Global Genome Biodiversity Network

Jonathan Coddington
GGBN Executive Committee Member
Smithsonian Institution

Members of the Interim Executive Committee

Global Genome Initiative

GLOBAL GENOME INITIATIVE

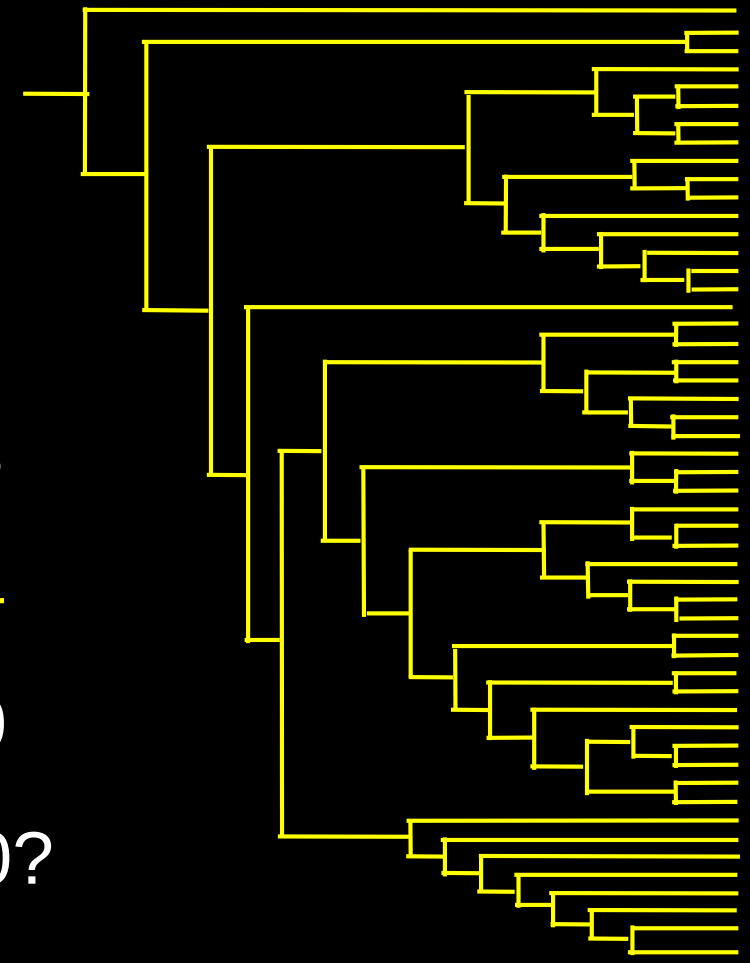
GGBN

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

GLOBAL GENOME INITIATIVE

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

GGBN Goals

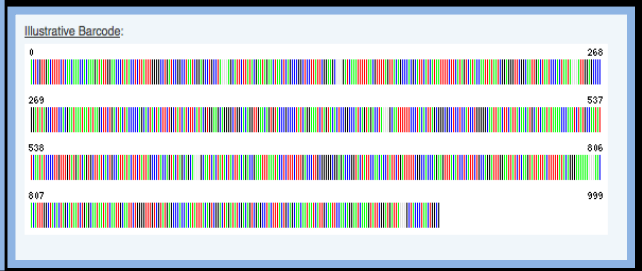
- Standard for sharing tissue and DNA information
- Institutional directory (GRBio)
- Platform for aggregating member data and portal
- Best practices for management and stewardship of genomic samples
- Partners with different regional and taxonomic focus
- Identify global gaps in GGBN collections

Smithsonian Institution Forest Earth Observatory

54 plots, 10,500 species, 4,346 genera (“trees”)
~60% world total?

