

# BREEDBASE:

a digital ecosystem for modern  
plant breeding

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

ASPB 2023



# AFFILIATION

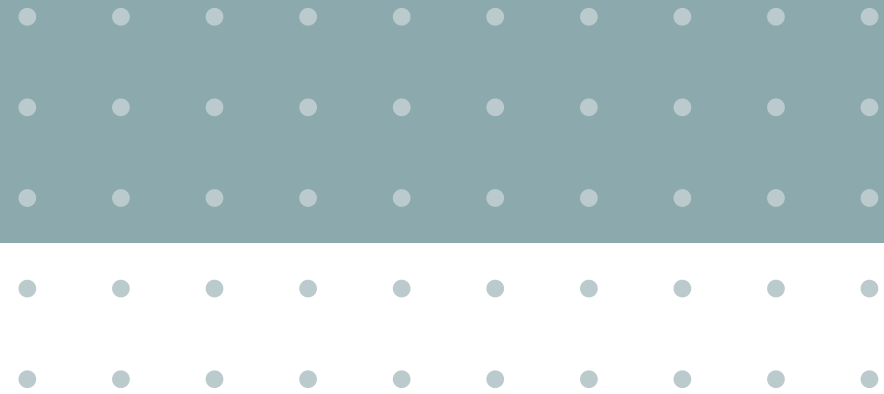
Boyce Thompson Institute,  
Cornell University



01.

# PROBLEM VS SOLUTION

*Breedbase*



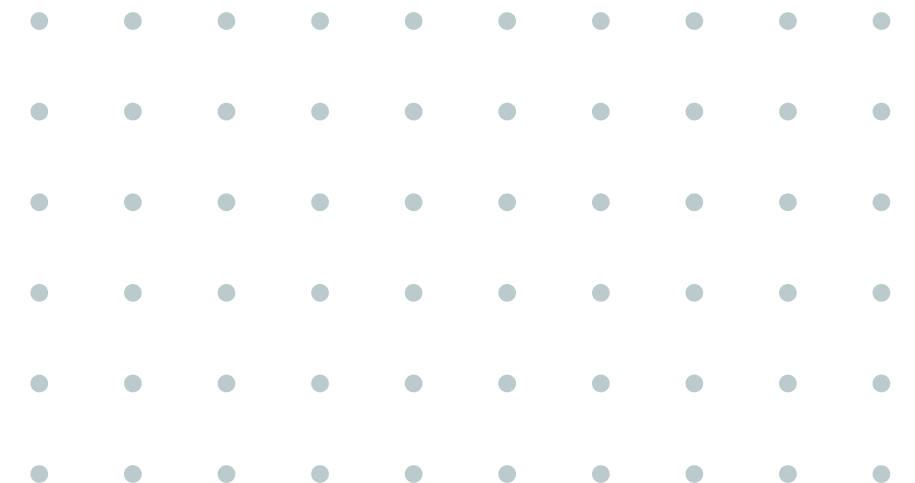


# PROBLEM

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

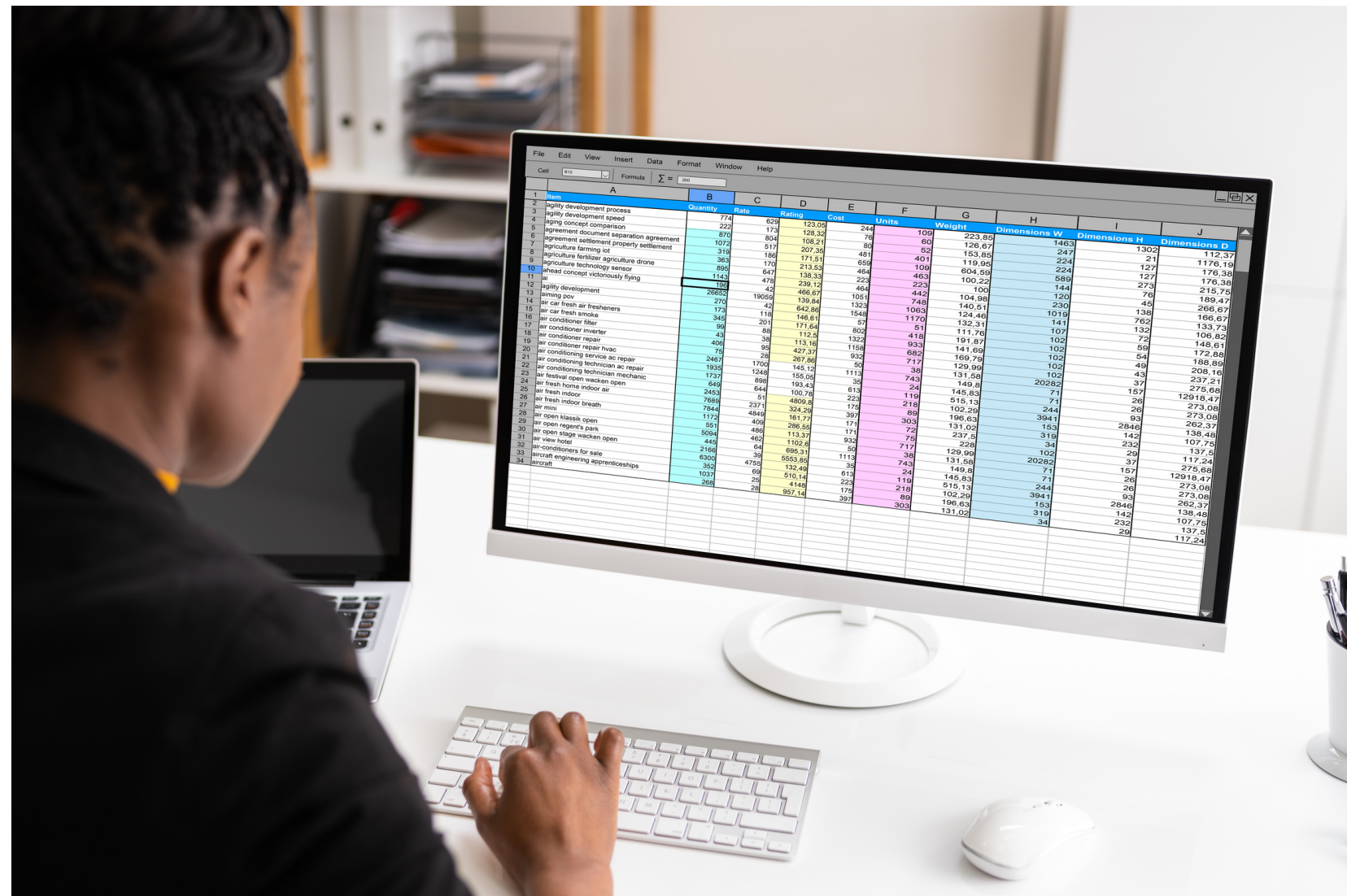
Require **efficient** data management, quality control and analytics

