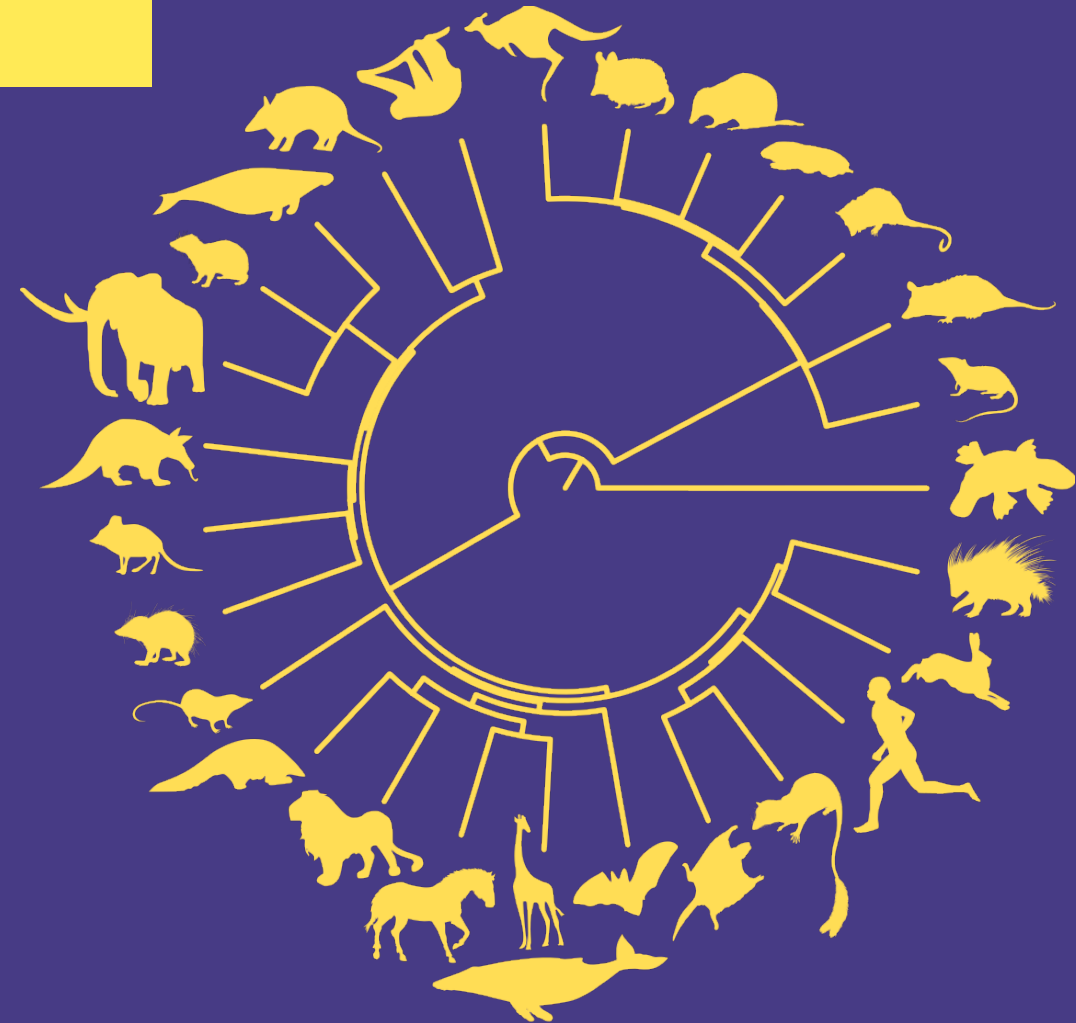


# Mammal-centric spillover risk modeling: species meaning and ecological context are key

**Nathan S. Upham**  
*Arizona State University*

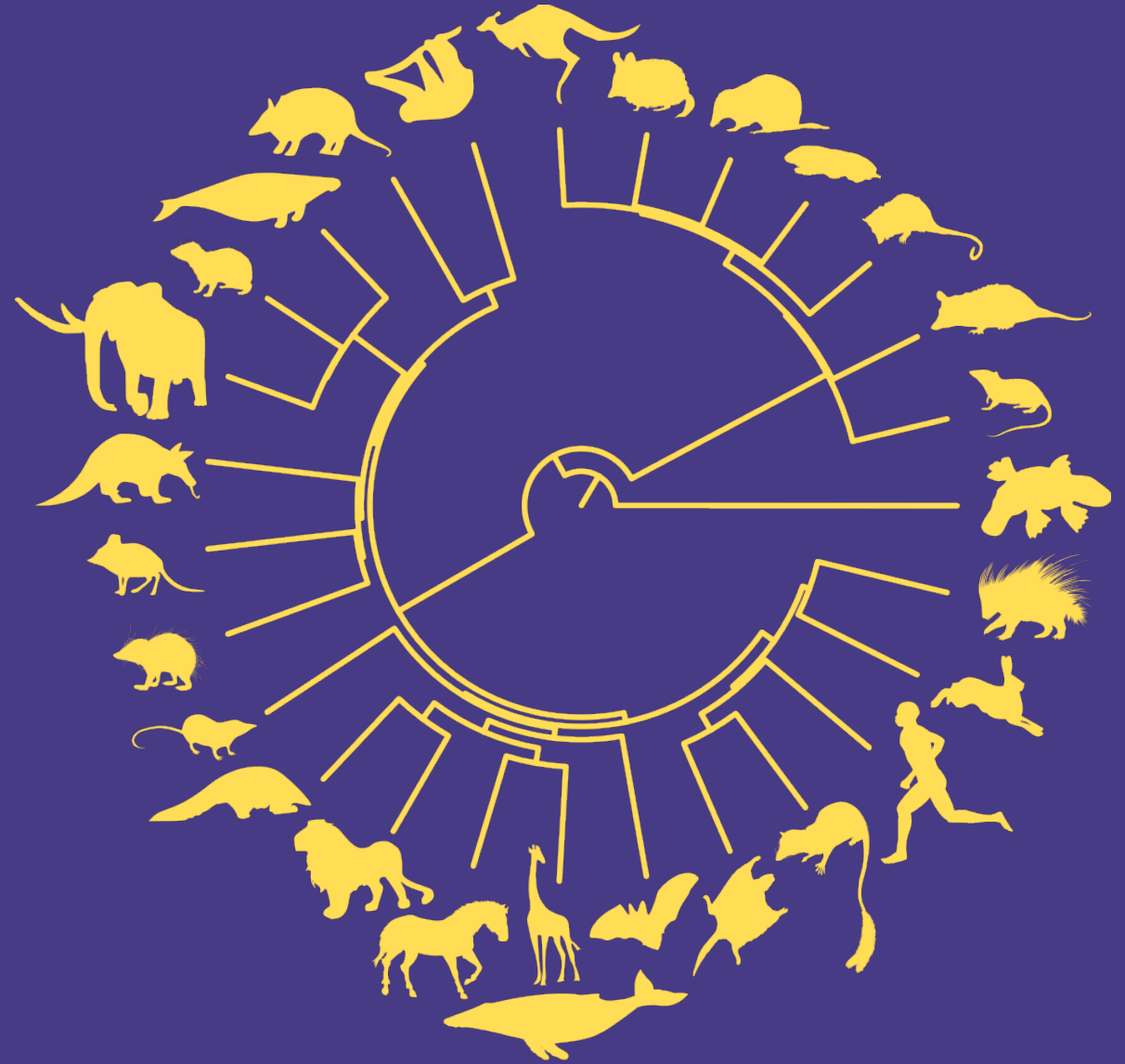
*17 Nov 2022*

*EcoHealth Alliance*





***Taxonomic and ecological  
data curation are the keys  
to accurate risk modeling***





Pathogen  
spillover  
to humans











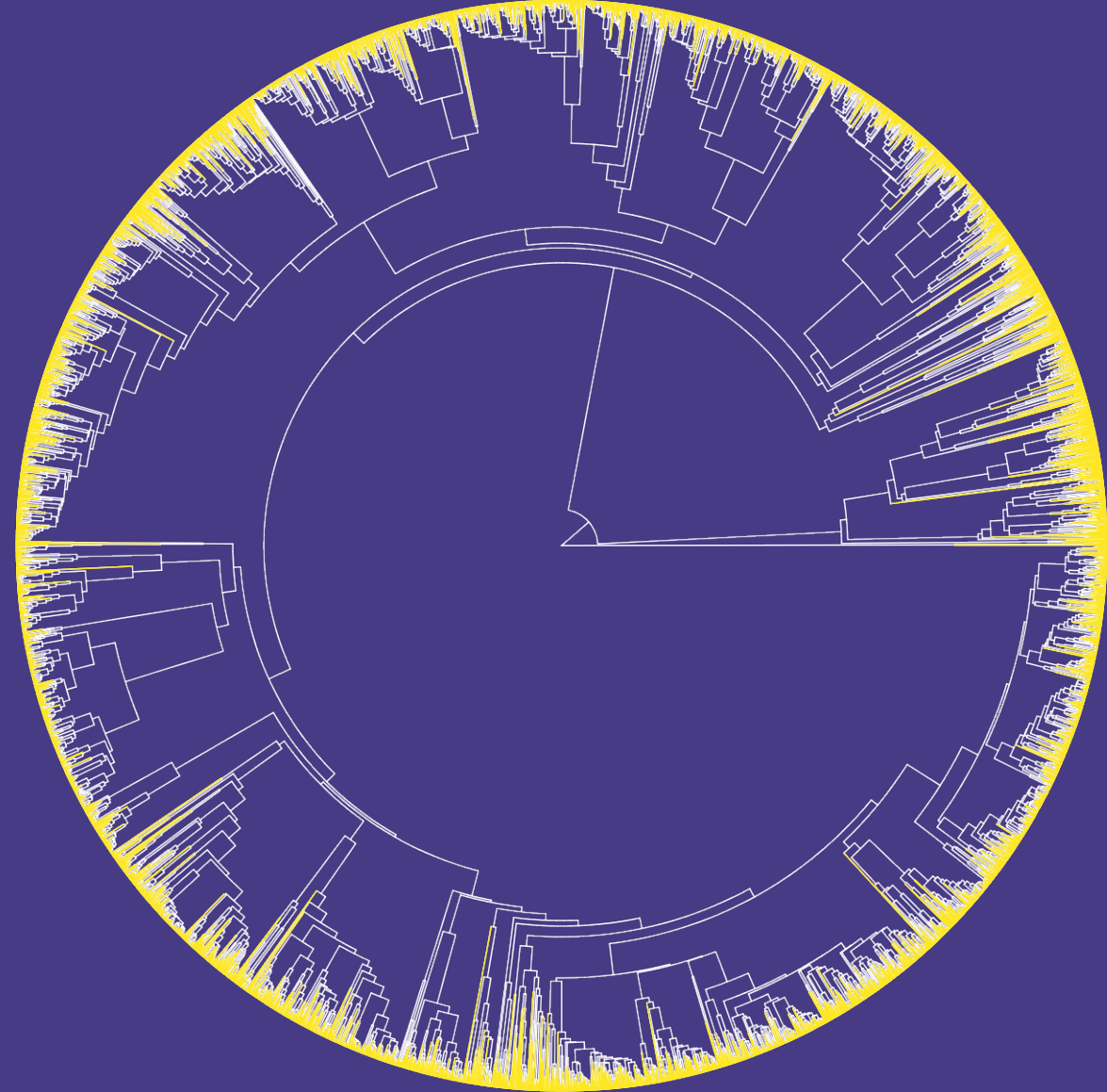
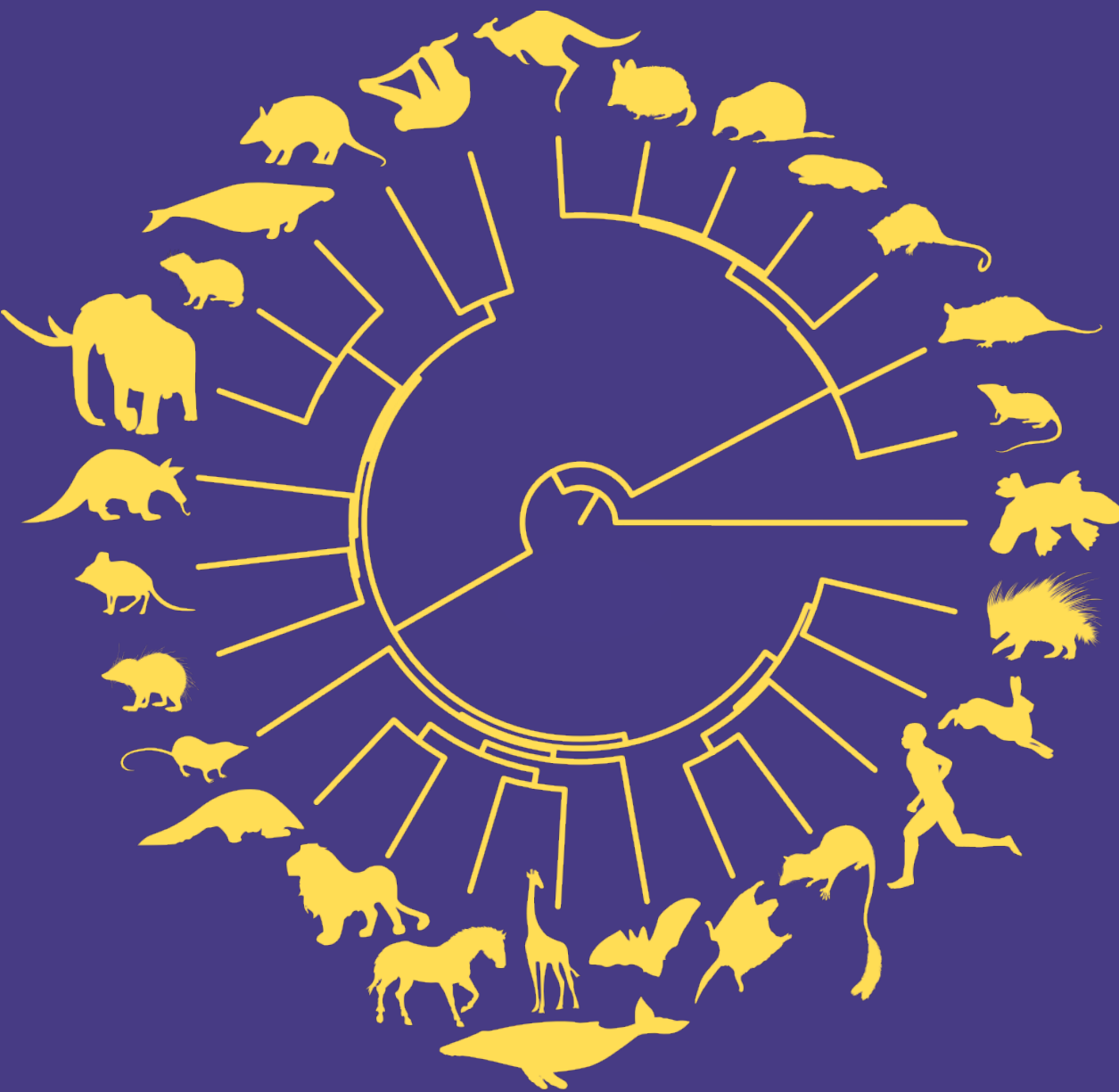
What **processes** generated this diversity of **species & ecologies**?



# 27 orders

# MAMMALIA

~6,500 species

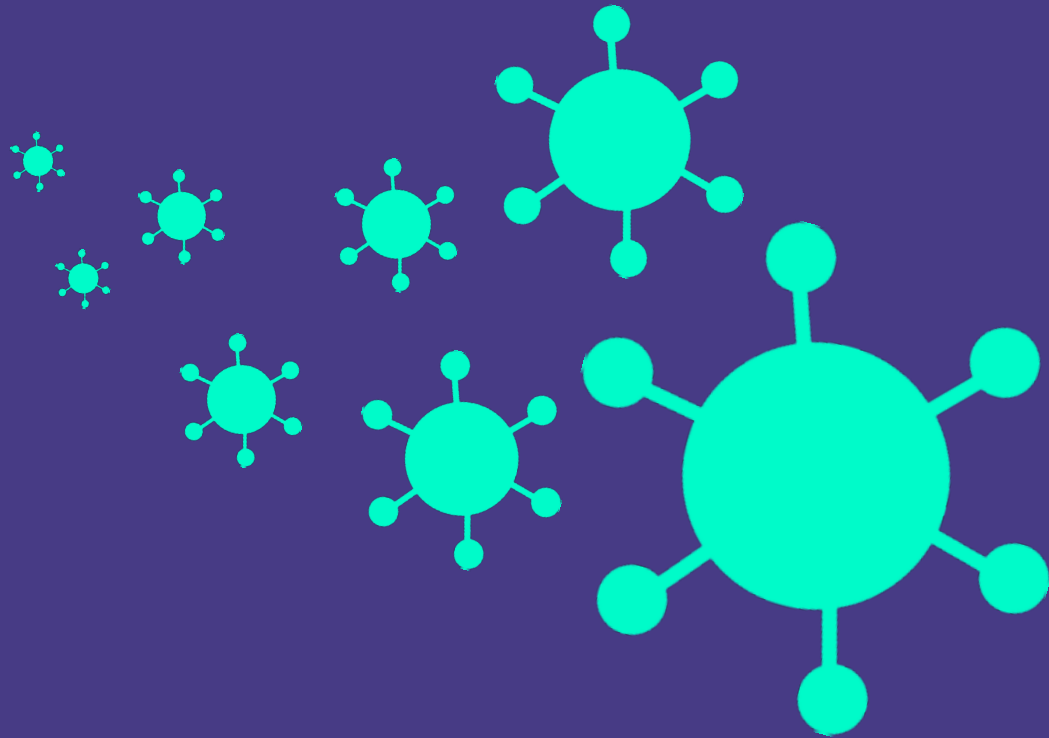


Upham et al. (2019, *PLOS Biology*)  
Burgin, Colella, Kahn, and Upham (2018, *J Mamm*)



# VIRUSES

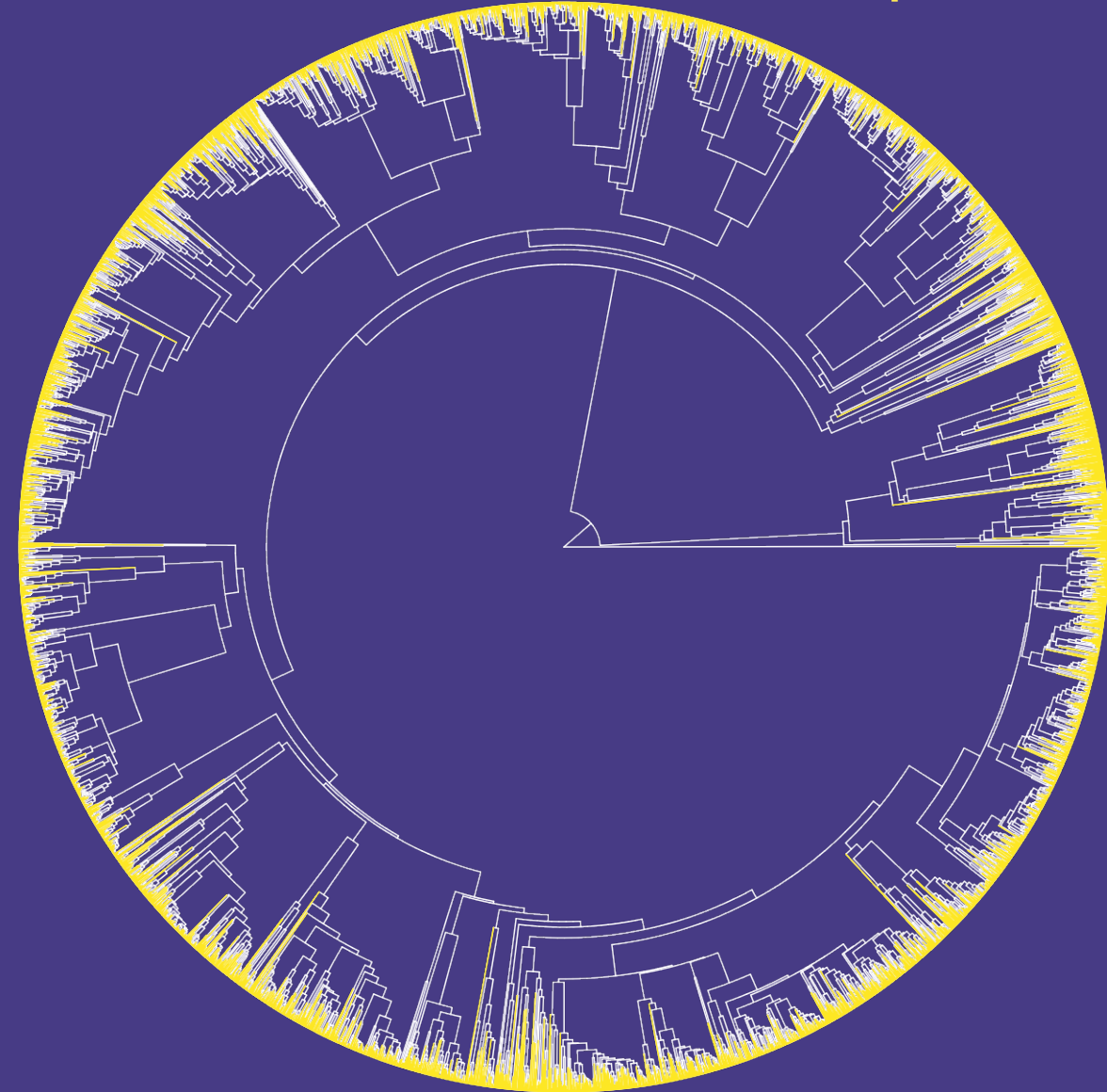
10,000–40,000  
species in mammals  
(~1,000 described)



Carlson et al. (2019, *Nat E&E*)

# MAMMALIA

~6,500 species

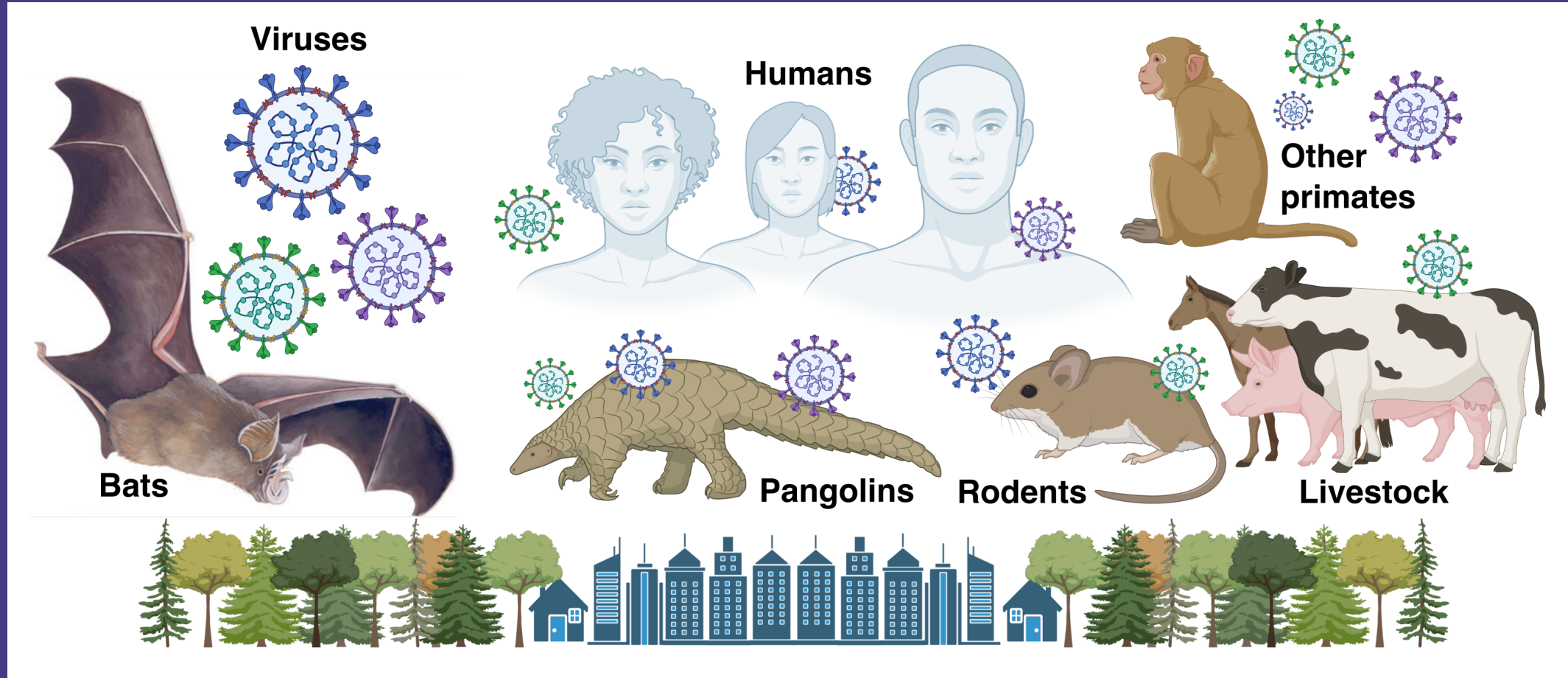


Upham et al. (2019, *PLOS Biology*)  
Burgin, Colella, Kahn, and Upham (2018, *J Mamm*)



# What is the distribution of **spillover risk**?

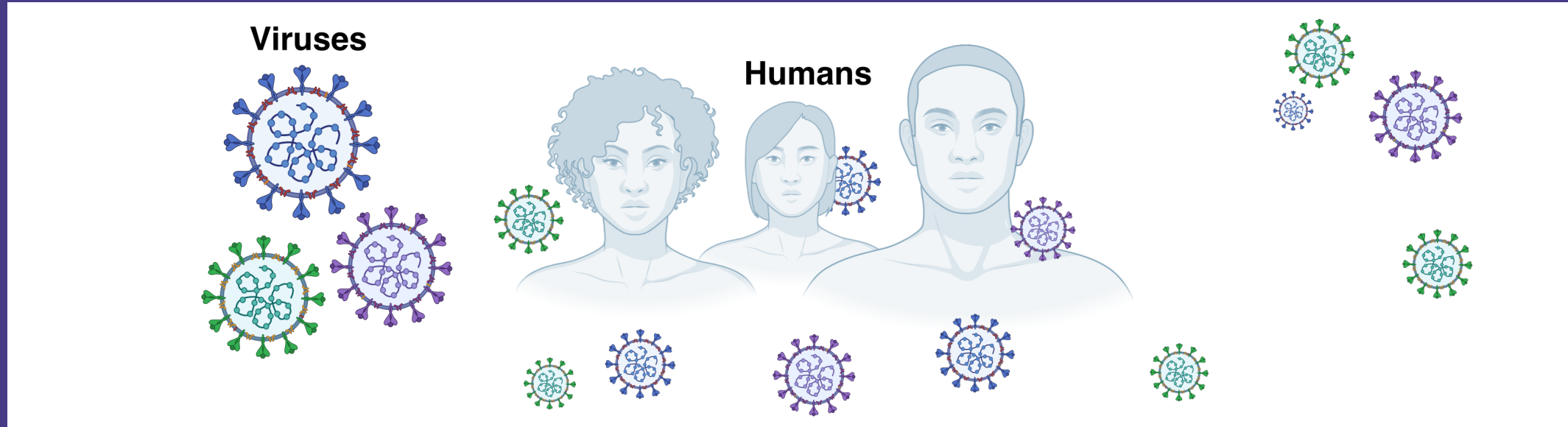
Per **species** — Per **region**



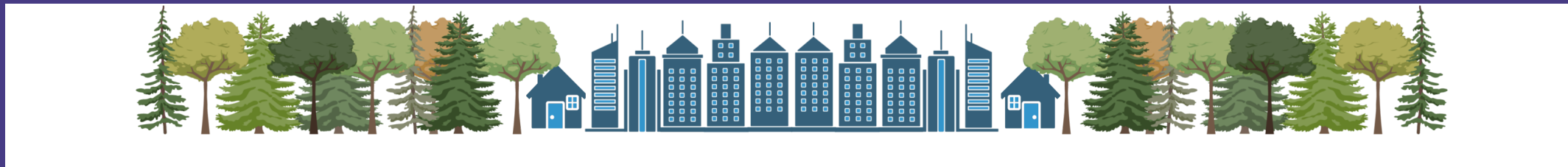


# What is the distribution of **spillover risk**?

Per **species** — Per **region**



Ecology of host species has been neglected



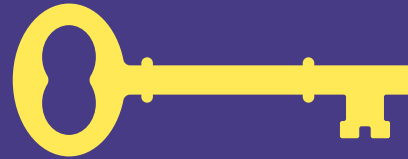


Phylogeny

Ecology



Taxonomy



Knowing *which species are present* is the first step to knowing many other things...

