

The GGBN Data Standard and its Darwin Core Archive implementation

Droege, G., Barker, K., Flemons, P., Orrell, T. & Whitacre, J.



*Global Genome
Biodiversity
Network (GGBN)*

SYNTHESYS 
Synthesis of systematic resources



Botanic Garden and Botanical Museum Berlin-Dahlem
Freie Universität Berlin



Global Genome Biodiversity Network

- Making genomic collections accessible via a shared portal
- Develop standards for sharing tissue and DNA information
- Develop best practices related to management and stewardship of genomic samples
- Knowledge platform for biodiversity biobanking

Global Genome Biodiversity Network

- Technical secretariat at BGBM
- General secretariat at NMNH Smithsonian
- Started with 4 German partners with its precursor DNA Bank Network 7 years ago
- Current Funding:



GGBN today

- Major impact by defining policies and practices for sharing genomic data and samples
 - GGBN Nagoya protocol recommendations to be published soon
 - GGBN Data Standard draft version published
 - Many Institutions currently updating/changing their databases, workflows and policies following GGBN recommendations

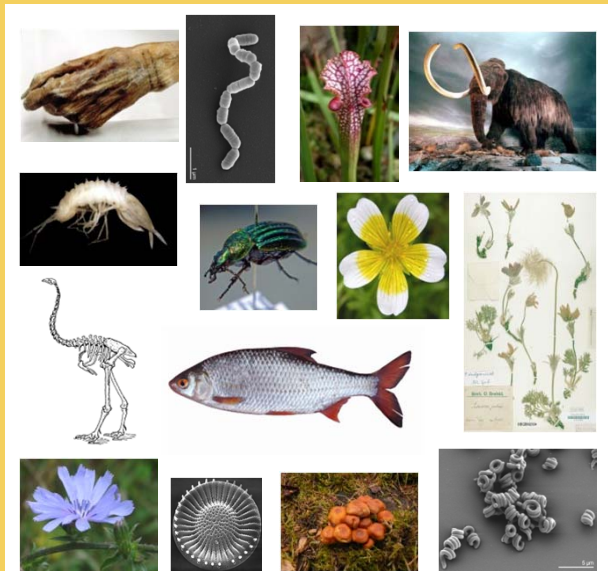
GGBN Data Portal

- <http://data.ggbn.org>
 - Based on GBIF infrastructure (ABCD and DwC-A)
 - Specimen Data pipelines (BioCAsE and IPT) used for both GBIF and GGBN
 - Additional mappings for tissues and DNAs
 - Scientists can order tissue and DNA samples
 - Current portal ABCD only
- > DFG project -> enable DwC-A -> **new portal release 11/15**

Basic Architecture

Global Genome
Biodiversity
Network (GGBN)

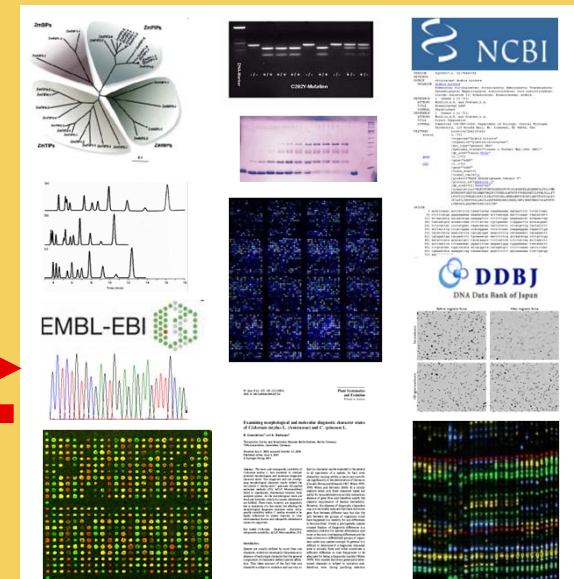
Source Material / Specimens



DNA & Tissue



Molecular Analysis Data



Global Genome
Biodiversity
Network (GGBN)



