



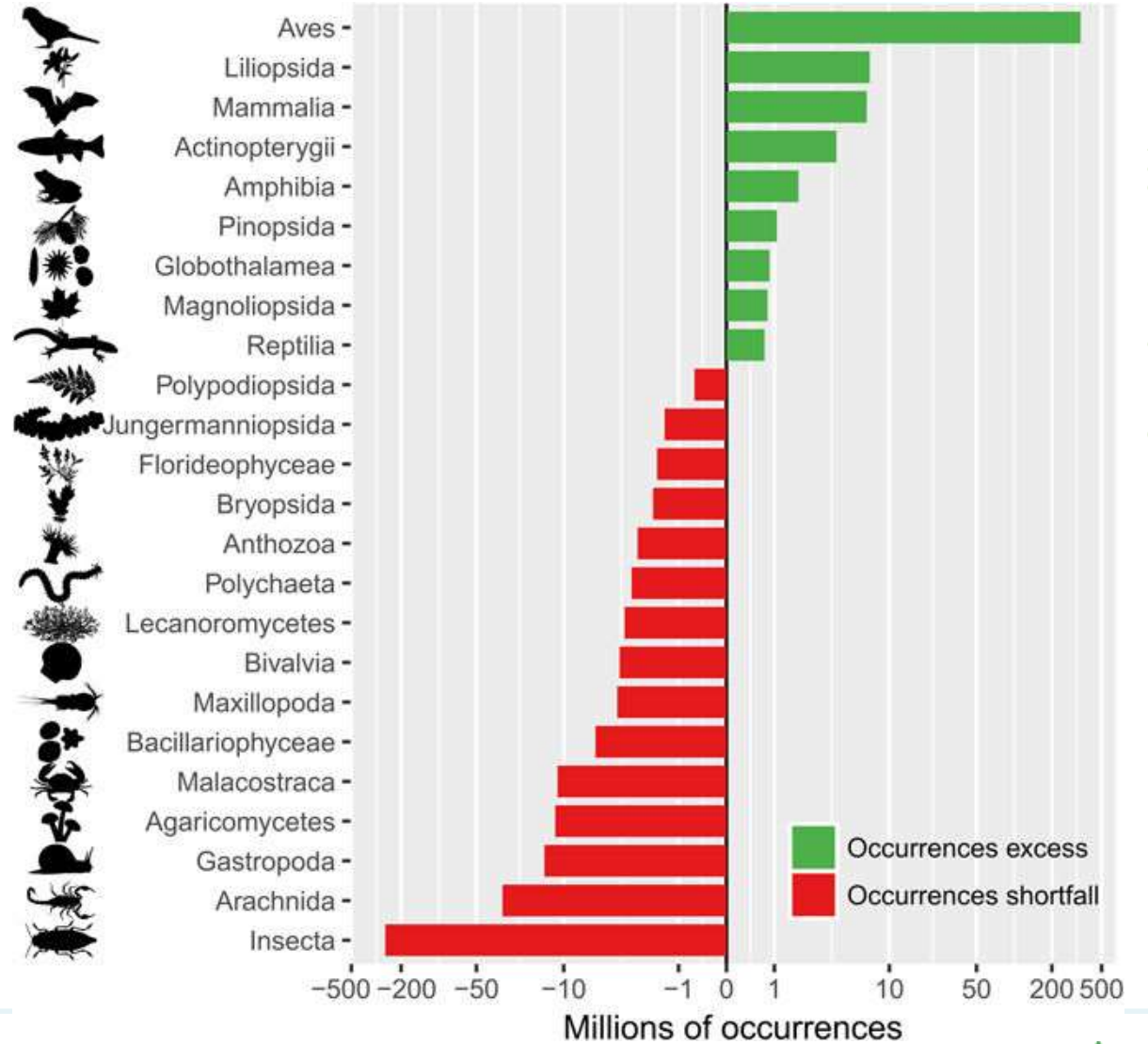
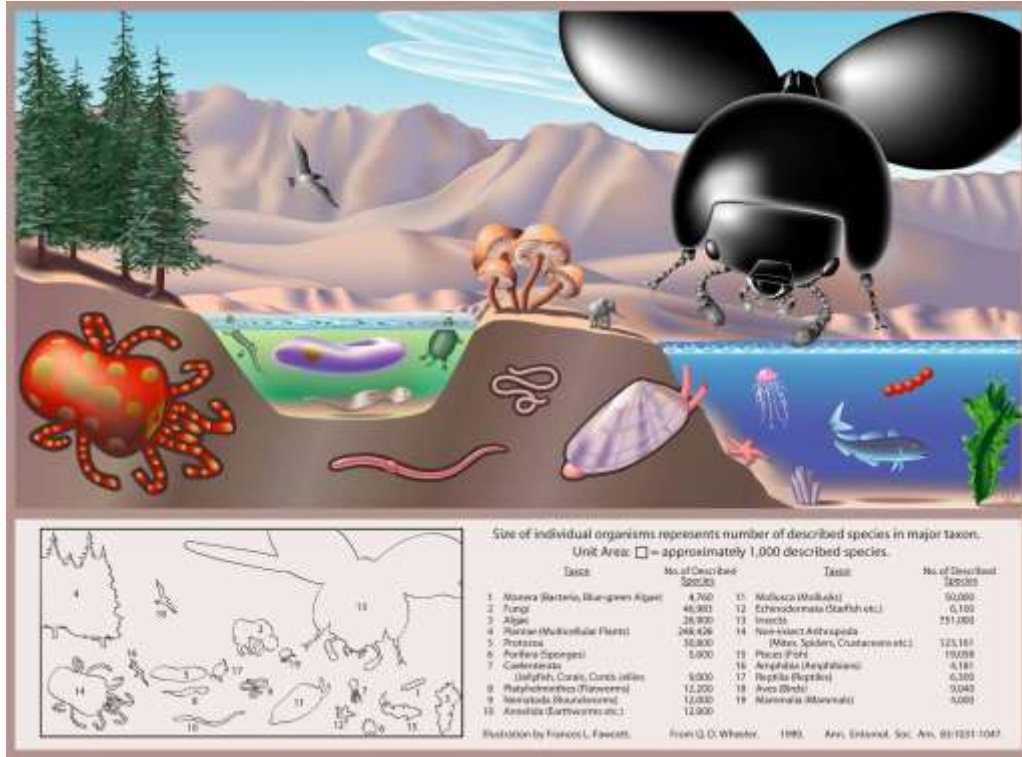
Update on the forthcoming GBIF guide to publishing sequence-based biodiversity data

Dmitry Schigel | Scientific officer



GLOBAL BIODIVERSITY VS. DIGITALLY AVAILABLE DATA

Image: FL Fawcett in Wheller Ann. Entomol. Soc. Am. 1990



Troutlet et al. Nature Scientific Reports 2017

CRYPTIC BIODIVERSITY IS UNDERREPRESENTED IN GBIF

- **Detection** difficulties: size, habitats, methods
 - Only emerging in natural history outside academia
- **Identification** difficulties: Collin's guide won't do!
 - Barcoding and molecular identification techniques
- **Attention** issues: few citizen science projects, lack of popularization, no flowers & feathers
 - Molecular citizen science, microbial social media
- **Isolation** of professional communities
 - Better interlinking happens as we speak
- **Technical** developments at GBIF.org are new OTU and DNA work line to grow