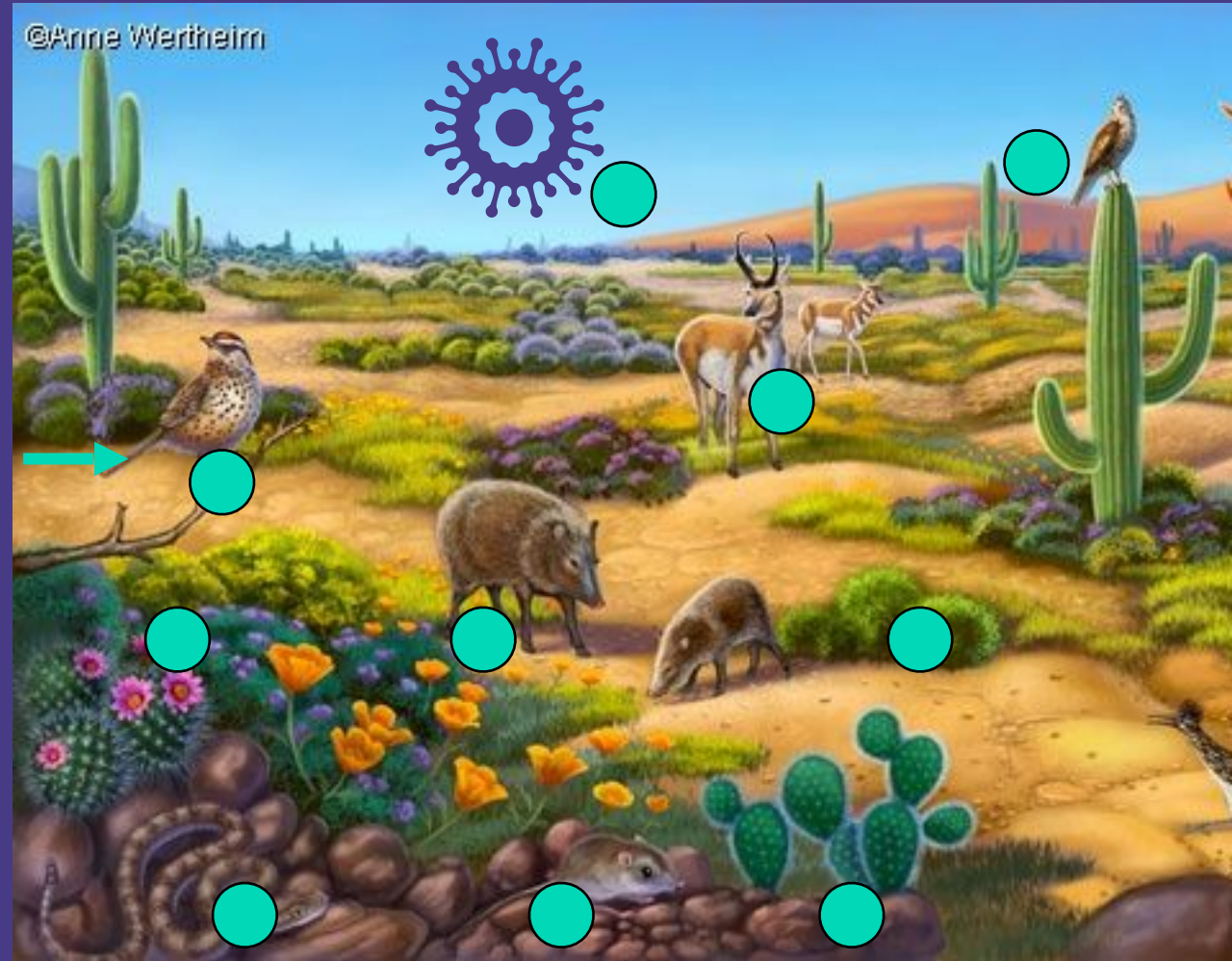
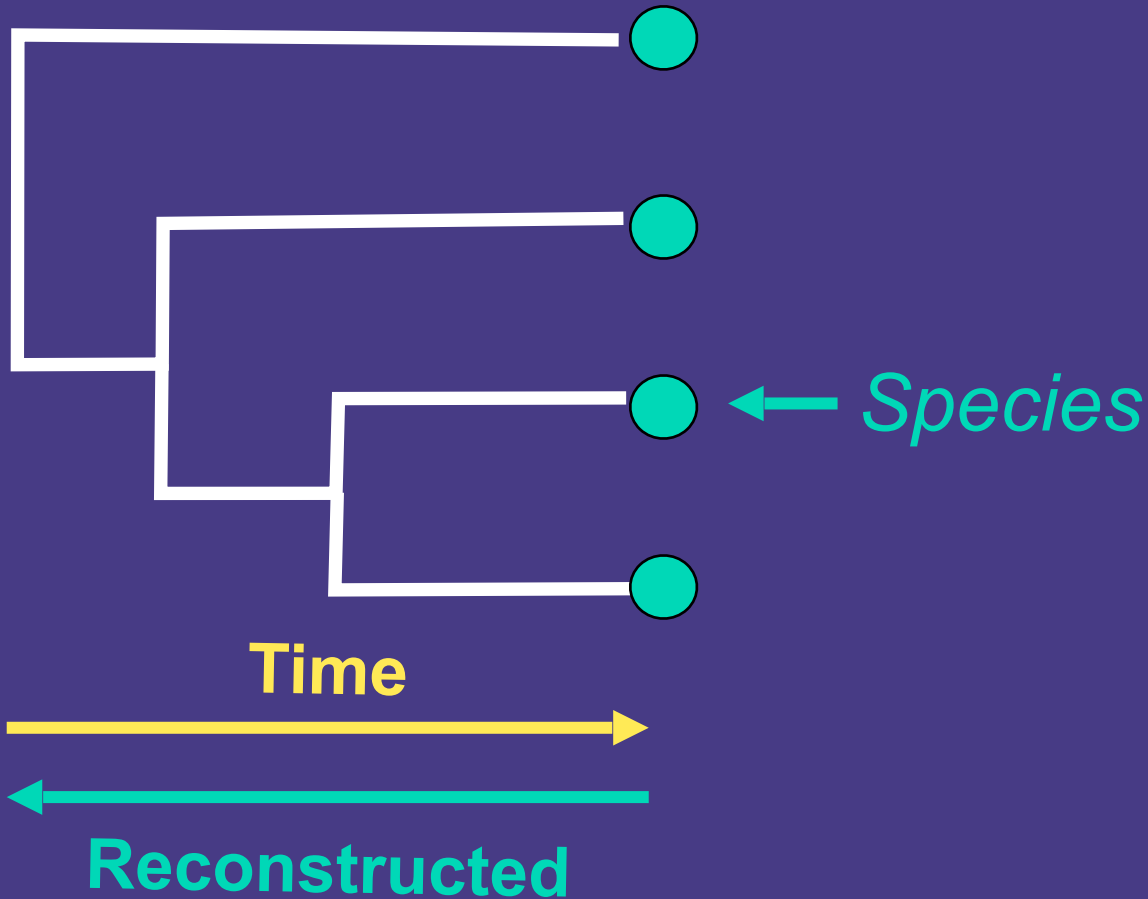


# evolutionary past



# ecological present

*Phylogeny:* common ancestry



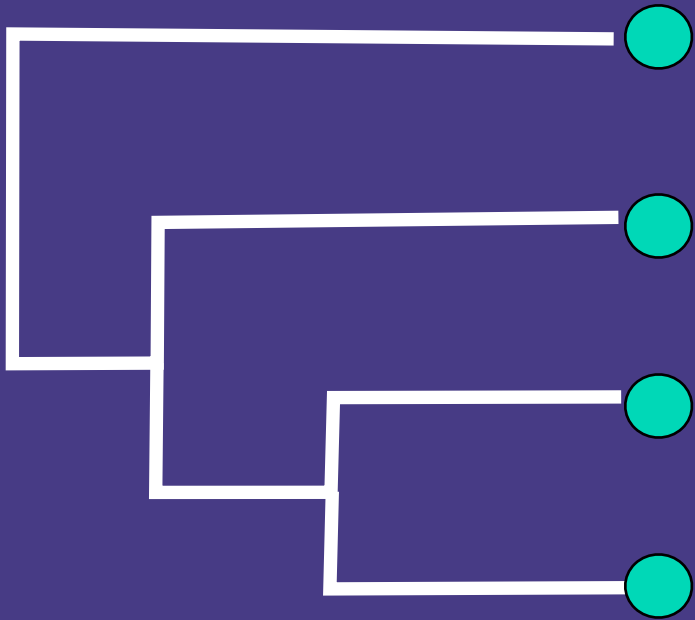


# evolutionary past

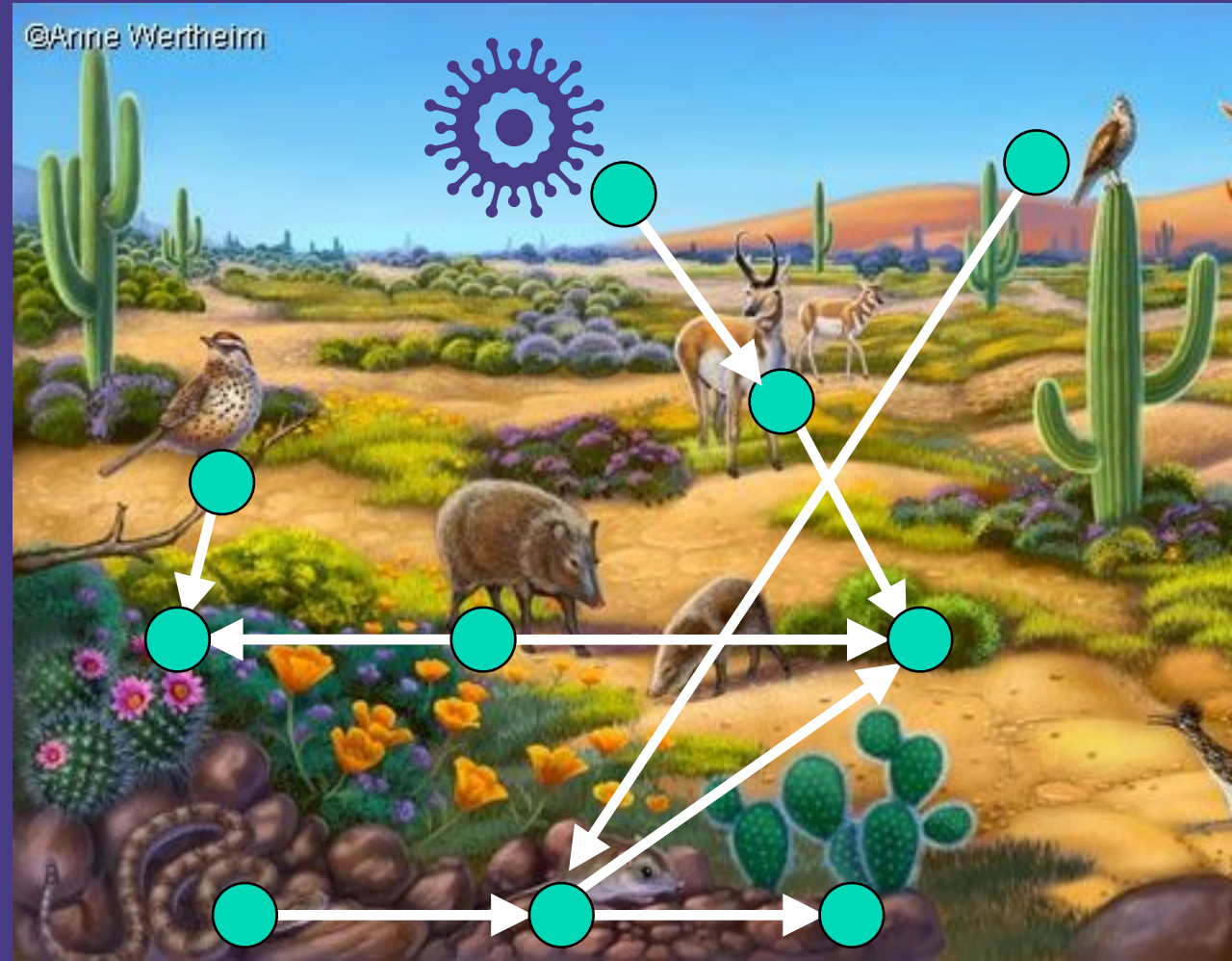


# ecological present

*Phylogeny:* common ancestry

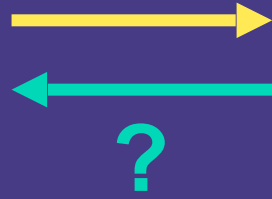


*Ecology:* species interactions



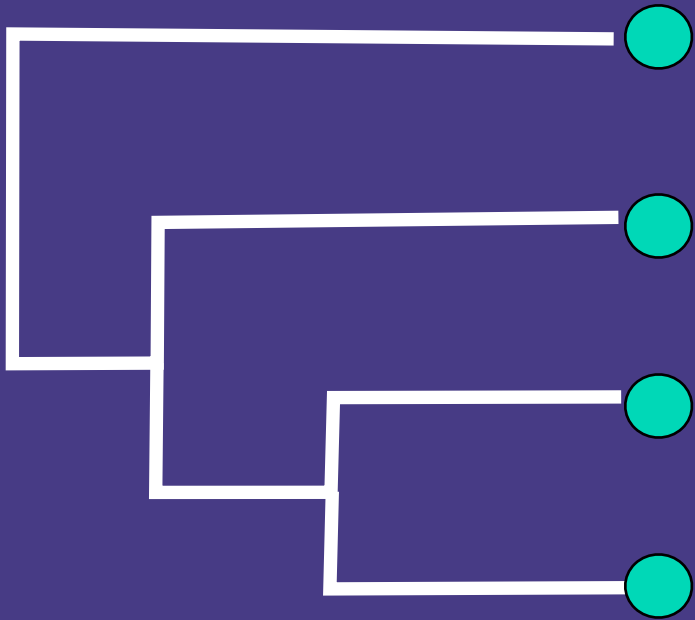


speciation & extinction

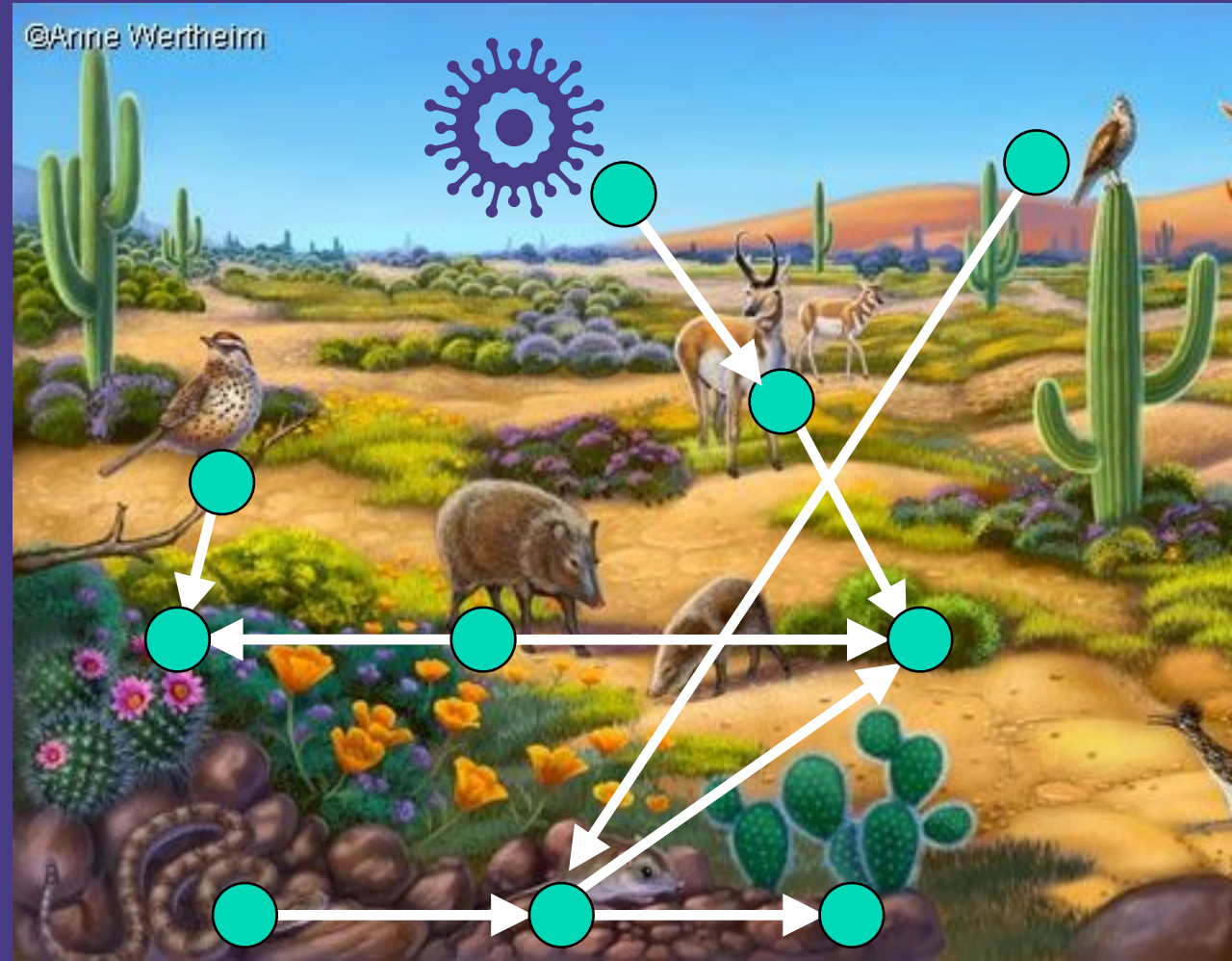


ecological present

*Future shape of the phylogeny*



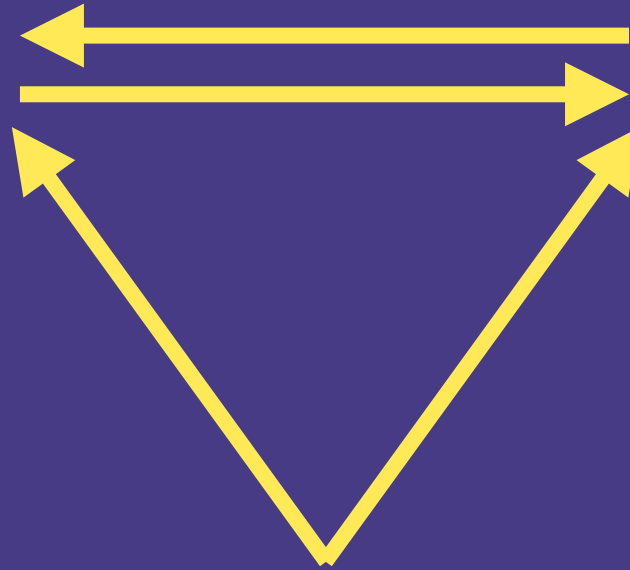
*Ecology:* species interactions



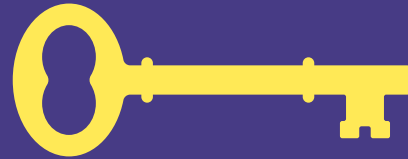


**2. Phylogeny**

**3. Ecology**



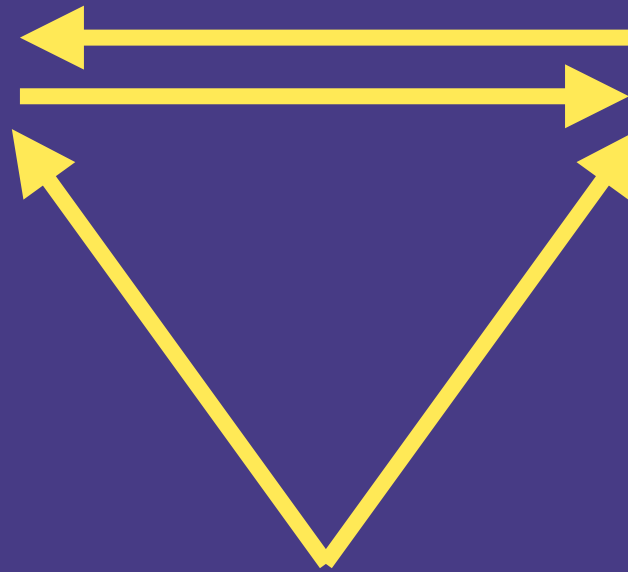
**1. Taxonomy**



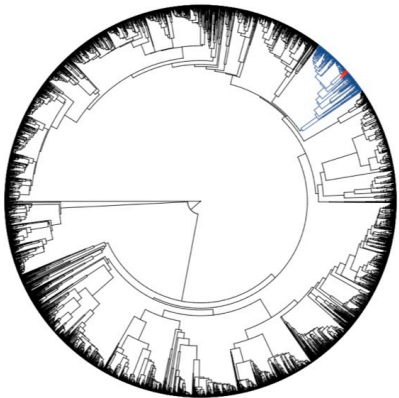


2. Phylogeny

3. Ecology



1. Taxonomy



Upham lab of  
Phylogenetic  
Ecology



**ASU**® Biodiversity Knowledge  
Integration Center  
Arizona State University



# TODAY

- I. Host-virus 'dark data'
  - Main approach to improving risk models





# TODAY

## I. Host-virus 'dark data'

## II. Data liberation & collection

- Progress so far
- Initial risk models



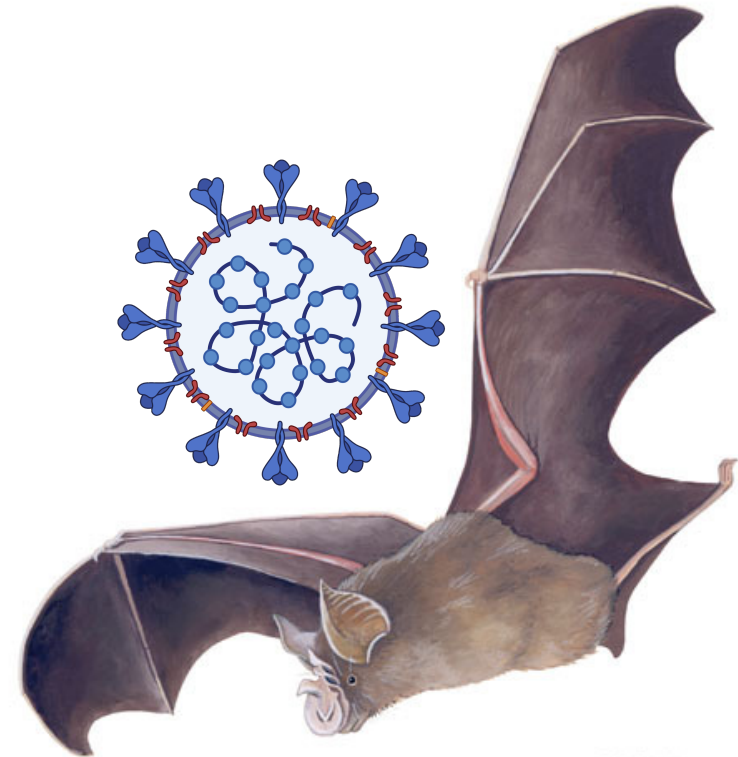


# April 2020...

- **Novel coronavirus:** SARS-CoV-2
- **Likely host reservoirs:** mammals ➤ bats ➤ horseshoe bats (*Rhinolophus* sp.)

## Major questions

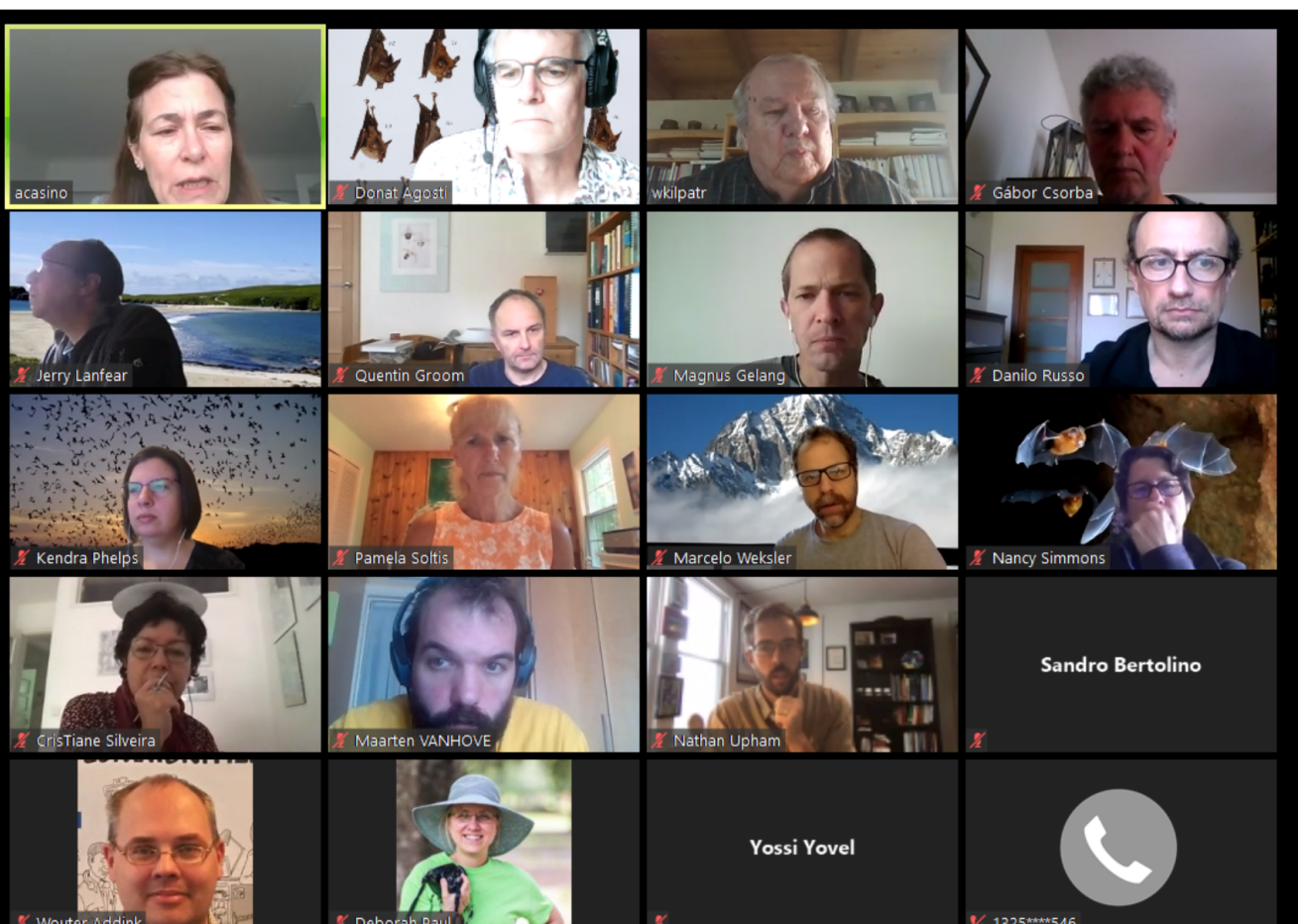
- How did viral spillover to humans occur?
- What was the exact host species (& host immunity)?
- **What is the risk of future spillover events?**



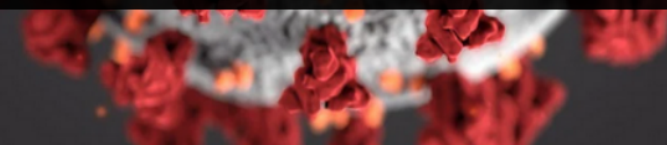


# April 2020... Goal: liberate host-virus data

- **Novel coronavirus: SARS-CoV-2**
- **Likely host reservoirs:** mammals ➤ bats ➤ horseshoe bats (*Rhinolophus* sp.)



**CETAF-DISSCO COVID-19 TASKFORCE**



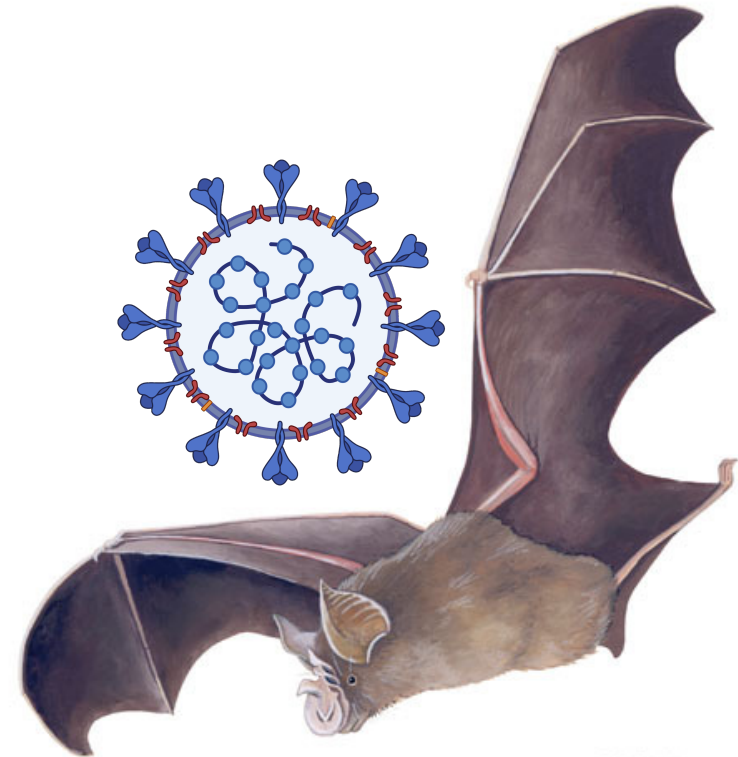


# April 2020... Goal: liberate host-virus data

- **Novel coronavirus:** SARS-CoV-2
- **Likely host reservoirs:** mammals ➤ bats ➤ horseshoe bats (*Rhinolophus* sp.)



