

Opinion note

It's time to celebrate! Linking genetic resources to the Mexican National Biological Collections custodied by the Institute of Biology of the National Autonomous University of Mexico

¡Es tiempo de celebrar! Vinculando recursos genéticos a las Colecciones Biológicas Nacionales mexicanas custodiadas por el Instituto de Biología de la Universidad Nacional Autónoma de México

Carolina Granados Mendoza ^{a, *}, Miguel Murguía-Romero ^b,
Gerardo A. Salazar ^a

^a Universidad Nacional Autónoma de México, Instituto de Biología, Departamento de Botánica, Circuito Zona Deportiva s/n, Ciudad Universitaria, 04510 Ciudad de México, Mexico

^b Universidad Nacional Autónoma de México, Instituto de Biología, Unidad de Informática para la Biodiversidad, Circuito Zona Deportiva s/n, Ciudad Universitaria, 04510 Ciudad de México, Mexico

*Corresponding author: carolina.granados@ib.unam.mx (C. Granados Mendoza)

Received: 26 May 2024; accepted: 06 August 2024

Abstract

The Instituto de Biología of the Universidad Nacional Autónoma de México (IBUNAM) celebrates 95 years as a leading institution on biodiversity research. Besides developing frontier research, training of human resources, and public science communication, IBUNAM houses 10 National Zoological Collections and the National Herbarium of Mexico (MEXU). The specimens deposited in these collections are the foundation of numerous biological studies, from which ample genetic resources have been generated. Here we present an initial effort to link the specimens deposited at the National Biological Collections housed at IBUNAM to their public genetic resources, using MEXU's Collection of Types of Vascular Plants as proof of concept. First, a list of the type specimens was retrieved from IBdata, the web system for consulting the records of the biological collections housed at IBUNAM. Then, the interface Entrez Programming Utilities of GenBank was used to search for the available genetic resources associated with the type specimens. New fields were incorporated into IBdata to facilitate access to the identified genetic resources. Future initiatives should promote access to the public metadata (e.g., molecular, morphological) associated to specimens of the biological collections housed at IBUNAM.

Keywords: Metadata; GenBank; IBdata; Type specimens; Digitalized information

Resumen

El Instituto de Biología de la Universidad Nacional Autónoma de México (IBUNAM) cumple 95 años como institución líder en investigación de la biodiversidad. Además de desarrollar ciencia de frontera, formación de recursos humanos y comunicación pública de la ciencia, el IBUNAM alberga 10 Colecciones Zoológicas Nacionales y el Herbario Nacional de México (MEXU). Los ejemplares depositados en estas colecciones son fundamento de numerosos estudios biológicos, de los cuales se han generado amplios recursos genéticos. Aquí se presenta un primer esfuerzo para vincular los ejemplares depositados en las Colecciones Biológicas Nacionales albergadas en el IBUNAM con sus recursos genéticos públicos, utilizando como prueba de concepto la Colección de Tipos de Plantas Vasculares del MEXU. Primero, se recuperó una lista de especímenes tipo de IBdata, el sistema web que permite consultar los registros de las colecciones biológicas alojadas en el IBUNAM. Luego, se utilizó la interfaz Entrez Programming Utilities de GenBank para buscar los recursos genéticos disponibles asociados a los especímenes tipo. Se incorporaron nuevos campos a IBdata para facilitar el acceso a los recursos genéticos identificados. Iniciativas futuras deberían promover el acceso a los metadatos públicos (e.g., moleculares, morfológicos) asociados a los especímenes de las colecciones biológicas albergadas en el IBUNAM.