GGBN

GLOBAL GENOME INITIATIVE



	Chesapeake Bay	Fort Pierce FL
Sequences	572,290	409,613
Species	1,204	1,391
Unknown	40.9	28.3

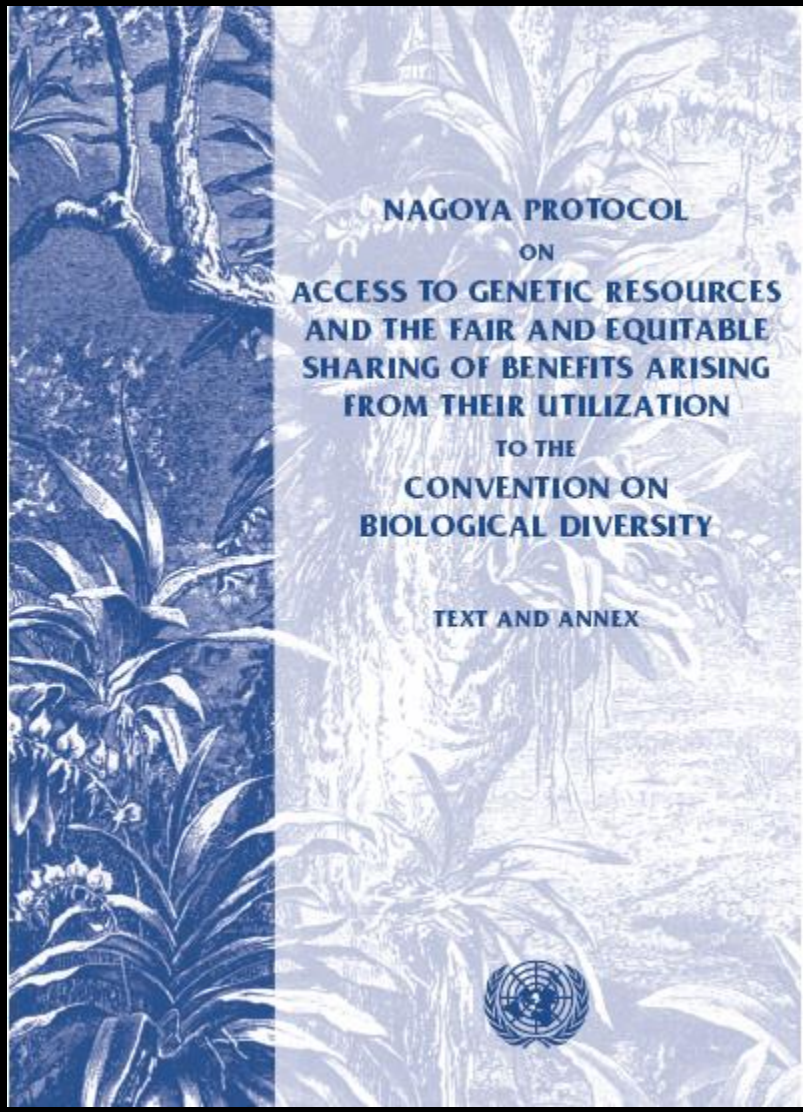
…91% correct for families; 85% for genera

GLOBAL GENOME INITIATIVE

Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits

(Convention on Biological Diversity)

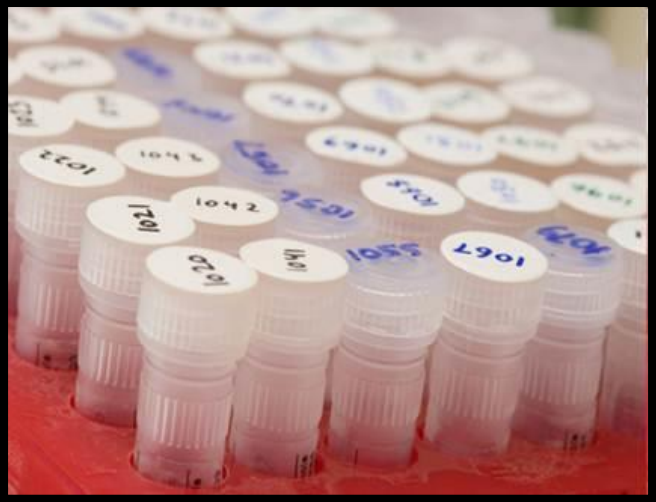
- Entered into force Oct. 2014
- Greater legal certainty and transparency
- Promote and encourage research
- Awareness-raising & technology transfer
- In-country research capability and institutions



