GOBii™

Genomic Open-source Breeding informatics initiative

Bill & Melinda Gates foundation

Cornell University Founded A.D. 1865

Institute of Biotechnology

BTI Boyce Thompson Institute

USDA

CIMMYT International Maize and Wheat Improvement Center

ICRISAT

IRRI

The James Hutton Institute

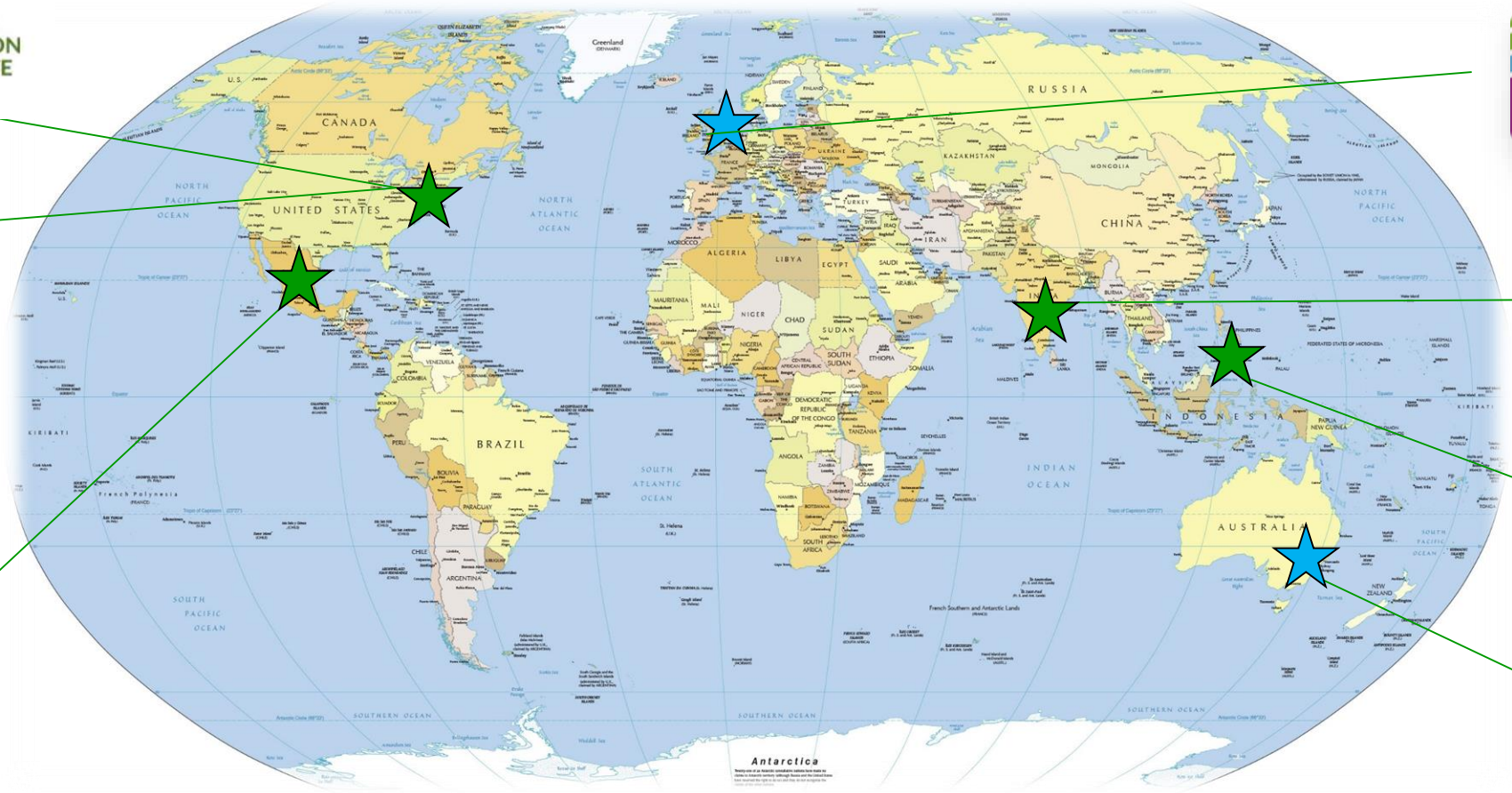
Diversity Arrays Technology
IT'S A BULLSEYE WITH DArT

Genomic Open-source Breeding informatics initiative

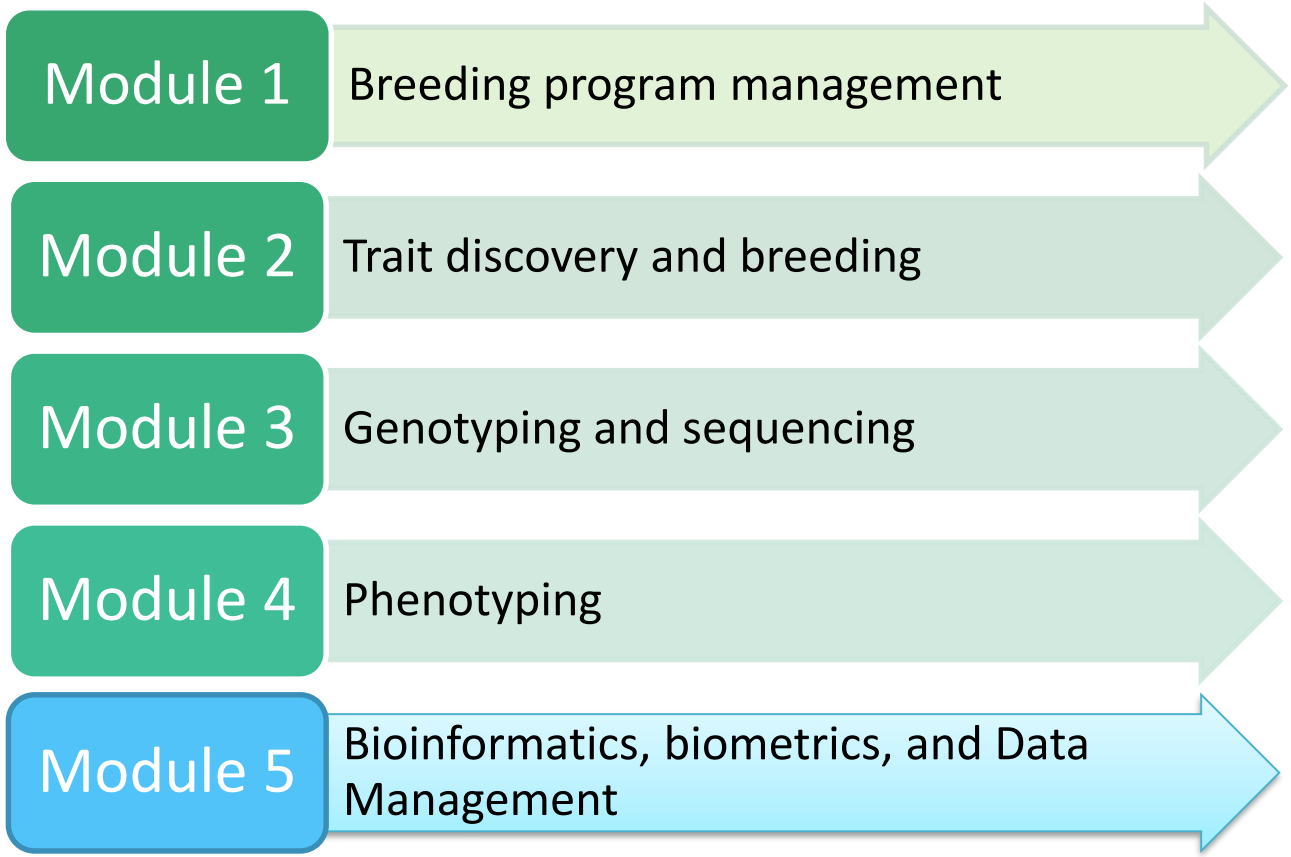
To transform breeding by enabling the implementation of genomic and marker-assisted selection as part of routine breeding programs in developing countries

