Weekly Notes - Bat Eco-Interactions Working Group

Meeting zoom: https://deic.zoom.us/j/7588563077
Meeting time: 5pm CET / 11am ET / 8am PT
Whimsical board for brainstorming:
https://whimsical.com/biodiversity-knowledge-hub-WyqgfhD1V4rMg5W7d6CnYY
GloBI Project page - https://globalbioticinteractions.org/gbatnet

18 December 2024

Present: Aja, Connor, Nancy, Quintin, DeeAnn, Jelle, Ariadna, Kendra, Anna, Jorrit (joined late), Cullen, Donat

- Aja update on IUCN 200+ taxonomic change recommendations
    - (Nancy) submitted files need to be in exact format- IUCN strict on this
    - Submitting to general group for broad taxonomic shifts, not order-specific groups
        - https://www.iucnredlist.org/assessment/authorities
    - Also providing detailed guidance to bat specialist group for once taxonomic changes are implemented
    - Need to get contact info from Dave
    -

11 December 2024

Present: Cecilia, Nancy, Jelle, Donat, Aja, Jorrit, Hernani, Ariadna, Iroro, DeeAnn, Connor, Kendra, Anna

Agenda:
1. Australian Bat Symposium The 20th International Bat Research Conference (Aja)
2. BigBatDB update (Jorrit, Ariadna)

Ariadna is planning to go to https://www.ausbats.org.au/ibrc2025.html 3 - 8 August, 2025 – Cairns, Australia and is asking about some email.

Cecilia was contacted by organisers of a symposium inviting her as a speaker for a digital databases symposium. She forwarded email to Aja and Nate who are better suited to submit some titles for potential talks on behalf of Eco-interactions groups.

Ariadna on the BigBatDB - we decided to collect dental formulas and index it by Encyclopedia of Life (EOL) - (step 1) transcribe dental formula in a Google Sheet https://docs.google.com/spreadsheets/d/1xGPXuwEuqEao2ezya7kW2DWixp4pejqVEjKFD1sqG9g/edit?gid=0#gid=0 . Ariadna created a google form to help facilitate data entry https://docs.google.com/forms/d/1w4_ttyMr5r1cfQQ-XqYBGZ-B5NEgapyqoaMsEVSQWaY/edit . Ariadna is planning to take the form responses and reformat them into the dental formula suited for EOL indexing and name alignment.

Hernani mentioned a reptile trait database that may be a source of inspiration. http://www.reptile-database.org/

A link directly to a copy of Koopman
https://linker.bio/hash://md5/70c3e0fba7379e09e95a38569fe29da7

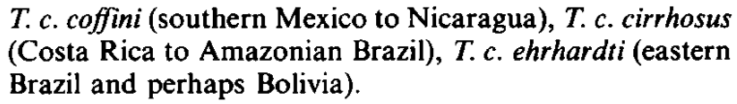
same, but with page 77
https://linker.bio/pdf:hash://md5/70c3e0fba7379e09e95a38569fe29da7!/p77 renders a single page

Similar derived pdf for koopman page 77-80 ranges.
https://linker.bio/pdf:hash://md5/70c3e0fba7379e09e95a38569fe29da7!/p77-p80

Cecilia / Hernani suggests including the date and locality information as a separate field to help with the name alignment.

Arianda - how to best capture the locality information

Anna - suggest to include verbatim (what is in the text, e.g., "southern Mexico to Nicaragua"))



T. c. coffini (southern Mexico to Nicaragua), *T. c. cirrhosus* (Costa Rica to Amazonian Brazil), *T. c. ehrhardti* (eastern Brazil and perhaps Bolivia).

Anna - do keep track of the basisOfRecord for the dental formula

https://github.com/BigBatDB/bigbatdb/issues - contains all the issues

https://github.com/BigBatDB/bigbatdb/issues/5 - Katja Schulz (of EOL) suggested to include Lifestage - Nancy says that (for bats) are only for adults due to their specific dental development, different from other mammals.

Discussion of extracting trait data using OCR or similar. No script for doing this from Koopman 1994 exists yet. There is structured data from the HMW series https://jhpoelen.nl/hmw/ (Jorrit: this work was extracted from already semi-structured data)

What do we get out of doing this inside BigBatDB? Idea is to build an expandable database that can be extended into more data, instead of a single spreadsheet that can be forgotten. This is the test case for BigBatDB.

Name alignment. Jorrit explains the data format for BigBatDB releases. Standardized (Darwin Core) format allows doing name alignment using standard tools.
github.com/bigbatDB/bigbatdb-name-alignment
Allows aligning with a variety of other sources. Run a GitHub Action that produces an alignment report summarizing the results.

Aja: Taxonomic alignment for Bat1K.

4 December 2024

Present: Cecilia (left early), Nancy, Nate, Jelle, Donat, Aja, Jorrit, Hernani, Ariadna, Cullen, Iroro

Aja action items for IUCN-MDD matchup

- https://docs.google.com/spreadsheets/d/164fcpl7i36MC4ZUEIrzd2Zj06fiQndlBO9k-wlSX4aE/edit?gid=204177465#gid=204177465
- To enable the IUCN update for bats
- Linking to BatLit at the same time
    - Good way to ensure that BatLit is complete
    - Can do this for the BTA more generally
- Wanting to mark down the per species assessment date + change date in the literature
- Will follow up further in a separate call

Cecilia wants titles for potential talks for a symposium at IBRC in Australia

- https://www.ausbats.org.au/ibrc2025.html
- 3 - 8 August, 2025 – Cairns, Australia
- suggests to come up with a title / abstract
    - Jorrit suggests: "IUCN Conservation as a driver to preserve and develop digital research infrastructure."
    - Pragmatic, practical means of implementing now
- Email notice:
    - "I am putting together a symposia proposal tiled “Digital resources for bat researchers and managers” and was hoping I could include you or someone in your group to talk about your online resources.
    - The symposia is designed to advertise and discuss issues around developing these sorts of tools and databases"

- Could plan to have the BatLit be more advanced
    - Should apply to everyone – who uses literature
    - Live tools, with BiodiversityPMC
        - Data organized, links to publications
        - Not just ideas, but tools
- The Taxonomy work has a smaller user group and interest base

IUCN reassessment list

- 200+ species that are valid but not in the IUCN list
- Geographic ranges are not matched to these new lists – we do have some MDD maps, but not precise – could still be used
- Body size
- In 5 years or so we could do a much better job at this
- Then can look at predictive power to greater extent
- Iroro has a paper looking at bat extinction risk
    - Using IUCN data as current

Linking the BTA to BatLit more generally → taxonomic alignment tool, how could we use it for GBIF / CoL, which is more directly used for research

- Much more interested in alternative taxonomies at GBIF now
- Publications – but then taxonomy from the publications
- May need to get more specific when say “GBIF”
    - Not just one thing
    - Put datasets in one place
    - Tools are getting better so we can zoom in more
    - GBIF is accumulating data – not producing data – data model used to be only occurrences, but now it is broader
        - iDigBio, DiSSCo are doing lots of digitization, but don’t care about the linking as much
        - Name system is how we do the linking
- How to use BTA to link datasets?
    - Can implement this in Nomer perhaps
    - Coming up with a super simple use case – that is the best place to get started
- We can document how Aja is doing this
- Nomer
    - A name alignment tool
    - https://github.com/globalbioticinteractions/nomer
    - Could be key for doing the name-based linking part of this
    - Use a versioned copy of the BTA for this – well-defined sandbox to test/play
- Name- vs concept-based matching
    - Taxonomic Data Objects for Communicating the Meaning of Species Names
    - Nathan S. Upham, Jorrit Poelen
    - https://doi.org/10.3897/biss.8.139413

Example

- Trachops cirrhosus
- https://synospecies.plazi.org/#Trachops+cirrhosus
- We could track this through manually through the BTA
    - From Jorrit to Ariadna - please send me the row in the BigBatDB you'd like to align (the name) to BTA

Jorrit wonders about our current strategy to coordinate activities of Bat1k, Big Bat DB with projects like BatLit, BTA, GloBI, Plazi's BLR, MDD etc.

20 November 2024

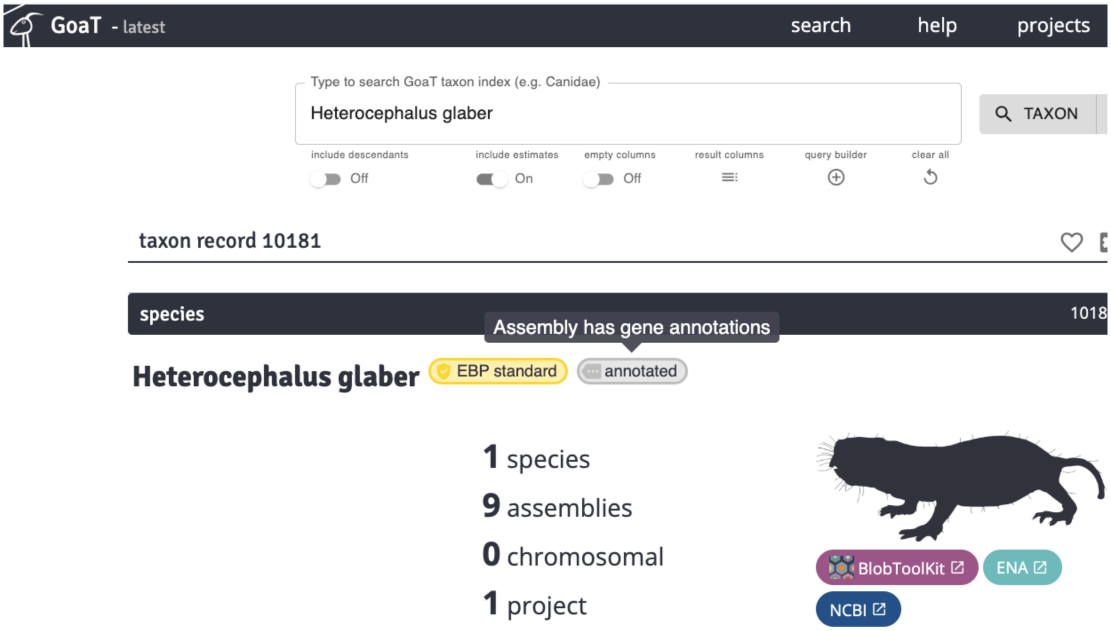
Present: Kendra, Cibele, Cecilia, Nancy, Nate, DeeAnn, Anna, Jelle, Donat, Connor, Iroro

Cibele joining from Darwin Tree of Life / Wellcome Sanger Institute

- Part of time is on the Earth Biogenome Project
- Genomes on a Tree (GoaT)
    - https://goat.genomehubs.org/
    - Large data system that underlies
    - Runs on NCBI taxonomy (as backbone)
    - Stores lots of data related to genome assemblies
        - Chromosome number, genome size
        - Inferred ploidy
    - Runs on Bioproject IDs
    - Working closely with NCBI taxonomy curators
        - Now seem willing to change some aspects of taxonomy
        - But needs to be done at the *record* level rather than wholesale taxonomic changes
            - Would result in issues of taxonomic misIDs
    - ToL IDs
        - Known to be a problematic issue
        - Not robust to taxonomic change
        - Moving to align these with NCBI taxon IDs
            - Works as well as the NCBI taxonomy does
        - Article on using the TaxIDs – https://wellcomeopenresearch.org/articles/9-591
            - Developing a pipeline for submitting batch changes to the taxonomy of a given set of records
            - Could work for individual sequence records also
    - Similar to issues encountering with the IUCN

Linking to taxonomy → but also want to link to taxonomic treatments

- Plazi linking to the genomes themselves: http://plazi.org/treatmentbank/
- Existing links with GBIF
- ENA – european NCBI
- Assigning a name to a genome – are referring to one concept of a given name – should be explicit about this >>
    - On the one hand a small thing – but becomes much larger when talk about AI linking of biodiversity knowledge – need explicit links of this specimen / genome is a representative of taxon A according to author B
- Potential to do for all EBP – but starting with the bats
    - BatLit project
    - BatLit on Zenodo: https://zenodo.org/communities/batlit/records?q=&l=list&p=1&s=10
- https://biotxplorer.sibils.org/?species1=Pholidota&size=1
- Exploring taxonomic names via Synospecies, a triple store based on taxonomic names in Catalogue of Life and TreatmentBank. See eg [Rhinolophus sinicus]()
- Possible meeting to be arranged with Checklistbank to discuss their approach to alternative taxonomies and how this could be combined with the Bat-eco group members taxonomic names subprojects?
- CSC Post meeting thoughts: Ways GoaT could immediately help:
    - Workflow 1: a)The working group could double check bat names and ncbi taxIDs that have genomes. b) Associate specific BioSample used to produce the genome with most relevant treatments. c) If there is a validation for used name = ncbi taxID = specific treatment, goat can add a validation tag "valid Name" or "Taxonomy verified" in the record's page. (like the tags in the image).



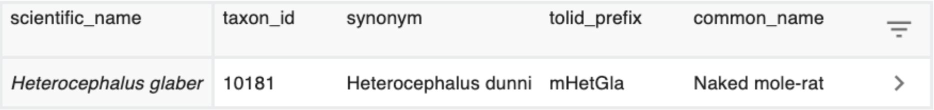
Type to search GoaT taxon index (e.g. Canidae): Heterocephalus glaber

include descendants: Off
include estimates: On
empty columns: Off

taxon record 10181

Heterocephalus glaber

1 species
9 assemblies
0 chromosomal
1 project

  - Create a new column for the “number of treatments” that would contain a link-out to the equivalent search result for that name, like in the *Rhinolophus_sinicus* example. That would at least inform the user that disambiguation might be needed.
  - Current columns under “names” that can already be filled if needed:
  - 

scientific_name	taxon_id	synonym	tolid_prefix	common_name	
Heterocephalus glaber	10181	Heterocephalus dunni	mHetGla	Naked mole-rat	>

  - GoaT can produce a list of names not found on NCBI taxonomy that can be used to prioritize the names Bat1K needs to request from ENA/NCBI. This is publicly available in our S3 store, but currently outdated: Not found | spellcheck suggested. Next update scheduled for Jan25.