GGBN

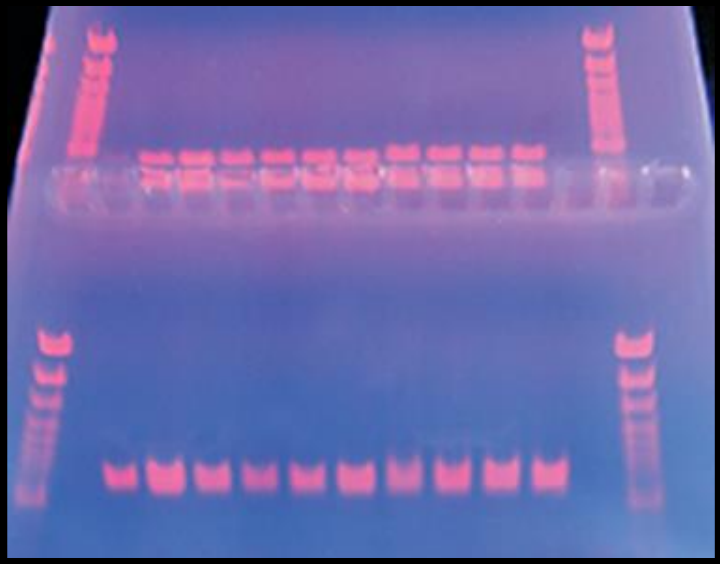
GGBN Value Proposition One

- **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.

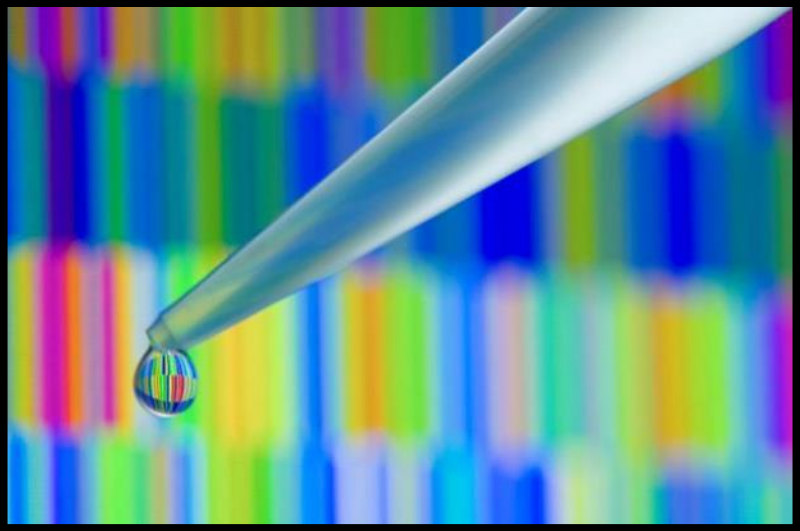


GGBN Value Proposition Two

- 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)

