


















BrAPI Success Stories

This manuscript ([permalink](#)) was automatically generated from [plantbreeding/BrAPI-Manuscript2@7bafa2d](#) on February 29, 2024.

Authors

- **Peter “BrapMan” Selby** 
 [0000-0001-7151-4445](#) ·  [BrapiCoordinatorSelby](#)
Cornell University · Funded by NIFA-DSFAS 2022-67021-37024
- **Trevor “Cool Kid” Rife**
 [0000-0002-5974-6523](#) ·  [trife](#)
Clemson University
- **Khaled Al-Shamaa**
·  [khaled-alshamaa](#)
ICARDA
- **Isabelle Alic**
 [0000-0002-8961-6068](#) ·  [Isabelle-inrae](#)
INRAE
- **Sebastian “Baz” Raubach**
 [0000-0001-5659-247X](#) ·  [sebastian-raubach](#)
The James Hutton Institute
- **Iain Milne**
 [0000-0002-4126-0859](#) ·  [imilne](#)
The James Hutton Institute
- **Cyril Pommier**
 [0000-0002-9040-8733](#) ·  [cpommier](#)
Université Paris-Saclay, INRAE, BioinfOmics, Plant Bioinformatics Facility, Versailles, France; Université Paris-Saclay, INRAE, URGI, Versailles, France · Funded by French Infrastructure en Biologie Santé ‘Phenome-FPPN’ supported by the French National Research Agency (ANR-11-INBS-0012)
- **Becky Smith**
 [0000-0002-8968-3383](#) ·  [Batbaby91](#)
The James Hutton Institute
- **Chaney Courtney**
·  [chaneylc](#)
Clemson University

✉ — Correspondence possible via [GitHub Issues](#) or email to Peter “BrapMan” Selby <ps664@cornell.edu>.

Abstract

The Breeding API (BrAPI) project is an effort to enable interoperability among plant breeding databases. BrAPI is a standardized RESTful web service API specification for communicating plant breeding data. This community driven standard is free to be used by anyone interested in plant breeding data management. This manuscript describes updates and outlook for the current version of BrAPI.

Introduction

Plant and animal breeding is an incredibly important part of today's society. Almost every country in the world has some kind of breeding program supporting the agricultural community to produce bigger, better, healthier, more sustainable crops. Modern breeding techniques require large amounts of high quality data to be effective. In the digital age, that breeding data is being collected, managed, and analyzed with computer software. Interoperability between breeding software tools, systems, and databases can substantially increase the efficiency of a breeding program. The ability to share tools gives each program a boost in computational power. The ability to share data means everyone has access to larger, more complete, datasets and get build more accurate computational models and produce more accurate predictions.

The Breeding API (BrAPI) project is an effort to enable interoperability among breeding tools, systems, and databases. BrAPI is a standardized Representational State Transfer (REST), web service, Application Programming Interface (API), specification for breeding and related agricultural data. [1] By using the BrAPI standard, breeding software can more easily become interoperable, allowing groups to more easily share data and software tools.

How it works

An Application Programming Interface (API) is a technical connection between two pieces of software. Just as a Graphical User Interface (GUI) or a Command Line Interface (CLI) allows a human user to interact with a piece of software, an API allows one software application to interact with another. A GUI or CLI might allow a user to input data, read data, and start processes within an application. An API allows one piece of software (sometimes called a client, user agent, or service consumer) to programmatically input data, read data, and start process within another piece of software (sometimes called a server or service provider).

A Representational State Transfer (REST) web service is a type of API commonly used in today's modern web infrastructure. REST is a technical architecture that describes the stateless transmission of data between applications. Typically, REST systems are implemented using the standard HTTP protocol that most of the modern internet is built upon. REST implementations also generally use JavaScript Object Notation (JSON) to represent the data being transferred. Both HTTP and JSON are programming language agnostic, very stable, and very flexible. This means BrAPI can be implemented in almost any piece of software, and can solve a wide range of use cases.