13 November 2024

Present: Aja, Quentin, Cecilia, Nancy, Nate, Connor, Donat, Meike, Kendra, Anna, Ariadna
Big bat database

- https://docs.google.com/spreadsheets/d/1gtENmce4tv2AJWduWrj6dKnDFE0O-tLNZVAG6BAn5r4/edit?gid=0#gid=0
- How to keep track of changes to taxonomy in the context of traits
  - Record locality, date
- Focus on primary versus secondary literature
- Donat: https://zenodo.org/records/13505746 or https://doi.org/10.5281/zenodo.13505746 for the publication
  - Also include the BatLit link to the Zenodo record – allows for linking to the extraction of trait via machine learning – subsequently
  -

Similar to dynamic of the Bat1K data – how to keep/organize the metadata

- Meike – tracking the metadata / sample coordination
- All genomes derived from only one organism
- Already collecting lots of metadata
  - Geographic location, not all with photos
- They use https://goat.genomehubs.org/ – GoaT
  - Uses NCBI taxonomy likely – could have more taxonomic data maybe from MDD?
- Quentin: legacy of mess of thinking that have a genome of species not a specimen
- Backwards data curation is what’s happening
  - Important to talk to the people who collected the specimens
  - Collector of the tissues / submitter needs to be the one to link the taxonomic treatment to the sample
- Need for a *review / perspective* article to clarify what types of metadata should be associated with a given new genome sequence
  - We could work on this as a group

  - MIXS: https://github.com/GenomicsStandardsConsortium/mixs
    - MIxS: Minimum Information about any (X) Sequence
  - https://www.gensc.org/ – Genomic Standards Consortium
  - Data nerds already know about this
    - But what's needed is a broader paper written in plain language
  - Science Forum: The critical importance of vouchers in genomics
    - https://elifesciences.org/articles/68264

- 3 letter codes for genome names

- Current input form for Bat1K: https://docs.google.com/spreadsheets/d/10cz-E3aLc1vA_IqYFgQQKPdCdQwIdZTt/edit?usp=sharing&ouid=113418377508345956042&rtpof=true&sd=true

From Jorrit:
"We are currently visiting extended family in Totnes, Devon, UK: in my absence, I hope that you'll give Meike Mai and Cibele Ciao of Bat1K a warm welcome (should they decide to attend) and chat about to ways to collaborate and help establish bidirectional links between https://bat1k.com genomic information with MDD/BatNames as well as BalLit and possibly other projects. For some ideas, please see https://github.com/jhpoelen/bat1k-dwca and https://github.com/jhpoelen/bat1k-talk-2024-10-27 . Enjoy!"

Linking sequences to other data including the voucher
- https://biodiversityknowledgehub.eu/wp-content/uploads/2024/07/ENA-Cross-Reference-Search-Factsheet.pdf
- https://www.ebi.ac.uk/ena/browser/xref
- https://ena-docs.readthedocs.io/en/latest/retrieval/programmatic-access/cross-reference.html

6 November 2024

Present: Aja, Jorrit, Quentin, Cullen, DeeAnn, Cecilia, Nancy

Jorrit - Bat1k - Nancy on the steering committee

Bat1k -[: https://bat1k.com/about/ refer

Bat1k -[:uses names from]-> MDD/ BatNames

Bat1k (genomes)
MDD / Batnames / BTA (taxonomic stuff)
BigBatDB (traits)
BatLit (literature)
BatBase/GloBI (interactions)

How do our activities connect to other related initiatives like: GBatNet working groups.

https://jhpoelen.nl/bat1k-talk-2024-10-27/#/title-slide

slides sources and data https://github.com/jhpoelen/bat1k-talk-2024-10-27

Jorrit explained the Bat1K interactions he had at NASBR and the scripts (Preston, Nomer) he used to compare names in Bat1K with MDD (see above). Cullen suggested including name and location. It was clarified that more data will be collected to update the dataset with needed location and other metadata, like how the data were obtained. Next steps to turn metadata into archive so it can be indexed by GBIF to see what assemblies are in Bat1K without exposing them directly. Nancy said there is a lot of complexity with assemblies which is why they are using Globus data storage, including annotations. Time is an issue because they have a bandwidth problem while they are preparing the paper. Jorrit is seeing a lot of connections between MDD/BatNames and how he can help them link to other places and make the links more explicit. Nancy said they are now dealing with the issue that bat names will change overtime and these need to be included in their dataset. Jorrit planted the idea of data repackaging and wants to know what's next: Jorrit and Nate are on the Bat1K slack, invited Bat1K to come to one of our meetings…..

Aja pointed out that using automation to compare name lists have been problematic and it would be good to explore how we can improve it. Jorrit recommended adding BTA into the Nomer search (?). DeeAnn wants to use BTA in a paper. Group discussed publishing short paper on the tool. Nancy said we could publish in Biodiversity Data Journal and then DeeAnn could cite it, etc. We could use a few examples (Pteronotus, Miniopterus).

Note that https://github.com/globalbioticinteractions/nomer/issues/157 already documented a desire to integrate BTA in Nomer. DeeAnn's list would be make for a good test case to implement this automation.

Next step - create diagrams to help show connection between the different projects

30 October 2024

Present: Aja, Jorrit, Nate, Connor, Hernani, Jelle, Kendra, Anna, Quentin, Jorrit, Anna, Cullen, Donat

Nate netted bats with Nancy near Tempe in a *Macrotus californicus* (insectivore, see also https://www.mammaldiversity.org/taxon/1004950) .

Anna – bats that are eating hornets in Georgia

- Invasive hornets (*Vespa velutina*)
- https://www.globalbioticinteractions.org/?interactionType=interactsWith&sourceTaxon=Vespa%20velatina&targetTaxon=Chiroptera - no indexed interactions in GloBI Bat <> Vespa
-

Gary McCracken (https://eeb.utk.edu/people/gary-mccracken/) interaction at the airport

Asked where certain papers were in BatLit – he's writing a book right now, so close to access for references

- Idea to have all references freely available within the BatLit repo
- Broader access to BatLit
    - Few people requesting to become members, but not a ton
    - Select folks that we can slowly assemble

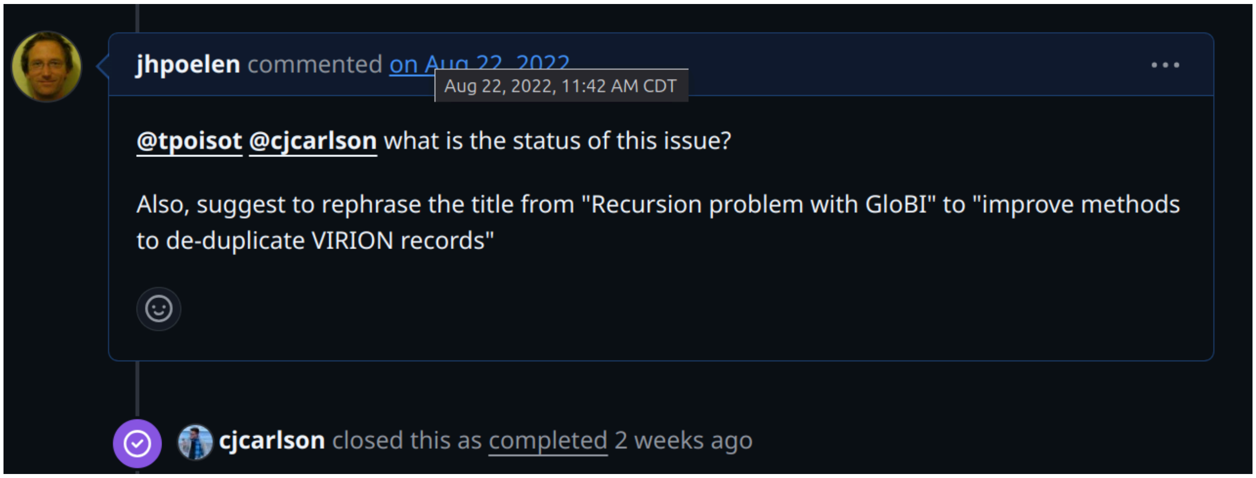
Kendra is working on a paper with Iroro on a trait paper for all IUCN assessed species

- Can integrate that effort with the Big Bat Database

BigBatDB v0.1 triggered ideas on structuring data by Katja Schulz of Encyclopedia of Life - https://github.com/jhpoelen/bigbatdb/issues/1

GloBI <> Virion - https://github.com/globalbioticinteractions/globalbioticinteractions/issues/665
https://github.com/viralemergence/virion/issues/67

as of 2024-10-30 - (with weeks ago being 2024-10-17)
https://github.com/viralemergence/virion/issues/67#issuecomment-1222619458


jhpoelen commented on Aug 22, 2022
Aug 22, 2022, 11:42 AM CDT

@tpoisot @cjcarlson what is the status of this issue?

Also, suggest to rephrase the title from "Recursion problem with GloBI" to "improve methods to de-duplicate VIRION records"

cjcarlson closed this as completed 2 weeks ago

Lots and lots of databases in Verena: https://www.viralemergence.org/data

"**October 17, 2024**: The GLOBI dataset has been de-indexed from VIRION, due to a growing number of concerns about the reliability of text-mined data. This represents an important departure from the pipeline described in the *mBio* publication. The current pipeline is described at the top of this README." – bottom of https://github.com/viralemergence/virion
- Very strange
- GloBI does not focus solely on text-mined data
- It indexes existing datasets see https://globalbioticinteractions.org/datasets .
- suggestion to update GloBI's description/issues at https://github.com/viralemergence/virion/issues/93

BHL and the internet archive going down
- Relative to Plazi converting BHL articles into Biodiversity PMC – current approach is using a professional vendor to do the scan and OCR – then Plazi is extracting from that
- Very few people support / use BHL (Jorrit: could you be more specific?)
- But it is clearly essential to the taxonomic enterprise
- BHL operates like a library – just digitizing whole library catalogs, but some of it is less valuable (Jorrit: less valuable because ...?)
- How does one really cite BHL? Tends to just be mentioned in the methods –
- For mammals
    - Have opportunity to get all the original descriptions
    - And then get all the modern revisions
    - And then we step through all of this – showing the value of reconstructing the conceptual and nomenclatural changes
    - Anna *did* this for ringtails
        - Citation method doubles the number of papers in the reviews
        - https://docs.google.com/spreadsheets/d/1smvZJh4V2Av_CVD55evzYBeUWI7eyGsevEYMkdeUvQ8/edit?gid=2112296708#gid=2112296708
- BatLit

    - Has the value of showing how you turn a big collection of literature – into something that you can compute with
    - Count all the words that start with A, or something more relevant
    - Aja: goal to make sure that the abstracts are included in the metadata

Two members of the Bat Eco-Interactions group won awards at NASBR

COP16 in Colombia – NASBR folks were there
- https://www.science.org/doi/10.1126/science.adu0409
- https://www.science.org/doi/10.1126/science.adg0344

Jorrit: I am thinking about ways to have batlit <> bat1k <> MDD <> BTA linked through literature records and taxonomic names, genome sequences. Hoping to discuss next time.
- Mos def

23 October 2024

Present:
- In person! Aja, Jorrit, Cecilia, Cullen, Nancy, Ariadna, Nate, Connor, Hernani, DeeAnn (left early?)
- Online: Jelle, Kendra, Anna, Quentin

How to discuss the BatLit service while at NASBR
- Legal framework
    - Anything that is 2000 and before is open access
    - Anyone can jump into that by going to batlit.org
        - You can help curate
- Password access can be given to certain researchers
    - They can reach out to info@batbase.org
    - Contribute and use the repo
        - How do we decide this?
    - Communities of practice
        - We want to establish this
        -
- Reach out to people
    - Ask for theses, gray literature, reports – stuff that can actually disappear
    - Anna: theses in Mexican public university library could be a nice start.
        - This is a list of repositories I've put together:
            - https://docs.google.com/spreadsheets/d/1k2Ob-OKEqaZogg5NDv8vEpfFe1fNofQhOv4bejR7Kss/edit?usp=sharing

        - UNAM has 568+ bat theses https://repositorio.unam.mx/contenidos?&q=murcielago
            - Creative Commons BY-NC-ND 4.0 Internacional
                - https://creativecommons.org/licenses/by-nc-nd/4.0/deed.en
- Membership at BatLit?
    - CV too much?
    - Tiered system, multiple ways to engage
    - A lot is already open
    - Example for liberal membership from UGA's Center for Ecology of Infectious Diseases
        - https://www.ceid.uga.edu/about/information-for-members/

- Co-roosting talk
    - Extrapolation of the 300 publications to the expected full diversity of bat species interactions → 5200 total publications on co-roosting, many of which don't exist
    - AlterNate shared some slides of how to use the co-roosting archive
        - Showing how we would extract these data
        - Comparing the ML data extraction aspects with the rule-based vs GPT contexts – validating – human in the loop too
        - Important to build in checks to the workflow
            - >> Education
            - Important to have an *intention* to understand the technology at every stage
            - Make sure that have off-ramps, backups
    - Anna: can rerun summary scripts from many years ago: Output: https://covid19taf.slack.com/files/U01P3UH96KA/F07TKAR1W81/status-of-lit-extractions.pdf
    - Script: https://covid19taf.slack.com/files/U01P3UH96KA/F07SS48S6KZ/status-of-lit-extractions.rmd

16 October 2024

Present:
Jelle, Donat, Aja, Kendra, Jorrit, Cecilia, Cullen, Nancy, Anna, Ariadna, Quentin

Data review issues: it always needs more people to review and "clean" datasets

NASBR

- Where to share the files: GoogleDrive
- Our symposium is Friday morning

- Cullen will let everyone know where to meet before the actual symposium, may be for breakfast or another time before.

IUCN

- IUCN Species Specialist Group
    - Bat
    - Bat oneHealth
    - Redlist group
    - Taxonomic Backbone - Jorrit: is the IUCN taxonomic backbone openly available? Who is managing it? Should I make an effect to add IUCN taxonomic backbone to the Nomer Corpus of Taxonomic Resources? Answer: The Redlist is the taxonomic backbone and is available online.