EMBL



BOLD
SYSTEMS

http://data.ggbn.org – aggregating data

DNA Details

Preferred Taxon Name: **Lecythis confertiflora**
Unit ID/DNA Bank No: 1287461
Institution Code: NYBG
Collection Code: DNA Bank
Record Basis: DNA Sample
Kind of Unit: Nuclear, plastid, mitochondrial

Relation to voucher:

DNA from specimen

Tissue: Leaf

Preservation: Silica

Extraction:

Extraction Method: Qiagen DNeasy
Extraction Staff: Y.-Y. Huang

Quality:

Notes:

DNA provided by: Y.-Y. Huang

Molecular details

Amplification(s):

Genetic Locus/Fragment: **ndhF**

Sequencing(s):

GenBank Number(s):

1. DQ388210 | [Link](#)

Sequencing References:

Mori, S.A. et al. (2007): Evolution of Lecythidaceae combined ndhF and trnL-F sequence data. *Am. J. Bot.* 94 (3): 289-301 | [Link](#)
Genetic Locus/Fragment: trnL, trnL-trnF, trnF

Sequencing(s):

GenBank Number(s):

1. DQ418014 | [Link](#)

Sequencing References:

Mori, S.A. et al. (2007): Evolution of Lecythidaceae with an emphasis on the circumscription of neotropical genera: information from combined ndhF and trnL-F sequence data. *Am. J. Bot.* 94 (3): 289-301 | [Link](#)

DNA/Molecular dataset summary

Specimen Details (voucher)

Unit ID: 612293
Institution Code: NY
Collection Code: Herbarium
Record Basis: Specimen

Identification(s):

Name: **Lecythis confertiflora**
Taxonomy: **Lecythidaceae (familia) ;**

Collection Info:

Collector(s): S. A. Mori
Field number: 25411
Collection Date Begin: 19.02.2002

Locality

Country: French Guiana (GF)
Region: Middle and South America
Locality: , Nouragues Field Station, on trail to Crique Kwak, ca. 200 m NE of Camp Inselberg
Coordinates (Lat|Lon): 4.2 | 52.41

Other info



Specimen dataset summary

GGBN Data Standard

- http://terms.tdwg.org/wiki/GGBN_Data_Standard
- Based on ABCDDNA
- Is meant to be used with ABCD or DwC -> all occurrence terms are excluded (geography, scientificname etc.)

GGBN Data Standard

The [Global Genome Biodiversity Network \(GGBN\)](#) is a global network of well-managed collections of genomic tissue samples from across the Tree of Life, benefiting society through biodiversity research, development and conservation. This network will foster collaborations among repositories of molecular biodiversity in order to ensure quality standards, improve best practices, secure interoperability, and harmonize exchange of material in accordance with national and international legislation and conventions.

The GGBN Data Standard is a set of vocabularies designed to represent tissue, DNA or RNA samples associated to voucher specimens, tissue samples and collections. *Contributors: Gabriele Droege, Birgit Gemeinholzer, Holger Zetzsche, Astrid Schories, Jörg Holetschek, Enrique Arbeláez Cortés, Katharine Barker, Sean Brady, Boyke Bunk, Margaret Casey, Jonathan Coddington, John Deck, René Dekker, Sonya Dyhrman, Elisabeth Haring, Hans-Peter Klenk, Patricia Kelbert, Thomas Kneibelsberger, Renzo Kortmann, Christopher Lewis, Jacqueline Mackenzie-Dodds, Christopher Meyer, Jon Norenburg, Michael Raupach, Thomas von Rintelen, Ole Seberg, Larissa Smirnova, Carola Söhngen, Sun Ying, Lee A. Weigt, Kenneth Wurdack, Pelin Yilmaz, Elizabeth Zimmer, Thomas Orrell, Jamie Whitacre, Xin Zhou.*

Contents [\[hide\]](#)

[1 Introduction](#)

[2 GGBN Data Standard Terms](#)

[3 Dynamic Term List](#)

[3.1 Index to concepts \(all collections combined\)](#)

[3.2 GGBN Amplification Vocabulary](#)

[3.3 GGBN DNA Cloning Vocabulary](#)

[3.4 GGBN Gel Image Vocabulary](#)

[3.5 GGBN Loan Vocabulary](#)

