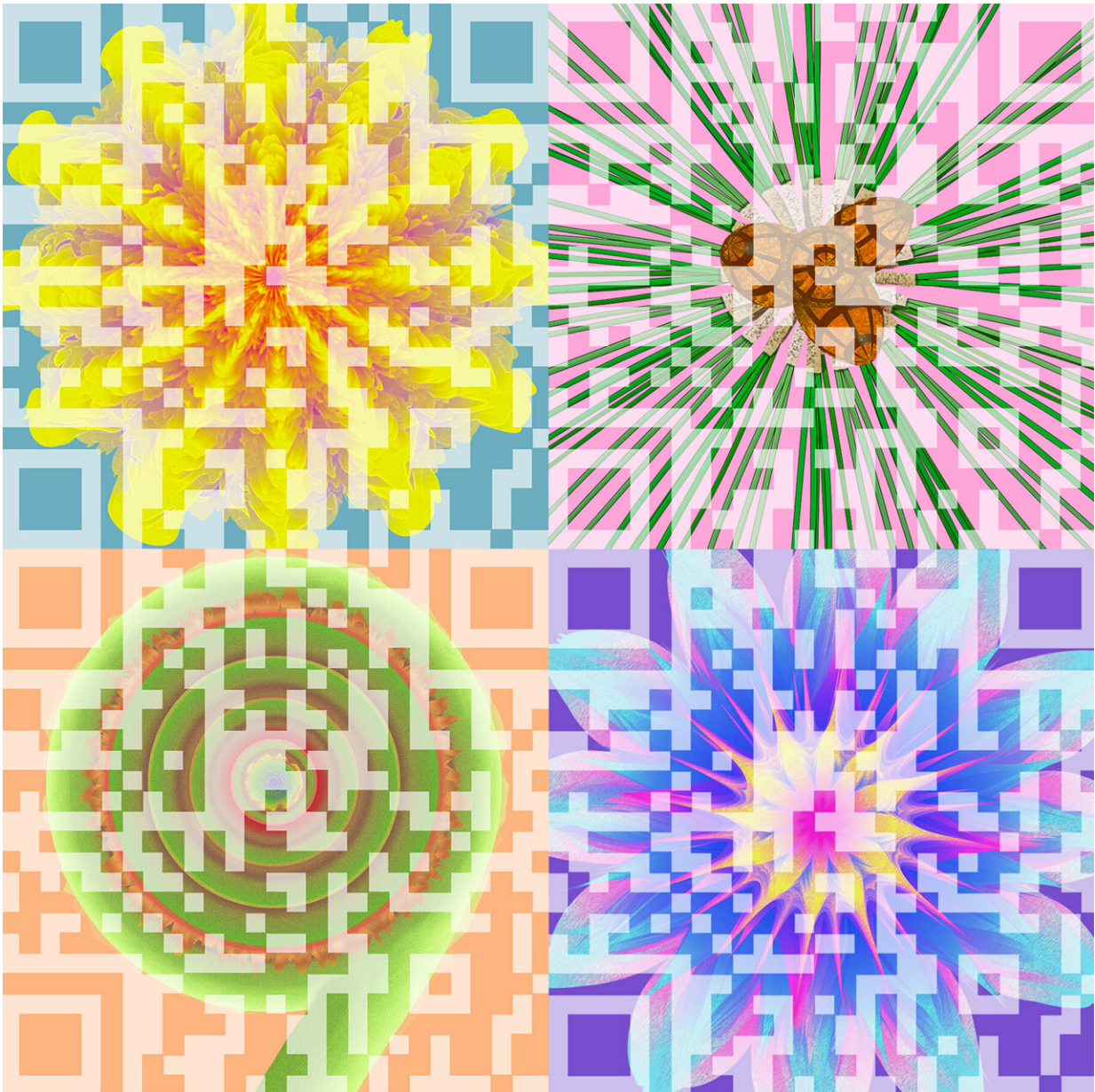


Unlocking herbaria biodiversity using a QR code sampling-to-sequencing workflow

April 29 2021



A new automated collection process utilizing unique object identifiers and QR codes has the potential to revolutionize the way in which biologists acquire data from museum specimens. Credit: Jerald Pinson

Within the past decade, next-generation sequencing technologies have revolutionized the way in which genetic data are generated and analyzed. In the field of phylogenetics, this has meant that researchers are rapidly reconstructing the tree of life, a goal that biologists have been working toward since Darwin sketched the first phylogeny in his notebook in 1837.

