GGBN – Strategies for standardized exchange of genetic resources on a global scale

Jonathan Coddington
Smithsonian Institution
Global Genome Biodiversity Network

• **Network for non-human biobanks** (e.g. DNA, tissue), founded in 2011
• Based on DNA Bank Network, 2007
• General secretariat: Smithsonian Institution
• Technical secretariat: Botanic Garden and Botanical Museum Berlin

• [http://www.ggbn.org](http://www.ggbn.org)
High Molecular Weight DNA
The Problem

1. Adapted from (1)

Existing Solutions

Established Platforms:
GBIF, NCBI, EMBL, DDBJ, EOL, Index Herbariorum, GRBio, ...

Community Standards:
Darwin Core, ABCD, MlxS, EML, SPREC, BRISQ, ...

Open Source Collection Management Software:
Specify, BRAHMS, Diversity Workbench, JACQ, ...

Major Gaps

3.

- No data standard to cover all types of genomic samples.
- Not all existing standards already have technical implementations as xml, json or something similar.
- No open source software for managing DNA and tissue collections.
- No portal for genomic sample data.
- Not enough communication between different communities.

What GGBN offers

4.

- The GGBN Data Standard to complement other community standards.
- Two technical implementations of the GGBN Data Standard for BioCASE and IPT.
- An open source software for managing DNA and tissue collections.
- The GGBN Data Portal to enable standardized access to genomic sample and accompanying data.
- The GGBN conferences to bring together different communities working with biodiversity biobanks

Figure 1. Bridging the gaps. Schematic representation of (1) Low percentage of available sequence data in public repositories with proper information where the voucher and/or sample is deposited. (2) Existing tools and platforms for standardized management and access to biodiversity data. (3) Major gaps identified by GGBN and (4) what GGBN has developed to fill these gaps.
This page describes the GGBN Data Standard version 1 (v1), released on 04. October 2016.

**Citation:** Droge et al. (2016): The Global Genome Biodiversity Network (GGBN) Standard Specification. DATABASE. doi: 10.1093/database/baw125

**Note:** Ongoing discussions and updates of the GGBN Data Standard will be documented at [http://terms.tdwg.org/GGBN_Data_Standard](http://terms.tdwg.org/GGBN_Data_Standard)

The GGBN Data Standard is intended to be used with ABCD or Darwin Core and is not a stand-alone solution! Implementations are available for both ABCD and DarwinCore-Archive. For ABCD two implementations have been made. One for single species samples ABCDGGBN and one for environmental samples ABCDGGBN-Enviro. Please check out the TDWG wiki for updates.

A new stable version of the standard will be released earliest at the end of 2019.

The GGBN Data Standard is a set of terms and controlled vocabularies designed to represent sample facts. It does not cover e.g. scientific name, geography, or physiological facts. This allows combining the GGBN Data Standard with other complementary standards, such as DwC, ABCD or MxS. Potentially, the standard can be used not only for non-human genomic samples but also human samples. It builds upon existing standards commonly used within the communities extending them with the capability to exchange data on tissue, environmental and DNA sample as well as sequences. The GGBN Data Standard incorporates all molecular terms of MxS and...
50 members, 22 countries: 518,000 samples, 2200 families, 11,000 genera, 32,000 species
GGBN Growth: 17 Biobanks to date
# GGBN Distribution of Families (2329)

<table>
<thead>
<tr>
<th>N Families</th>
<th>N_BioBanks</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1704</td>
<td>N_BioBanks=1</td>
<td>0.73</td>
</tr>
<tr>
<td>442</td>
<td>N_BioBanks=2</td>
<td>0.19</td>
</tr>
<tr>
<td>121</td>
<td>N_BioBanks=3</td>
<td>0.05</td>
</tr>
<tr>
<td>38</td>
<td>N_BioBanks=4</td>
<td>0.02</td>
</tr>
<tr>
<td>20</td>
<td>N_BioBanks=5</td>
<td>0.01</td>
</tr>
<tr>
<td>4</td>
<td>N_BioBanks=6</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Thanks!