[3.6 GGBN Material Sample Vocabulary](#)

[3.7 GGBN Permit Vocabulary](#)

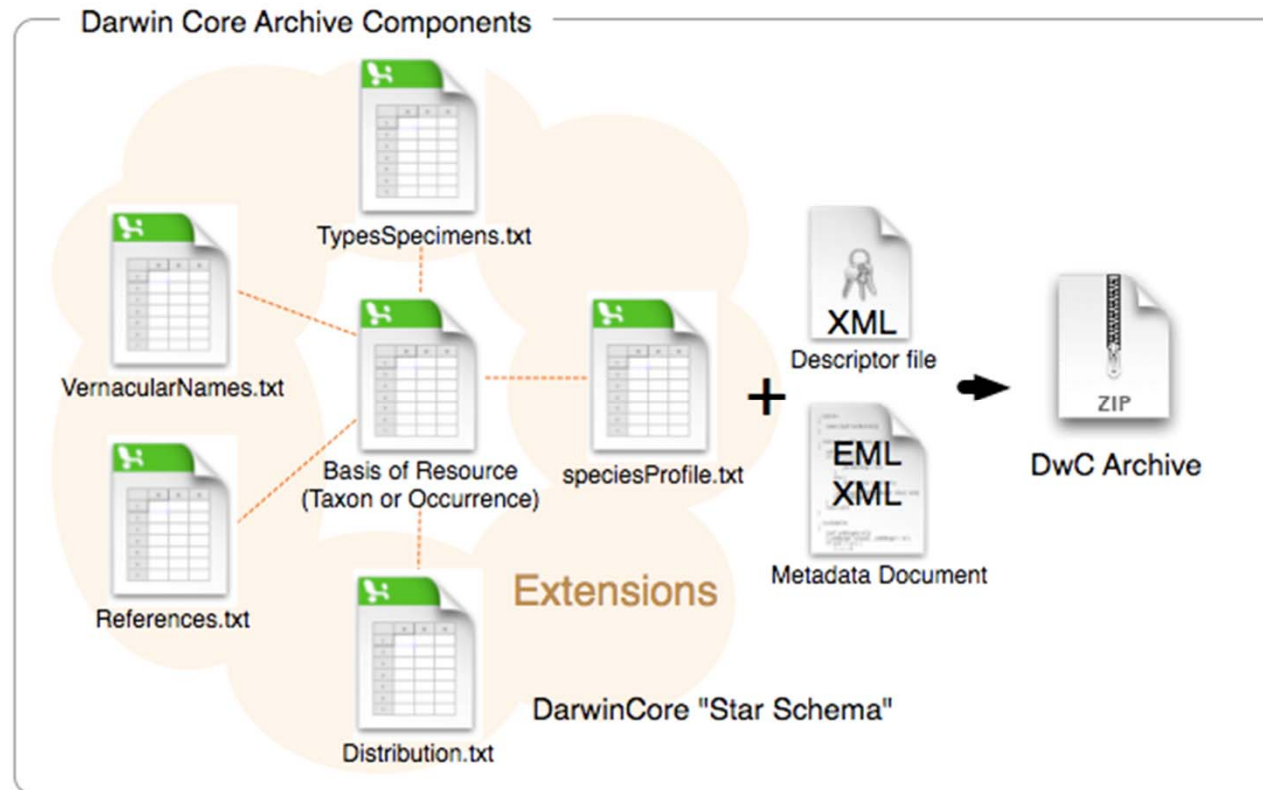
[3.8 GGBN Preparation Vocabulary](#)

[3.9 GGBN Preservation Vocabulary](#)