**Complex** and expensive process



# TRADITIONAL APPROACHES

## 1. Spreadsheets



## 2. Book Records





**01.**

**Difficulties in merging data across different spreadsheets**

**02.**

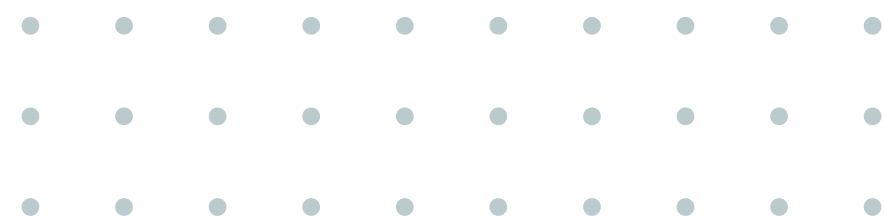
**Lack of centralized storage; often stored on personal computers and laptops**

**03.**

**Difficulties in visualizing or analyzing data across spreadsheets**

**04.**

**Limited backup strategies and little recourse if accidental data loss occurs**



# **PROBLEMS WITH TRADITIONAL APPROACHES**

02.

# BREEDBASE



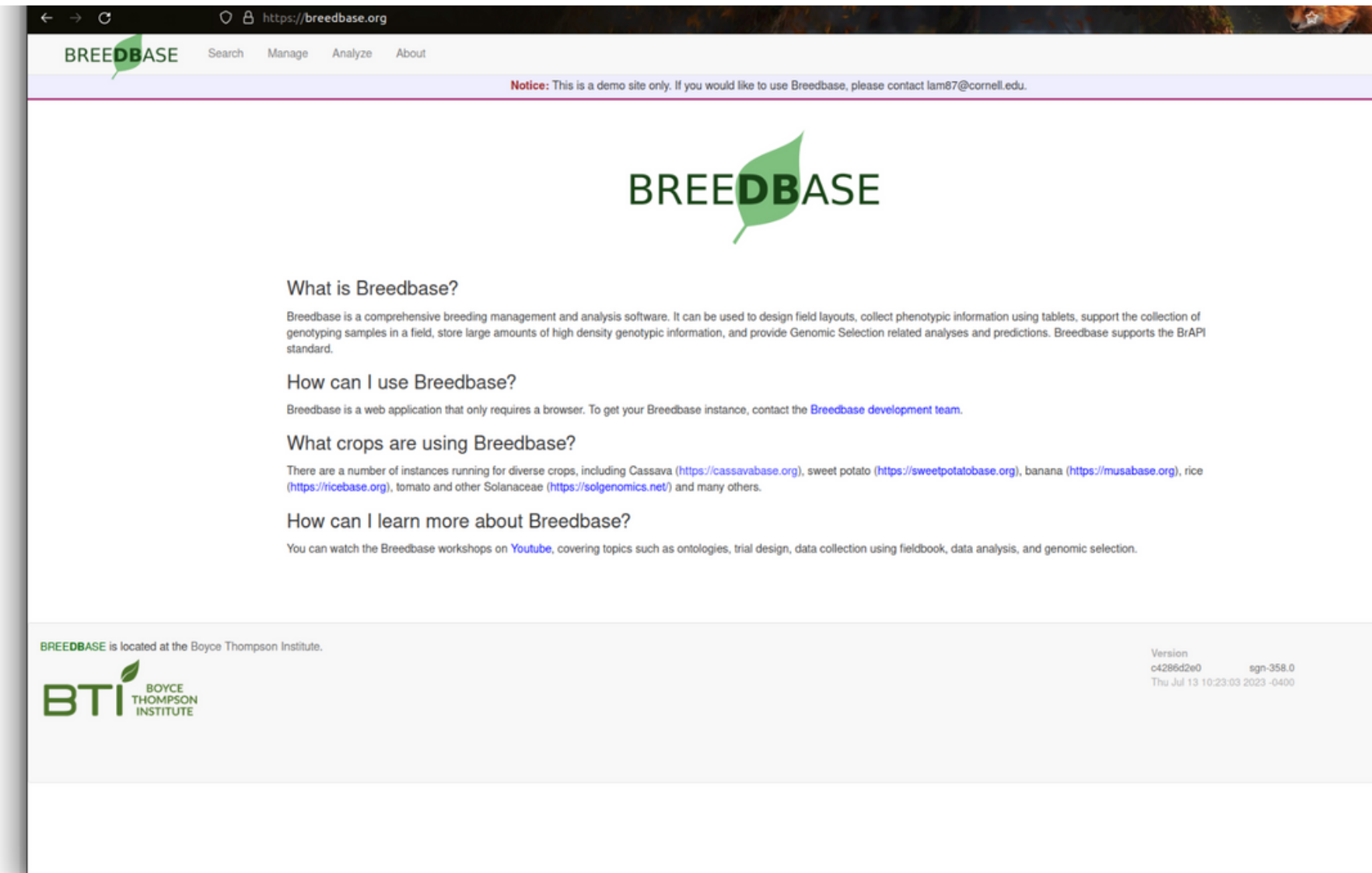
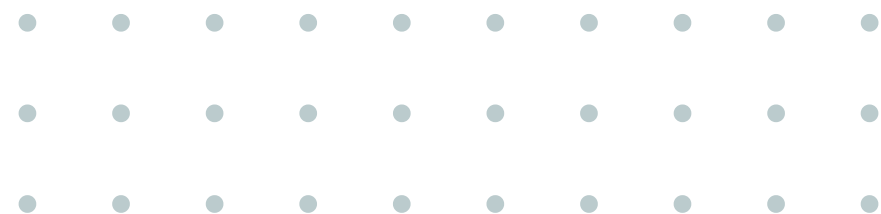


