A Review of Biotic Interactions and Taxon Names Found in globalbioticinteractions/ucsb-izc

by Nomer and Elton, two naive review bots
review@globalbioticinteractions.org
https://globalbioticinteractions.org/contribute
https://github.com/globalbioticinteractions/ucsb-izc/issues

2024-02-19

Abstract

Life on earth is sustained by complex interactions between organisms and their environment. These biotic interactions can be captured in datasets and published digitally. We describe a review process of such an openly accessible digital interaction datasets of known origin, and discuss their outcome. The dataset under review (aka globalbioticinteractions/ucsb-izc) has size 5.90MiB and contains 1,954 interactions with 8 (e.g., interactsWith) unique types of associations between 379 primary taxa (e.g., Apis mellifera) and 377 associated taxa (e.g., Lupinus bicolor). The report includes detailed summaries of interactions data as well as a taxonomic review from multiple perspectives.

Contents

Introduction

Data Review

Data review can be a time consuming process, especially when done manually. This review report aims to help facilitate data review of species interaction claims made in datasets registered with Global Biotic Interactions (Poelen, Simons, and Mungall 2014). The review includes summary statistics of, and observations about, the dataset under review:

> University of California Santa Barbara Invertebrate Zoology Collection https://github.com/globalbioticinteractions/ucsb-izc/archive/42a422d5369aafe3a41c6b4d4c9f905fe2a74b2f.zip 2024-02-17T05:53:26.163Z 6d5c8c051a264b07a767631b35d2208dfaac509ab332921a4394d06d5b726af3

For additional metadata related to this dataset, please visit https://github.com/globalbioticinteractions/ucsb-izc and inspect associated metadata files including, but not limited to, *README.md*, *eml.xml*, and/or *globi.json*.

Methods

The review is performed through programmatic scripts that leverage tools like Preston, Elton, Nomer combined with third-party tools like grep, mlr, tail and head.

Table 1: Tools used in this review process

tool name	version
elton	0.13.2
nomer	0.5.6
mlr	6.0.0
pandoc	3.1.6.1

The review process can be described in the form of the script below [1].

```
# get versioned copy of the dataset (size approx. 5.90MiB) under review
elton pull globalbioticinteractions/ucsb-izc

# generate review notes
```

[1] Note that you have to first get the data (e.g., via elton pull globalbioticinteractions/ucsb-izc) before being able to generate reviews (e.g., elton review globalbioticinteractions/ucsb-izc), extract interaction claims (e.g., elton interactions globalbioticinteractions/ucsb-izc), or list taxonomic names (e.g., elton names globalbioticinteractions/ucsb-izc)

```
elton review globalbioticinteractions/ucsb-izc\
 > review.tsv

# export indexed interaction records
elton interactions globalbioticinteractions/ucsb-izc\
 > interactions.tsv

# export names and align them with the Catalogue of Life using Nomer
elton names globalbioticinteractions/ucsb-izc\
 | nomer append col\
 > name-alignment.tsv
```

or visually, in a process diagram.

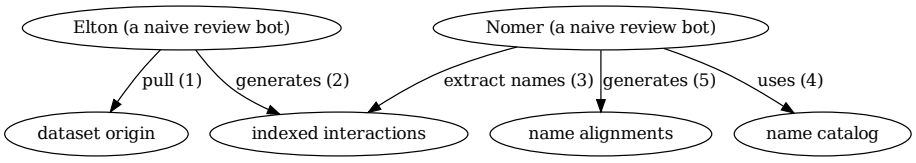

Figure 1: Review Process Overview

You can find a recent copy of the full review script at check-data.sh.

Results

In the following sections, the results of the review are summarized [2]. Then, links to the detailed review reports are provided.

Biotic Interactions

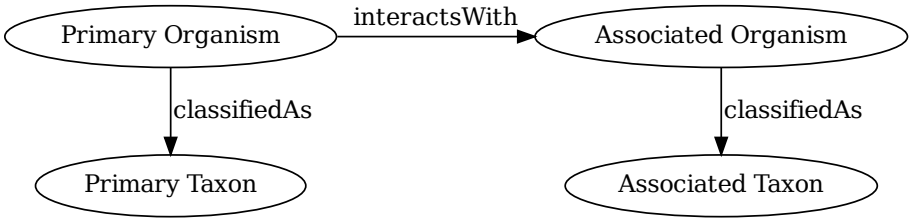

Figure 2: Biotic Interaction Data Model

In this review, biotic interactions (or biotic associations) are modeled as a primary (aka subject, source) organism interacting with an associate (aka object,

[2]Disclaimer: The results in this review should be considered friendly, yet naive, notes from an unsophisticated robot. Please keep that in mind when considering the review results.

target) organism. The dataset under review classified the primary/associate organisms with specific taxa. The primary and associate organisms The kind of interaction is documented as an interaction type.