Data repositories and service providers can choose to represent their data as a BrAPI compatible API. By mapping the internal data structures to the standard models, data repositories can easily expose data to the outside world. Similarly, they can accept new data from external sources and automatically map the new data into the existing database. Client application developers can take advantage of this standardization by building tools that can easily integrate with all other BrAPI compatible data repositories. Visualization, reporting, analytics, data collection, and quality control

tools can be built once and shared with other organizations following the standards. As the number of BrAPI compatible databases, tools, and organizations grows, so does the value added by implementing the standard into a given application.

Project Updates

Over its lifetime, the BrAPI project has grown and changed substantially. The latest stable version of the specification (v2.1) looks vastly different from the original version (v1.0) released in 2017. The total size of the specification has almost quadrupled in that time, going from 51 endpoints documented in v1.0 to 201 endpoints documented in v2.1. Because of this growth, the specification documents were reorganized into four modules: BrAPI-Core, BrAPI-Germplasm, BrAPI-Genotyping, and BrAPI-Phenotyping. Figure 1 shows a simplified domain map of the whole BrAPI v2.1 data model, divided into the organizational modules. The early versions of the specification focused on read-only phenotype data, with a small consideration to the other domains. Now the specification has a full representation of most of the major concepts applicable to the breeding process. The new specification is also internally consistent, easier to navigate, and allows for read, write, and update capabilities. None of those qualities were a guarantee for the earlier versions.