- ✓ 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



Excellence in Breeding Platform



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

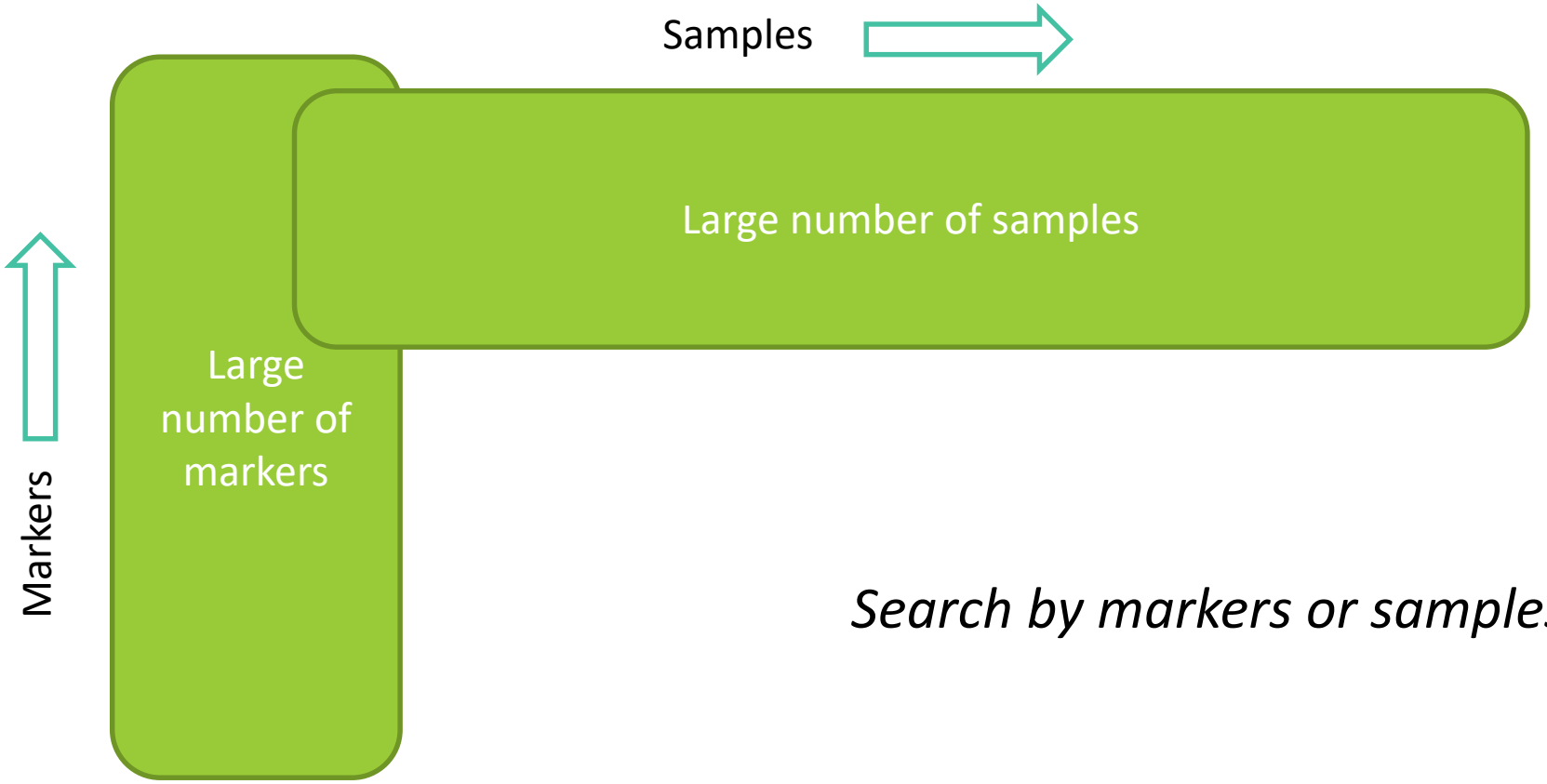
Genotype Management (GOBii)

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

Trial Analytics

Advancement Decision

Genotyping Data Management – Dimensionality Issues



Search by markers or samples

GOBii Genomic Data Management

Designed with extract performance in mind

LOAD

LOAD

IUPAC, SNP, Indel, diploid, tetraploid, codominant, dominant, SSRs, Vcf,.txt, .csv, .hmp

