BREEDBASE:
a digital ecosystem for modern plant breeding

Christine Nyaga
Ph. D. (Plant Breeding and Genetics)
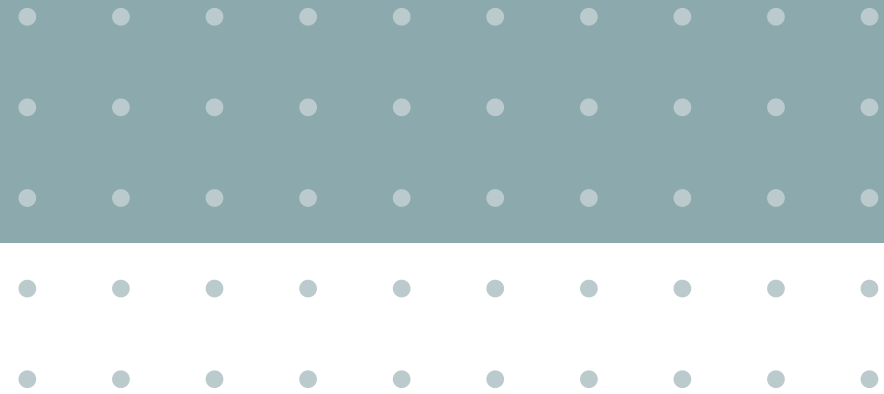Cornell University

ASPB 2023

AFFILIATION

Boyce Thompson Institute,
Cornell Univesity

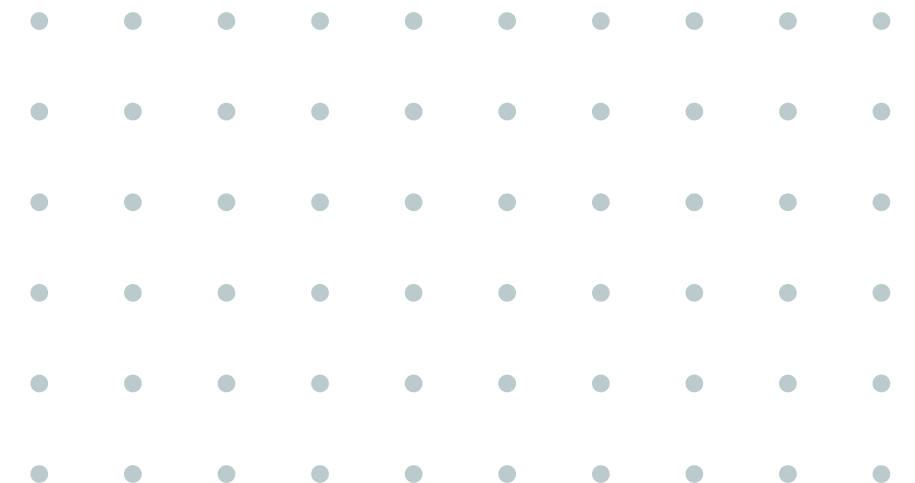# 01.

PROBLEM VS SOLUTION

Breedbase

PROBLEM

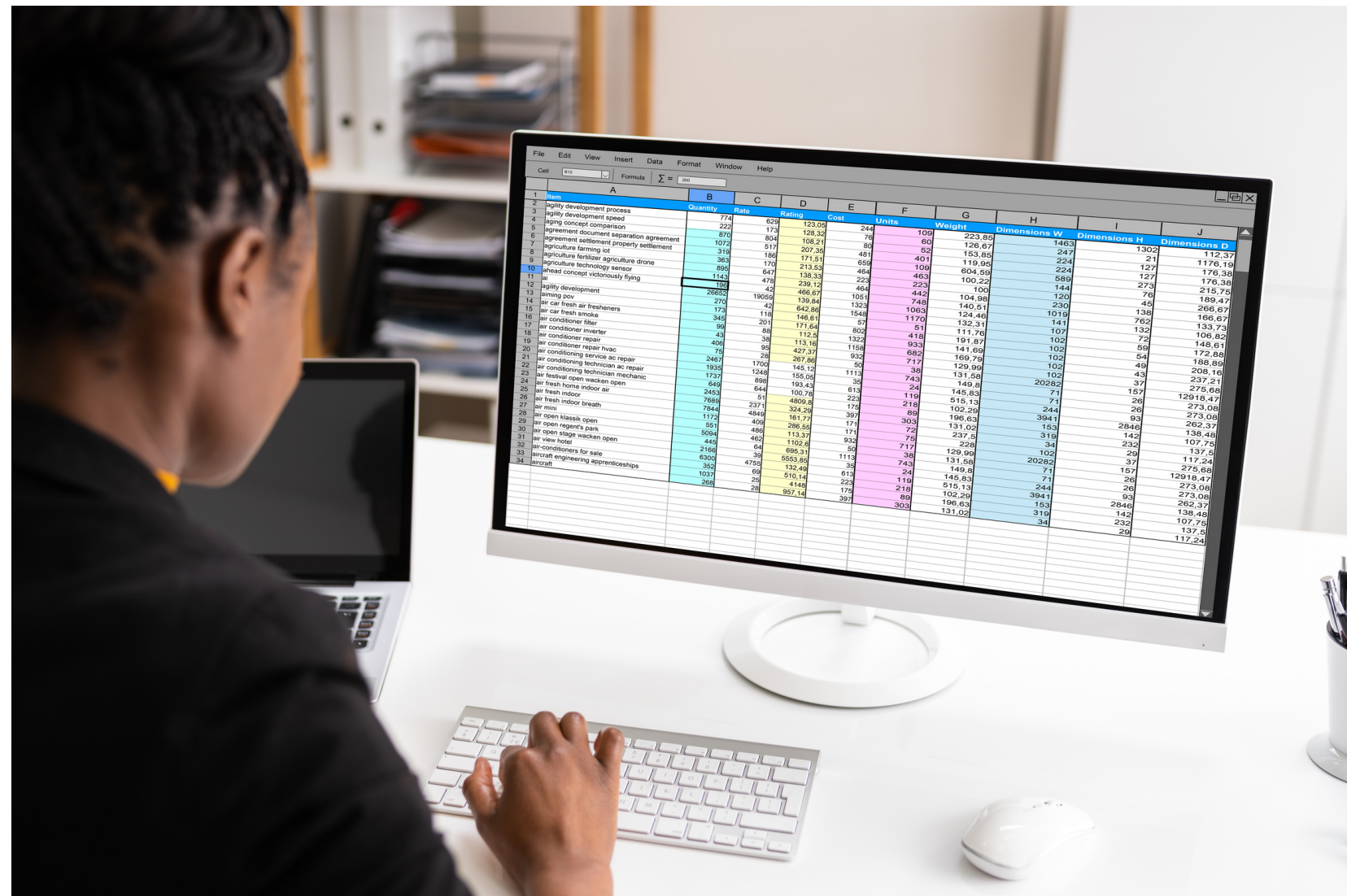Advances in modern plant breeding result to **high data volume**