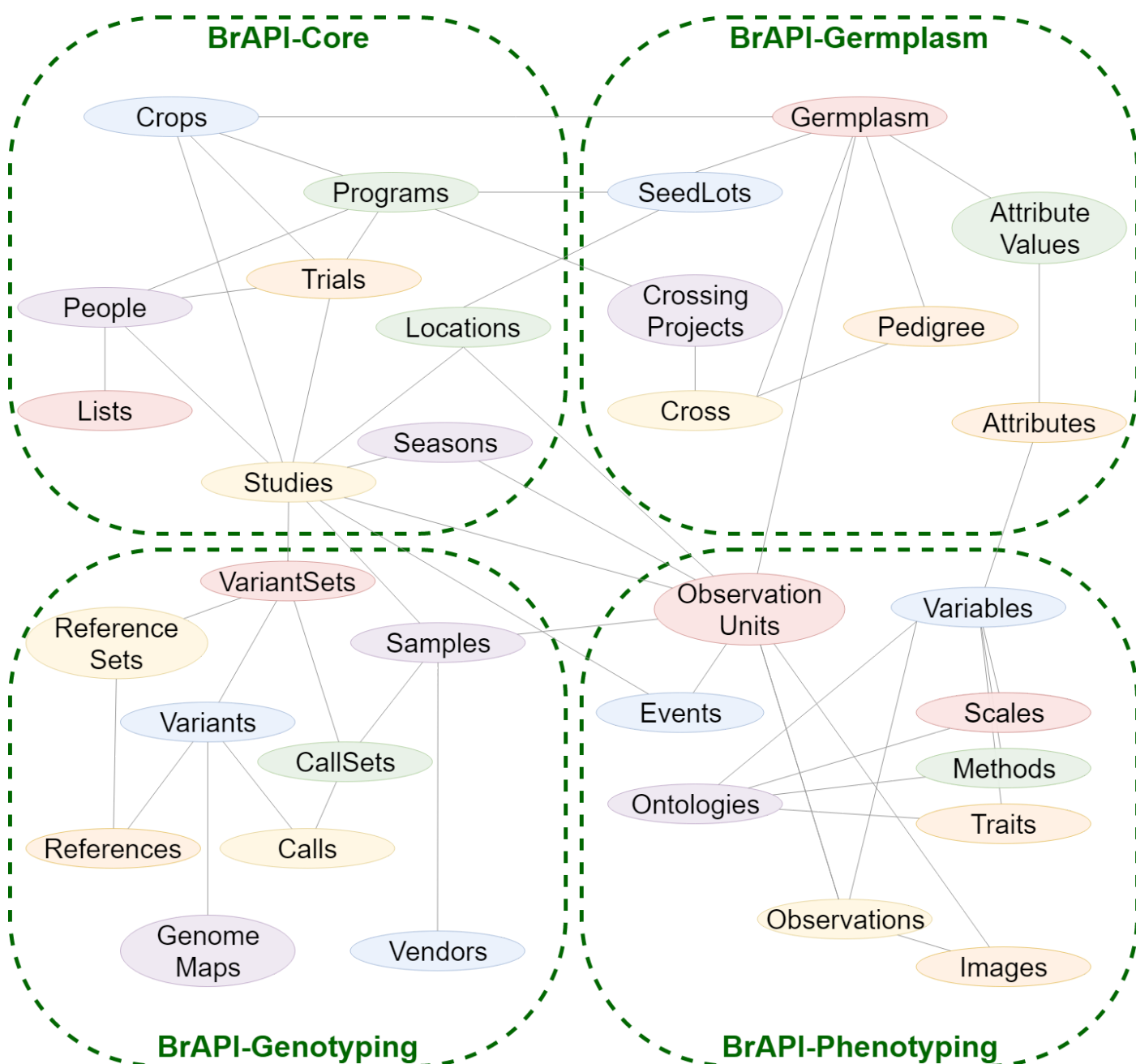


Figure 1: A simplified domain map of the whole BrAPI data model, divided into organizational modules. A more detailed Entity Relationship Diagram (ERD) is available on brapi.org.

As the specification has matured, so have the tools, services, and libraries available to the community to work with the specification. Every version of the specification is now released with a change log to guide developers upgrading from a previous version, an Entity Relationship Diagram (ERD) to describe the whole data model visually, and a JSON Schema version of the model to be used in some automated development efforts. For groups who are using Java, Java Script, Python, R, or Drupal, there are community maintained libraries available that contain full BrAPI implementations ready to be added to some existing code. The BrAPI Test Server and the BRAVA validation tool are both still available to the community for testing purposes, and they have been maintained to support every version of the specification. Finally, the three new resources list pages on brapi.org advertise the other BrAPI compatible software available in the community. The BrAPPs list displays the 10 standalone, plug-and-play, applications available to the community. The servers list displays the 27 registered public data servers, their current status, and a form for registering additional servers. The compatible software list shows the 31 software applications that are BrAPI compliant, and again, a form for registering additional applications. The process for registering new items for these lists is completely voluntary, so the totals represent a lower bound for the number of BrAPPs, data repositories, and applications available in the the community.

Community Growth

The international BrAPI Community consists of the software developers, breeders, and related scientists working on BrAPI related projects and data sources. This community is what sustains the BrAPI project, building implementations, maintaining development tools, and providing input to enhance the specification. As the project has grown, so has the community. The BrAPI project started in June 2014 with less than ten people coming together to discuss the idea. Over the next nine years, the community has grown to between 200 and 250 members. The community mailing list has 208 members, and the BrAPI Slack workspace has 234 members. The project leadership uses the mailing list to broadcast newsletters, announcements, and updates to the community. The BrAPI Slack workspace allows members of the community to discuss specific topics and collaborate directly with each other.

The BrAPI Hackathons are a major staple of the BrAPI community. Twice a year, the community gathers to discuss the specification and collaborate on BrAPI related projects. This time is very valuable to the community; for some organizations, the hackathon is the only time during the year when they have time to work on anything related to BrAPI. During the COVID-19 pandemic, virtual hackathons took the place of in person events. While the virtual hackathons do not provide the same level of face-to-face time that is crucial to collaborative work, they did allow for more attendees to gather and share their opinions. The typical virtual hackathon has about twice as many registered attendees compared to an in-person hackathon. However, attendees have reported much more productive work time during the in-person events. As a compromise, going forward, the community leadership has decided to have one in-person hackathon and one virtual hackathon each year, to balance the advantages of both.