**Nancy Simmons**  
AMNH Mammal Curator  
Global bat expert  
[batnames.org](http://batnames.org)



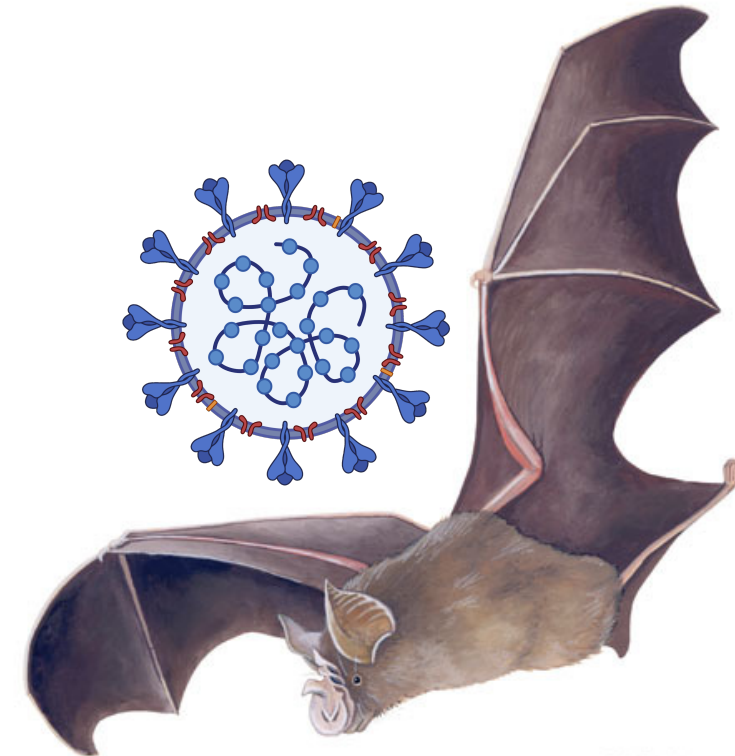


# April 2020... Goal: liberate host-virus data

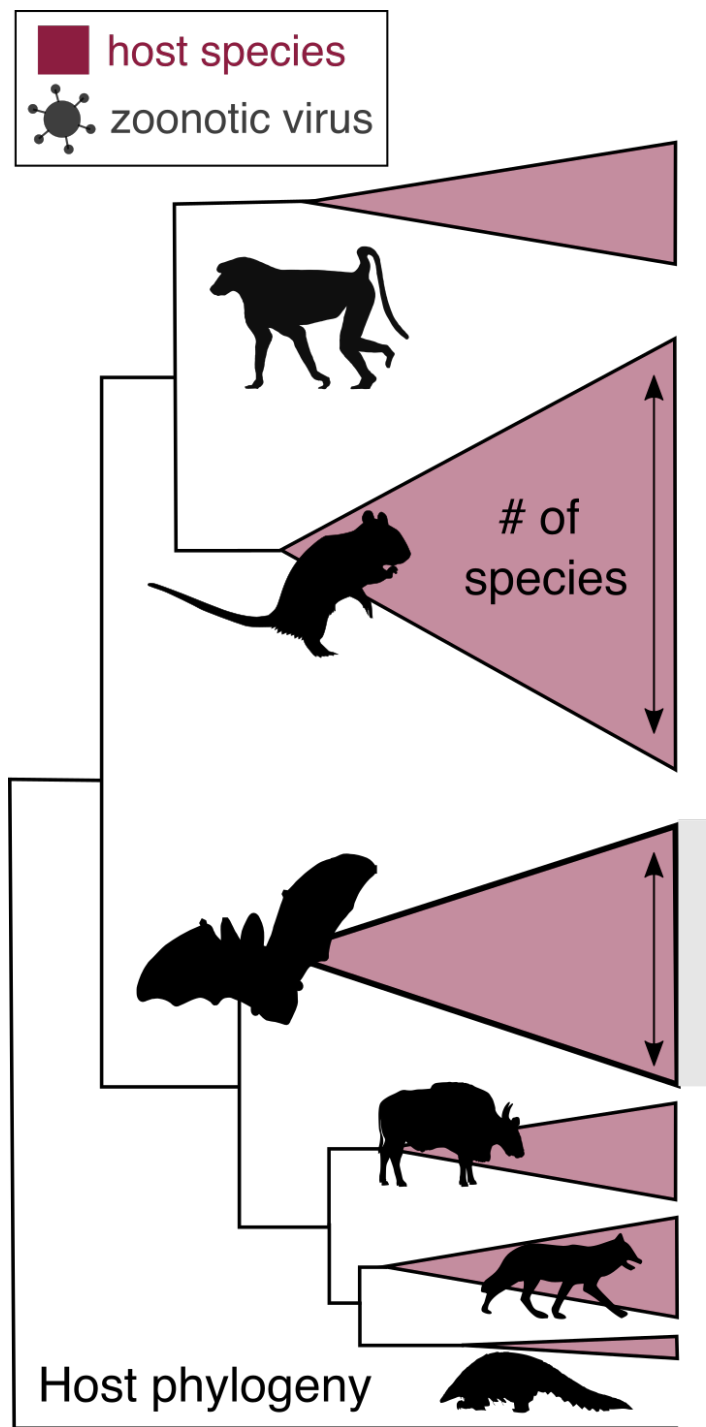
- **Novel coronavirus:** SARS-CoV-2
- **Likely host reservoirs:** mammals ➤ bats ➤ horseshoe bats (*Rhinolophus* sp.)



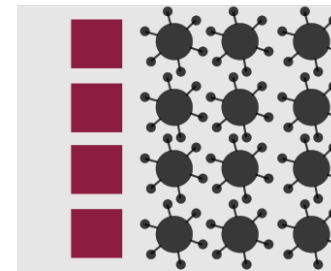
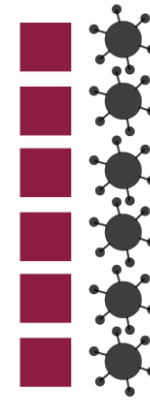
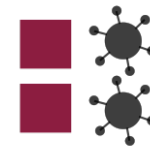
**Nancy Simmons**  
AMNH Mammal Curator  
Global bat expert  
[batnames.org](http://batnames.org)





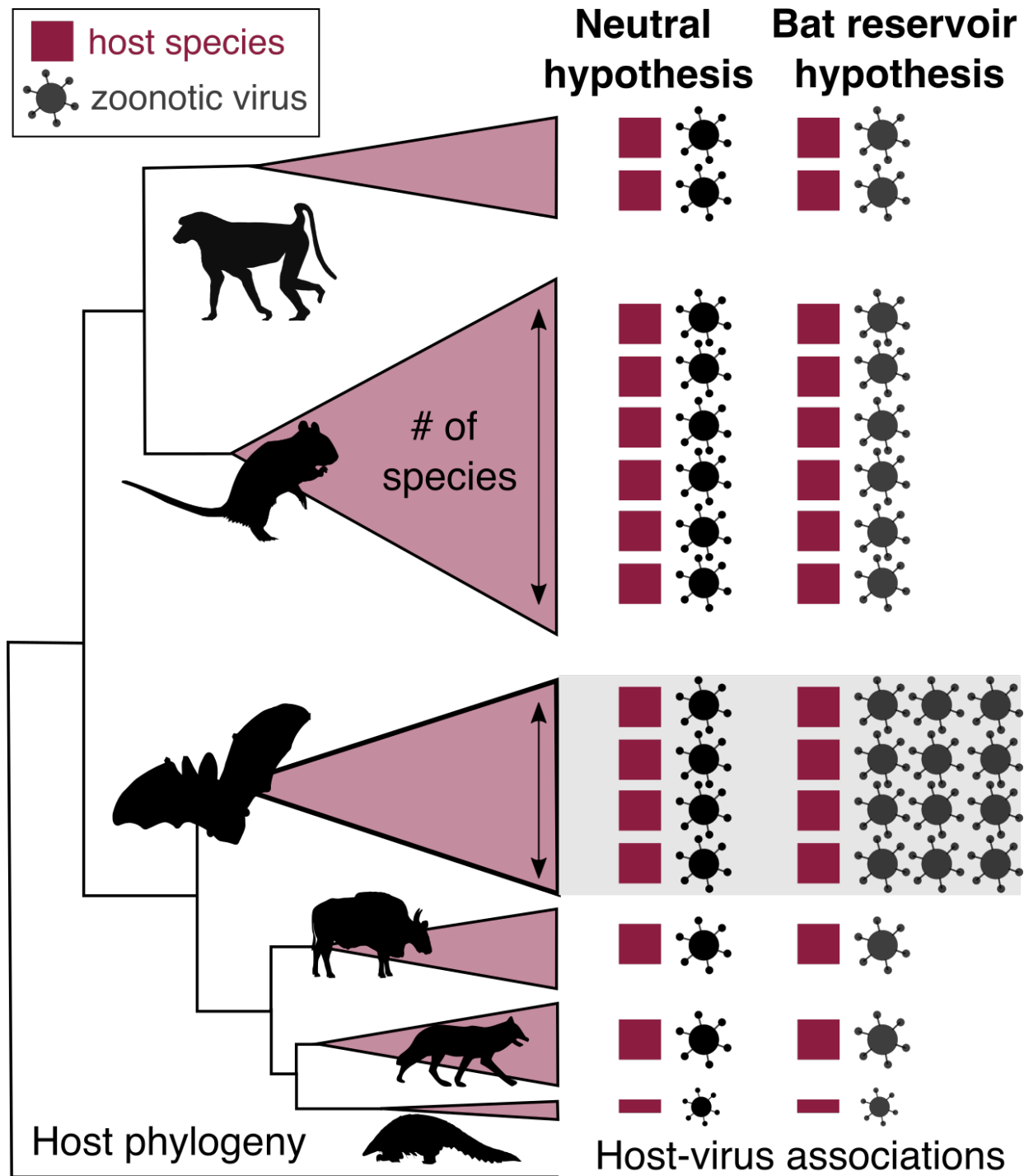


## Bat reservoir hypothesis



Host-virus associations







---

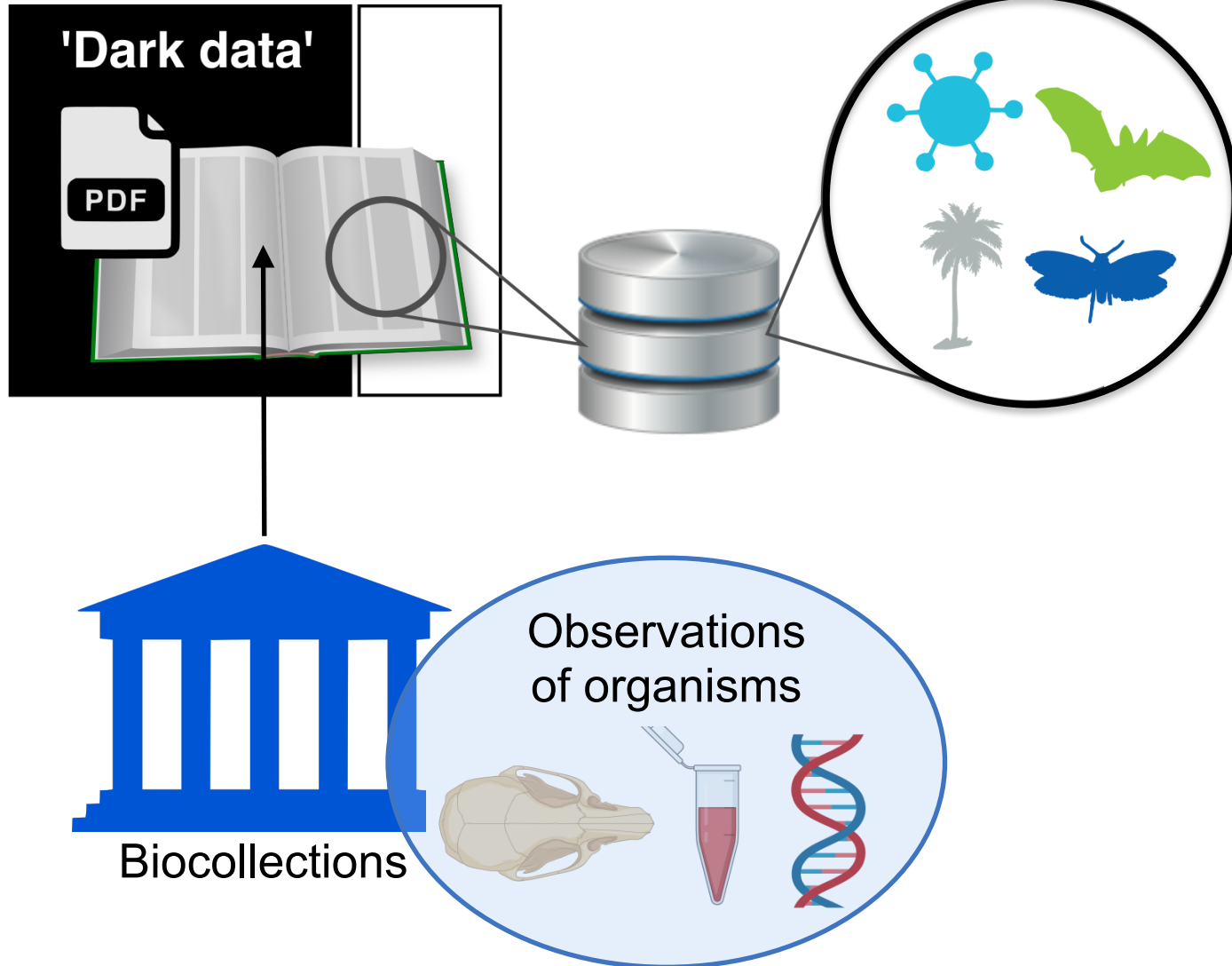
In a perfect world, **biocollections data** and  
**written knowledge** would already be  
digitized, extracted, and freely available  
for anyone to analyze...

---



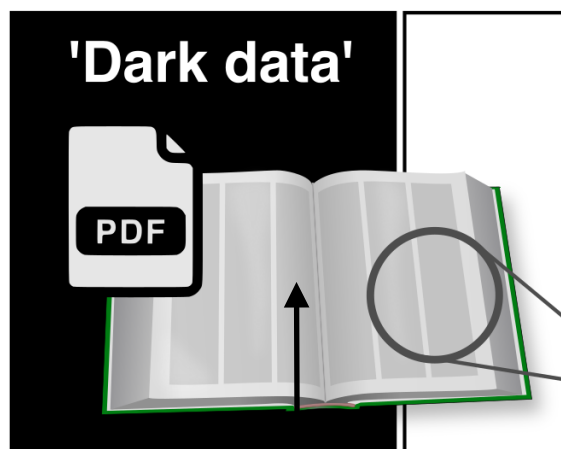
Written knowledge

Digital data

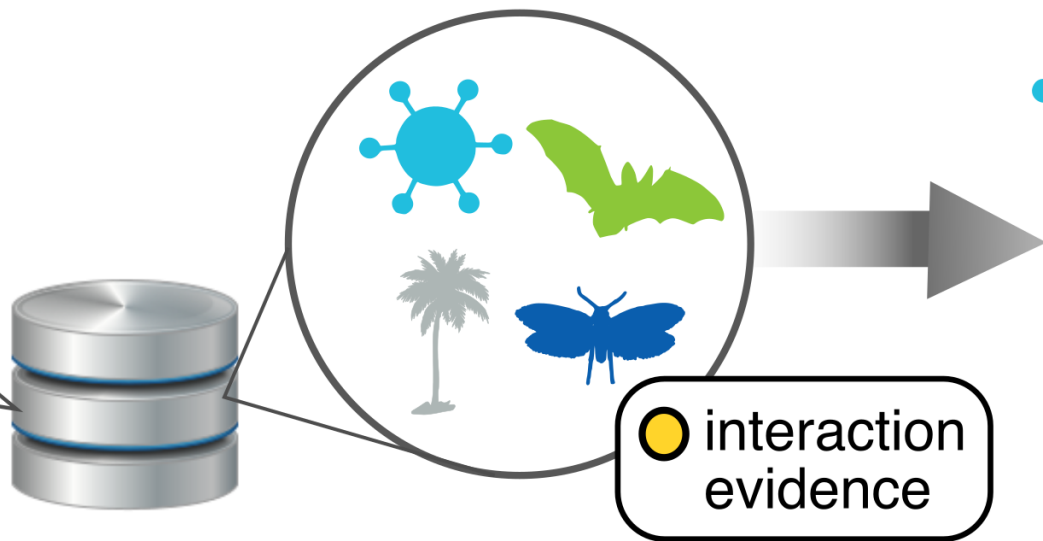




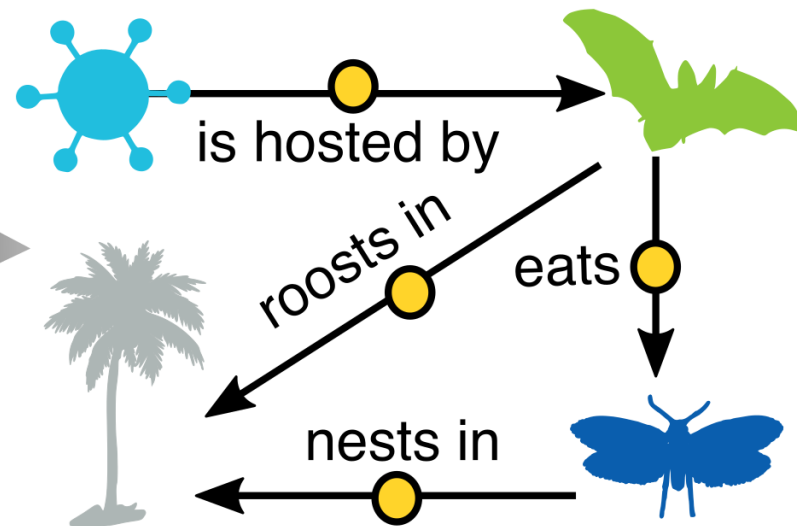
Written knowledge



Digital data

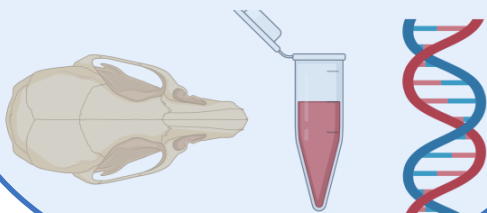


Host-virus knowledge graph



Biocollections

Observations  
of organisms

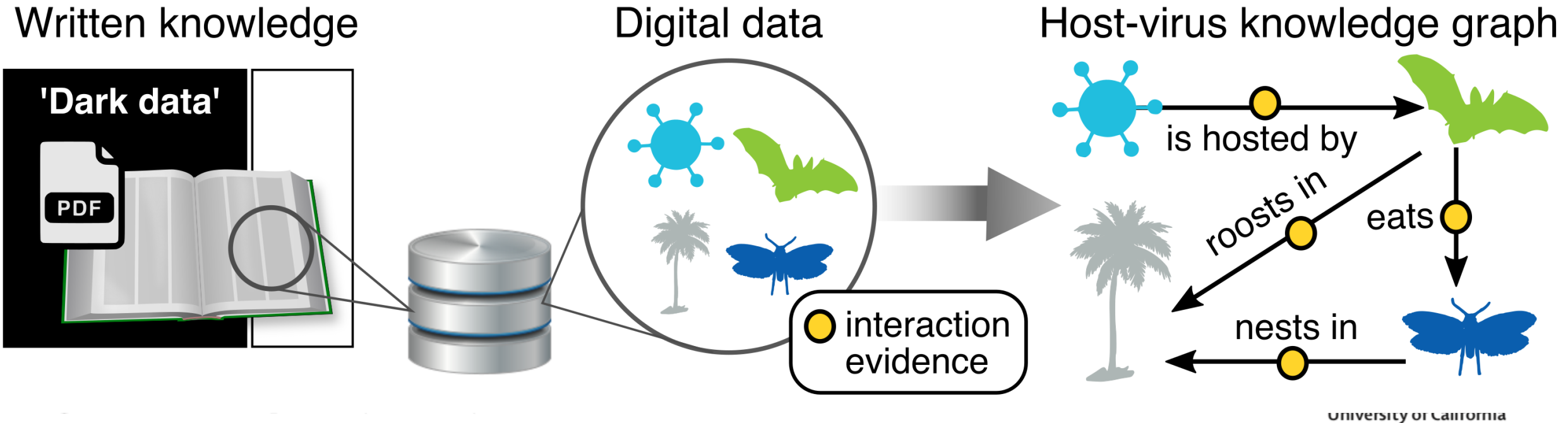




# Liberating host-virus knowledge from biological dark data

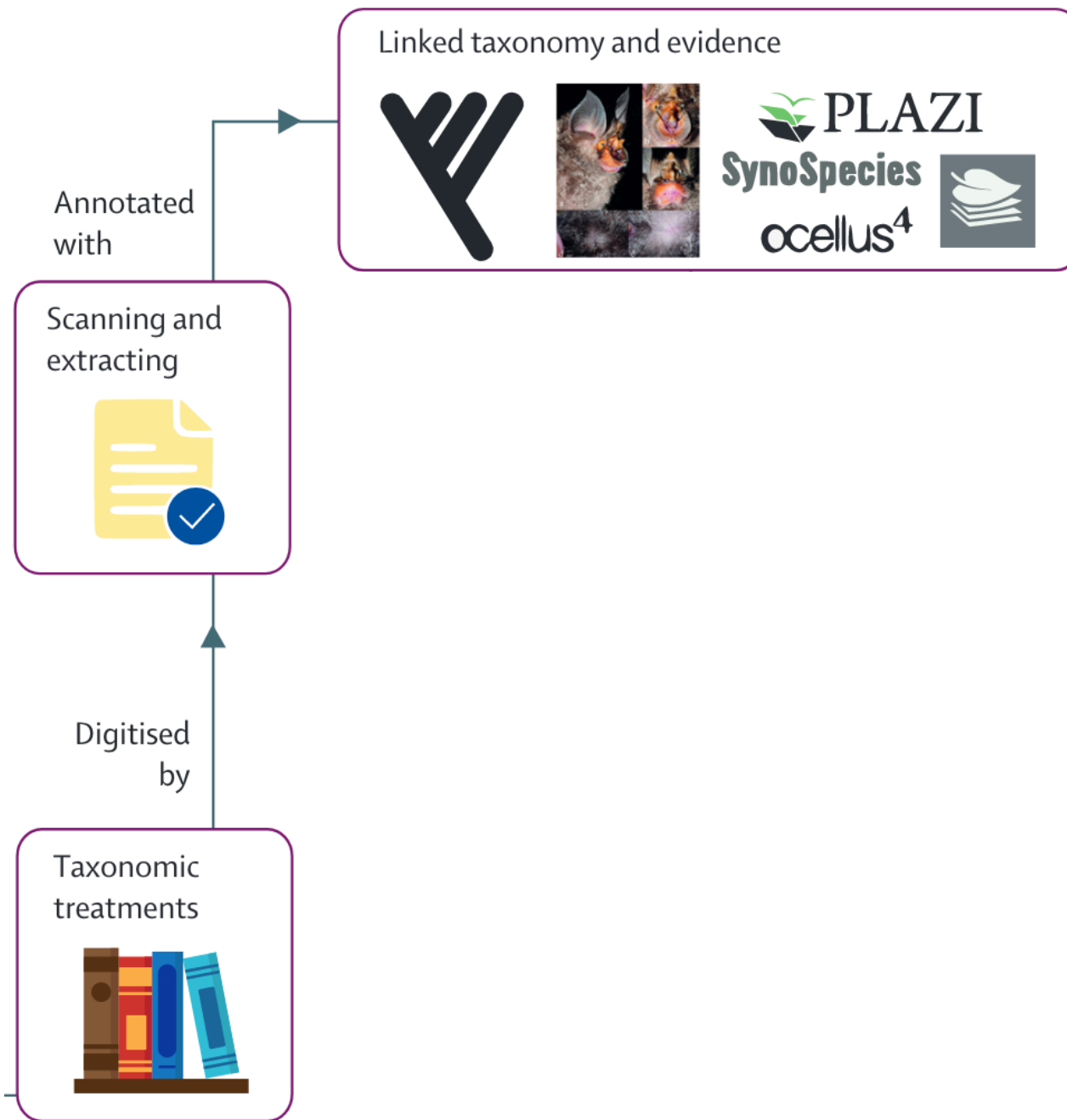


Nathan S Upham, Jorrit H Poelen, Deborah Paul, Quentin J Groom, Nancy B Simmons, Maarten P M Vanhove, Sandro Bertolino, DeeAnn M Reeder, Cristiane Bastos-Silveira, Atriya Sen, Beckett Sterner, Nico M Franz, Marcus Guidoti, Lyubomir Penev, Donat Agosti

