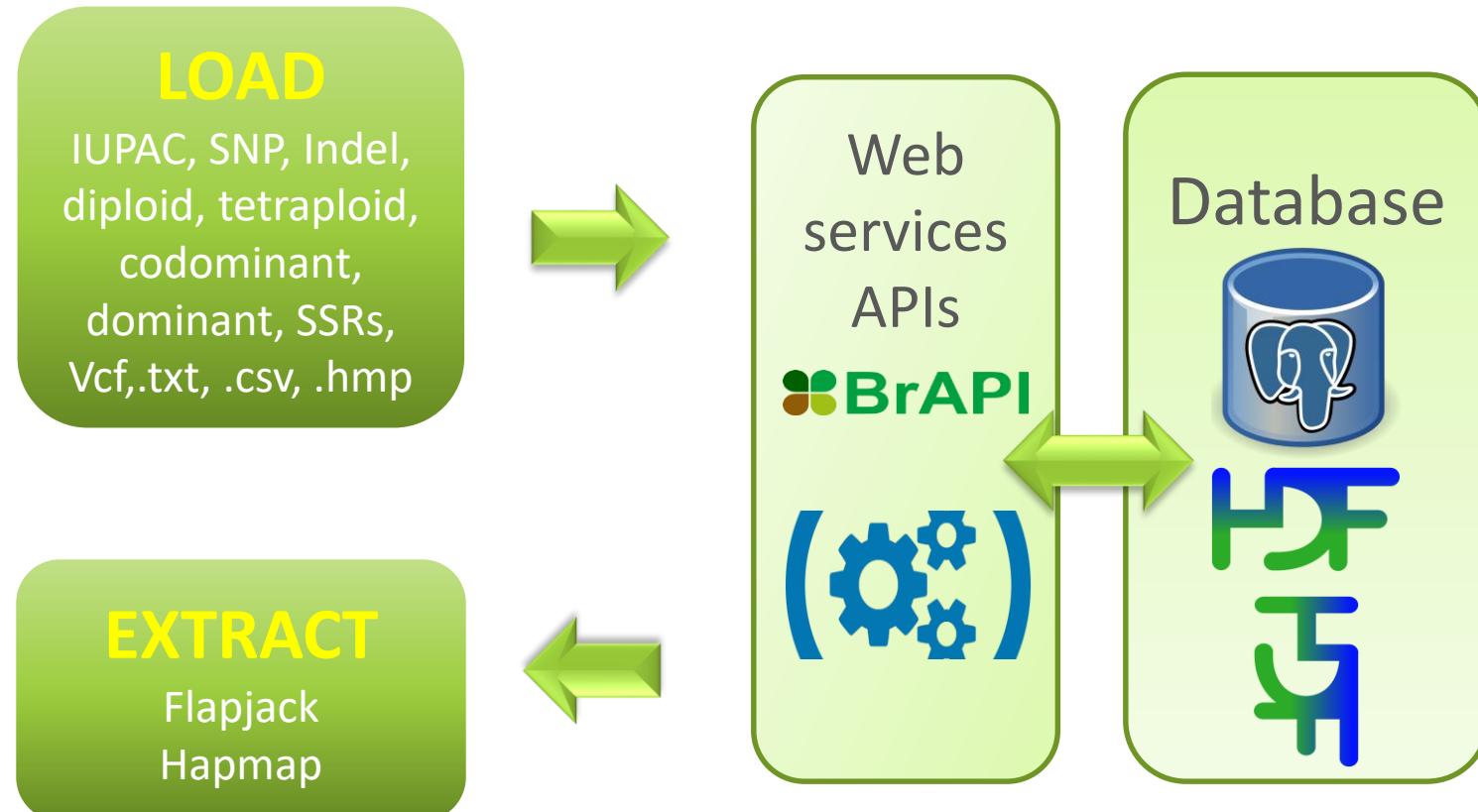
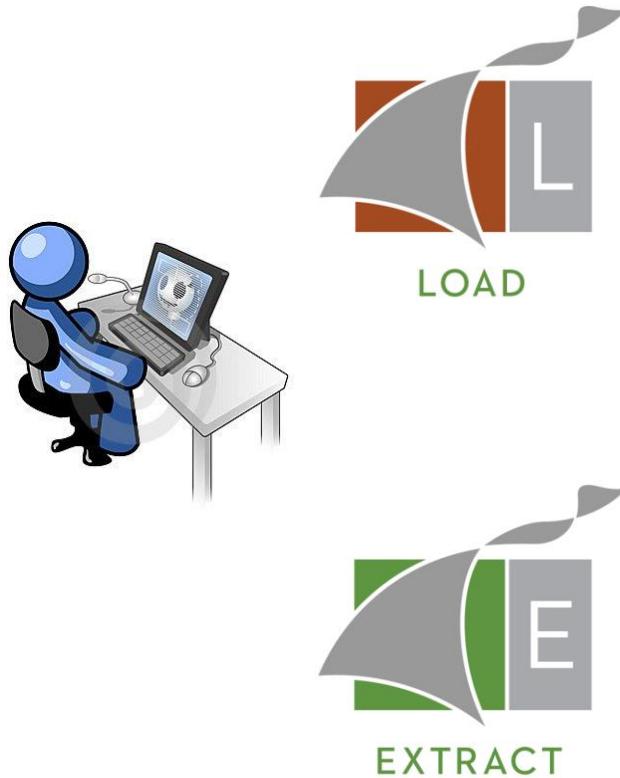
Advancement Decision