EXTRACT

EXTRACT

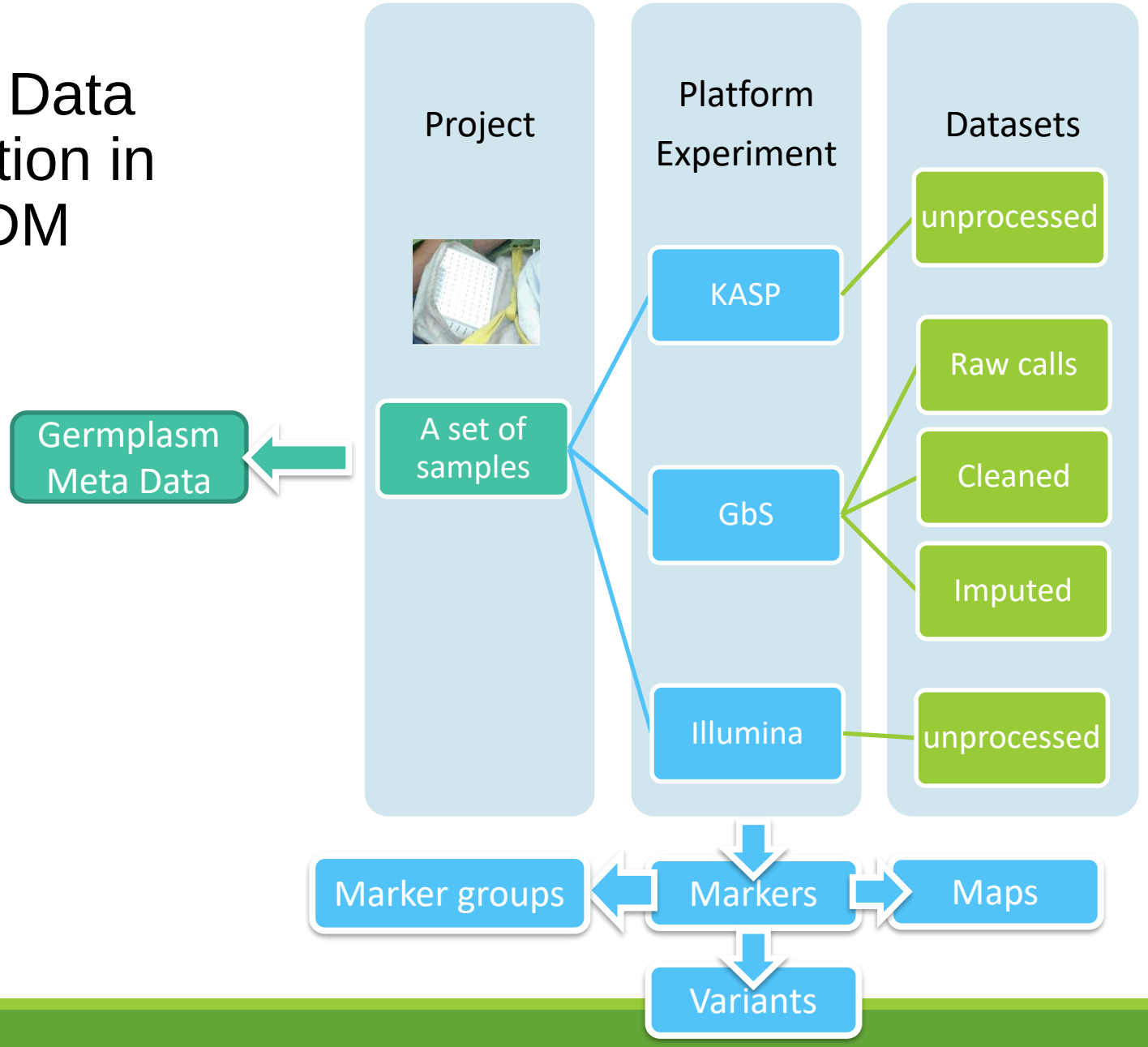
Flapjack
Hapmap

Web services APIs

BrAPI

Database

Levels of Data Organization in GOBii-GDM

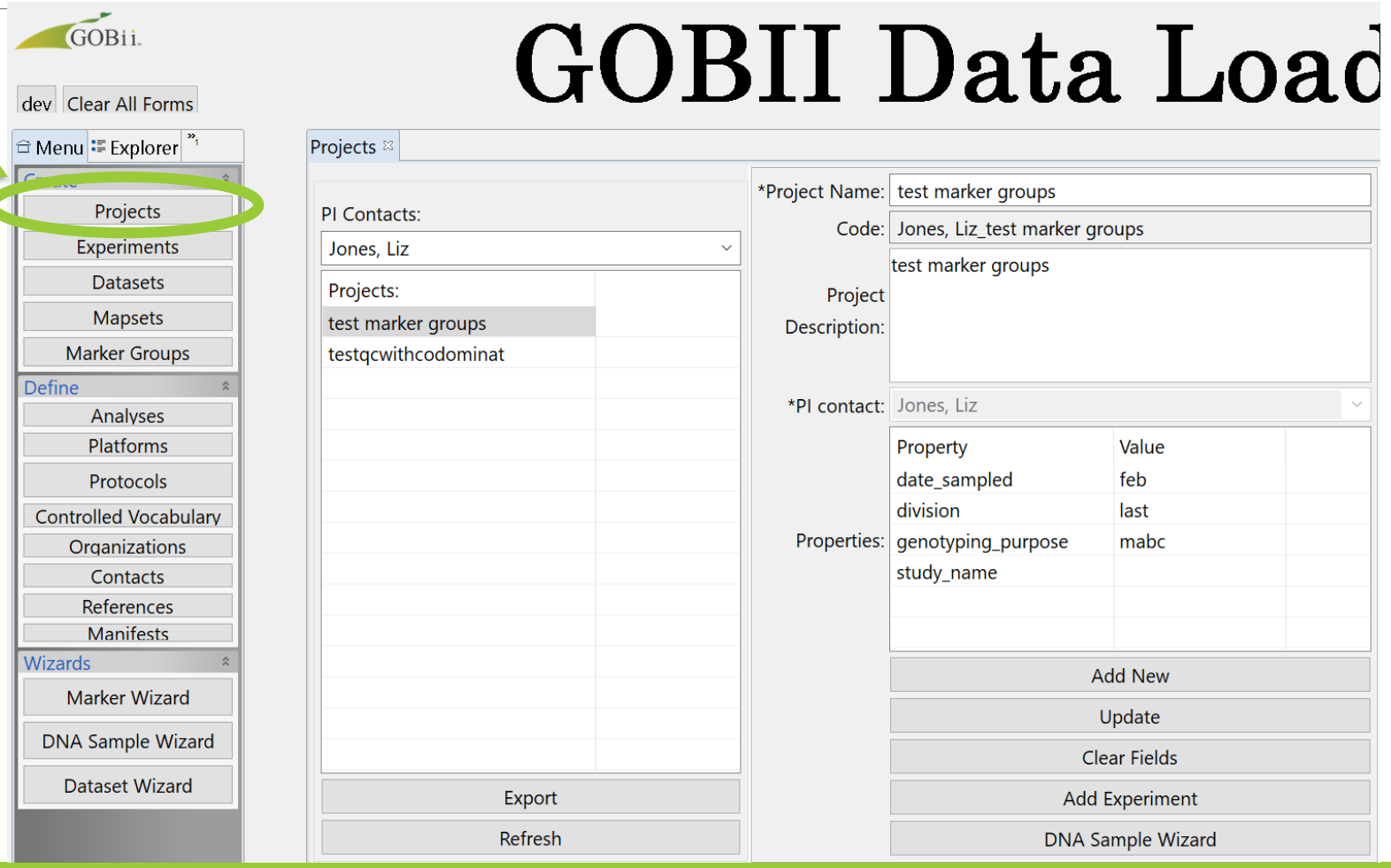


GOBii Germplasm levels

- 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)

