

## GGBN: Making Genomic Collections Discoverable for Research through a Networked Community of Biodiversity Repositories<sup>1</sup>

Jonathan Coddington<sup>2</sup>, Katharine Barker<sup>2</sup>, Gabriele Droege<sup>3</sup>, Jonas J. Astrin<sup>4</sup>, Paul Bartels<sup>5</sup>, Carol Butler<sup>2</sup>, David Cantrill<sup>6</sup>, Félix Forest<sup>7</sup>, Birgit Gemeinholzer<sup>8</sup>, Donald Hobern<sup>9</sup>, Jacqueline Mackenzie-Dodds<sup>10</sup>, Éamonn Ó Tuama<sup>9</sup>, Gitte Petersen<sup>11</sup>, Oris Sanjur<sup>12</sup>, David Schindel<sup>2</sup> and Ole Seberg<sup>11</sup>

### ABSTRACT

Advances in DNA extraction techniques, combined with next generation sequencing technologies, provide new tools for the study of non-human genetics and evolution. However, the accelerating loss of biodiversity and the demand to support and improve conservation efforts in a time of increasing sensibilities worldwide regarding collecting activities, increasing sensitivity regarding national biodiversity patrimony and unsolicited commercialization has increased the demand for accessible information on non-human genomic collections worldwide. It is increasingly urgent for biodiversity collection-holding institutions to set a standard of collaboration towards excellence in collections stewardship, information access and sharing, and responsible and ethical use of collections. The Global Genome Biodiversity Network (GGBN) seeks to meet this need by making genomic collections discoverable for research through a networked community of biorepositories and providing trusted and transparent access to genomic samples for users and contributors through an access and benefit sharing framework.

### INTRODUCTION

Genome sequencing for biodiversity analysis is at the forefront of innovation and discovery due to technological advances and the sequencing of whole genomes in the last 10 years. Information generated from biodiversity genomics will revolutionize our approach to taxonomy, phylogeny, conservation, ecological monitoring, wildlife management, agriculture, drug development, zoonotic disease forecasting and even aspects

of national security. Consequently the demand is rapidly increasing for professionally preserved, managed and document samples that yield high-molecular weight DNA and RNA from throughout the tree of life (e.g., Pawlowski et al., 2012; Brown, 2013). Many ambitious projects with the potential to revolutionize biodiversity research, such as the Ten Thousand Vertebrate Genomes Project (Genome 10K, <https://genome10k.soe.ucsc.edu/>), the Global Invertebrate Genomics Alliance (GIGA, <https://genome10k.soe>

<sup>1</sup> This work represents a partial re-use of Droege, G., K. Barker, J. Astrin, P. Bartels, C. Butler, D. Cantrill, et al. 2014. The Global Genome Biodiversity Network (GGBN) Data Portal. *Nucleic Acids Res.* 42(D1): D607–D612. doi: 10.1093/nar/gkt928

<sup>2</sup> National Museum of Natural History, Smithsonian Institution, Washington DC 20013, USA

<sup>3</sup> Botanic Garden and Botanical Museum Berlin-Dahlem, Freie Universität Berlin, Berlin 14195, Germany

<sup>4</sup> Zoological Research Museum Alexander Koenig, Bonn 53113, Germany

<sup>5</sup> Wildlife & Environment Society of South Africa, Pretoria 0001, South Africa

<sup>6</sup> National Herbarium of Victoria, Royal Botanic Gardens Melbourne, South Yarra, VIC 3141, Australia

<sup>7</sup> Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3DS, UK

<sup>8</sup> Justus-Liebig-Universität Giessen 35392, Germany

<sup>9</sup> Global Biodiversity Information Facility (GBIF), Copenhagen Ø DK-2100, Denmark

<sup>10</sup> Natural History Museum, London SW7 5BD, UK

<sup>11</sup> Natural History Museum of Denmark, Copenhagen K DK-1307, Denmark

<sup>12</sup> Smithsonian Tropical Research Institute, Balboa Ancon, Unit 0948, Panama

[ucsc.edu/](http://ucsc.edu/)) and the International Barcode of Life Project (<http://ibol.org/>), are finding access to adequate samples needed for genome sequencing to be a major bottleneck in their workflow. A major goal of the Global Genome Biodiversity Network (GGBN) is to effectively bridge this gap as well as to encourage and enable scientists to link vouchers, tissues, physical DNA, sequences and publications.

Natural history collections play an important role in catering to the need for access to high-quality well-documented samples for genome sequencing. They have archived biodiversity specimens over centuries, often including taxa or populations that have now vanished from the wild. These collections allow researchers to access biological specimens and their data efficiently and often make further expensive field trips and subsequent time-consuming species identifications unnecessary. Biodiversity biobanks form an emerging collection type that complements classical natural history collections. While often conveniently associated with these traditional collections, biodiversity biobanks are much better suited for modern downstream molecular applications (e.g. second- and third-generation sequencing requiring high-quality DNA), as their focus lies explicitly on conserving the molecular structure of their samples. Biodiversity biobanks adhere to standard operating procedures within a sustainable collection management infrastructure to guarantee integrity, identity and ethical use of samples. They commit themselves to making samples lastingly available and accessible. By using internationally accepted standards for biodiversity information, federation and synthesis of data among collection databases is possible.