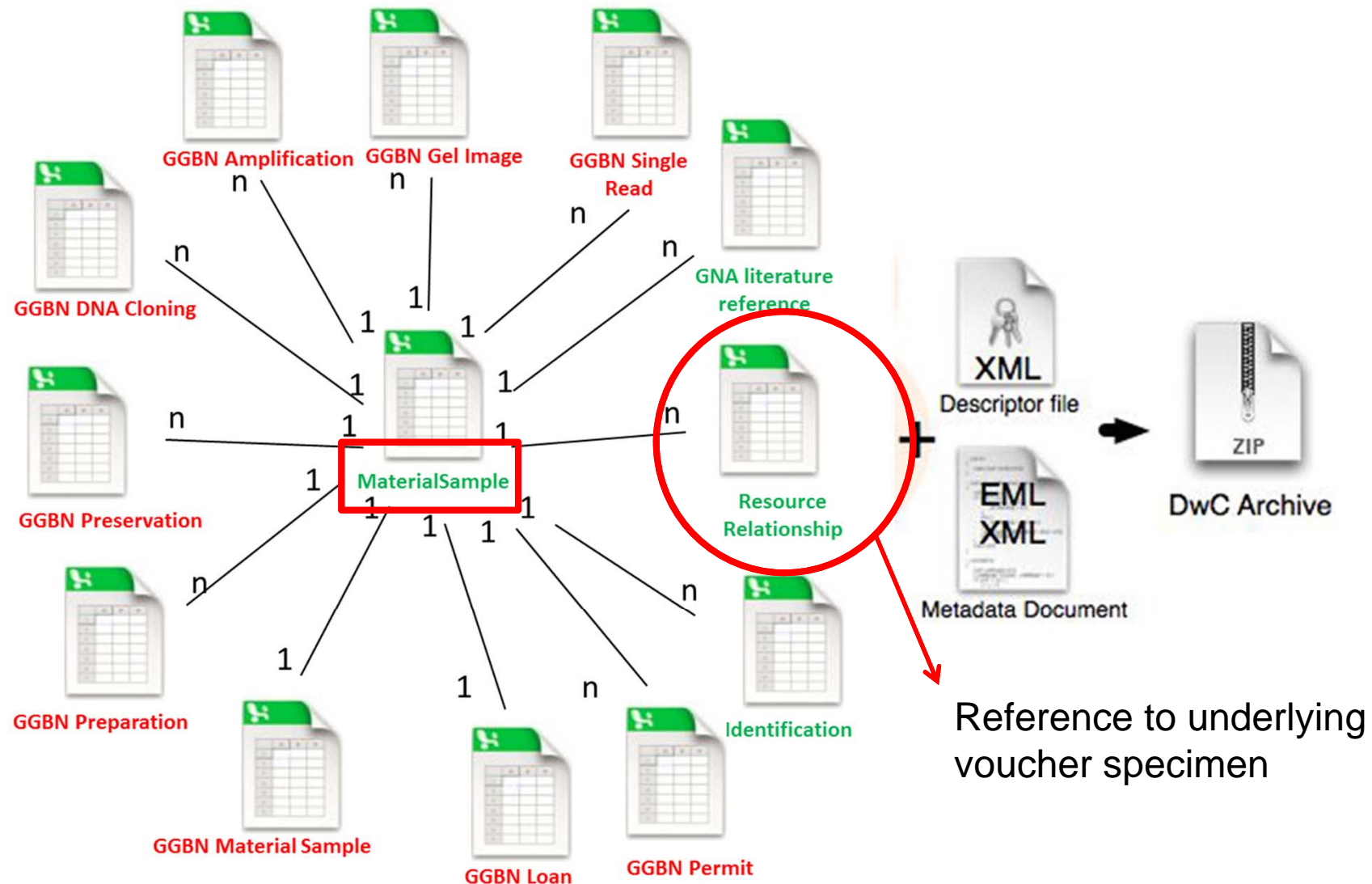
[3.10 GGBN Single Read Vocabulary](#)

DwC-A

*Global Genome
Biodiversity
Network (GGBN)*



GGBN Data Standard and DwC-A



GGBN Data Standard and DwC-A

- Next IPT release end of November
 - instructions on how to define MaterialSample as core and example mapping will be available via <http://www.ggbn.org>
- > <http://wiki.bgblm.org/dnabankwiki/index.php/DwC>

GGBN Data Standard and MlxS

- alignment with MlxS terms related to molecular issues

Page **Discussion** Read View form View source View history Search

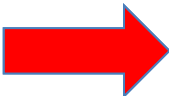
ggbn:preparationMaterials

Preparation Materials: Materials and chemicals used in the preparation of the specimen, tissue, DNA or RNA sample

Example(s): for DNA: DNeasy blood and tissue kit, CTAB;




Constraints in the context of [GGBN Data Standard](#):


- cardinality** 0..n (=optional multiple occurrence)



Preparation Materials

- URI:** <http://data.ggbn.org/schemas/ggbn/terms/preparationMaterials>
- skos: has exact match** [mixs:nucl_acid_ext](#)

RDF feed  | Browse properties  | SMW-prop. 