Require **efficient** data management, quality control and analytics

Complex and expensive process

TRADITIONAL APROACHES

1. Spreadsheets

2. Book Records

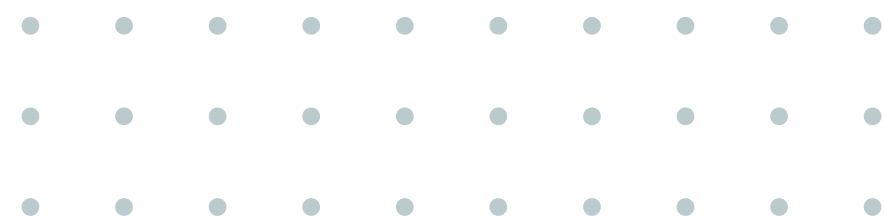**01.** Difficulties in merging data across different spreadsheets

02. Lack of centralized storage; often stored on personal computers and laptops

03. Difficulties in visualizing or analyzing data across spreadsheets

04. Limited backup strategies and little recourse if accidental data loss occurs

PROBLEMS WITH TRADITIONAL APROACHES

02.

BREEDBASE

BREEDBASE
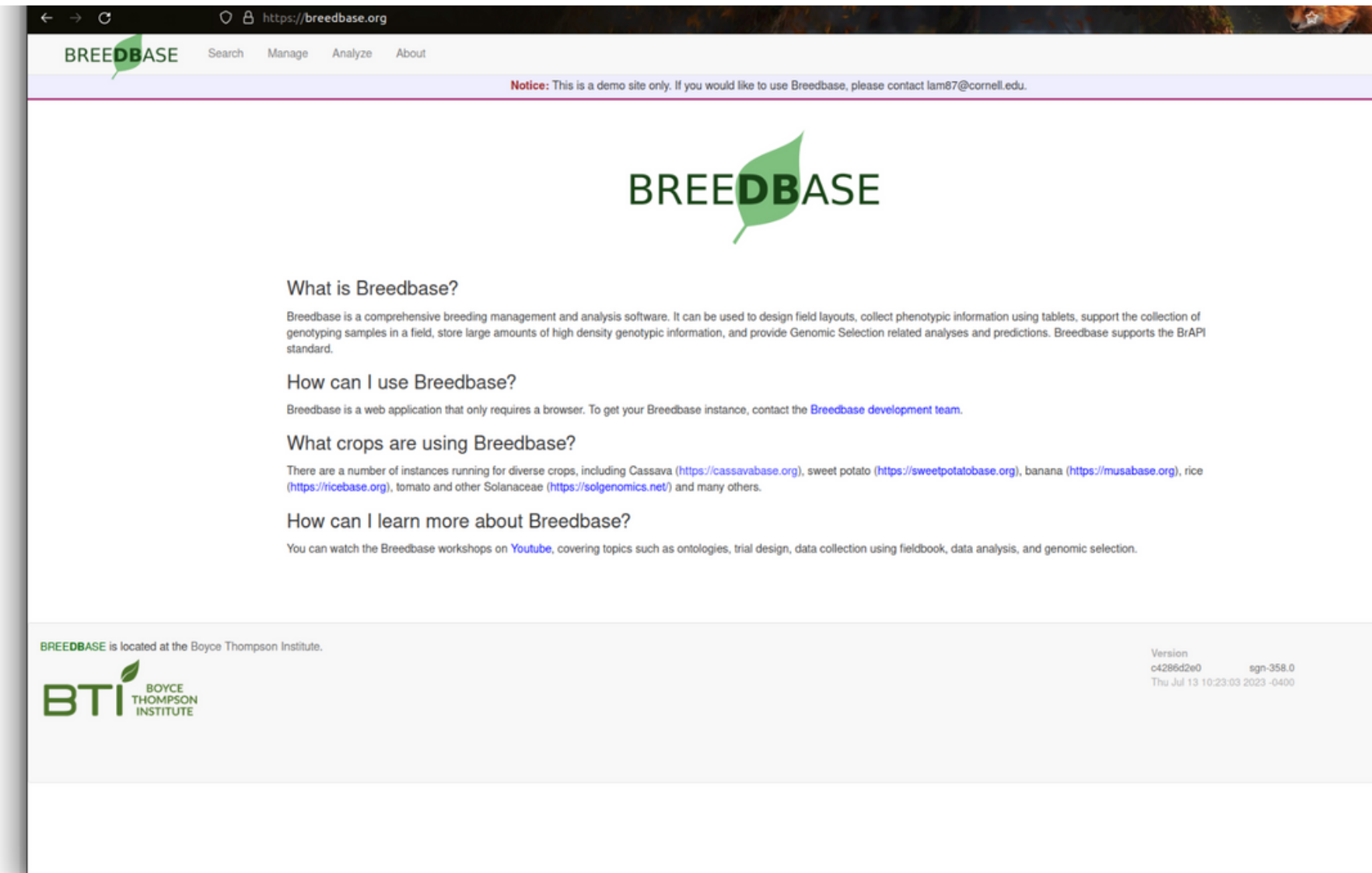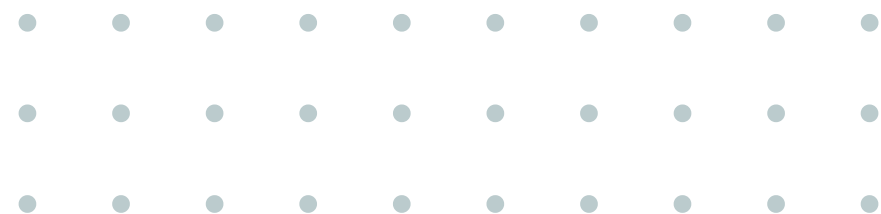

Web-based (https://breedbase.org/)

also available as open-source on GitHub
(https://github.com/solgenomics/).