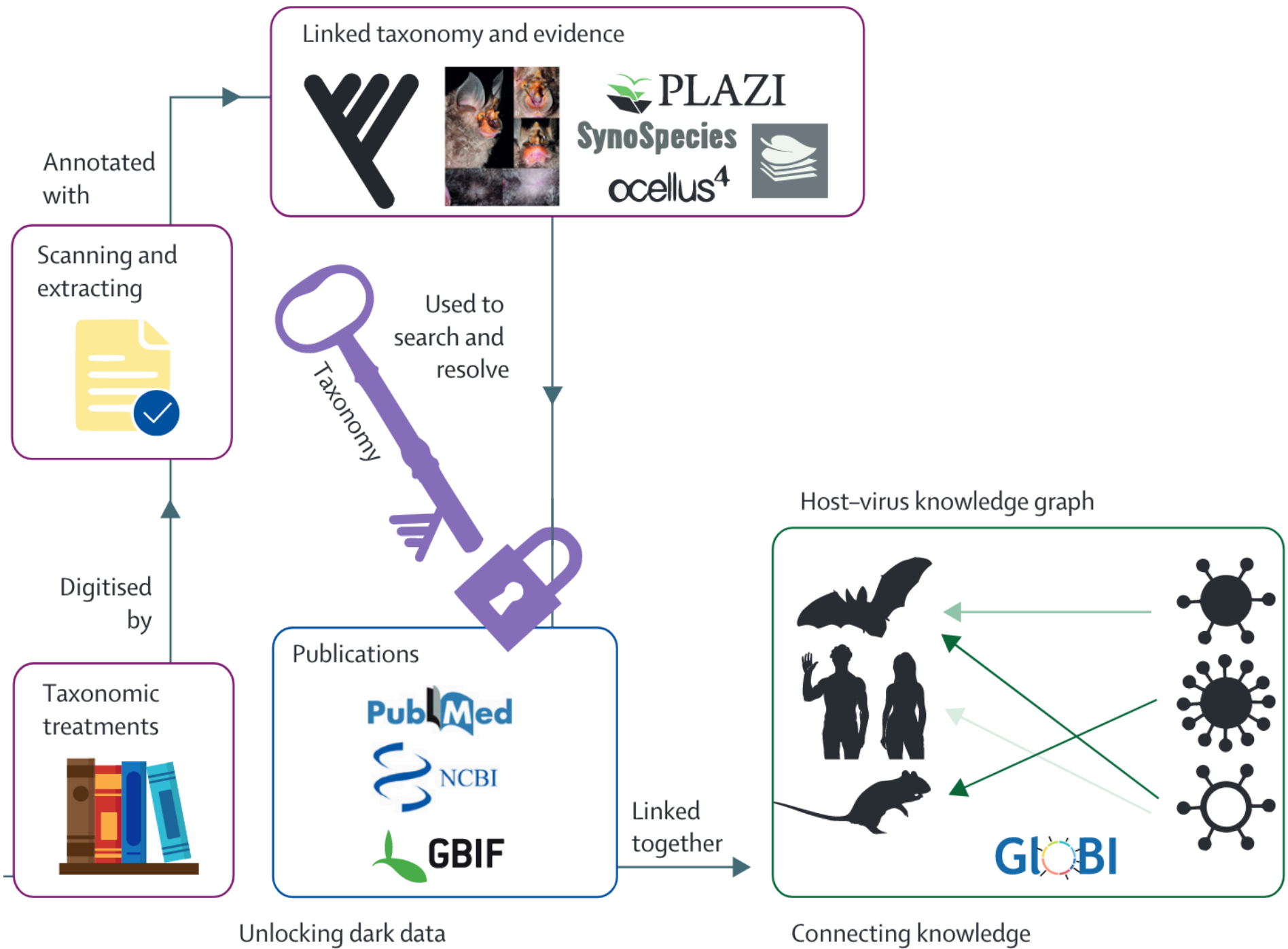


Upham et al. (2021) *The Lancet Planetary Health*

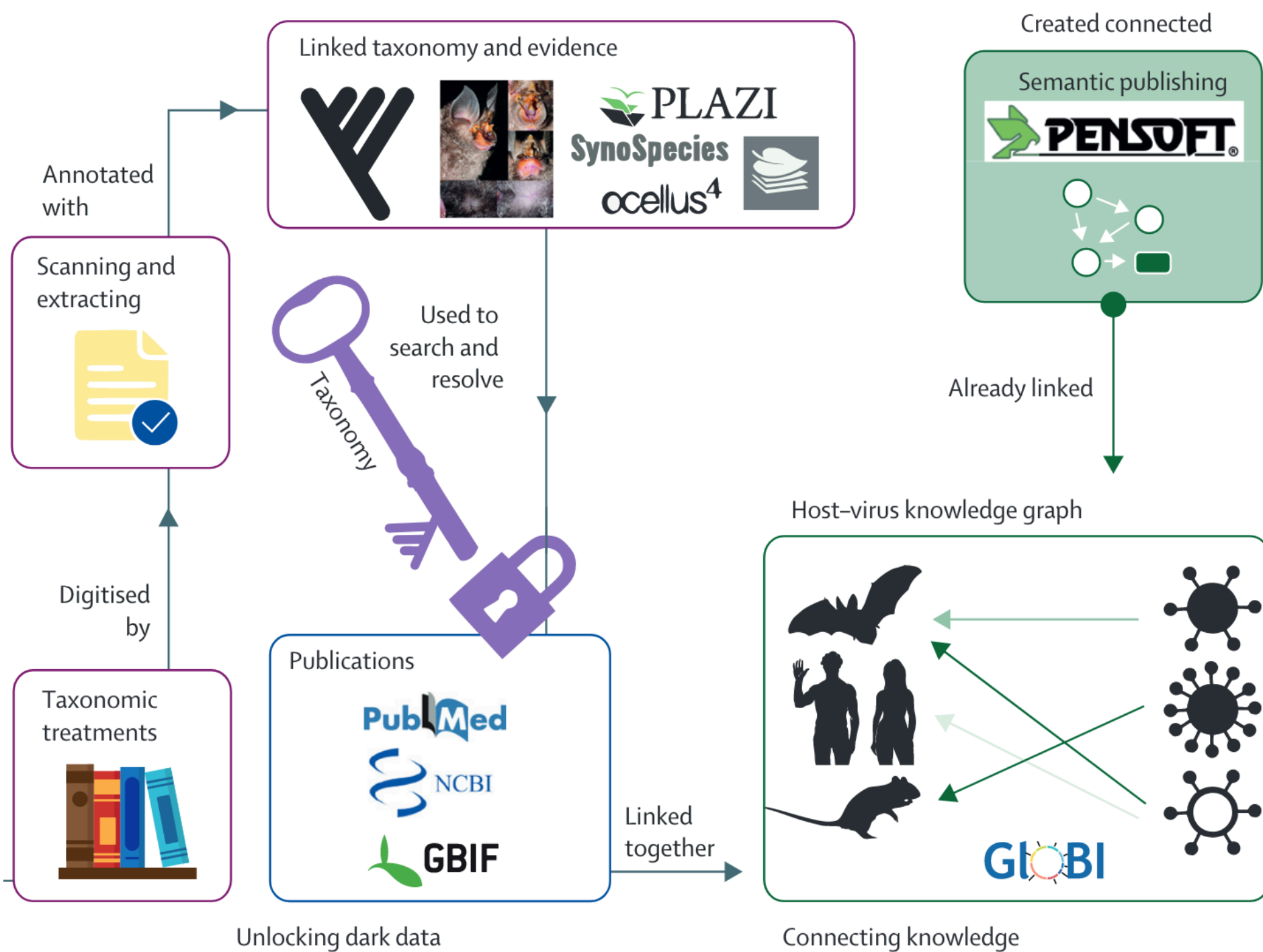














# U.S. National Institutes of Health — R21 funding



DeeAnn Reeder  
Bucknell Univ



Beckett Sterner  
ASU



Arvind Varsani  
ASU



Nico Franz  
ASU



Atriya Sen  
Univ New Orleans



Ángel Robles  
ASU PhD student



Jorrit Poelen  
GloBI

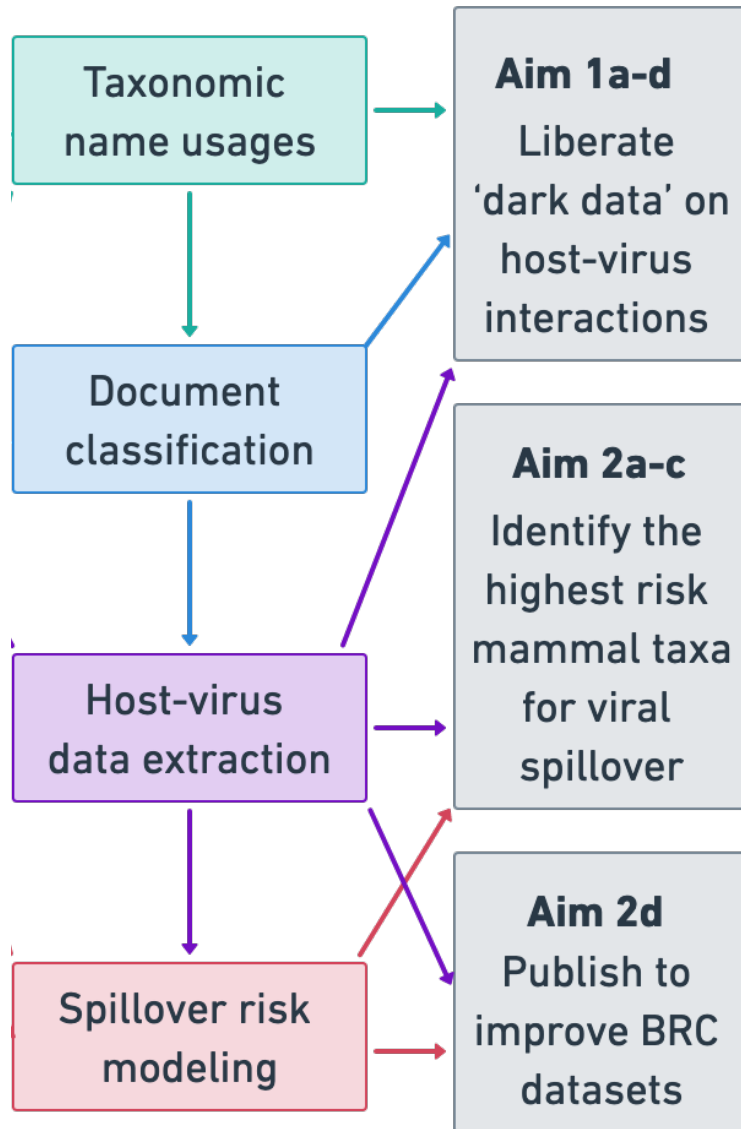


Donat Agosti  
Plazi

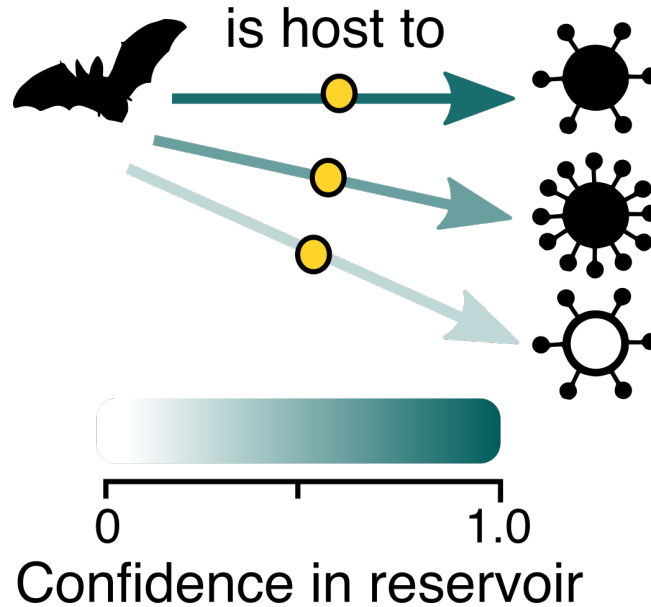


## SUBTEAMS

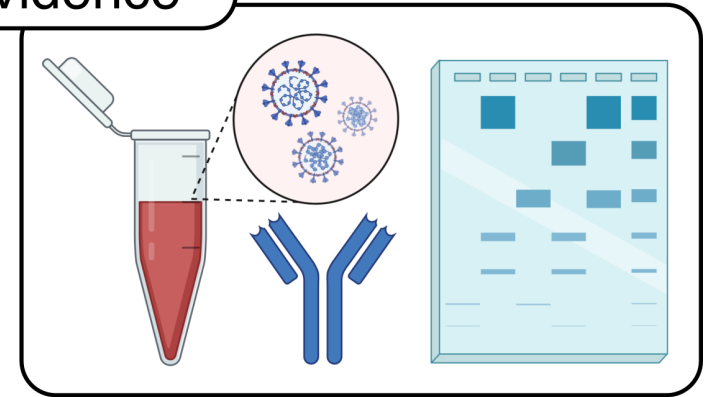
## Aims



# Confidence (or uncertainty) in the evidence of host reservoir status



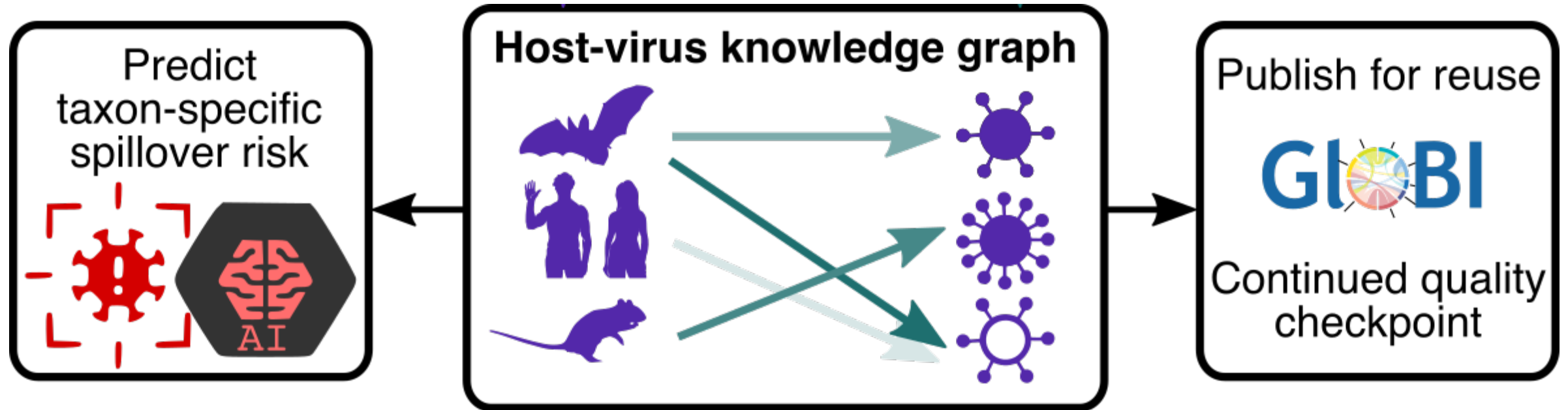
● interaction evidence





## *Understanding taxon-specific spillover risk*

---





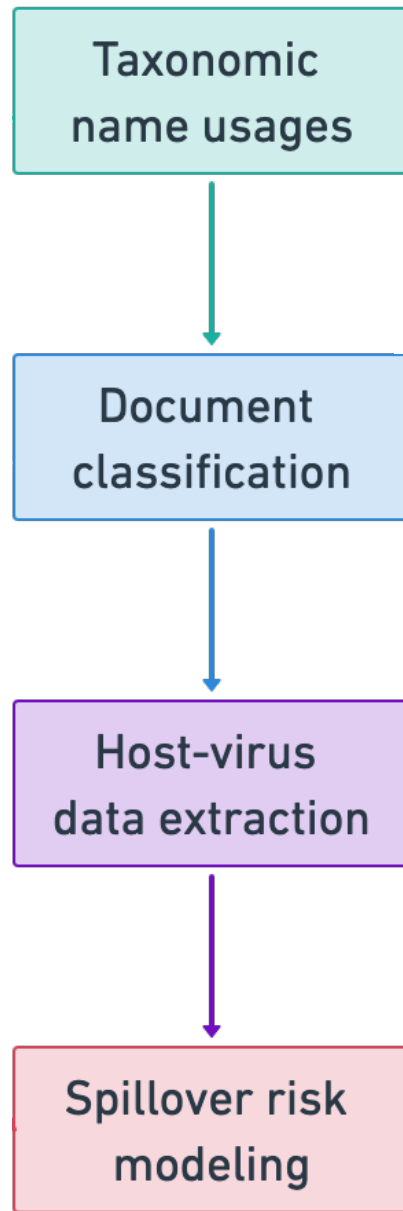
# I. Host-virus 'dark data'

## II. Data liberation & collection

- Progress so far
- Initial risk models



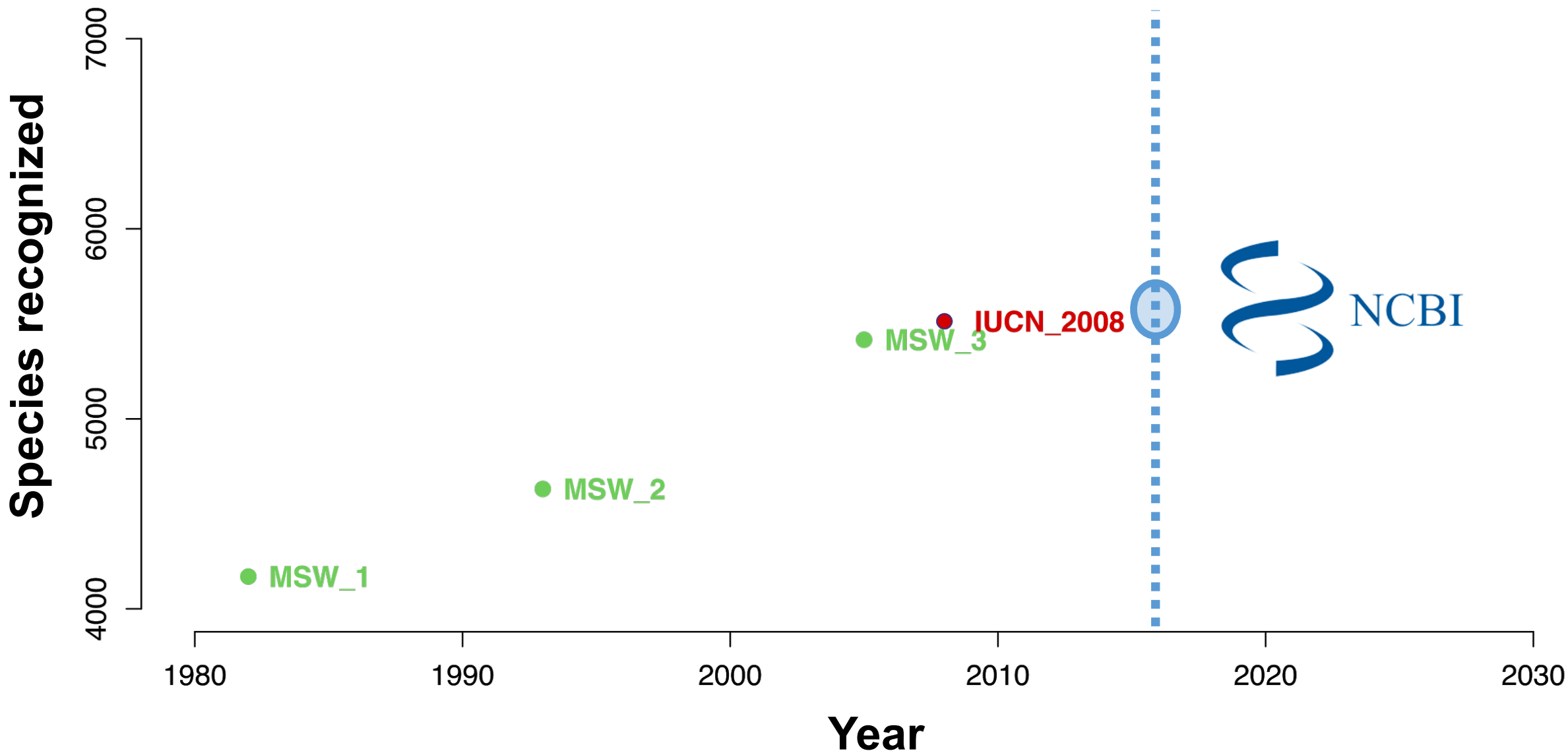




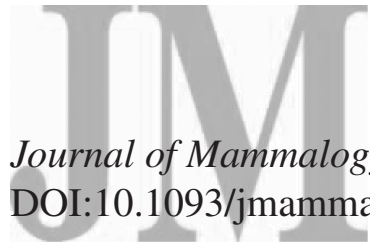


# 1. What are the mammal species?

MSW = Mammal Species of the World  
IUCN = Int'l Union of Cons. Nature







*Journal of Mammalogy*, 99(1):1–14, 2018  
DOI:10.1093/jmammal/gyx147

INVITED PAPER

# How many species of mammals are there?

CONNOR J. BURGIN,<sup>1</sup> JOCELYN P. COLELLA,<sup>1</sup> PHILIP L. KAHN, AND NATHAN S. UPHAM\*



## Since 2004

- 1,251 new species recognitions
- At least 172 unions
- 88 genera and 14 families newly recognized





<https://mammaldiversity.org>

ASM Mammal Diversity Database

American Society  
of Mammalogists

- Home
- Download the Database
- Explore Taxonomy
- Search Species
- About

Welcome to ASM's Mammal Diversity Database

Welcome!

The Mammal Diversity Database of the [American Society of Mammalogists](#) (ASM) is your home base for tracking the latest taxonomic changes to living and recently extinct (i.e., since ~1500 CE) species and higher taxa of mammals.

Here we are curating the taxonomic implications of new research publications in real time — with the goal of promoting rigorous study of mammal biodiversity worldwide.

6,495 species of  
recognized living  
mammals, and  
counting... (v1.9)

Taxa	MSW3 2005	MDD 2018	Current
Species			
Total	5,416	6,495	6,596
Extinct	75	96	101
Living	5,341	6,399	6,495
Domestic	0	16	19
Living wild	5,338	6,382	6,476
Genera	1,230	1,314	1,342
Families	153	167	167



Primates - Cercopithecidae  
*Rungwecebus kipunji*  
Kipunji  
TRB Davenport  
ASM - MIL








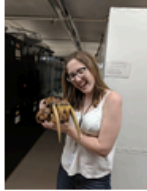



<https://mammaldiversity.org>

ASM Mammal Diversity Database

American Society  
of Mammalogists

Home Download the Database Explore Taxonomy Search Species About

### Our curation team

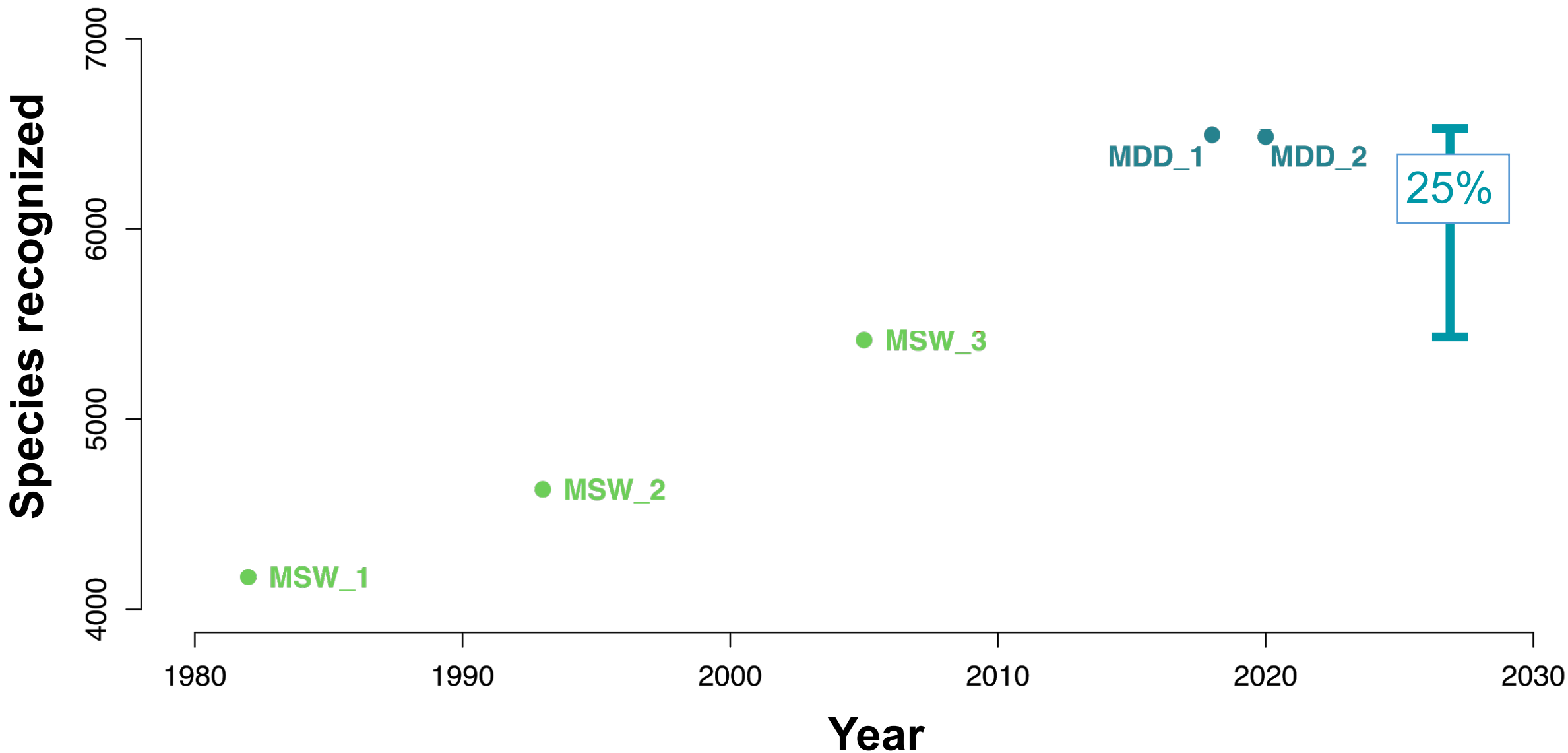
<a href="#">Nate Upham</a>	Connor Burgin	Jane Widness	Madeleine Becker	Camilla Parker	Schuyler Liphardt	David Huckaby
						
Chair, Biodiversity Committee	Student Research Assistant	Student Research Assistant	Student Research Assistant	Student Research Assistant	Student Web Developer	Chair, Mammal Images Library Committee
Arizona State University	University of New Mexico	Yale University	George Mason University	Central New Mexico Community College	University of New Mexico	California State University, Long Beach

6,495 species of  
recognized living  
mammals, and  
counting... (v1.9)



# 1. What are the mammal species?

MSW = Mammal Species of the World  
MDD = Mammal Diversity Database

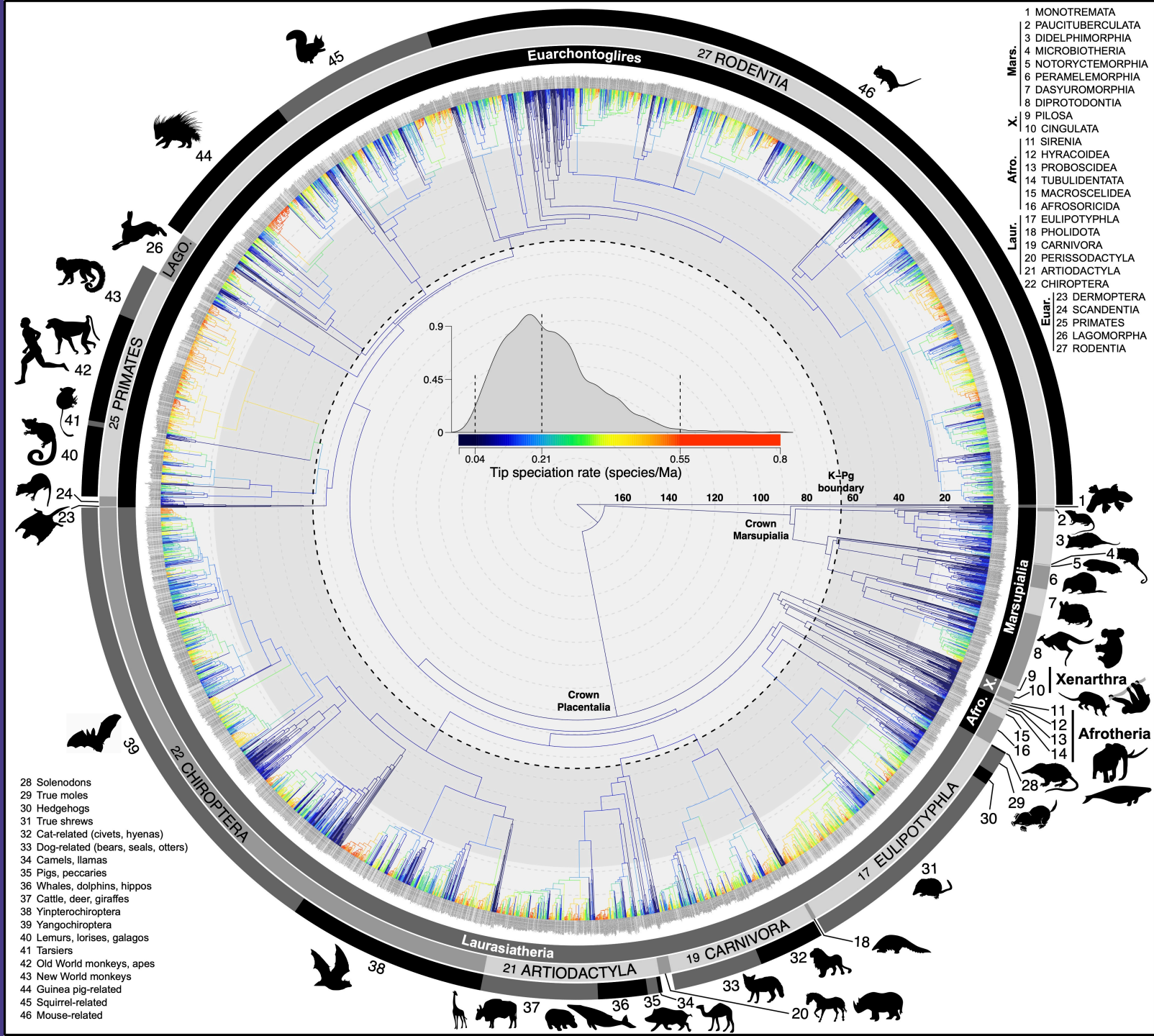




# Tree of life for all living mammals

## Consensus tree of 10,000

Upham et al. (2019)  
PLOS Biology

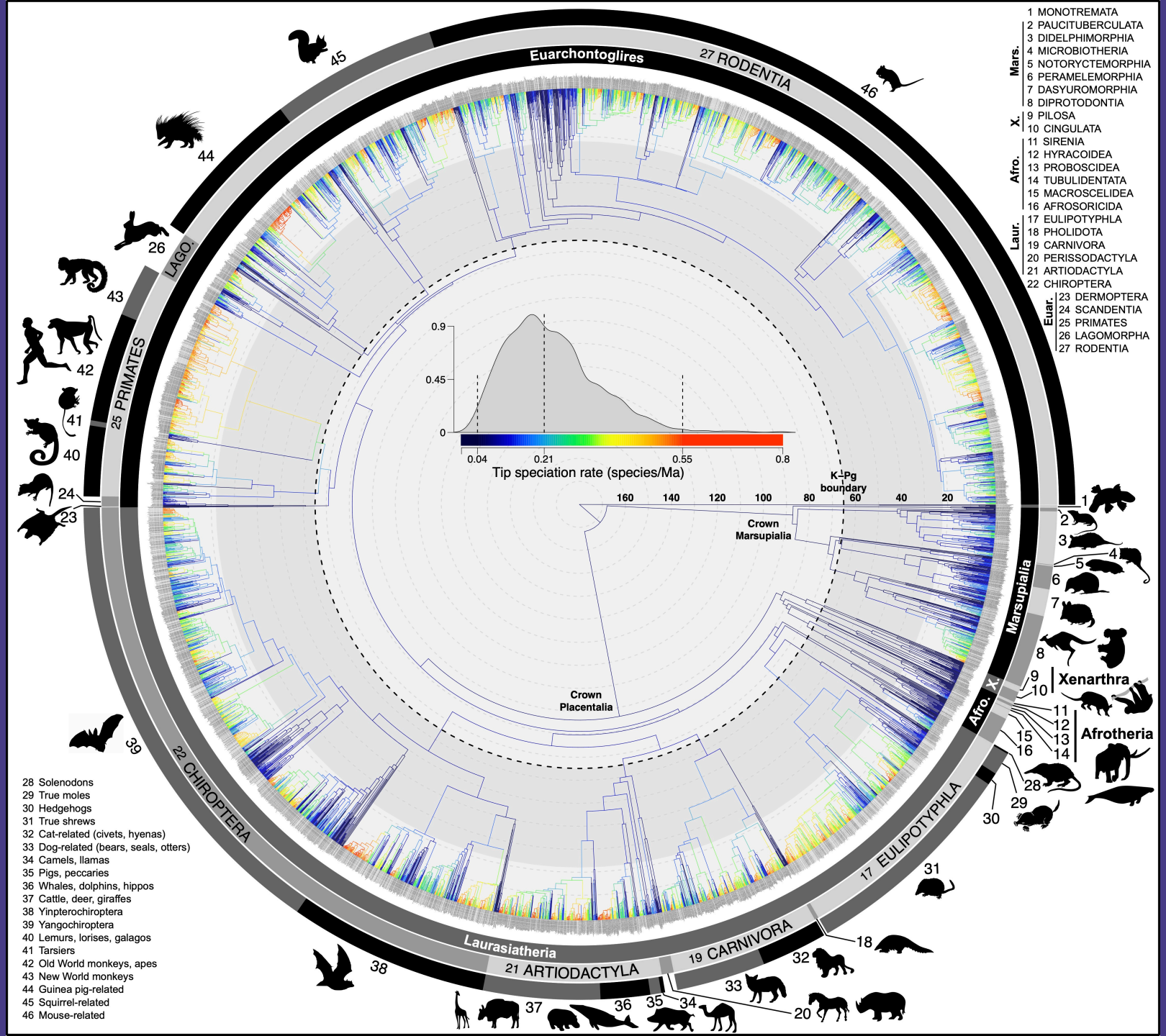




# Much remains unknown.

- 30% of species have no DNA sequenced
- Taxonomy continues to change rapidly
- Ecological data is unlinked to changing taxonomy

Upham et al. (2019)  
*PLOS Biology*



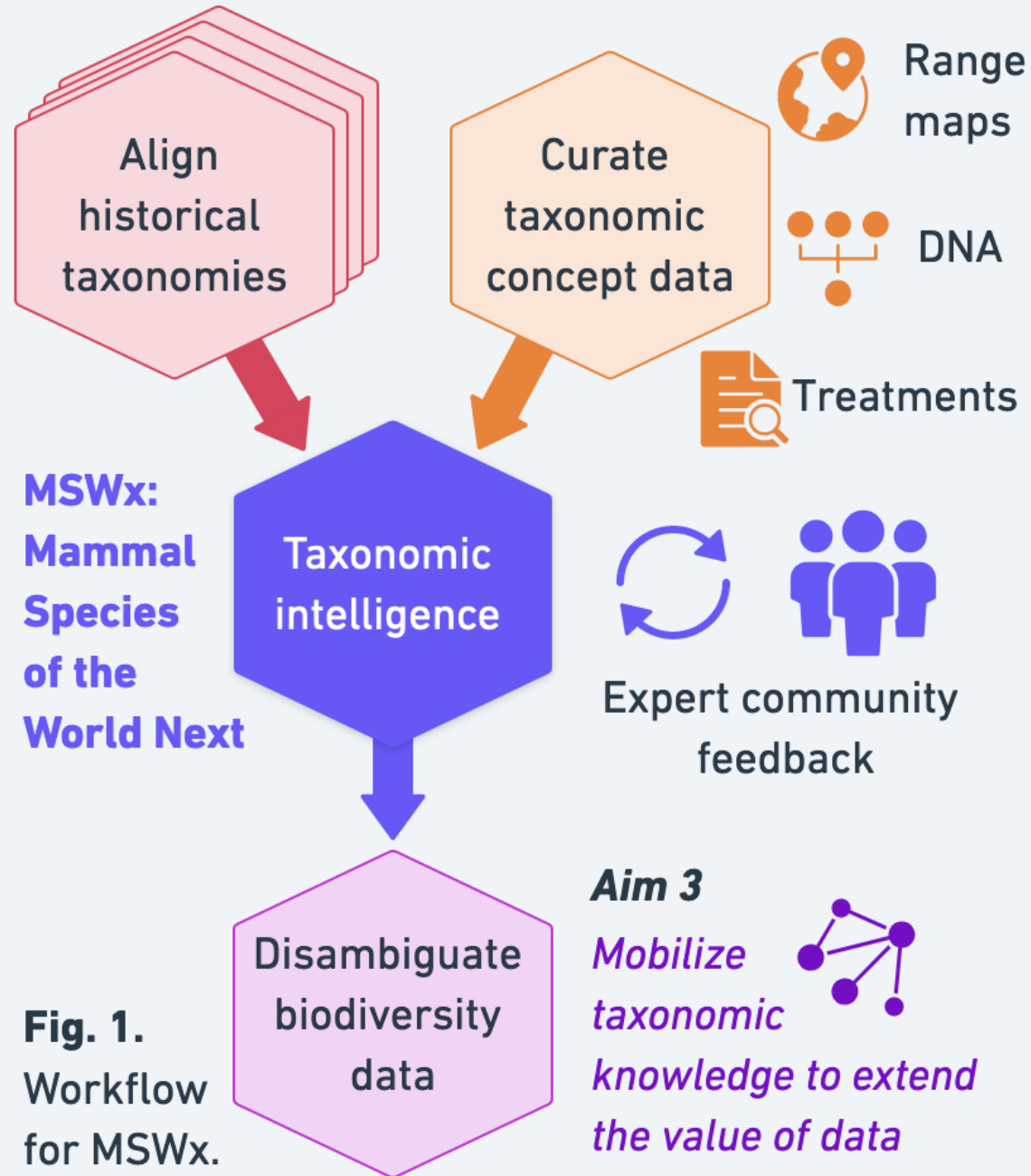


Mammal Diversity Database  
Mammal Species of the World  
Bats of the World



**Mammal Species  
of the World Next  
(MSWx)**

**Aim 1** *Know your history*   **Aim 2** *Link data & meaning*







***Peromyscus***  
(North America)

**52 species in 1993**  
**77 species in 2021**

## *Taxonomic intelligence*

Keeping track of the **meaning** of  
taxonomic names through time

### Species Meaning Artifacts (**SMArts**)



***Microcebus***  
(Madagascar)

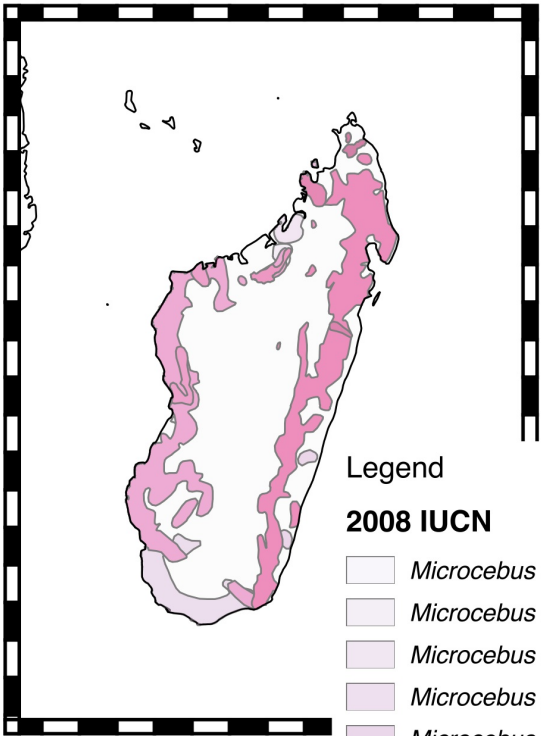
**2 species in 1993**  
**25 species in 2021**