Genotyping Data Management – Dimensionality Issues

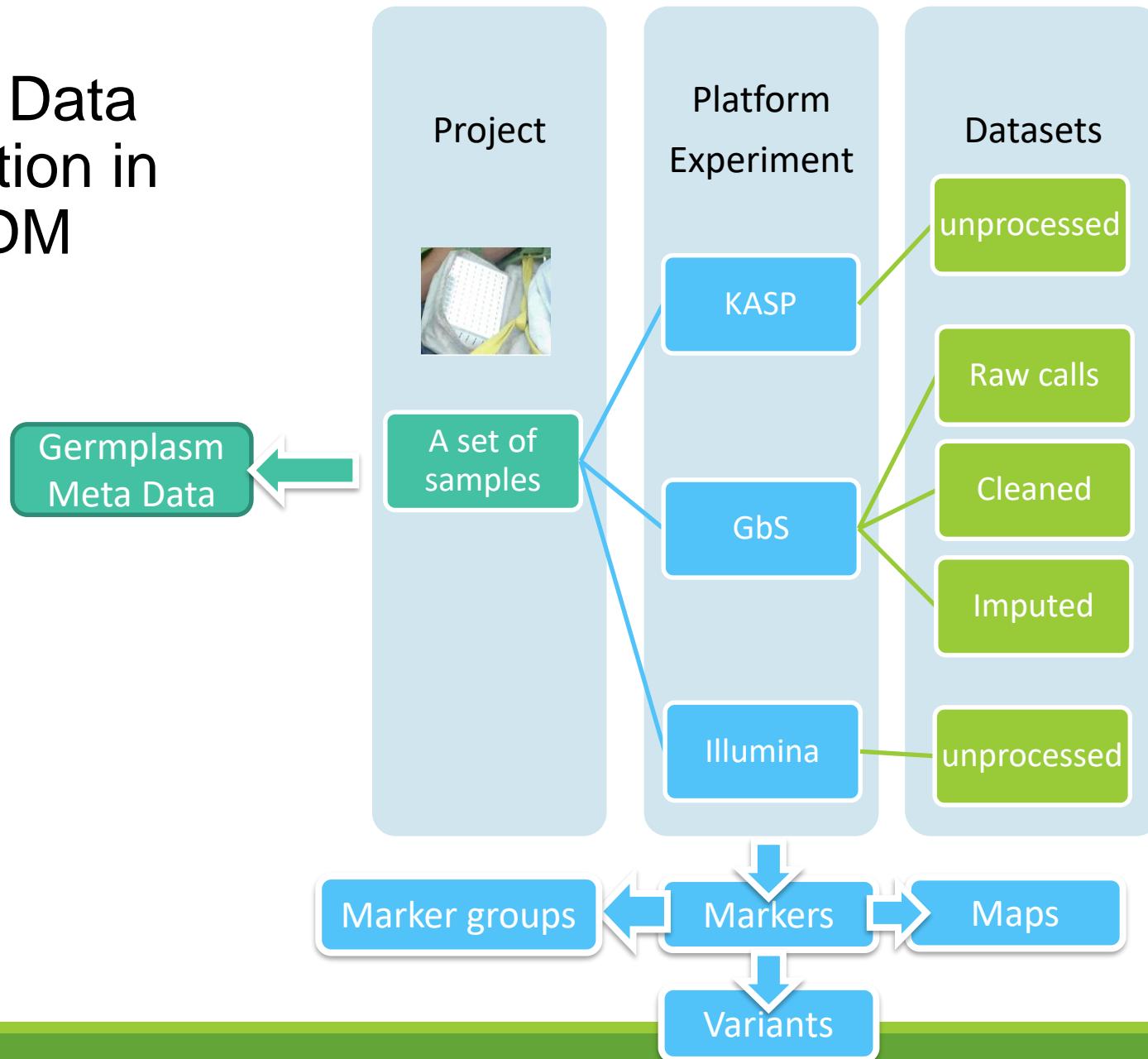


GOBii Genomic Data Management

Designed with extract performance in mind



Levels of Data Organization in GOBii-GDM



GOBii Germplasm levels

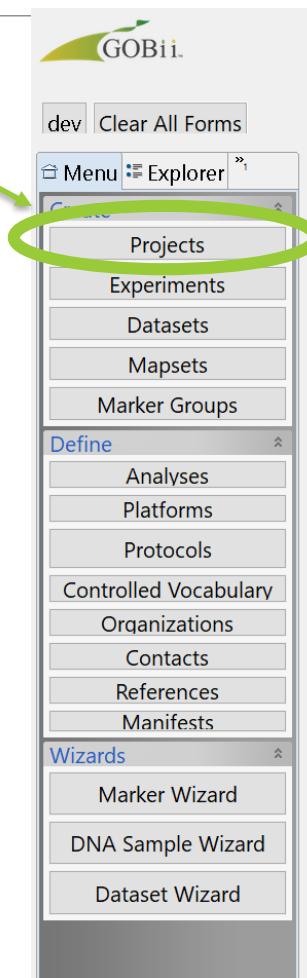


- Germplasm name (accession, common name)
 - External code (eg plot ID)
 - **Sample name (name sent to vendor)**
 - **Sample UUID (Universal unique identifier)**
 - **Sample number**
 - DNA run name (name from vendor)

GOBii-GDM Data Loader



- Define projects
- Map markers to multiple maps
- Define marker groups
- Define entities to support data eg platform, analyses, maps etc



GOBII Data Loader

Projects

PI Contacts: Jones, Liz

Projects: test marker groups, testqcwithcodominat

*Project Name: test marker groups
Code: Jones, Liz_test marker groups
test marker groups
Project Description:

*PI contact: Jones, Liz

Properties:

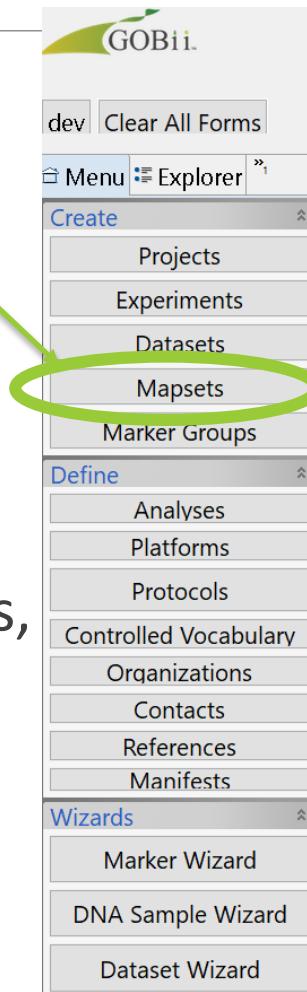
Property	Value
date_sampled	feb
division	last