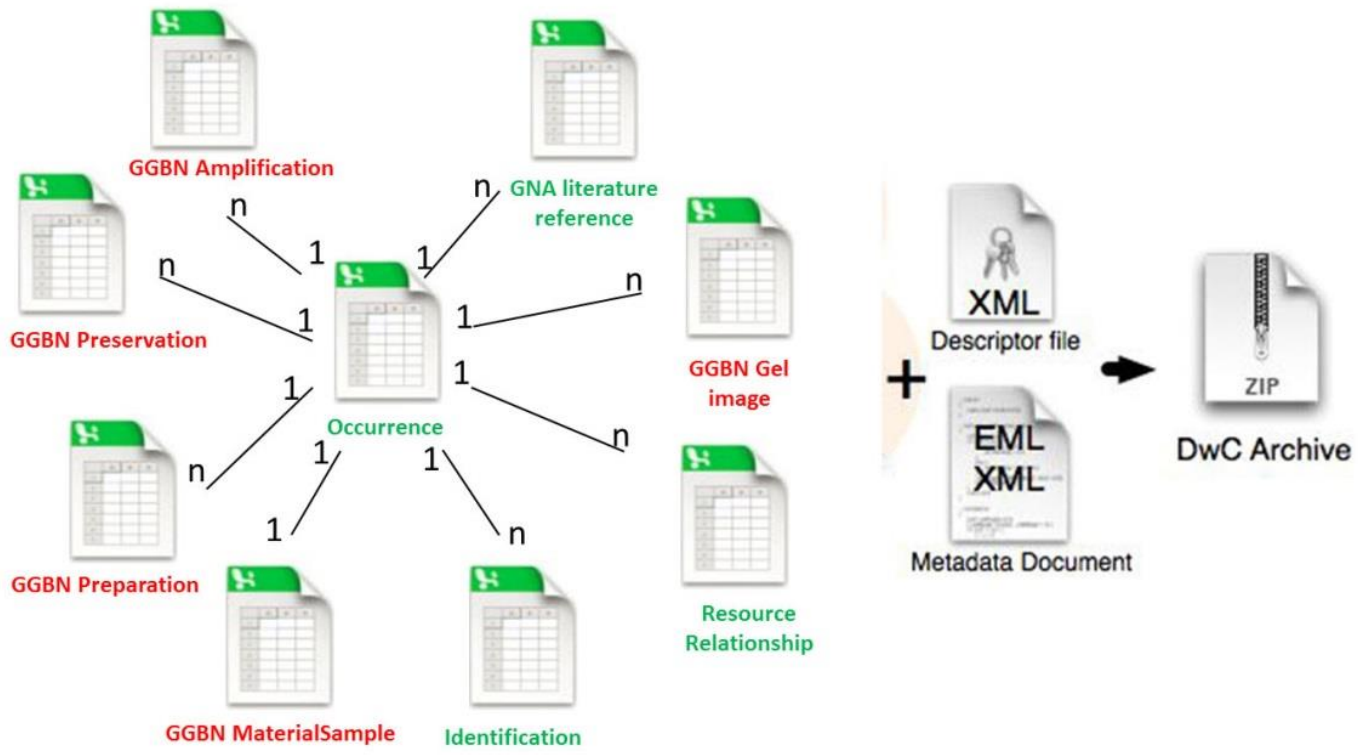
November 2014: Private beta release of GGBN Data Portal

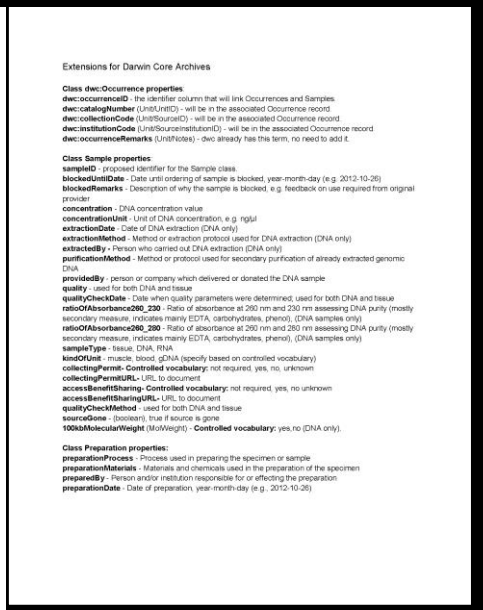
June 2015: 1st public release of GGBN Data Portal

November 2015: Submit GGBN Data Standard to TDWG review and ratification process

November 2015: 2nd public release of GGBN Data Portal

GGBN Data Standard





Extensions for Darwin Core Archives

Class dwc:Occurrence properties:
dwc:occurrenceID - the identifier column that will link Occurrences and Samples.
dwc:catalogNumber (UnitIdUnitID) - will be in the associated Occurrence record.
dwc:collectionCode (UnitISourceID) - will be in the associated Occurrence record.
dwc:institutionCode (UnitISourceInstitutionID) - will be in the associated Occurrence record.
dwc:occurrenceRemarks (UnitNote) - dwc already has this term, no need to add it.