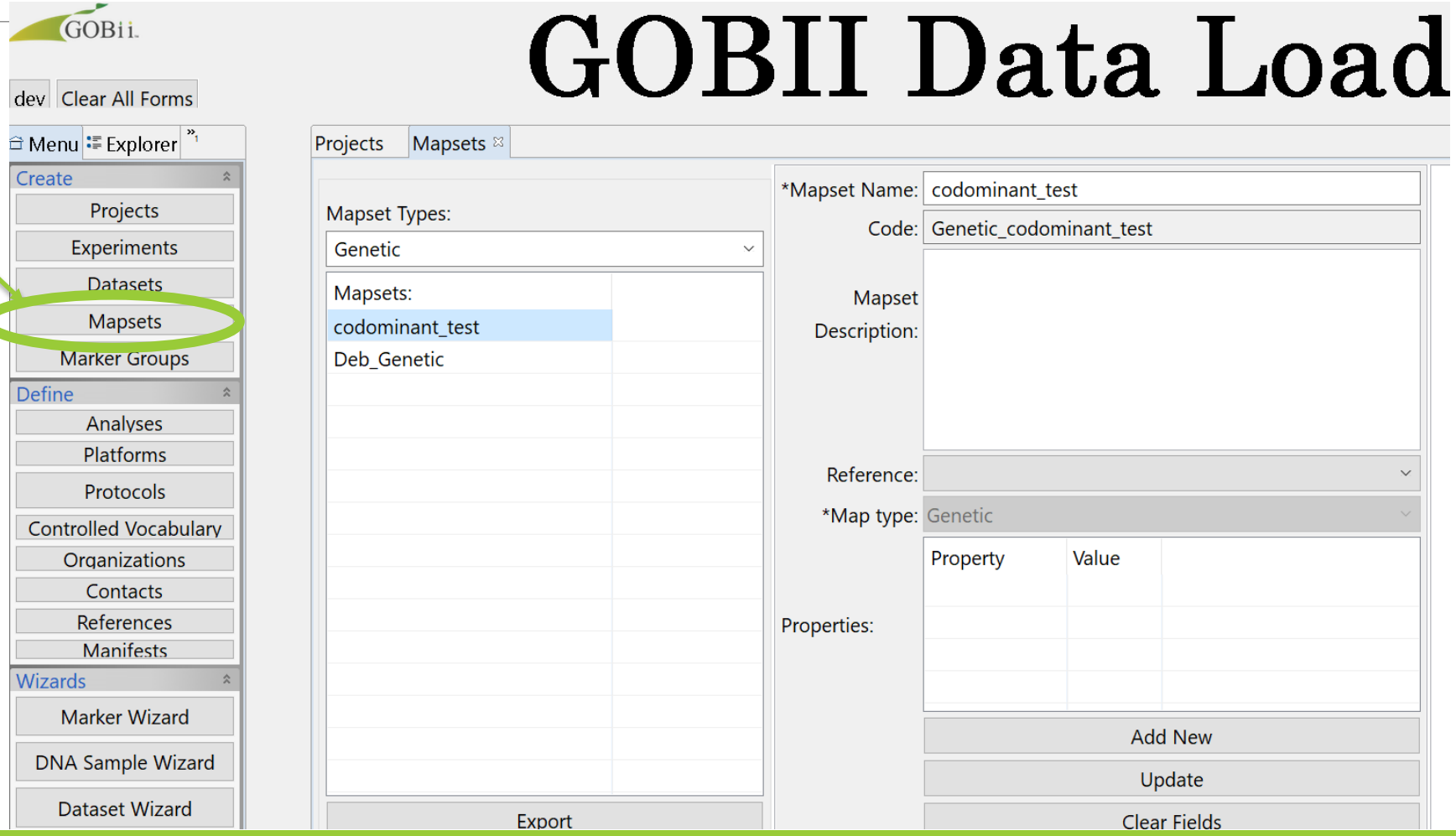
GOBii-GDM Data Loader

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



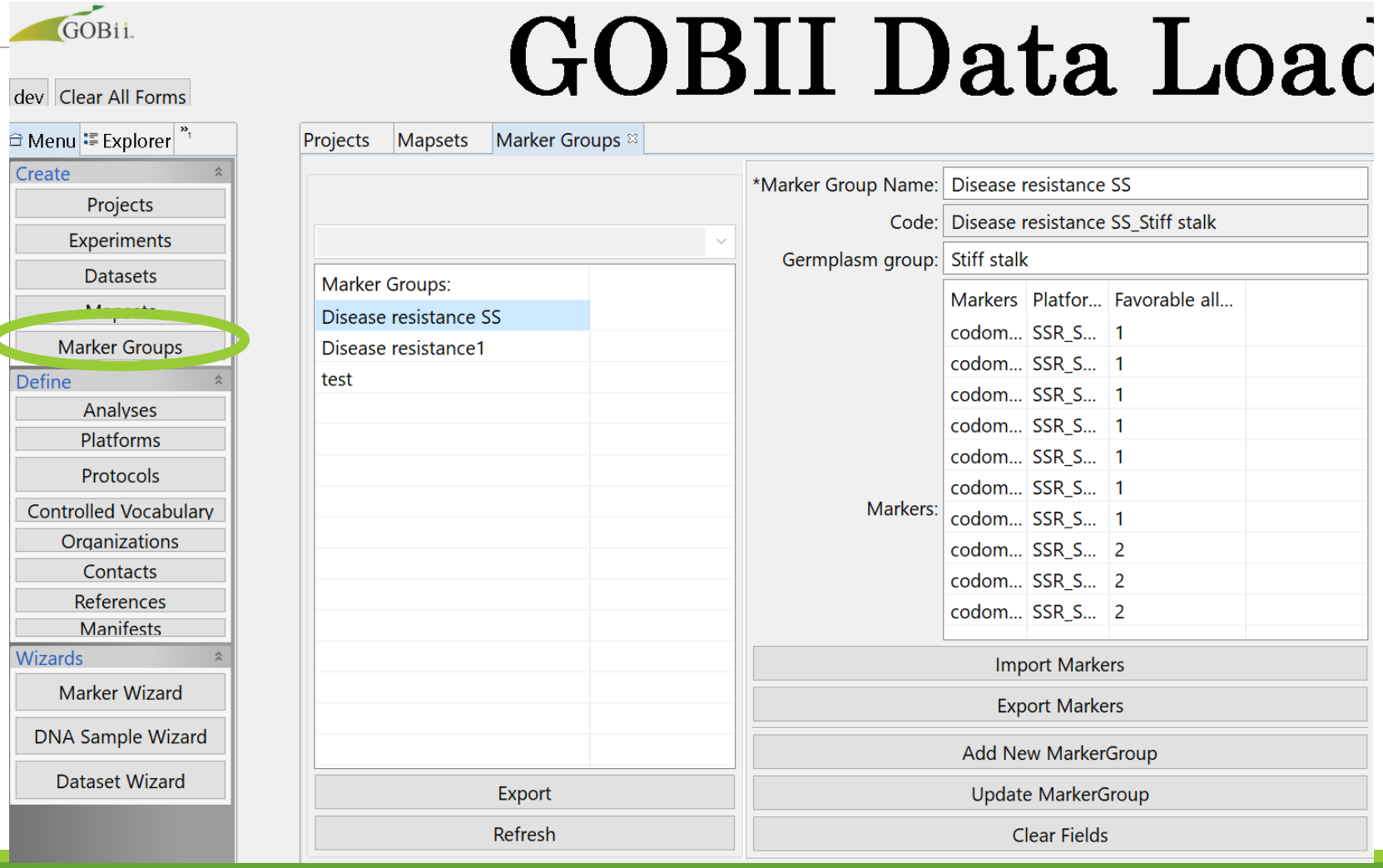
GOBii-GDM Data Loader

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



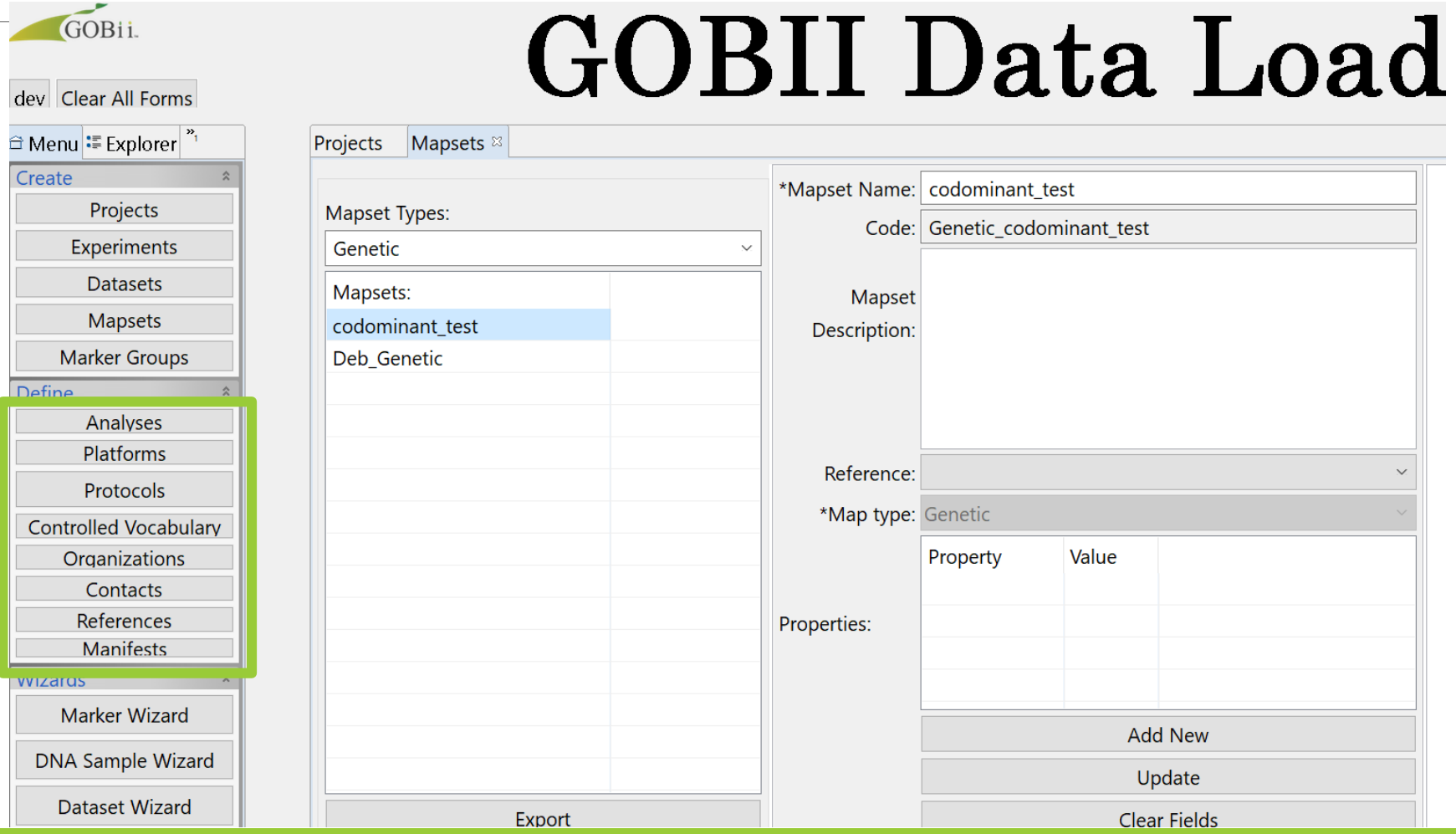
GOBii-GDM Data Loader

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