Web-based (<https://breedbase.org/>)

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

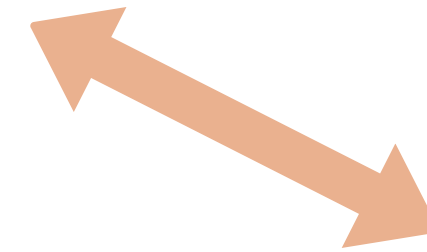
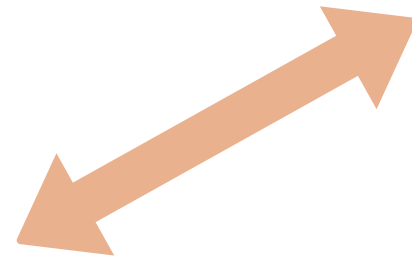
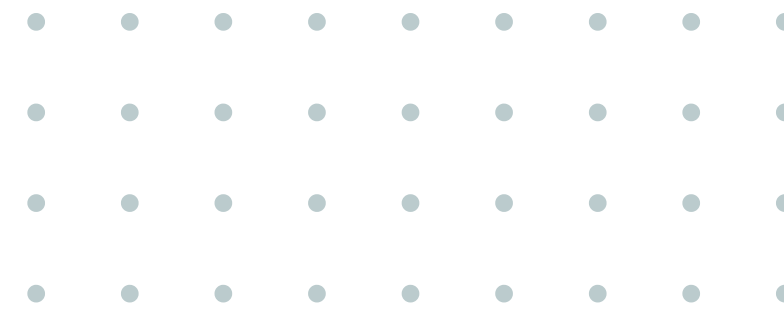
Originally initiated as cassavabase (<https://cassavabase.org/>)

Changed to Breedbase – amenable to any crop





# "DIGITAL ECOSYSTEM"



manage germplasm,  
fields, inventory &  
pedigrees



collect phenotypic and  
genotypic data, evaluate  
progeny (Fieldbook)



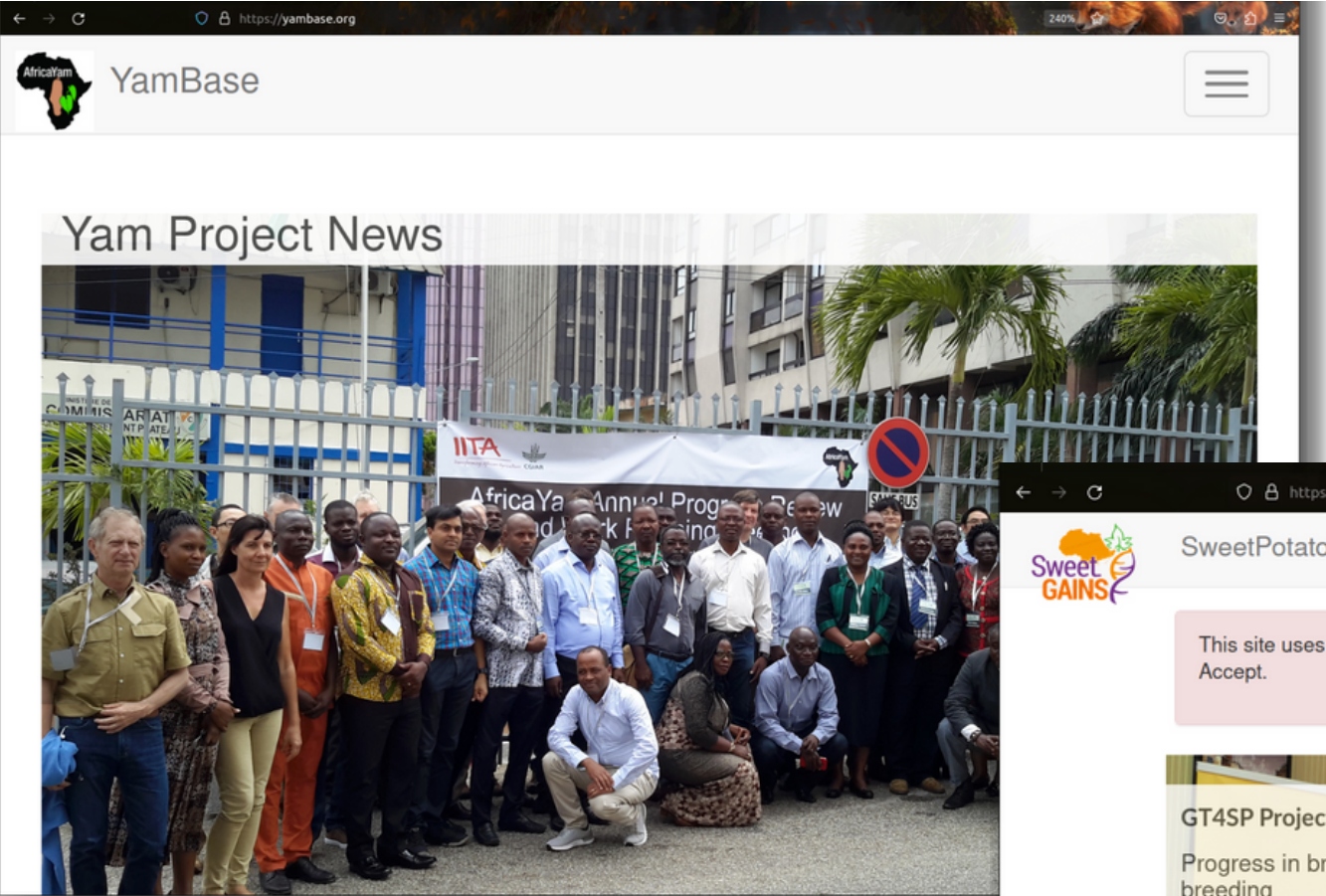
analyses, make  
selections (genomic  
selection etc)



