A Public API for Crop Breeding Data

David E. Matthews
USDA-ARS, Dept. Plant Breeding and Genetics, Cornell University, Ithaca NY

PAG-XXIII, 12 Jan 2015

API:
Application Programming Interface

- Specification of syntax and semantics for data requests and responses
- Request via HTTP GET or POST, like web forms
- Response data is structured, using JSON format (Javascript Object Notation)

Breeding API (brapi)
Example

- Request
  http://<server>/brapi/0.1/germplasm/find?q=STEPTOE
- Response
  ```
  [
    {
      "queryName": "STEPTOE",
      "germplasmId": 437,
      "uniqueName": "STEPTOE"
    }
  ]
  ```

Breeding API
Goals

- A public, shared, open API for multiple repositories of data needed by crop breeding software applications
- Foster cooperation and communication among software developers and reduce unuseful duplication of effort
- Better software for breeders in south Asia and sub-Saharan Africa
- Initial targets: phenotype and genotype data
- Example client apps: Breeding View, ISMU
- Example data repositories: BMS, GOBII
Example Application

T3 Breeding API: genetic distances

Create a list of genotypes. Data on their genetic distance will be fetched from the desired server, and then genetic distances from each other will be included.

Endpoint "traits/list"

Requests and Responses

Genetic Distance Calculation with R

Shared Development and Documentation on Apiary

Project History

- May 2014, Seattle: Convened by Gary Atlin
- Participants: BMS, Leafnode, BIMS, KDDart, Cassavabase, GOBII, Triticeae Toolbox (T3), James Hutton Institute, iPlant
- Documentation site created on Apiary and GitHub
- Conference calls weekly since then
- Additions: Android Fieldbook, Genomes to Fields Initiative
- Workshop to follow PAG on Thursday
**Endpoint "Allele Count"**

**Current Status**

<table>
<thead>
<tr>
<th>Endpoint</th>
<th>Specified</th>
<th>Implemented</th>
<th>Sample Client App</th>
</tr>
</thead>
<tbody>
<tr>
<td>genotype/count</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>genotype/{id}</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>germplasm/find</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>study/list</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>study/{id}</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>traits/list</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>traits/{ids}</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

**A Public, Shared, Open API**

- Public API examples:
  - KDDart, Ensembl Genomes, Crop Ontology, 23andMe
  - Public, Shared, Open:
    - Global Alliance for Genomics and Health (GA4GH)
      - genotype data only
      - EBI, NCBI, Sanger Inst., UC Santa Cruz, Google

**Global Alliance for Genomics and Health, GA4GH**

**Links and related PAG events**

- Documentation: [http://docs.breeding.apiary.io](http://docs.breeding.apiary.io)
- Mon 3:30 "Genomics Tools and Services", BMGF, Fairfield Room
- Tue 1:50 "EnsemblAPI", California Room
- Thu 9am-5pm "Breeding API Workshop", Convene 4 Room, Marriott Courtyard
“Co-authors”

- Jan Erik Backlund, Integrated Breeding Program
- Gary Atlin, Bill and Melinda Gates Foundation
- Lukas Mueller, Cassavabase
- Clay Birkett, The Triticeae Toolbox (T3)

Documentation on GitHub