As the project has matured, a formal project leadership structure became increasingly important. As described in Figure 2, the project governance is divided into two groups. The Project Management team is responsible for the day to day operations of the project. The PI and Co-PI are responsible for the project funds, and responsible for hiring the BrAPI Project Coordinator who is paid by the project funds. The Advisory Board is a group of elected officials representing the community. The board is responsible for long term planning of the project, as well as quick decision making on behalf of the community. The two groups meet quarterly to report on progress and stay synchronized.

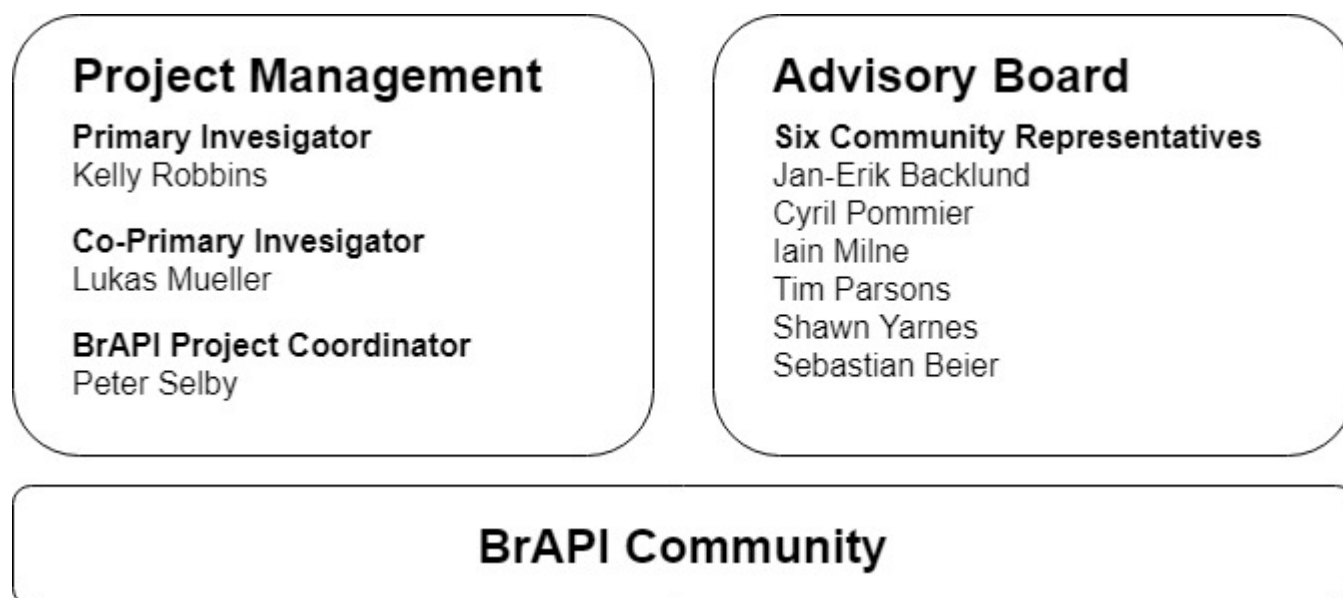


Figure 2: The formal governance structure of the BrAPI Project Community

Success Stories

Below are a number of short success stories from the BrAPI community. These tools, applications, and infrastructure projects serve as another indicator of community growth and success over the past 5-10 years. These stories clearly illustrate all the different ways the BrAPI Standard can be used productively and in practice.

Data Collection

Field Book

Phenotypic data collection is an essential part of the breeding process. Historically, gathering data in the field was done with pen and paper, or perhaps some version of a digital spreadsheet. The abundance and prevalence of smart phones has allowed the Field Book mobile app to enhance data collection. Field Book can create well formed digital observation records from the moment they are taken. This can improve the efficiency of data collection and reduce human error.

