Global Genome Biodiversity Network

Jonathan Coddington
GGBN Executive Committee Member
Smithsonian Institution

Members of the Interim Executive Committee
Global Genome Initiative

Preserving and Understanding the Genomic Diversity of Life

Just One Genome!
Tree of Life

- DOMAINS/KINGDOMS: 3
- PHYLA / DIVISIONS: 94
- CLASSES: 364
- ORDERS: 1413
- FAMILIES: ~ 9654
- “GENERA”: ~ 200,000
- SPECIES: >15,000,000?
## GGI Goals

<table>
<thead>
<tr>
<th>Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hard-to-find tissues</td>
<td>Public</td>
</tr>
<tr>
<td>Ambiguous quality</td>
<td>Genome-quality</td>
</tr>
<tr>
<td>Ambiguous ownership</td>
<td>Enterprise biorepositories</td>
</tr>
<tr>
<td>Individual scientists</td>
<td>Benefit-sharing &amp; access</td>
</tr>
<tr>
<td>Expensive</td>
<td>Affordable</td>
</tr>
<tr>
<td>“Boutique”</td>
<td>Coordinated, strategic</td>
</tr>
<tr>
<td>Mostly model organisms</td>
<td>All branches of Life</td>
</tr>
<tr>
<td>Classical taxonomy</td>
<td>Gene-based identification</td>
</tr>
<tr>
<td>Experts only</td>
<td>Citizen technology</td>
</tr>
<tr>
<td>Limits environmental monitoring, conservation, biotech, basic research</td>
<td>Most organisms anywhere</td>
</tr>
<tr>
<td></td>
<td>Accurate, scalable, cheap tech to address important challenges</td>
</tr>
</tbody>
</table>
GGBN Goals

- Standard for sharing tissue and DNA information
- Institutional directory (GRBio)
- Platform for aggregating member data and portal
- Best practices for management and stewardship of genomic samples
- Partners with different regional and taxonomic focus
- Identify global gaps in GGBN collections
Smithsonian Institution Forest Earth Observatory

54 plots, 10,500 species, 4,346 genera ("trees")
~60% world total?
Chesapeake Bay

Sequences: 572,290
Species: 1,204
Unknown: 40.9

Fort Pierce FL

Sequences: 409,613
Species: 1,391
Unknown: 28.3

...91% correct for families; 85% for genera
Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits
(Convention on Biological Diversity)

- Entered into force Oct. 2014
- Greater legal certainty and transparency
- Promote and encourage research
- Awareness-raising & technology transfer
- In-country research capability and institutions
GGBN Value Proposition One

- Make genomic collections *discoverable* for research through a networked community of biorepositories
  - Provide biorepositories with standardized methods for making genomic collections discoverable;
  - Provide biorepositories and contributors with community standards and best practices for the collection and data management of genomic samples;
  - Provide contributors with access to a community that provides storage facilities for and access to information on their genomic collections.
GGBN Value Proposition Two

• Provide trusted and transparent access to genomic samples for users and contributors through an ABS framework—supporting the trust from biodiversity-rich countries and organizations.

• Guide global biorepository growth strategically to preserve Life’s genome
Associate Membership

Institutional Biorepositories plan to become core members and:

- Commit to preservation of genomic collections and associated metadata for research;
- Registered as a biorepository (GRBio);
- Enterprise level, reliable data system;
- Will contribute data to GGBN complying with GGBN standards

Core Membership

- An Associate Member that is contributing data to GGBN.
GGBN Data Portal Milestones

Oct 2011  1st Intl. workshop, DC
July 2012  2nd Intl. workshop, Copenhagen
2013   BGI, China National Gene Bank, Ocean Sampling Day
Spring 2013  NSF Workshop (USA,)
August 2013: GGBN & DNA Bank Network sites merged
Fall 2013  DFG Grant (3 FTE’s 4 years)
November 2014:  Private beta release of GGBN Data Portal
June 2015:  1st public release of GGBN Data Portal
November 2015:  Submit GGBN Data Standard to TDWG review and ratification process
November 2015:  2nd public release of GGBN Data Portal
GGBN Data Standard

July 2014: Released for Public Review, implemented in IPT/DwC-A and BioCASE/ABCDDNA

November 2015: Submit GGBN Data Standard to TDWG review and ratification process
32 Biorepository Members & Collaborators (17 countries)
Taxa and Samples Per Year

- Samples
- Taxa

Year: 2009-2014

Number of Samples and Taxa over the years.
The Global Genome Biodiversity Network (GGBN) Data Portal

Gabrielle Droegemueller, Katharine Barker, Jonas J. Astrin, Paul Bartels, Carol Butler, David Cantrill, Jonathan Coddington, Félix Forest, Birgit Gemeinholzer, Donald Hobern, Jacqueline Mackenzie-Dodd, Éamonn Ó Tuama, Gitte Petersen, Oris Sanjurjo, David Schindel, and Ole Seberg

The Global Genome Biodiversity Network (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. This is achieved through the GGBN Data Portal (http://data.ggbn.org), which links globally distributed databases and bridges the gap between biodiversity repositories, sequence databases, and research results. Advances in DNA extraction techniques combined with next-generation sequencing technologies provide new tools for genome sequencing. Many ambitious genome sequencing projects with the potential to revolutionize biodiversity research consider access to adequate samples to be a major bottleneck in their workflow. This is linked not only to accelerating biodiversity loss and demands to improve conservation efforts but also to a lack of standardized methods for providing access to genomic samples. Biodiversity biobank-holding institutions urgently need to set a standard of collaboration towards excellence in collections stewardship, information access and sharing and responsible and ethical use of such collections. GGBN meets these needs by enabling and supporting accessibility and the efficient coordinated expansion of biodiversity biobanks worldwide.

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 documented samples that yield high-molecular weight DNA and RNA from throughout the tree of life (e.g. [4,7]). Many ambitious projects with the potential to revolutionize biodiversity research are finding access to adequate samples needed for genome sequencing to be a major bottleneck in their workflow. Examples of these projects include the Ten Thousand Vertebrate Genomes Project (Genome10K).
GGBN collaborates with other stakeholders and across domains to reach this goal.
Aggregate multiple sources

Explore *Moenkhausia*

- 9284 specimens
- 219 nucleotide sequences
- 63 barcode sequences
- Taxon page
- 22 DNA samples
- 47 tissue samples

Getting live counts from other biodiversity portals for each record
2nd GGBN International Conference

Berlin
June
2013
w
SPNHC
Next Steps

• Begin concrete USA collaborations
  – Specify
  – Arctos
  – Recruit new Associate/Core Members

• IDigBio Working Group?

• Other suggestions?