Genomic Open-source Breeding informatics initiative

GOBii: Connecting genotyping data, breeding data and analysis tools to facilitate breeding decisions
GOBii Mission

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

[gobiiproject.org](http://gobiiproject.org)
GOBii A Global Community
The GOBii Project

- Provide database tools to manage genotyping data from multiple genotyping platforms for any crops
- Provide user interfaces to query across datasets by samples and markers
- Provide analyses and visualizing tools to support data curation and breeding decisions
- Provide solutions to integrate genotyping data management to adjacent breeding management, and sample tracking systems, and downstream tools
- Develop a community of knowledge through training and consulting
GOBii Genomic Data Management (GDM)
‘Designed with extract performance in mind’

IUPAC
Vcf
Codominant
Dominant
SSR alleles
.txt, .csv, .hmp

Flapjack
Hapmap

GOBii Data Load

GOBii Data Extract

Web services APIs

Compute & digest

Database
GOBii Genomic Data Management (GDM) 'Designed with extract performance in mind'

- IUPAC Vcf Codominant Dominant SSR alleles .txt, .csv, .hmp
- Flapjack Hapmap
- GOBii Data Load
- GOBii Data Extract
- Web services APIs
- Compute & digest
- Database

Bioinformatics for Large Scale Genotyping Data Management and Analytics: Mon Jan 14th 4pm

Demo GOBii Database: Digital Tools and Resources Tues Jan 15, 12:00pm
GOBii Tools

Curator tools: Data QC

MABC Ped ver Forward Breeding

<table>
<thead>
<tr>
<th></th>
<th>Par2</th>
<th>Par5</th>
<th>Testline1</th>
<th>Testline2</th>
<th>Testline3</th>
<th>Testline4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotype</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>A</td>
<td>T</td>
<td>C</td>
<td>T</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Demo Galaxy Pipeline: Digital Tools and Resources Tues Jan 15, 11.15am
APIs

Analysis and decision tools

Genotyping Data Management

Genomic selection

Suitable genotyping platform

Sample tracking

Breeding Management System

MAS
GOBii Integration

Breeding management
Sample tracking
Vendors

BrAPI

GOBii

Data visualization
Analysis pipelines
Decision Support
The sample tracking use case: from field to lab and back
Tracking samples back to plants

GOOD

BETTER

UUID - Universal Unique Identifier
Guaranteed to be different from any other UUID generated until 3400 A.D
Sample UUIDs

- Uniquely identify samples generated across databases and systems
- Can be generated in breeding management systems or in mini-apps ‘Sample Tracker’, ‘Simple Tracker’
- Downside - hard to read
- Not always usable in vendor submission systems
- Can use an associated human readable name tracked by a sample tracking system
- ‘Sample orchestrator’ will connect sample UUID between systems
Sample orchestrator

- PUT Plates, samples
- GET Project status
  - Samples received
  - In progress
  - Completed
- GET Data URL

Vendor

- Vendor ID
- GET Project status
- URL

Connecting sample UUIDs across vendors and databases

BrAPI
Sample orchestrator

Vendor

POST GOBii Project, samples
GOBii project ID
POST Vendor URL
GET vendor files
GET Data loading status
In progress
Done

GOBii
From GOBii to Flapjack analysis

Data extract
- Genotype file
- Map file
- Sample file
- QTL file

Marker group extract

User

MABC, FB
F1 and line pedigree verification
BrAPI Dataset View in Flapjack
GOBii Integration

BrAPI

BrAPI
Acknowledgement
Core System - GDM

Load, store, extract

Tools
Genomic selection
Imputation
Haplotype analysis
MABC, Pedigree Verification
Data QC

Integration
Tools
Vendors
Breeding management systems

Community
seminars, meetings, hackathons