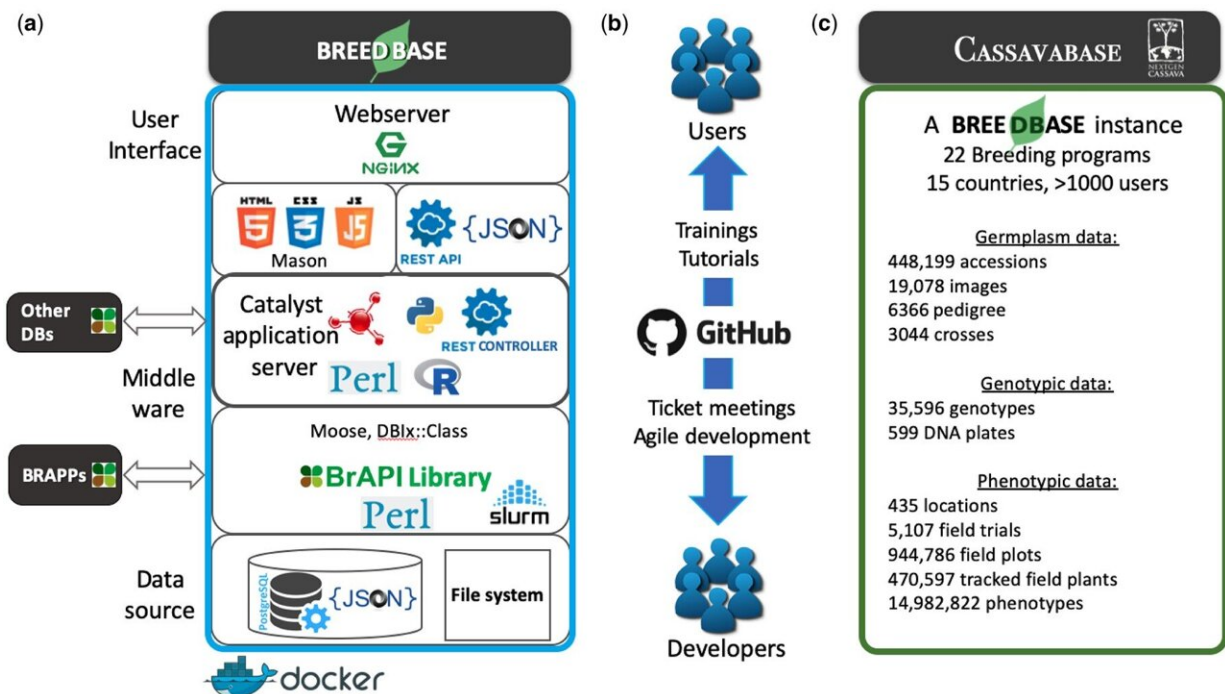


# Breedbase software to help speed crop improvement

October 14 2022



a) Breedbase platform architecture. User interface: To offer a dynamic, highly interactive user interface, several JavaScript libraries are implemented including D3, JQuery, and Bootstrap. RESTful APIs, including a full BrAPI 2.0 implementation, handle the communication between the front and back end, allowing fast calculations without reloading the website. HTML5 for interactive graphical display, allowing instant reorganization of visual elements. The Bootstrap framework is used for modern and dynamic page templating. Middleware layer: A Perl software stack including Mason components to connect to the user interface, a Catalyst a web application framework, Moose an object oriented perl library and DBIX::Class an object-relational mapper to connect to SQL code. In addition, BrAPI libraries are used. Finally a job cluster

scheduler, Slurm is implemented to allocate server resources and ensure scalability. Data source layer: Breedbase operates on a relational database using Postgres. Postgres 12.0 offers “Big data” solutions including parallel query execution and optimized binary JSON data type handling. Binary JSON (JSONB) is a simple data structure designed to be storage space and scan-speed efficient. In Breedbase, JSONB is used in various data types including genotypic (marker) information. In addition to the relational database a standard file system space is available for flat files. Finally, other databases can communicate to a Breedbase instance to provide additional back-end for marker data [i.e. Genomic Open Source Informatic Initiative (GOBii)] or to exchange germplasm information for example. b) Breedbase codevelopment process. User–developers interactions are promoted using various media. Users have online access to documentation (<https://solgenomics.github.io/sgn/>, last accessed 4/18/2022), video tutorials, or through onsite training. Software development goals are extensively discussed between developers, data managers, breeders, and other appropriate stakeholders. Agile development allows short-term product release. Suggested improvements, issues, and bugs discovered in Breedbase are submitted and tracked on the public GitHub issue tracking software (<https://github.com/>, last accessed 4/18/2022). Software development progress is tracked using a version control system and Docker releases. c) Cassavabase, a breedbase instance: data content overview. Cassavabase involves national and international breeding programs (22) from various African and South American countries (15) and currently has 1,131 registered users. Cassavabase hosts various data types including high-density and low-density genotyping assays (35,000), plot-based phenotypic data points (near 15 million), images from plants and plots from trials (5107) and locations (435). Credit: *G3 Genes|Genomes|Genetics* (2022). DOI: 10.1093/g3journal/jkac078