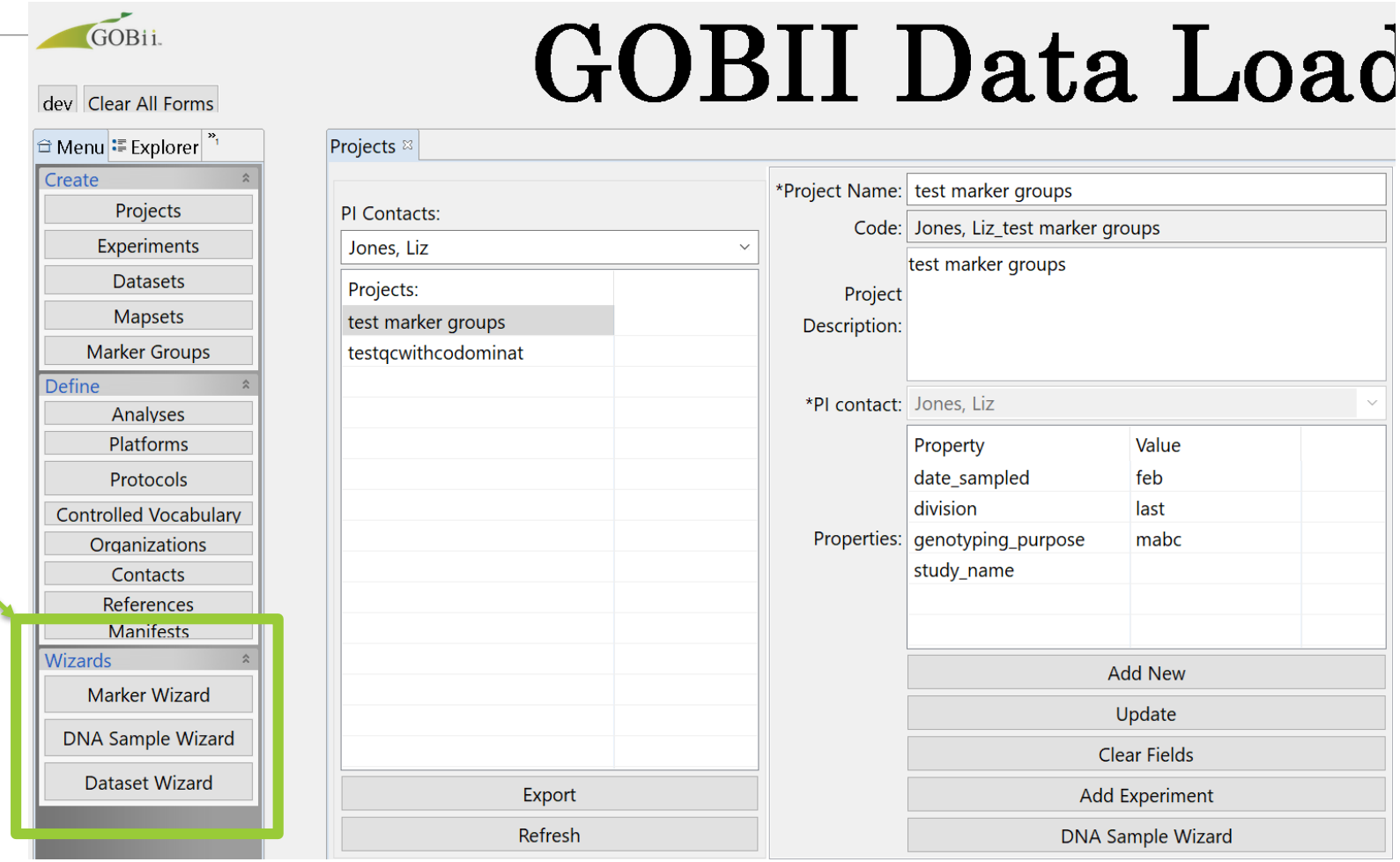
GOBii-GDM Data Loader

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



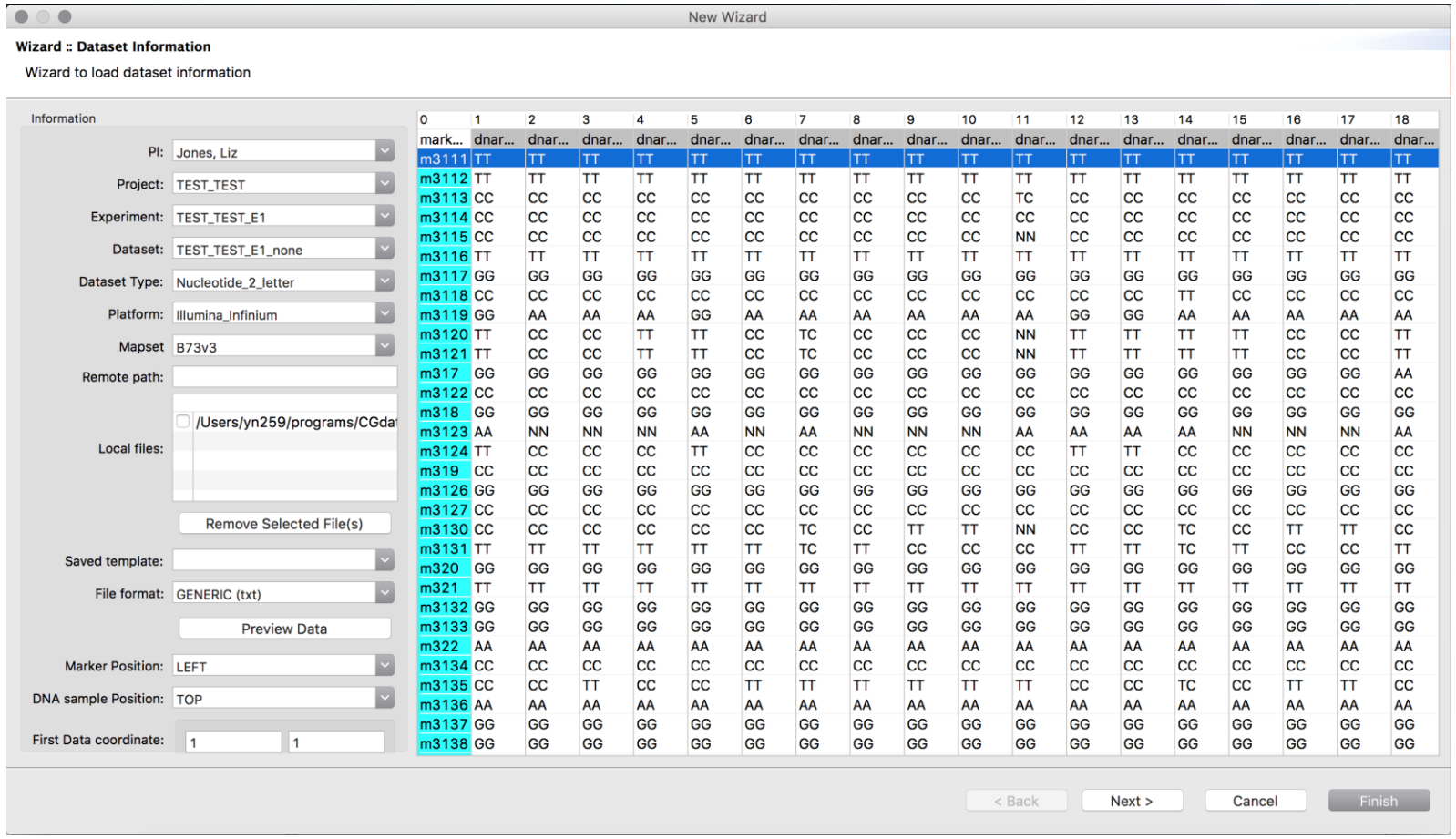
GOBii-GDM Data Loader

- Map fields in dataset to database fields



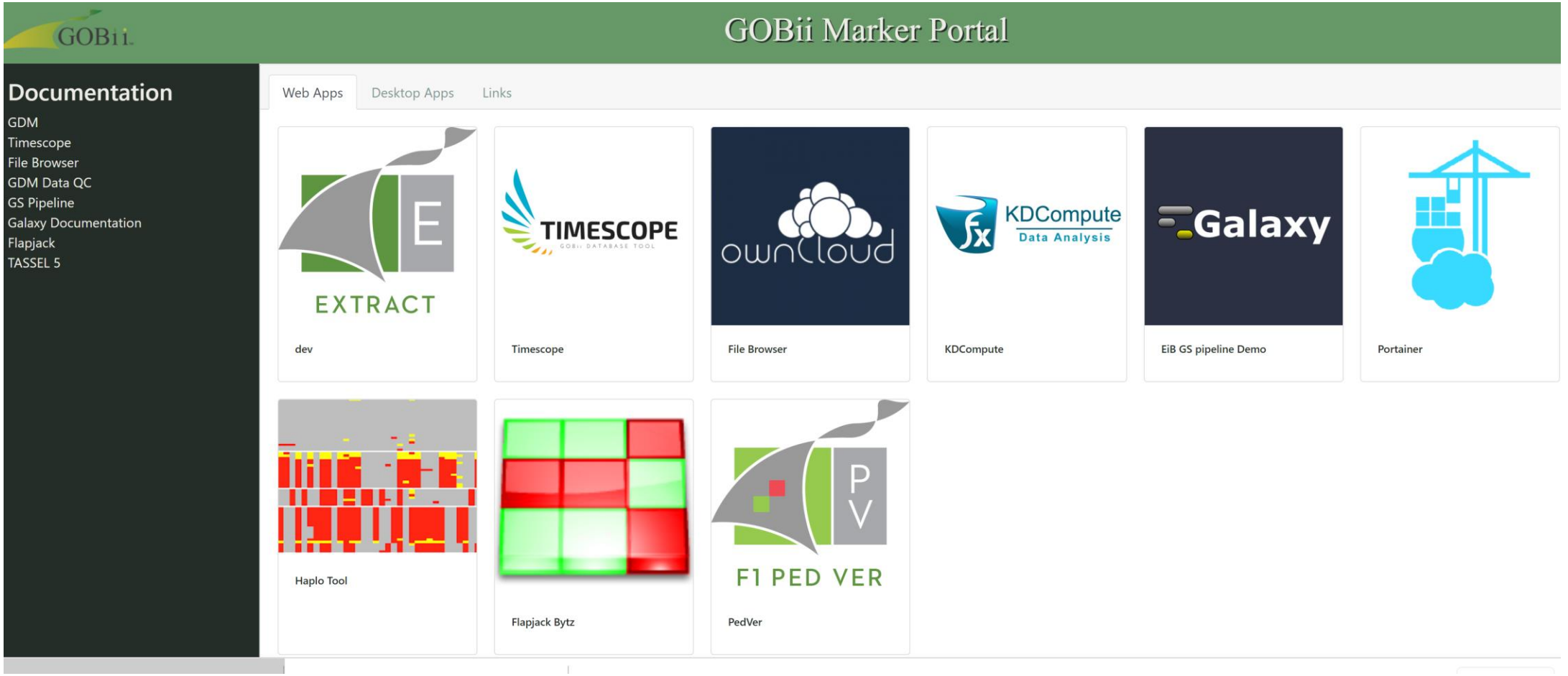
GOBii-GDM Data Loader