**Microcebus**  
(Madagascar)

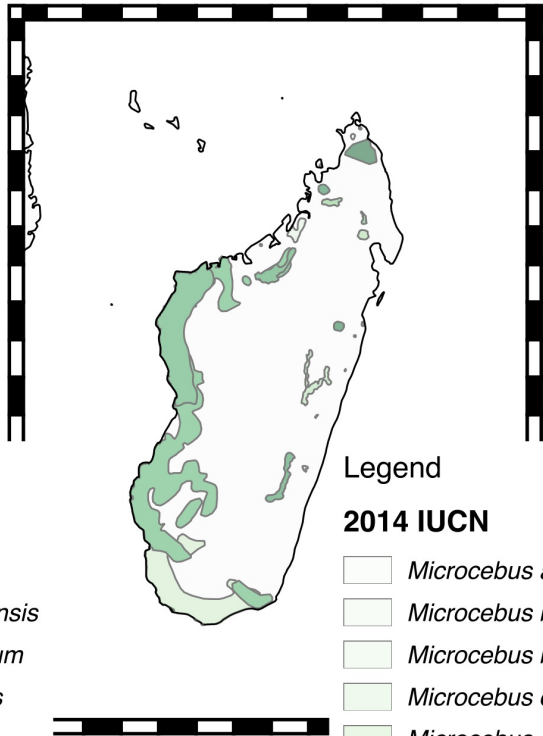
e.g., range maps  
as **SMArts**



Legend

**2008 IUCN**

- Microcebus berthae*
- Microcebus bongolavensis*
- Microcebus danfossorum*
- Microcebus griseorufus*
- Microcebus jollyae*
- Microcebus lehilahytsara*
- Microcebus mampiratra*
- Microcebus mittermeieri*
- Microcebus murinus*
- Microcebus myoxinus*
- Microcebus ravelobensis*
- Microcebus rufus*
- Microcebus sambiranensis*
- Microcebus simmonsii*
- Microcebus tavaratra*



Legend

**2014 IUCN**

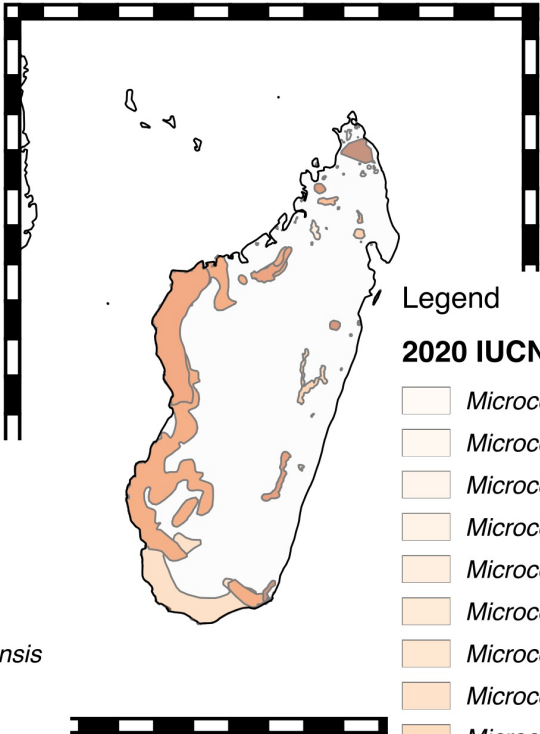
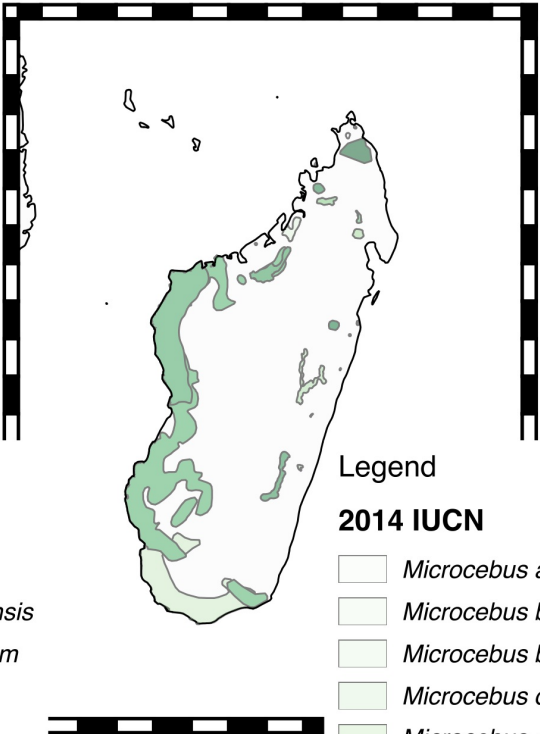
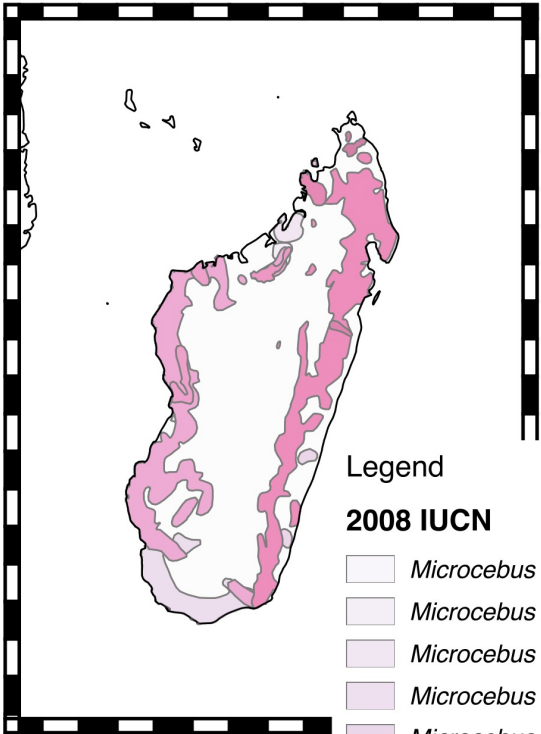
- Microcebus arnholdi*
- Microcebus berthae*
- Microcebus bongolavensis*
- Microcebus danfossi*
- Microcebus gerpi*
- Microcebus griseorufus*
- Microcebus jollyae*
- Microcebus lehilahytsara*
- Microcebus macarthurii*
- Microcebus mampiratra*
- Microcebus margotmarshae*
- Microcebus marohita*
- Microcebus mittermeieri*
- Microcebus murinus*
- Microcebus myoxinus*
- Microcebus ravelobensis*
- Microcebus rufus*
- Microcebus sambiranensis*
- Microcebus simmonsii*
- Microcebus tavaratra*





**Microcebus**  
(Madagascar)

e.g., range maps  
as **SMArts**

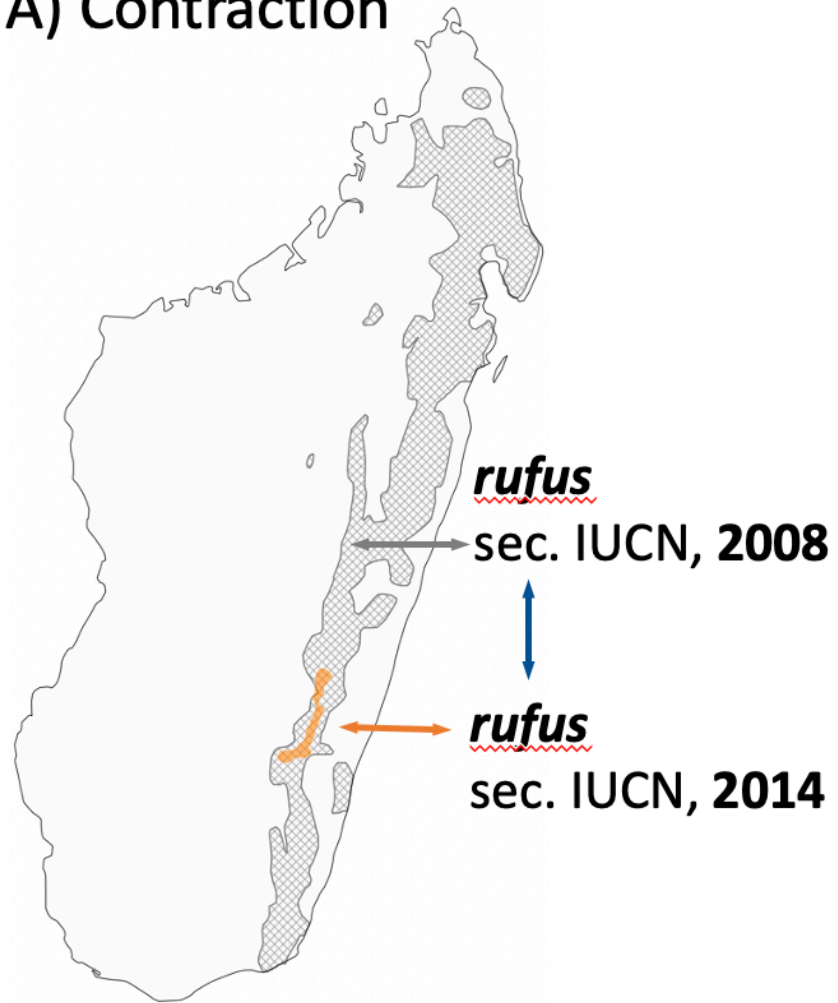






***Microcebus***  
(Madagascar)

### A) Contraction



**Species meanings**  
(name usages) change  
according to (“sec.”) different  
versions of the IUCN Red List



## Evidence of organisms



Organismal  
occurrence  
record



Sighting  
(photo, expert ID)

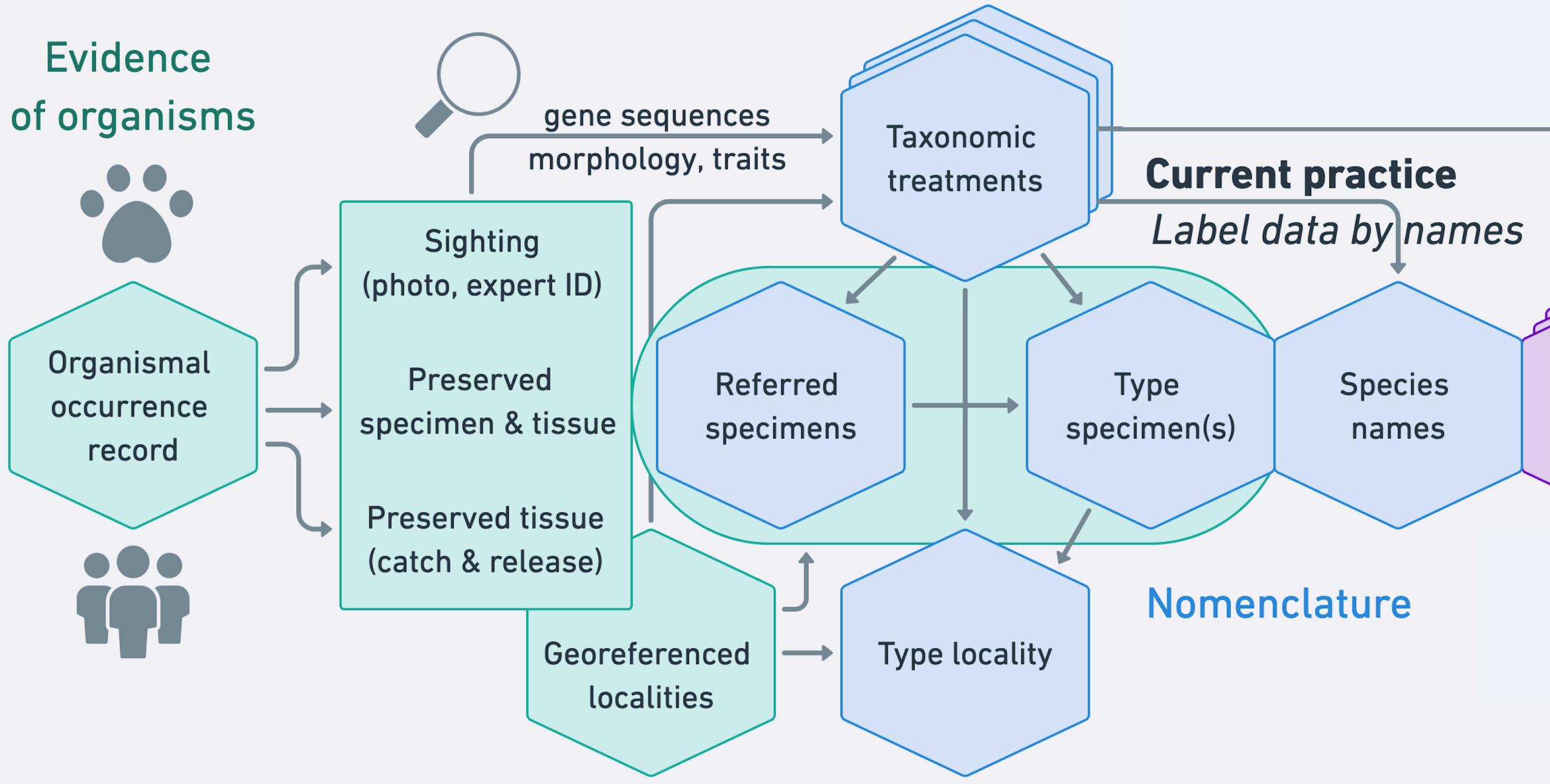
Preserved  
specimen & tissue

Preserved tissue  
(catch & release)

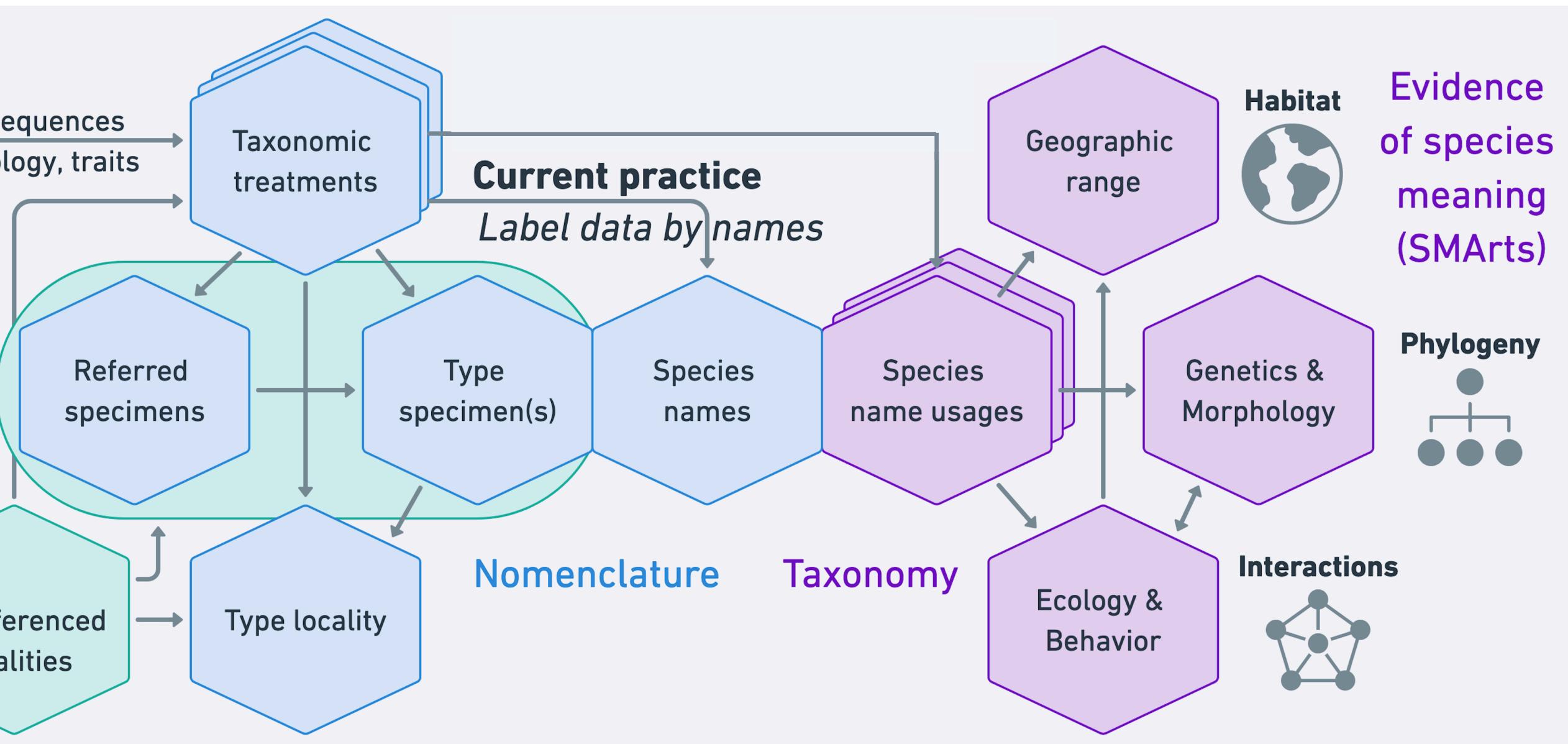
Georeferenced  
localities



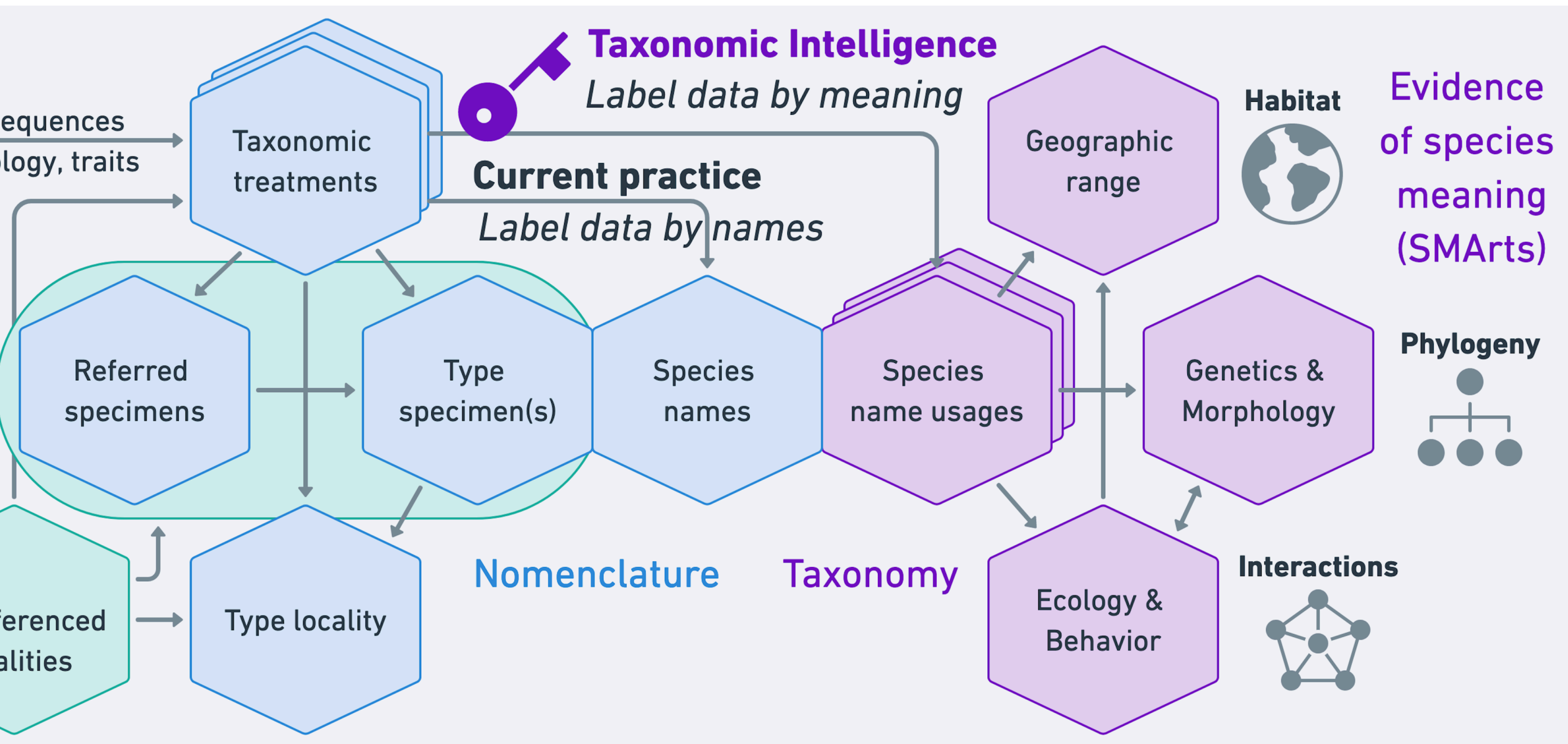
## Evidence of organisms



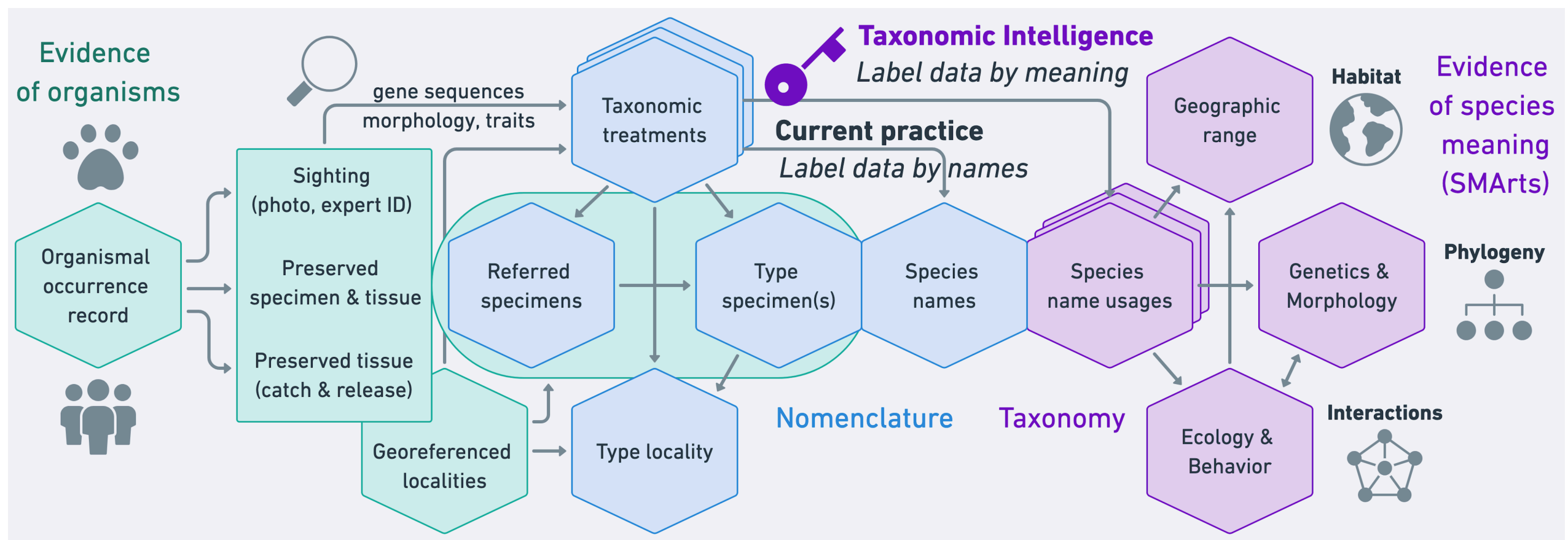








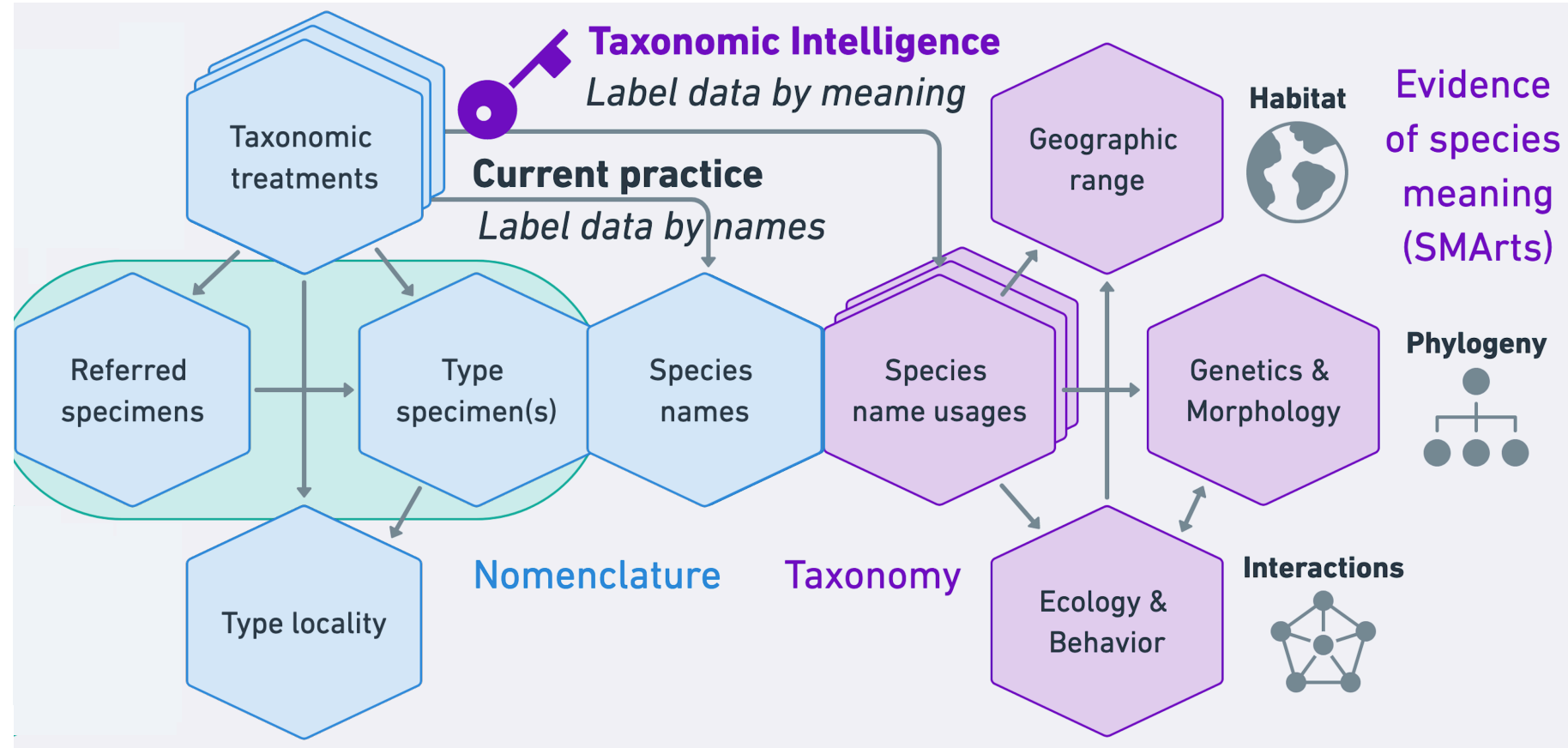




Currently, these updates only happen **manually and irregularly**, if ever



# Extended taxonomic curation



Curating  
**SMArt data**  
for species names

We need to be  
\*curating\* all of this!