Palabras clave: Metadatos; GenBank; IBdata; Especímenes tipo; Información digitalizada

The Instituto de Biología of the Universidad Nacional Autónoma de México (IBUNAM) celebrates its 95th anniversary this year. Faculty and students at IBUNAM are devoted to the study, conservation, and sustainable use of the biota of Mexico, but also from other regions of the world. The research performed at IBUNAM touches on virtually all branches of the tree of life and uses a wide variety of methodological and analytical tools to discover, describe, document, and understand biological diversity. Among other research institutions in Mexico, IBUNAM stands out for housing several National Biological Collections, including 10 National Zoological Collections and the National Herbarium of Mexico (MEXU). The specimens deposited at the IBUNAM's National Biological collections are the foundation of a myriad of taxonomic, evolutionary, ecological, biogeographic, social, and conservation studies, from which an enormous amount of associated data (hereafter referred to as "metadata") is generated.

Every day, these collections are actively consulted, both in person and virtually through IBdata (<http://ibdata4.ib.unam.mx>), a web system for consulting the records of the National Biological Collections housed at IBUNAM (Murguía-Romero et al., 2024). Under UNAM's open data policy (<http://www.datosabiertos.unam.mx/informacion/terminosdeuso.html>), IBdata currently provides free, easy, and continuous access to digitalized information of over 1.7 million biological specimens, allowing the dissemination of knowledge and transdisciplinary research, thus benefiting the scientific, governmental, and educational society sectors, as well as private users. For each physical specimen, the digitalized information available in IBdata usually includes high-resolution digital images along with data on the locality where the specimen

was collected (including geographic coordinates, when available), date of collecting, collector(s), as well as notes on habitat, morphological, and socio-cultural aspects recorded by the collectors.

One of the commonly generated metadata are genetic resources, which are often made publicly available through the International Nucleotide Sequence Database Collaboration (INSDC; <https://www.insdc.org/>), which includes 3 international databases that exchange data every day, namely the DNA Databank of Japan (DDBJ; <https://www.ddbj.nig.ac.jp/index-e.html>), the European Nucleotide Archive (ENA; <https://www.ebi.ac.uk/ena/browser/>), and GenBank of the USA National Center for Biotechnology Information (NCBI; <https://www.ncbi.nlm.nih.gov/genbank/>). When properly submitted, these genetic resources include information about the voucher specimens of the genetic data, as well as where the specimens are deposited. Access to such genetic information is essential for the sustainable use and conservation of global biodiversity (Cowell et al., 2022).

Here we used the information of MEXU's Collection of Types of Vascular Plants available in IBdata (10,972 records) to search and link the specimens to their public genetic resources available at GenBank. For this, we downloaded all the type records and built URL calls for the interface Entrez Programming Utilities (E-utilities; <https://www.ncbi.nlm.nih.gov/books/NBK25501/>) of GenBank. Query searches used the species' scientific name and the collection number assigned by the collector or the unique identifier of the specimens (MEXU's catalogue number) as in the following example: "[https://eutils.ncbi.nlm.nih.gov/entrez/eutils/esearch.fcgi?db=nucleotide&term=%22Agave%20isthmensis%22\[organism\]+AND+\(4177+OR+628489\)](https://eutils.ncbi.nlm.nih.gov/entrez/eutils/esearch.fcgi?db=nucleotide&term=%22Agave%20isthmensis%22[organism]+AND+(4177+OR+628489))".

In cases where the collection number contained non-numerical characters or spaces, query searches used instead the main collector's last name (only the first last name was used when 2 last names were present). URL calls were submitted to the nucleotide NCBI database, with a 1 s delay between each search request to avoid overload of the NCBI E-utility servers, with the Bulk URL Opener extension of Google Chrome. Calls with hits were saved, and the corresponding query translations were used to download the associated GenBank accession numbers. Retrieved accession numbers were merged into a text file to perform a NCBI Batch Entrez search (<https://www.ncbi.nlm.nih.gov/sites/batchentrez>). The resulting records were further filtered with a custom filter using the flag "MEXU." The filtered results were revised individually to confirm their association with one vascular plant type specimen deposited at MEXU. To facilitate linking back the type specimens to their associated genetic metadata, IBdata records of type specimens having genetic information available at GenBank were complemented with a new data field called "GenBank", which displays a list of available molecular markers and their corresponding GenBank accession numbers. Additionally, a web link leading to the corresponding NCBI records was added as an additional data field named "GenBank Search" in IBdata (Fig. 1).