Originally ininitiated as cassavabase
(https://cassavabase.org/)

Changed to Breedbase - amenable to any
crop

"DIGITAL ECOSYSTEM"

BREEDBASE
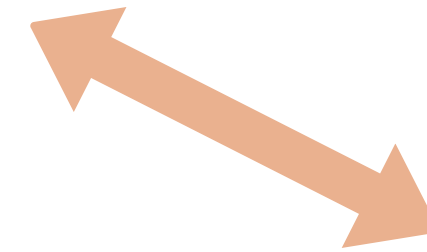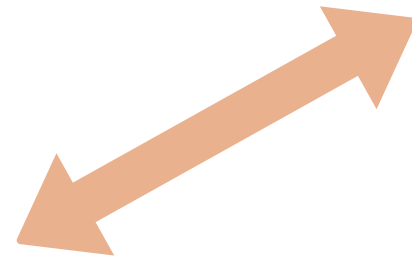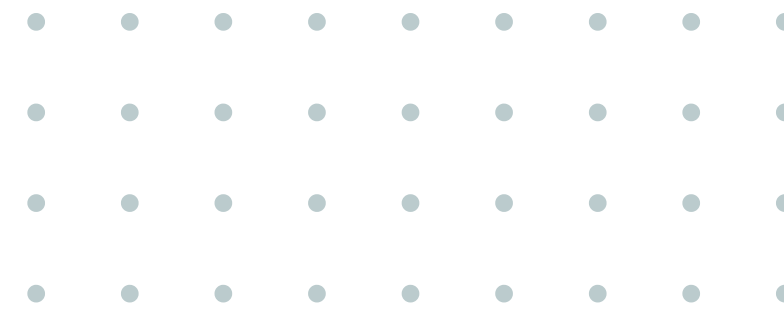
manage germplasm, fields, inventory & pedigrees