genotyping_purpose	mabc
study_name	

Add New
Update
Clear Fields
Add Experiment
DNA Sample Wizard

GOBii-GDM Data Loader



- Define projects
- Map markers to multiple maps
- Define marker groups
- Define entities to support data eg platform, analyses, maps etc



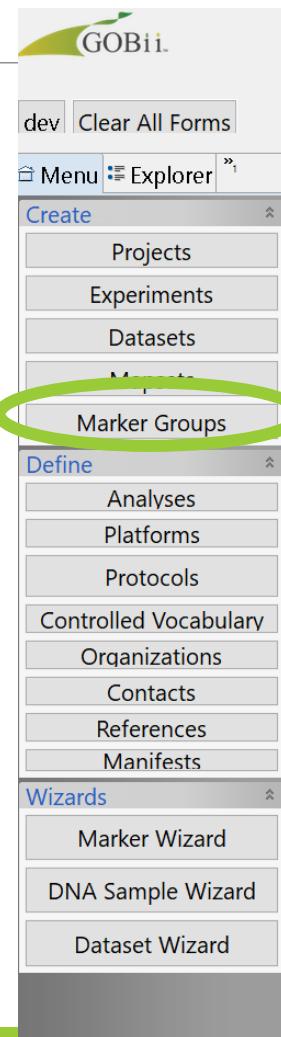
GOBII Data Load

The screenshot shows the 'Mapsets' tab of the GOBII Data Loader. On the left, a list of mapset types is shown, with 'Genetic' selected. Below this is a list of mapsets: 'codominant_test' (which is highlighted with a blue selection bar) and 'Deb_Genetic'. To the right, detailed information for the selected mapset is displayed. The 'Mapset Name' is 'codominant_test' and the 'Code' is 'Genetic_codominant_test'. The 'Mapset Description' and 'Reference' fields are empty. The 'Map type' is set to 'Genetic'. A 'Properties' table is present, with columns 'Property' and 'Value', but it is currently empty. At the bottom are buttons for 'Add New', 'Update', and 'Clear Fields'.