Class Sample properties:
sampleID - proposed identifier for the Sample class.
blockedUntilDate - Date until ordering of sample is blocked, year-month-day (e.g. 2012-10-26)
blockedRemarks - Description of why the sample is blocked, e.g. feedback on use required from original provider
concentration - DNA concentration value
concentrationUnit - Unit of DNA concentration, e.g. ng/µl
extractionDate - Date of DNA extraction (DNA only)
extractionMethod - Method or extraction protocol used for DNA extraction (DNA only)
extractedBy - Person who carried out DNA extraction (DNA only)
purificationMethod - Method or protocol used for secondary purification of already extracted genomic DNA
providedBy - person or company which delivered or donated the DNA sample
quality - used for both DNA and tissue
qualityCheckDate - Date when quality parameters were determined; used for both DNA and tissue
ratioOfAbsorbance260_230 - Ratio of absorbance at 260 nm and 230 nm assessing DNA purity (mostly secondary measure, indicates mainly EDTA, carbohydrates, phenol), (DNA samples only)
ratioOfAbsorbance260_280 - Ratio of absorbance at 260 nm and 280 nm assessing DNA purity (mostly secondary measure, indicates mainly EDTA, carbohydrates, phenol), (DNA samples only)
sampleType - tissue, DNA, RNA
kindOfUnit - muscle, blood, gDNA (specify based on controlled vocabulary)
collectingPermit- Controlled vocabulary: not required, yes, no, unknown
collectingPermitURL- URL to document
accessBenefitSharing- Controlled vocabulary: not required, yes, no unknown
accessBenefitSharingURL- URL to document
qualityCheckMethod - used for both DNA and tissue
sourceGone - (boolean), true if source is gone
100kbMolecularWeight (MolWeight) - **Controlled vocabulary:** yes,no (DNA only).

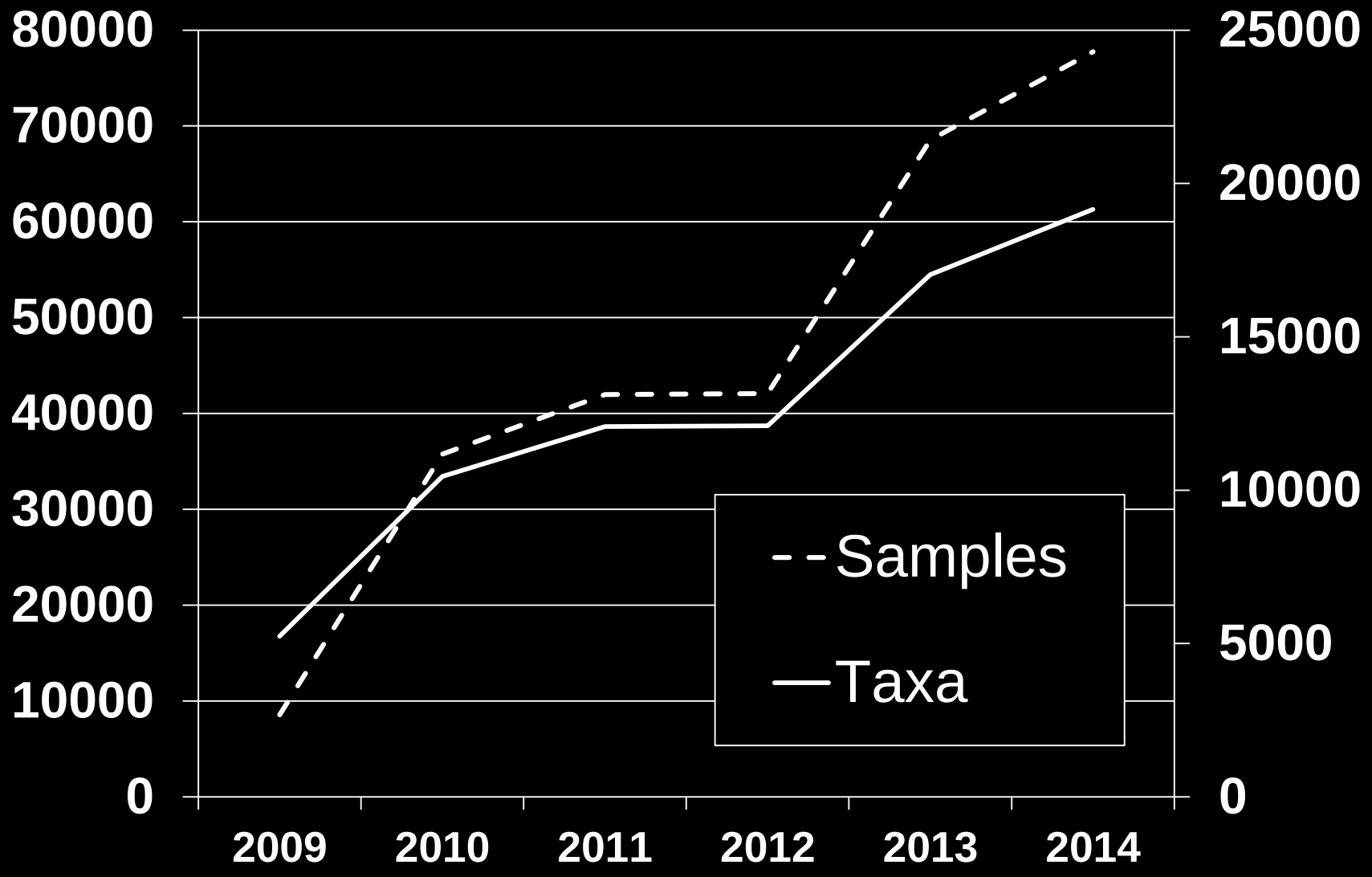
Class Preparation properties:
preparationProcess - Process used in preparing the specimen or sample
preparationMaterials - Materials and chemicals used in the preparation of the specimen
preparedBy - Person and/or institution responsible for or effecting the preparation
preparationDate - Date of preparation, year-month-day (e.g., 2012-10-26)