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



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

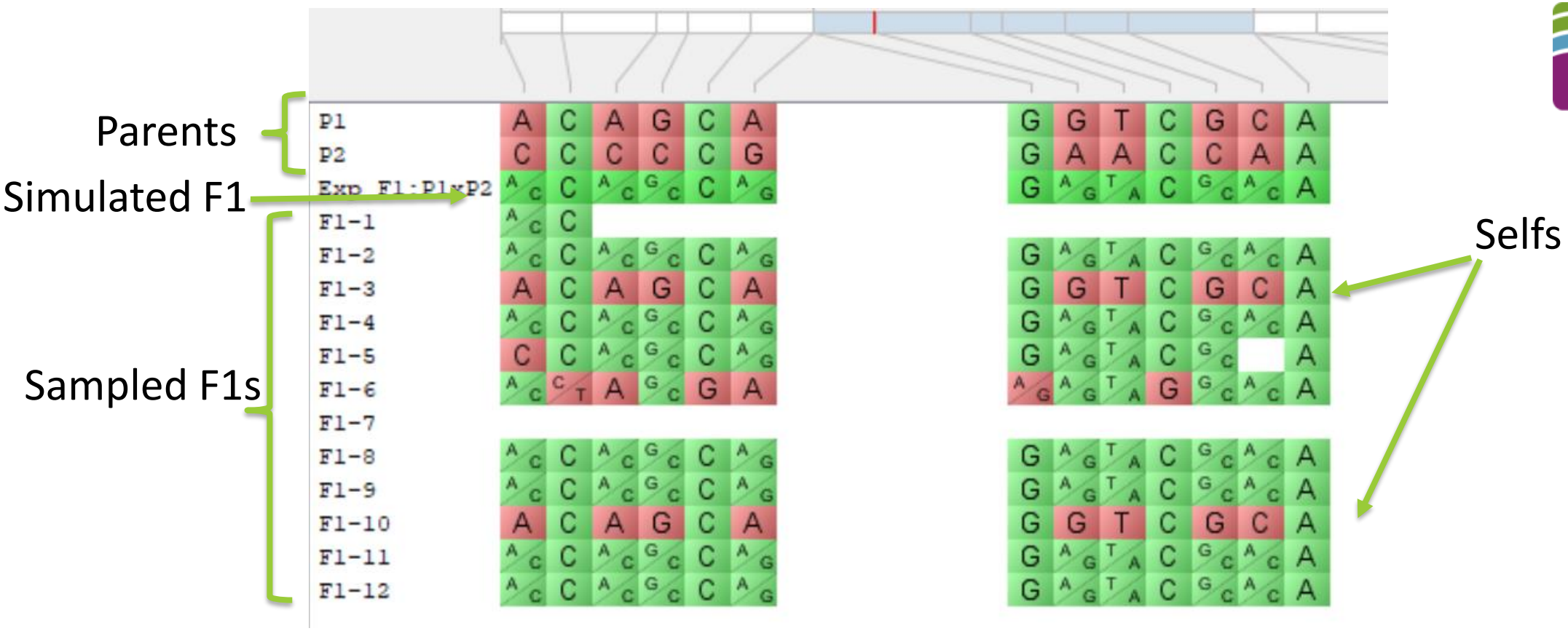
GOBii Marker Portal



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



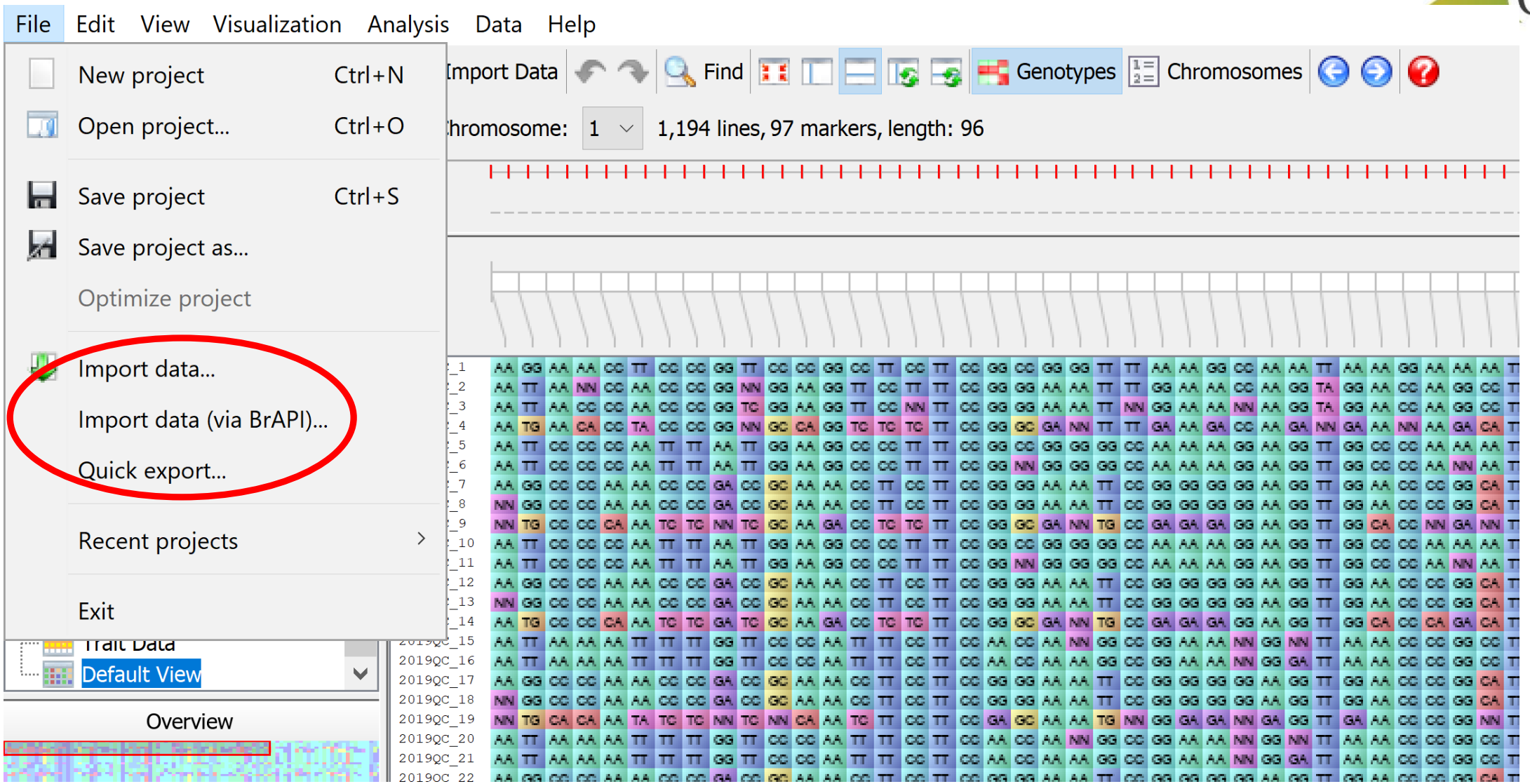
- The ability to use higher quality seed can improve yields by two to three times