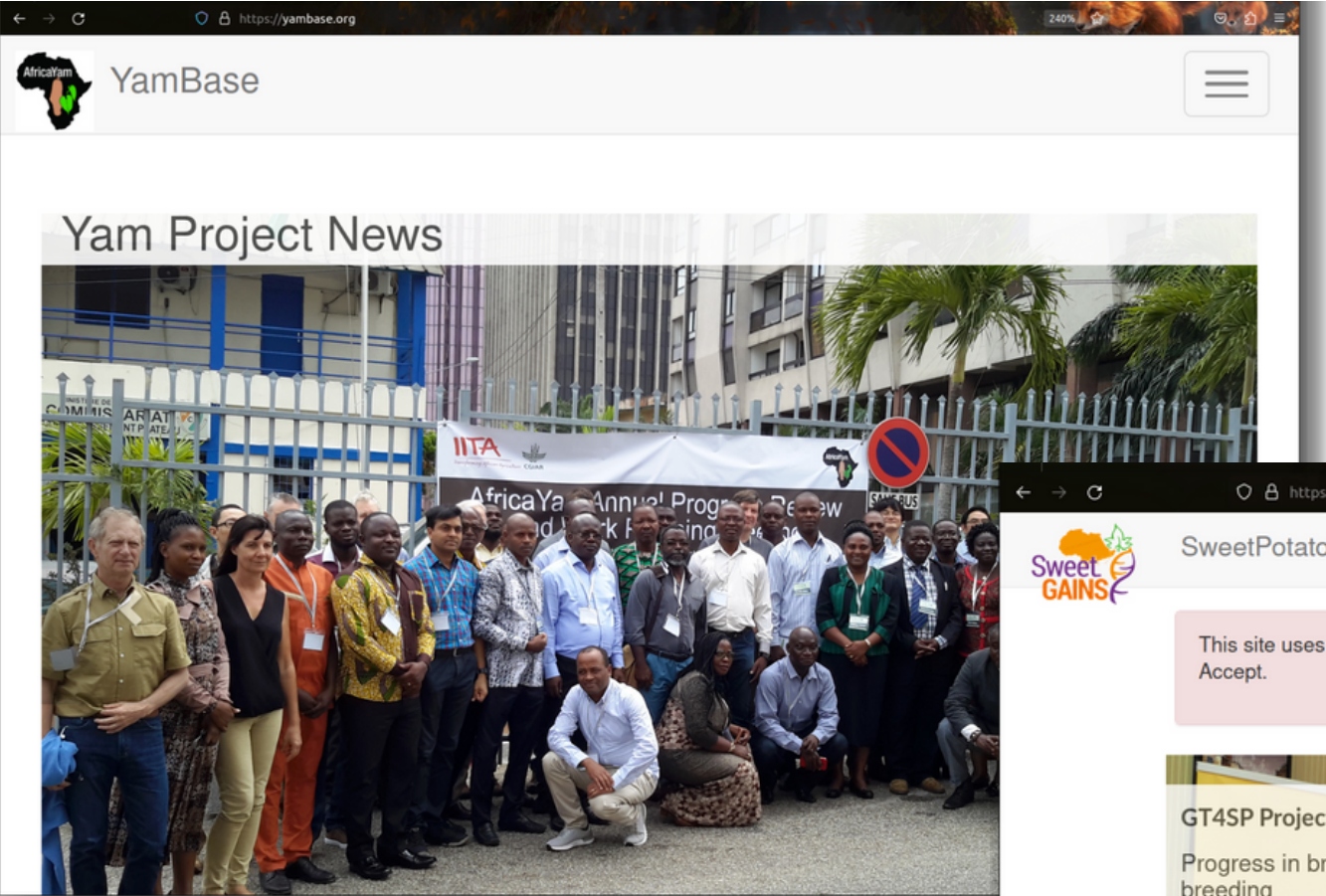**collect phenotypic and genotypic data, evaluate progeny (Fieldbook)**
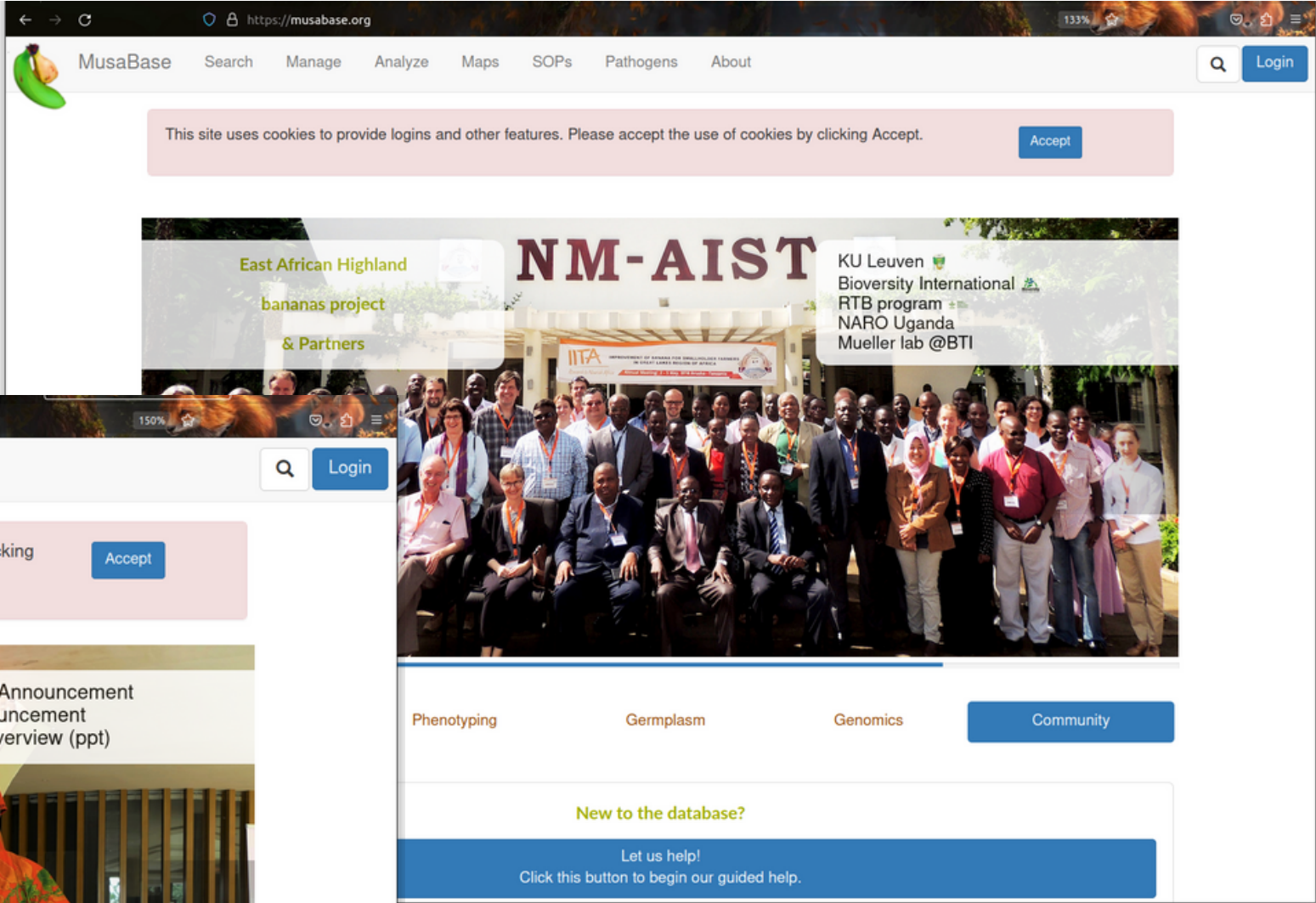
analyses, make selections (genomic selection etc)

