

An entire botanical garden of genomes

January 25 2019



Researchers worked with experts from the Ruili Forestry Bureau to collect plant materials growing in the Ruili area. This was a 45-day effort to collect 1,093 plant samples, in addition to samples, geographic coordinates were recorded for future assessment. Credit: China National GeneBank

An article published today in the Open-Access journal *GigaScience* provides data that effectively triples the number of plant species with available genome data. This mammoth amount of work comes on the back of the growing efforts of the scientific community to sequence more plant genomes to aid in understanding their complex evolution and provide practical information for improving agricultural yield. To date, around 350 land plant genomes have been sequenced. The desire for more plant genome sequences has recently been highlighted with the announcement of the 10KP project, which aims to ultimately sequence 10,000 plant genomes to resolve the evolution of all the major branches of the plant tree of life. The work here provides images, raw sequencing data, assembled chloroplast genomes, and preliminary nuclear genome assemblies- all freely available. Effectively this work is a digital representation of an entire botanical garden.

Researchers from the China National GeneBank, BGI, and the Forestry Bureau of Ruili, China have sampled and sequenced 761 samples, representing 689 vascular plant species from 137 families and 49 orders. The plant samples are all from in and around the 500-hectare Botanical Garden in Ruili, a subtropical part of China bordering Myanmar. Being in a biologically rich part of China, the garden is committed to protecting endangered and Chinese-endemic plants, including the preservation and archiving of these germplasm resources to assist with their long-term conservation. This project is the world's first scientific and systematic attempt to digitize a whole botanical garden based on genomic as well as voucher specimen information.

On the scientific potential of this resource, BGI's CEO and author on the paper Xun Xu highlights that: "Current understanding of the evolution of plants and their diversity in a phylogenomic context is limited because of the lack of [genome](#)-scale information across phylogenetically diverse species. This innovative project integrates a new way of thinking about the digitization of all the plant species to augment evolutionary and

ecological research in botanical gardens."

In total, the researchers produced 54 terabytes of sequencing data, with an average sequencing depth of 60X per species. In addition to the basic challenge of carrying out DNA sequencing on this number of species, another major task was scaling up the species identification, digitizing images of the specimens, and building a new herbarium for their storage at a new China National GeneBank (CNGB) herbarium in Shenzhen. So far, of the 761 specimens, sequence and chloroplast data has enabled the identification of 257 [plants](#) at the species level and 504 at the family level. Deep learning has also been successful applied to 181 species to enable them to be identified to the species level.



The collection of seeds from the sampled plants that will be used for the China Nation GeneBank Herbarium that is currently under construction. Credit: China National GeneBank

Author Ting Yang says that this was "the largest amount of data I have ever processed. During the data analyses, I think the biggest challenges was sequence checking and results examination." This required researchers to individually check each of the 761 sample's sequencing data, and compare the chloroplast gene sequences with herbarium specimens for species identification.

Another difficulty relating to simply getting to the point of being able to do the sequencing work was collecting all the samples. Author Jinpu Wei states: "We cooperated with experts from the Ruili Forestry Bureau to collect plant materials distributed in the area of Ruili for the establishment of a digital botanical garden. After 45-days of tiring effort, we collected 1,093 plant materials. Although it was challenging for us to transport the materials properly, we finally managed to ensure the high quality of these plant materials for future research."

Corresponding author, Xin Liu, adds that the project "was a baseline project to fine tune and standardize the sampling, methodologies, and the data accumulation and analyses techniques for large-scale genome projects like the 10KP (10 thousand Plant Genome Project). From this project, we have gained considerable and useful experience for subsequent sample collection, sequencing, and assembly. At the same time, the data produced from this study can be effectively used in subsequent genome projects."

Despite having constructed only one sequencing library for each species, the authors were able to assemble preliminary genomes for 17 of them, reflecting the quality and reuse potential of the DNA. Researchers at the Chinese University of Hong Kong have already independently assembled the genomes of species of particular interest to them. The potential for the wider research community to study their species of interest, improve other genomes, develop tools and methods, and provide education opportunities for new generations of scientists is enormous.



A plant sample that has been prepared and catalogued for imaging. Another form of digital data that is available as component of the sequencing and sampling

data. Credit: China National GeneBank

Lead author Huan Liu added that "Genomic characterization will provide a large amount of basic data for plant genome assembly, which will be an excellent start for the 10KP project. At the same time, it lays a good foundation for the future research on the correlation mechanism from macroscopic ecology and biodiversity to microscopic molecular level."

To promote more extensive data sharing than just making sequence data available, the researchers are also making the digitized images available and providing access to the herbarium. The Herbarium (HCNGB) serves as a living plant database that records the position of species grown in the Ruili Botanical Garden and monitors the status of each species.

All the digital data generated here (images, raw sequencing data, assembled chloroplast genomes, and preliminary nuclear genome assemblies) are available via the NCBI SRA, *GigaScience* GigaDB database and China National GeneBank CNSA. Additionally, to enable the data to be searched and genomes and [species](#) identification to be updated, metadata is indexed and linked via Datacite and GigaDB. And all resources are released without restriction under a CC0 waiver. Author Dr. Sunil Kumar Sahu highlighted that this is the most important legacy of the project "This dataset is of great value to plant researchers, and more importantly, can serve as a reference for future planetary-scale genome sequencing projects including the Earth BioGenome Project (EBP) and 10 thousand Plant Genome Project (10KP)."

More information: Liu H et al. (2019): Genomic and Imaging Data Supporting the Digitization of Ruili Botanical Garden GigaScience Database. [DOI: 10.5524/100502](https://doi.org/10.5524/100502)

Huan Liu et al. Molecular digitization of a botanical garden: high-depth whole genome sequencing of 689 vascular plant species from the Ruili Botanical Garden, *GigaScience* (2019). [DOI: 10.1093/gigascience/giz007](https://doi.org/10.1093/gigascience/giz007)

Provided by GigaScience

Citation: An entire botanical garden of genomes (2019, January 25) retrieved 25 April 2024 from <https://phys.org/news/2019-01-entire-botanical-garden-genomes.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.