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



White Paper on Data Portal

Nucleic Acids Research Advance Access published November 19, 2013

Nucleic Acids Research, 2013, 1–6
doi:10.1093/nar/gkt928

The Global Genome Biodiversity Network (GGBN) Data Portal

Gabriele Droege[1,*], Katharine Barker[2], Jonas J. Astrin[3], Paul Bartels[4], Carol Butler[2], David Cantrill[5], Jonathan Coddington[2], Félix Forest[6], Birgit Gemeinholzer[7], Donald Hobern[8], Jacqueline Mackenzie-Dodds[9], Éamonn Ó Tuama[8], Gitte Petersen[10], Oris Sanjur[11], David Schindel[2] and Ole Seberg[10]

[1]Botanic Garden and Botanical Museum Berlin-Dahlem, Freie Universität Berlin, Berlin 14195, Germany, [2]National Museum of Natural History Smithsonian Institution, Washington DC 20013, USA, [3]Zoological Research Museum Alexander Koenig, Bonn 53113, Germany, [4]Wildlife & Environment Society of South Africa, Pretoria 0001, South Africa, [5]National Herbarium of Victoria, Royal Botanic Gardens Melbourne, South Yarra, VIC 3141, Australia, [6]Molecular Systematics Section, Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3DS, UK, [7]Systematic Botany, Justus-Liebig-Universität, Giessen 35392, Germany, [8]Global Biodiversity Information Facility (GBIF), Copenhagen Ø DK-2100, Denmark, [9]Department of Zoology, The Natural History Museum, London SW7 5BD, UK, [10]Natural History Museum of Denmark, Copenhagen K DK-1307, Denmark and [11]Smithsonian Tropical Research Institute, Balboa Ancon, Unit 0948, Panama

Received August 15, 2013; Revised September 16, 2013; Accepted September 20, 2013

ABSTRACT

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.