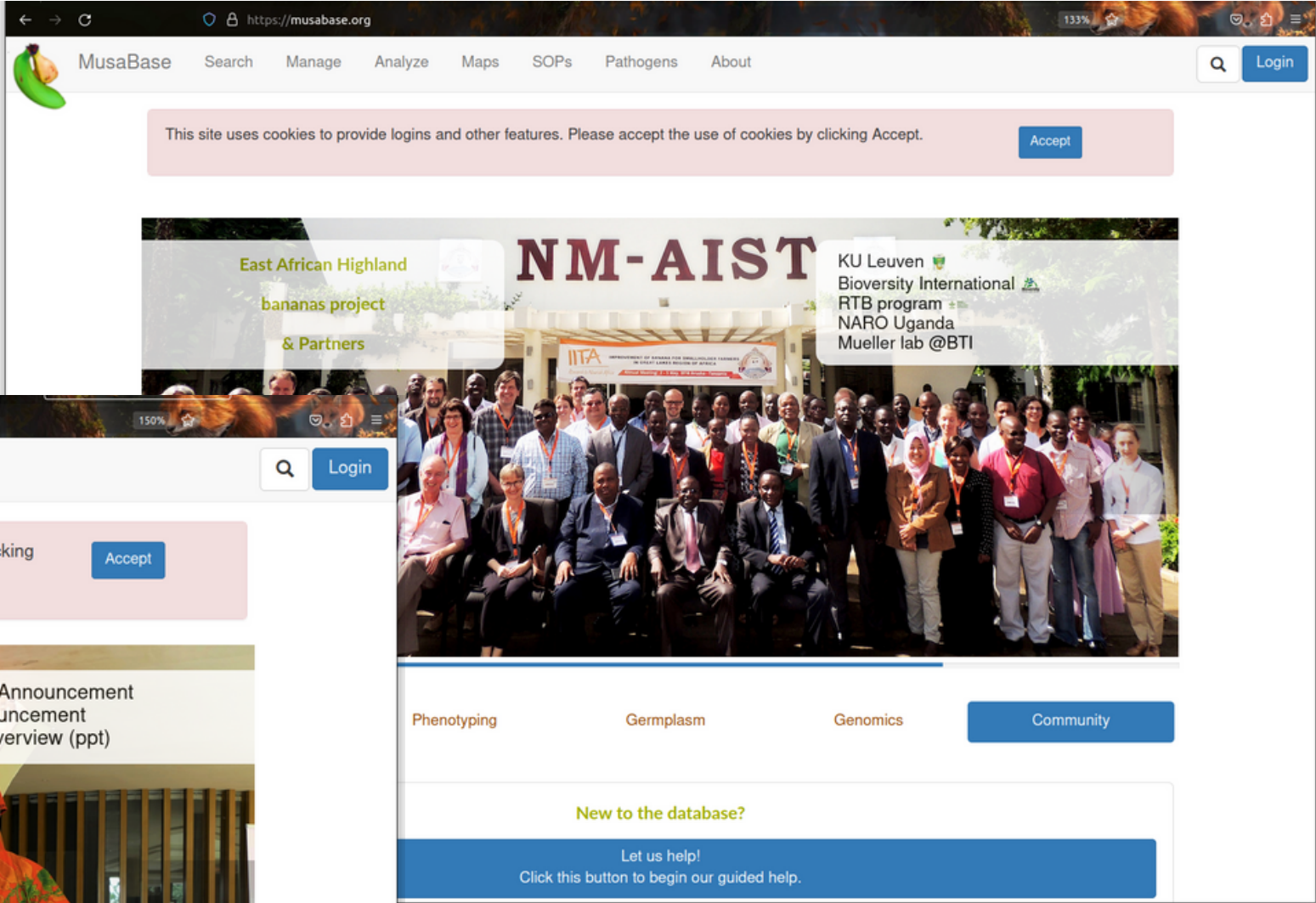
**“Better manage and leverage data for decision making within a fully integrated breeding workflow.”**