**Extended  
taxonomic  
curation**

Pilot work in  
global Mammalia

**Mammal Species  
of the World Next  
(MSWx)**



Beckett Sterner  
ASU



Nico Franz  
ASU



Caleb Powell  
ASU



Laura Rocha Prado  
ASU



Ángel Robles  
ASU



DeeAnn Reeder  
Bucknell Univ



Nancy Simmons  
AMNH



Joe Cook  
UNM / Arctos



Connor Burgin  
UNM



Jonathan Rees  
EoL



Donat Agosti  
Plazi



# Taxonomic name usages

## Plazi and GloBI

- Digitized and extracted all taxonomic treatments in volumes 1-9 of the *Handbook of the Mammals of the World* series (HBW; 2009-2019).
- Ongoing QC

[TREATMENTBANK](#) [BIODIVERSITY LITERATURE REPOSITORY](#) [SERVICES](#) [HOW TO PARTICIPATE](#) [ABOUT](#)

**Burgin, Connor, 2019, Family Rhinolophidae (Horseshoe Bats)**

[Specimens](#)  
[Downloads](#)

Burgin, Connor, 2019, Family Rhinolophidae (Horseshoe Bats), *Handbook of the Mammals of the World – Volume 9 Bats*, Lynx Edicions, pp. 280-332

publication ID	
document provided by	Plazi (2020-04-08 08:13:16, last updated by Felipe 2022-05-07 21:19:04)

**Treatments (110)**

<b>Rhinolophidae</b>			260
<b>Rhinolophus alcyone</b>			281-282
<b>Rhinolophus odami</b>			281



# Taxonomic name usages

## Plazi and GloBI

- Digitized and extracted all taxonomic treatments in volumes 1-9 of the *Handbook of the Mammals of the World* series (HBW; 2009-2019).
- Ongoing QC

### Treatment

3 [View On](#). Halcyon Horseshoe Bat

#### **Rhinolophus alcyone**

French: **Rhinolophe alcyone** / German: Alcyon-Hufeisennase / Spanish: Herradura alción

Taxonomy. **Rhinolophus** aZcyon « Temminck, 1853, “ La rivière Boutry à la Guinée [= Boutry River , Guinea] .”

**Rhinolophus alcyone** is in the **landerl** species group with **R landerl** , **R guineensis** , **R blasli** , and **R lobatus** . It is genetically close to **R landerl** , although additional sampling is needed to fully resolve its placement Although no subspecies are recognized, there are distinct pelage differences between eastern and western populations and the Bioko Island population, suggesting taxonomic revision is needed. Specimens from Gabon might represent **R silvestris** , which requires additional investigation. Monotypic.

Distribution. W & C Africa from Senegal to Togo and from S Nigeria, S Cameroon, and SW Central African Republic to W Republic of the Congo, Equatorial Guinea and Bioko I, and Gabon, also patchily in Congo Basin, SW South Sudan, and W Uganda. [View Figure](#)

Descriptive notes. Head—body c. 57—68 mm, tail 18—32 mm, ear 19—25 mm, hindfoot 11-13 mm, forearm 48-56 mm; weight 14—23 g. Typical dorsal pelage is dark to medium brown (hairs are



# Taxonomic name usages

<https://github.com/jhpoelen/hmw>

## Plazi and GloBI

- Digitized and extracted all taxonomic treatments in volumes 1-9 of the *Handbook of the Mammals of the World* series (HBW; 2009-2019).
- Ongoing QC

Search or jump to... Pull requests Issues Codespaces Marketplace Explore

Public

Watch 2 Fork 1

<> Code Issues 4 Pull requests Actions Projects Security Insights

main 1 branch 4 tags Go to file Add file <> Code

jhpoelen update hmw.csv as discussed in #10 2d6466c on Sep 12 33 commits

.gitignore	add make script	5 months ago
LICENSE	Initial commit	5 months ago
README.md	add description and generation of names.txt	4 months ago
check.jq	add checks related to #7	4 months ago
check.sh	make check.sh executable	4 months ago
hmw-prov.nq	update hmw.csv as discussed in #10	2 months ago
hmw-sample-pretty.json	update hmw.csv as discussed in #10	2 months ago
hmw-sample.csv	update hmw.csv as discussed in #10	2 months ago
hmw-sample.json	update hmw.csv as discussed in #10	2 months ago
hmw-volume-1.csv	update hmw.csv as discussed in #10	2 months ago
hmw-volume-2.csv	update hmw.csv as discussed in #10	2 months ago
hmw-volume-3.csv	update hmw.csv as discussed in #10	2 months ago
hmw-volume-4.csv	update hmw.csv as discussed in #10	2 months ago
hmw-volume-5.csv	update hmw.csv as discussed in #10	2 months ago
hmw-volume-6.csv	update hmw.csv as discussed in #10	2 months ago
hmw-volume-7.csv	update hmw.csv as discussed in #10	2 months ago

About

(experimental) Machine readable version of Handbook of the Mammals of the World

Readme

CC0-1.0 license

1 star

2 watching

1 fork

Releases 4

0.4 Latest on Sep 12

+ 3 releases

Packages

No packages published

Contributors 2

jhpoelen Jorrit Poelen

myrmoteras Donat Agosti



# Taxonomic name usages

<https://github.com/jhpoelen/hmw>

## Plazi and GloBI

- Digitized and extracted all taxonomic treatments in volumes 1-9 of the *Handbook of the Mammals of the World* series (HBW; 2009-2019).
- Ongoing QC

The screenshot shows the GitHub interface for the repository 'jhpoelen / hmw'. The file 'hmw-sample.csv' is selected, showing its commit history and a preview of its contents. The file is 75.1 KB and contains 11 lines of code. The preview shows a table with taxonomic data.

name	interpretedGenus	interpretedSpecies	interpretedAuthorityName	interpretedAuthorityYear	commonNames
Geomys pinetis	Geomys	pinetis	Rafinesque	1817	Gaufre des pinédes @fr   Stidostliche Taschenratte @de   Tuza hirsu
Heterogeomys hispidus	Heterogeomys	hispidus			Gaufre hérissé @fr   Borstige Taschenratte @de   Tuza hirsu
Orthogeomys grandis	Orthogeomys	grandis			Grand Gaufre @fr   Riesentaschenratte @de   Tuza gigante
Zygogeomys trichopus	Zygogeomys	trichopus	Merriam	1895	Gaufre du Michoacan @fr   Michoacéan-Taschenratte @de
Heterogeomys cavator	Heterogeomys	cavator			Gaufre du Chiriqui @fr   Chiriqui-Taschenratte @de   Tuza d
Heterogeomys lanius	Heterogeomys	lanius			Gaufre de Xuchil @fr   Wollige Taschenratte @de   Tuza gran
Heterogeomys dariensis	Heterogeomys	dariensis	Goldman	1912	Gaufre du Darien @fr   Darien-Taschenratte @de   Tuza de D
Geomys jugossicularis	Geomys	jugossicularis	Hooper	1940	Gaufre de Hall @fr   Hall-Taschenratte @de   Tuza de Hall @
Geomys bursarius	Geomys	bursarius	Shaw	1800	Gaufre brun @fr   Flachland-Taschenratte @de   Tuza de Ilar
Geomys knoxjonesi	Geomys	knoxjonesi	Baker & Genoways	1975	Gaufre de Jones @fr   Jones-Taschenratte @de   Tuza de J



Go



6 result(s), took 5.33s



Rhinolophus alcyone Temminck, 1852

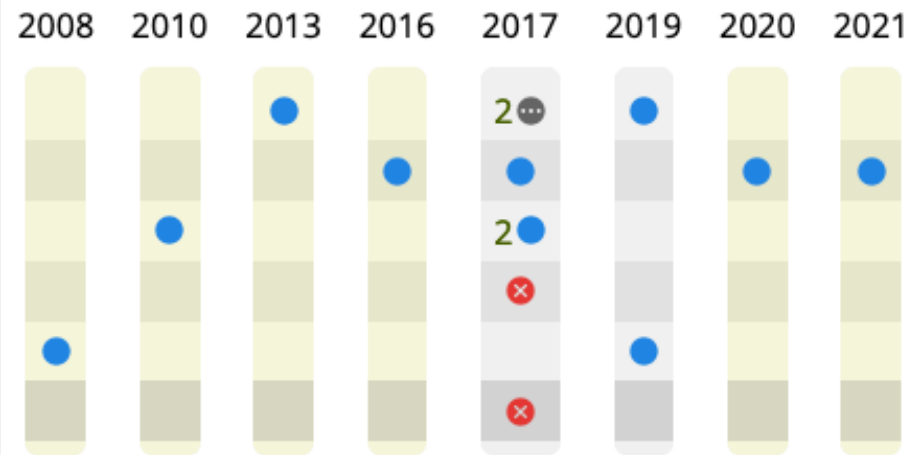
Rhinolophus alcyone Temminck, 1853

Rhinolophus blasii Peters, 1867

Rhinolophus alcyone alcyone Temminck, 1852

Rhinolophus blasii Peters, 1866

Rhinolophus blasii empusa Andersen, 1904



Animalia Rhinolophus alcyone



GBIF ID: 2432630



## Rhinolophus alcyone Temminck 1852

► Justification

► Treatments (●●●●●)

Phylum	Class	Order	Family
Chordata	Mammalia	Chiroptera	Rhinolophidae

## Rhinolophus alcyone Temminck 1853

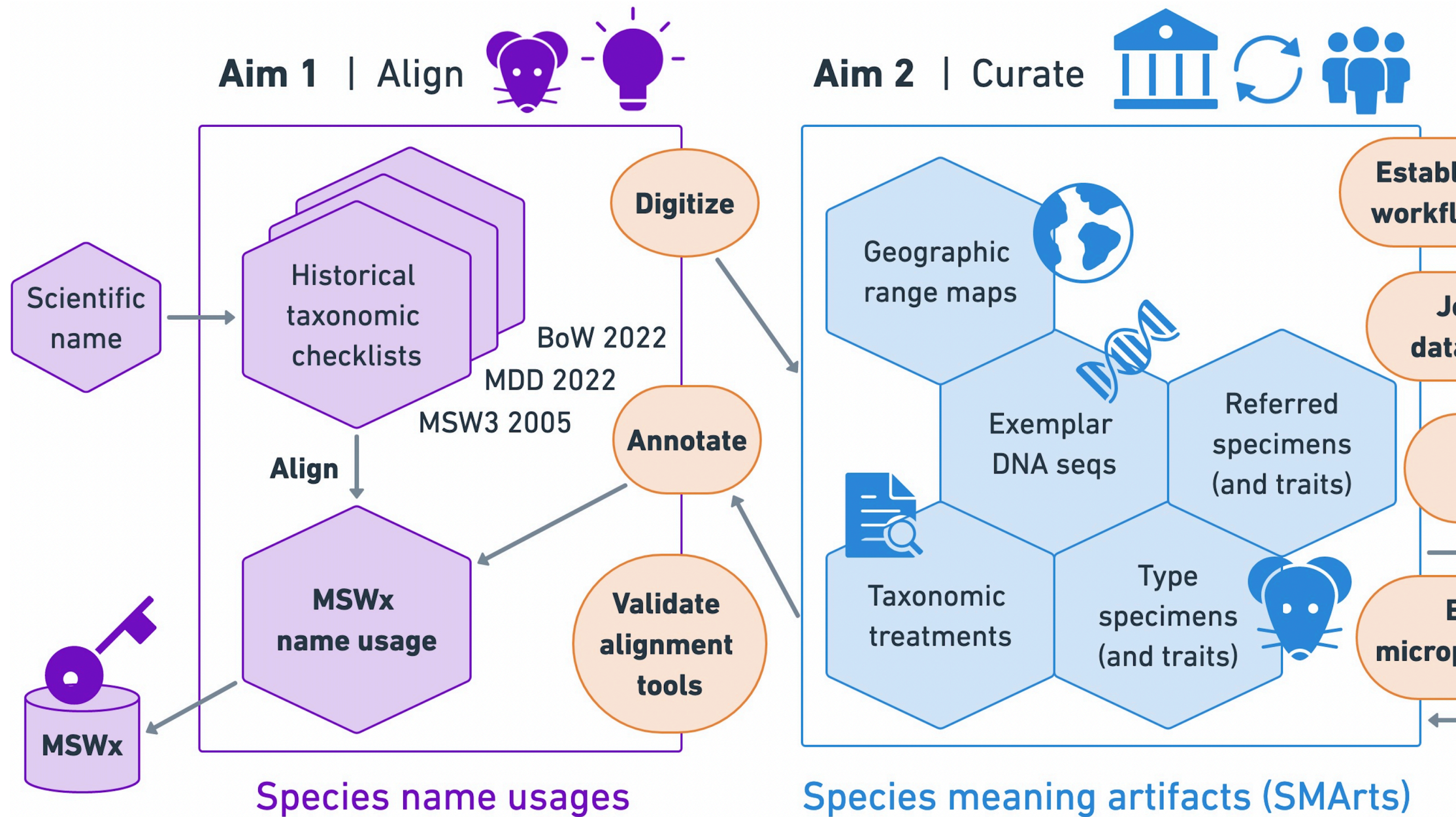
► Justification

► Treatments (●●●●●)

Phylum	Class	Order	Family
Chordata	Mammalia	Chiroptera	Rhinolophidae



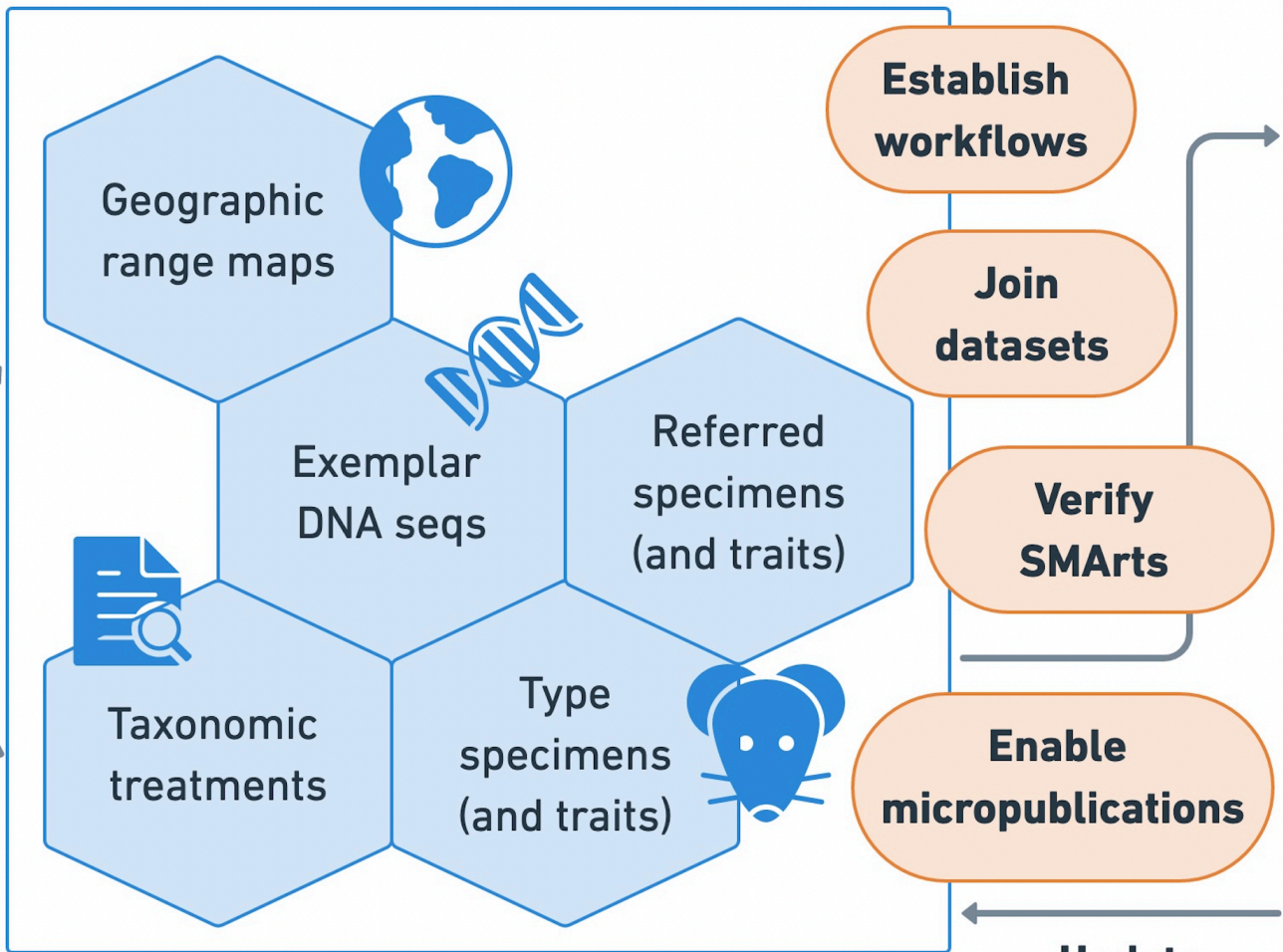
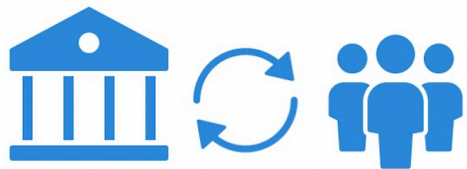
# Community curation of taxonomic meaning





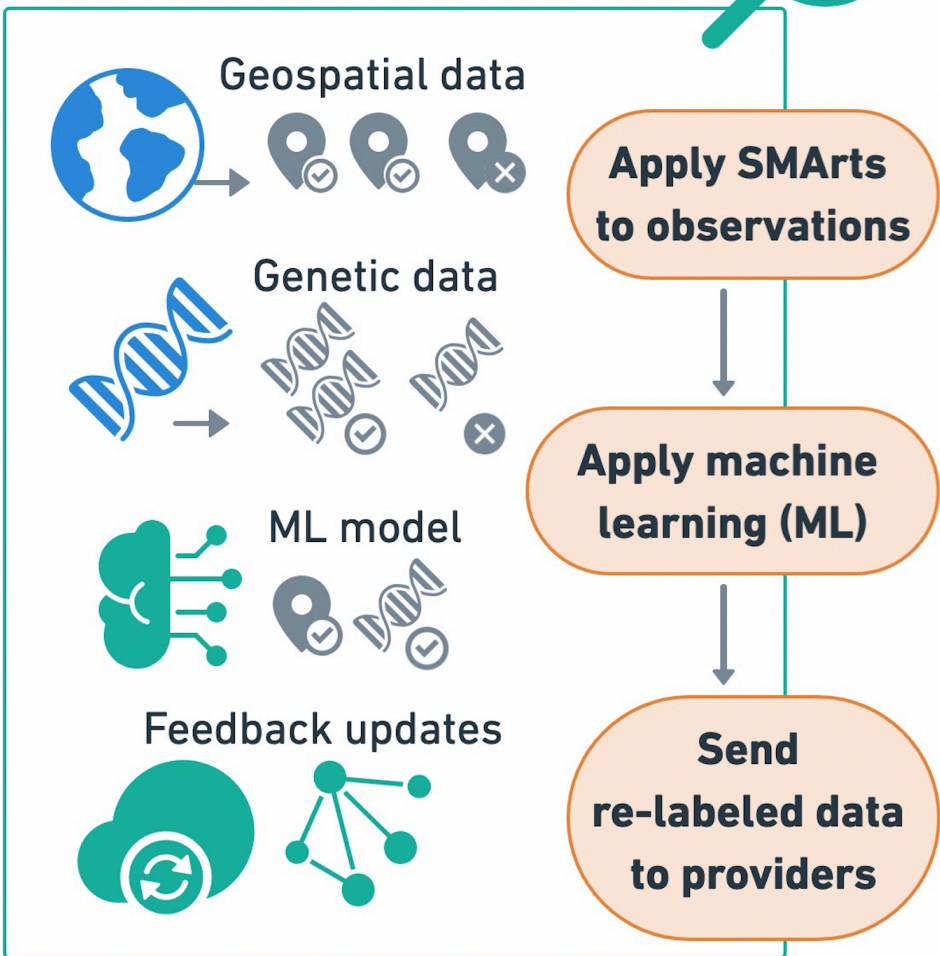
# Community curation of taxonomic meaning

## Aim 2 | Curate



Species meaning artifacts (SMARts)

## Aim 3 | Disambiguate

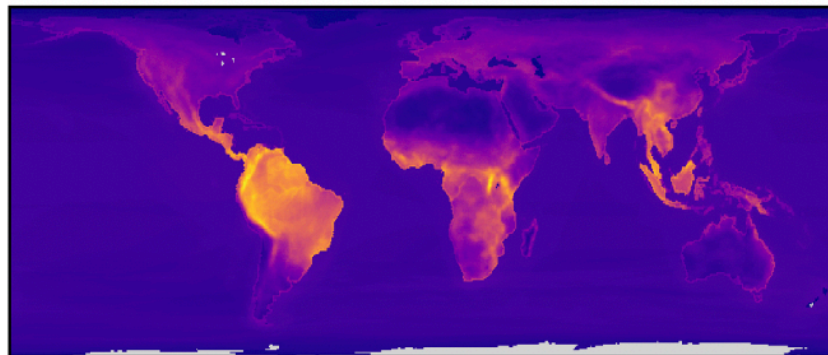


Biodiversity observations

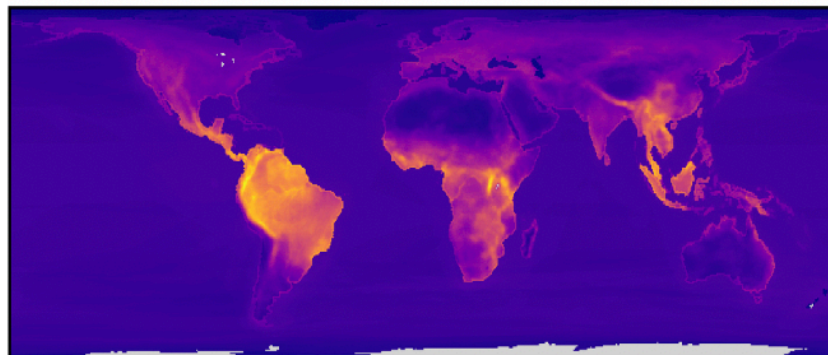


# Community curation of range maps

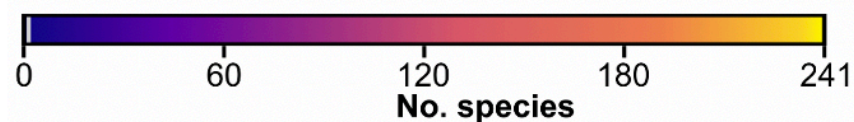
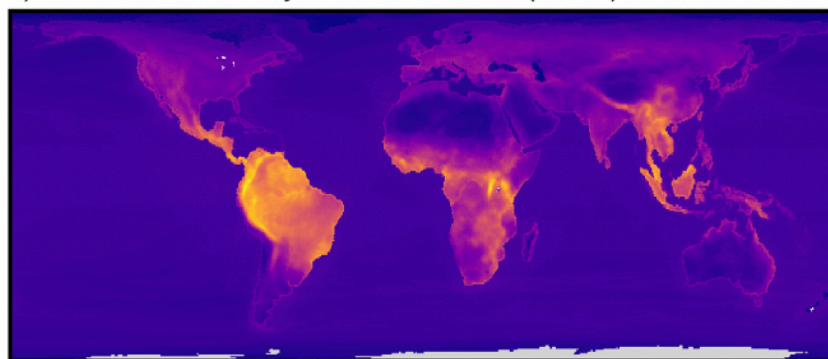
a) Handbook of the Mammals of the World (HMW)



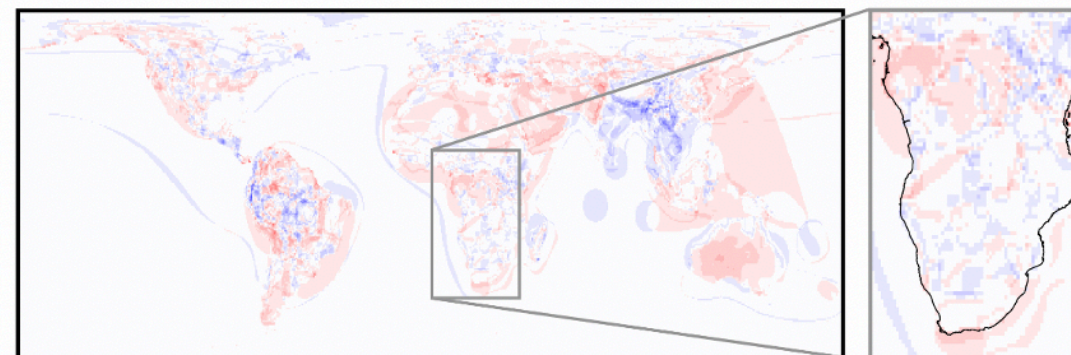
b) Illustrated Checklist of the Mammals of the World (CMW)



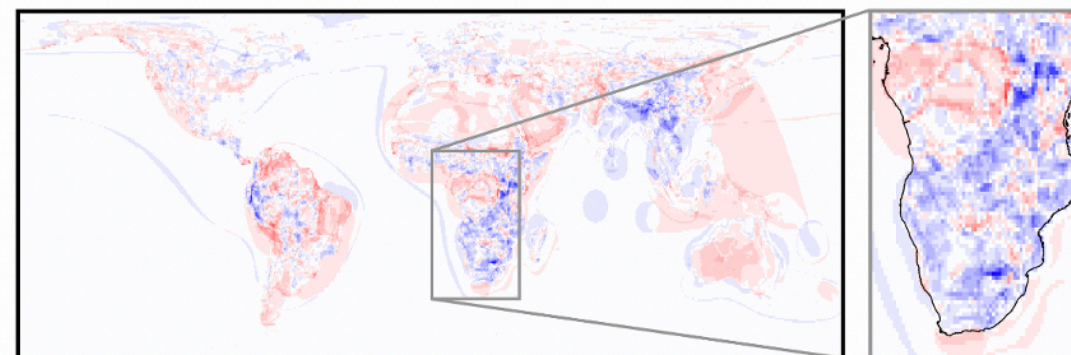
c) Mammal Diversity Database v1.2 (MDD)



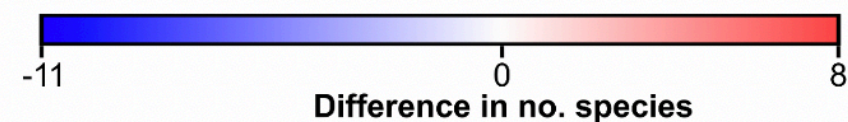
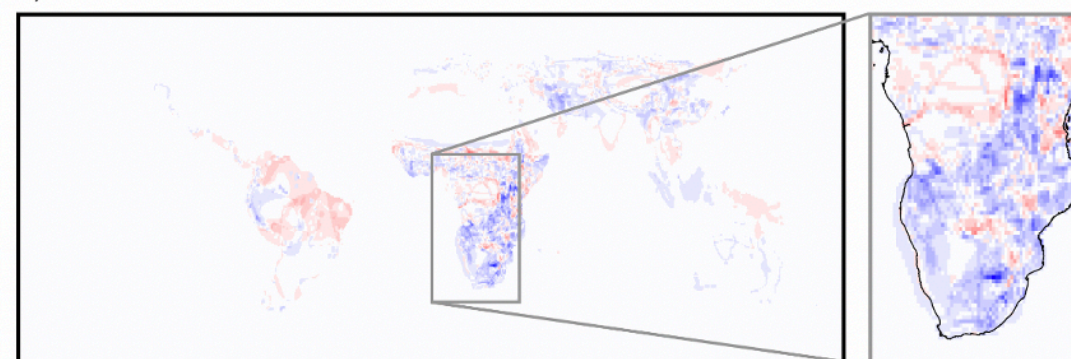
d) HMW vs CMW



e) HMW vs MDD



f) CMW vs MDD





**Deer mice:** most common mammal in North America

## Geospatial disambiguation

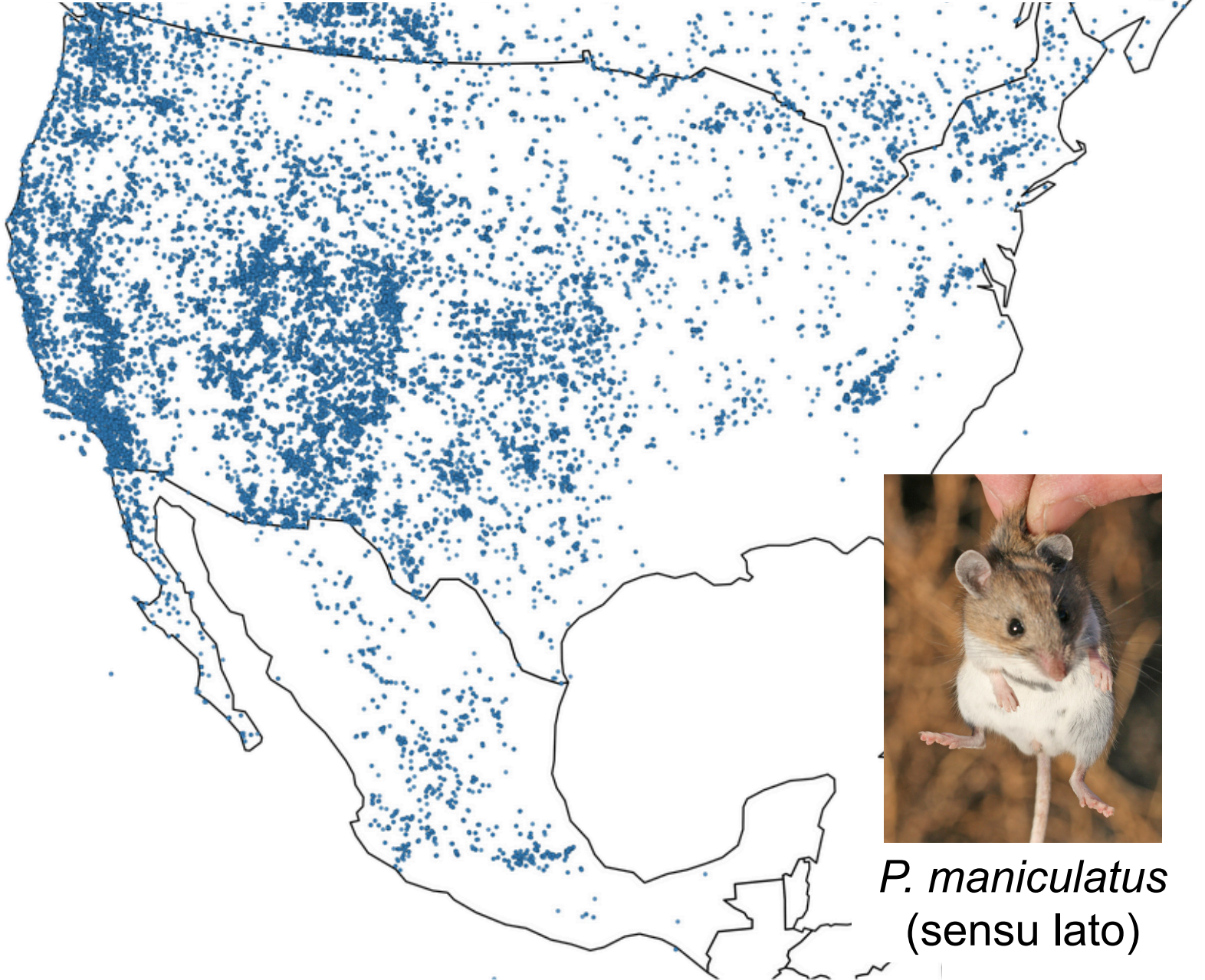
Disambiguating the  
taxonomy of  
**242,663 records** for  
*Peromyscus maniculatus*