Yet despite the relative ease with which DNA can now be sequenced in large quantities, scientists must first extract that DNA from an organism, often relying on vast numbers of curated specimens in museums and herbaria. With over 250,000 species in the [plant kingdom](#) alone, the acquisition and documentation of specimen material is now by far the most time-consuming and error-prone process in large studies.

In a research article published in a recent issue of *Applications in Plant Sciences*, researchers undertook the goal of automating the collecting process by using a combination of unique object identifiers, QR codes, and citizen science.

"Our goal was to create a resource for the scientific community," said lead author Ryan Folk, an assistant professor and herbarium curator at Mississippi State University. "In the future, we hope that all such collection information will be available online, where it's easy to find and the work won't need to be repeated."

Folk and his colleagues are working to create a partial—50 percent coverage—phylogeny of seed plants that harbor nitrogen-fixing bacteria

in specialized root nodules. This symbiotic relationship spans several disparate seed plant groups that collectively contain more than 30,000 species, for which the team relied entirely on [herbarium specimens](#).

For a project of this scale, that meant members of the team worked for weeks at a time at multiple herbaria, focusing their efforts primarily at the New York and Missouri Botanical Gardens and the California Academy of Sciences.

Ordinarily, this would involve sifting through specimen material and manually transcribing or copying voucher information (such as the specimen locality, date, and name of the collector(s)) into a spreadsheet, as well as manually copying labels onto the samples themselves.

This process is absolutely essential for downstream analyses, but it requires large amounts of time and creates the potential for error. The biggest drawback, however, is the information garnered at this stage is typically only useful to the researchers who collected it and cannot be easily shared between groups. If a different group of researchers wanted to extract DNA from the same specimens, they would have to re-collect all of the same data.

Folk and his colleagues, wanting to curtail this duplication of effort, devised a digitization workflow whereby voucher information would only need to be collected once.



This herbarium specimen of *Neptunia oleracea* was among those used in the present study, with the QR code prominently displayed on the destructive sampling bag. Photographs of each specimen and the corresponding QR code were taken and deposited in a database for anatomical reference. Credit: California Academy of Sciences

"Our workflow is made up of essentially several strategies—more or less connected—that takes you straight from walking into a museum all the way through data analysis and publication," said Folk.

At a glance, the process has three steps. First, a unique object identifier is assigned and physically attached to each herbarium voucher, and the specimen is photographed with the identifier clearly visible.

To transcribe the data from the roughly 15,000 specimens they used for the study, they used the citizen science platform [Notes From Nature](#), which offers an online, interactive workspace where volunteers can join specific projects from home and communicate with researchers involved in the study.

By itself, this digitization of specimen information with unique object identifiers will be a valuable tool for researchers in the future and may ultimately complement the monumental effort being undertaken in museums around the world to digitize collections. But the researchers didn't stop there.

In the third step, QR codes were assigned to each specimen to further streamline data collection. This meant that when extracting and amplifying DNA in the lab, the researchers no longer had to manually

enter specimen data into spreadsheets. Instead, that information was auto-populated when they scanned a given QR code.

"Scanning QR codes was basically effortless and didn't require any training or set up," said study co-author Heather Kates, a post-doctoral associate at the University of Florida. "But more important than the time-saving aspect of this approach was the reduction of errors. The errors introduced through illegibility and typos are a pain when you're doing a set of 20 extractions; you can imagine how important it is with 15,000 that those errors are avoided as much as possible."

Rather than keeping all of this information in spreadsheets, which have limited utility for specialized queries or data analysis, the team also created their own database using Python scripts.

The detailed workflow and Python scripts have been made freely available online so that researchers can access and fine-tune them for personalized use in their own studies.

"Although we have seen large DNA datasets published in recent years, reuse of data from many projects has been minimal due to the difficulty of accessing data online," said Folk. "In the long-term, we will release the data to the community in a form more easily amenable to future researchers. My hope is that our efforts will jump-start major projects focused on other organisms as well as establish a new baseline for biodiversity in the nitrogen-fixing clade of seed plants."

More information: Ryan A. Folk et al, High-throughput methods for efficiently building massive phylogenies from natural history collections, *Applications in Plant Sciences* (2021). [DOI: 10.1002/aps3.11410](https://doi.org/10.1002/aps3.11410)

Provided by Botanical Society of America

Citation: Unlocking herbaria biodiversity using a QR code sampling-to-sequencing workflow (2021, April 29) retrieved 24 June 2024 from <https://phys.org/news/2021-04-herbaria-biodiversity-qr-code-sampling-to-sequencing.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.