GOBii-GDM Data Loader



- Define projects
- Map markers to multiple maps
- Define marker groups
- Define entities to support data eg platform, analyses, maps etc



GOBII Data Loader

The screenshot shows the 'Marker Groups' tab of the GOBII Data Loader. The interface includes:

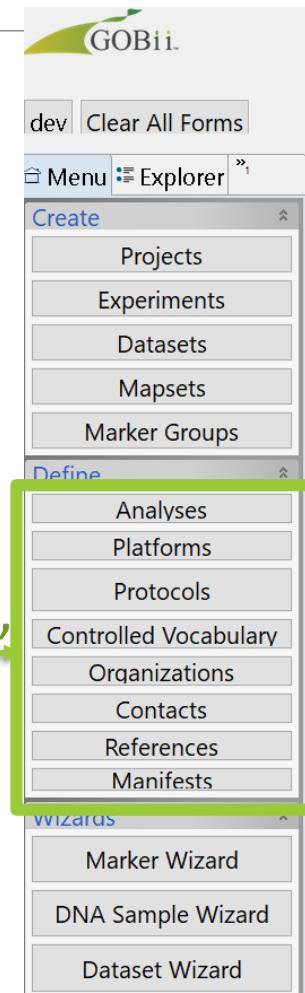
- Marker Groups:** A list of marker groups: Disease resistance SS, Disease resistance1, and test. The 'Disease resistance SS' group is selected and highlighted in blue.
- Marker Group Details:** A panel on the right shows details for the selected group:
 - *Marker Group Name: Disease resistance SS
 - Code: Disease resistance SS_Stiff stalk
 - Germplasm group: Stiff stalk
- Markers:** A table showing marker details:

Markers	Platform	Favorable allele
codom...	SSR_S...	1
codom...	SSR_S...	2
codom...	SSR_S...	2
codom...	SSR_S...	2
- Actions:** Buttons at the bottom include Import Markers, Export Markers, Add New MarkerGroup, Update MarkerGroup, and Clear Fields.

GOBii-GDM Data Loader



- Define projects
- Map markers to multiple maps
- Define marker groups
- Define entities to support data eg platform, analyses, maps etc



GOBII Data Load

Projects Mapsets

Mapset Types: Genetic

Mapsets: codominant_test Deb_Genetic

*Mapset Name: codominant_test
Code: Genetic_codominant_test

Mapset Description:

Reference:

*Map type: Genetic

Properties:

Property	Value

Add New Update Clear Fields

Export

GOBii-GDM Data Loader



- Map fields in dataset to database fields

The screenshot shows the GOBII Data Loader interface. On the left, a sidebar menu includes options like Projects, Experiments, Datasets, Mapsets, Marker Groups, Analyses, Platforms, Protocols, Controlled Vocabulary, Organizations, Contacts, References, Manifests, and Wizards. The Wizards section is highlighted with a green box and arrow, containing options for Marker Wizard, DNA Sample Wizard, and Dataset Wizard. The main area displays a 'Projects' form with fields for PI Contacts (set to Jones, Liz), Projects (listing 'test marker groups' and 'testqcwithcodominat'), and Properties (listing date_sampled: feb, division: last, genotyping_purpose: mabc, study_name). Buttons at the bottom include Export, Refresh, Add New, Update, Clear Fields, Add Experiment, and DNA Sample Wizard.

PI Contacts:
Jones, Liz

Projects:
test marker groups
testqcwithcodominat

*Project Name: test marker groups
Code: Jones, Liz_test marker groups
test marker groups

Project Description:

*PI contact: Jones, Liz

Properties:

Property	Value
date_sampled	feb
division	last
genotyping_purpose	mabc
study_name	

Add New
Update
Clear Fields
Add Experiment
DNA Sample Wizard

GOBii-GDM Data Loader



- Upload data from local machine or remote server
- Preview data
- Map fields in dataset to database fields
- Apply templates

New Wizard

Wizard :: Dataset Information

Wizard to load dataset information

Information

PL: Jones, Liz

Project: TEST_TEST

Experiment: TEST_TEST_E1

Dataset: TEST_TEST_E1_none

Dataset Type: Nucleotide_2_letter

Platform: Illumina_Infinium

Mapset: B73v3

Remote path:

/Users/yn259/programs/CGda

Local files:

Remove Selected File(s)

Saved template:

File format: GENERIC (txt)

Preview Data

Marker Position: LEFT

DNA sample Position: TOP

First Data coordinate: 1 1

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
m3111	TT																	
m3112	TT																	
m3113	CC	TC	CC	CC	CC	CC	CC	CC										
m3114	CC																	
m3115	CC	NN	CC															
m3116	TT																	
m3117	GG																	
m3118	CC																	
m3119	GG	AA	AA	GG	AA	GG	GG	AA	AA	AA	AA							
m3120	TT	CC	TT	CC	TC	CC	CC	CC	CC	NN	TT	TT	TT	TT	TT	CC	CC	TT
m3121	TT	CC	TT	CC	TC	CC	CC	CC	CC	NN	TT	TT	TT	TT	TT	CC	CC	TT
m3122	GG																	
m3123	AA	NN	NN	AA	NN	AA	NN	NN	NN	AA	AA	AA	AA	AA	NN	NN	NN	AA
m3124	TT	CC	CC	CC	TT	CC												
m3125	CC																	
m3126	GG																	
m3127	CC																	
m3128	CC																	
m3129	TT																	
m3130	GG																	
m3131	TT																	
m3132	GG																	
m3133	CC																	
m3134	AA																	
m3135	CC																	
m3136	AA																	
m3137	GG																	
m3138	GG																	