Aja's presentation:

- Bat taxonomy under control by Nancy; there are no processes in place; IUCN is sometimes reassessing the taxa in the Redlist. An effort is on the way to create a pipeline to update names in the IUCN backbone taxonomy.
- IUCN versioned?
    - Each assessment has a date, which allows to get back
    - Yes, but the back-log isn't readily available

Reference local species/databases:

- https://enciclovida.mx/especies/33942-anoura-geoffroyi
- More complete language translations

AI perspectives

How to continue the discussion with Nate and Patrick

- llearn more about the black boxes
- Skin the cat

Proposal by Nate

1.1. Data Schema Design
2.1. OCR (Optical Character Recognition)
2.2. Output Structuring
2.3. Establish Performance Metrics

"Draft a grant" to progress?
e.g.
https://new.nsf.gov/funding/opportunities/innovation-infrastructure-innovation-biological-research

- First define the research question that you are trying to answer (not that you could imagine that could be answered - this because answering needs a commitment later, once a system is built)
    - Go back to the original question driving the groups:

        - outbreak of a pandemic
        - Prediction of a pandemic, using interactions, geography, behavior, roosting near human populations, reservoirs
        - Disease risk
    - What are the databases needed for above?
    - Build a system to get data out publications
    - Data integration
    - Create a data format that allows reuse of data extracted from publications

Next meeting:
Same time, just for some of use in the same room at NASBR

9 October 2024

Present: Aja, Nancy, Donat, Jelle, Nate, Cullen, Jorrit, Kendra, Ariadna, Anna, Cecilia, Nate L, Patrick Ruch, Quentin

Focus today on text-mining
- Patrick Ruch: https://www.sib.swiss/patrick-ruch-group
- Nate Layman: https://www.ecohealthalliance.org/personnel/dr-nathan-layman

BatLit community in Zenodo >> staging area for text mining

BiotXplorer, SIB approach to data extraction and interaction labeling
- https://biotxplorer.sibils.org/
- Two species and an interaction type
- Brings the passages from the text with it, and the PMID
- How good is the tool?
    - Did two benchmarks – relative to 100 manually curated triplets and 100 triplets from GloBI
        - GloBI data is about accessing existing species datasets, from very heterogenous sources – citing hundreds of datasets that are cited
            - All datasets in GloBI: https://www.globalbioticinteractions.org/datasets
            - ML-derived dataset: https://depot.globalbioticinteractions.org/reviews/EOL/pseudonitzchia/
        - Result: BiotXplorer has a lower accuracy with the manually curated data, as compared to the GloBI data (with known provenance)
        - Could scale this result to 100k triplets rather than 100, and then GloBI would be the most relevant
        - Positive side: all datasets in this comparison have known provenance
    - https://github.com/globalbioticinteractions/globalbioticinteractions/issues/556
- Using the ROBI relationship ontology

    - Interaction types – synonyms of those interactions – some improved, but some increased false positives
- Next step: looking at the agents of the interaction – species / taxonomy – but the *intersection* of the taxonomic category and the interaction type
- Overall
    - Uses a traditional NLP approach – sentence splitting – regex matching – stemming of words / fuzzy matching – expecting 74% precision currently
    - Classifiers rather than LLMs – binary classification
    - Sensitivity – not estimating this currently – would require a human to go through a 20 page paper and list all the interactions

- Tokenizing is the first step in a lot of LLMs – but BiotXplorer is not using this currently (using regex searching)
    - Are using BERT – and tokenizing at that level
    - Tokens:
        - Might be able to access these through the paywall >> but could also yield full sentences that would still be protected

- Would ideally be looking at the *network* aspect of how that is changing based on tweaks in the model – could measure differences
    - Adding visualization as well

- LLM approach – has advantage that the interaction could be 3 sentences later – not all info just in one sentence

Comparing across methods

- Possible to compare / contrast the strengths and weaknesses of these different methods? Should be yes
    - >> How combine, ideally
- Could use the different methods to check / verify each other
    - One approach is more traceable than another – scalability differs too – also with human-in-the-loop as well
- https://github.com/globalbioticinteractions/globalbioticinteractions/issues/1021
- https://en.wikipedia.org/wiki/Blackboard_system
- At some point we will be able to use fully read / tagged papers for use in calculating both sensitivity and precision
    - E.g., the rodent-virus interaction data from ASU / Beckett Sterners group
-

Broader

- you may be interested to https://github.com/big-bee-network/bif - we are compiling a giant list of interaction terms and their RO mapping from ~10-20 millions interaction candidate claims mined from GBIF indexed datasets. - see https://github.com/Big-Bee-Network/bif/issues/6

NASBR

- We can share our NASBR presentations for team edits using this shared file; https://drive.google.com/drive/folders/1nwD1eafKZFdc82I9H4kVnzVPLNxVBV7B
    - Can't wait to see everyone in real life
- Planning to have the Oct 23rd meeting in person

2 October 2024

Present: Aja, Nancy, Jelle, Nate, Cullen, Donat, Nathan Layman, Jorrit, Kendrai, Anna, Ariadna, Cecilia

Cullen: BatLit access to the DB

- need some onboarding process for getting into the Zenodo "batlit" group to get access to the restricted publications.

- Parallel with BatBase – everyone is approved for someone that can search the DB
- Nancy - we need a document for folks to agree with before they become a member. To make sure to do legal
- Terms of Agreement / Terms of Use
    - Example from the World Spider Catalog we could build from
    - Not easy to change the terms, although perhaps possible
- We made bridge between Zotero and Zenodo – and opened up large amounts of literature
    - We want to share this sooner rather than later
- 20k records in BatLit
    - >> this version 0.6 is the one that will be discussed from the

Aja working on IUCN bat RedList updates

- Wanting to host a taxonomy call to discuss in more detail

Nice search box, Jorrit! https://batlit.org/

- four lines of html code, thanks to Zenodo -
    - from https://github.com/bat-literature/bat-literature.github.io/blob/f43d595e61eb95f03f56bd598f2b60ed53791956/README.md?plain=1#L1

```
<form class="searchbox" action="https://zenodo.org/communities/batlit">
  <input type="search" id="name" name="q" placeholder="Search BatLit..."/>
  <button type="submit">Search</button>
</form>
```

NABR presentation guide - https://nasbr.org/presentations24

tentative schedule
https://docs.google.com/spreadsheets/d/1h-pFB6et0ryGUaKfkZDrpC_CCNpxYartpd8fI5TVxuc/edit?gid=0#gid=0

Paperpile (Ariadna), Zotero (Aja), Endnote (Cullen), Custom software (Jelle)

Nomen nescio. 2024. https://api.openai.com/v1/ accessed at 2024-09-27.

25 September 2024

Present: Aja, Nancy, Jelle, DeeAnn, Nate, Kendra, Cullen, Donat, Jorrit, Nate Layman (alter-nate, EHA), Hernani, Anna

Introductions

- Nate Layman has done work on transferable / transmissible vaccines in wild animals – degree in evolutionary biology – worked with Scott Nuismer at UIdaho as a postdoc – https://imci.uidaho.edu/all-of-our-people/name/nathan-layman/
- Kendra now a research associate at AMNH

ML-based extraction of data from the bat co-roosting dataset

- Already started working some with Aja on this
- World Organization for Animal Health – the ML work is derived from that in a prior EHA contract >> goal was to estimate the number of cases / outbreaks that they missed
    - At the time ChatGPT / LLMs were developing
    - Grabbed 100k abstracts, fed to GPT, asked it to read the abstracts and pull out all the outbreaks of given viruses in given years
        - Similar in several ways the approach we are using
- Slides on ecological data modeling
    - Aiming for tabular data as the output from data extraction
    - Function calling schema
        - How you prompt ChatGPT – using a javascript function for generating prompts – ask for return to be tabular / structured
        - Respond with co-roosting interactions + page number + snippet of text
        - Context allows for still having the human check it
        - Structured response engineering
            - This is critical for downstream usage
    - How to evaluate performance?
        - Ranking metric – accuracy, specificity, precision, recall
        - Computer vision can now extract data from figures
    - Work in progress for sure, lots of errors still
        - Now working on fine tuning – domain specific knowledge
        - Want to be able to re-use the pipeline – bat-plant interactions
- Jorrit: What data was used to train ChatGPT version you used?
    - Likely no
    - Entire internet used
    - ChatGPT is a black box – doesn't cite sources
    - Answers today could be different than answers tomorrow
    - Approaches to this
        - Ask multiple times – see if gives same result
        - Include humans in the workflow >> reducing work, not eliminating
    - Is ChatGPT repeatable and suitable for academic research?
        - Don't know who created, what data went in, paid service,
    - (Jorrit notes) Lama https://research.facebook.com/publications/llama-open-and-efficient-foundation-language-models/ opens the model, but **not** the data. ChatGPT offers to interact with the model in a pay-to-play "cloud" model, but not the model itself. Both, do not clearly define what data corpus (Jorrit: likely the provenance of the data is undefined because the LLaMA/ChatGPT folks don't know how or have no interest in citing their data sources, likely both) is used to train the models.
        - We could endeavor to build our own LLM / fine-tuned model for ecological data extraction
            - But not a trivial exercise
            - Need to scope how much work would indeed be involved

- Could bring in Patrick Ruch’s group to discuss with Nate Layman – compare approaches – the accuracy metrics are a central part of this
    - https://www.sib.swiss/patrick-ruch-group
    - https://sibils.text-analytics.ch/search/
    - https://biotxplorer.sibils.org/
    - https://biotxplorer.sibils.org/?species1=Rhinolophus&interaction=co-roosts%20with&size=1 here is an example

- Co-roosting of bat species >> couldn’t a rule-based NLP model be used? (Anna)
    - Context could be missed – LLM will provide more nuance to the output (presumably), flexibility of LLM is the power – but NLP
    - Compare to BiotXplorer approach

- Live coding of pipeline in R Studio

18 September 2024

Present: Aja, Nancy, Jelle, DeeAnn, Nate, Kendra, Cullen, Quentin

NASBR plans for 23-26 Oct 20204
- Nancy already making her talk
    - Introducing the full symposium
    - We can send Nancy content, aim to set the context – to help set up the talks
- 5 weeks until the first day!
-

EcoHealth Alliance
- Kendra no longer works there, lots of people are leaving
- Wow, because have no money, no scientists
    - NIH is barring them from receiving funding
- World OneHealth Alliance in South Africa
    - Some EHA folks are still going
- Organization is folding
- Many folks looking for new jobs
    - Incl Kendra, directly contacting people

Aja – batLit update – de-duplicating 30k records that were found
- No more empty PDFs
- Near ready, will send to Jorrit >> previously had the Trash file indexed – now cleaned up
    - V0.5 of the Bat Lit Corpus
- Aim is for this all to “go live” while at NASBR

        - Need to think more about the governance structure – what are folks agreeing to with respect to becoming members / users of the repository
    - Search features
        - Could just keep with Zenodo for now
        - Then advocate for a step-wise approach of using Google Scholar – then to BatLit to find the full text
    - Versioning
        - Many reasons to want to be able to cite
        - E.g., “cited v3.2 of Bat Lit for doing my text mining routine, etc”
        - We want this to live on beyond this group of curators though too
        - Zenodo – each PDF has it’s own page, but then they come together as a corpus (Zorpus?)
            - https://batlit.org/
        - Came to a consensus that we think as a group that versioning is important, including content-based versioning

IUCN is the next issue coming

- Aja will report on next week

EHA generated a pipeline for doing LLMs with the co-roosting dataset

- Nate Lehman at EHA
    - Still works there for now
    - https://github.com/n8layman/ehallm
    - Already looking at the data, can pull out different aspects of the dataset
    - Using OpenAlex also – for finding papers
- Aiming to understand what can be automated more easily >> for NASBR talk

Symposium planning

- Abstracts: https://docs.google.com/document/d/1TFPWLNxf5KrczCNGQzybiru93AwRtbx7I6NNWVgRcbY/edit
- Having a meeting at Weds Oct 23

11 September 2024

Present: Aja, Nancy, Donat, Jelle,
How can we get most efficiently more colleagues be included

Strategies on how to deduplicate new paper collections
Jelle is looking into how to deduplicate the existing zotero file

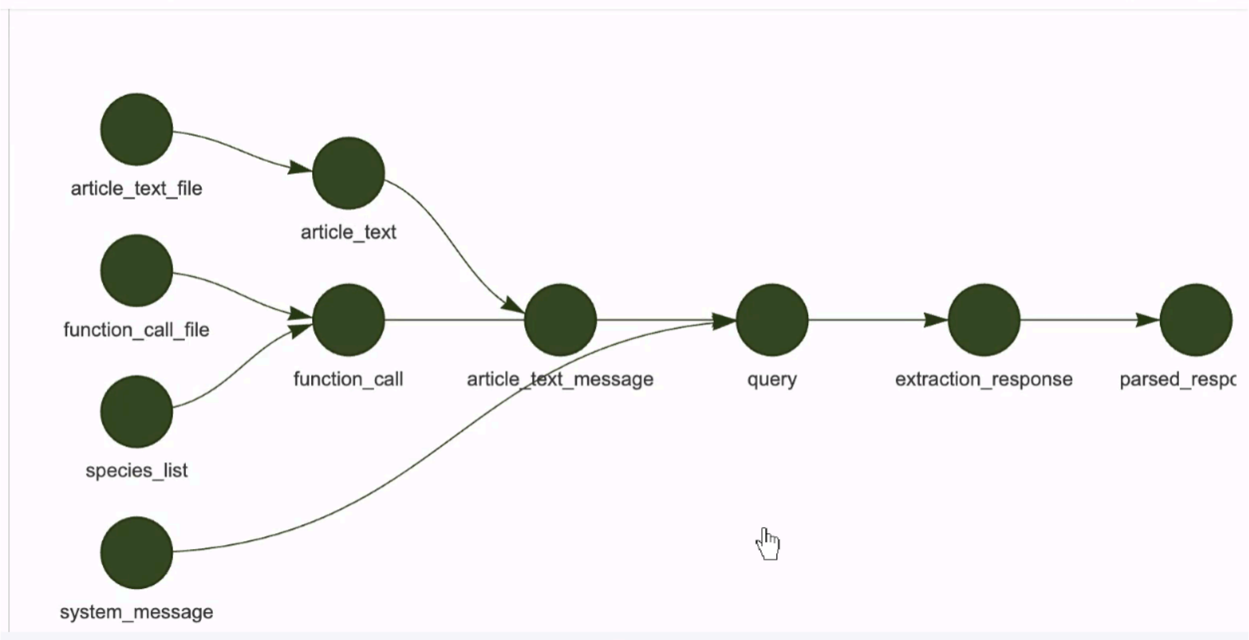
We need to include a column in Zotero that includes the Zenodo ID of the uploaded files.

Each Zotero that has been uploaded to Zenodo should be kept when dedupliacting

Question arises what to do when an old bad PDF is replaced by a better one?

Value of openAlex for next generation of bat literature, and how much back is openAlex helpful to find old bat publications

Co-roosting: new tool by Nate Leman (?) https://github.com/ecohealthalliance/ehallm
announced by Aja, each query running at 8 cts.



This could complement the development in biodiversityPMC

https://biotxplorer.text-analytics.ch/?species1=Pholidota&size=1

Add here is the list of terms used in biotexplorer for biotic interactions. Additional terms, or synonyms can be added.

4 September 2024

Present: Aja, Nancy, Jorrit, Kendra, Cullen, Donat, Jelle
Agenda:
Soft launch of the BatLit Corpus
Issues arising

BatLit -

Jorrit asks what to do with publications without a publication date
https://github.com/bat-literature/bat-literature.github.io/issues/27 .

Soft launch result can be: https://zenodo.org/communities/batlit .

Open Question:

Donat: Should we keep versions of BatLit in both Zenodo and Zotero?

Cullen: For starters, we'll consider the Zenodo records be "readonly" and the Zotero records are going to be read-write.

Simplified Workflow

Aja curates records in Zotero **-[:versioned_and_translated_into]*** -> Zenodo BatLit community

https://zenodo.org/records/13538164 - example of restricted batlit publication

https://zenodo.org/records/13538144 - example of open batlit publication

"Amazon" BatLit package consists of three parts:
1. Packaging slip (describing the contents and their origin, e.g., I saw pdf x at location y on date z)
2. Zotero Metadata (structure data as retrieved through the Zotero Web API, for example, see https://linker.bio/cut:hash://md5/03118e04a34ea876a0b48c511cd24a8d!/b300600-303154)
3. PDFs download from the restricted Zotero Group using proper credentials.

Only 1. and 2. are open access, and are published on https://github.com/bat-literature/bat-literature.github.io .

3. are stored offline and in restricted member access only locations. So, they are *not* open available.

On releasing a future BatLit version, the updated (meta)data would be added to the existing BatLit Archive and a new version "hash" would be created to help updated the Zenodo records in a systematic and controlled manner.

Zotero -> BatLit "Amazon" Package -[:translated] -> Zenodo speak -[:update existing or create new] - > Zenodo

Next Steps: Aja to divvy up the work to correct the publication dates.