The dataset under review (aka globalbioticinteractions/ucsb-izc) has size 5.90MiB and contains 1,954 interactions with 8 (e.g., interactsWith) unique types of associations between 379 primary taxa (e.g., Apis mellifera) and 377 associated taxa (e.g., Lupinus bicolor).

An exhaustive list of indexed interaction claims can be found in csv and tsv archives. To facilitate discovery, the first 500 claims available on the html page at indexed-interactions.html.

The exhaustive list was used to create the following data summaries below.

Table 2: Sample of Indexed Interaction Claims

sourceTaxonName	interactionTypeName	targetTaxonName	referenceCitation
Lasioglossum	interactsWith	inside flower of Eschscholzia californica	UCSB-IZC00038170 https://ecdysis.org/collections/individual/index.p
Diptera	adjacentTo	flower of Mimulus auranticus	UCSB-IZC00038207 https://ecdysis.org/collections/individual/index.p
Diptera	adjacentTo	flower of Mimulus auranticus	UCSB-IZC00038169 https://ecdysis.org/collections/individual/index.p
Diptera	adjacentTo	flower of Eschscholzia californica	UCSB-IZC00038248 https://ecdysis.org/collections/individual/index.p

Table 3: Most Frequently Mentioned Interaction Types (up to 20 most frequent)

interactionTypeName	count
interactsWith	1392
adjacentTo	341
visits	150
visitsFlowersOf	44
hasHost	18
hostOf	4
eats	3
coOccursWith	2

Table 4: Most Frequently Mentioned Primary Taxa (up to 20 most frequent)

sourceTaxonName	count
Apis mellifera	105
Linepithema humile	87
Cicadellidae	71
Lasioglossum	61
Solenopsis	46
Augochlorella pomoniella	41
Hemiptera	39
Heleomyzidae	39
Ceratina acantha	35
Temnothorax andrei	35
Bombus vosnesenskii	34
Halictus tripartitus	32
Agapostemon texanus	28
Aphididae	27
Lygus	26
Syrphidae	26
Lasioglossum (Evylaeus)	23
Anthomyiidae	21
Miridae	20

Table 5: Most Frequently Mentioned Associate Taxa (up to 20 most frequent)

targetTaxonName	count
Lupinus bicolor	282
Lupinus nipomensis	99
Quercus agrifolia	96
Venegasia carpesioides	63
Erigonium	59
Arctostaphylos	49
Marrubium	46
Populus	42
Brassica	40
Salix	39
Encelia californica	36
Foeniculum	36
Atriplex lentiformis	32
Cupressus macrocarpa	31
Lupinus succulentus	26
Convolvulus arvensis	26

targetTaxonName	count
Salvia mellifera	24
Heteromeles arbutifolia	23
Astragalus pycnostachyus var. lanosissimus	21

Table 6: Most Frequent Interactions between Primary and Associate Taxa (up to 20 most frequent)

sourceTaxonName	interactionTypeName	targetTaxonName	count
Cicadellidae	interactsWith	Lupinus bicolor	52
Solenopsis	interactsWith	Quercus agrifolia	32
Heleomyzidae	interactsWith	Lupinus bicolor	30
Lasioglossum	interactsWith	Venegasia carpesioides	30
Linepithema humile	interactsWith	Populus	29
Lygus	interactsWith	Lupinus bicolor	24
Linepithema humile	interactsWith	Salix	24
Aphididae	interactsWith	Lupinus bicolor	22
Hemiptera	interactsWith	Lupinus bicolor	21
Apis mellifera	interactsWith	Arctostaphylos	19
Temnothorax andrei	interactsWith	Quercus agrifolia	17
Apis mellifera	interactsWith	Marrubium	16
Closterocoris amoenus	adjacentTo	Lupinus nipomensis	15
Apis mellifera	interactsWith	Erigonium	15
Syrphidae	interactsWith	Lupinus bicolor	14
Andrena principalis	interactsWith	Arctostaphylos	13
Thrips	adjacentTo	Lupinus nipomensis	13
Camponotus clarithorax	interactsWith	Quercus agrifolia	13
Apis mellifera	interactsWith	Brassica	13

Interaction Networks

The figures below provide a graph view on the dataset under review. The first shows a summary network on the kingdom level, and the second shows how interactions on the family level. Note that both network graphs were first aligned taxonomically via the Catalogue of Life. Please refer to the original (or verbatim) taxonomic names for a more original view on the interaction data.