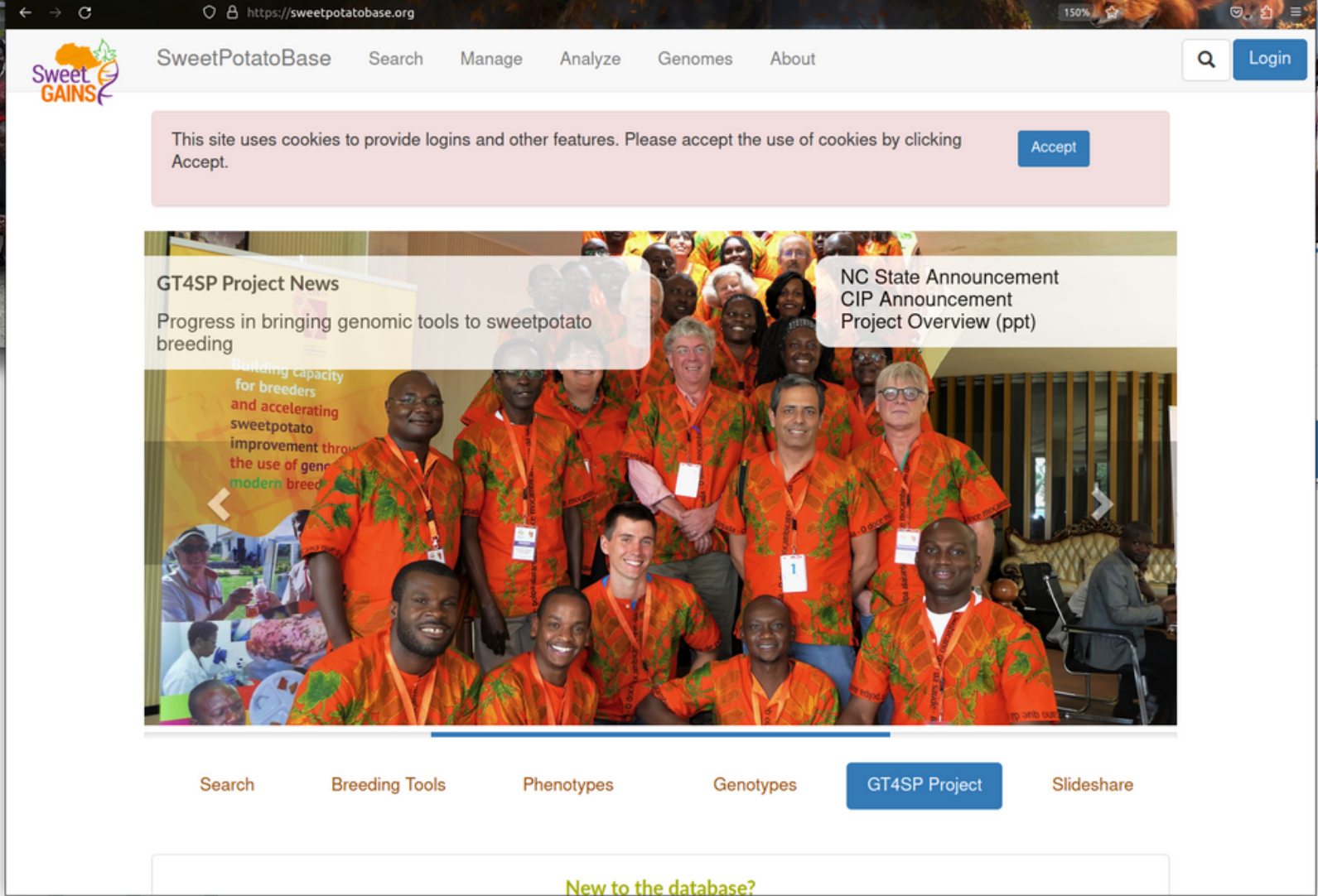
# ADOPTED TO OTHER CROPS



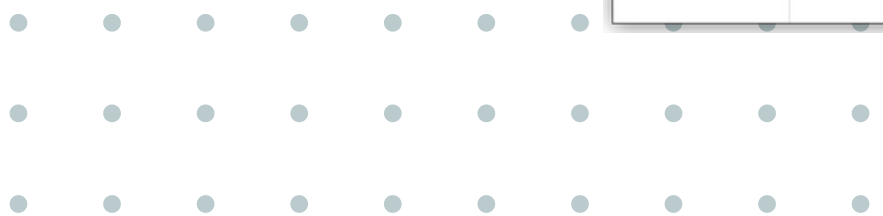
Yambase  
(<https://yambase.org>)



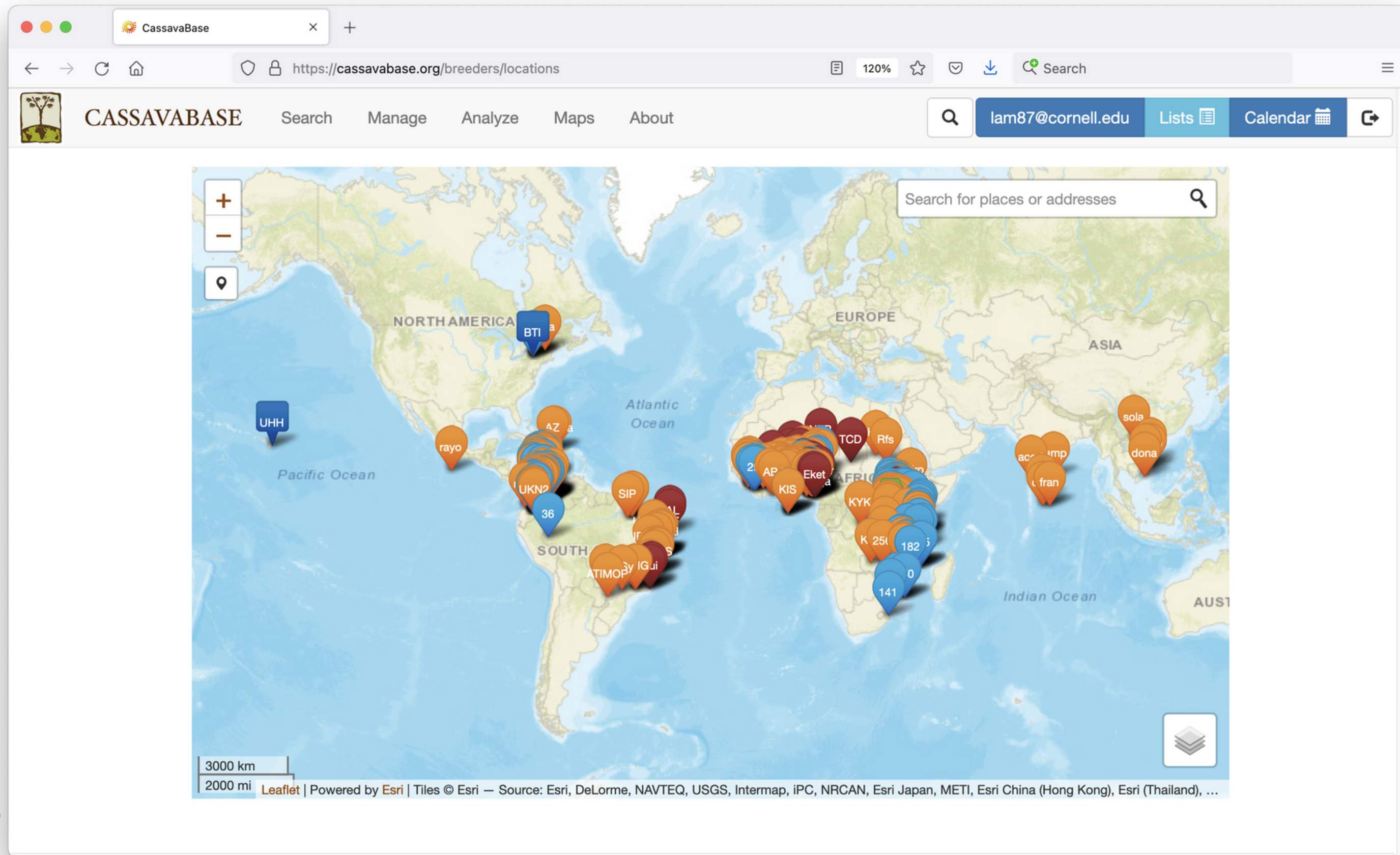
Musabase  
(<https://musabase.org>)



SweetPotatoBase  
(<https://sweetpotatobase.org>)



# CASSAVA BREEDING LOCATIONS



03.