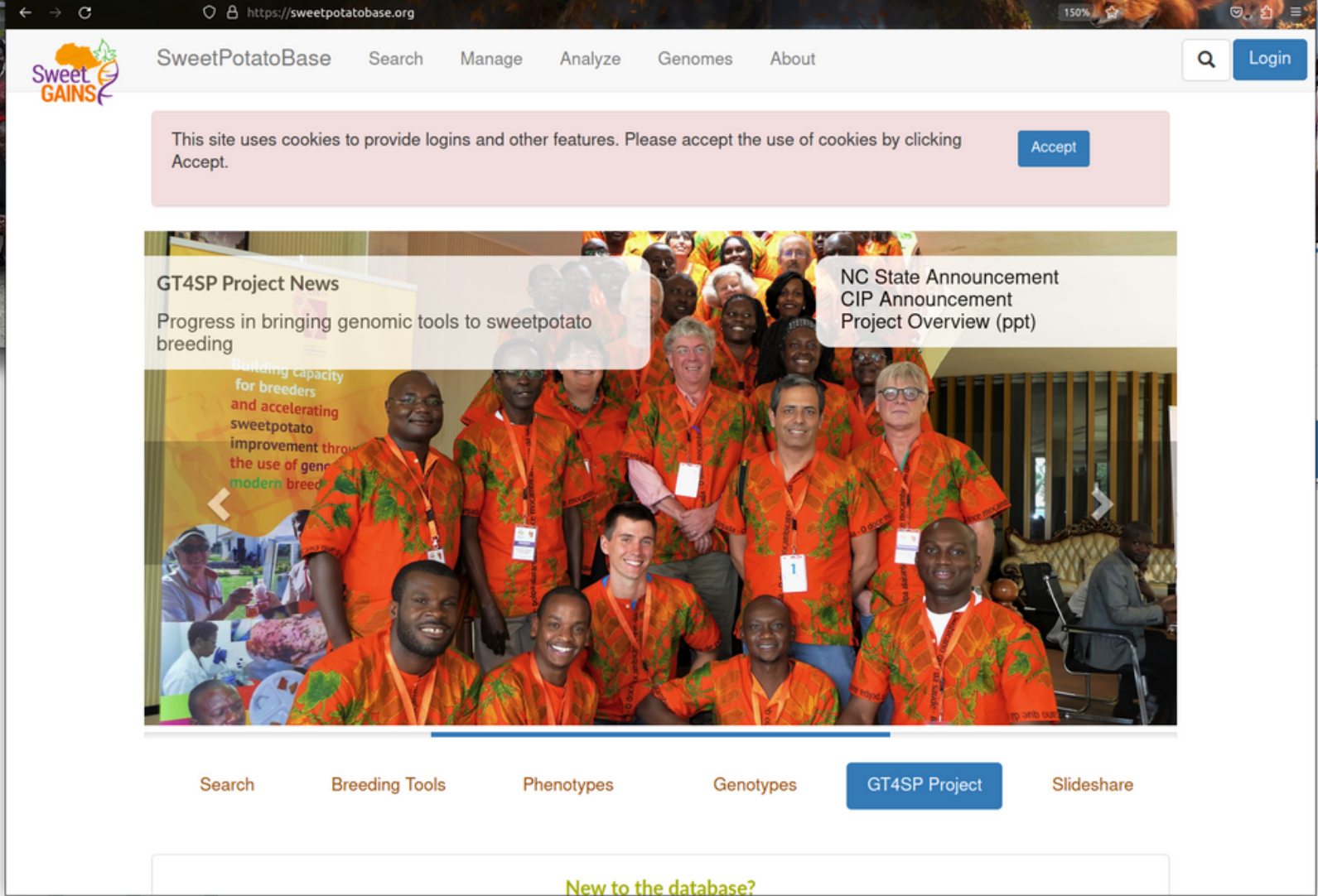"Better manage and leverage data for decision making within a fully integrated breeding workflow."
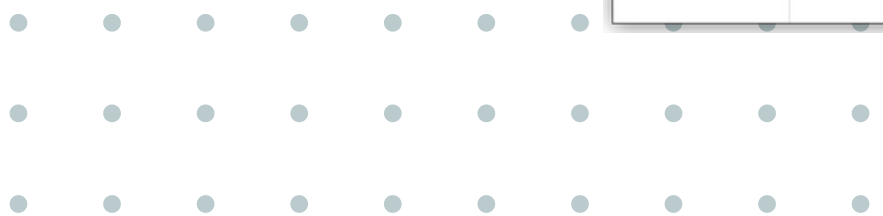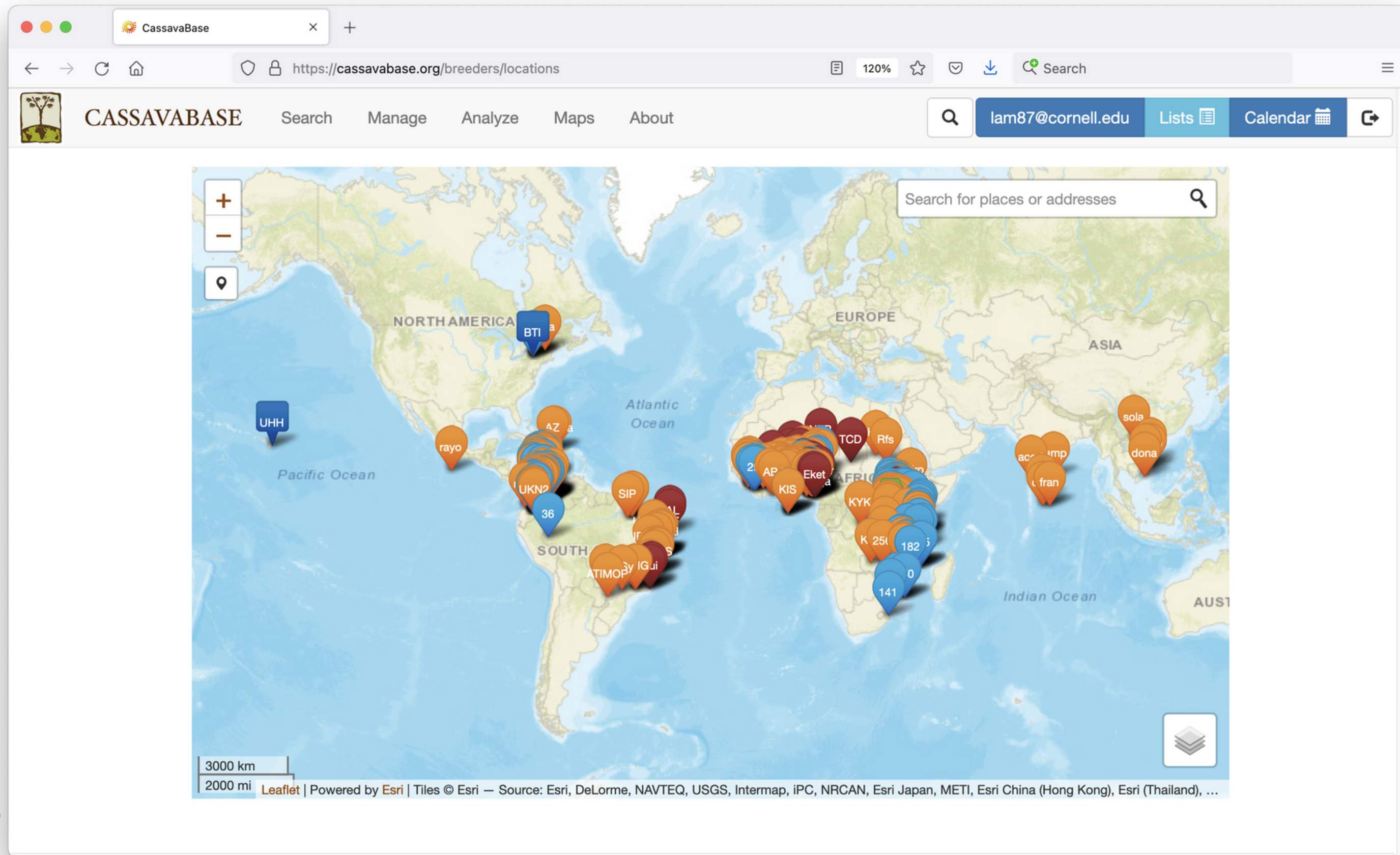
ADOPTED TO OTHER CROPS