Search for values 

GGBN Data Standard and MlxS

- add MlxS term if missing

By concept name: abcd:AcquiredFrom • dcterms:description • dcterms:identifier • dwc:associatedReferences • dwc:disposition • dwc:materialSampleID • ggbn:BOLDProcessID • ggbn:DNADNAHybridization • ggbn:DNAMeltingPoint • ggbn:GC-content • ggbn:amplificationDate • ggbn:amplificationMethod • ggbn:amplificationStaff • ggbn:amplificationSuccess • ggbn:amplificationSuccessDetails • ggbn:barcodeSequence • ggbn:blocked • ggbn:blockedUntil • ggbn:chromatogramFileURI • ggbn:cloneStrain • ggbn:cloningDate • ggbn:cloningMethod • ggbn:cloningStaff • ggbn:concentration • ggbn:concentrationUnit • ggbn:consensusSequence • ggbn:consensusSequenceChromatogramFileURI • ggbn:consensusSequenceLength • ggbn:fragmentLength • ggbn:gelConcentration • ggbn:gelDuration • ggbn:gelLadder • ggbn:gelStain • ggbn:gelVoltage • ggbn:geneticAccessionNumber • ggbn:geneticAccessionURI • ggbn:haplotype • ggbn:loanConditions • ggbn:marker • ggbn:markerAccordance • ggbn:markerSubfragment • ggbn:methodDeterminationConcentrationAndRatios • ggbn:methodDeterminationWeight • ggbn:permitStatus • ggbn:permitText • ggbn:permitType • ggbn:permitURI • ggbn:plasmid • ggbn:preparationDate • ggbn:preparationMaterials • ggbn:preparationProcess • ggbn:preparationType • ggbn:preparedBy • ggbn:preservationDateBegin • ggbn:preservationTemperature • ggbn:preservationType • ggbn:primerName • ggbn:primerNameForward • ggbn:primerNameReverse • ggbn:primerReferenceCitation • ggbn:primerReferenceCitationForward • ggbn:primerReferenceCitationReverse • ggbn:primerReferenceLink • ggbn:primerReferenceLinkForward • ggbn:primerReferenceLinkReverse • ggbn:primerSequence • ggbn:primerSequenceForward • ggbn:primerSequenceReverse • ggbn:purificationMethod • ggbn:quality • ggbn:qualityCheckDate • ggbn:qualityRemarks • ggbn:ratioOfAbsorbance260 230 • ggbn:ratioOfAbsorbance260 280 • ggbn:sequence • ggbn:sequencingDate • ggbn:sequencingDirection • ggbn:sequencingMethod • ggbn:sequencingStaff • ggbn:singleSequence • ggbn:singleSequenceLength • ggbn:volume • ggbn:volumeUnit • ggbn:weight • ggbn:weightUnit • ggbnvoc:high • ggbnvoc:low • ggbnvoc:medium • ggbnvoc:quality vocabulary • mixs:adapters • mixs:annotationSource • mixs:assembly • mixs:chimeraCheck • mixs:estimatedSize • mixs:finishingStrategy • mixs:libConstMeth • mixs:libReadsSeqd • mixs:libSize • mixs:libScreen • mixs:libVector • mixs:mid • mixs:poolDNAExtracts • mixs:sampleSize • mixs:seqQualityCheck • mixs:sieving • mixs:sop

GGBN Data Standard and DwC-A

Source Data

Keine Datei ausgewählt.

Your source data files for generating a Darwin Core Archive. You can upload delimited text files (csv, tab, and files using any other delimiter) either directly or compressed (zip or gzip). Excel files are also supported. To (re)upload a file, please select the local file then click "Add".