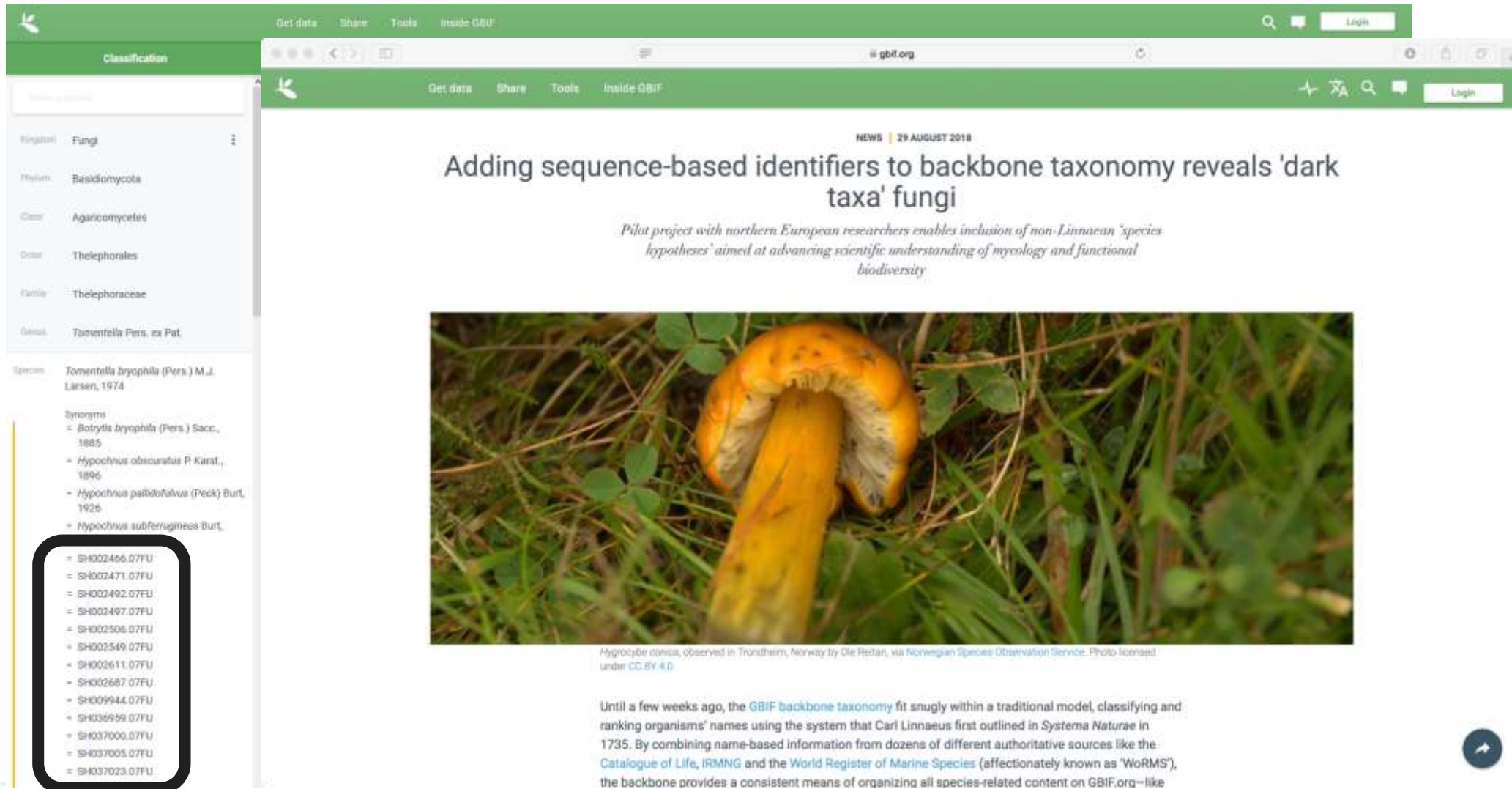
BEYOND LATIN NAMES

UNITE and BOLD

p. 9, cum ic. — SACCARDO, V, 364. — QUELET, FLORE, MYC. 192. — PATOUILLARD, Tab. no. 6. — BIGEARD et GUILLEMIN, II, 124. — FRIES, H. E., 182. — [Non Panus craterellus in KALCHBRENNER, Icones Fung. Hung. p. 48, t. XXI, f. 4! = ? P. violaceofulvus Batsch. f. Delastrei (Mont.) Malkovský.] — **Pleurotus canus** f. **griseo-rosella** BRESADOLA in herbario (h. S.). Exsiccata: RABENHORST: Fungi Europei, No. 2510. — SACCARDO: Mycotheca italica, No. 403. — SACCARDO: Mycotheca Veneta, No. 1106.

CHAPEAU 3—6 mm., sessile, orbiculaire, caliciforme, à la fin incliné, mince, cilié et hérissé de poils soyeux, puis frisé blanc de neige; ressemblent plus aux Discomy

UNITE'S MOLECULAR NON-LINNAEAN NAMES OF FUNGI



Get data Share Tools Inside GBIF

Classification

Kingdom: Fungi

Phylum: Basidiomycota

Class: Agaricomycetes

Order: Thelephorales

Family: Thelephoraceae

Genus: Tomentella Pers. ex Pat.

Species: Tomentella bryophila (Pers.) M.J. Larsen, 1974

Synonyms

- Botrytis bryophila (Pers.) Sacc., 1885
- Hypochnus obscuratus P. Karst., 1896
- Hypochnus pallidofusca (Peck) Burl., 1926
- Hypochnus subferrugineus Burl., 1926