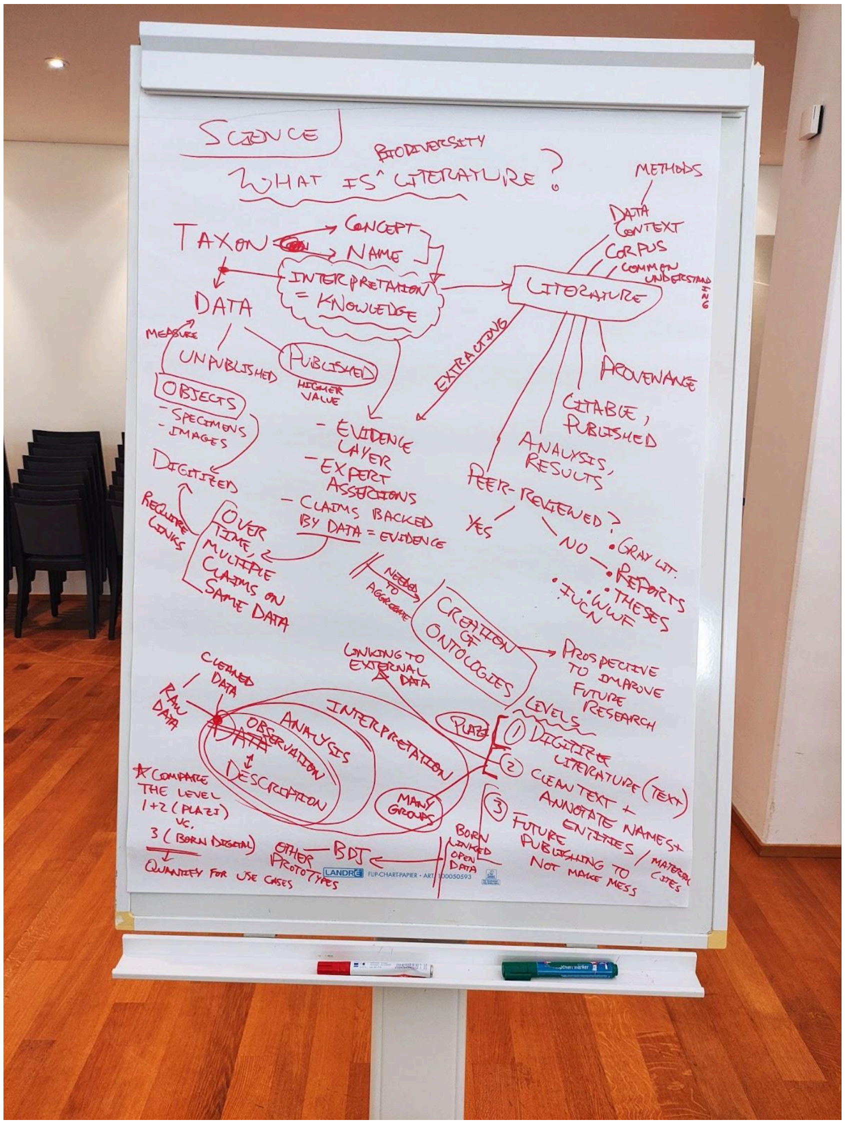
28 Aug 2024

Present: Aja, Nancy, Ariadna, Nate, Jorrit, Kendra, Cullen, Donat

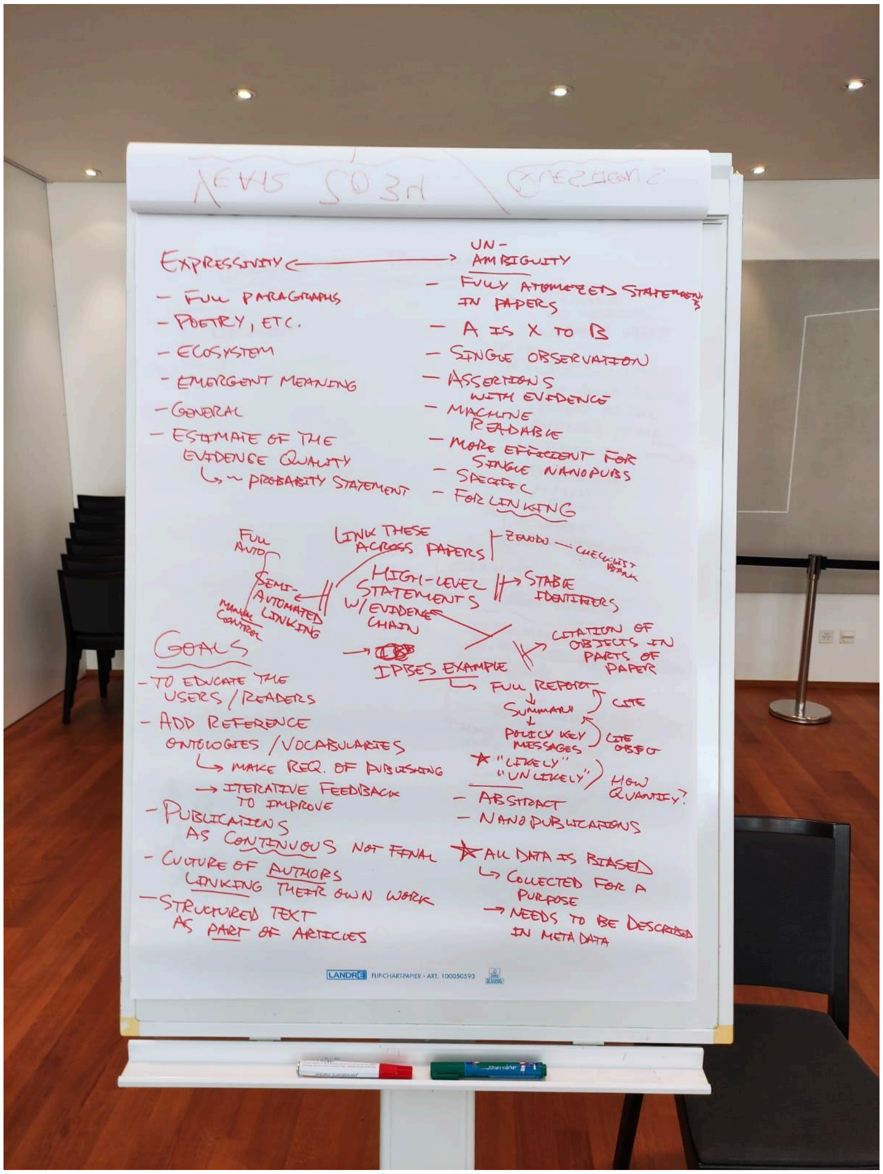
- Year of the Bat 2026
    - Bane or Blessing? Reviewing Cultural Values of Bats across the Asia-Pacific Region https://journals.sagepub.com/doi/full/10.2993/0278-0771-41.1.18

- Agenda
    - ? BatLit v0.5 release
    - Updates on the forthcoming 'Distentis Roadmap' that stem from the meeting Donat organized last week (the BatLit project was viewed as an excellent use case)
    - BatLit data reviews, governance
    - Issues arising
- shared by Kendra following Aja's announcement around "2026: Year of the Bat" - cultural significance of Low, M.-R., Hoong, W. Z., Shen, Z., Murugavel, B., Mariner, N., Paguntalan, L. M., Tanalgo, K., Aung, M. M., Sheherazade, ., Bansa, L. A., Sritongchuay, T., Preble, J. H., & Aziz, S. A. (2021). Bane or Blessing? Reviewing Cultural Values of Bats across the Asia-Pacific Region. Journal of Ethnobiology, 41(1), 18-34. https://doi.org/10.2993/0278-0771-41.1.18

- Summary from the Disentis meeting
    - 10-year plan for the forthcoming path of liberating global biodiversity knowledge from literature
    - Day 1 - Group 1 (Scientific) Brainstorming Session

- 
SCIENCE
WHAT IS BIODIVERSITY LITERATURE?
TAXON
CONCEPT
NAME
INTERPRETATION = KNOWLEDGE
DATA
MEASURE
UNPUBLISHED
PUBLISHED
HIGHER VALUE
OBJECTS
- SPECIMENS
- IMAGES
DIGITIZED
REQUIRE LINKS
OVER TIME, MULTIPLE CLAIMS ON SAME DATA
- EVIDENCE LAYER
- EXPERT ASSERTIONS
- CLAIMS BACKED BY DATA = EVIDENCE
LITERATURE
METHODS
DATA CONTEXT
CORPUS
COMMON UNDERSTANDING
EXTRACTING
PROVENANCE
CITABLE, PUBLISHED
ANALYSIS, RESULTS
PEER-REVIEWED?
YES
NO
• GRAY LIT.
• REPORTS
• THESES
• WWF
• IUCN
NEEDED TO AGGREGATE
CREATION OF ONTOLOGIES
PROSPECTIVE TO IMPROVE FUTURE RESEARCH
LINKING TO EXTERNAL DATA
CLEANED DATA
RAW DATA
OBSERVATION
~~DATA~~
DESCRIPTION
ANALYSIS
INTERPRETATION
PLAZI
MANY GROUPS
LEVELS
① DIGITIZE LITERATURE (TEXT)
② CLEAN TEXT + ANNOTATE NAMES + ENTITIES / MATERIAL CITES
③ FUTURE PUBLISHING TO NOT MAKE MESS
BORN LINKED OPEN DATA
BDJ
OTHER PROTOTYPES
★ COMPARE THE LEVEL 1+2 (PLAZI) VS. 3 (BORN DIGITAL)
QUANTIFY FOR USE CASES
LANDRÉ FLIP-CHART-PAPIER · ART. 100050593

- Day 2

- 

EXPRESSIVITY ⟷ UN-AMBIGUITY

- FULL PARAGRAPHS
- POETRY, ETC.
- ECOSYSTEM
- EMERGENT MEANING
- GENERAL
- ESTIMATE OF THE EVIDENCE QUALITY
  ↳ ~ PROBABITY STATEMENT

- FULLY ATOMIZED STATEMENTS IN PAPERS
- A IS X TO B
- SINGLE OBSERVATION
- ASSERTIONS WITH EVIDENCE
- MACHINE READABLE
- MORE EFFICIENT FOR SINGLE NANOPUBS
- SPECIFIC
- FOR LINKING

FULL AUTO

SEMI-AUTOMATED LINKING

MANUAL CONTROL

LINK THESE ACROSS PAPERS

ZENODO — CHECKLIST BANK

HIGH-LEVEL STATEMENTS W/ EVIDENCE CHAIN

→ STABLE IDENTIFIERS

→ CITATION OF OBJECTS IN PARTS OF PAPER

→ IPBES EXAMPLE

↳ FULL REPORT

↓

SUMMARY

↓

POLICY KEY MESSAGES

CITE

CITE OBJECT

★ "LIKELY" "UNLIKELY"

HOW QUANTIFY?

- ABSTRACT
- NANOPUBLICATIONS

★ ALL DATA IS BIASED

↳ COLLECTED FOR A PURPOSE

→ NEEDS TO BE DESCRIBED IN META DATA

GOALS

- TO EDUCATE THE USERS / READERS
- ADD REFERENCE ONTOLOGIES / VOCABULARIES
  ↳ MAKE REQ. OF PUBLISHING
  → ITERATIVE FEEDBACK TO IMPROVE
- PUBLICATIONS AS CONTINUOUS NOT FINAL
- CULTURE OF AUTHORS LINKING THEIR OWN WORK
- STRUCTURED TEXT AS PART OF ARTICLES

- Learned intuition from the use case of actually doing this work
  - Plazi process of extracting from literature
    - Initial work starting in 2002
    - Old papers – harder because they don't template well – start from the bibliographies
  - Zenodo repos
    - Storage and metadata
  - Biodiversity PMC
    - For use in additional annotation
  - Zotero use for gathering bibliographic metadata
  - GBIF folks integrating these data
    - Material citations
  - Data Futures

        - Peter Cornwall
        - “Green Field” tool for extracting citation metadata from publications
            - Just a prototype now – last year of Arcadia funding
- Main question:
    - Where do we invest?
        - Opening up as many new journals as possible
        - Use cases like BatLit
        - Working on old literature and the metadata of this
    - Likely depends on where / how the money comes from
        - Use cases can drive some of this funding and innovation
    - Ultimately want to open up 500 million pages
        - No one best way to do this

BatLit v0.5

- Data publication near ready to go
- Nate and Ariadna reviewed the administrative / machine-readable part of the BatLit
    - The digital concrete == the software architecture part of it
- https://github.com/bat-literature/bat-literature.github.io/issues/23
- Sandbox: https://sandbox.zenodo.org/communities/batlit-review-md5-26f7ce5dd404e33c6570edd4ba250d20/records?q=&l=list&p=1&s=10
    - 17 items included as a pilot
    - Linker.bio output is coming from server in Germany – and/or Zenodo –
    - But there is some core dependency on Jorrit knowing where these things are – talk about how to keep these communities alive and sustainable
        - Bat Lit community, others
    - Through Zenodo
        - We cannot make everything fully downloadable – but can can cite things that are closed access in a way that maintains the privacy but still confirm that the dataset is complete – hiding portions of it
    - Some concern about the keeping of extra copies outside of Zenodo
        - Does the server in Germany count as an extra copy?