< Back Next > Cancel Finish

<http://gdm-demo.gobii.org:8081/gobii-portal/>

GOBii Marker Portal



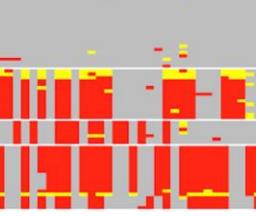
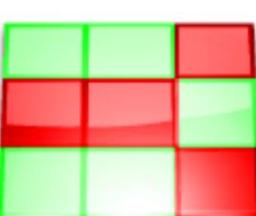
GOBii Marker Portal

Documentation

GDM

- Timescope
- File Browser
- GDM Data QC
- GS Pipeline
- Galaxy Documentation
- Flapjack
- TASSEL 5

Web Apps Desktop Apps Links

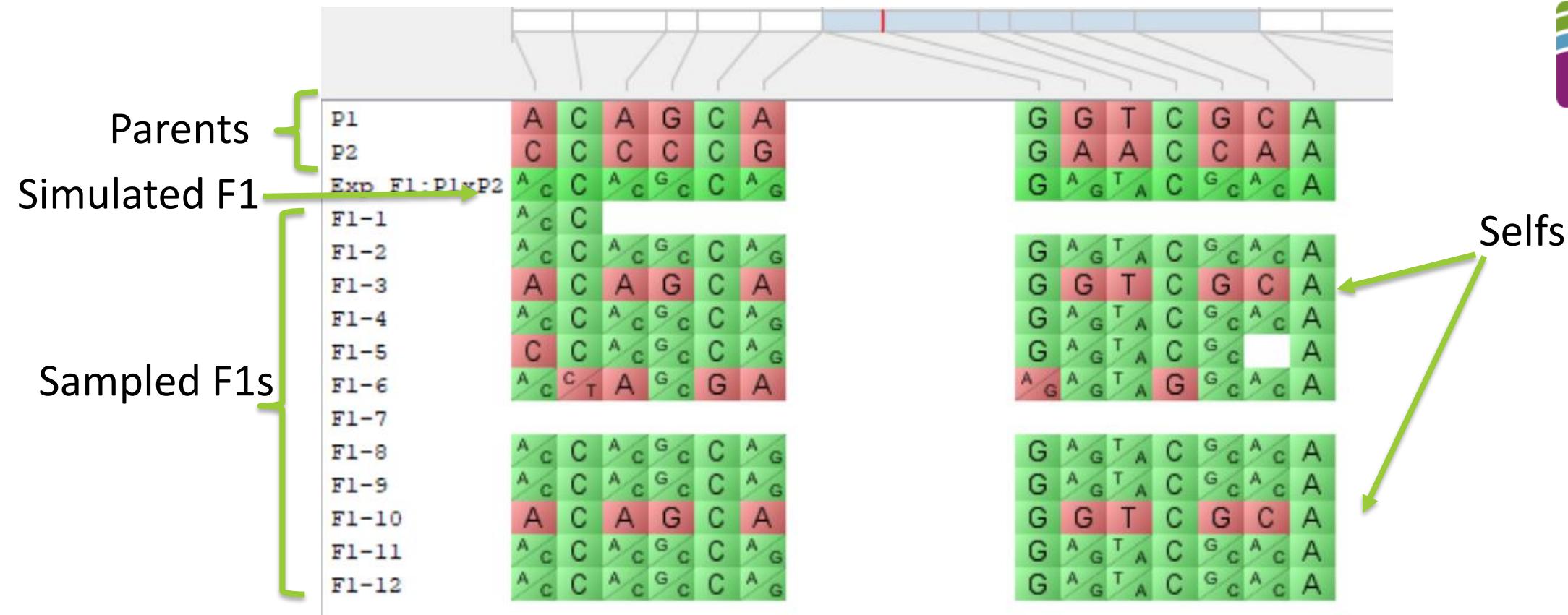
 EXTRACT dev	 TIMESCOPE GOBii DATABASE TOOL	 ownCloud	 KDCompute Data Analysis	 Galaxy EiB GS pipeline Demo	 Portainer
 Haplo Tool	 Flapjack Bytz	 F1 PED VER PedVer			

Open-source



- Request access to a demo <http://cbsugobii05.biohpc.cornell.edu/wordpress/index.php/gdm-access/>
- Public documentation <https://gobiiproject.atlassian.net/wiki/spaces/GD/overview>
- Code to Genomics Data Manager <http://gobiin1.bti.cornell.edu>
- Flapjack <https://ics.hutton.ac.uk/flapjack/download-flapjack/>
- Galaxy GS pipeline <http://galaxy-demo.excellenceinbreeding.org>
 - New: <http://cropgalaxy.excellenceinbreeding.org/>
- DArTView: <https://software.kddart.com/kdxplore/dartview/>

GOBii Tools: Pedigree Verification in Flapjack



- The ability to use higher quality seed can improve yields by two to three times

Gary Vaughan-Smith 2017

What hinders tool adoption?