- SH002466.07FU

- SH002471.07FU

- SH002492.07FU

- SH002497.07FU

- SH002506.07FU

- SH002549.07FU

- SH002611.07FU

- SH002687.07FU

- SH009944.07FU

- SH036959.07FU

- SH037000.07FU


- SH037005.07FU

- SH037023.07FU

NEWS | 29 AUGUST 2018

Adding sequence-based identifiers to backbone taxonomy reveals 'dark taxa' fungi

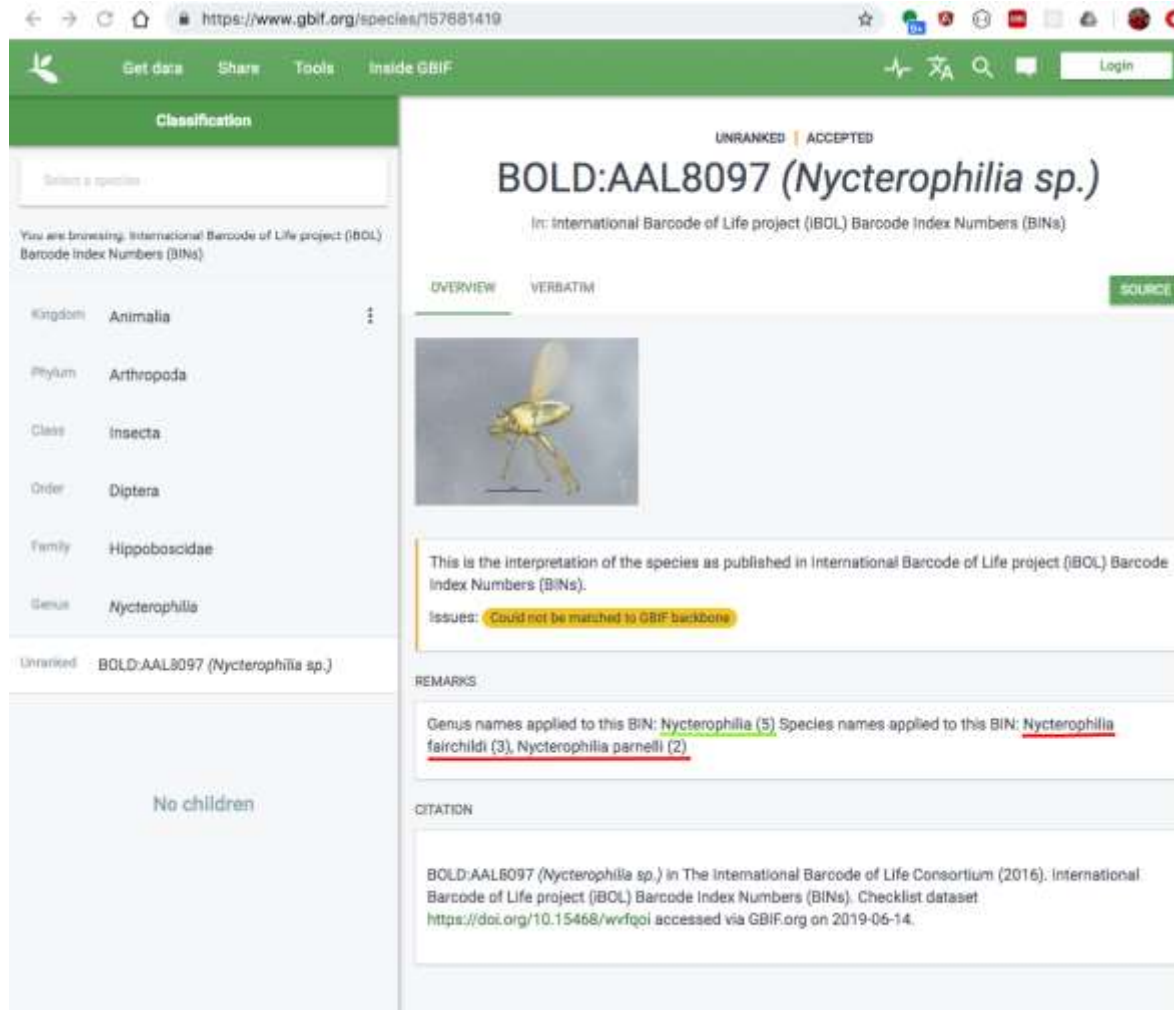
Pilot project with northern European researchers enables inclusion of non-Linnaean 'species hypotheses' aimed at advancing scientific understanding of mycology and functional biodiversity



Hypocyber conica, observed in Trondheim, Norway by Ole Retan, via Norwegian Species Observation Service. Photo licensed under CC-BY 4.0.

Until a few weeks ago, the GBIF backbone taxonomy fit snugly within a traditional model, classifying and ranking organisms' names using the system that Carl Linnaeus first outlined in *Systema Naturae* in 1735. By combining name-based information from dozens of different authoritative sources like the [Catalogue of Life](#), [IRMNG](#) and the [World Register of Marine Species](#) (affectionately known as 'WoRMS'), the backbone provides a consistent means of organizing all species-related content on GBIF.org—like

BINs IN GBIF AS HANDLES FOR ANIMAL OTUs



https://www.gbif.org/species/157681419

Classification

Select a specimen

You are browsing: International Barcode of Life project (IBOL) Barcode Index Numbers (BINs)

Kingdom: Animalia

Phylum: Arthropoda

Class: Insecta

Order: Diptera

Family: Hippoboscidae

Genus: Nycterophila

Unranked: BOLD:AAL8097 (Nycterophila sp.)


No children

UNRANKED | ACCEPTED

BOLD:AAL8097 (*Nycterophila* sp.)

In: International Barcode of Life project (IBOL) Barcode Index Numbers (BINs)

OVERVIEW VERBATIM SOURCE



This is the interpretation of the species as published in International Barcode of Life project (IBOL) Barcode Index Numbers (BINs).

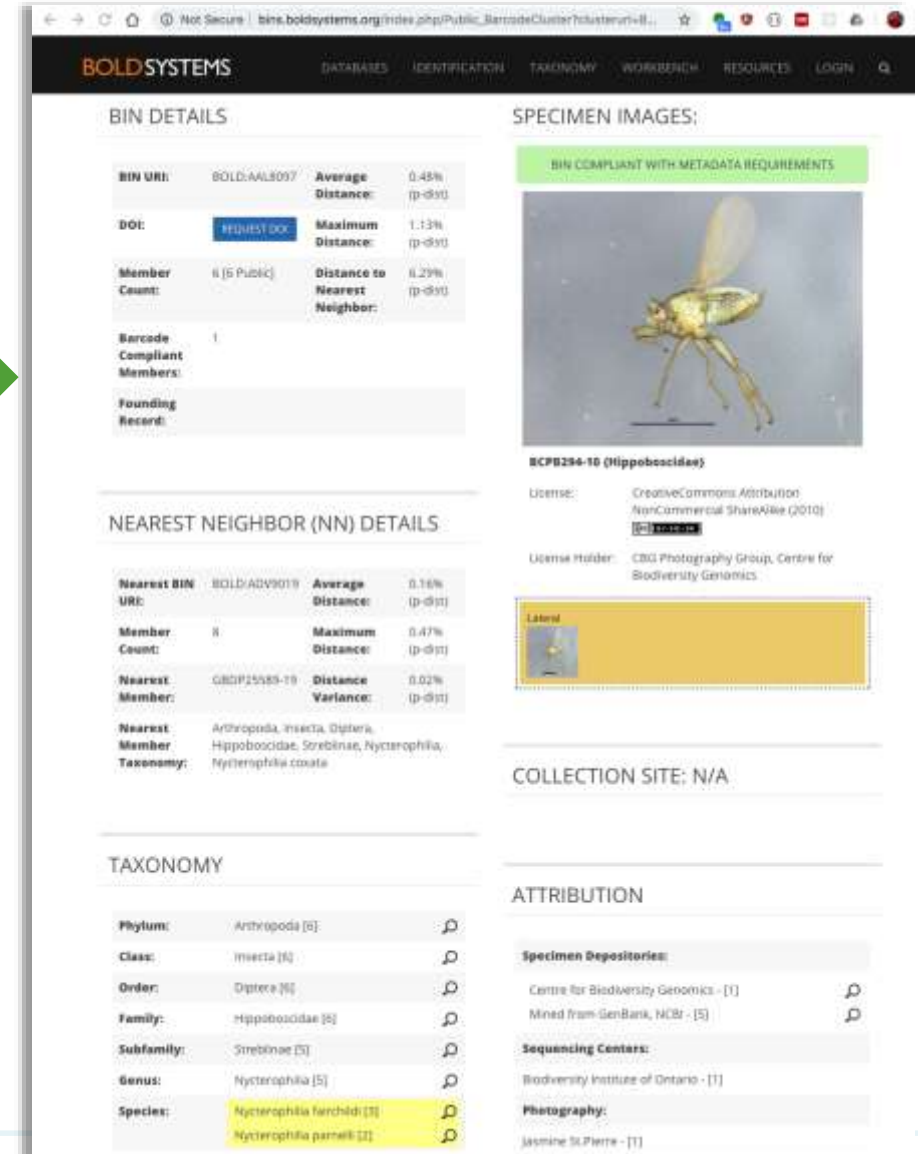
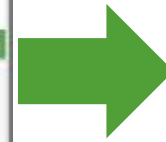
Issues: **Could not be matched to GBIF backbone**