BrAPI was able to take things a step further by automating the flow of data from the Field Book mobile app to a central database server. In this work flow, data is collected and stored completely digitally with little-to-no human involvement.

ClimMob

Not all data can be collected by a single person, or even by a single organization. ClimMob is a tool to easily allow citizen scientists to assist in the data collection process. Although this data may not be as detailed as a focused scientific program, it can be very useful to collect simple data from a wide range of locations and environments.

When it comes to BrAPI compatibility, ClimMob follows the same patterns established by Field Book. During a survey, all the farmer collected data is stored in a central ClimMob node. When the survey is complete, all the data is uploaded automatically via BrAPI to a central breeding database for long term storage and analysis.

ImageBreed

High-throughput phenotyping has been gaining significant traction lately as a way to collect lots of data very quickly. Image collection from unmanned aerial and ground vehicles (UAVs and UGVs) are a great way to collect a lot of raw data all at once, then analyze it later. ImageBreed is a image collection pipeline tool to support regular use of UAVs and UGVs.

When the raw images have been processed through the standardization pipelines in ImageBreed, useful phenotypes can be extracted from the images. The BrAPI standard is used to push these phenotypes back to a central breeding database where they can be analyzed with other data. In addition to this, ImageBreed also has the option to use BrAPI to upload the raw images to the central breeding database, or any other BrAPI compatible long term storage service. The BrAPI models in the current version of the standard (V2.1) are rudimentary, but effective. The ImageBreed team has put in some work to enhance the BrAPI image data standards.

Data Management

PHIS

The Hybrid Phenotyping Information System ([PHIS](#)), based on the [OpenSILEX](#) framework, assumes responsibility for the systematic collection and management of data from phenotyping and high-throughput phenotyping experiments on a day-to-day basis. PHIS has the ability to efficiently store, organize and manage a wide range of data sets, including images, spectra and growth curves. This functionality extends to data at multiple spatial and temporal scales, from leaf to canopy, from a variety of sources such as field and greenhouse environments.

A key feature of PHIS is the unambiguous identification of all objects and traits within an experiment, establishing consistent relationships between them through the application of ontologies and semantics. This approach is designed to adapt to variations in experimental conditions, whether in the field or in controlled environments. PHIS's ontology-driven architecture emerges as a robust tool for integrating and managing data derived from diverse experiments and platforms, facilitating the creation of meaningful relationships between objects and augmenting datasets with relevant knowledge and metadata.

Furthermore, PHIS adheres to the Minimal Information About a Plant Phenotyping Experiment ([MIAPPE](#)) and the Breeding API (BrAPI) standards.

The system recommends specific naming conventions, fostering a standardized approach for users to declare their resources. Notably, PHIS is widely adopted by various experimental platforms of the national [PHENOME](#) and European [EMPHASIS](#) infrastructure, serving as a hub for data management. Moreover, dedicated instances of PHIS have been established for the explicit purpose of resource sharing, encompassing projects, genetic resources, and variables, thereby fostering collaborative engagement and the dissemination of knowledge pertaining to studied concepts.

PHIS & BrAPI

PHIS offers a RESTful API designed to streamline interaction with data within a platform. Within this API, various services aligning with the Breeding API (BrAPI) standards have been implemented, encompassing the Core, Phenotyping, and Germplasm modules. Comprehensive documentation for these services is available on the PHIS Swagger interface. This integration with Swagger ensures that users can easily access, understand, and utilize the functionalities provided by the BrAPI-compliant web services, fostering transparency and facilitating effective engagement with the PHIS platform.

The design and development of PHIS have been meticulously tailored, incorporating the explicit constraint of aligning with BrAPI requirements. This intentional alignment ensures that PHIS adheres to the specified standards and protocols outlined by the Breeding API (BrAPI), thereby fostering seamless integration and compatibility with BrAPI-compliant systems and platforms. This prerequisite served as a substantial foundation for formalizing the data model, simultaneously facilitating compatibility with other standards, such as MIAPPE. Thus, by consciously incorporating BrAPI requirements into its structure, PHIS not only meets the phenotyping domain standards but also enhances its capacity for interoperability and effective collaboration within the broader context of plant breeding and related domains.