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



Consensus calling and splitting tool

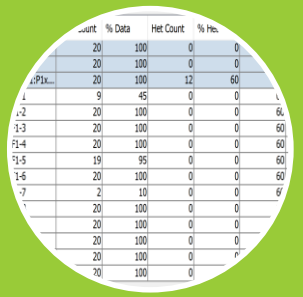
Get data and metadata

Consensus calling for replicated samples

Split populations

Create Flapjack file

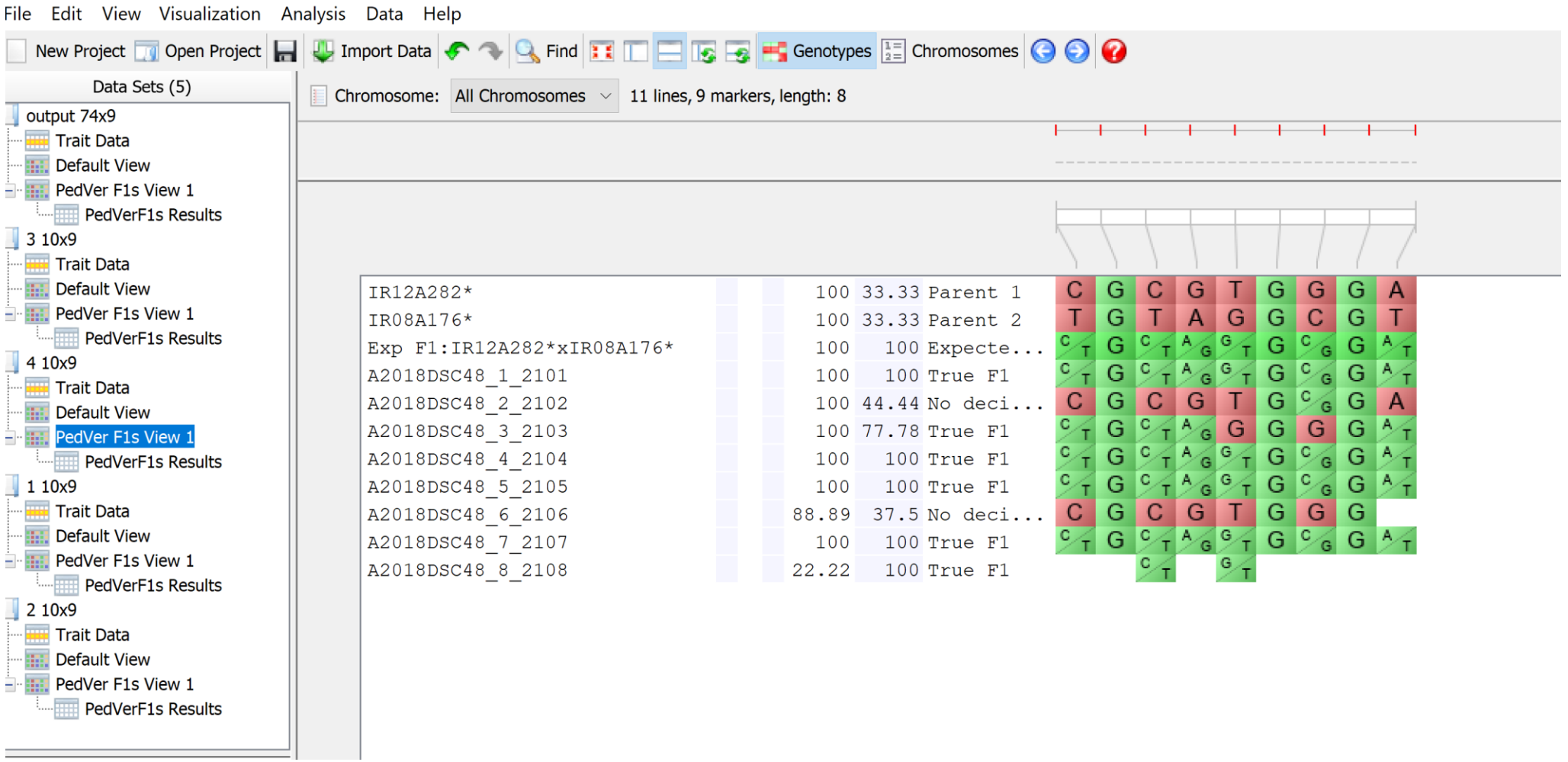
Analyze in Flapjack



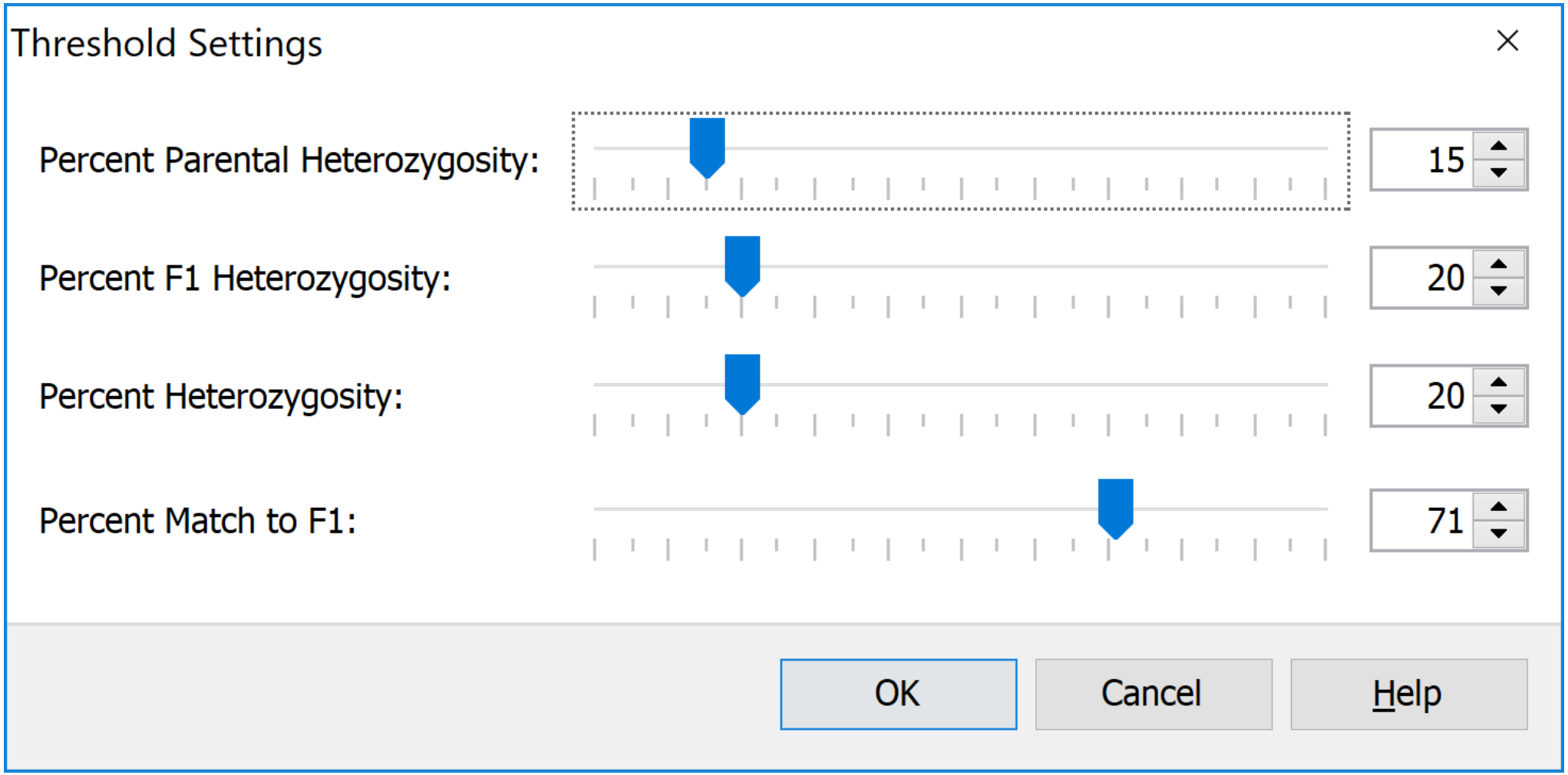
http://gdm-demo.gobii.org:8081/gobii-portal/