Alternatively, you can configure SQL views to databases in your local network. To create a new SQL source, please click "Connect to database" without any file chosen.

amplifications [sql]	db=dnabank, 8 columns.	<input type="button" value="Edit"/>
dna_bank_local [sql]	db=dnabank, 77 columns.	<input type="button" value="Edit"/>
dna_bank_local_previous_ident [sql]	db=dnabank, 10 columns.	<input type="button" value="Edit"/>
loan [sql]	db=dnabank, 4 columns.	<input type="button" value="Edit"/>

Darwin Core Mappings

Darwin Core Material Sample ▼

Your mapping between the source data and Darwin Core terms.

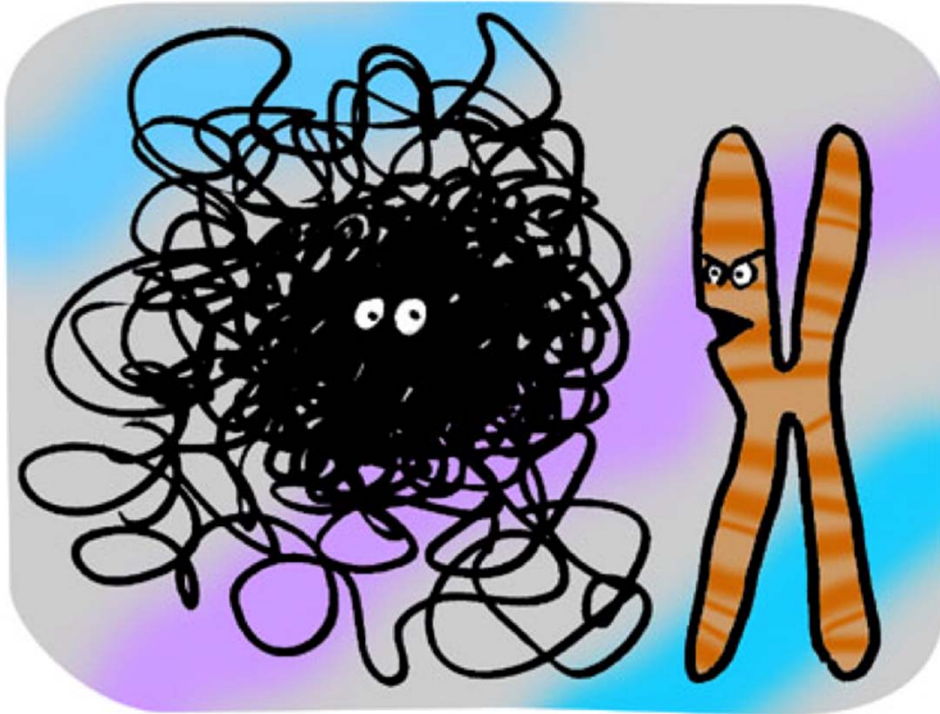
Core

Darwin Core Material Sample 22 terms mapped to dna_bank_local

Extensions

GGBN Material Sample Extension	7 terms mapped to dna_bank_local	<input type="button" value="Edit"/>
GGBN Preparation Extension	3 terms mapped to dna_bank_local	<input type="button" value="Edit"/>
Darwin Core Resource Relationship	2 terms mapped to dna_bank_local	<input type="button" value="Edit"/>
GGBN Preservation Extension	2 terms mapped to dna_bank_local	<input type="button" value="Edit"/>
GGBN Amplification Extension	6 terms mapped to amplifications	<input type="button" value="Edit"/>
Darwin Core Identification History	5 terms mapped to dna_bank_local_previous_ident	<input type="button" value="Edit"/>
GGBN Loan Extension	2 terms mapped to loan	<input type="button" value="Edit"/>

Thank you



Dude, mitosis starts in five minutes...
I can't believe you're not condensed yet.

Thanks to:

GGBN

GBIF

GSC

Funded by:

DFG & SYNTHESYS

<http://www.ggbn.org>

<http://wiki.bgbm.org/dnabankwiki/index.php/DwC>