REMARKS

Genus names applied to this BIN: Nycterophila (5) Species names applied to this BIN: Nycterophila fairchildi (3), Nycterophila parnellii (2)

CITATION

BOLD:AAL8097 (*Nycterophila* sp.) in The International Barcode of Life Consortium (2016). International Barcode of Life project (IBOL) Barcode Index Numbers (BINs). Checklist dataset <https://doi.org/10.15468/wvfqoi> accessed via GBIF.org on 2019-06-14.



BOLD SYSTEMS


DATABASES IDENTIFICATION TAXONOMY WORKBENCH RESOURCES LOGIN

BIN DETAILS

BIN URI:	BOLD:AAL8097	Average Distance:	0.45% (p=0.0)
DOI:	REQUEST DOI	Maximum Distance:	1.13% (p=0.0)
Member Count:	6 (5 Public)	Distance to Nearest Neighbor:	6.29% (p=0.0)
Barcode Compliant Members:	1		
Founding Record:			

SPECIMEN IMAGES:

BIN COMPLIANT WITH METADATA REQUIREMENTS



BCPB236-10 (Hippoboscidae)

License: Creative Commons Attribution NonCommercial ShareAlike (2010) [CC BY-NC-SA](#)

License Holder: CBG Photography Group, Centre for Biodiversity Genomics

NEAREST NEIGHBOR (NN) DETAILS

Nearest BIN URI:	BOLD:ADV9019	Average Distance:	0.16% (p=0.0)
Member Count:	6	Maximum Distance:	0.47% (p=0.0)
Nearest Member:	GBOP25585-19	Distance Variance:	0.02% (p=0.0)
Nearest Member Taxonomy:	Arthropoda, Insecta, Diptera, Hippoboscidae, Streblinae, Nycterophila, Nycterophila coxate		

TAXONOMY

Phylum:	Arthropoda [6]	
Class:	Insecta [6]	
Order:	Diptera [6]	
Family:	Hippoboscidae [6]	
Subfamily:	Streblinae [5]	
Genus:	Nycterophila [5]	
Species:	Nycterophila fairchildi [3]	
	Nycterophila parnellii [2]	

COLLECTION SITE: N/A

ATTRIBUTION

Specimen Depositories:

- Centre for Biodiversity Genomics - [1]
- Mined from GenBank, NCBI - [5]

Sequencing Centers:

- Biodiversity Institute of Ontario - [1]

Photography:

- Jasmine St-Pierre - [1]



GBIF
backbone taxonomy

SH ABC0001



OTU = **SH**,
Species hypothesis

BIN DEF0002



international
BARCODE
OF LIFE



OTU = **BIN**,
Barcode
identification number

WILD SEQUENCES WITH COORDINATES

EBI: ENA and BOLD

Occurrences

SEARCH OCCURRENCES | 1,309,506 WITH IMAGES

Search all fields

Simple Advanced

License

Scientific name

Basis of record

Location

Coordinate uncertainty in meters

Year

Month

Dataset

International Barcode of Life project (iBOL)

Country or area


































Continent

Issues and flags

Media type

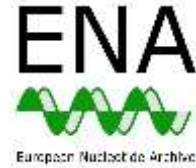
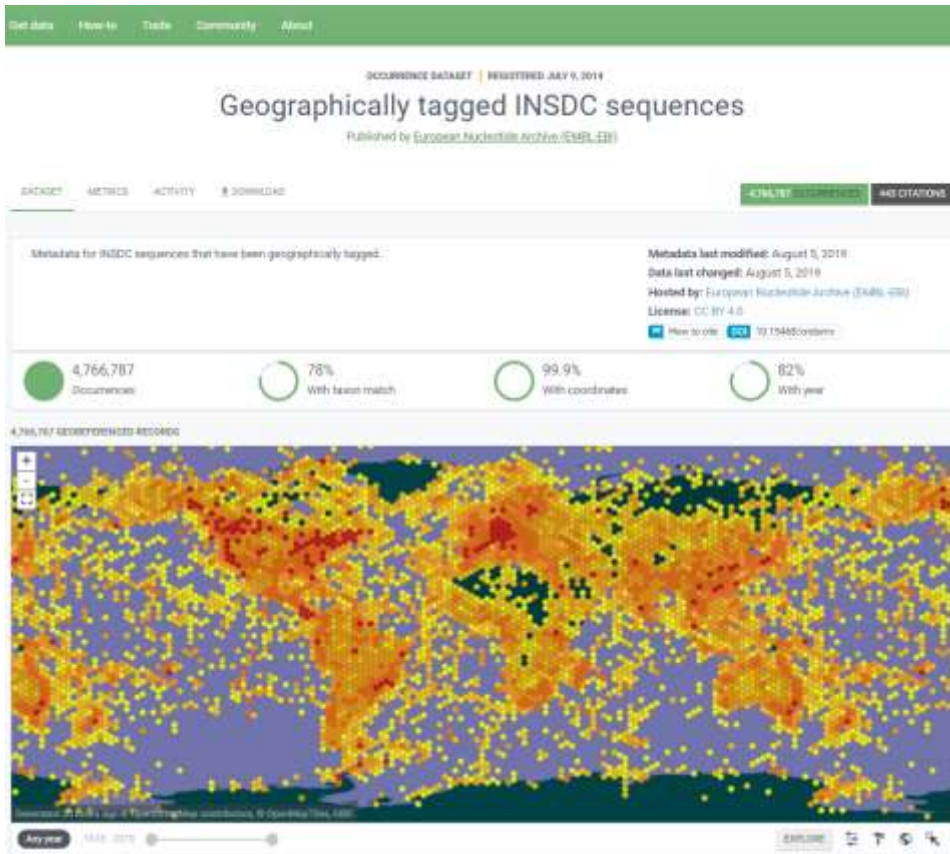
Publisher