You can download the indexed dataset under review at indexed-interactions.csv. A tab-separated file can be found at indexed-interactions.tsv

Learn more about the structure of this download at GloBI website, by opening a GitHub issue, or by sending an email.

Another way to discover the dataset under review is by searching for it on the

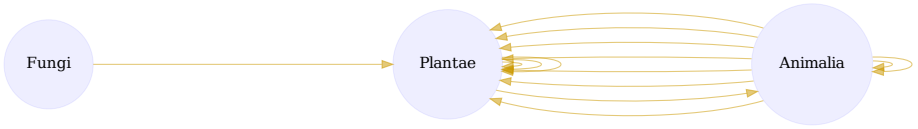

Figure 3: Interactions on taxonomic kingdom rank as interpreted by the Catalogue of Life download svg

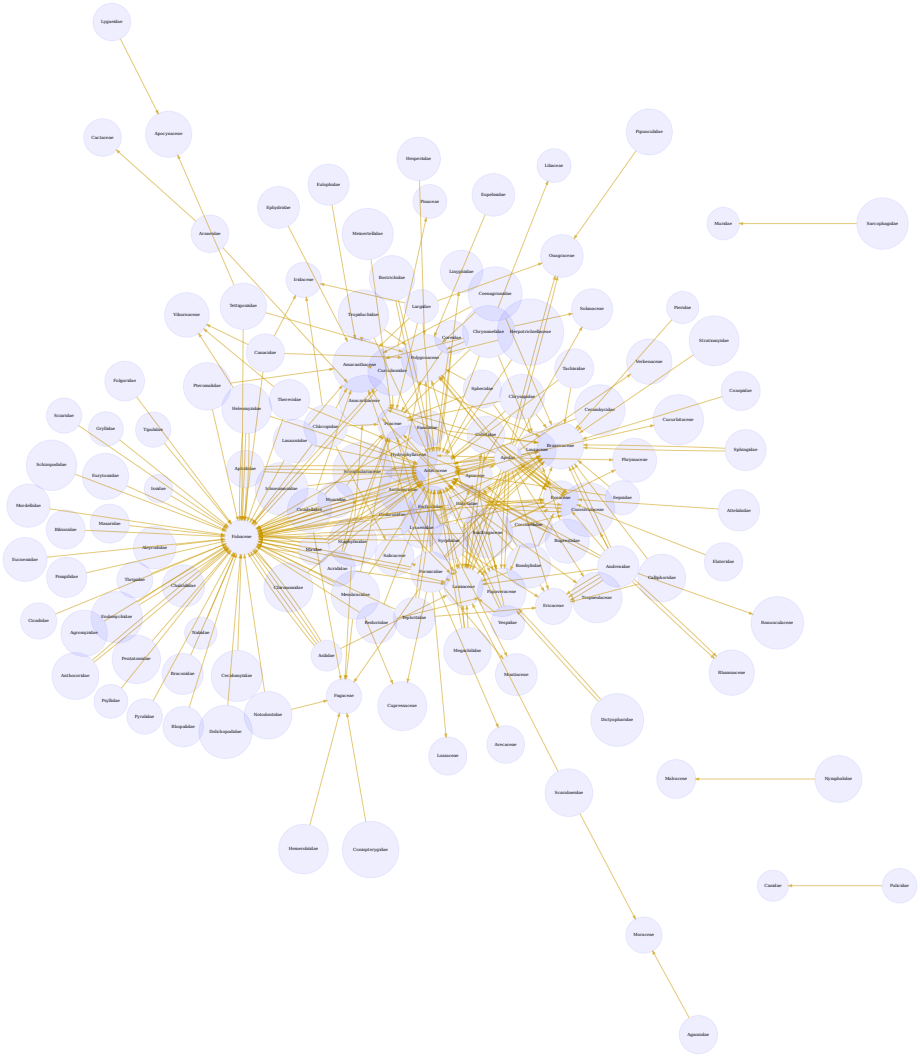

Figure 4: Interactions on the taxonomic family rank as interpreted by the Catalogue of Life. download svg

GloBI website.

Taxonomic Alignment

As part of the review, all names are aligned against various name catalogs (e.g., col ncbi discoverlife gbif itis globi mdd tpt pbdb). These alignments may serve as a way to review name usage or aid in selecting of a suitable taxonomic name resource to use.