21 Aug 2024

Present: Aja, Hernani, DeeAnn, Nancy, Ariadna, Nate, Jorrit, Anna

- Kendra and Aja will review Nate Layman’s recommendation for a workflow for our co-roosting dataset autom
- BatLit v0.5 was released https://batlit.org - request for review available at https://github.com/bat-literature/bat-literature.github.io/issues/23

- IUCN name matching for bats
    - Relative to the MDD, and Batnames
    - https://docs.google.com/spreadsheets/d/164fcpl7i36MC4ZUEIrzd2Zj06fiQndlBO9k-wlSX4aE/edit?gid=0#gid=0
    - With the purpose of updating the IUCN taxonomy for bats
    - (Jorrit) Which version of BTA did you use? How'd you like IUCN to cite this BTA version?
        - (jorrit) Still not sure which version of BTA is used for this IUCN comparison.
            - was it "Sherman, A., Geiselman, C., Poelen, J., Simmons, N., Reeder, D., Upham, N., Phelps, K., & Agosti, D. (2024). A Standardized Review of Bat Names Across Multiple Taxonomic Authorities [Data set]. Zenodo. (https://doi.org/10.5281/zenodo.11193643)" ?
    - Which MDD version? Aiming for the v1.13 MDD version, published Jul 13, 2024
    - IUCN has their own system
        - What we are sending – just the google sheet
        - Informally we'll send
        - But then that IUCN-to-MDD alignment would not get published
            - We COULD publish this on Zenodo or similar – why not do so?
    - What is IUCN's process?
        - Different sub authorities are in charge of assessments
        - But they cannot assess anything until the backbone taxonomy in the IUCN is changed
            - But that is a major bottleneck – someone has to notice the change, fill out a spreadsheet, then someone accepts, and then it is possible for a reassessment to occur
            - Solution - make more batch-level changes to the IUCN backbone taxonomy
    - What we need is a formal data publication that can be iteratively sent to the IUCN >> and to cite the metadata resources involved in generating this
        - Hash-based versioning
        -

14 Aug 2024

Present: Jelle, Aja, Hernani, Kendra, Cullen, Jorrit, Donat, DeeAnn, Nancy, Nate

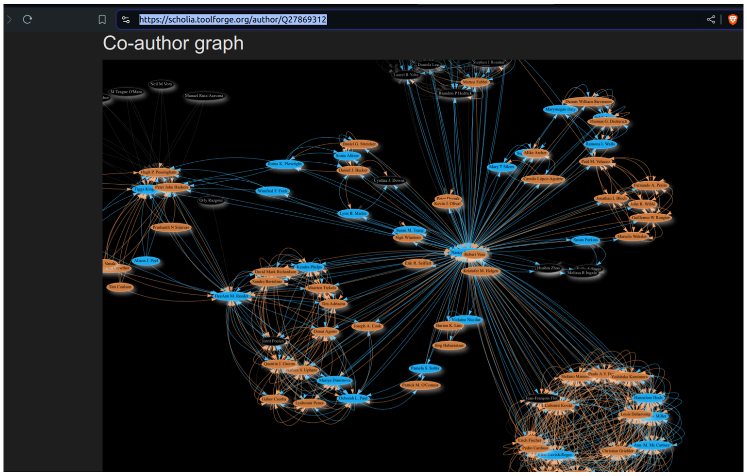
- Kendra and Aja will review Nate Layman's recommendation for a workflow for our co-roosting dataset automation (https://github.com/ecohealthalliance/ehallm)

- Jorrit’s recommendation to start a new journal “BioKnowDar”: https://pad.carpentries.org/bioknowdar
    - Challenges of siloed projects
    - Datasets are always coming online – how to keep aware of
    - Digital scholarship initiatives
        - How to deal with tenure review boards?
        - Rigor of digital publications need to be acknowledged
        - Tension about academic credit
            - BUT tools for data review will also make it easier to write ‘traditional’ papers
            - Bot can publish naive reviews, but then humans can come along and write
    - Peer review is big within this
        - Is what you are producing being checked by a peer review?
        - Dynamic digital publication – with a peer review system
        - But peer review is also a big time and
    - Academic credit models
        - Versioning
        - Elaborate schemes of contribution half-life
    - Hernani has an in-progress MS along similar lines:
        - https://docs.google.com/document/d/1FSsG2GQoclEsBX__f4RTU8_YI3x1GqXauG7cWTWrUFM/edit
        - “This is what we have been drafting so far…I was going to send an invitation to everyone when things were more structured”

BatLit – governance doc:
https://docs.google.com/document/d/1kuDALvMIfQrIh9mhHnzOdsEm88_Xjzvl/edit

- We are doing a 5 min brain dump regarding some of the highest value use cases for the project
- Accessibility
    - PDFs are difficult to access in the Global South, especially older literature, even smaller schools in the US
    - Instead of all doing this independently, we pool resources and create a “friends” community around the world
- Finding complete metadata for known papers
    - More accurate, faster, better tools, especially for old literature
    - Abbreviated journal names
    - Authors names with one or two initial
        - Author indexes as well – e.g., Nancy Simmons in Hesperomys: https://hesperomys.com/h/3643
        - Helpful for determining if these are the same individual

        - Scholia also does this: https://scholia.toolforge.org/author/Q27869312



(Nancy's co-author graph according to wikidata visualized via Scholia, note the cluster of co-authors associated with the GBatNet/Covid19 group).

        - ORCID is trying to do this – so is WikiData
- Broader goal of BatLit – trying to find + connect *all* the publications about bats, ever
    - Because of that, we should not allow for mission creep there
    - Cannot solve all the world's problems there
- Other databases can cite the BatLit papers
    - "Provide a durable literature resource for others to link to and find!"
    - Stop re-inventing the wheel in this context
    - Centralization – but in the sense of "let's use the same data and methods to refer to the same literature"
        - PDF repo that everyone can access and cite
        - Some risks of centralization too – if a few people control this, then those folks gain outsized power, and the power can lead to negative consequences
    - What we want is a place to start collecting these data

Need for a "terms of use" on the BatLit

- Example from the World Spider Catalog – basically saying that cannot replicate the entire data and serve elsewhere
- BUT – from a technical perspective, actually we DO want the ability to translate the entire corpus
- We want to show that the corpus can be re-used – more than just being "open" in terms of "allowed to download"
    - Ideally – have batches of 5k pdfs available for download, etc – but for own personal purposes

Broader thing – how can we ultimately unite these different literature repos?

-

7 Aug 2024

Present: Nancy, Aja, Jelle, Cullen, Hernani, Jorrit, Kendra, Nate, Donat, Anna

- Plantar fasciitis
    - Approaches – good shoes

- BatLit Corpus Governance Doc:
    - Reviewing this as a group, Cullen initially drafted
    - Steps
        - Zotero for organizing the PDFs and Metadata
        - Versioned archives whenever Aja puts up the BatSignal
        - Zenodo then reflects – the Zenodo is just an archiving service
            - If we make edits there, then they do not feed back to Zotero
    - Who can upload and from where?
        - Currently need to upload to Zotero then Zenodo reflects that

    - UI for exploring the PDFs, searching
        - Not going to be Zenodo
            - But Zenodo does have a search feature
        - Could be the batlit.org interface
            - Requires a relational database for the backend
            - If using Zenodo only as the file repository
        - Use the Zenodo web API
            - Re-use this API to create own webApp wrapper
            - Can use the web URL to be writing those queries
            - Build an index

    - Different projects
        - Curatorial workflow for creating / uploading PDFs to the BatLit
        - Web UI for users to explore the batLit Corpus

    - Hernani: Could Arctos be a potential good solution for this issue? https://arctosdb.org/about/
        - Anna: paid service that operates from member fees

    - Jelle: built a custom search for Hesperomys that searches through PDF text, e.g. http://hesperomys.com/search?q=trachops+coffini . Definitely could be a lot better but as Jorrit says it wouldn't be too hard to build something similar for BatLit.

    - BiodiversityPMC
        - https://sibils.text-analytics.ch/search/
        - Based on MedLine / Plazi TreatmentBank / Supplemental data
        - Accessibility?
            - Can you provide access to closed pubs

    - Game: how to use the minimal effort to make these resources accessible?
        - Re-use Jelle's search tools? Uses Amazon web-services to host currently
            - But this does not require the PDFs to be on Zenodo – so they are somewhat less open/stable
            - Costs money to keep this running
        - BiodiversityPMC
            - Is this a separate upload? Or directly from Zenodo – ideally this happens automatically, but not yet in place

    - What are the requirements for a particular use case?
        - Search a name – get PDFs for those
            - Well defined, finish that, then add more functionality after that
        - Next small problem

    - Each of us can think of use cases >> can think of commonalities – for next time

    - Parallel project with TaxoDros
        - They are taking a similar approach – Jorrit took PDFs and metadata that was custom – decoded – similar workflow as with BatLit – box it, export to Zenodo – they built a web application at https://www.taxodros.uzh.ch/search/lit.php for the search

    - Here's another example from Donat: World Spider Catalogue https://wsc.nmbe.ch/
        - Full community – requires a signup – they have a governance document – have to agree to not reproduce and sell the information
        - This is awesome, and would be a great target for the BatLit Corpus

- BatLit v0.4 release
    - Up to 20k papers now – does this include duplicates too?
    - https://batlit.org/
        - More attachments (29k) than references (20k) – can have multiple files per reference, sometimes supplemental materials
        - Metadata travels together with the PDF
    - Data review
        - Invites feedback, data errors can be caught

- Others
    -

31 Jul 2024

Present: Jorrit, Jelle, Aja, Nate, Hernani, Cullen

IUCN updates relative to bat taxonomy

- Adding inform from http://hesperomys.com/ – Jelle Zijlstra is joining
- Raised issue that Small Mammal Specialist group could also benefit from this type of update

1-page document for the GBatNet

- Drafted based on the images taken at the meeting
- Then collected feedback
- Start based on what they had emailed
- Focus on genome-based metadata standards
    - Will write this up
    - Eye on enabling future taxonomic changes for the published genomes

Jelle is working on importing classifications into Hesperomys

- Imported the HMW files
- Aligned with the MDD taxonomy as well
- And relative to IUCN
    - Aja had been doing that somewhat manually = Bat Taxonomic Alignment / BTA
    - She was also adding alignments to taxonomic meaning – considering the geographic locations / range maps to some extent
    - Now wanting to review all names again
        - Jelle helped with this matchup + the links to the new records
        - IUCN will have the original + current links to the info
- IUCN in general
    - They do not allow or practice versioning
    - They lack funding in a way
    - How could we better version these? Using the GloBI approach
        - Upload older versions of the taxonomy + range maps
    - Citation hygiene brings a lot of these issues to the front

        - IUCN has lots of these older data – but they are broadly unwilling to share it / feel it is outdated and not useful

BatLit work
- Aja is having to manually update metadata – due to some duplicates
    - Hard to know that have duplicates until enter the metadata
    - Some OCR is misidentifying authors
    - What can we do here
        - One solution: Tracking the metadata for each specific PDF
            - Treat the raw PDFs as one publication
            - Then next publication is v1.0 of the metadata on those PDFs
                - Versioned copies of Zotero basically
            - Helps understand how our knowledge of the metadata on these PDFs has changed through time
    - Can we help Aja compare the journal names / info relative to what Zotero is extracting from the PDFs
        - Use the DOIs → CrossRef as a check
        - Check which years the journal was published

Raw PDFs → automated annotations (e.g., OCR, DOI extraction) → manual annotations

About 500 away from 20k PDFs ingested

24 Jul 2024

Present: Jorrit, Anna, Nancy, Ariadna, Aja, Cullen, Connor, Kendra, Harnani

Review changes for GBatNet IUCN
https://docs.google.com/document/d/1au0_bSgZMn5b43qxKb0XEZYJca2PMeHF50pYLKLoPi0/edit
Project: Updating Bat Taxonomy for IUCN Red List

Participating Networks and Working Groups:
- Bat Eco-Interactions Working Group
- Global Bat Taxonomy Working Group
- IUCN SSC Bat Specialist Group

Rationale:
The IUCN Red List has not kept pace with current bat taxonomy, meaning that many currently recognized species are not listed and therefore their conservation status has not been appropriately assessed. Using the Bat Taxonomic Alignment Tool, we propose to provide an updated species list and associated data – and build a pipeline for future updates – to facilitate the inclusion of new and revised taxa in the IUCN assessment process.

Alignment with GBatNet Strategic Directions:

- 3.3 Help maintain accurate species assessments
- 1.6 Assist networks in prioritizing strategic actions

Approach:
The taxonomic boundaries of species and higher-level taxa frequently change as knowledge of population structure and evolutionary relationships improves, and older hypotheses are reframed. We have collected and aligned treatments across primary taxonomic authorities for Chiroptera to create a single dynamic and versioned name translation tool. The Bat Taxonomic Alignment (BTA) tool integrates data from three of the most utilized treatments: The Handbook of the Mammals of the World Volume 9 Bats (HMW), Bat Species of the World: A Taxonomic and Geographic Database (Batnames database), and Mammal Diversity Database (MDD). The BTA reveals that the IUCN Red List (version 2022) lacked over 180 bat species currently accepted by MDD and Batnames, along with over 120 superseded epithets. The BTA allows easy identification of missing species and will be used to propose updates to the Red List to facilitate more accurate species status assessments. We will align the IUCN Red List species list using the BTA to identify missing or outdated treatments and will determine the nature of the discrepancy (split, merge, new species described, etc.), the alternative names associated with the current binomial, and any journal publications supporting the change. This information will be provided to the IUCN in the format required by them for taxonomic updates to their lists. Once IUCN has ingested the new names, species can be prioritized for updated assessments. This project will benefit all GBatNet networks and working groups by ensuring the globally recognized list of species of conservation concern is accurate and useful.

Key targets and timeline:

- Obtain most recent list of bat species recognized by IUCN from Dave Waldien (done)
- Align Red List species with BTA and identify gaps (done)
- Obtain IUCN forms to propose species updates (waiting for response from Dave Waldien)
- Prepare IUCN forms for new species and submit to IUCN (Fall 2024)
- Prioritize species assessment updates through the Bat Specialist Group (Spring 2025)

Key Contacts: Aja Sherman aja@batbase.org; Nancy Simmons simmons@amnh.org; Ricardo Moratelli rimoratelli@gmail.com; Connor Burgin connorjburgin@gmail.com; Julie Shapiro julie.teresa.shapiro@gmail.com; Nate Upham nathan.upham@asu.edu; Dave Waldien dwaldien@gmail.com

Nancy: There are no regular updates of the IUCN taxonomy. No mechanism is place across the IUCN taxonomic or bats.

Conner: In IUCN non-bat groups there have been "well done" (e.g., primate group).

other topics beyond the GBatNet project one pager -

1. How to keep track of all the activities / outcomes of this group (https://globalbioticinteractions.org/gbatnet)?

GBatNet has a quarterly activity reporting form - required by National Science Foundation (NSF) Nancy to share the form.

2. Cullen - batlit corpus update - we need some kind of governance document to share how this will live on in perpetuity.
3. Aja - > 300 hours uploading pdfs etc. lots of duplicates. Many of them were abstracts, deduplicated ~1/3 replaced. ~30s to process a pdf for automated process, with manual updates processing takes longer (2 minutes). Estimated release BatLit v0.4 https://batlit.org .
Aja priority is to get the basic metadata out (author names, date, publication, title).
Anna - Is the literature corpus addition protocol written down somewhere? Can I help review? Suggest to setup a curation decision tree.

Connor - sent intro email to Aja with cc to Jelle Zijlstra. Jelle keeps a collection of pdfs related to taxonomies.

Aja goes over the batlit pdf addition protocol. Google drive -> download as zip -> import into Zotero

Jorrit: did you consider rclone ? This tool allows for access / syncing data in various "cloud" storage solutions, including google drive. I've been using it to sync files in dropbox .

Anna - suggests to create a journal reference lists .

Jorrit - suggests to have a dedicated review meeting when BatLit v0.4 has been released.