TABLE GALLERY MAP TAXONOMY METRICS DOWNLOAD

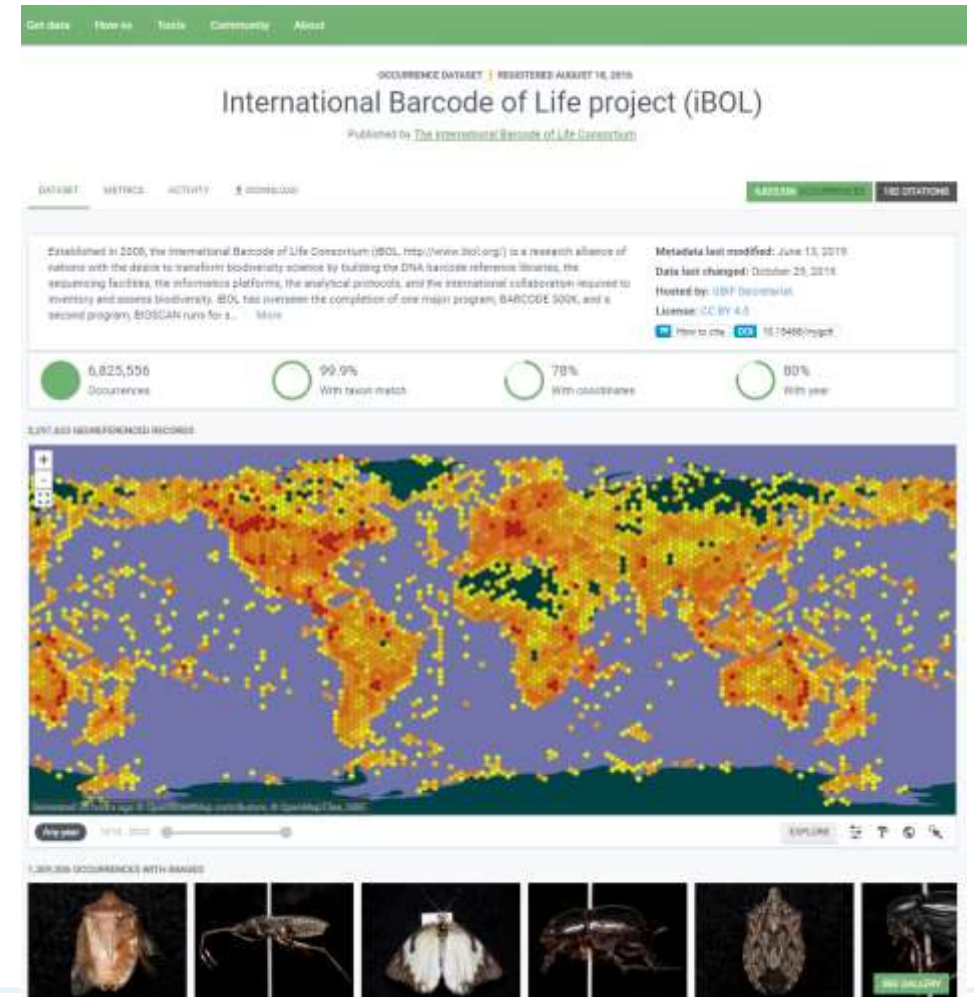
 <p>Hemiptera</p>	 <p>Hemiptera</p>	 <p>Belenois aurata (Fabricius, 1793)</p>	 <p>Coleoptera</p>	 <p>Hemiptera</p>	 <p>Coleoptera</p>	 <p>Garra ceylonensis Bleeker, 1863</p>	 <p>Trochilocharax Zarske, 2010</p>	 <p>Pristella maxillaris (Ulrey, 1894)</p>	 <p>Cichlidae</p>	 <p>Hyphessobrycon columbianus Zarske & Géry, 2002</p>
 <p>Nebria brevicollis (Fabricius, 1792)</p>	 <p>Enterobius vermicularis (Linnaeus, 1758)</p>	 <p>BOLD-ADF5939</p>	 <p>Rekoa manius (Lucas, 1857)</p>	 <p>Penaincisalia curva</p>	 <p>BOLD-AA18030</p>	 <p>Heterusia Hübnér, 1831</p>	 <p>BOLD-AD18930</p>	 <p>BOLD-ADD2179</p>	 <p>BOLD-ABX0486</p>	 <p>Thecla Fabricius, 1807</p>
 <p>Penaincisalia curva</p>	 <p>BOLD-AAF0156</p>	 <p>BOLD-AEB6602</p>	 <p>BOLD-ABX0486</p>	 <p>BOLD-ADR4993</p>	 <p>BOLD-ABX0486</p>	 <p>Cerodontha lapplandica (Rydén, 1956)</p>	 <p>BOLD-ACL0132</p>	 <p>Papaver somniferum L.</p>	 <p>BOLD-AEB0921</p>	 <p>BOLD-AEB0921</p>

INDIVIDUAL SEQUENCES WITH COORDINATES

European Nucleotide Archive: 4.8M records



International Barcode of Life: 6.8M records



METABARCODING AND METAGENOMICS

EBI: MGnify; BLOWIDE (Denmark)



PUBLISHING AND INDEXING DATA FROM eDNA SAMPLING EVENTS

eDNA based dataset



Hybrid Linnaean and OTU based taxonomy

https://www.gbif.org/dataset/5b3c1e18-b6c2-4264-ac52-95d772c05e9f

Get data Share Tools Inside GBIF

OCCURRENCE DATASET | REGISTERED JULY 12, 2016

BIOWIDE eDNA Fungi dataset

Published by Danish Biodiversity Information Facility
Tobias Freese • Rasmus Emsw

DATASET PROJECT METRICS ACTIVITY DOWNLOAD HOW TO ASK

The Biowide project (2014-2018) was a project aiming at collecting biodiversity data from 130 terrestrial sampling sites across Denmark. Data was collected with both classical means (observation and trapping) and by eDNA metabarcoding (sequencing of amplified marker genes). Data was also collected on environmental variables. Biowide (Biodiversity in Width and Depth) took place in collaboration between Århus University (main project holder) and the two major Danish natural history museums and the... [More](#)

Metadata last modified: July 12, 2018
Data last changed: December 12, 2018
Hosted by: Danish Biodiversity Information Facility
License: CC BY-NC 4.0
[How to cite](#) [DOI](#) [10.15468/biowide](#)

30,091 Occurrences
100% With taxon match
100% With coordinates
100% With year

30,091 GEOREFERENCED RECORDS

Year 2014 EXPLORE AREA

Event ID	Event date	Sampling protocol	Occurrences count
BIOWIDE-SV088	23 October 2014	eDNA sampling from soil	476
BIOWIDE-FL119	28 October 2014	eDNA sampling from soil	497
BIOWIDE-SV084	23 October 2014	eDNA sampling from soil	434
BIOWIDE-FM034	13 November 2014	eDNA sampling from soil	434
BIOWIDE-FF108	05 October 2014	eDNA sampling from soil	401

Get data Share Tools Inside GBIF Admin

Classification

- Hygrophorus conicus var. conicopalustris R.Haller Aar.
- Hygrophorus conicus var. conicopalustris R.Haller Aar. ex Arnolds, 1974
- Hygrophorus conicus var. conicus (Schaeff.) Fl., 1836
- Hygrophorus conicus var. olivaceoniger (P.D.Orton) Arnolds
- Hygrophorus conicus var. peradenyca Berk. & Broome, 1871
- Hygrophorus conicus var. peradenyca Sacc., 1887
- Hygrophorus conicus var. tristicus
- Hygrophorus nigrescens var. brevissporus Dennis, 1961
- Hygrophorus olivaceoniger P.D.Orton

Species | ACCEPTED

Hygrocybe conica (Schaeff.) P.Kumm.