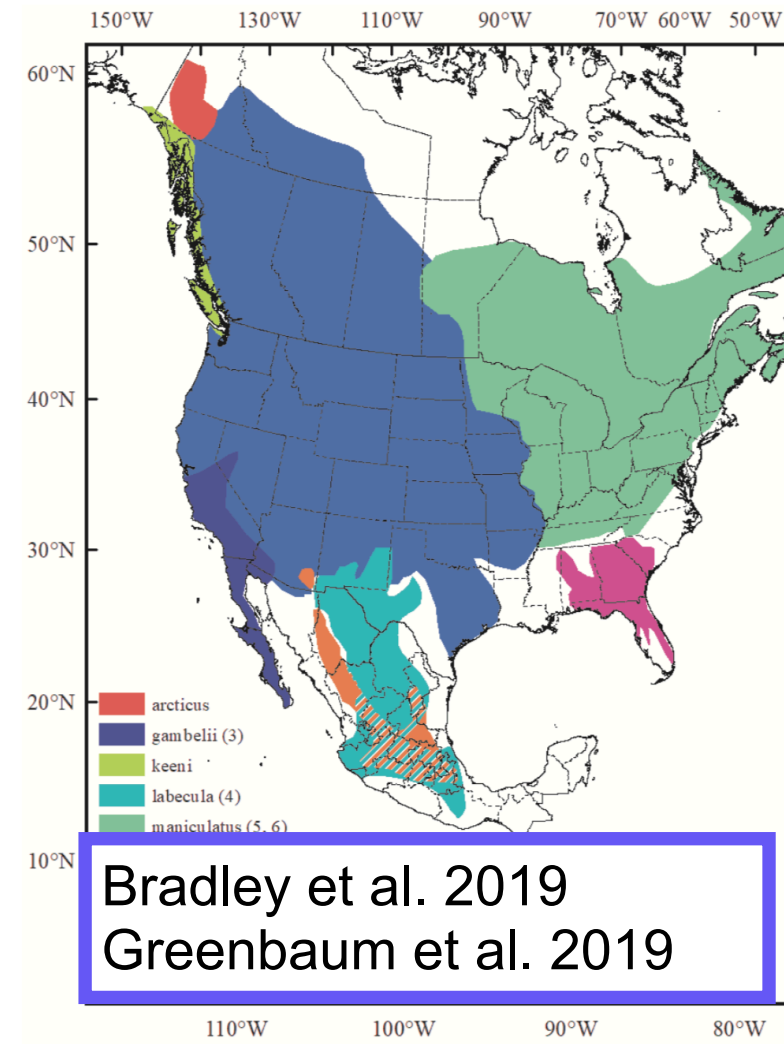
*P. maniculatus*



# Deer mice: most common mammal in North America



*P. maniculatus*  
(sensu lato)



Bradley et al. 2019  
Greenbaum et al. 2019

*P. maniculatus* (sensu stricto)  
*P. gambelii*  
*P. labecula*  
*P. sonoriensis*  
*P. arcticus*



## Geospatial disambiguation

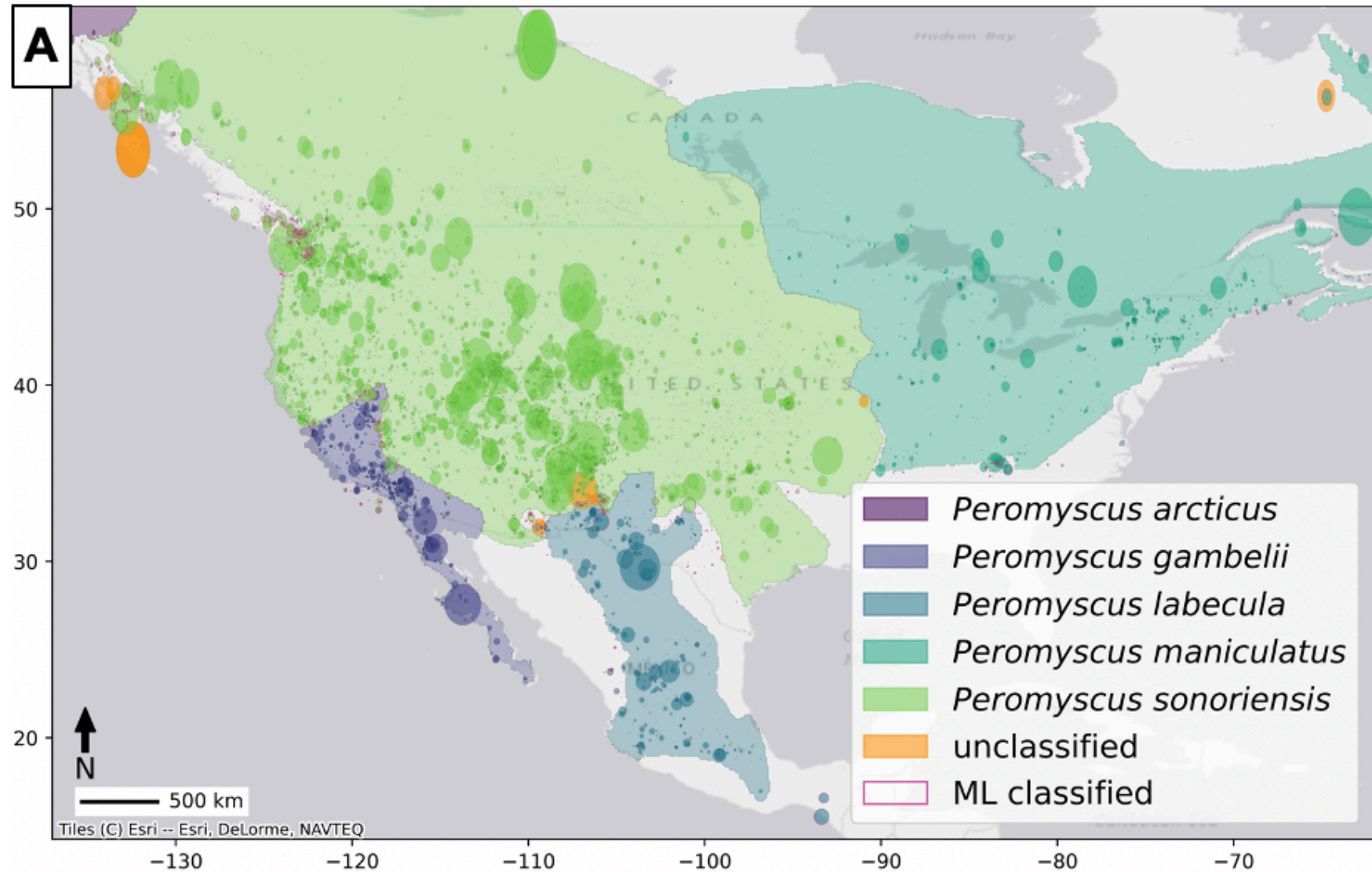
99% re-labeled to  
updated taxonomy using  
range maps  
+  
machine learning (ML)



Mammal Species  
of the World Next  
(MSWx)

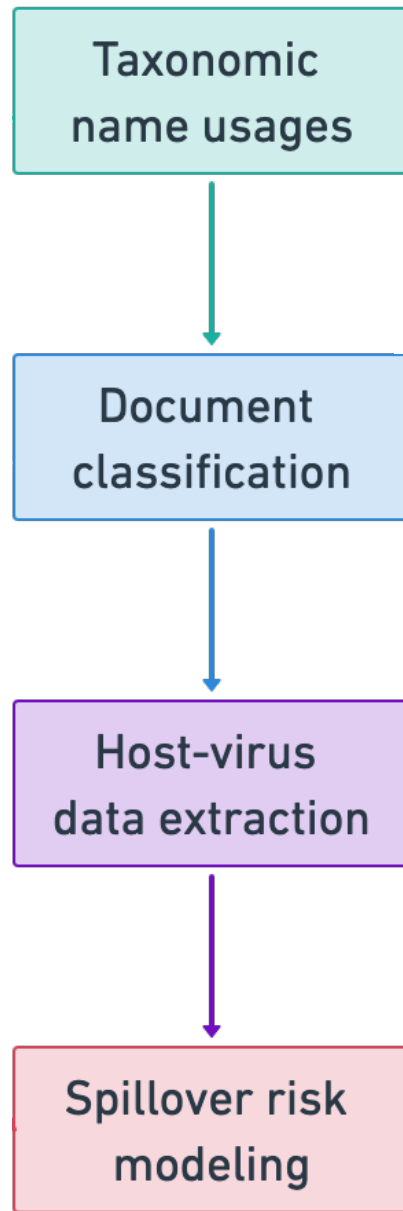


Arctos Collections  
Management Software



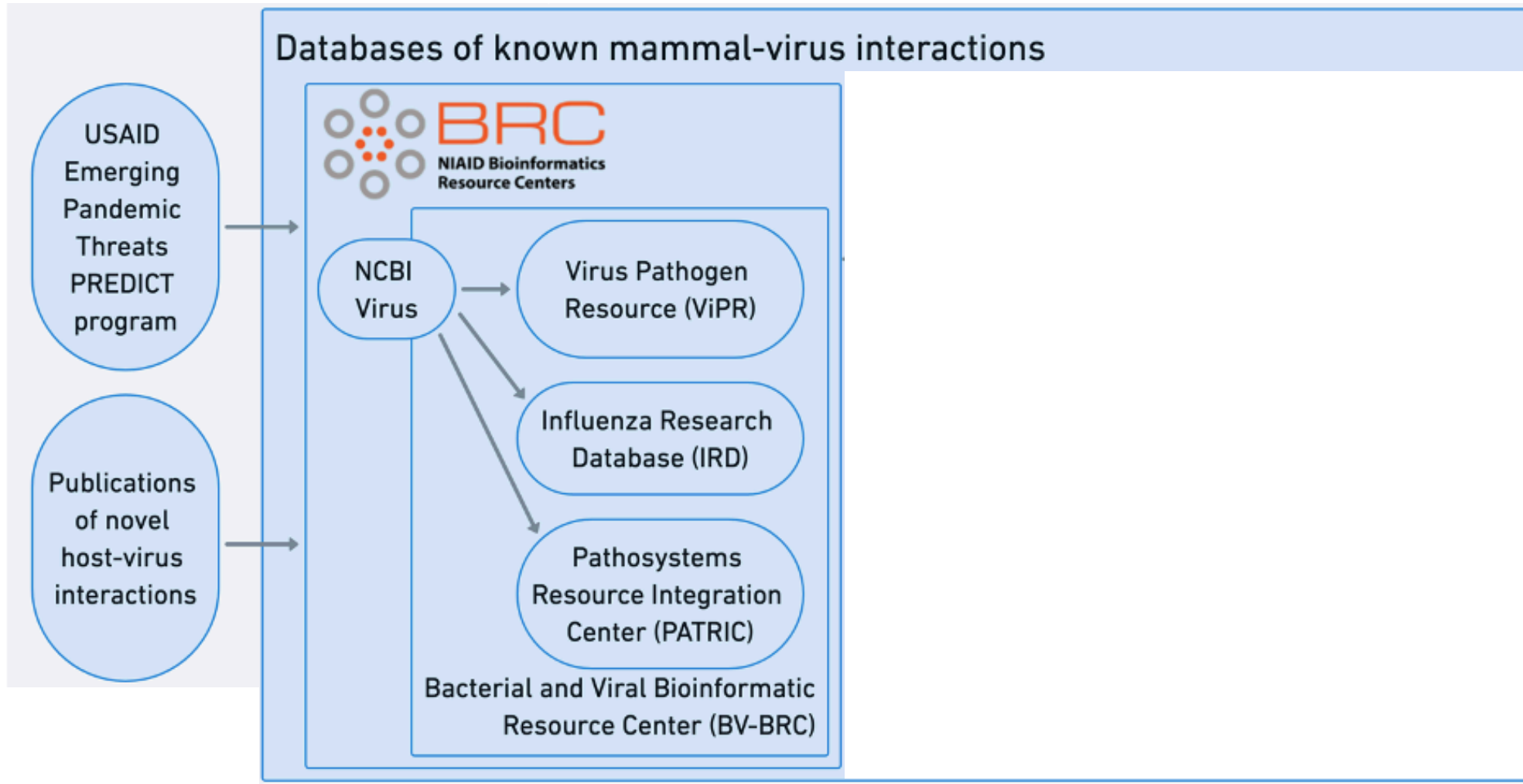
Upham et al. (in prep)





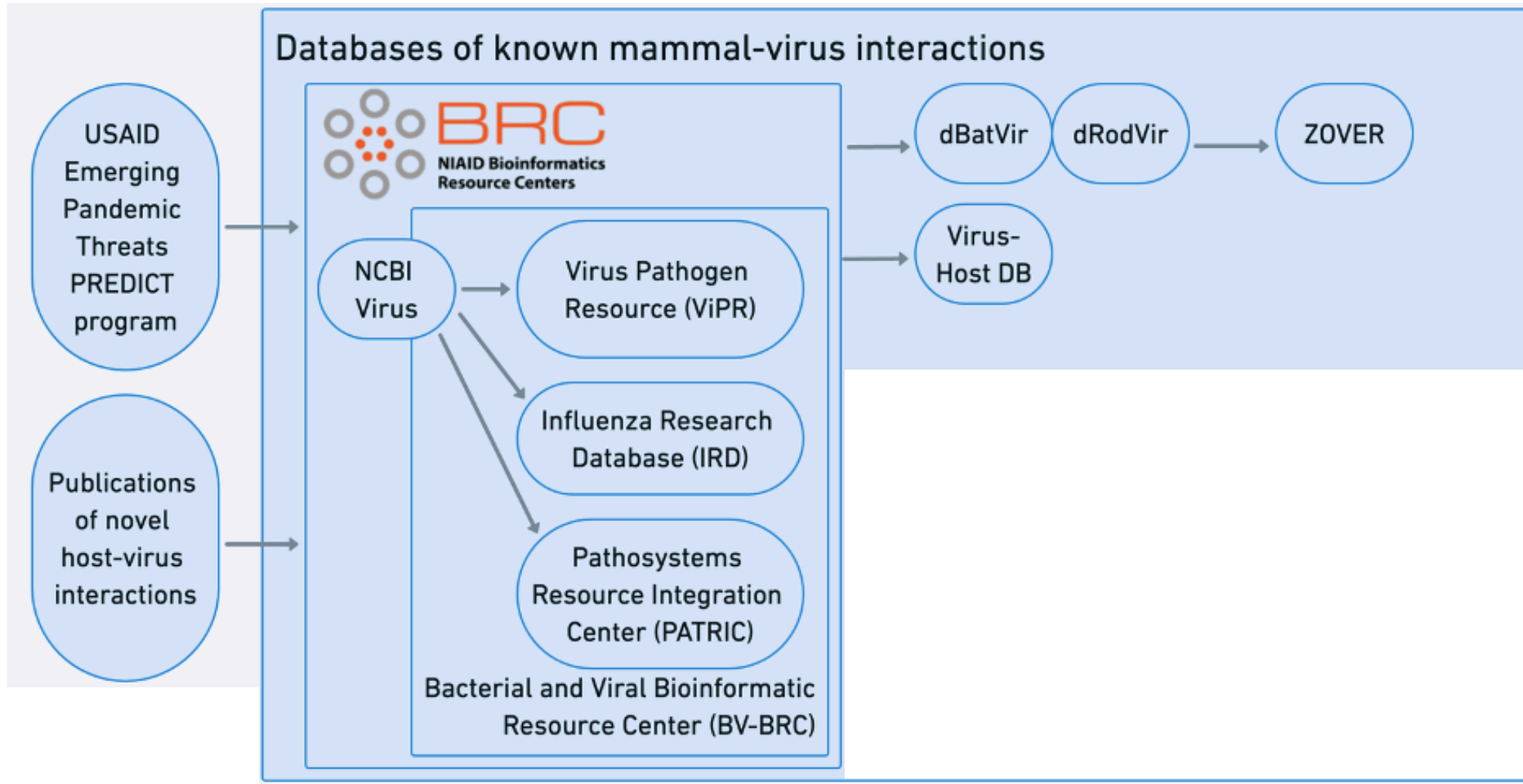


# Literature Abstract Classification System (LACS)



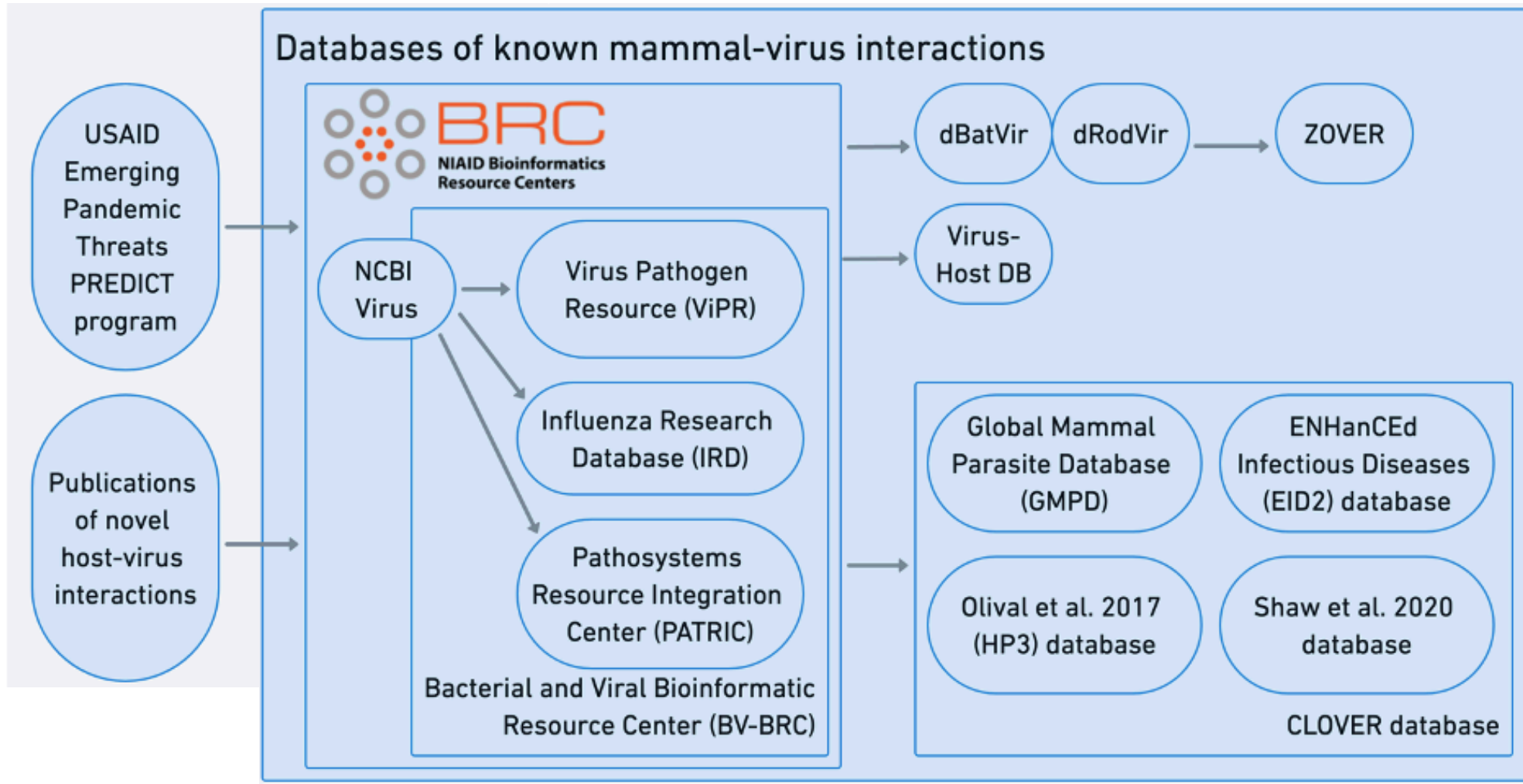


# Literature Abstract Classification System (LACS)





# Literature Abstract Classification System (LACS)





## Databases of known mammal-virus interactions



NCBI  
Virus

Virus Pathogen  
Resource (ViPR)

Influenza Research  
Database (IRD)

Pathosystems  
Resource Integration  
Center (PATRIC)

Bacterial and Viral Bioinformatic  
Resource Center (BV-BRC)

dBatVir

dRodVir

ZOVER

Virus-  
Host DB

Global Mammal  
Parasite Database  
(GMPD)

ENHanCed  
Infectious Diseases  
(EID2) database

Olival et al. 2017  
(HP3) database

Shaw et al. 2020  
database

CLOVER database

USAID  
Emerging  
Pandemic  
Threats  
PREDICT  
program

Publications  
of novel  
host-virus  
interactions

Global Biotic  
Interactions  
(GloBI)  
database

Extract

Digital Object  
Identifiers (DOIs)  
of source articles



## Databases of known mammal-virus interactions



NCBI  
Virus

Virus Pathogen  
Resource (ViPR)

Influenza Research  
Database (IRD)

Pathosystems  
Resource Integration  
Center (PATRIC)

Bacterial and Viral Bioinformatic  
Resource Center (BV-BRC)

dBatVir

dRodVir

ZOVER

Virus-  
Host DB

Global Mammal  
Parasite Database  
(GMPD)