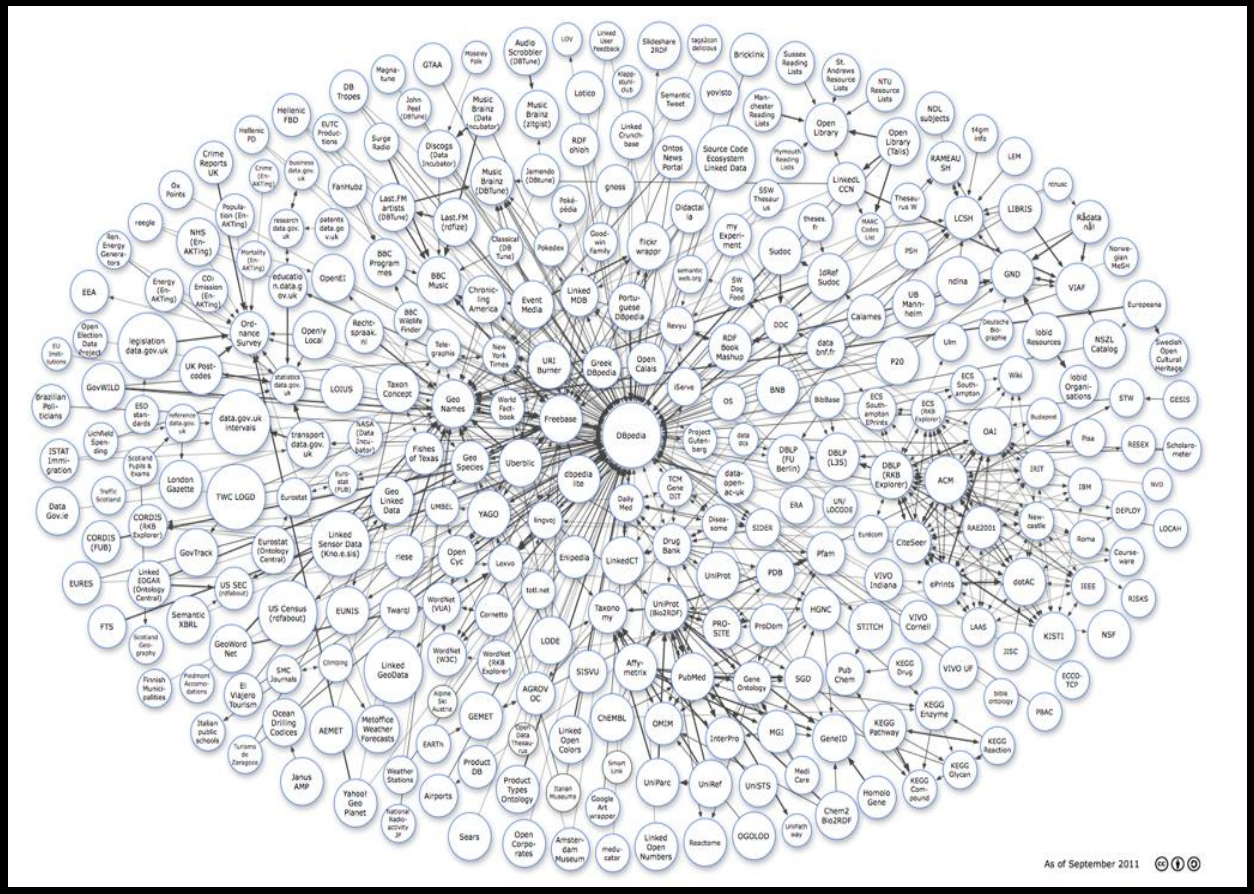
INTRODUCTION

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. (1,2)]. 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 (Genome10 K),

*To whom correspondence should be addressed. Tel: +49 30 838 50 139; Fax: +49 30 838 50 186; Email: g.droege@bgbm.org
Present address:
Gabriele Droege, Botanic Garden and Botanical Museum Berlin-Dahlem, Freie Universität Berlin, Berlin, 14195, Germany.