Yambase
(https://yambase.org)

SweetPotatoBase
(https://sweetpotatobase.org)

Musabase
(https://musabase.org)

CASSAVA BREEDING LOCATIONS

03.

GENOMIC PREDICTION TOOLS

Breedbase

1. SolGS – Genomic Selection

https://cassavabase.org/solgs/

Predicts genomic estimated breeding values of individuals

Uses Ridge–Regression Best LInear Unbiased Predictor (rr–BLUP) model in R

Intitutively select training population

Estimates phenotypic correlation and heritability of traits and selection indices

Analysis output is graphically visualized and dowloable as txt formart

Isaak Tecle

2. GPCP – Genomic Prediction of Cross Performance

https://cassavabase.org/tools/gcpc
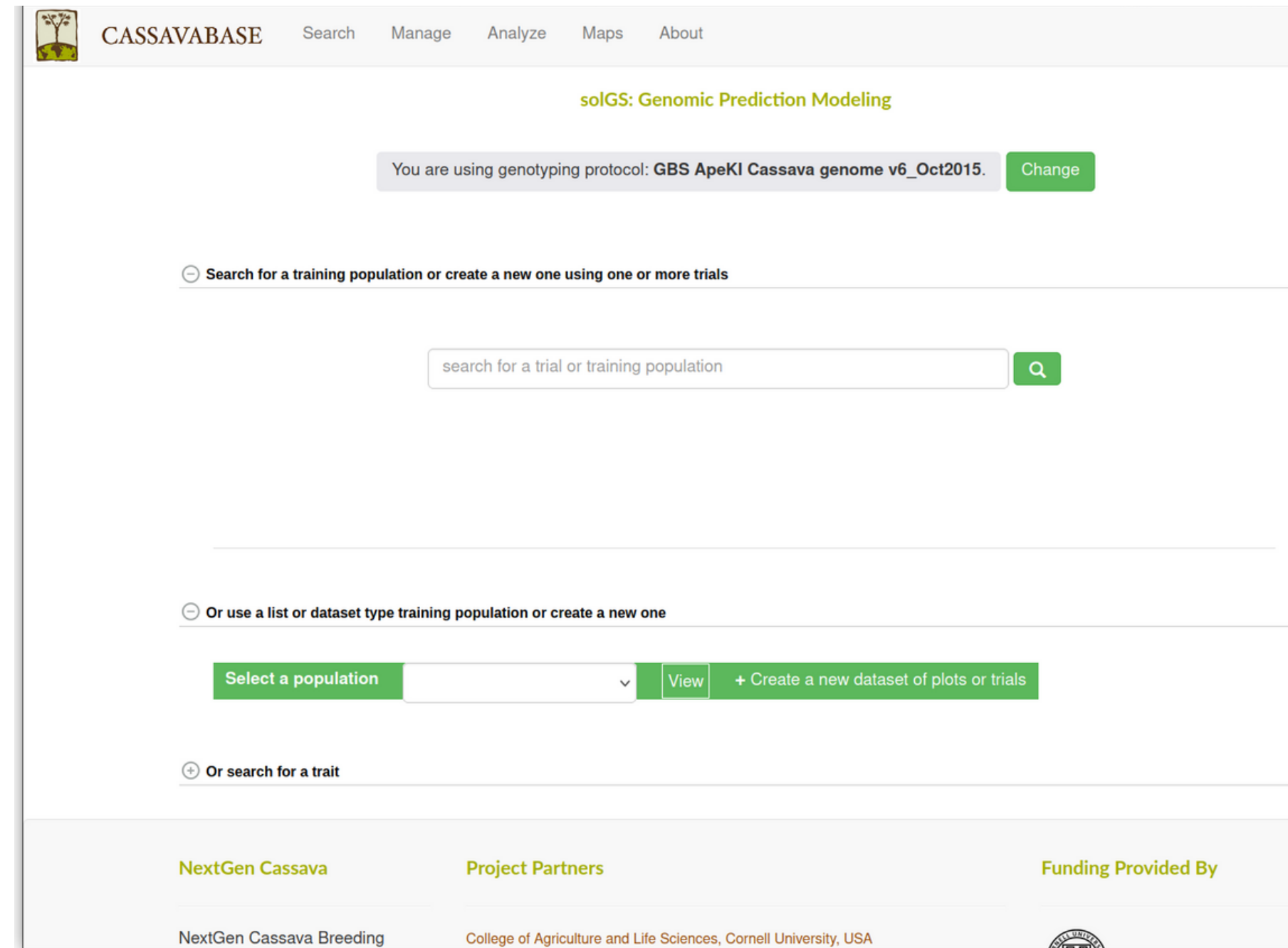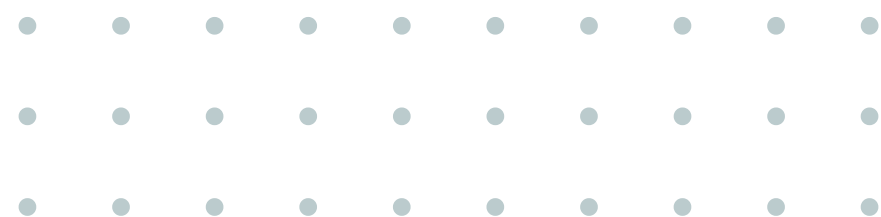
Clonal diploids with appreciable inbreeding depression and heterosis in the trait index

implements genomic prediction of cross performance

by including additive and directional dominance in mixed linear model in R

Inputs are a genotyped dataset and selection indices for the traits

Christine Nyaga

2. GPCP – Genomic Prediction of Cross Performance

https://cassavabase.org/tools/gcpc

Outputs a table of the predicted crosses sorted in descending order;

