Science Societies

Making *waves* in breedbase

Free, Open-Source R-Package Makes Handheld Spectrometers Easier to Use

By Dj McCauley

March 19, 2021

Jenna Hershberger interacts with students and plant breeders at the Uganda National Crops Re

Jenna Hershberger interacts with students and plant breeders at the Uganda National Crops Resources Research Institute. Photo courtesy of Jenna Hershberger.

- Handheld spectrometers are fast becoming affordable, widespread tools for plant breeders to non-destructively analyze samples in the field.
- Data analysis and storage can be difficult and unwieldy, but researchers from Cornell and the Boyce Thompson Institute created a free, open-source R-package, *waves*, so breeders can analyze and share their data and models, within or outside of the existing Breedbase platform.
- Here, authors of the recent article in *The Plant Phenome Journal* discuss the *waves* function and its creation.

Jenna Hershberger didn't mean to make *waves*.

Instead, like many of the most revolutionary tools, *waves* was born of necessity. It's a free, open-source R-package for statistical analysis of data collected from handheld spectrometers. It works on its own, or dovetails neatly with **Breedbase**, an existing platform for storing and analyzing data from a variety of crops around the world.

Here, Hershberger, an ASA, CSSA, and SSSA member, details how scientists in the Plant Breeding and Genetics Section at Cornell collaborated with the Boyce Thompson Institute team that runs and maintains Breedbase. The outcome of their collaboration is the formal release of the *waves* R-package through a recent article in *The Plant Phenome Journal* (**https://doi.org/10.1002/ppj2.20012**.

Analyzing Spectral Data

When Hershberger began her doctoral work at Cornell under adviser Michael Gore, handheld spectrometers weren't really on her radar.

"We were trying to think of the most useful addition we could make to existing Android smartphone apps for plant phenotyping," Hershberger says, "but we talked with a lot of cassava breeders, and they kept coming up with quality traits you can't just count or measure with a normal image."

Enter spectral data. In a technique called visible and near-infrared spectroscopy (vis-NIRS), researchers blast a sample of plant matter with wavelengths of light energy, then analyze the radiation that sample emits in return. Depending on both the color and chemical composition of the sample, researchers can identify spectral characteristics that correspond to certain traits. In the case of our cassava breeders, they can gather information about both water and carotenoid content through spectral imaging.

Breeders have long been using benchtop spectrometers to make selection decisions, but analyzing samples in the lab requires breeders to destroy the sample in preparation for measurements. Cassava breeders, for example, often shred storage roots before analysis. Handheld spectrometers create a range of opportunities for selecting samples without destroying them.

Enoch Webabzi, a graduate student at West Africa Center for Crop Improvement, uses a handheld spectrometer to analyze cassava roots at the Uganda National Crops Resources Research Institute. Photo by Jenna Hershberger.

“You can do things like test fruit while it’s still developing or scan your leaves while they’re still growing,” Hershberger says. “But there are still a lot of moving parts you have to think through before you can effectively use it.”

Some of those moving parts involve data storage. Many handheld spectrometers—including the highly affordable $300 models the Cornell team tested—sync via Bluetooth to a smartphone. If you’ve ever tried to organize all the photos of your pets, kids, soil cores, and test fields that end up in your camera roll, you can probably sympathize with their frustrations.

But the small price tag means researchers can potentially employ a whole fleet of handheld spectrometers in the field for way less than the cost of a typical benchtop spectrometer.

“Benchtop machines are so expensive [that] you can’t even find the price online. You have to call and get a quote!” Hershberger says.

The typical benchtop spectrometer sets a laboratory back $50,000 to $100,000. Plus, benchtop models are often repositories for data that may never make it out of the lab.

While the Cornell team worked to solve these issues facing plant breeders, Hershberger was diligently analyzing spectral data using R, the statistical analysis language and software. She was working with existing models that calibrate handheld spectrometers, comparing indicators from spectral data with the outcome of laboratory analyses. Her end goal was to make sure the indicators really do correlate with the plant characteristics breeders are really trying to measure.

Making *Waves*

“I wrote and re-wrote a *bunch* of code,” Hershberger says. “After a while, it just made sense to make it into an R-package, even if for just my own use. But other people were asking me if they could use it, and some weren’t as well versed in R.”

Enter the Boyce Thompson Institute (BTI). Also housed on the Cornell campus, the institute hosts a software called Breedbase that plant breeders worldwide use to analyze and store data. It has servers for a variety of crops already, including cassava, rice, banana, sweetpotato, and tomatoes. It’s free to use and open source—breeders can submit packages, like Hershberger’s *waves*, and BTI integrates new features and tools into the system for other people to use. Plus, it works in your internet browser.

“It just made sense to integrate *waves* with Breedbase,” Hershberger says.

The Breedbase platform solves two problems for handheld spectrometer users. By incorporating *waves* as a function within the Breedbase system, researchers can both store their spectral data in the same place as corresponding laboratory data that backs it up. For example, spectral data that correlate certain signals to high levels of carotenoids are tied to laboratory analysis from samples that demonstrate carotenoid content on a chemical level. Over time, researchers will calibrate models so that they can rely on the spectrometer data alone, cutting out the time-consuming process of analyzing samples in the lab.

But Hershberger couldn't do it on her own. One programmer served a critical role in bringing *waves* to users who are not on great terms with the R coding language. Nicolas Morales was a programmer at BTI when his principal investigator, Lukas Mueller, tasked him with creating an interface that brings *waves* to Breedbase users.

"Jenna wrote the R-package—the back-end code—but I took it and built a web interface that calls forward the functions Jenna wrote," Morales explains.

Indispensably, the work Morales and his colleagues at BTI completed integrates the *waves* function in an existing suite of tools that researchers already use within the Breedbase platform.

Sharing Data, Enabling Collaboration

"The web database really enables collaboration," Morales says. "If all this information just sits on someone's desktop computer in a lab somewhere, you would never move forward. Now, there are people accessing the same data from all different places, and we can better understand how to use it."

For a nascent technology like handheld spectrometers, sharing data is an invaluable tool for researchers in the field, sparking ideas between labs and providing new means for researchers to understand how their phenotypic results correlate with models and how they can increase the speed and accuracy of selections within their breeding programs without destroying samples to analyze them.

“This is an idea from plant breeders themselves that democratizes both the equipment and its functions, and there’s a lot of protocol development that will have to take place in a breeding program before you can actually put it in your routine,” Hershberger says. “But once it’s there, it has the potential to really increase the throughput of phenotyping.”

If tools like *waves* and Breedbase can help breeding programs on a tight budget (and what program isn’t on a tight budget?) find more efficient ways to make selections, we can decrease the time and energy it takes to create new plant varieties. In a rapidly changing environment, any tool that can help breeders select for the phenotypes we most need to see—from drought tolerance to nutritional content to disease resistance—all the better.

Dig Deeper

“Making Waves in Breedbase: An Integrated Spectral Data Storage and Analysis Pipeline for Plant Breeding Programs,” in *The Plant Phenome Journal* at **https://doi.org/10.1002/ppj2.20012**. Also check out the article on p. 9 of this issue comparing handheld and benchtop near-infrared spectrometers (see **https://doi.org/10.1002/csan.20439**).

More science articles

Back to issue

Back to home