Published in: Führ. Pilzk. (Zwickau) 111 (1871) source: Catalogue of Life
In: GBIF Backbone Taxonomy
Toppraxskivling in Swedish Basionym: Agaricus conicus Schaeff.

OVERVIEW METRICS REFERENCE TAXON

22,302 OCCURRENCES 12 SUBSPECIES

1,091 OCCURRENCES WITH IMAGES

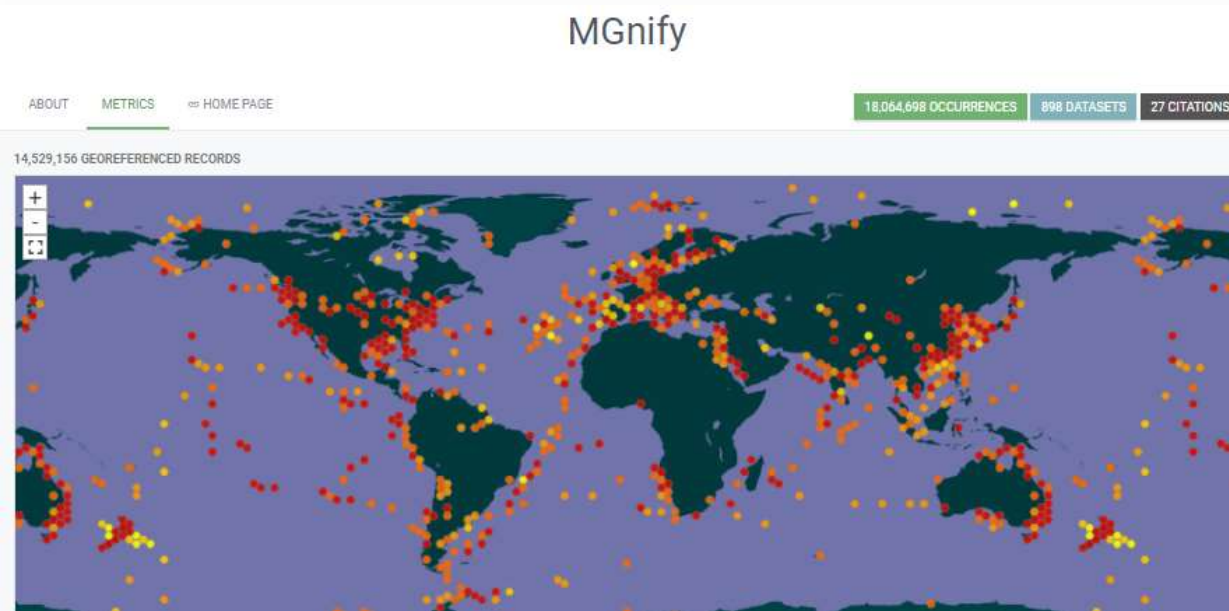
21,036 GEOREFERENCED RECORDS

Immediate children

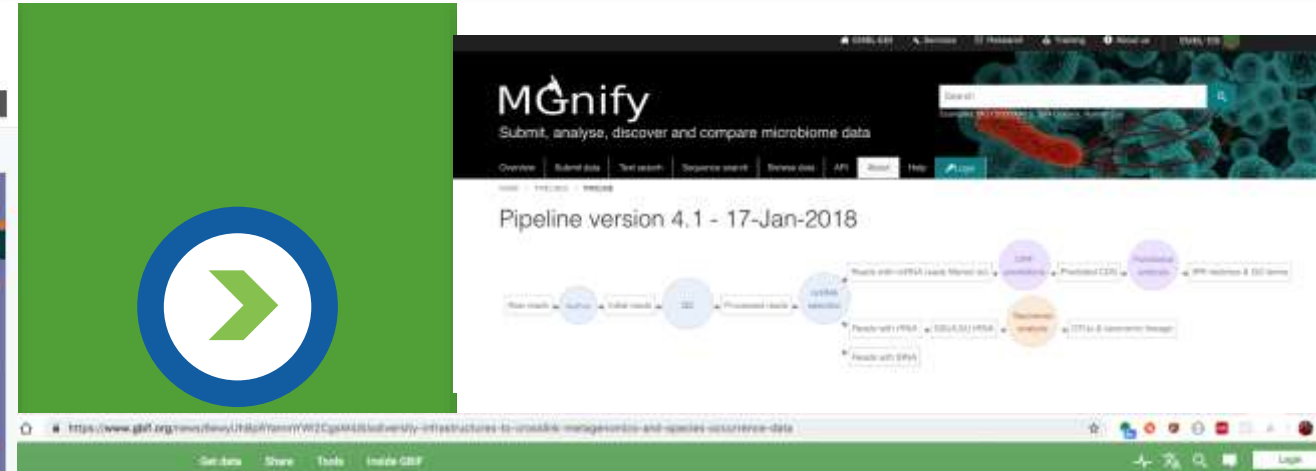
- Variety: Hygrocybe conica var. pseudoconica J.E.Lange
- Form: Hygrocybe olivaceonigra f. luteomarginata Bon & B.Lafabvre
- Form: Hygrocybe olivaceonigra f. olivaceonigra
- Unranked: SH176826.07FU (cf. Hygrocybe conica)
- Unranked: SH176829.07FU (cf. Hygrocybe conica)
- Unranked: SH176830.07FU (cf. Hygrocybe conica)
- Unranked: SH176833.07FU (cf. Hygrocybe conica)
- Unranked: SH176834.07FU (cf. Hygrocybe conica)
- Unranked: SH176839.07FU (cf. Hygrocybe conica)
- Unranked: SH176845.07FU (cf. Hygrocybe conica)
- Unranked: SH176849.07FU (cf. Hygrocybe conica)
- Unranked: SH529723.07FU (cf. Hygrocybe conica)

Any year 1800 - 2019 EXPLORE

MGNIFY INCREASES COVERAGE OF UNDERREPRESENTED TAXA

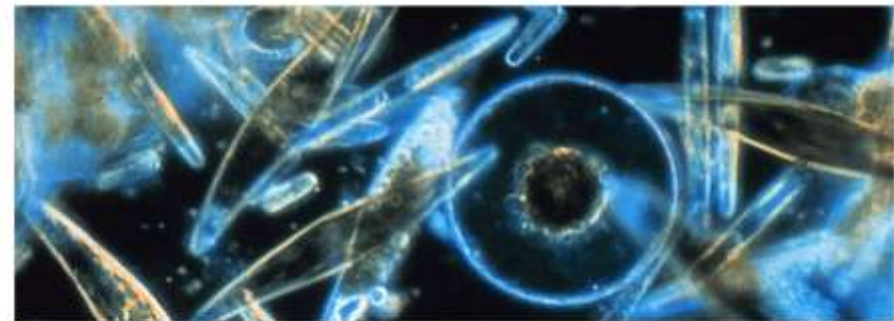


898 datasets: 18.1M new records:
13.1M Bacteria
2.2M Chromista



Biodiversity infrastructures to crosslink metagenomics and species occurrence data

Deepening collaboration between GBIF and EMBL-EBI to address taxonomic bias and reduce barriers between big FAIR and open data for biodiversity and metagenomics



GBIF and the European Bioinformatics Institute (EMBL-EBI) will extend their collaboration by sharing evidence (species occurrence records) of living creatures and communities known only from their genetic material. This collaboration around metagenomic data adds a significant new data stream to GBIF.org and marks an important step in bridging the gap between biodiversity studies based on

DATA STREAMS IN GBIF

sequence-based evidence

natural history collections

human observations etc.