17 Jul 2024

Present: Hernani, Jorrit, Anna, Nancy, Donat, Ariadna, Cecilia

- Jorrit working with Brazilians on World Fair project: https://www.globalbioticinteractions.org/worldfair/

Recap of last week

- EHA funding changes
- BatNet domain name https://batlit.org
- BTA use cases https://jhpoelen.nl/bat-taxonomic-alignment/
    - People don’t know where to look
        - Frequently Nancy can point to sources in 5 min
    - Will be putting together BTA tutorial videos
    - How to tract taxonomic changes? Eg phylogeny
        - Quentin has worked on this
        - Synospecies
    - Other taxa communities
- Tools can be hosted by GBAT net as long-term commitment for storage and dissemination.
    - Eurobat traits Shiny: https://jasja.shinyapps.io/ClimBats/
- publications vs. publish website - articles appear to be more static, less likely to include video or dynamic elements / links.
- Nancy

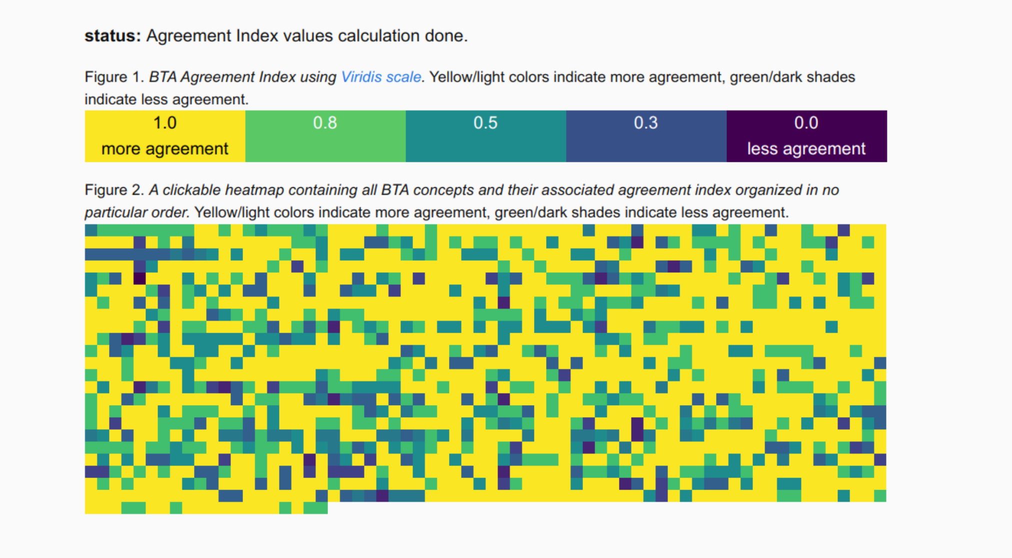
- Harnani - how to spot overlap between trait databases?
    - Jorrit - perhaps register your dataset with https://opentraits.org - see also https://opentraits.org/best-practices.html - “Ten (mostly) simple rules to future-proof trait data in ecological and evolutionary sciences” in Methods in Ecology and Evolution. doi:10.1111/2041-210X.14033
    - Jorrit - did you consider integrating with Encyclopedia of Life TraitBank ? https://eol.org/traitbank
    -
- ClimaBats data browser https://jasja.shinyapps.io/ClimBats/
- Anna - suggest a way to list which traits have been extracted for specific publications. So, that others don't have to redo the work for those publications.
- Anna - mentions - Stephens PR, Pappalardo P, Huang S, Byers JE, Farrell MJ, Gehman A, Ghai RR, Haas SE, Han B, Park AW, Schmidt JP, Altizer S, Ezenwa VO, Nunn CL. Global Mammal Parasite Database version 2.0. Ecology. 2017 May;98(5):1476. doi: 10.1002/ecy.1799. PMID: 28273333.
- https://www.nature.com/articles/s41597-023-02157-4
- Linking traits to the sources, where is the corpus stored?
    - ClimaBats pdfs are likely with the first author:
        - jeremy.froidevaux(.at.)gmail.com
- ClimaBats Viewer vs. the Publication
    - Viewer should have citation, methods
        - references are currently in the download (Traits_references.csv), but not through the viewer
        - https://jasja.shinyapps.io/ClimBats/_w_c316e7bb/EuroBaTrait_v1.0.zip
- How do you retain the detail? Eg. individual with a morphometric trait
    - Aja’s method of co-roosting
        - is to retain page number and text/sentence of trait export as a data cell in the database
        - Running list of edits
- How to grow project
    - Short term goals, what is achievable
    - Work packages, can be cited - (e.g., WorldFAIR 10 had 3 pubs 10.1/10.2/10.3 see https://globalbioticinteractions.org/worldfair https://doi.org/10.5281/zenodo.10719265)

10 Jul 2024

Present: Hernani, Kendra, Aja, Donat, Jorrit, Cecilia, Anna, Quentin

- EHA’s debarment from federal funds
    - In the media:
        - https://www.nytimes.com/2024/05/15/health/ecohealth-alliance-peter-daszak-nih-grants.html
        - https://www.nytimes.com/2024/05/01/science/covid-lab-leak-ecohealth-nih.html

    - Unclear timeline (officially 3-5 years; possibly longer or investigator specific)
    - Kendra transitioning after August to somewhere new
- Aja got 20k Bat literature reference, working towards v0.4 of https://batlit.org .
- Aja mentioned that the IUCN project is a little delayed - bat names review of IUCN RedList to point out missing names, or name updates needed. Cullen and Aja are figuring out how to help IUCN to update their names. Reusing method
- Kendra - re: IUCN batnames list - how did Aja select the names from IUCN. Aja used the name agreement index https://jhpoelen.nl/bat-taxonomic-alignment/ .



status: Agreement Index values calculation done.

Figure 1. *BTA Agreement Index using Viridis scale*. Yellow/light colors indicate more agreement, green/dark shades indicate less agreement.

1.0 more agreement	0.8	0.5	0.3	0.0 less agreement

Figure 2. *A clickable heatmap containing all BTA concepts and their associated agreement index organized in no particular order.* Yellow/light colors indicate more agreement, green/dark shades indicate less agreement.

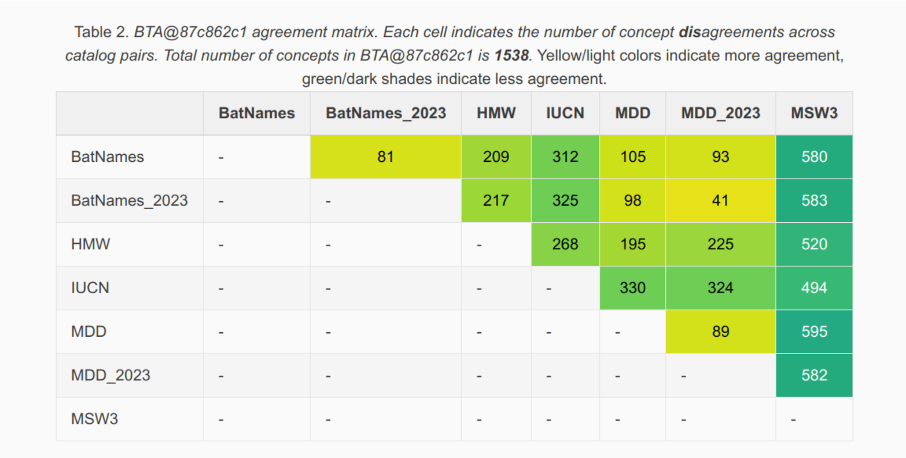


Table 2. *BTA@87c862c1 agreement matrix. Each cell indicates the number of concept **dis**agreements across catalog pairs. Total number of concepts in BTA@87c862c1 is **1538**.* Yellow/light colors indicate more agreement, green/dark shades indicate less agreement.

	BatNames	BatNames_2023	HMW	IUCN	MDD	MDD_2023	MSW3
BatNames	-	81	209	312	105	93	580
BatNames_2023	-	-	217	325	98	41	583
HMW	-	-	-	268	195	225	520
IUCN	-	-	-	-	330	324	494
MDD	-	-	-	-	-	89	595
MDD_2023	-	-	-	-	-	-	582
MSW3	-	-	-	-	-	-	-

-
- Kendra asked a question about *Pteropus giganteus* - *Pteropus medius/vampyrus*, from [MDD](MDD) “the name medius is used instead of giganteus because giganteus is now considered a synonym of P. vampyrum”. Reference to record - https://linker.bio/line:hash://sha256/87c862c149f6a33c7d8c344568a9a7f80310cbeed324395aa94fff2d5248a3e6!/L2,L1182.tsv csv version: https://linker.bio/line:hash://sha256/029b12ab3a7d5c2cef01df411ea3ecb4a1fa5c199acca3bee4544a6ee59691c7!/L1,L1182.csv

See reference for *P. giganteus* change to *P. medius*: Mlíkovský, J. (2012). Correct name for the Indian flying fox (Pteropodidae). Vespertilio, 16, 203-4. https://www.ceson.org/vespertilio/16/203_204_Mlikovsky.pdf

- Donat asks about input for the selection of the BHL publications to be imported into BLR / BiodiversityPMC https://docs.google.com/spreadsheets/d/1fxi-ndcZJymKiAJx0XiIllx29UuIO5Ndzo0vo9uEZxg/edit?gid=0#gid=0
- Aja/Quentin - re: co-roosting database - no progress yet
- Bat Literature Project: https://sandbox.zenodo.org/communities/batlit/records?q=&l=list&p=1&s=10&sort=publication-desc
-
    - What is the Purpose?
        - Anna: a corpus should help provide a way to link a copy of an article that has a DOI. Document the workflow, which paper have been contributed in what way
            - How do we bridge in trait/content-based data to the corpus
        - Aja/Cullen:
            - Use keywords to build subcorpor. Identifies: What is a bat publication: taxonomy, just mentioning of a .
        - Donat:
            - Get immediate access to papers from any bibliography we are publishing in the future
        - Aja:
            - Provide access to students
    - Currently lacks definition/scope
        - Create a policy of what is a bat publications
        - Nancy's current goal is Koopman paper digitization (taxonomy)
        - Drosophila
        - What is the scope of the community: publications, supplements, data sets, PhD thesis, grey literature, reports, etc.?
        -
    - Currently lacks procedure:
        - Who can contribute?
        - What is the metadata? How complete do you need to be? Does curator need to do a check?
        - How do you attach content-derived metadata? How complete do you need to be?
    - Is this a hosting platform or will it point to already accessible platforms (e.g., BHL)?
        - Donat: hosting
    - Legal issues?
        - Pre-2000 open, the rest according to the existing open / closed access

        -

3 Jul 2024

Present: Hernani, Nancy, Nate, Cullen, Cecilia, Donat, Kendra, Aja

NASBR symposium abstracts

- Deadline now to Jul 10th
- We currently have 8 symposium abstracts
- Impact of taxonomic splits on bat-viral data
- https://docs.google.com/document/d/1TFPWLNxf5KrczCNGQzybiru93AwRtbx7l6NNWVgRcbY/edit – Current doc

- BatLit abstract
    - Could be edited more
    - Have the website: https://batlit.org/
        - Now 5.5k of the 20k+ PDFs from Van Cakenburge
        - All now uploaded via the cloud
        - AfricaBat Report, e.g., 2016: https://www.researchgate.net/publication/311888337_African_Chiroptera_Report_2016
            - Lots of data here, difficult to access
            - I think this is the latest one: https://www.researchgate.net/publication/364943183_African_Chiroptera_Report_2022_Complete_Report
            - Perhaps there is a structured database that underlies this?

Jorrit is additionally processing the BHL bat publications

- 25k articles that cover bats, but also some other taxa (Jorrit: 2024-07-05 an initial assessment suggests about 6k works that mention "Chiroptera". https://github.com/bat-literature/bat-literature.github.io/issues/15 in BHL . Please cite sources for the 25k articles so that I can understand the discrepancies)
- Part of the Plazi workflow – but Jorrit is helping to optimize in some ways
- Data in supplements in PubMedCentral – OCR and using the figures and tables
- Could import these via JATS etc
- OR import the articles as supplements – into Biodiversity PMC
    - Shortcut for getting data in – still searchable and parseable
    - Jorrit helping with versioning of those data
    - Donat wanting to make the data accessible for search
- Cullen and Donat are sponsoring for now – until we get further funding (R01?)

- Targets: https://docs.google.com/spreadsheets/d/1fxi-ndcZJymKiAJx0XiIllx29UuIO5Ndzo0vo9uEZxg/edit?gid=0#gid=0
-

Initial BatLit uploads

- https://github.com/bat-literature/bat-literature.github.io/issues/14
- In Zenodo Sandbox, with metadata – for checking now
- Could we also add the Supplementary Files?
    - E.g., scan documents for “supplement” and then have a part of the pipeline that brings in those files as well
    - Currently not being dealt with – but it is a big issue
- Could we pull out full text of all these articles in a batch way – if OCRed
-

Cullen linked to – https://www.nature.com/articles/s41597-023-02472-w

- A dataset on African bats’ functional traits
- “Here, we present the AfroBaT dataset, a compilation of trait data on 320 African bat species containing 76,914 values for 86 traits focusing on morphology, reproduction, life-history, trophic ecology, and species distributions.”

26 June 2024

Present: Hernani, Aja, Nancy, Nate, Ariadna, Anna, Cullen, Cecilia, Donat, Jorrit

Bat literature

- Data from Cakenburghe
    - 1400 articles just starting with E
- OSF library is having difficulty with the size of the data – 50 Gb, likely more than that, of PDFs
    - Jorrit has 10 TB for 20 euros a month https://www.hetzner.com/de/storage/storage-box/.
- Zotero shared library
    - Jorrit is indexing from here: https://bat-literature.github.io/#version-history
        - https://bat-literature.github.io v0.3 just released today.
        - refs.csv has limited metadata of ~1000 paper

- Swiss-based bat literature repository
    - These were all file cabinets of non-digitized literature
    - Similar to what is at major museums
    - From Manuel Ruedi
        - “Well, as I said, I don't have so many pdfs, mostly from the last 15 years (which you certainly already have). Before that, had my private collection of reprints, but not digitalized, and not so many (if you want them all, you can can get them; most will be destroyed anyway upon my retirement, in 3 years from now...). But the most complete library of paper reprints of bats of the World (former private colelction of Villy Aellen + many more since Pascal's rule), are in our Museum library; for this, contact our chief librarian Laura.Zbinden@ville-ge.ch. As our museum in now under repare for the next few months, she will also tell you if and when those bat reprints will be again available (currently, they are in boxes and inaccessible).Hope it helps. Manuel Ruedi”
        - “Dear Manuel– Do you happen to have Pascal Möschler’s email – or is he not anymore involved in bat and the global bat library? His email seems not to work anymore. pascal.moeschler@ville-ge.ch Thanks for a hint Donat”

- Searching DOIs
    - Looking up open access or not:
        - R Package Open Alex: https://docs.ropensci.org/openalexR/
    - Also the RCrossRef package
        - https://ciakovx.github.io/rcrossref.html
    - Jorrit already did this query in OpenAlex: https://github.com/bat-literature/bat-literature.github.io/issues/2
        - Not 100% reliable though
        - Donat helped follow this up:
            - https://github.com/bat-literature/bat-literature.github.io/issues/2#issuecomment-2153428532
        - Unpaywall is another option
            - https://unpaywall.org/products/simple-query-tool
            - Up to 1000 queries at a time
    - “Reference-extract sheet that is incomplete and using Zotero tools/platform to fill gaps. This open access column is not in the Zotero references.csv https://docs.google.com/spreadsheets/d/1y5uBKvyzDQgQUtyRHvn5f20IO

        A7LoMho/edit?usp=sharing&ouid=111971835458035348757&rtpof=true&sd=true ”

- Aim is to emulate this TaxoDros:
    - https://zenodo.org/communities/taxodros/records?q=&l=list&p=1&s=10&sort=newest
    - https://taxodros.github.io/ << view on how to query these things
- Donat: “cakenberg has 34480 bibliographic references” !