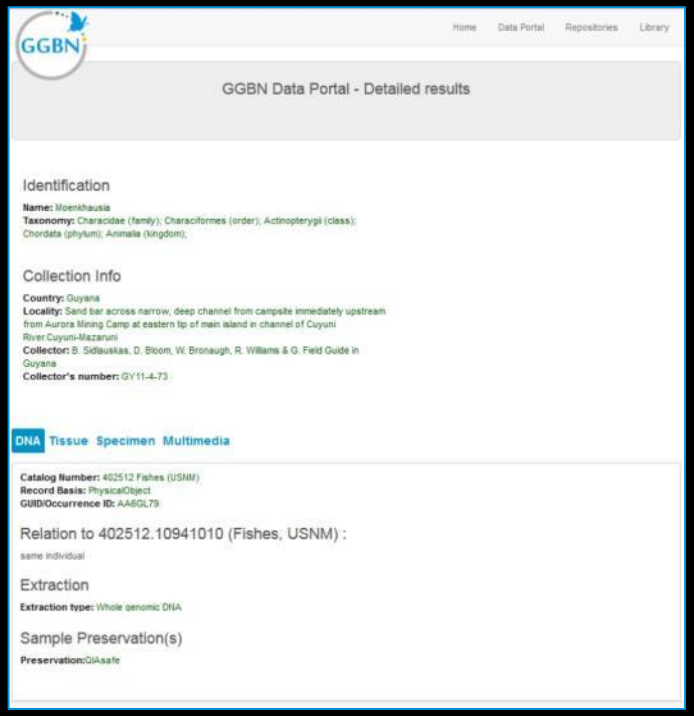
The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

Linked Data and Stable Identifiers



GGBN collaborates with other stakeholders and across domains to reach this goal.

Aggregate multiple sources



Getting live counts from other biodiversity portals for each record

Explore *Moenkhausia*

GBIF — 9284 specimens

NCBI — 219 nucleotide sequences

BOLD SYSTEMS — 63 barcode sequences

eol Encyclopedia of Life — taxon page

GGBN — 22 DNA samples

47 tissue samples

2nd GGBN International Conference

Berlin
June
2013
w
SPNHC

Thanks!


GLOBAL
GENOME
INITIATIVE


GGBN


Australian
museum


ASSOCIATION OF AMERICAN UNIVERSITIES
AAU
1900


NATURAL HISTORY MUSEUM OF DENMARK
UNIVERSITY OF COPENHAGEN


AMERICAN MUSEUM & NATURAL HISTORY


Kew
ROYAL BOTANIC GARDENS


RB


NATURAL
HISTORY
MUSEUM


OCEAN
GENOME
LEGACY


iDigBio
Integrated Digitized Biocollections


DSMZ


museum für
naturkunde
berlin


AFRD


国家基因库
CHINA NATIONAL
GENEBANK


SENCKENBERG
world of biodiversity

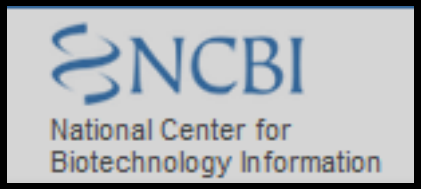

HUMBOLDT


UNICAMP


Z·S·M


HUNGARIAN
NATURAL
HISTORY
MUSEUM


Smithsonian
National Museum of Natural History


NCBI
National Center for
Biotechnology Information


NSF


OPEN
Tree of Life


NRF
National Research Foundation
NZG
National Zoological Gardens of South Africa


MUSEUMVICTORIA


ITIS
Integrated Taxonomic
Information System


FORSCHUNGS
museum
KOENIG


South
Australian
Museum


Convention on
Biological Diversity


BGBM


THE NEW YORK BOTANICAL GARDEN


Barcode of Life
Identifying Species with DNA Barcoding


eol
Encyclopedia of Life


Smithsonian
National Zoological Park


CONSORTIUM FOR THE
BARCODE OF LIFE


BIOCODE
moorea

Smithsonian Tropical Research Institute
Center for Tropical Forest Science

Next Steps

- Begin concrete USA collaborations
  - Specify
  - Arctos
  - Recruit new Associate/Core Members
- IDigBio Working Group?
- Other suggestions?