- **Integration: easily importing data into application**
- **Formatting data**
 - Genotyping projects contain multiple populations that have to be split
 - Genotyping projects contain multiple replicates of parents that need consensus calling
 - Populations and parents across different genotyping datasets

Import GOBii data into Flapjack using BrAPI



File Edit View Visualization Analysis Data Help

New project Ctrl+N Import Data Find Genotypes Chromosomes

Open project... Ctrl+O Chromosome: 1 1,194 lines, 97 markers, length: 96

Save project Ctrl+S

Save project as...

Optimize project

Import data... (highlighted with a red oval)

Import data (via BrAPI)... (highlighted with a red oval)

Quick export...

Recent projects >

Exit

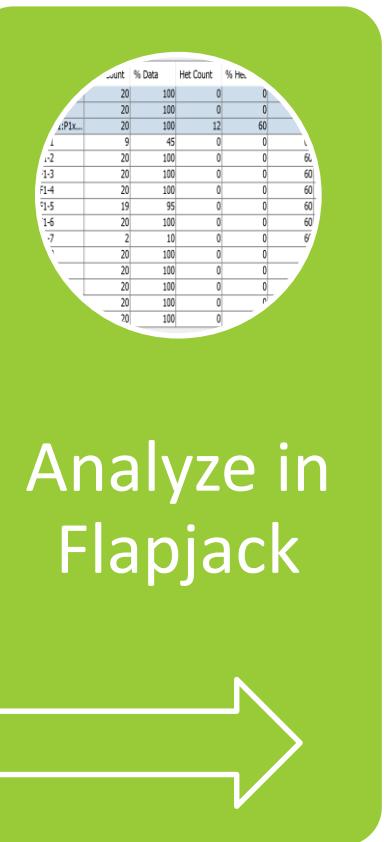
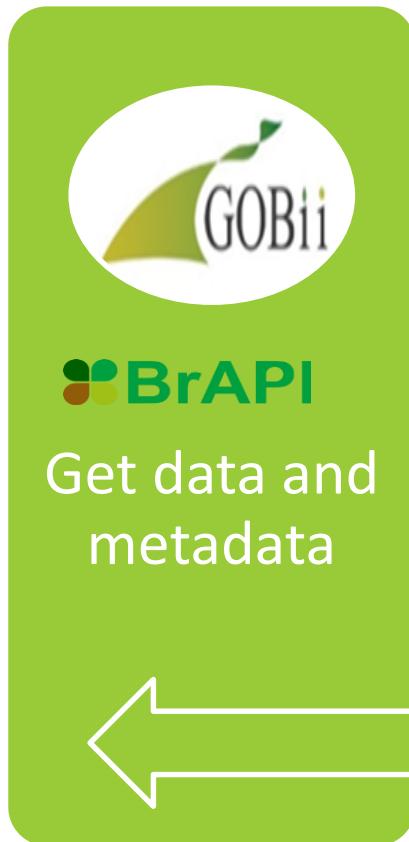
TTDIL Data
Default View

Overview

2019QC_15 2019QC_16 2019QC_17 2019QC_18 2019QC_19 2019QC_20 2019QC_21 2019QC_22

The screenshot shows the Flapjack software interface. The menu bar includes File, Edit, View, Visualization, Analysis, Data, and Help. The 'File' menu is open, showing options like New project, Open project..., Save project, Save project as..., Optimize project, Import data..., Import data (via BrAPI)... (both highlighted with a red oval), Quick export..., Recent projects, and Exit. Below the menu is a toolbar with icons for Import Data, Find, Genotypes, and Chromosomes. The main workspace displays a genotype matrix for chromosome 1, showing 1,194 lines (samples) and 97 markers. The matrix consists of a header row with marker names and a body of data where each cell represents a genotype (e.g., AA, GG, NN). A red oval highlights the 'Import data...' and 'Import data (via BrAPI)...' options in the 'File' menu.

Consensus calling and splitting tool



GOBii Tools: Batch Analysis, Flapjack



Batch analysis of four pops

File Edit View Visualization Analysis Data Help

New Project Open Project Import Data Find Genotypes Chromosomes

Data Sets (5)

- output 74x9
 - Trait Data
 - Default View
 - PedVer F1s View 1
 - PedVerF1s Results
- 3 10x9
 - Trait Data
 - Default View
 - PedVer F1s View 1
 - PedVerF1s Results
- 4 10x9
 - Trait Data
 - Default View
 - PedVer F1s View 1
 - PedVerF1s Results
- 1 10x9
 - Trait Data
 - Default View
 - PedVer F1s View 1
 - PedVerF1s Results
- 2 10x9
 - Trait Data
 - Default View
 - PedVer F1s View 1
 - PedVerF1s Results