Histogram with summary statistics

Plant Sex Information
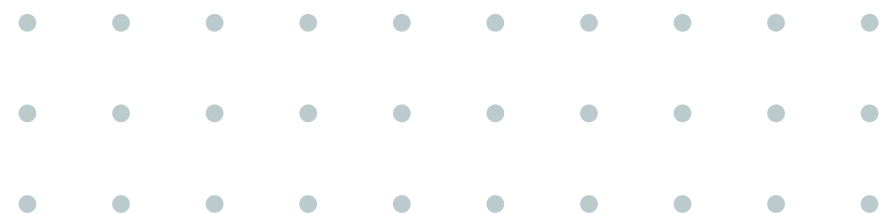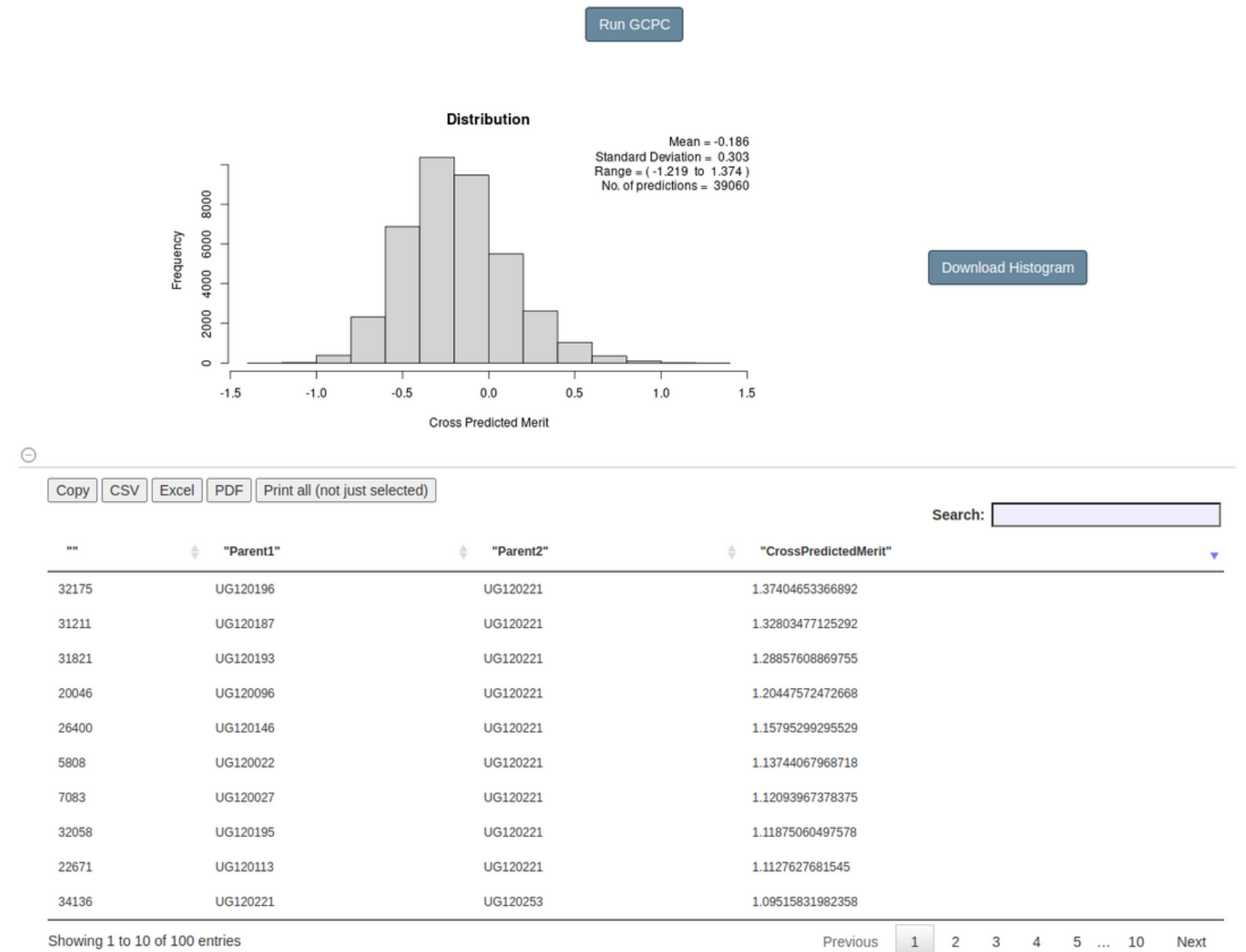
Christine Nyaga

04.