# GENOMIC PREDICTION TOOLS

*Breedbase*



# 1. SolGS – Genomic Selection

<https://cassavabase.org/solgs/>

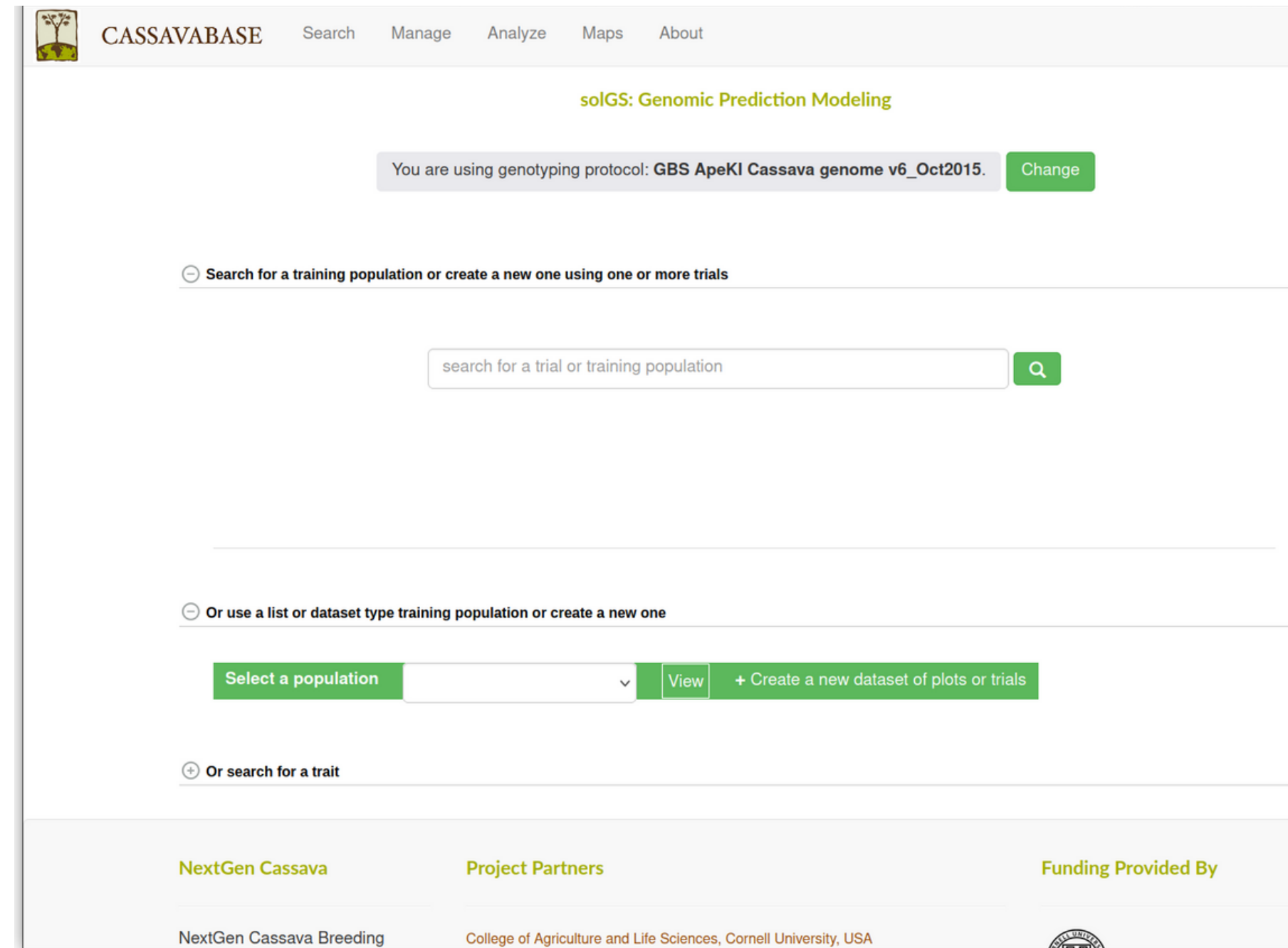
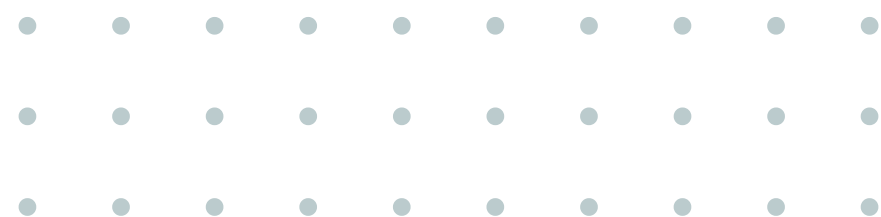
Predicts genomic estimated breeding values of individuals

Uses Ridge-Regression Best Linear Unbiased Predictor (rr-BLUP) model in R

Intuitively select training population

Estimates phenotypic correlation and heritability of traits and selection indices

Analysis output is graphically visualized and downloadable as txt format



Isaak Tecele

# 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

The screenshot shows the web interface for the Genomic Prediction of Cross Performance (GPCP) tool. The browser address bar shows the URL <https://cassavabase.org/tools/gcpc>. The page title is "Genomic Prediction of Cross Performance (GPCP)".

The main section is titled "Select a Genotyped Dataset". It features a table of available datasets:

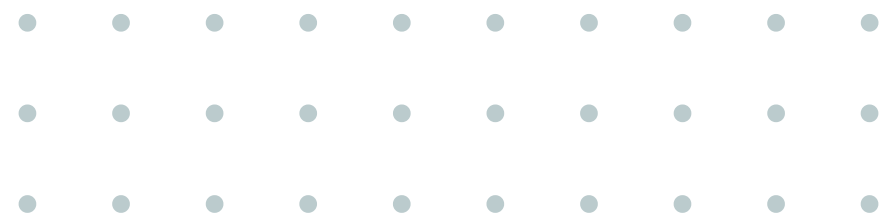
Select	Dataset Name	Contents						
<input type="checkbox"/>	my favs	<table border="1"><thead><tr><th>Breeding_programs</th><th>Years</th><th>Traits</th></tr></thead><tbody><tr><td>NaCRRI</td><td>2014 2016</td><td>cassava bacterial blight severity 6-month evaluation CO_334:0000176 cassava brown streak disease leaf incidence 3-month evaluation CO_334:0000208</td></tr></tbody></table>	Breeding_programs	Years	Traits	NaCRRI	2014 2016	cassava bacterial blight severity 6-month evaluation CO_334:0000176 cassava brown streak disease leaf incidence 3-month evaluation CO_334:0000208
Breeding_programs	Years	Traits						
NaCRRI	2014 2016	cassava bacterial blight severity 6-month evaluation CO_334:0000176 cassava brown streak disease leaf incidence 3-month evaluation CO_334:0000208						
<input type="checkbox"/>	ciat trials	<table border="1"><thead><tr><th>Breeding_programs</th><th>Years</th><th>Locations</th></tr></thead><tbody><tr><td>CIAT</td><td>2019</td><td>Cerete, Cordoba, Colombia Sevilla</td></tr></tbody></table>	Breeding_programs	Years	Locations	CIAT	2019	Cerete, Cordoba, Colombia Sevilla
Breeding_programs	Years	Locations						
CIAT	2019	Cerete, Cordoba, Colombia Sevilla						

Below the table, it says "Showing 9 to 10 of 21 entries". There are navigation links for "Previous", "1", "4", "5", "6", "11", and "Next".