Chromosome: All Chromosomes 11 lines, 9 markers, length: 8

IR12A282* 100 33.33 Parent 1 C G C G T G G G A
IR08A176* 100 33.33 Parent 2 T G T A G G C G T
Exp F1:IR12A282*xIR08A176* 100 100 Expecte... C T G C T A G G T G C G G A T
A2018DSC48_1_2101 100 100 True F1 C T G C T A G G T G C G G A T
A2018DSC48_2_2102 100 44.44 No deci... C G C G T G C G G A
A2018DSC48_3_2103 100 77.78 True F1 C T G C T A G G T G C G G A T
A2018DSC48_4_2104 100 100 True F1 C T G C T A G G T G C G G A T
A2018DSC48_5_2105 100 100 True F1 C T G C T A G G T G C G G A T
A2018DSC48_6_2106 88.89 37.5 No deci... C G C G T G G G G
A2018DSC48_7_2107 100 100 True F1 C T G C T A G G T G C G G A T
A2018DSC48_8_2108 22.22 100 True F1 C T G C T A G G T G C G G A T

GOBii Tools: Threshold Setting, Flapjack



Threshold Settings

Percent Parental Heterozygosity:



15

Percent F1 Heterozygosity:



20

Percent Heterozygosity:



20

Percent Match to F1:



71

OK

Cancel

Help

GOBii Tools: Pedigree Verification, Flapjack

Parents {
IR12A282*
IR08A176*

Simulated F1 → Exp F1: IR12A282*xIR08A176*
A2018DSC48_1_2101
A2018DSC48_2_2102
A2018DSC48_3_2103
A2018DSC48_4_2104
A2018DSC48_5_2105
A2018DSC48_6_2106
A2018DSC48_7_2107
A2018DSC48_8_2108

100	33.33	Parent 1
100	33.33	Parent 2
100	100	Expecte...
100	100	True F1
100	44.44	No deci...
100	77.78	True F1
100	100	True F1
100	100	True F1
88.89	37.5	No deci...
100	100	True F1
22.22	100	True F1

C	G	C	G	T	G	G	G	A
T	G	T	A	G	G	C	G	T
c	T	G	C	T	A	G	T	G
c	T	G	C	T	A	G	T	G
C	G	C	G	T	G	c	G	A
C	T	G	C	T	A	G	G	T
c	T	G	C	T	A	G	G	A
c	T	G	C	T	A	G	G	T
C	G	C	G	T	G	c	G	A
C	T	G	C	T	A	G	T	G
c	T	G	C	T	A	G	T	G
C	G	C	G	T	G	G	G	A
C	T	G	C	T	A	G	T	G
c	T	G	C	T	A	G	T	G