Table 7: Sample of Name Alignments

providedName	relationName	resolvedCatalogName	resolvedName
	NONE	ncbi	
	NONE	discoverlife	
	NONE	globi	
Acalyptratae	NONE	col	Acalyptratae

Table 8: Distribution of Taxonomic Ranks of Aligned Names by Catalog. Names that were not aligned with a catalog are counted as NAs. So, the total number of unaligned names for a catalog will be listed in their NA row.

resolvedCatalogName	resolvedRank	count
tpt	NA	705
tpt	species	3
tpt	genus	1
pbdb	NA	498
pbdb	family	104
pbdb	genus	56
pbdb	subfamily	15
pbdb	order	13
pbdb	species	10
pbdb	superfamily	6
pbdb	unranked clade	3
pbdb	suborder	3
pbdb	class	2
pbdb	infraorder	1
pbdb	infraclass	1
pbdb	kingdom	1
ncbi	NA	251
ncbi	species	202
ncbi	genus	110
ncbi	family	104
ncbi	subfamily	16

resolvedCatalogName	resolvedRank	count
ncbi	order	12
ncbi	subgenus	8
ncbi	superfamily	6
ncbi	subspecies	3
ncbi	infraorder	2
ncbi	suborder	2
ncbi	subclass	1
ncbi	varietas	1
ncbi	class	1
ncbi	cohort	1
mdd	NA	709
itis	NA	243
itis	species	217
itis	family	103
itis	genus	99
itis	order	12
itis	subfamily	10
itis	subspecies	10
itis	suborder	6
itis	superfamily	4
itis	infraorder	2
itis	subclass	1
itis	class	1
itis	superorder	1
itis	kingdom	1
globi	NA	437
globi	species	256
globi	genus	207
globi	family	109
globi	subspecies	23
globi	subfamily	18
globi	subgenus	17
globi	order	16
globi	variety	11
globi	suborder	7
globi	superfamily	7
globi	class	4
globi	subclass	2
globi	infraorder	2
globi	phylum	2
globi	subsection	1
globi	infraclass	1
globi	superorder	1
globi	kingdom	1

resolvedCatalogName	resolvedRank	count
gbif	NA	233
gbif	species	229
gbif	genus	127
gbif	family	106
gbif	subspecies	16
gbif	order	12
gbif	variety	5
gbif	class	1
gbif	kingdom	1
discoverlife	NA	648
discoverlife	species	62
col	NA	258
col	species	207
col	genus	110
col	family	103
col	order	12
col	subfamily	9
col	subspecies	8
col	superfamily	4
col	subgenus	3
col	tribe	2
col	subtribe	1
col	class	1
col	variety	1
col	kingdom	1
col	suborder	1

Table 9: Name relationship types per catalog. Name relationship type "NONE" means that a name was not recognized by the associated catalog. "SAME_AS" indicates either a "HAS_ACCEPTED_NAME" or "SYNONYM_OF" name relationship type. We recognize that "SYNONYM_OF" encompasses many types of nomenclatural synonymies (ICZN 1999) (e.g., junior synonym, senior synonyms).

resolvedCatalogName	relationName	count
col	NONE	260
col	HAS_ACCEPTED_NAME	469
col	SYNONYM_OF	62
discoverlife	NONE	663
discoverlife	HAS_ACCEPTED_NAME	59
discoverlife	SYNONYM_OF	16
discoverlife	HOMONYM_OF	2

resolvedCatalogName	relationName	count
gbif	NONE	235
gbif	HAS_ACCEPTED_NAME	596
gbif	SYNONYM_OF	106
globi	NONE	180
globi	SAME_AS	8286
itis	SYNONYM_OF	30
itis	HAS_ACCEPTED_NAME	465
itis	NONE	244
mdd	NONE	723
mdd	HAS_ACCEPTED_NAME	1
ncbi	NONE	250
ncbi	SAME_AS	483
ncbi	SYNONYM_OF	18
ncbi	COMMON_NAME_OF	2
pbdb	HAS_ACCEPTED_NAME	233
pbdb	SYNONYM_OF	15
pbdb	NONE	501
tpt	NONE	719
tpt	HAS_ACCEPTED_NAME	5