GOBii Tools: Batch Analysis, Flapjack

Batch analysis of four pops



GOBii Tools: Threshold Setting, Flapjack



GOBii Tools: Pedigree Verification, Flapjack

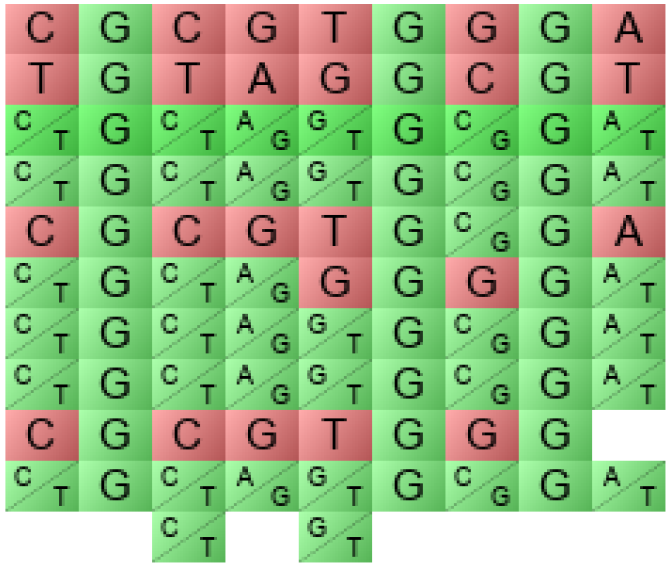
Parents

Simulated F1

Sampled F1s

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

Decision on selection

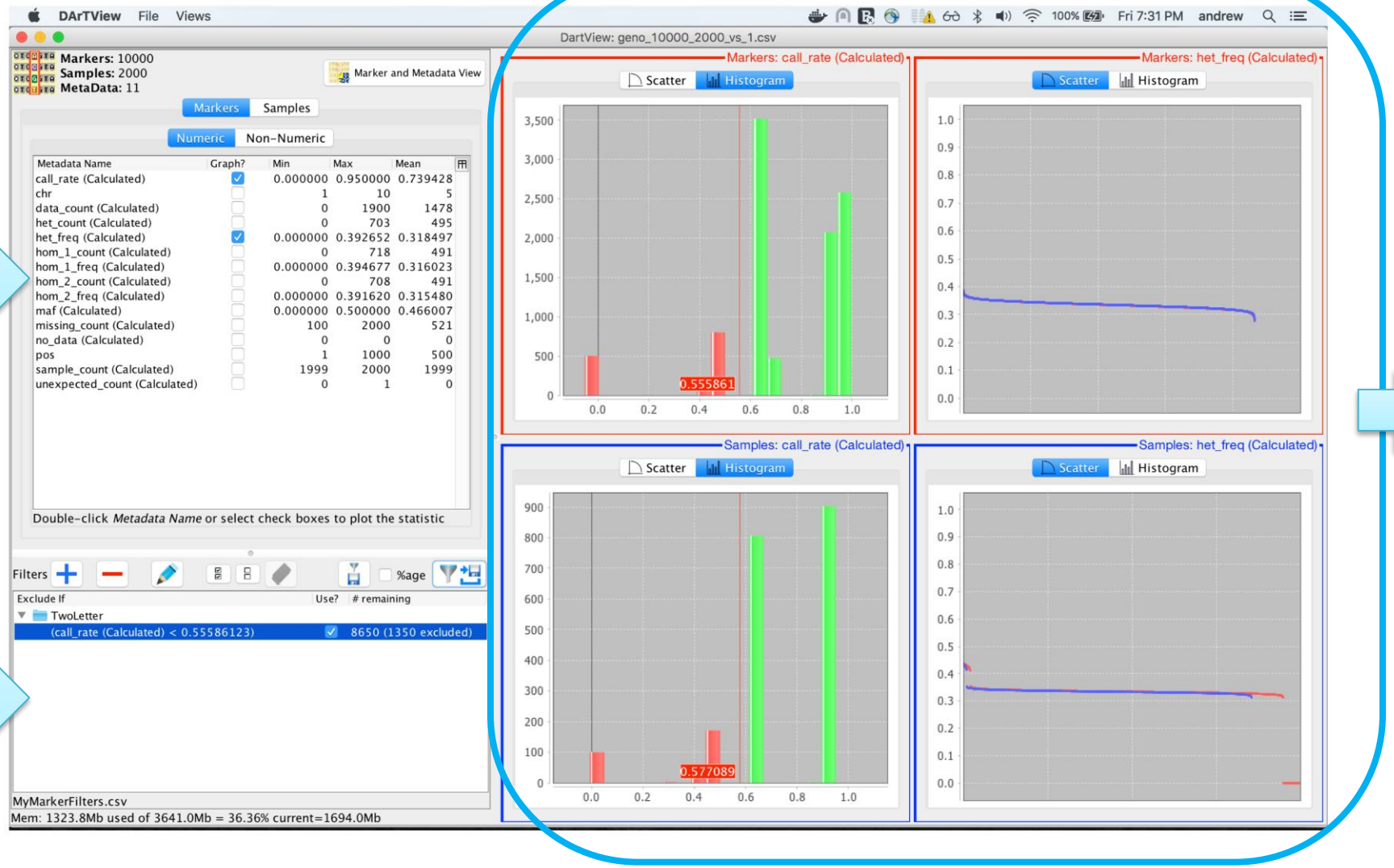


Selfs

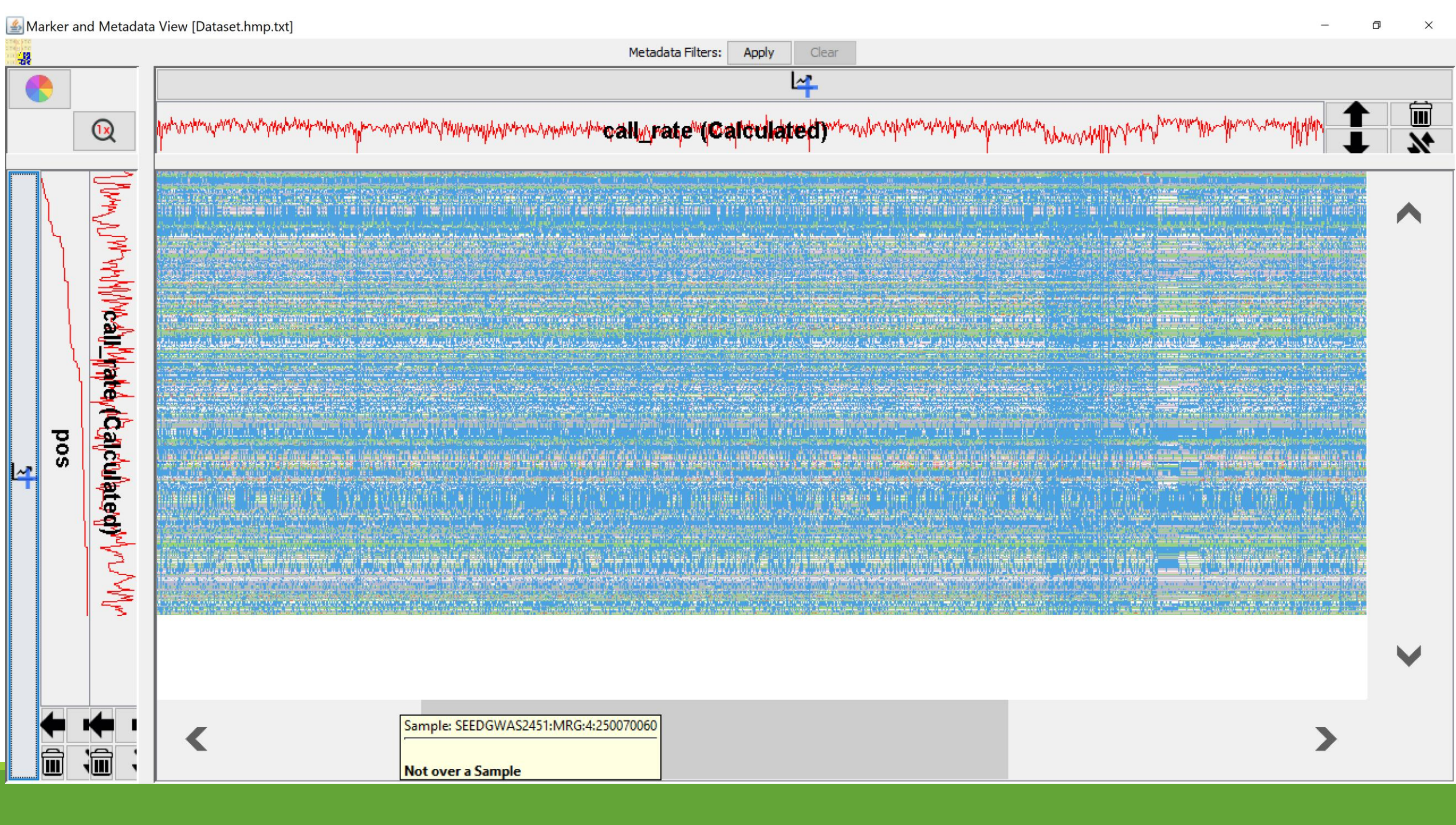
GOBii Genotyping Data QC

Select traits to view

Filter data



View distributions for selected traits as scatter plots or histograms


Marker and Metadata View [Dataset.hmp.txt]
Metadata Filters:
Apply
Clear
call_rate (Calculated)
pos
call_rate (Calculated)
Sample: SEEDGWAS2451:MRG:4:250070060
Not over a Sample