Big Bat Database work

- They have 100 databases that they are extracting from
- Taking all the verbatim names and locations from the databases
- Not as broad as PanTHERIA, etc
- More primary syntheses of individual traits from the literature
- @Hernani (he/him) @Ariadna Morales (she/her) @Cecilia Montauban et al. do you consider species interactions a trait? (e.g., parasite, fruit plant, prey etc.)
    - Yes, to some extent – not an entire focus, but does include this
    - Initially related to the genomes of the bats – for the Bat1K folks

Abstracts for the symposium

- https://docs.google.com/document/d/1TFPWLNxf5KrczCNGQzybiru93AwRtbx7l6NNWVgRcbY/edit
- Guidelines doc: https://cdn.wildapricot.com/214854/resources/docs/meetings/year/2017/2017%20NASBR%20Abstract%20Checklist.pdf?version=1495429744000&Policy=eyJTdGF0ZW1lbnQiOiBbeyJSZXNvdXJjZSI6Imh0dHBzOi8vY2RuLndpbGRhcHJpY290LmNvbS8yMTQ4NTQvcmVzb3VyY2VzL2RvY3MvbWVldGluZ3MveWVhci8yMDE3LzIwMTclMjBOQVNCUiUyMEFic3RyYWN0JTIwQ2hlY2tsaXN0LnBkZj92ZXJzaW9uPTE0OTU0Mjk3NDQwMDAiLCJDb25kaXRpb24iOnsiRGF0ZUxlc3NUaGFuIjp7IkFXUzpFcG9jaFRpbWUiOjE3MTk0MTQ5NDd9LCJJcEFkZHJlc3MiOnsiQVdTOlNvdXJjZUlwIjoiMC4wLjAuMC8wIn19fV19&Signature=YODu-S8zff4028Koms3ZRbZIw8WfSVUSTa0SNLJJaYL4O-Hk9k8a~Qh2yZi8GhsQTYcZ4-VR7fgRfsMP63bcaF3WLGlDy7zWfCrmQH-6MQbPb7RpP1kJQC-foxCrjyZj4MGlbdrccqvpuexA23qrkksTWnPUR~Vz87jSfeQNCMoitX9F-k4jSub3kPvNCrwgYst3UB1Y9OE6cgdFc5CM9~IdccoGYKu1MNeH-6OP0sx~RahYD3RiTGXxHQeD0ZM853X8Oz9he1e01DoFAryQL4CzcPVyKi1VpuhfJTsYgrgrwr5m-6Fo7~QPpf2V9gLMkMRLg3DHvuDBFVreSNx~1w__&Key-Pair-Id=K27MGQSHTHAGGF
-

19 June 2024

Present: Quentin, Hernani, Aja, Nancy, Nate, Ariadna, Anna, Cullen, Hernani, Cecilia, Kendra, Donat

NASBR Symposium abstracts
- Cullen to submit collation in two weeks (July 3 – but abstracts to her by July 1)
- Most pressing

Jorrit is open to going to NASBR
- But we could also re-invest those funds into more data extraction
- Work has been to get all those publications up on Zenodo
  - Then following the Drosophila workflow
- We could have a pre-meeting workday
- Oct 24-26 are the main meeting days (https://www.nasbr.org/annual-meetings)
  - Oct 23 we could have a meeting everyone together
- Ask the group about to what extent want to interact in person versus keep that into the repository
  - If he goes – maximize his impact – advertise the initiative
  - Could make a poster for the purposes
  - Recruit PDFs from meeting attendees

Bat literature digitization work
- Uploading to BiodiversityPMC
- Not only taxonomic papers
  - Ecological papers
  - Shared archive of Bat Cons International
- OSF folder where collecting PDFs: https://osf.io/28cvq/files/googledrive
  - Want to start uploading documents
  - Do we have a quality control pre-check? Could be above the technical capabilities of the folks with the PDF libraries
- Journals:
  - From web of science journal list:
    - African Bat Conservation News https://www.africanbats-npc.org/abcn
    - AFRICAN CHIROPTERA REPORT https://www.africanbats-npc.org/acr
    - Bat Research News
  -
- Donat is at Davos meeting - https://www.worldbiodiversityforum.org/en/wbf-2024
- Market out there for environmental assessments

- Map of Life folks are there
    - But not currently sending data back to GBIF
- GBIF, CoL folks are at the meeting – Tim Robertson
- Plazi aims to get new names into CoL
- Question
    - How to resolve all the names
    - Need to cite taxonomic treatments
- COI as possible ‘type’ for names / taxa

Bouchout Declaration meeting
- Hackathon-like meeting
- Integrating publications into CoL
- https://bouchoutdeclaration2024.org/
- Tim Robertson, Marcus Deiring, Olaf Banki
- Goals
    - Get more names in CoL
    - Make names databases more FAIR
    - Trying to get nomenclature folks to talk to each other
- Plant folks working on this too
    - Schellenberger Costa, D., Boehnisch, G., Freiberg, M., Govaerts, R., Grenié, M., Hassler, M., Kattge, J., Muellner-Riehl, A.N., Rojas Andrés, B.M., Winter, M., Watson, M., Zizka, A. and Wirth, C. (2023), The big four of plant taxonomy – a comparison of global checklists of vascular plant names. New Phytol, 240: 1687-1702. https://doi.org/10.1111/nph.18961
- CoL is trying to enable cross-linking / translation across taxonomies
    - Not going for “one taxonomy” to rule them all
- Don;t want just another wishful thinking list
    - Want actual needs from stakeholders
    - What is needed, how to achieve
    - Time to actually make an impact

12 June 2024

Present: Kendra, Quentin, Aja, Cullen, Ariadna, Hernani

Quentin scored $6 million grant to study invasive species for 3.5 years

Dave Waldien shared a spreadsheet of IUCN bat species - Aja has updated the taxonomic alignment

Ariadna possibly to attend NASBR, would present on the integration of Bat1K and other genomic data with the Big Bat Database for data mining

05 June 2024

Present: Nancy, DeeAnn, Kendra, Quentin, Aja, Ariadna Morales, Hernani Oliveira, Cecilia Montauban, Nate (late)

- Introductions for guests
- Discussing NASBR 2024 Symposium:

  Potential Presenters and Proposed Titles:

  Presenter: Nancy B. Simmons – American Museum of Natural History (USA)
  Proposed Title: Key links in the global bat data chain

  Presenter: Hernani F. M. Oliveira - Universidade de Brasília (Brazil).
  Proposed title: Challenges and perspectives in collating bat trait data worldwide.

  Presenter: Winifred Frick - Bat Conservation International (USA)
  Proposed Title: Bat data for conservation planning: the What, How, and Why

  Presenter: Nathan Upham - Arizona State University (USA)
  Proposed Title: When do taxonomic splits matter for bat biodiversity and viral risk analyses?

  Presenter: Aja Sherman - Bat Eco-Interactions Project (USA)
  Proposed title: The Bat Taxonomic Alignment (BTA), A Standardized Review of Bat Names Across Multiple Taxonomic Authorities

  Presenter: Cullen Geiselman - Bat Eco-Interactions Project (USA)
  Proposed Title: Liberating literature and the data within: the Bat Literature Repository and evolution of data extraction and curation

- Cullen suggested offline “Perhaps we should consider the main points we'd like participants that attend the whole thing will take away instead of talking about BTA, Co-roosting dataset, etc in separate talks.”
- Aja suggested one talk on liberating inaccessible literature and dark data (as well as linking) and another on taxonomy for this process, the BTA, and struggles with keeping up with taxonomic changes.

- Hernani will present on Big Bat Database, traits that colleagues are measuring, and how to link.
- Nancy was concerned that Nate and Aja's talks will be too similar. Let's discuss this with Nate when he returns. Also concern that 3 taxonomy talks might be redundant. Agreed that Nancy's talk will still be relevant since some friends don't know about existing databases, darwin core, or other available resources.
- Kendra asked about the status of the Big Bat Database. Members gave us background and informed us that changes are happening with new management…
- Nancy recommends Hernani presents on the Big Bat Database, the history, and current challenges
- Aja introduced the co-roosting
- Hernani will share literature associated with best practices for data mining
- Hernani asked if we document unpublished anecdotal interactions, Nancy recommends to leave these out unless published evidence
- Cecilia interested in how we keep track of where the interactions are documented from, Aja shared citation lists and internal identifiers
- https://www.nature.com/articles/s44185-024-00043-9
- https://libguides.derby.ac.uk/literature-reviews/prisma-lr
- https://biotxplorer.text-analytics.ch/?species1=Rhinolophus&size=1
- https://zenodo.org/communities/coviho/records?q=&l=list&p=1&s=10&sort=newest
- DeeAnn emphasizing gray literature from conservation organizations
    - How to access this?
    - E.g., interactions noted from non-peer reviewed literature
    - Could flag this in the data matrix
    - https://smartconservationtools.org/en-us/
        - "SMART platform consists of a set of software and analysis tools designed to help conservationists manage and protect wildlife and wild places. SMART can help standardize and streamline data collection, analysis, and reporting, making it easier for key information to get from the field to decision-makers."
- Nate parsing his talk relative to Aja's
    - Will be a use case of the BTA on bat-virus interaction data – so will follow logically after Aja's BTA talk
    - Interfacing with Kendra too
- Nancy and Aja mentioned the IUCN integration of the Batnames.org / MDD taxonomy for bats – feeding this into conservation action

10 May 2024

Present: Nate

Brainstorming for the GBatNet meeting – name of group and notes on the "short blurb" that we need prior to Weds to be included in the printed materials:

- Name ideas
    - Bat Biotic Interactions for Organizing Mammal Ecology (BatBIOME)
    - Bio Biotic Interactions Organized by Species (BatBIOS)

- Blurb [sent to Susan Tsang on 16 May]
    - Our mission is to digitally liberate and organize all knowledge of bat biotic interactions – including that contained in taxonomic and ecological literature, databases, biocollections, and other archives – for studying the multifaceted ecological roles of bats worldwide. We are a group of biodiversity collections-related scientists that have been meeting weekly since April 2020. The Covid-19 pandemic reminded us that much of the knowledge of bats interacting with other bats, as well as with parasites, predators, prey, and habitats, are locked in offline books or paywalled PDFs and thus constitute 'dark data' that are digitally unavailable for synthetic analyses. These data are central to building robust models of viral spillover risk and other complex ecological dynamics, the predictive abilities of which are increasingly critical to society. Additionally, the pace at which taxonomic changes to bats and other taxa are occurring has brought considerable instability to sources of genetic and ecological data that are linked by species names and inform these models. Therefore, our group endeavors to both improve the digital access and taxonomic linking of biotic interaction data for bat species globally.

8 May 2024

Present: Aja, Kendra, DeeAnn, Cullen, Quentin

Aja caught DeeAnn up on what happened last week

Kendra notes that she will be attending the iDigBio meeting - will give a presentation at the "Digital Collections Data and Tracking Disease Workshop" - really interested in bat synonymy and in co-species data [so, this is us]

Kendra asked for slides to help - Aja shared prior slides, Cullen also has some…

DeeAnn - we really need to settle on our working group name.

- Bat Biotic Interactions and Taxonomy Working Group [or are we a network - GBatNet categorization] - Cullen will ask Tigga for advice on working group or network. Or **consortium**?
- Bat Biotic Interactions and Taxonomic Alignment Consortium?
- Need greater group input; should have this settled by next week's meeting

Aja/ChatGPT suggestions:
Global Bat Taxonomy and Ecology Network (GB-TEN)
Bat Taxonomic Interactions Initiative (BTII)
International Bat Biodiversity and Interaction Consortium (IB-BIC)
Bat Taxonomy and Ecological Network (BTEN)
Global Bat Diversity and Interaction Project (GBDIP)
Bat Taxonomy and Ecological Dynamics Consortium (B-TEDC)
International Bat Taxonomy and Ecological Nexus (IB-TEN)
Bat Taxonomy and Interaction Collaborative Hub (BATICH)
Worldwide Bat Taxonomy and Ecological Exchange (W-BTEE)
Bat Taxonomy and Interaction Assessment Network (BTIAN)
As a grou, we reviewed the current GBatNet networks and "Interdisciplinary Projects" [AKA working groups]. We are akin to the "Bat Phenotypes and Evolution Network (BPEN)"??
Aja news: BTA done (will step aside for a bit, then review); will work on making it a relational database as next steps. Co-roosting and BTA manuscripts are in the works. She is also creating how-to videos and tutorials for the BTA.
Aja invited the team to review both datasets while she is writing, but plans to review one more time before publication.
We are still tackling the landing site for the searchable platform for the literature collection and BTA (probably batbase).

Co-roosting data and BTA are done!
Zotero goals: 20K more by end of summer

Aja - to DeeAnn and Nancy - need more examples of lineages with lots of change/unusual scenarios (e.g., Pteronotus, Miniopterus)

Aja will work with Quentin on Gephi interaction models, species accumulation curve, and possible data mining automation resources.

Aja asked the group: When making co-roosing dataset - was there a plan for automating it? Cullen thought we would be able to use our data as a training set…

Cullen - provided some funds for Jorrit to work with Zotero collection - pdfs to zenodo etc.…

Discussed NASBR symposium proposal: https://docs.google.com/document/d/1TFPWLNxf5KrczCNGQzybiru93AwRtbx7l6NNWVgRcbY/edit

Will submit the 14th

Quentin new business: recently had a meeting with a group that wants to push forward a IUCN working group on parasites. Interesting to think about what IUCN status might look like for parasites. Endo and ectoparasites etc… will keep our group apprised as to how it progresses.