Table 10: List of Available Name Alignment Reports

catalog name	alignment results
col	associated names alignments (first 500, full csv/tsv)
ncbi	associated names alignments (first 500, full csv/tsv)
discoverlife	associated names alignments (first 500, full csv/tsv)
gbif	associated names alignments (first 500, full csv/tsv)
itis	associated names alignments (first 500, full csv/tsv)
globi	associated names alignments (first 500, full csv/tsv)
mdd	associated names alignments (first 500, full csv/tsv)
tpt	associated names alignments (first 500, full csv/tsv)
pbdb	associated names alignments (first 500, full csv/tsv)

Additional Reviews

Elton, Nomer, and other tools may have difficulties interpreting existing species interaction datasets. Or, they may misbehave, or otherwise show unexpected behavior. As part of the review process, detailed review notes are kept that document possibly misbehaving, or confused, review bots. An sample of review notes associated with this review can be found below.

Table 11: First few lines in the review notes.

reviewDate	reviewCommentType	reviewComment
2024-02-19T06:57:24Z	note	source taxon name missing: using institution-Code/collectionCode/collectionId/catalogNumber/occur as placeholder
2024-02-19T06:57:25Z	note	found unsupported interaction type with name: [11]
2024-02-19T06:57:25Z	note	found unsupported interaction type with name: [11]
2024-02-19T06:57:25Z	note	found unsupported interaction type with name: [11]

In addtion, you can find the most frequently occurring notes in the table below.

Table 12: Most frequently occurring review notes, if any.

reviewComment	count
source taxon name missing: using institution-Code/collectionCode/collectionId/catalogNumber/occurrenceId as placeholder	11
found unsupported interaction type with name: [11]	3
found unsupported interaction type with name: [Hovering over]	3
found unsupported interaction type with name: [attacks]	1

For addition information on review notes, please have a look at the first 500 Review Notes or the download full csv or tsv archives.

GloBI Review Badge

As part of the review, a review badge is generated. This review badge can be included in webpages to indicate the review status of the dataset under review.

Figure 5: Sample of a GloBI Review Badge [3]

Note that if the badge is green, no review notes were generated. If the badge is yellow, the review bots may need some help with interpreting the species interaction data.

GloBI Index Badge

If the dataset under review has been registered with GloBI, and has been succesfully indexed by GloBI, the GloBI Index Status Badge will turn green. This means that the dataset under review was indexed by GloBI and is available through GloBI services and derived data products.

Figure 6: Sample of a GloBI Index Badge [4]

If you'd like to keep track of reviews or index status of the dataset under review, please visit GloBI's dataset index [5] for badge examples.

Discussion

This review is intended to provide a perspective on the dataset to aid understanding of species interaction claims discovered. However, this review should *not* be considered as fitness of use or other kind of quality assessment. Instead, the review may be used as in indication of the open-ness[6] and FAIRness (Wilkinson et al. 2016; Trekels et al. 2023) of the dataset: in order to perform this review, the data was likely openly available, **F**indable, **A**ccessible, **I**nteroperable and **R**eusable. Currently, this Open-FAIR assessment is qualitative, and with measurement units specified, a more quantitative approach can be implemented.

[3] Up-to-date status of the GloBI Review Badge can be retrieved from the GloBI Review Depot

[4] Up-to-date status of the GloBI Index Badge can be retrieved from GloBI's API

[5] At time of writing (2024-02-19) the version of the GloBI dataset index was available at https://globalbioticinteractions.org/datasets

[6] According to http://opendefinition.org/: "Open data is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike."

Acknowledgements

We thank the many humans that created us and those who created and maintained the data, software and other intellectual resources that were used for producing this review. In addition, we are grateful for the natural resources providing the basis for these human and bot activities.

References

ICZN. 1999. "International Code of Zoological Nomenclature." The International Trust for Zoological Nomenclature, London, UK. https://www.iczn.org/the-code/the-code-online/.

Poelen, Jorrit H., James D. Simons, and Chris J. Mungall. 2014. "Global Biotic Interactions: An Open Infrastructure to Share and Analyze Species-Interaction Datasets." *Ecological Informatics* 24 (November): 148–59. https://doi.org/10.1016/j.ecoinf.2014.08.005.

Trekels, Maarten, Debora Pignatari Drucker, José Augusto Salim, Jeff Ollerton, Jorrit Poelen, Filipi Miranda Soares, Max Rünzel, Muo Kasina, Quentin Groom, and Mariano Devoto. 2023. "WorldFAIR Project (D10.1) Agriculture-related pollinator data standards use cases report." Zenodo. https://doi.org/10.5281/zenodo.8176978.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3 (1). https://doi.org/10.1038/sdata.2016.18.