Until now, no central platform existed that would globally aggregate biodiversity biobank data. While the need for data sharing

and samples access increases, information on genomic samples is still as fragmented as the geographic distribution of the repositories that maintain them. Without a central data portal that aggregates genomic sample data, biodiversity genomics studies cannot reach their full potential. The German Research Foundation (DFG)-supported DNA Bank Network was established in 2007 as a first step towards this goal (Gemeinholzer et al., 2011). The GGBN aims to close this gap on a global scale by building on the DNA Bank Network's data model. GGBN was formed in 2011 with the principal aim of making high-quality well-documented and vouchered collections that store DNA or tissue samples of biodiversity, discoverable for research through a networked community of biodiversity repositories.

#### THE NETWORK

GGBN is an open network of currently 24 biodiversity repositories from across all continents that came together with the aim to endorse biodiversity biobank stewardship, information sharing and ethical use of collections in compliance with national and international conventions and regulations (Convention on Biological Diversity, <http://www.cbd.int/convention/text/> and Nagoya Protocol on access and benefit-sharing (ABS), <http://www.cbd.int/abs/doc/protocol/nagoya-protocol-en.pdf>).

Key objectives of this global endeavor include establishing and adopting standards and best practices, accelerating access to specimens and data and enabling targeted collection. GGBN's most central mission is to promote access to information on the genomic samples maintained by its members. Goals include to:

- Provide genome-quality DNA and tissue samples from across the tree of

life for research and training;

- Provide free access to a global Data Portal hosting the aggregated primary specimen and sample data as well as metadata of all member institutions;
- Develop standards for sharing DNA and tissue information;
- Develop best practices related to management and stewardship of genomic samples and their derivatives, including appropriate access and benefit-sharing;
- Establish an infrastructure for deposition of genomic samples;
- Promote targeted collection and preservation of genomic samples representing a synoptic sample of life on earth;
- Provide a platform for biodiversity biobanking knowledge exchange;
- Recruit partners with different regional and taxonomic foci, to preserve global genetic diversity in an efficient, planned and concerted effort.

Global biodiversity and collections initiatives around the world will both contribute to and rely on GGBN resources for reaching their research goals.

#### GGBN DATA PORTAL

GGBN's Data Portal (<http://data.ggbn.org>) bridges the gap between biodiversity repositories, sequence databases and research results by linking globally distributed biodiversity databases of genomic samples, representing the biodiversity of life on Earth to voucher specimens, sequence data and publications. At present, more than 70,000 DNA and tissue samples from almost 17,000 taxa are available. For detailed information on GGBN's data portal, see *The Global Genome Biodiversity Network (GGBN) Data Portal*.

#### MEMBERSHIP

Membership in GGBN is open to natural history museums, herbaria, botanical gardens, culture collections, seedbanks, gene banks, biobanks, frozen tissue collections, other types of biological repositories, and representatives of government agencies, academic, and other organizations and individuals involved in genomic biodiversity. GGBN has three membership categories for institutions and individuals:

- Core members are institutional biorepositories that fulfill all the requirements of Associate Membership and are also actively contributing data to the GGBN Data Portal in compliance with the standards established by GGBN.
- Associate members are institutional biorepositories that have initiated the process of becoming core members and therefore have an enterprise level data system that can provide data on a reliable basis, are represented in an institutional registry of biorepositories such as The Global Registry of Biorepositories (GRBio, <http://grbio.org/>), intend to prepare and contribute data to the GGBN Data Portal in compliance with the standards established by GGBN, and have made a commitment to long-term preservation of genomic collections and associated data, for the purpose of providing access for research.
- Observer Institutions and Individual Members have an interest in genome biorepositories and biodiversity informatics, collect and contribute material to Core and Associate members in compliance with the relevant national/international laws

and regulations, support the goals of GGBN, provide technical expertise to GGBN, and participate in GGBN activities.

For more information on GGBN membership and standards, see the *GGBN Memorandum Of Cooperation* ([http://ggbn.org/docs/GGBNMOC\\_2013\\_web.pdf](http://ggbn.org/docs/GGBNMOC_2013_web.pdf)) or contact the secretariat ([ggbn@si.edu](mailto:ggbn@si.edu)).

## CONCLUSIONS

Discoverability and coordinated expansion of biodiversity biobanks worldwide only becomes possible through a networked community of biodiversity biorepositories that provide trusted and transparent access to genomic samples. GGBN opens up the repositories of its member institutions to the scientific community via a shared central Data Portal that exposes collection information through a federated information infrastructure. GGBN furthermore provides a collaboration platform that helps in developing best practices for biodiversity repositories, ensures quality standards and harmonizes the exchange of samples in accordance with national and international legislation and conventions.

In times of rapid biodiversity loss and

increased awareness of ABS requirements, as exemplified by the Nagoya Protocol, the GGBN integrates widely distributed resources; it multiplies sample visibility of biodiversity repositories worldwide and unites the people who make them functional. Sharing their expertise, these people work towards the common goals of enabling more efficient and more sustainable research on biodiversity. The synergistic effect of this collaboration increases with each individual member, and GGBN welcomes and needs new members with a long-term commitment to responsible and ethical use of material deposited in biodiversity biobanks.

## LITERATURE CITED

- Brown, W.Y. 2013. DNA Net Earth. Global Views Policy Paper 2013-01. Brookings Institution, Washington, D.C.
- Gemeinholzer, B., G. Dröge, H. Zetzsche, G. Haszprunar, H.-P. Klenk, A. Güntsch, W. G. Berendsohn & J.-W. Wägele. 2011. The DNA Bank Network: the start from a German initiative. *Bio-preserv. Biobank*. 9: 51–55.
- Pawlowski, J., S. Audic, S. Adl, D. Bass, L. Belbahri, C. Berney, et al. 2012. CBOL protist working group: barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. *PLoS Biol.* 10: e1001419. doi: 10/1371/journal.pbio.1001419