1 May 2024

Present: Aja, Anna, Nate, Kendra, Cullen

Anna planning to defend her PhD in early Fall
- Still doing some data collection
- Looking at isotope signatures of processed vs natural food

Kendra taking notes through this OtterPop app: https://otter.ai/u/haFT-sl_mGZPjC_o7IJJUg8SxqA?utm_source=va_chat_link_1

AI translation / transcripts have gotten very good lately
- https://sonix.ai/ this is the website I was using
- https://www.linguee.com/
    - Seems much better than Google Translate
    - Helpful for her Zooniverse project: Canyon Critters

NASBR meetings in October
- Cullen has been organizing
- We have 6 talks, could be ready to submit already
- Potential for Jorrit to be invited too
    - Digitization of bat literature
    - Lots of PDFs and metadata available now
    - Bringing the publications to Zenodo → Plazi method for this
        - But then additional steps needed to index and search
    - Modeling it off of Drosophila:
        - https://zenodo.org/communities/taxodros/records?q=&l=list&p=1&s=10&sort=newest
        - https://www.taxodros.uzh.ch/

        - Building the search feature is the part that Jorrit will help with – also creating the digital signatures from that

- NA bat people at doing their own symposium
- EuroBat people don't have their own funding
    - https://www.nature.com/articles/s41597-023-02157-4

24 April 2024

Present: Aja, Nancy, Nate, DeeAnn

- DeeAnn is back from fieldwork in NW corner of Uganda
    - Flight cages, molecular lab – now just need bats to eat /

- We will all be in Houston for the GbatNet meetings in ~1 month
    - Prep?
    - Everyone is representing a working group here
        - Bringing these together with the networks

- Symposium at NASBR
    - Bringing bat datasets together
    - Projects
        - Aja
            - BTA
        - Cullen:
            - Co-roosting dataset
        - Jorrit:
            - Bat Literature group
    - Other talks
        - Nancy
            - Intro to symposium, progress on BatNames and MDD alignment
        - Nate
            - When do taxonomic splits matter / range maps changing through time and intersecting that with observational data
    - Outside folks
        - BCI – Fred Frick – Teague O'Mara

        - Ash Malmlov; Anna Fagre – ash@bathealthfoundation.org; anna.fagre@gmail.com

- BTA updates
    - Making a website for the BTA –
    - Setting up a peer review system for that
        - Flagging the content by Nancy, DeeAnn, Nate – Connor B too
    - Making subsets of the BTA too
        - Automate this process
        - Use SQL to select fields – the underlying database would always be the same
        -

17 April 2024

Present: Aja, Nancy, Cullen, Donat, Nate,

- Nancy doing fieldwork in Belize in May
    - 85 people going over the course of 2-3 weeks
    - Becker lab, Melissa Ingala, Kelly Spear, all taking samples at different time points – aiming for recaptures to get data through time
        - How many re-captures are they getting?
        - PIT tagging started in 2021 – 13% recapture rate – arm banding the vampire bats
            - Bats tend to have 5-6 years life span in some species, or 40+ years in Myotis in the wild
    - Collecting animals for whole genome work – 23 genomes being targeted
    - Air samplers for collecting eDNA in the caves / hollow trees / human structures

Other agenda items:

- Potential symposium at NASBR in Oct 2024
    - Guadalajara, Jalisco between 23-26rd October 2024
    - Due May 15th: https://form.jotform.com/231825735583160
    - Ideas
        - Considering how the databases can be used in planning and decision making
        - Databases in bat biodiversity science

        - Inviting Jorrit
        - Topic – bats and databases – linking networking databasing – what questions are uniquely enabled by aggregating data
            - Uses for informing bat research priorities
            - Liberating, linking, leveraging bat data for research and conservation priorities
            - Leveraging = taking advantage of something existing for a new purpose / scale of questions
            - Liberating side – how could I contribute to this – how could I use it?
        - Roundtable as part of this
            - Symposium + workshop
            - Workshop could focus more on "how to" of liberating and linking
            - Panel discussion – bringing ideas together
    - Hernani Olivera

3 April 2024

Present: Aja, Nancy, Cullen, Kendra

- GBatNet meeting
    - Possibly include NABat, Hernani Oliveira (oliveira.hfm@gmail.com), Orly Razgour, and/or Ariadna Morales (ariadna.biologia@gmail.com)
    - There will be a questionnaire soon for travel details. Reservations will be made through GBatNet, not individuals
    - Need catchy title
        - Big Bat Databases: Easier Said Than Done
    - Speaking with facilitators
- Bat Literature Project
    - If Aja finds manuscripts, she will omit
    - Kendra will add literature
- BTA
    - Spoke about species country lists and if it was possible to solve island only issues - maybe revisit this topic, maybe not
- Kendra brought to our attention this grant opportunity
    - https://www.nsf.gov/pubs/2024/nsf24069/nsf24069.jsp
    - She also apologized for using the saying "negative Nancy" during previous meeting

27 March 2024

Present: Quentin, Aja, Nancy, DeeAnn, Nate, Anna

GBatNet connections

- Meeting in Houston – Week of the 20th of May – 21-24/25
- New name needed
  - Bat-Pathogen Working Group?
  - Other groups: https://www.gbatnet.org/interdisciplinary-projects/
    - Big Bat Database is taking off
      - Hernani Olvera, and Ariadna Morales leading this
    - How old is this bat?
    - Evo-devo group
    - Population sizes of bats
    - Bats in habitats – parasites/microbiomes
    - Stress in bats (DeeAnn)
    - Immune systems in bats (Dan Becker)
    - Key habitats
    - SciComm of bat science
    - Socioeconomics group
    - 11 of original 15 –
  - Our group
    - Co-roosting dataset
    - Bat taxonomic alignment
    - Ecological interaction
      - Bat Biotic Interactions Working Group
      - Biodiversity Exchange on Bat Biotic Interactions
      - Chirocosmos!
      - Bat Interactions and Digital Knowledge
      - Bat Biotic Interactions and Taxonomy Exchange (Bat BITE)

- Asdell's Patterns of Mammalian Reproduction
  - Punch cards burned in a fire
  - Digitized?
    - Nate has mammae number and litter size data digitized – and collaborator Tom Stewart does have a scanned copy of the volume – planning to upload to Zenodo in the near future

Moth defenses against bats

- New employee at AMNH: Jesse Barber
  - Barber, J.R., Plotkin, D., Rubin, J.J., Homziak, N.T., Leavell, B.C., Houlihan, P.R., Miner, K.A., Breinholt, J.W., Quirk-Royal, B., Padrón, P.S. and Nunez, M., 2022. Anti-bat ultrasound production in moths is globally and phylogenetically widespread. *Proceedings of the National Academy of Sciences*, *119*(25), p.e2117485119. https://www.pnas.org/doi/10.1073/pnas.2117485119

DeeAnn wrote a book:
- “Lives of” Bats; series Princeton University Press
- “Lives of Bats” book
    - Out in October 2024!

Co-roosting dataset
- https://docs.google.com/spreadsheets/d/1y5uBKvyzDQgQUtyRHvn5f20IOA7LoMho/edit#gid=691336520
- Looking at the GBIF matched names vs others
    - Record secondary authorities in “interpretted_name” column
    - Migrate (*) in GBIF
    - Plants: Quentin
    - Macroparasites: Anna
    - Ants: Donat?
- Aja working on manuscripts:
    - Methods for co-roosting
    - Methods for bax taxonomic alignment

20 March 2024

Present: Quentin, Aja, Cullen, Nate, Donat at the end

IPBES reports – species assessment
- Figuring out baselines vs monitoring of species

Co-roosting data – Aja is cleaning up the names relative to the GBIF backbone

iDigBio meeting 2024 in Kansas:
https://www.idigbio.org/content/digital-data-biodiversity-research-conference-2024

TDWG in Japan in Sept: https://www.tdwg.org/conferences/2024/call-for-abstracts/

NASBR in Mexico / Guadalajara in Oct: https://www.nasbr.org/
- We could do a symposium? https://www.nasbr.org/symposia24
    - NA bat people
        - Occurrence data
    - Data aggregating – movement data aggregation
    - Ask Nancy who all could be giving talks
    - Way for our talks to not be lost
    - Roundtable possibly
    - Bat morphometric data possibly

    - Bat data aggregation and synthesis generally
- May 15 – deadline – time to assemble

Oscar project
- https://oscars-project.eu/
- Covers all disciplines (not just biodiversity) but there is a life sciences cluster
- Could be good for Jorrit –

iNaturalist data on degree of establishment of invasive species in a given country
- Current issue with iNaturalist not pushing these data to GBIF

Donat talked to Jorrit – going to get all of these bat publications in Zenodo
- Scope for upload:
    - Taxonomy and ecology initially
    - Nancy, DeeAnn, Nate (MDD), Victor Van Cakenburgh
    - Gabor Scorba and Sandro Bertolino would be good to contact too

6 March 2024

Present: Quenting, DeeAnn, Aja, Nancy, Kendra, Donat

Aja explaining
- the biotic interactions table and what she needs from the team
- Cleaning up the mammal database synonyms, priority
- Starts writing the ms biotic and coroosting
- Andrea has not change document for BTA
- Writing the BTA paper by summer

Bat literature list
- Zootero add keywords automatically? If not possible do it later

28 Feb 2024

Present: Quentin, Donat, Aja, Cullen, Kendra

Reviewed top priorities with the group that wasn't present last week

Aja compared taxonomy (of interaction dataset) against Checklist Bank & GBIF for agreement

BiCIKL+ submission by Donat & Quentin (will know if funded by July)

Donat to send template to Aja for organizing citations stored on Zotero

21 Feb 2024

Present: Quentin, Aja, Nate

Camping gear

- https://www.outdoorgearlab.com/
- https://www.bioliteenergy.com/products/campstove-2-plus

Top priorities right now

- Taxonomy for the co-roosting dataset
    - Did a check against the GBIF backbone – using this to verify
    - Bat-bat interactions
        - We can dig into this part of the dataset to a greater extent
        - Plot the two phylogenies facing each other
        - Make line thickness proportional to the frequency of interactions
    - Rarefaction analysis as well
        - Are these sampling curves saturating? Or will it go on and on depending on increased sampling
- BTA manuscript
- How-to for the BTA

14 Feb 2024

Present: DeeAnn, Nancy, Aja, Nate

AirTags – helpful for tagging the field supplies when traveling

Ebola virus-like particles – just have the protein on the outside to be recognized by the bats –

Using DNA/RNA shield – also BAMM Banker for tissue culture

- Slowly make way into freezing
- Needs new -80C

Aja update
- BTA versions – batNames –

7 Feb 2024
Present: Cullen, Kendra, Nate, Nancy, Anna, Donat, Quentin

Anna presenting on ringtail omnivory at the University of Georgia, Odum School of Ecology annual student symposium
- Rank abundance curves of dietary items in scat
- Being an omnivore is on a big spectrum
- https://www.inaturalist.org/observations?place_id=any&taxon_id=41676
    - Includes lots of feces images
- Early ASM poster about this work (Speir et al. 2022)

Cullen / Donat / Aja / Nancy
- Digitization of bat literature – all literature is the plan there –
- First working on 20k Drosophila publications
    - Will figure out full workflow there
- Bat literature library in Switzerland – in Geneva – Can order PDFs from them
    - Muséum in Geneva https://www.ville-ge.ch/musinfo/bd/mhng/cco/accueil.php , and was founded by the now retired Pascal Moeschler.
- Lots of low hanging fruit here
    - Victor Van Calkenburgh has a massive PDF library that he gave access to
- What are the goals?
    - Every publication ever on a bat?
    - Can be a parallel bibliographic effort and PDF ect
    - Multiple goals
        - 1. Get full bibliography reference for all publications
        - 2. Get PDFs for those papers
        - 3. Upload PDFs to Zenodo
        - 4. Tag / annotate / extract PDFs for forming digital knowledge
        - For 2-3-4 steps, can target a specific topic / clade
- What metadata do we need?
    - Some columns already developed on the “reference-extract” sheet of the co-roosting data.

- Nate / MDD can send literature on bats from the Zotero of the MDD
    - Will send to Aja

Bat Libraries:

- Bat bibliographies generated from the General Index subset:
        - Special Index to Species: https://archive.org/details/IndexGenerium

Other Internet Archive scans:
- BioStor https://archive.org/details/biostor-127588/page/438/mode/2up
- Big issue with bioheritage library is the lack of separated articles. Volumes are scanned and not parsed by article metadata.
    - Lack of funding/staff to do this. Priority is Smithosonians own stuff

24 Jan 2024
Present: Donat, Cullen, Kendra, Nate, Nancy, Aja, Quentin

Our collective bodily injuries and aging
- Nancy retiring – getting rid of administrative things

Liberating large amounts of bat literature
- Could we take all the PDFs from Nancy's computer – plus the files from batBase
- Possible to have a drag and drop?
    - Nancy has a combined folder for the first time – just joined things across computers
- Aja can help to create the Zotero record – BibTex or other –
- Then Donat can work with Plazi team to automate the uploads to Zenodo
- Also add the MDD references in this context
    - 196 "bat" references included – not updated for the v1.12 version

Philippines project work – aiming to get more funding via EcoHealth
- https://www.washingtonpost.com/national-security/2023/09/07/deep-vzn-virus-wild-usaid-program/
- https://stopspillover.org/ (program that is funded by USAID)
- Lots of in country work being done –
    - Capacity building ultimately

10 Jan 2024
Present: Quentin, Cullen, Kendra, Aja, Nate, Donat

Updating the taxonomy of the bat co-roosting data

- Quentin took all the names included + matched it against the GBIF taxonomy backbone
- https://docs.google.com/spreadsheets/d/1y5uBKvyzDQgQUtyRHvn5f20IOA7LoMho/edit#gid=691336520
    - WorkingDoc_Raw-Data-Extraction_ExpertLiterature_8_8
- Steps were to use the GBIF taxonomy matcher tool
    - Species lookup: https://www.gbif.org/tools/species-lookup
    - ChecklistBank has something different – provides stable identifiers –
        - https://www.checklistbank.org/tools/name-match
        - Will return “valid” species names – but the species could still definitely be split
            - Kendra example with Miniopterus schreibersii – used to be widespread species, now restricted to Europe – so all the records in the Philippines are inaccurate
- Sources for digitizing all the bat literature
    - Try different sources
    - Private funding
    - Swiss Govt – RDF version of checklist bank – $300k for these conversion services – just for the Swiss bats for the starter
        - Proof of what can be done for the swiss species
        - BiCikL – Arcadia
    - Working with Nancy – to get all the bat literature – leverage to get more funds to expand further
        - Gathering PDFs from her, batnames, MDD
        - Some of the bibliographic data contain errors currently

    - Knowledge Hub with the BiCiKL
        - https://bicikl-project.eu/biodiversity-knowledge-hub
        - Pensoft is involved because of the JATS conversion needed for PMC – Biodiversity PMC
            - https://sibils.text-analytics.ch/search/ (SibIs portal)
    - Also this “Knowledge Centre for Biodiversity”: https://knowledge4policy.ec.europa.eu/biodiversity_en

- Locality information for the bat co-roosting information

    - There is a need to consider the “localityCountry” field in comparing to the BTA for taxonomic translation
        - Compare this to the “countryDistribution” field in the MDD for example – create a pairwise country matrix for doing this
        - >> Then have logic for the subsequent comparisons
        - E.g., if nameA exists in countryX in taxonomyY, then we have the possibility of properly translating taxonomic splits

03 Jan 2024

Present: Kendra, Nancy, Quintin, Anna

- Hellos from the holidays
- Quintin working on
    - flora online : no global one available that’s great; need for more specific areas (e.g., Congo)
        - Difficult to acquire funding for the projects
    - IUCN invasive species as threat
        - Data available through rredlist R package
- Role of reference books
    - ICTV virus taxonomy still heavily referenced book
    - Walker’s large bat book: almost 25 lbs
    - Walker’s Mammals of the World may have had last version in 2018
            - Interesting/Harsh Review broaching the topic https://www.mammalwatching.com/2018/07/17/a-review-a-long-review-of-the-latest-version-of-walkers-mammals-of-the-world/
    - Out of date by the time their published?
    - IUCN
        - embedded citations + species-specific authors can give idea of when updated
        - 2000 was when all animals started to be assessed
        - Many bats that were listed covered since the ‘80s
    - Taxonomy always an issue in any of the records
- How to deal with rejection/controversy