Selfs

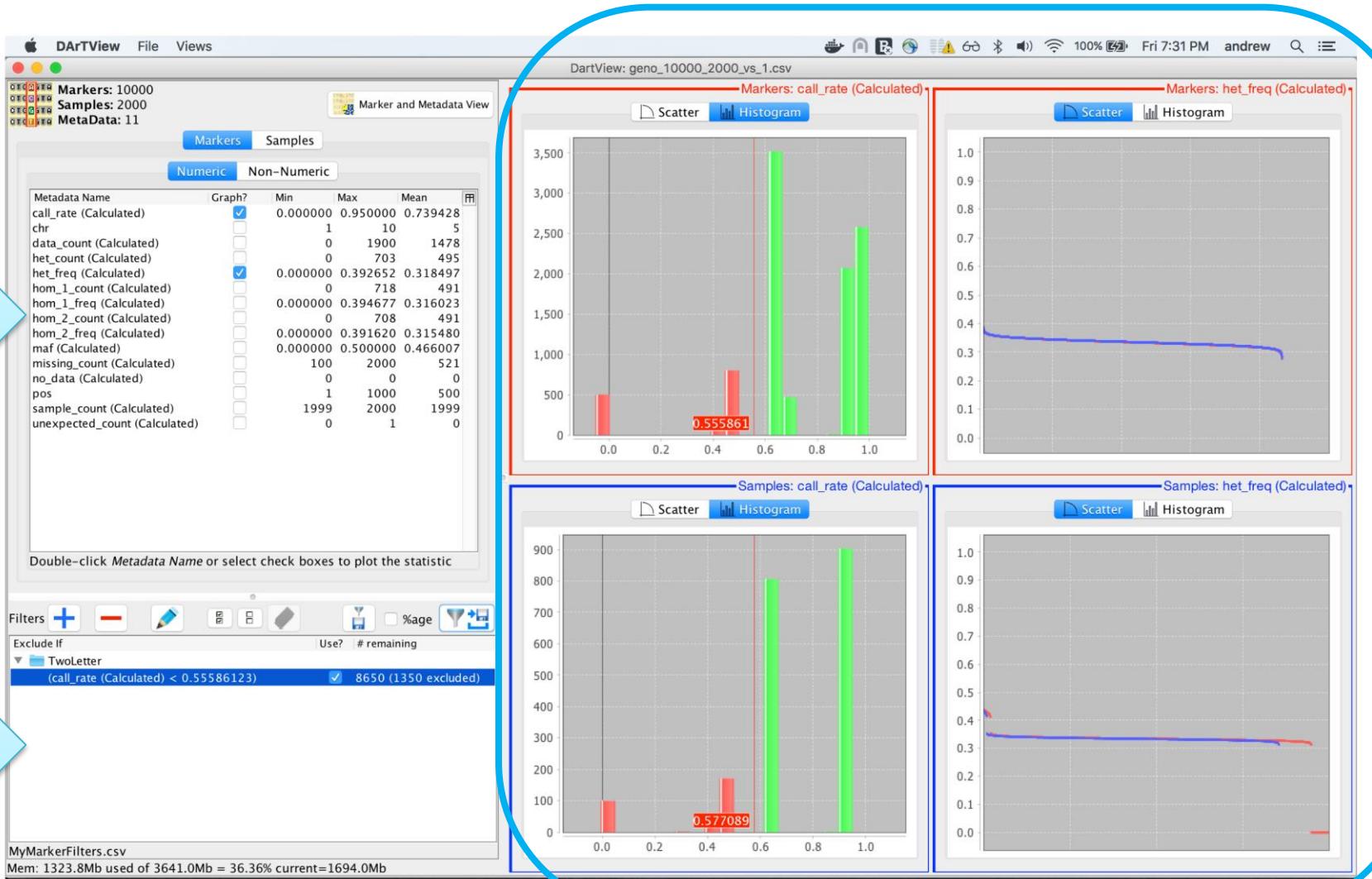
Decision on
selection

GOBii Genotyping Data QC



Select traits to view

Filter data

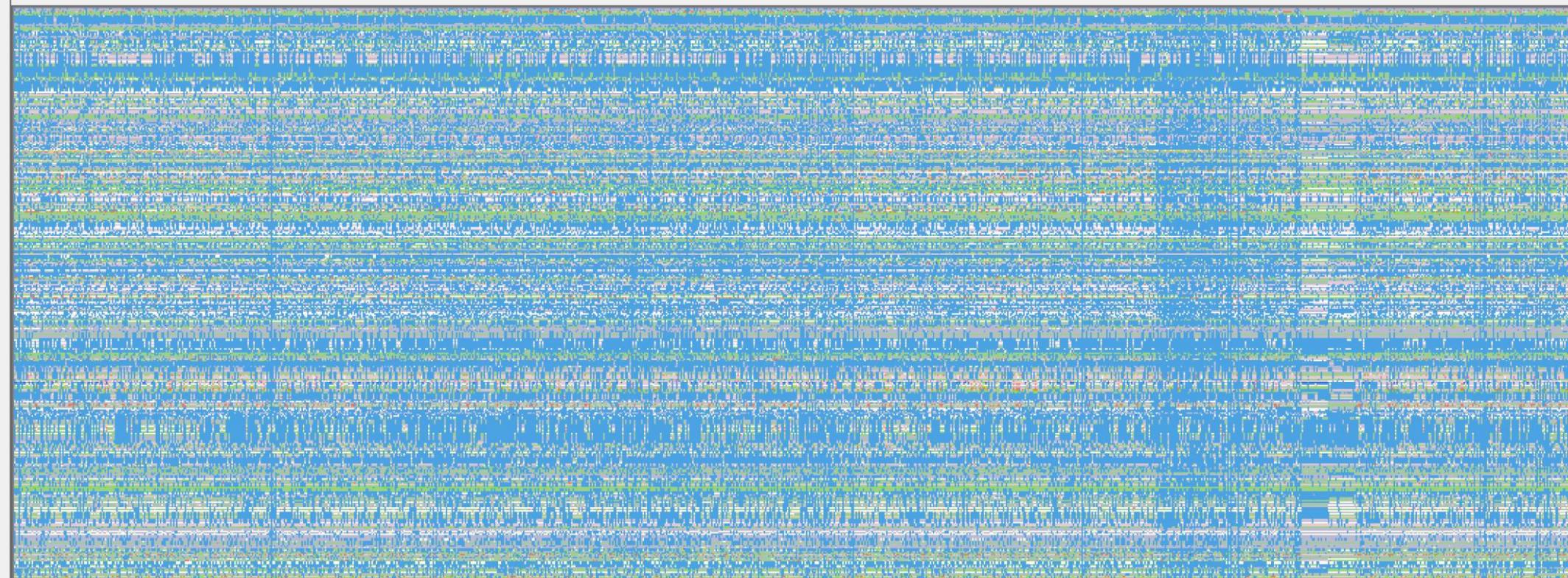


View distributions for selected traits as scatter plots or histograms

Metadata Filters: Apply Clear

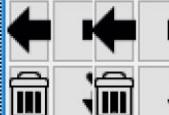


call_rate (Calculated)



pos

call_rate (Calculated)



Sample: SEEDGWA2451:MRG:4:250070060

Not over a Sample