Our search identified 71 GenBank accession numbers corresponding to 23 angiosperm species representing 20 genera, 7 families, and 5 orders (Table 1). Type specimens found associated with genetic resources at GenBank can be easily accessed in IBdata through the "Simple Search" option and the keyword "genbanksearch". The sequenced

molecular markers include 18 plastid regions: genes *accD*, *atpA*, *matK*, *ndhF*, *psbA*, *rbcL*, and *ycf1*; the introns of the genes *rpl16*, *trnL*; the intergenic spacers *rpl20-rps12*, *rpl32-trnL*, *rps16-trnQ*, *trnS-trnG*, and *ycf6-psbM*; the regions *trnD-trnT*, *trnH-psbA*, *trnK-matK*, and *trnL-trnF*; and different portions of the nuclear-ribosomal Internal Transcribed Spacer (ITS) region. Each type specimen had 1 to 7 associated sequences, being the most frequently sequenced markers the plastid regions *trnL-trnF* and *trnK-matK* and the nuclear-ribosomal ITS region. Most sequenced type specimens were collected in the 1980's, 1990's, and 2000's. However, the sequenced isotype specimen of *Epiphyllum chrysocardium* Alexander (Cactaceae) was collected in 1951 (Fig. 1). Authors of the sequences of the latter type specimen indicated us that the plant tissue used for DNA extraction came indeed from the type collection (MacDougall, 198), but from a division maintained under cultivation at the Botanical Garden of IBUNAM, explaining why sequencing was achieved from such an "old" specimen. Although explicitly stated for only 19 accessions, all the markers recovered seem to have been generated through capillary (Sanger) sequencing.

Previous studies have stressed the importance of open access to the digitalized information of type specimens (Nicolson et al., 2023), which are key reference elements of scientific names. The value of both the digitalized information of type specimens and the genetic information derived from them increases when both elements can be linked and easily accessed. Including genetic sequences from type specimens into molecular taxonomic studies often plays an important role in the circumscription of taxa or their placement at a particular place of the tree

Table 1
MEXU's types of vascular plants with available genetic information at GenBank.

Scientific name	GenBank accession number	Collector, collection number	Type category
Asparagales			
Asparagaceae			
<i>Agave isthmensis</i> García-Mend. & F. Palma	MN900422.1	García Mendoza, 4177	Holotype
<i>Agave rzadowskiana</i> P. Carrillo, Vega & R. Delgad.	MN900449.1	Carrillo-Reyes, 1503	Isotype
<i>Agave tenuifolia</i> Zamudio & E. Sanchez	MN900461.1	Carranza, 1905	Isotype
<i>Yucca mixtecana</i> García-Mend.	MN900508.1, MN893703.1	García Mendoza, 6198	Holotype
<i>Milla valliflora</i> J. Gut. & E. Solano	MF189697.1, MF189646.1, MF189596.1	Gutiérrez, 1151	Holotype
Orchidaceae			
<i>Bletia riparia</i> Sosa & Palestina	KU054381.1, KU054368.1, KU054356.1, KU054344.1	Palestina, 590	Isotype