CHECKLISTS

CHECKLISTS

Sequence names

“taxon hypotheses”, OTUs



Sequence occurrences

Latin names



individuals / isolates

species mixes (samples)

Amplicon (barcoding)
one or more marker genes

Genomic *shotgun* DNA

Transcriptomic *shotgun* RNA

Amplicon (metabarcoding)
one or more marker genes

Metagenomic *shotgun* DNA

Metatranscriptomic *shotgun* RNA



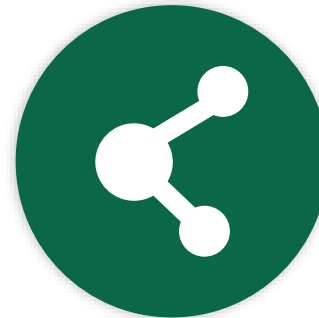
OCCURRENCE DATASETS

SAMPLING EVENT DATASETS

NEW DATA PUBLISHING GUIDE COMING IN 2020: KEY FACTS

Publishing sequence-derived data through biodiversity discovery platforms

- 12 authors: Australia, Norway, Sweden, Denmark, UNITE, and GBIF
- Based on practical mapping and data publishing experiences
- Cross-platform
- About 50 pages
 - ❖ Introduction
 - ❖ Categorization
 - ❖ Mapping
 - ❖ Visuals
 - ❖ Future prospects
 - ❖ Resources: glossary, links, references