The fact that data within a PHIS instance can be queried through BrAPI services makes the indexing of PHIS in [FAIDARE](#) very easy to implement.

Indeed, as PHIS offers BrAPI-compliant Web Services, this greatly simplifies the integration and data exchange with other European information systems that handle phenotyping data. The adherence to BrAPI standards ensures a common interface and compatibility, facilitating seamless communication and collaboration between PHIS and other systems within the European context. This interoperability not only streamlines data sharing but also promotes a more cohesive and effective approach to managing and utilizing phenotyping data across diverse platforms and research initiatives in the European scientific community.

DeltaBreed

DeltaBreed is an open-source data management system designed and developed by Breeding Insight to support USDA-ARS specialty crop and animal breeders. DeltaBreed is a unified system for managing breeding data that connects a variety of BrAPI applications (see list below). BrAPI integration allows the complexity underlying interoperability to be hidden, shielding users from multifactorial differences between diverse applications. DeltaBreed, adhering to the BrAPI model, establishes data standards and validations for users and provides a singular framework for data management and user training.

DeltaBreed users need not be aware of BrAPI or the specifics of underlying applications but will notice that BrAPI interoperability reduces the need for human-mediated file transfers and data manipulation. Field Book users, for example, can connect to their DeltaBreed program, authenticate, and pull studies and traits directly from DeltaBreed to Field Book on their data collection device. The subsequent step of pushing observations from Field Book to DeltaBreed is straightforward via BrAPI, but will not be implemented until repeated observation handling workflows are established to differentiate and validate repeated observations, such as accidental repeats, overwrite requests, time-series observations, and repeated sub-entity measures. Users can expect DeltaBreed observation handling to become more seamless with future development.

DeltaBreed Connected Applications << Submission is expected April 2024. We may need to trim this aspirational list down to reality in final edits.>>

- BIMS <https://www.breedwithbims.org/>
- BrAPI Java Server <https://test-server.brapi.org/brapi/v2/>
- BrAPI Sync <https://github.com/IntegratedBreedingPlatform/brapi-sync>
- BreedBase <https://breedbase.org/>
- Diversity Arrays Technologies (DART) genotyping services
- Field Book <https://play.google.com/store/apps/details?id=com.fieldbook.tracker>
- Gigwa <https://gigwa.southgreen.fr/gigwa/>
- Mr Bean <https://github.com/AparicioJohan/MrBeanApp>
- Pedigree Viewer <https://github.com/solgenomics/BrAPI-Pedigree-Viewer>

BMS

[brapi_sync](#) and other things

Breedbase

Since the inception of BrAPI in 2014, the Breedbase group has been a heavy contributor to the BrAPI community.

BIMS

BIMS is relatively new to the BrAPI community, with their primary focus being a link to the Field Book mobile data collection app.

Germinate

Germinate is a combination breeding database and germplasm management database.

Federated Data Management Infrastructures

AGENT

BraPI endpoints for AGENT

IPK-Genebank

Agrosystem Integration of germplasm collections in context of data trustee models among private economy and public research, integration of ex-situ genebanks (EU H2020 projects AGENT, INCREASING), integrated agrosystems and plant research infrastructure

MIAPPE “BrAPI to ISA” service

MIAPPE and BrAPI are designed to be inter-compatible. ISA-TAB is a file based implementation of MIAPPE. The “BrAPI to ISA” service is a converter between the ISA-TAB files and the BrAPI RESTful endpoints.

Data visualization

Flapjack

Flapjack is a desktop application for visualizing and analyzing genotype data. Data can be easily imported into Flapjack from any BrAPI compatible data source with genotype data available. Flapjack Bytes is the smaller, web based counterpart to Flapjack, which can be imbedded into a database website and provide the visualizations online.