Table 1. Continued

Scientific name	GenBank accession number	Collector, collection number	Type category
<i>Dichromanthus yucundaa</i> Salazar & García-Mend.	FN996950.1, FN996962.1	García Mendoza, 8774	Holotype
<i>Encyclia</i> × <i>nizandensis</i> Pérez-García & Hágsater	KP057187.1, KM385692.1, KM385889.1, KM386017.1	Pérez-García, 2085	Holotype
<i>Galeoglossum cactorum</i> Salazar & C. Chávez	FN645940.1, FN645939.1	Chávez-Rendón, 1604	Holotype
<i>Malaxis molotensis</i> Salazar & J.R. Santiago	HG970131.1, HG970153.1	Santiago, 1320	Holotype
<i>Myrmecophila christinae</i> Carnevali & Gómez-Juárez	EF065697.1	Carnevali, 4445	Isotype
Asterales			
Asteraceae			
<i>Sinclairia ismaelis</i> Panero & Villaseñor	JN837193.1, JN837373.1, JN837476.1, JN837283.1	Panero, 3572	Holotype
Caryophyllales			
Cactaceae			
<i>Epiphyllum chrysocardium</i> Alexander	KU598136.1, KU598186.1, KU597978.1, KU597925.1, KU598083.1, KU598030.1	MacDougall, 198	Isotype
<i>Selenicereus dorschianus</i> Ralf Bauer	LT745712.1, LT745480.1, LT745595.1	Böhme, s/n	Isotype
<i>Cephalocereus parvispinus</i> S. Arias, H.J. Tapia & U. Guzmán	MK165436.1, MK165437.1, MK165439.1, MK165435.1, MK165434.1, MK165433.1, MK165438.1	Tapia Héctor, 38	Holotype
Nyctaginaceae			
<i>Mirabilis polonii</i> Le Duc	KY952455.1	Le Duc, 259	Paratype
Cucurbitales			
Cucurbitaceae			
<i>Microsechium gonzalo-palomae</i> Lira	JN560193.1, JN560568.1, JN560294.1, JN560474.1, JN560640.1	Lira, 1230	Holotype
<i>Sicyos davilae</i> Rodrí.-Arév. & Lira	JN560230.1, JN560595.1, JN560330.1, JN560507.1, JN560663.1, JN560419.1	Lira, 949	Paratype
<i>Sicyos dieterleae</i> Rodrí.-Arév. & Lira	JN560232.1, JN560596.1, JN560332.1, JN560509.1, JN560664.1, JN560421.1	Lira, 1385	Isotype
Fabales			
Fabaceae			
<i>Caesalpinia oyamae</i> synonym of <i>Erythrostemon oyamae</i> (Sotuyo & G.P. Lewis) Gagnon & G.P. Lewis	KX373079.1, KX379300.1	Hawkins, 23	Holotype
<i>Phaseolus albescens</i> McVaugh ex R. Delgad. & A. Delgado	AF115150.1, DQ445955.1	Delgado, 1705	Holotype
<i>Platymiscium calyptratum</i> M. Sousa & Klitg.	EU735872.1, EU735933.1, EU735990.1, EU736047.1	Tenorio, 126	Holotype
<i>Harpalyce torresii</i> São-Mateus & M. Sousa	PP250089.1, PP238799.1	Téllez, 950	Paratype

Ficha resumen de datos del ejemplar

MEXU: TIPOS Plantas Vasculares



Nombre aceptado: *Epiphyllum chrysocardium*
Alexander

Última determinación: *Epiphyllum chrysocardium* Alexander

Familia: Cactaceae

Colector(es): Thomas MacDougall

Número de colecta: 198

Fecha de colecta (aaaa-mm-dd): 1951-02-09

País - Estado - Municipio: México - Chiapas -

Localidad: region of Selva Negra, between Tapalapa and Blanca Rosa near Santa Lucía and the Río Napak

Lat, Lon:

Elevación:

Número de catálogo: 72938

Identificador único (URN): IBUNAM:MEXU:T72938

GenBank:

- > matK gene: KU597925.1
- > rpl16 intron: KU598030.1
- > rps16-trnQ IGS: KU598136.1
- > trnH-psbA region: KU597978.1
- > trnL-trnF region: KU598083.1
- > ycf1 gene: KU598186.1

GenBank Search: [https://www.ncbi.nlm.nih.gov/search/all/?](https://www.ncbi.nlm.nih.gov/search/all/?term=KU598136.1,KU598186.1,KU597978.1,KU597925.1,KU598083.1,KU598030.1)

[term=KU598136.1,KU598186.1,KU597978.1,KU597925.1,KU598083.1,KU598030.1](https://www.ncbi.nlm.nih.gov/search/all/?term=KU598136.1,KU598186.1,KU597978.1,KU597925.1,KU598083.1,KU598030.1)

Figure 1. IBdata “Summary data sheet of the specimen” of the isotype of *Epiphyllum chrysocardium* (MEXU’s catalogue number: 72938) showing within a red rectangle two new implemented fields, “GenBank” and “GenBank Search”, which link the specimen to its available genetic information in GenBank.

of life. Explicit recognition of the inclusion of genetic sequences from type specimens in molecular studies can promote the progress of molecular systematics and taxonomy (Chakrabarty, 2010).