A button labeled "Proceed to Factor Selection" is visible. Below it, a green box contains the text "Factors will appear here..".

There is a dropdown menu labeled "Select a formula" with a refresh icon. Below it, a note reads: "This is a drop-down menu showing your pre-constructed selection indices. Select one and proceed to run the model." A "Run GPCP" button is located below the dropdown.

At the bottom, there is an "Output" section with the text: "Plant Sex info: 1= Male, 2=Female, 3= Monoecious male (m>f) and 4= Monoecious female (f-m)".



Christine Nyaga

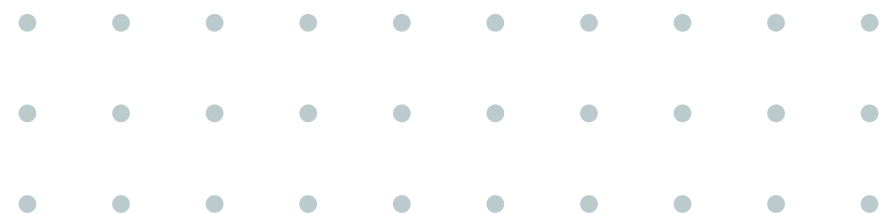
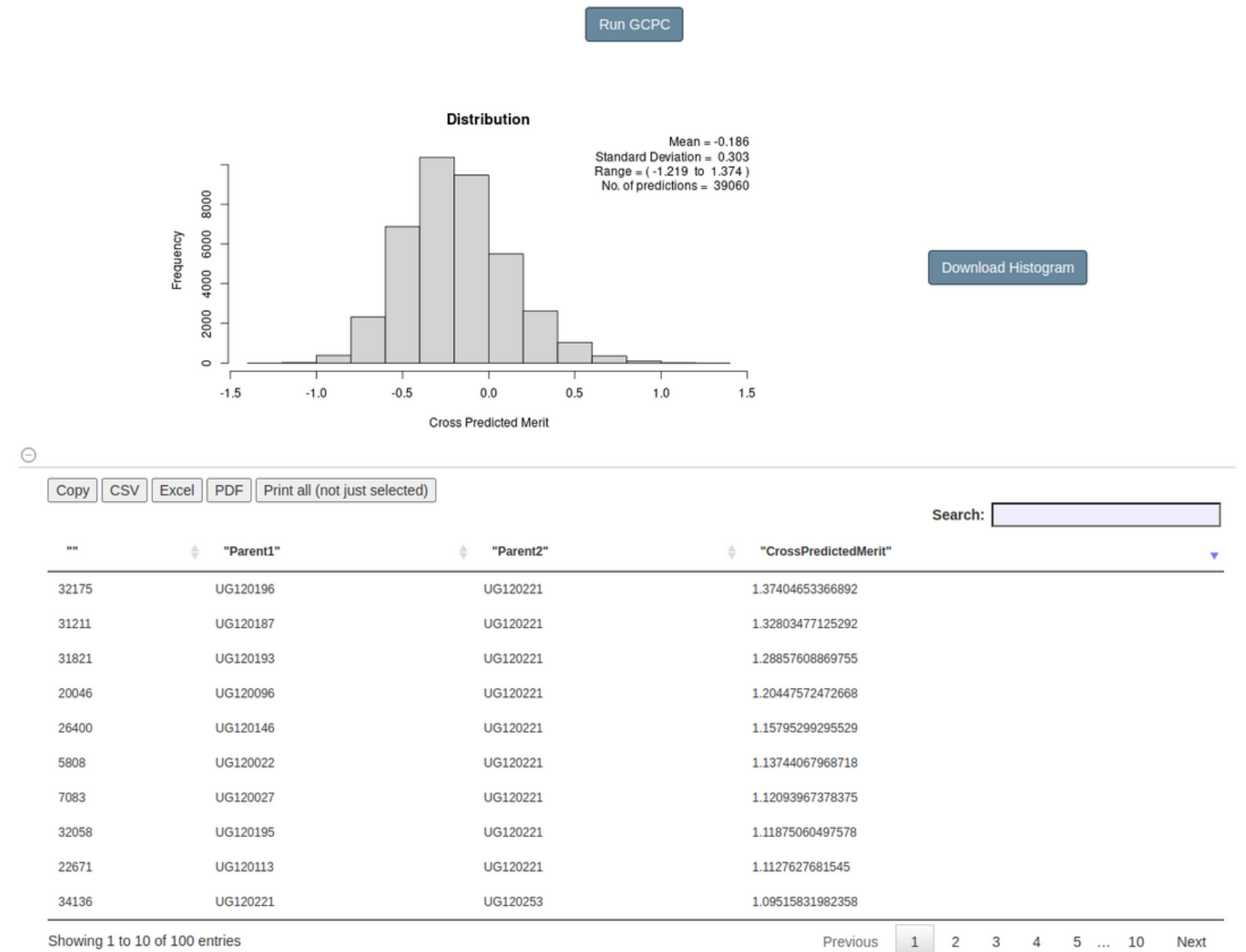
# 2. GPCP – Genomic Prediction of Cross Performance