PUBLISHING SEQUENCE-DERIVED DATA: THE “LEARN” SECTION

2. Introduction

2.1 Rationale

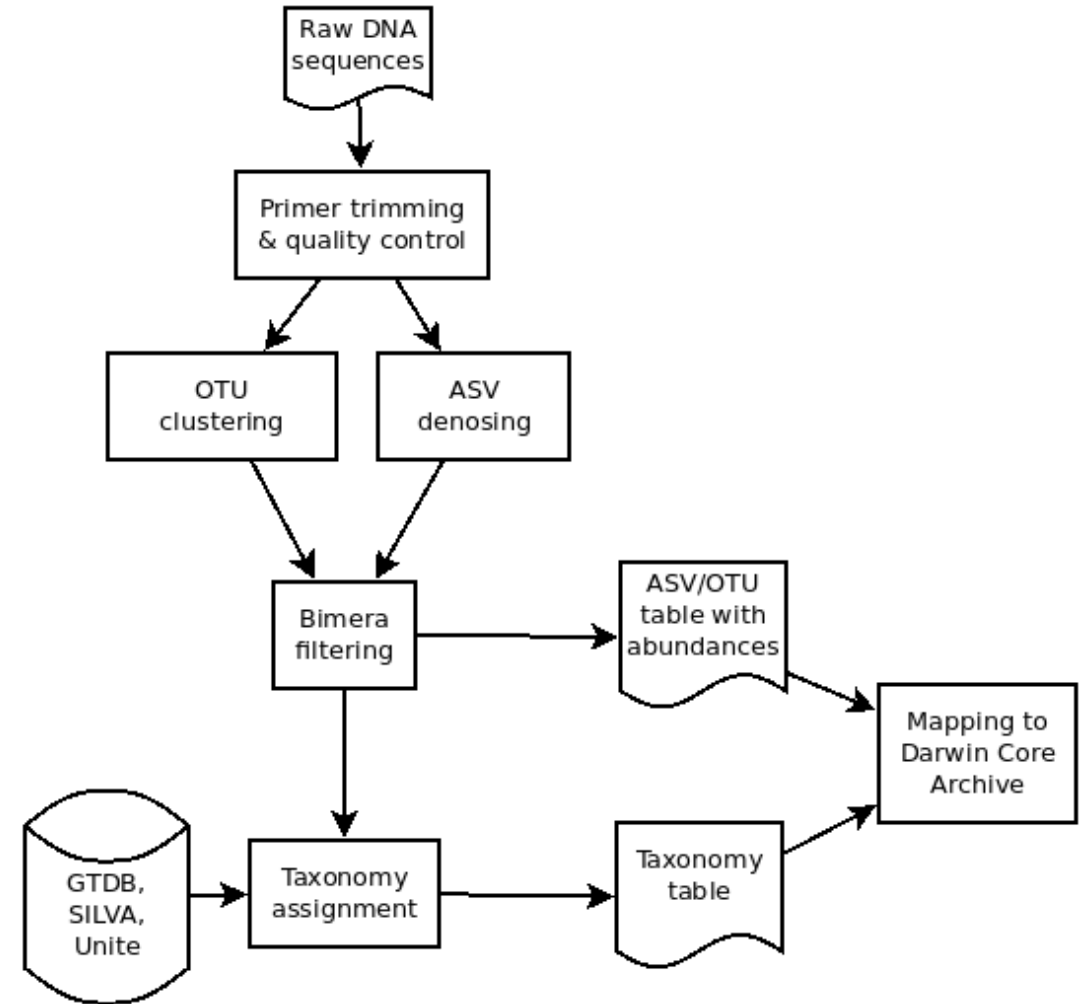
2.2 Target audiences

2.3 Introduction to sequence derived occurrence data

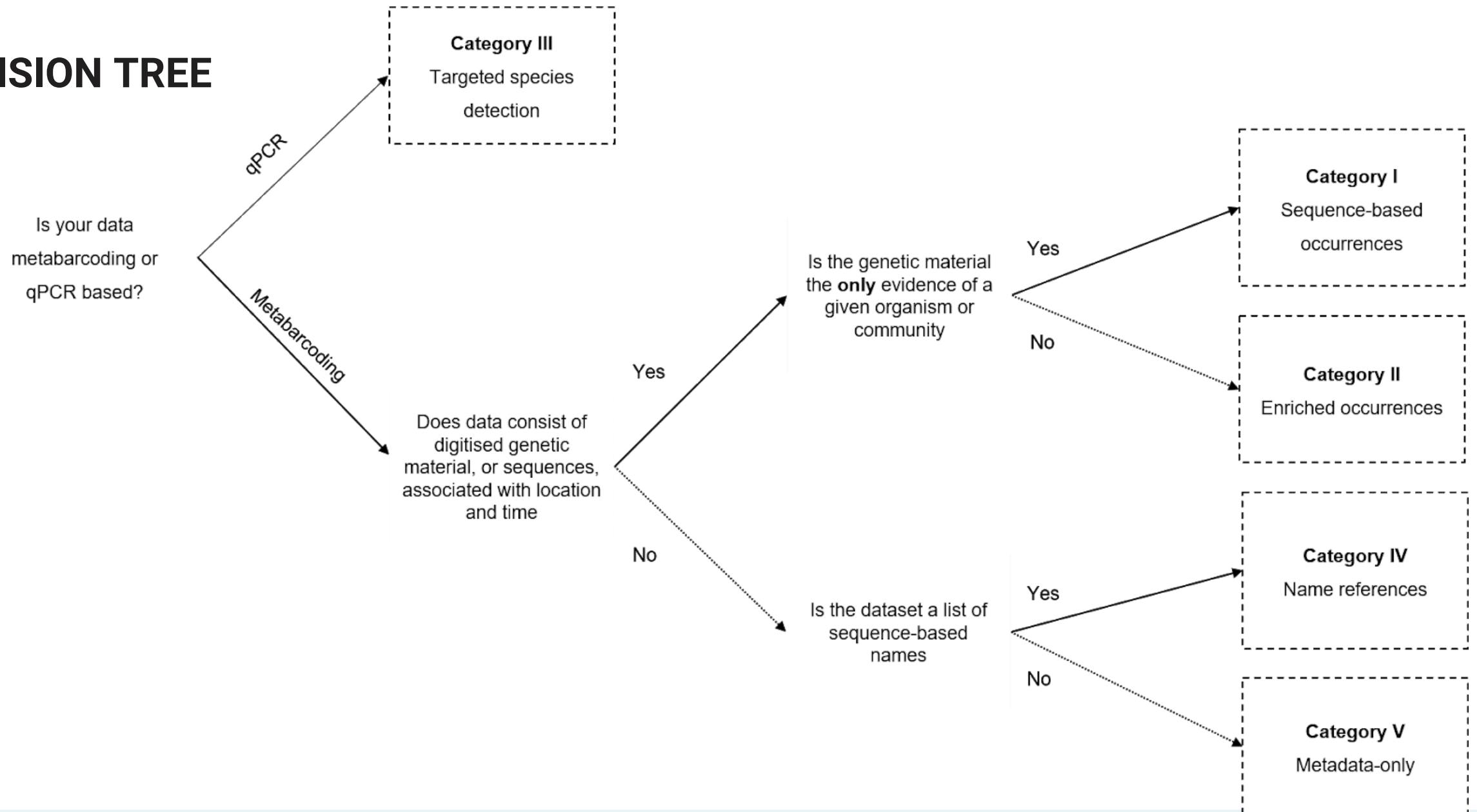
2.4 Introduction to biodiversity publishing

2.5 Processing workflows – from sample to ingestible data

2.6 Taxonomy of sequences



DECISION TREE



PUBLISHING SEQUENCE-DERIVED DATA: THE “DO” SECTION

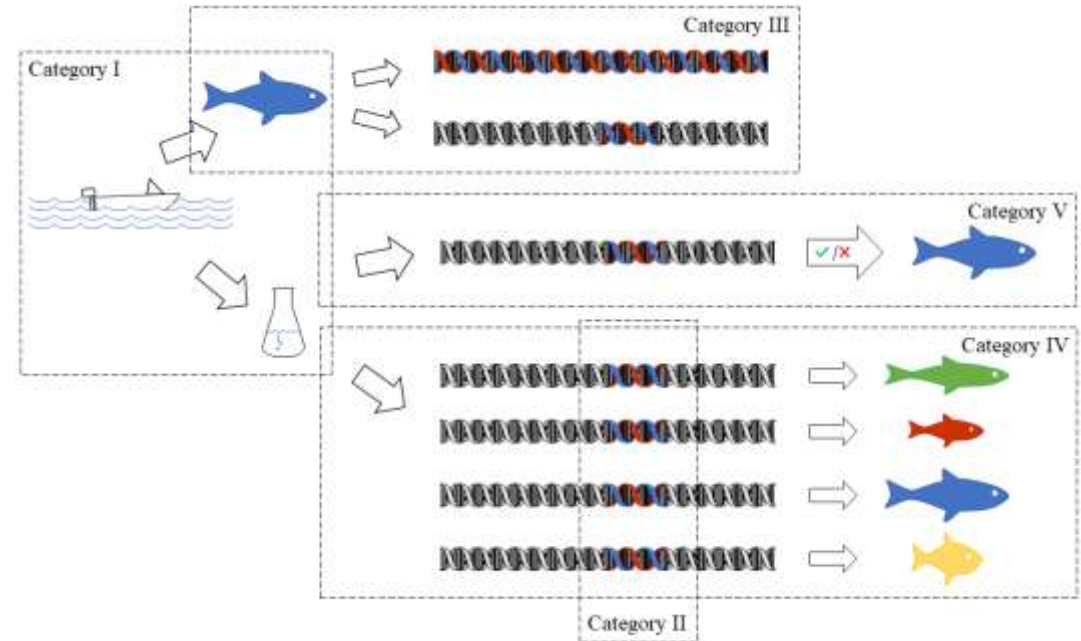
3. Data packaging and mapping

3.1 Categorisation of data - from samples to processed data

- I Sequence-based occurrences
- II Enriched occurrences
- III Targeted species detection (qPCR)
- IV Name references
- V Metadata-only

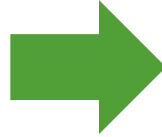
3.2 Mapping recommendations

Examples



TIMELINE AND NEXT STEPS

- Draft closed for edits today
- Copy-editing and formatting
- **Consultation and feedback (July)**
- Online release (autumn)
- Peer-reviewed publication of v 1.0
- Live guide, versioned as needed



PDF file
español
français

Contribute

- Create an issue
- Edit on GitHub

Table of Contents

Colophon

- [Suggested citation](#)
- [Contributors](#)
- [Licence](#)
- [Persistent URI](#)
- [Document Control](#)
- [Cover image](#)

Introduction

1. What is a GBIF Participant node?
 - Box 1. Definitions of key structures and roles in the GBIF network at the level of Participant country
 - Box 2. GBIF Head of Delegation and Node Manager: roles and responsibilities

Establishing an Effective Node *Concepts and general*

GBIF Secretariat – nodes@gbif.org – Version 1.0, 2020-06-17 14:40:58 UT

This document is also available in [PDF format](#) and in other languages



NEW HYBRID TRAINING: GBIF, BOLD, UNITE – WITH GBIF NORWAY & BIODATA

Share

New course to pilot in Tbilisi, Georgia in 2020 or 2021
4 days

* * *

Accelerating biodiversity research through
DNA barcodes, collection and observation data

- Publish to and retrieve data from GBIF, BOLD, and UNITE
- Use genetic sequence data as a DNA barcode to identify a species
- Basics of data capture, cleaning, georeferencing, and data citation
- Critically assess data quality and fitness for purpose
- Practice key tools
- FAIR data principles in biodiversity research and collaboration

Use

THANK YOU

Dmitry Schigel

dschigel@gbif.org