Helium

Helium is a platform for reviewing and managing pedigree data. It is BrAPI enabled, so pedigree data can be viewed straight from a database.

Tassel

I don't know much about Tassel or its BrAPI compliance. This is filler text for the layout of the manuscript.

Trait Selector BrAPP

BrAPPs are simple tools developed by the BrAPI community that are entirely reliant on BrAPI for their data requirements. This means a single BrAPI can be shared and used by many organizations, as long as those organizations have the standard BrAPI endpoints available. For example, the Trait Selector BrAPP is used to search and select useful traits, using a visual aid to help the user find exactly what they need. The BrAPP relies only on the BrAPI Trait endpoints to function, so any data resources with a the Trait endpoints available can add this tool to their system.

DArTView

DArTView is a desktop application for visualizing genotype variant data and looking for trends or correlations. It is newly BrAPI compatible and can use BrAPI as an input data source.

DivBrowse

DivBrowse is a BrAPI compatible tool for visualizing genotype variant data.

Analytics

QBMS

Modern breeding programs can utilize data management systems to maintain both phenotypic and genotypic data. Numerous systems are available for adoption. To fully leverage the benefits of digitalization in this ecosystem, breeders need to utilize data from different sources to make efficient data-driven decisions. With increased computational power at their disposal, scientists can construct more advanced analysis pipelines by combining various data sources.

To meet this demand, many breeding management systems have developed customized built-in analysis pipelines. However, these pipelines are static and may not accommodate evolving needs. As a result, API interfaces have been developed to facilitate data communication with other systems. The Breeding API (BrAPI) project specifies a standardized interface for plant phenotype/genotype databases, enabling them to share data with crop breeding applications. This promotes interoperability among plant breeding databases and allows third-party plugins to integrate with the ecosystem, delivering added value.

In the QBMS development team, we have identified a technical barrier between the breeding management systems' BrAPI interface and the scientists who create analysis scripts and pipelines. This barrier arises from the complexity of managing API backend processes, such as authentication, tokens, TCP/IP protocol, JSON format, pagination, stateless calls, asynchronous communication, database IDs, and more. To bridge this gap, we have developed the QBMS R package. This package abstracts the technical complexities, providing breeders (our end users) with stateful action verbs/functions that are familiar to them when navigating their GUI systems. This enables them to query and extract data into a standard data frame structure, consistent with their use of R language, one of the most common statistical tools in the breeding community.

Since its release on the official CRAN repository in October 2021, the QBMS R package has garnered over 7250 downloads. Several tools, such as MrBean, rely on the QBMS package as their source data adapter. Moreover, the community has started building extended solutions on top of it. QBMS can serve as a cornerstone in the breeding modernization revolution by providing access to actionable data and enabling the creation of dashboards to reduce the time between harvest and decision-making for the next breeding cycle.

Mr. Bean

Mr. Bean is an analytics and visualization tool. Mr. Bean has a data import tool which can pull data directly from a BrAPI compatible source.

G-Crunch

G-Crunch is a new general analysis pipeline tool, meant for automated, production level, analysis tasks. Part of that automation relies on BrAPI to access data.

Samples and Genotypes

DArT Sample Submission

The DArT genotyping lab is heavily used world wide when it comes to plant genotyping. Developers at DArT have worked with the BrAPI community to establish a standard API for sending sample metadata to the lab before genotyping. This eliminates much of the human error involved with sending samples to an external lab.

MGIS

MGIS has germplasm and genotype data stored for many musa accessions. Through BrAPI, users are able to access this data directly from MusaBase, for use in specific experiments.

GIGWA

GIGWA is an efficient storage system for genotype variant data. GIGWA uses BrAPI to query specific variant data out of the database. This allows for more efficient data transfer and analysis. Instead of transferring whole massive files, specific pieces, samples, markers, or chunks of data can be retrieved.