<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

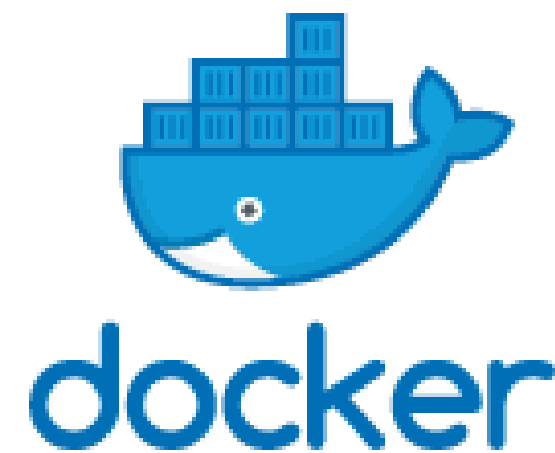
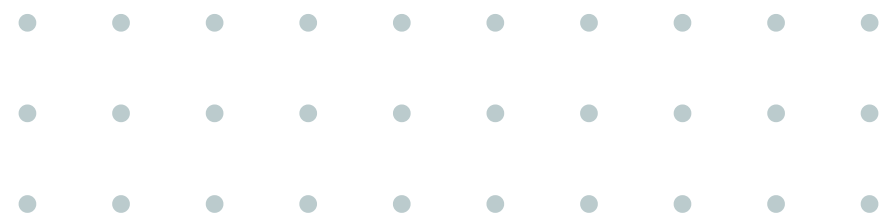
MIT License

Code available on GitHub

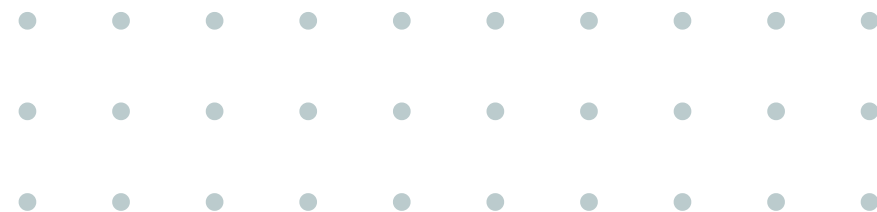
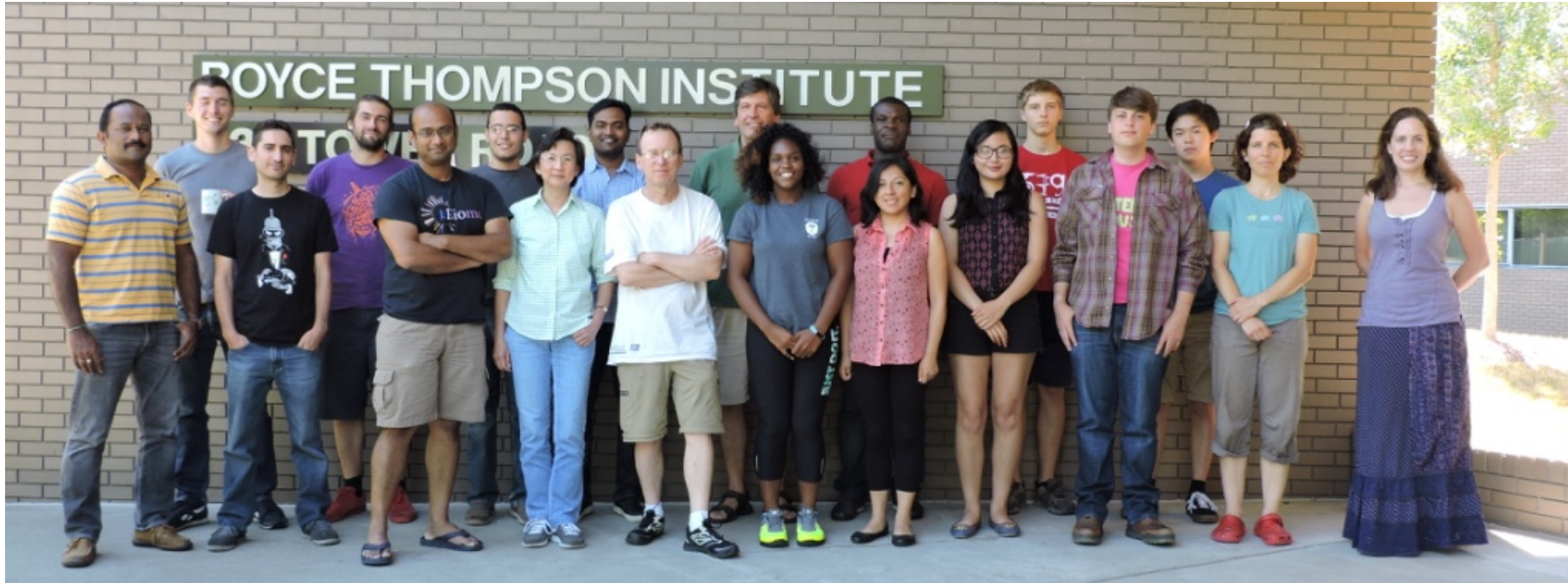
Issue & project tracking, Pull Requests

Deployment using docker, new build every week

Use open source standards such as BrAPI

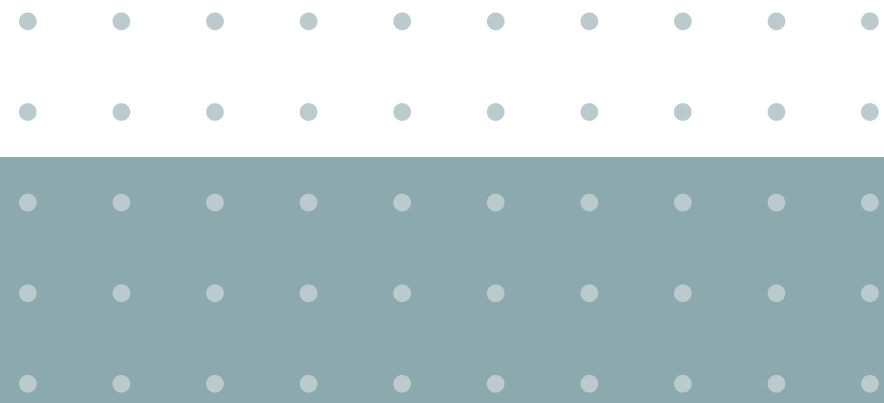


# Acknowledgements



BILL &  
MELINDA  
GATES  
*foundation*





# THANK YOU

<https://breedbase.org/>

