Breedbase: a digital ecosystem for plant breeding

AgBio seminar, December 2019
Guillaume Bauchet, Bryan Ellerbrock, David Lyon, Naama Menda, Nicolas Morales, Mirella Flores, Alex C. Ogbonna, Adrian Powell, Titima Tantikanjana and Isaak Y Tecle

Lukas Mueller lab
From Solgenomics...
Data:
5 reference genomes (tomato, potato, eggplant, pepper, wild tomatoes)
6 draft genomes (tobacco, petunia, wild tomato)
16 genetic maps
500+ re-sequenced genomes

Tools:
BLAST, Comparative maps, Expression Atlas, Jbrowse, Motif finder,
Pathway tools, Phylo tree browser, QTL tools, VIGS, WebApollo
...to Breedbase

https://github.com/solgenomics

https://cassavabase.org

https://yambase.org

https://sweetpotatobase.org

https://musabase.org

Breedbase
Plant Breeding

Analysis → Crossing → Selection → Evaluation

BreedBase
NEXTGEN CASSAVA

solGS: a web-based tool for genomic selection

Isaak Y TcLe, Jeremy D Edwards, Naama Menda, Chiedozie Egesi, Ismail Y Rabbi, Peter Kulakow, Robert Kawuki, Jean-Luc Jannink & Lukas A Mueller

BMC Bioinformatics 15, Article number: 398 (2014) | Cite this article
6304 Accesses | 3 Citations | 14 Altmetric | Metrics
BREEDBASE (Cassavabase) ROLES

• Project’s data repository

• Contribute to harmonize data, standardize procedure

• Enhance collaboration across partners, build a crop community
DEVELOPING A COMMON VOCABULARY: ONTOLOGIES
Formats:

TD (Trait Dictionary) - OBO (Open Biomedical Ontology) -
OWL (Web Ontology Language) -

Tools:

OBO-Edit, Protege, Web Protege
ONTObLOGLY TRAIT REQUEST

1. Term submission form
   submit.rtbbase.org

2. Github issue

3. Curation
   CO Curators
   Afolabi Aghona
   Naama Menda

4. Planteome
   GitHub

5. Release
   OBO
   OWL
   TD
   .xls,
   .xlsx

6. CO Website

7. CassavaBase

Credits: Bryan Ellebrock, Marie-Angelique Laporte
**Ontology Post Composing**

https://cassavabase.org/tools/compose/

### Trait Search

<table>
<thead>
<tr>
<th>Search Trait Name:</th>
<th>root</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search Definition:</td>
<td>Search any word or piece of definition</td>
</tr>
<tr>
<td>Subset Traits:</td>
<td>Leave blank to see all traits</td>
</tr>
</tbody>
</table>

### Results

<table>
<thead>
<tr>
<th>Trait ID</th>
<th>Trait Name</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMP-0000179</td>
<td>ash content in percentage/day</td>
<td>leaf from hardwood area (lower woody stem portion)</td>
</tr>
</tbody>
</table>
Breedbase: system overview
OVERVIEW

Manage Data
(1) Design
- List or a Dataset
- Trial experiment
- Barcoded labels

(2) Collect and Upload
- Phenotypes
- Tissue sample
- Genotypic data (vcf)
- Images

Phenotype-Genotype
- Summary statistics
- Selection Index
- ANOVA
- Trial comparison
- SoI5S- SoLGWAS (dev.)
- Mixed Modelling (dev.)
- HapMap JBrowse

Pedigree & Crossing
-> cross upload
-> seedlots

Search Data
(1) Single search:
- Traits (ontology)
- Trial
- Germplasm
- Genomic data
- Location

(2) Wizard search

Manage Data
(1) Design
- List or a Dataset
- Trial experiment
- Barcoded labels

(2) Collect and Upload
- Phenotypes
- Tissue sample
- Genotypic data (vcf)
- Images

Selection

Crossing

Analysis

Evaluation

BREEDBASE
This workflow will guide you through uploading genotypes into the database

Select a genotyping project on the next screen. This project can represent a series of genotyping plates sent to a genotyping facility.

Ideally the sample names in your VCF file will match sample names in genotyping plates in the database; however, the sample names in your file can also match accession names in the database.

Currently we support the VCF format for upload.

Print barcode labels for my experiment (for your plots or plants or tissue samples in the field, or for your 96 well plate and tissue samples)

Analyze phenotypic performance across trials
Breedbase – PhenoApps: field applications
REACHING THE FIELD: PHENOAPPS

http://phenoapps.org/

Credits: Trevor Rife, KSU
PHENOTYPIC DATA COLLECTION WORKFLOW
**Genotypic Data Collection Workflow**

- Tissue sampling

- Barcoded field plot level or plant level

- Plant field

- Export data

- Upload

- Create trial list of accessions and seedlots


- https://cassavabase.org/breeders/genotyping/: tissue sampling upload interface
Breedbase - Breeding API
Breeding Application Programming Interface (API)
  • Language support: Brapi.R interface and Brapi.JS
• Data exchange

• New way for coding breeding applications (BrAPPs)

• BrAPPs run on any data backend that supports BrAPI

http://brapi.org/
https://github.com/CIP-RIU/brapi
Graphical filtering tool

Trial comparison tool

Pedigree viewer

Field Map viewer

Boxplotter tool

Genetic Map viewer

BrAPI R package

EVALUATION

ANALYSIS

SELECTION

CROSSING

BRAPPS

https://brapi.org/brapps.php
Engaging communities

Cassavabase data content
(as of June 2018)

Germplasm:
217,500 accessions
2,450 images
6366 pedigree

Genotypes:
53,000 genotypes
120 DNA plates

Phenotypes:
207 locations (10 countries)
2,690 field trials
488,000 field plots
71,000 tracked field plants
11,000,000 phenotypes
A DIGITAL ECOSYSTEM FOR BREEDING

Databases and ontologies

BrAPI—an application programming interface for plant breeding applications


Image data analysis

Near Infrared Spectra (NIRS) analysis

Marker QC
What future looks like?

https://www.breedinginsight.org/

Cobb et al 2019, TAG
Online Resources

Looking for code?

BrAPI code: https://brapi.org/
BrApps: https://brapi.org/brapps.php
Breedbase code: https://github.com/solgenomics
Breedbase Docker instance: https://hub.docker.com/u/breedbase

Looking for phenoApps?

PhenoApps: https://github.com/PhenoApps
https://www.youtube.com/playlist?list=PLs7Y2nGwfz4E5 gv1H6Y4imeWDkFJDhIn

Looking for database tutorials or ontology request?

Online manual and videos https://solgenomics.github.io/sgn/
Request new traits: http://submit.rtbbase.org/
Slides: http://www.slideshare.net/solgenomics
Thank you!