To help plant breeders speed crop improvement around the world, Lukas Mueller of the Boyce Thompson Institute worked with an international team of 57 people to create Breedbase, a database software that was described in the July issue of *G3 Genes|Genomes|Genetics*.

"In the current era of population growth and [climate change](#), [plant](#)

[breeding](#) needs to be faster to ensure crops survive new pests and pathogens that are expanding their ranges, as well as unpredictable weather patterns," said Mueller.

Plant breeding is the process by which people improve plant traits, such as increasing yield or making them resistant to disease. Breeders and farmers traditionally have done this by crossing plants that have desirable traits, like larger and tastier fruit. But traditional plant breeding is a long and slow process, taking generations to achieve results.

In today's genomics era, plant breeding has undergone drastic changes. In an approach termed "Genomic Selection," breeders determine the genomic properties of plant lines and correlate them with traits, which allows them to predict traits based on genomic information. Plant breeders make decisions based on these predictions much faster than they would with the traditional approach of growing and observing the plants.

However, genomic approaches generate massive amounts of data that can be challenging to manage, especially for smaller breeding programs in developing countries. To be of use to breeders and researchers, the data need to be stored in specialized databases with precise organization, data management, quality control and analytics.

Up until now, plant breeders and researchers have typically collected data in non-standardized ways using spreadsheets, making it difficult to organize, share and analyze data with each other.

"Breedbase solves these problems by creating a common data language, and a free data tracking system that will change the way plant breeders communicate and archive all important breeding data," says Nicolas Morales, a graduate student in Mueller's group.

"Clear, organized data management and analysis is crucial to successful and efficient plant breeding. In order to grow plants that feed people in a nutritious and healthy way, plant breeding [data management](#) needs to be simplified, standardized and accessible to everyone who needs it," says Mueller, who is also an adjunct professor at Cornell University's School of Integrative Plant Science.

In addition to storing data, Breedbase includes algorithms that a breeder could run, such as predicting whether a plant variety has a particular trait, such as disease resistance or high yield.

"Breedbase makes complicated things easy. It's like a giant tool box with all the tools you need in one central place," says Morales, who is co-first author of the paper with former Mueller graduate student Alex Ogonna.

A key component of Breedbase is the Breeding Application Programming Interface (BrAPI), which standardizes how data are collected. This standardization allows plant breeders to more easily exchange data among disparate databases and computer-based breeding tools. For example, if a person wants to collect data in the field and has no internet connection, they can collect data on a tablet using an app, and then download the data to a database when they return to their computer, using a BrAPI interface behind the scenes.

Importantly, everything is standardized, so that a farmer growing corn in Iowa and another farmer growing corn in Africa will be able to easily share their data with each other, speeding up discoveries to improve crop traits.

"Empowering plant [breeders](#) in developing countries allows even smaller breeding programs to leverage genomic information to make breeding selections and help feed the world," says Mueller.

Breedbase is based on Cassavabase, which the Mueller Lab developed with NextGen Cassava, a project that brought cassava breeding to the next level at institutions in Africa and uses cutting-edge tools to efficiently deliver improved varieties of cassava.

In addition to cassava, at least 50 crop databases already use Breedbase, including yam, bananas, sweet potato, rice, tomatoes and carrots.

**More information:** Nicolas Morales et al, Breedbase: a digital ecosystem for modern plant breeding, *G3 Genes|Genomes|Genetics* (2022). [DOI: 10.1093/g3journal/jkac078](https://doi.org/10.1093/g3journal/jkac078)

Cassavabase: [www.cassavabase.org/](http://www.cassavabase.org/)

Provided by Boyce Thompson Institute

Citation: Breedbase software to help speed crop improvement (2022, October 14) retrieved 19 July 2024 from <https://phys.org/news/2022-10-breedbase-software-crop.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.