FAIR

Breedbase
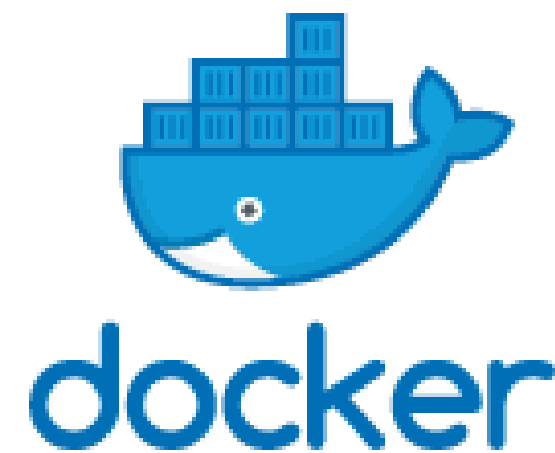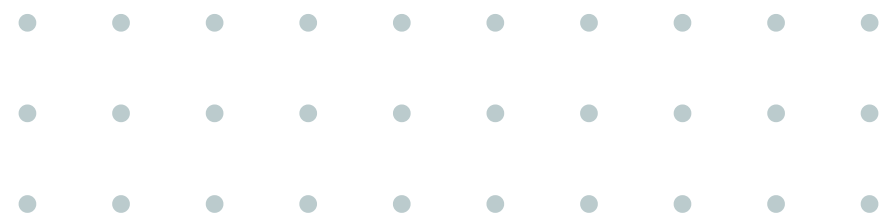
Breedbase Open Source Software Development
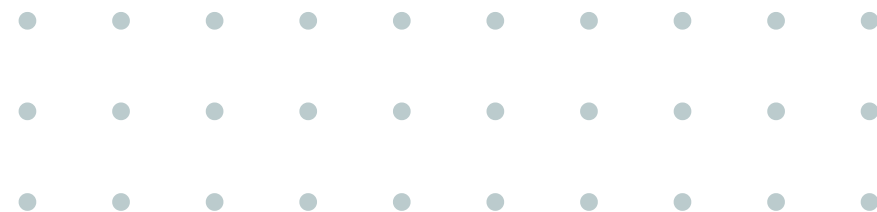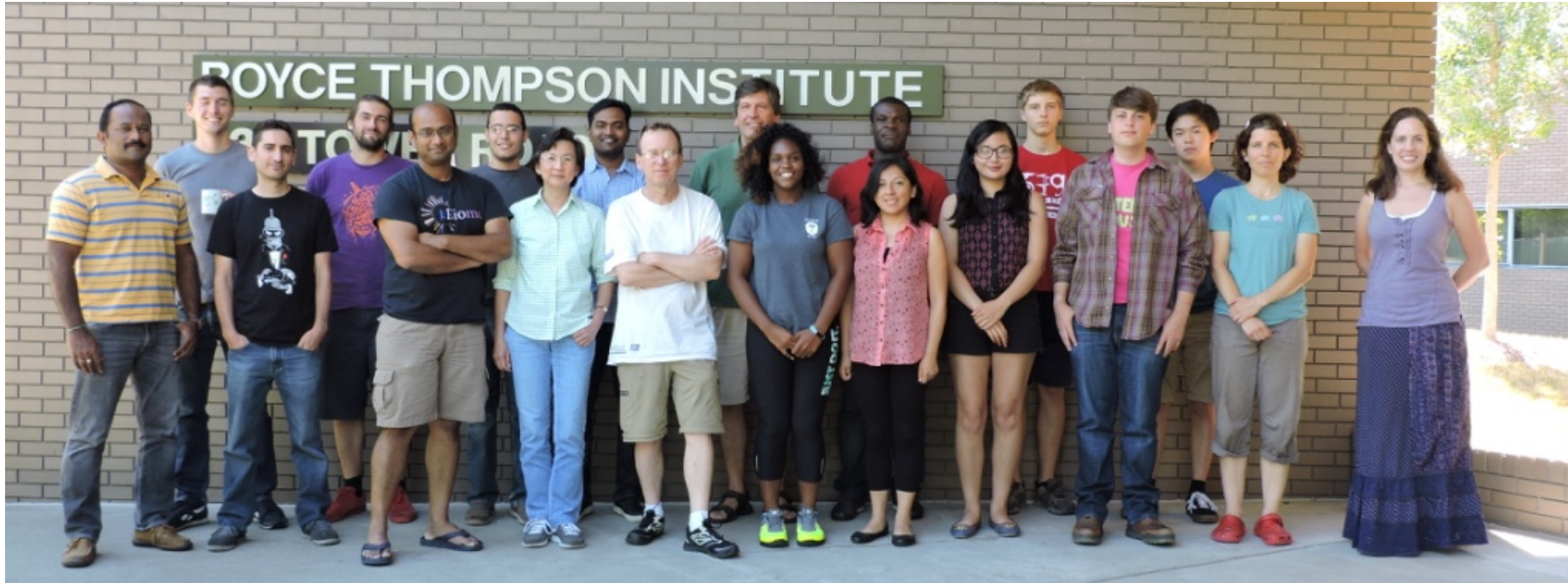
MIT License

Code available on GitHub

Issue & project tracking, Pull Requests

Deployment using docker, new build every week
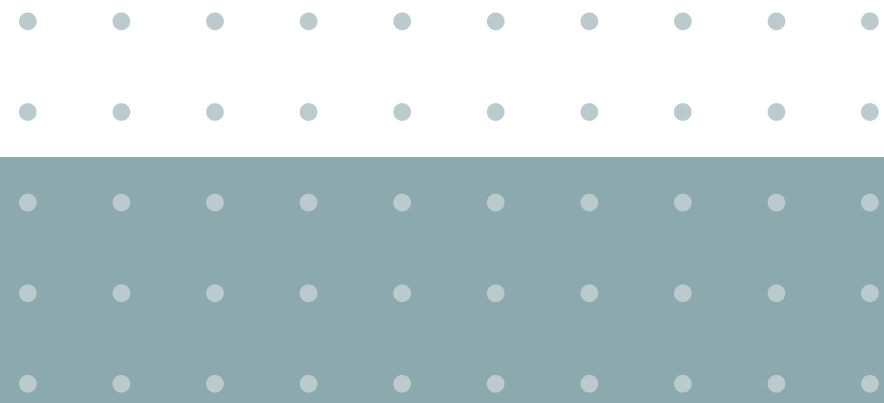
Use open source standards such as BrAPI

Acknowledgements

THANK YOU

https://breedbase.org/