Given the improvements in sequencing technology, sequencing of type and non-type herbarium specimens should seek to incorporate more efficient sequencing strategies that maximize the amount of generated sequence data. The combination of supervised sampling of herbarium specimens with high-throughput DNA sequencing and bioinformatics has given rise to “herbariomics” i.e., the access to genome-scale genetic

information from specimens maintained in herbaria. Such an approach opens the possibility of incorporating in genomic, phylogenomic, and population genetic studies taxa that otherwise may not be accessible, such as extinct or extremely rare species, or species that live in places difficult to access or subjected to regulations for collecting (Davis, 2023; Strijk et al., 2020). The wealth of already available, potential sources of new genomic information is informed by the recent report by Thiers (2023) on the world’s herbaria, based on data from the Index Herbariorum (<https://sweetgum.nybg.org/science/ih/>): the 3,567 active herbaria in the world hold over 396.7 million

specimens. It should be a priority to incorporate those valuable sources of already collected, curated specimens in world-wide initiatives such as the “global biodiversity cyberbank” (Wen et al., 2017), aimed at integrating all the existing resources to promote free access and generation of information on biological diversity.

References

- Chakrabarty, P. (2010). Genotypes: a concept to help integrate molecular phylogenetics and taxonomy. *Zootaxa*, 2632, 67–68. <https://doi.org/10.11646/zootaxa.2632.1.4>
- Cowell, C., Paton, A., Borrell, J. S., Williams, C., Wilkin, P., Antonelli, A. et al. (2022). Uses and benefits of digital sequence information from plant genetic resources: Lessons learnt from botanical collections. *Plants, People, Planet*, 4, 33–43. <https://doi.org/10.1002/ppp3.10216>
- Davis, C. C. (2023). The herbarium of the future. *Trends in Ecology & Evolution*, 38, 412–423. <https://doi.org/10.1016/j.tree.2022.11.015>
- Murguía-Romero, M., Serrano-Estrada, B., Salazar, G. A., Sánchez-González, G. E., Melo-Samper, U., Gernandt, D. S. et al. (2024). The IBdata Web System for Biological Collections: design focused on usability. *Biodiversity Informatics*, 18, 1–12. <https://doi.org/10.17161/bi.v18i.20516>
- Nicolson, N., Trekels, M., Groom, Q. J., Knapp, S., & Paton, A. J. (2023). Global access to nomenclatural botanical resources: Evaluating open access availability. *Plants, People, Planet*, 5, 899–907. <https://doi.org/10.1002/ppp3.10438>
- Strijk, J. S., Binh, H. T., Ngoc, N. V., Pereira, J. T., Slik, J. W. F., Sukri, R. S. et al. (2020). Museomics for reconstructing historical floristic exchanges: divergence of stone oaks across Wallacea. *Plos One*, 15, e0232936. <https://doi.org/10.1371/journal.pone.0232936>
- Thiers, B. M. (2023). *The world's herbaria 2022: a summary report based on data from Index Herbariorum*. Retrieved on May 17th, 2024 from https://sweetgum.nybg.org/science/wp-content/uploads/2023/11/The_Worlds_Herbaria_2022.pdf
- Wen, J., Harris, A., Ickert-Bond, S. M., Dikow, R., Wurdack, K., & Zimmer, E. A. (2017). Developing integrative systematics in the informatics and genomic era and calling for a global Biodiversity Cyberbank. *Journal of Systematics and Evolution*, 55, 308–321. <https://doi.org/10.1111/jse.12270>