PHG

The Practical Haplotype Graph (PHG) is a graph-based computational framework that represents large-scale genetic variation and is optimized for plant breeding and genetics. Using a pangenome approach, each PHG stores haplotypes (the sequence of part of an individual chromosome) to represent the collected genes of a species. This allows for a simplified approach for dealing with large scale variation in plant genomes. The PHG pipeline provides support for a range of genomic analyses and allows for the use of graph data to impute complete genomes from low density sequence or variant data.

Users access the crop databases either with direct calls to the PHG embedded server or indirectly using the rPHG library from an R environment. The PHG server accepts BrAPI endpoint queries to return information on sample lists and the variants used to define the graph's haplotypes. In addition, PHG users utilize the BrAPI variantsets endpoint query to return links to VCF files containing haplotype

data. Work on the PHG is ongoing. We expect to support additional BrAPI endpoints that allow for slicing genotypic data based on samples and regions.

Data Portal

FAIDARE

FAIDARE is an indexing and search portal for the wider breeding community. While it can extract data from various sources via a flat file, the preferred method uses the BrAPI standard APIs. As data is added or updated in data repositories around the world, FAIDARE keeps its index up to date automatically through BrAPI.

Phenospex - HortControl

HortControl, developed by Phenospex, is a data repository. HortControl has a BrAPI implementation to be used to automate workflows and analytics software.

GLIS

GLIS is a service for creating DOIs for agricultural purposes. It also holds basic germplasm information about each accession when a DOI is created for one. GLIS is BrAPI compatible, allowing external searches on the large GLIS database.

Discussion

BrAPI for Breeders and Scientists

The BrAPI technical specification document is meant to be read and used by software developers. However, the purpose of the specification, and the community around it, is to make things faster, easier, and cheaper for the breeders and scientists working to make the world a better place. BrAPI offers a convenient path to automation and data integration for software tools in the breeding domain. All of the example use cases described above can be achieved with manual effort, moving and editing data files by hand. However, when the basic structure and flow of data becomes automated, breeders and scientists can spend less time on data management and more time focussing on the science, doing what they do best. For many, the ultimate goal is the development of a digital ecosystem: a collection of software tools and applications that can all work together seamlessly. In this digital ecosystem, data is collected digitally from the beginning, reducing as much human error as possible. The data is checked by quality control and stored automatically, then can be sent to any internal tool or external lab for further analysis with just the click of a button. This idea might sound too good to be true, but as more tools start sharing a universal data standard, automating data flow becomes easier, and the community gets closer to total interoperability.

Looking Ahead

The BrAPI specification will continue to grow, enabling more use cases and new types of data. These new use cases might include newer scientific techniques and technologies. Things like drone imaging data, spectroscopy, LIDAR, metabolomics, transcriptomics, high-throughput phenotyping, and machine learning analysis. All of these technologies can open new avenues for research and development of new crop varieties. All of these technologies also generate more data, and require data sharing between different software applications and data repositories. The BrAPI project leadership and community is committed to building the standards to support these new use cases as

they arrive and become accepted by the scientific community. In fact, small groups within the BrAPI community have already start building generic data models and communication standards for many of the technologies listed above. These community efforts will eventually become part of the BrAPI standard in a future version of the specification document.

Conclusions and Impact

- High level summary of the project/consortium
- BrAPI is fitting into this gap, it doesn't need to fit these other gaps
- Call to action - Join us!

References

1. **BrAPI—an application programming interface for plant breeding applications**
Peter Selby, Rafael Abbeloos, Jan Erik Backlund, Martin Basterrechea Salido, Guillaume Bauchet, Omar E Benites-Alfaro, Clay Birkett, Viana C Calaminos, Pierre Carceller, Guillaume Cornut, ...
Bioinformatics (2019-03-23) <https://doi.org/gjgxxr>
DOI: [10.1093/bioinformatics/btz190](https://doi.org/10.1093/bioinformatics/btz190) · PMID: [30903186](https://pubmed.ncbi.nlm.nih.gov/30903186/) · PMCID: [PMC6792114](https://pubmed.ncbi.nlm.nih.gov/PMC6792114/)