ENHanCEd  
Infectious Diseases  
(EID2) database

Olival et al. 2017  
(HP3) database

Shaw et al. 2020  
database

CLOVER database

USAID  
Emerging  
Pandemic  
Threats  
PREDICT  
program

Publications  
of novel  
host-virus  
interactions

Global Biotic  
Interactions  
(GloBI)  
database

Extract

Digital Object  
Identifiers (DOIs)  
of source articles

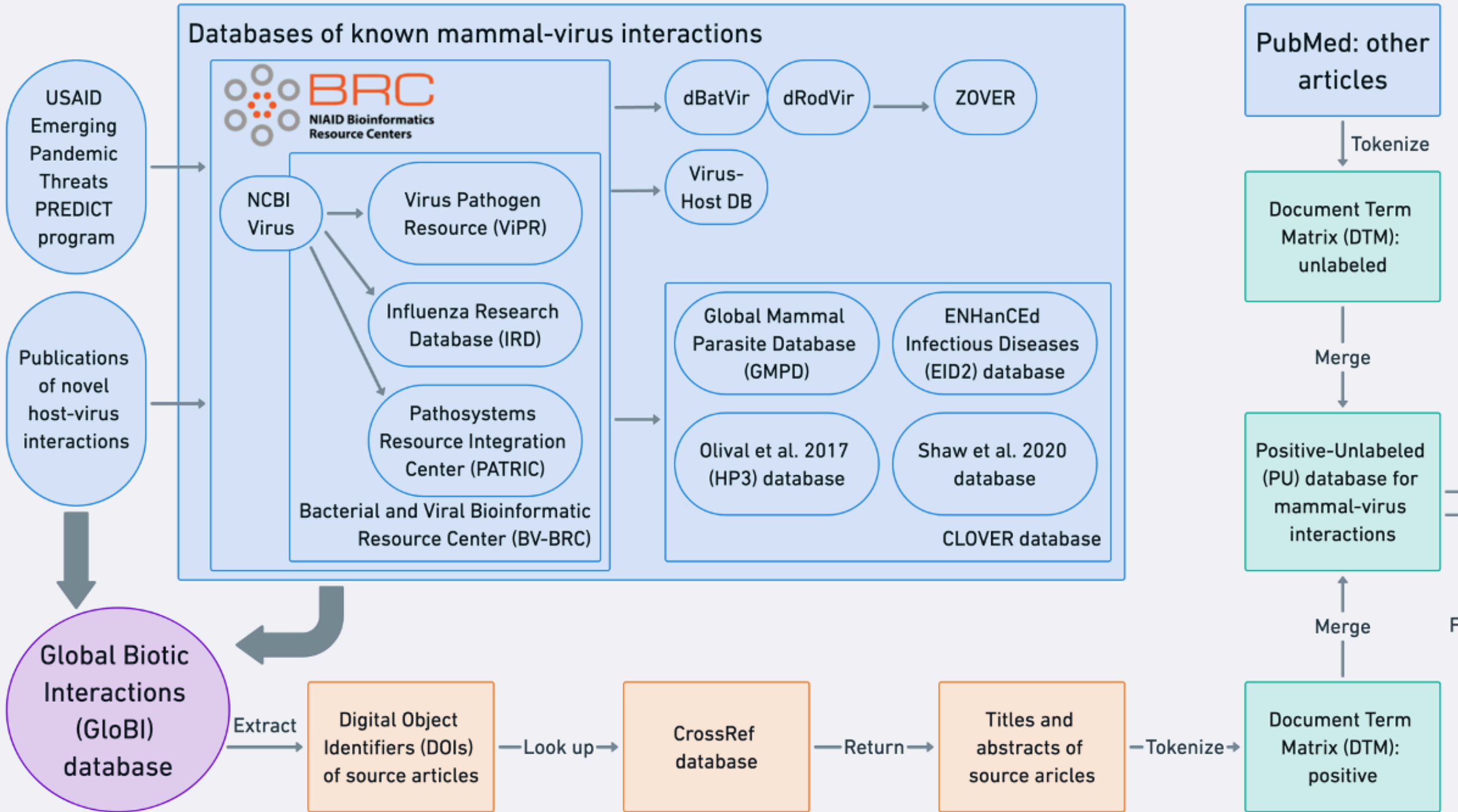
Look up

CrossRef  
database

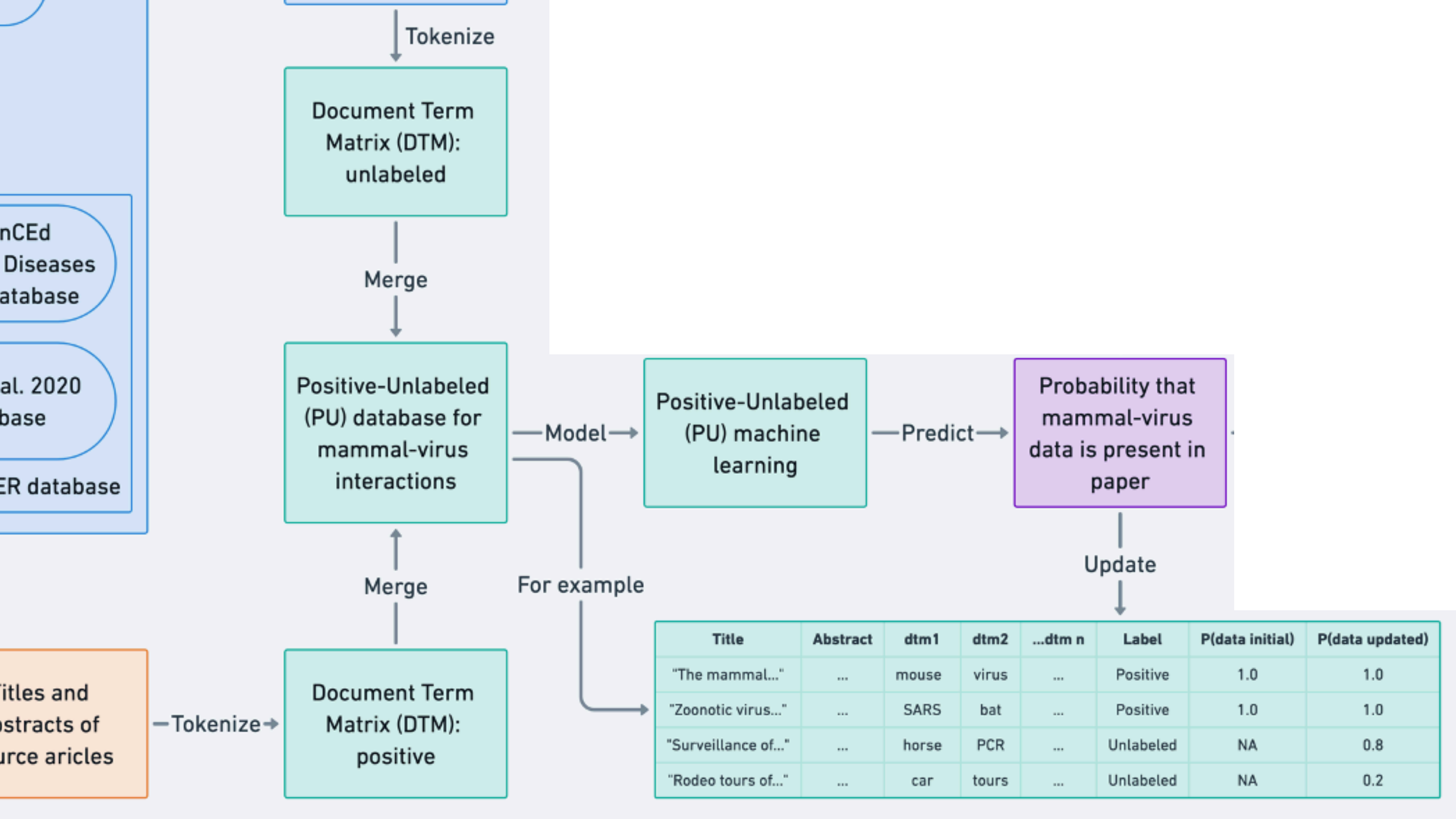
Return

Titles and  
abstracts of  
source articles

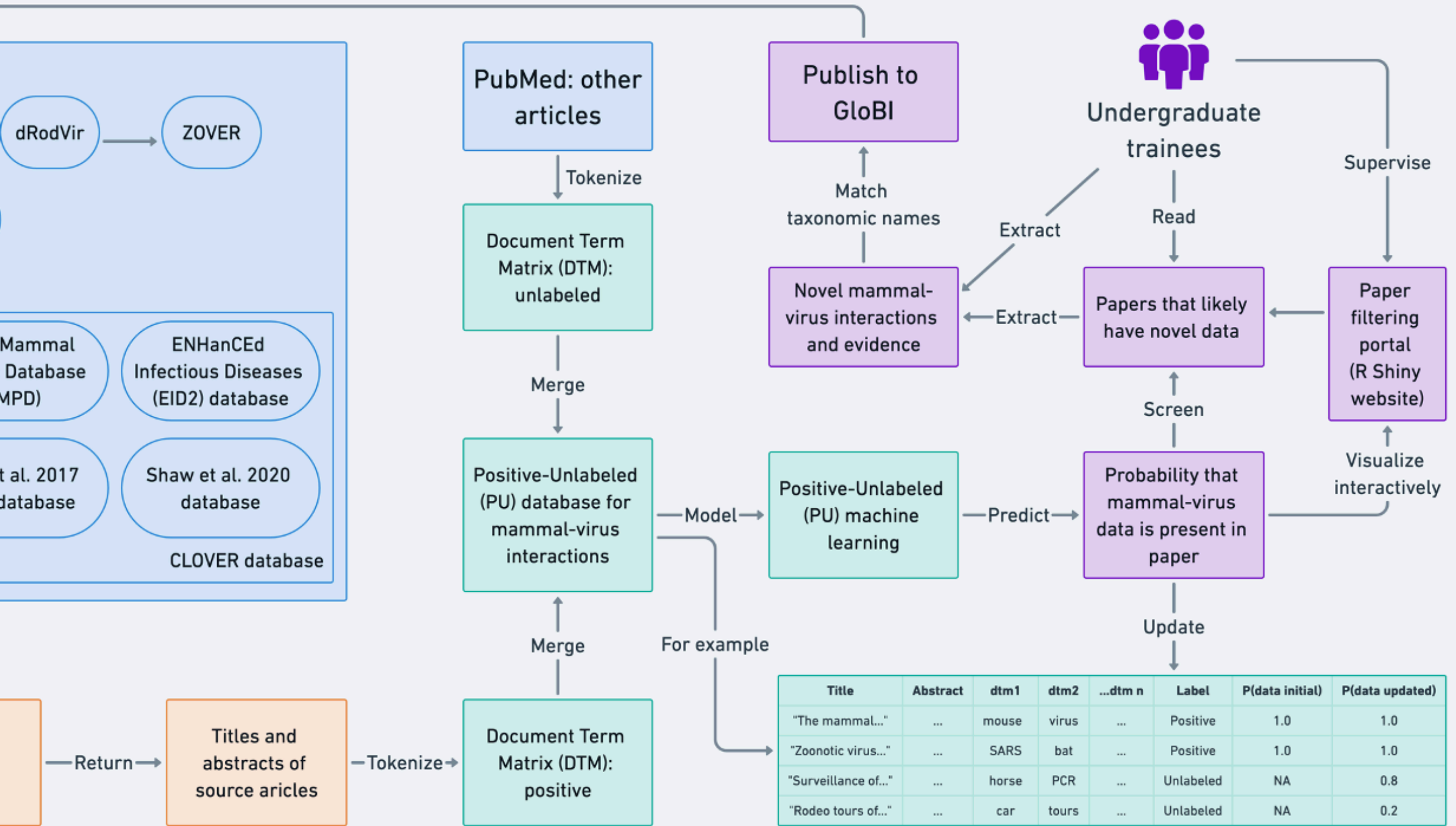






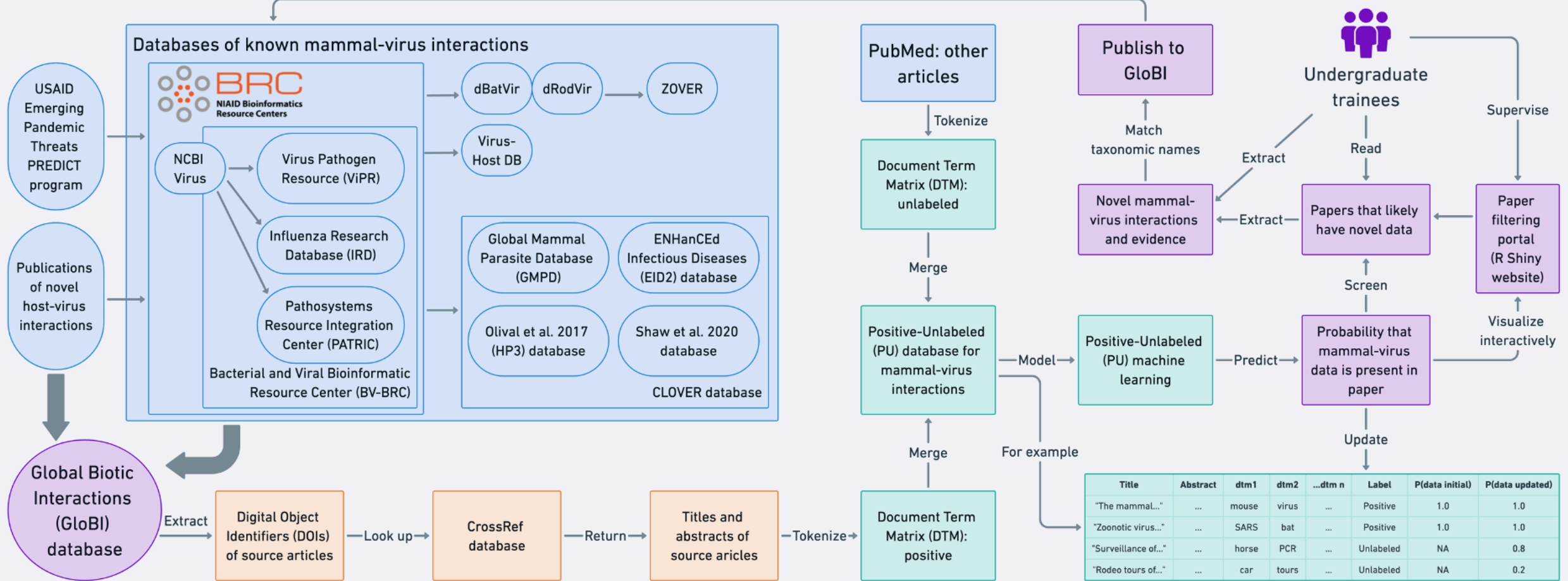








Improve and add to existing BRC data on mammal-virus interactions







Ángel Robles  
ASU PhD student

## Mammal-parasite paper recommender

Type in any string (e.g., 'hantavirus', 'rodent virus', 'bat').

The results will be PubMed papers sorted according to their 'classification\_score', which is the probability that a given paper contains mammal-parasite interactions.

### Search bar

Number of items searched:



You have selected peromyscus virus

This prototype service uses a machine learning model trained on titles and abstract from the Global Mammal Parasite Database (Stephens et al. 2017) to assess the similarity of given papers and thus the probability that they contain novel interaction data. Here is a description of our [methods](#). This work was funded by NIH ([1R21AI164268-01](#)) 'Intelligently predicting viral spillover risks from bats and other wild mammals' awarded to Arizona State University.

Search:

	search_title	search_abstract	pmid	pmcid	doi	classification_score	gmp
1	Tracing Transmission of Sin Nombre Virus and Discovery of Infection in Multiple Rodent Species.	Sin Nombre orthohantavirus (SNV), a negative-sense, single-stranded RNA virus that is carried and transmitted by the North American deer mouse Peromyscus maniculatus, can cause infection in humans through inhalation of aerosolized excreta from infected rodents. This infection can lead to hantavirus cardiopulmonary syndrome (HCPS), which has an ~36% case-fatality rate. We used reverse transcriptase quantitative PCR (RT-qPCR) to confirm SNV infection in a patient and identified SNV in lung tissues in wild-caught rodents from potential sites of exposure. Using viral whole-genome sequencing (WGS), we identified the likely site of transmission and discovered SNV in multiple rodent species not previously known to carry the virus. Here, we report, for the first time, the use of SNV WGS to pinpoint a likely site of human infection and identify SNV simultaneously in multiple rodent species in an area of known host-to-human transmission. These results will impact epidemiology and infection control for hantaviruses by tracing zoonotic transmission and investigating possible novel host reservoirs. IMPORTANCE Orthohantaviruses cause severe disease in humans and can be lethal in up to 40% of cases. Sin Nombre orthohantavirus (SNV) is the main cause of hantavirus disease in North America. In this study, we sequenced SNV from an infected patient and wild-caught rodents to trace the location of infection. We also discovered SNV in rodent species not previously known to carry SNV. These studies demonstrate for the first time the use of virus sequencing to trace the transmission of SNV and describe infection in novel rodent species.	34549977	PMC8577387	10.1128/JVI.01534-21	0.9956	NO



# GBIF Ebbe Nielsen, 2nd place

<https://gbif.lacs.info/>



Ángel Robles  
ASU PhD student



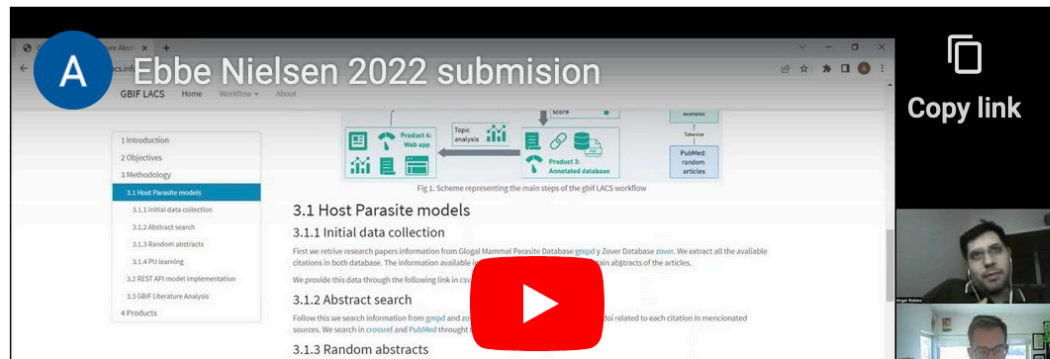
Home Workflow About



## GBIF LACS: GBIF Literature Abstract Classification System

Angel Luis Robles Fernandez & Nathan Upham - Arizona State University

This is GBIF Literature Abstract Classification System. Its designed to score if an abstract from scientific literature its related to host - parasite literature. We use this API to asses GBIF literature in order to find papers related with host - parasite interactions.





# Host-virus Data Extraction and Linking

- So far mostly manual work

## 1. Summary tables per paper

- North American rodent to virus interactions & methods of detection
- ASU undergraduates — creating a 'CURE' for Fall 2023



## 2. Bat co-roosting interactions

- Aja Sherman and Cullen Geiselman





August 7, 2022

Presentation Open Access

# What do we not know? Quantifying data gaps and biases in knowledge of bat co-roosting

ib Sherman, Aja C.; ib Geiselman, Cullen; ib Simons, Nancy B.; ib Upham, Nathan S.; ib Poelen, Jorrit H.; ib Reeder, DeeAnn M.; ib Bertolino, Sandro; ib Groom, Quentin; ib Phelps, Kendra; ib Agosti, Donat; ib Willoughby, Anna R.

Improved understanding of co-habitation of roosts by multiple species of bats is essential for estimating the risks of zoonotic disease transmission. However, ecological data on roosting environments, species richness, bat-bat interactions, viral infections, and other species interactions are scattered throughout the literature, making them difficult to study on a global scale. The research scope for most roost studies has been narrow, focusing on roost type, bat abundance, and locality data while failing to investigate interspecific roosting interactions. To meet this need, we have collaboratively built an open-access dataset of ecological interactions (including co-roosting, trophic, anthropogenic, and parasitic) extracted from the literature to improve our understanding of roost dynamics on a global scale, and to elucidate the role of shared roosts in disease transmission. As of April 2022, >11,500 interaction records involving >360 bat species from >137 countries encompassing a variety of habitats have been extracted from >175 publications spanning from 1860-2020, all accessible via the Coronavirus-Host community at Zenodo. With this benchmark dataset of open-access digitized interaction data, tools, and workflows, we provide evidence of co-roosting events that we aligned with multiple ontologies (interaction terms, taxonomies, administrative regions) and phylogenies suitable for high-throughput analysis. We followed open access and FAIR (Findable, Accessible, Interoperable, and Reusable) data principles for extracting data and choosing methodologies. We identify biases in the coverage of bat interaction records, suggest new tools for biodiversity informatics, and explore obstacles and opportunities in the mining of eco-interactions previously lost in the annals of scientific literature.

184

views

124

downloads

[See more details...](#)

Part of



Indexed in

OpenAIRE

Preview

Page: 1 of 43 Automatic Zoom



Aja C. Sherman<sup>1</sup>, Cullen Geiselman<sup>1</sup>, Nancy B. Simmons<sup>2</sup>, Nathan S. Upham<sup>3</sup>, Jorrit H. Poelen<sup>4,5</sup>, DeeAnn M. Reeder<sup>6</sup>, Sandro Bertolino<sup>7</sup>, Quentin Groom<sup>8</sup>, Kendra Phelps<sup>9</sup>, Donat Agosti<sup>10</sup> and Anna R. Willoughby<sup>11,12</sup>

Publication date:

August 7, 2022

DOI:

DOI 10.5281/zenodo.6959687

Keyword(s):

biodiversity, chiroptera, viral spillover

Meeting:

[19th International Bat Research Conference](#), Austin, Texas, USA, August 11, 2022

Related identifiers:

Has part

<https://vimeo.com/744642070> (Video/Audio)

Communities:

[Biodiversity Literature Repository](#)

As of April 2022

- >11,500 interaction records involving >360 bat species from >137 countries
- extracted from >175 publications spanning from 1860-2020






# Host-virus Data Extraction and Linking

## Collaborations to automate

- **SIBiLS** group  
(Swiss Institute of Bioinformatics Literature Services),  
Patrick Ruch
- **BICIKL** project  
(Biodiversity Community Integrated Knowledge Library),  
Donat Agosti



BICIKL  
*Biotic interactions browser*

Query

Species 1:

Example: pangolins

☒ Neilonella rodent [2851046]

Or select another taxonomic entity:

☐ unclassified Rodentia [4117428]

☐ Cryptosporidium sp. wild rodent strain [5986333]

☐ Hepatozoon sp. genotype rodent SP-1 [5986219]

☐ Hepatozoon sp. genotype rodent SP-2 [5986220]

Species 2:

Example: cats

☒ Plasmodium [474836]

Or select another taxonomic entity:

☐ Plasmodium vivax Sal-1 [427153]

☐ Plasmodium falciparum Mad20/Papua New Guinea [877115]

☐ Plasmodium billbrayi [252666]

☐ Plasmodium berghei ANKA [724400]

☐ Plasmodium sp. gorilla clade G3 [582802]

[\[+\] More](#)

Interaction:

Example: host of

Examples

Which species do/can interact with pangolins?

What is the interaction between Oncomelania hupensis and Schistosoma japonicum?

Which species are host of Biomphalaria glabrata?

Search

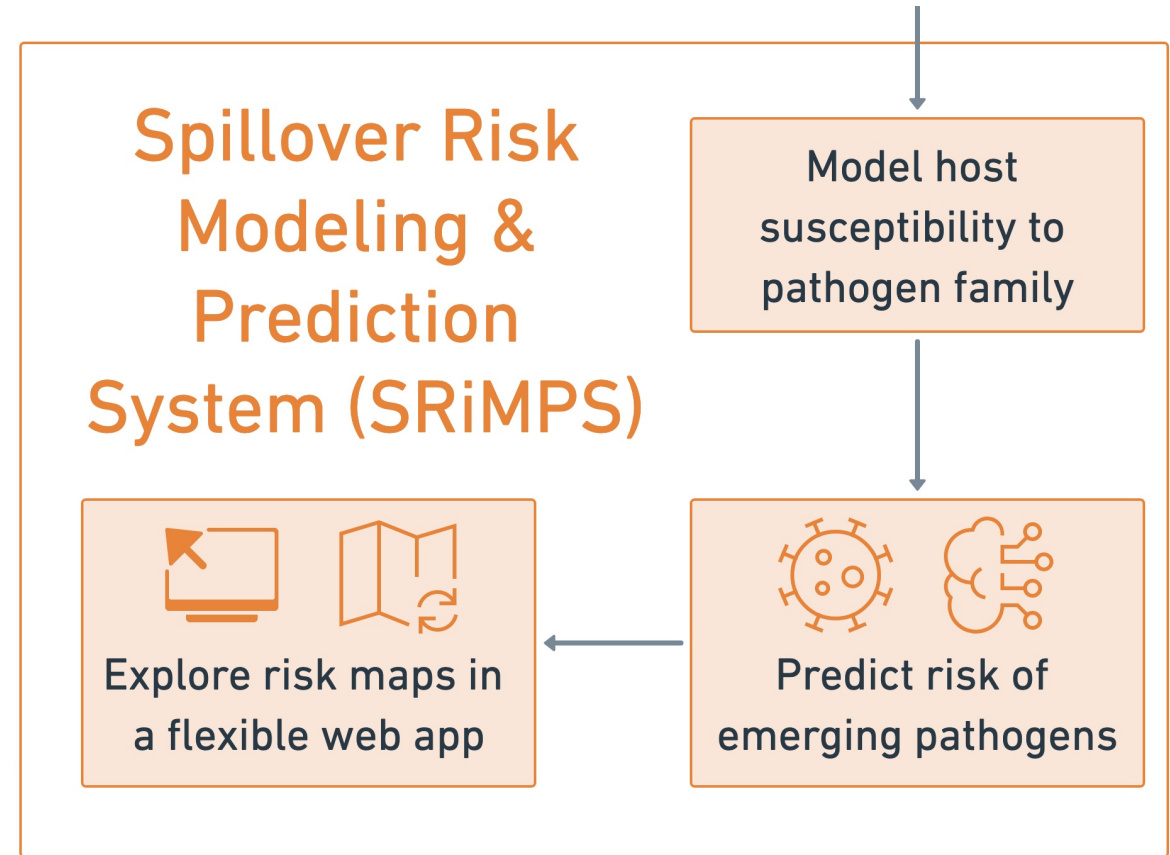


# Spillover Risk Modeling

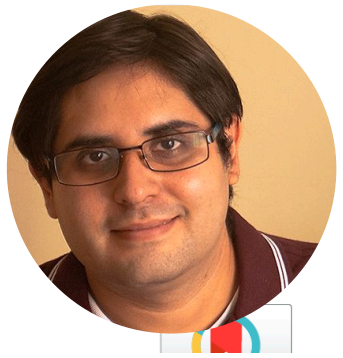
Examples:

- from Ángel's MS thesis in PNAS

- recent work at ASU on Rodentia-Hantavirus







Ángel Robles  
ASU PhD student

## Wildlife susceptibility to infectious diseases at global scales

Ángel L. Robles-Fernández<sup>a,b,1</sup> , Diego Santiago-Alarcon<sup>c,1</sup> , and Andrés Lira-Noriega<sup>d,1</sup> 

Edited by Nils Stenseth, Universitetet i Oslo, Oslo, Norway; received December 17, 2021; accepted July 11, 2022

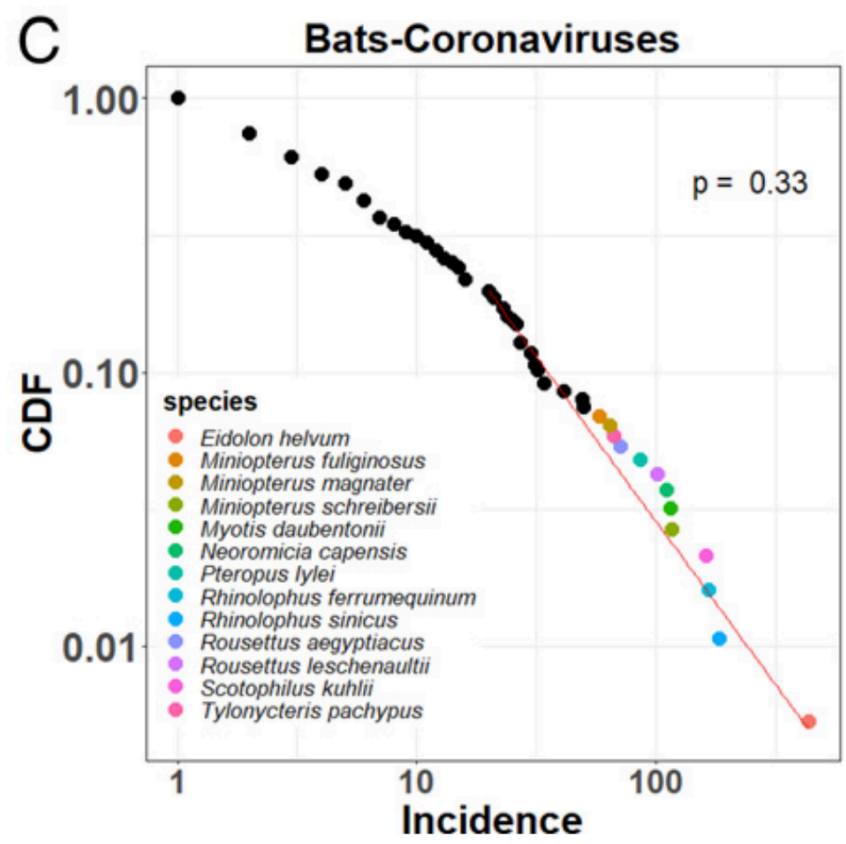
Disease transmission prediction across wildlife is crucial for risk assessment of emerging infectious diseases. Susceptibility of host species to pathogens is influenced by the geographic, environmental, and phylogenetic context of the specific system under study. We used machine learning to analyze how such variables influence pathogen incidence for multihost pathogen assemblages, including one of direct transmission (coronaviruses and bats) and two vector-borne systems (West Nile Virus [WNV] and birds, and malaria and birds). Here we show that this methodology is able to provide reliable global spatial susceptibility predictions for the studied host–pathogen systems, even when using a small amount of incidence information (i.e., <20% of information in a database). We found that avian malaria was mostly affected by environmental factors and by an interaction between phylogeny and geography, and WNV susceptibility was mostly influenced by phylogeny and by the interaction between geographic and

### Significance

Previous studies have investigated how environmental, phylogenetic, and geographic variables determine pathogen infection, particularly for human zoonosis. Yet, none of those previous studies have provided a methodology that can be applied

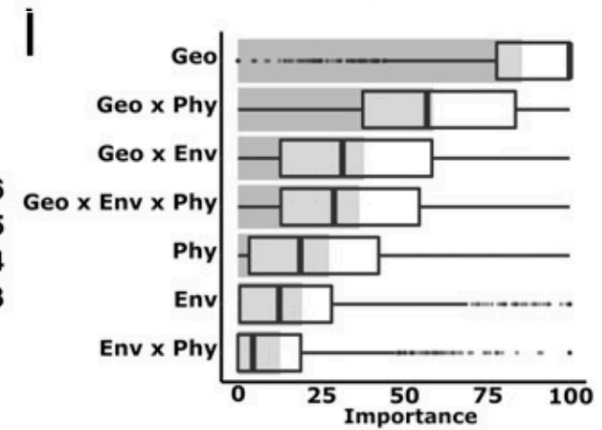
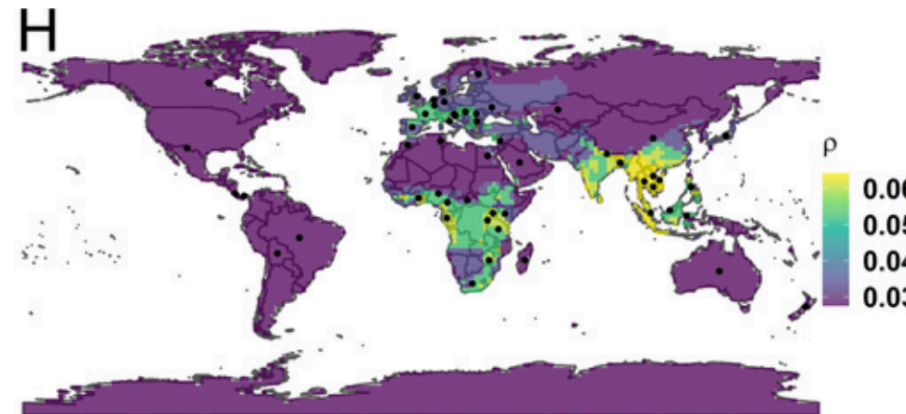
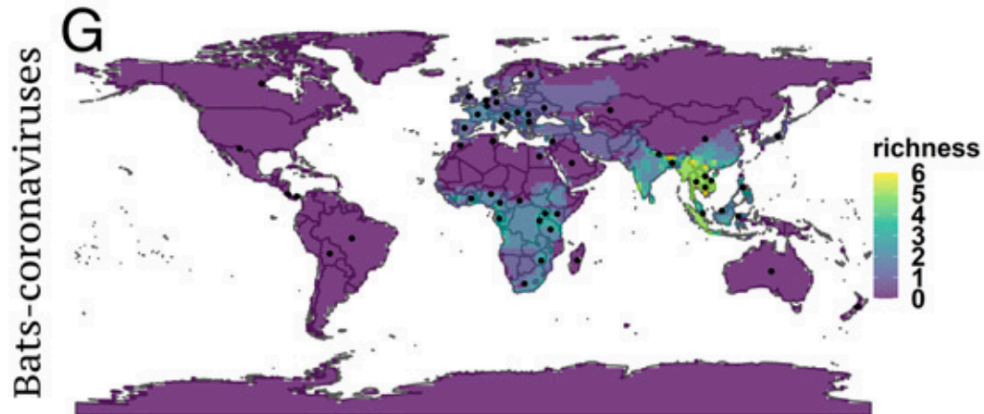
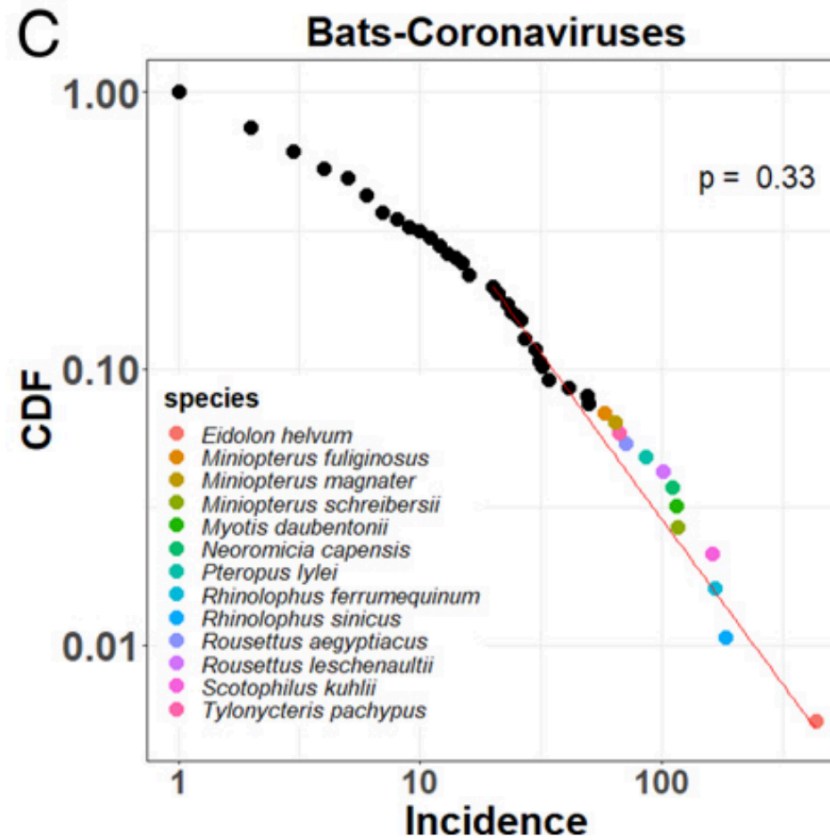


# Spillover Risk Modeling



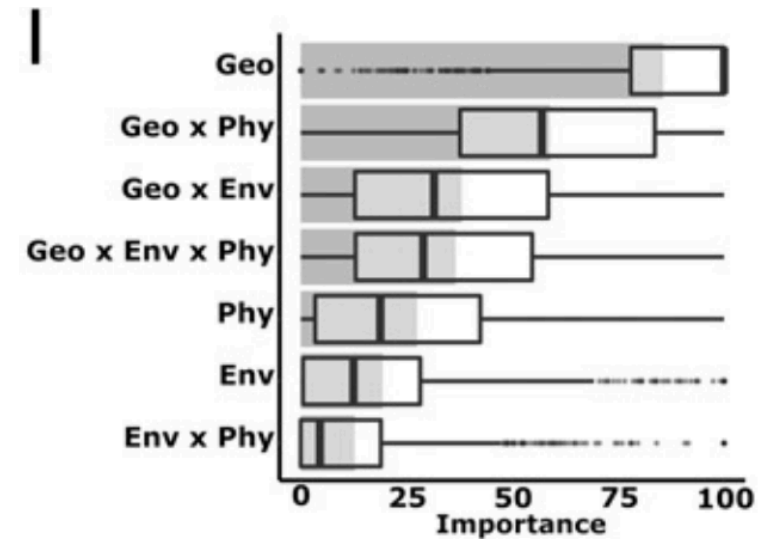
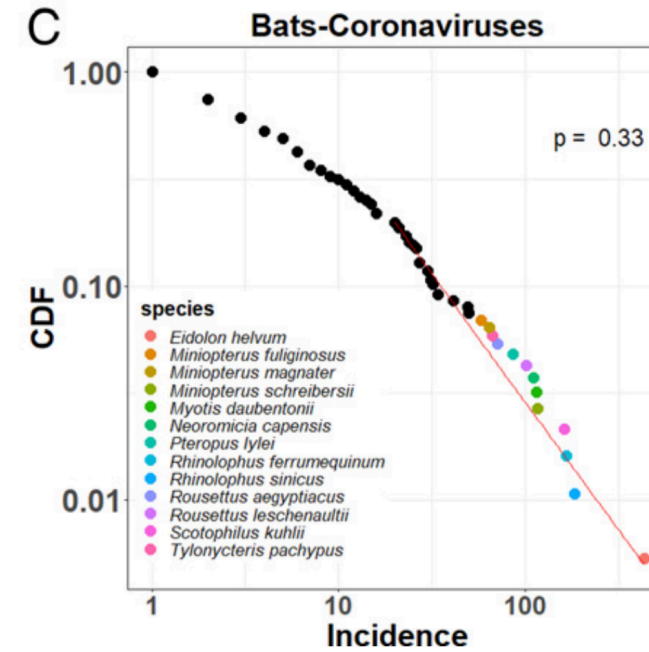
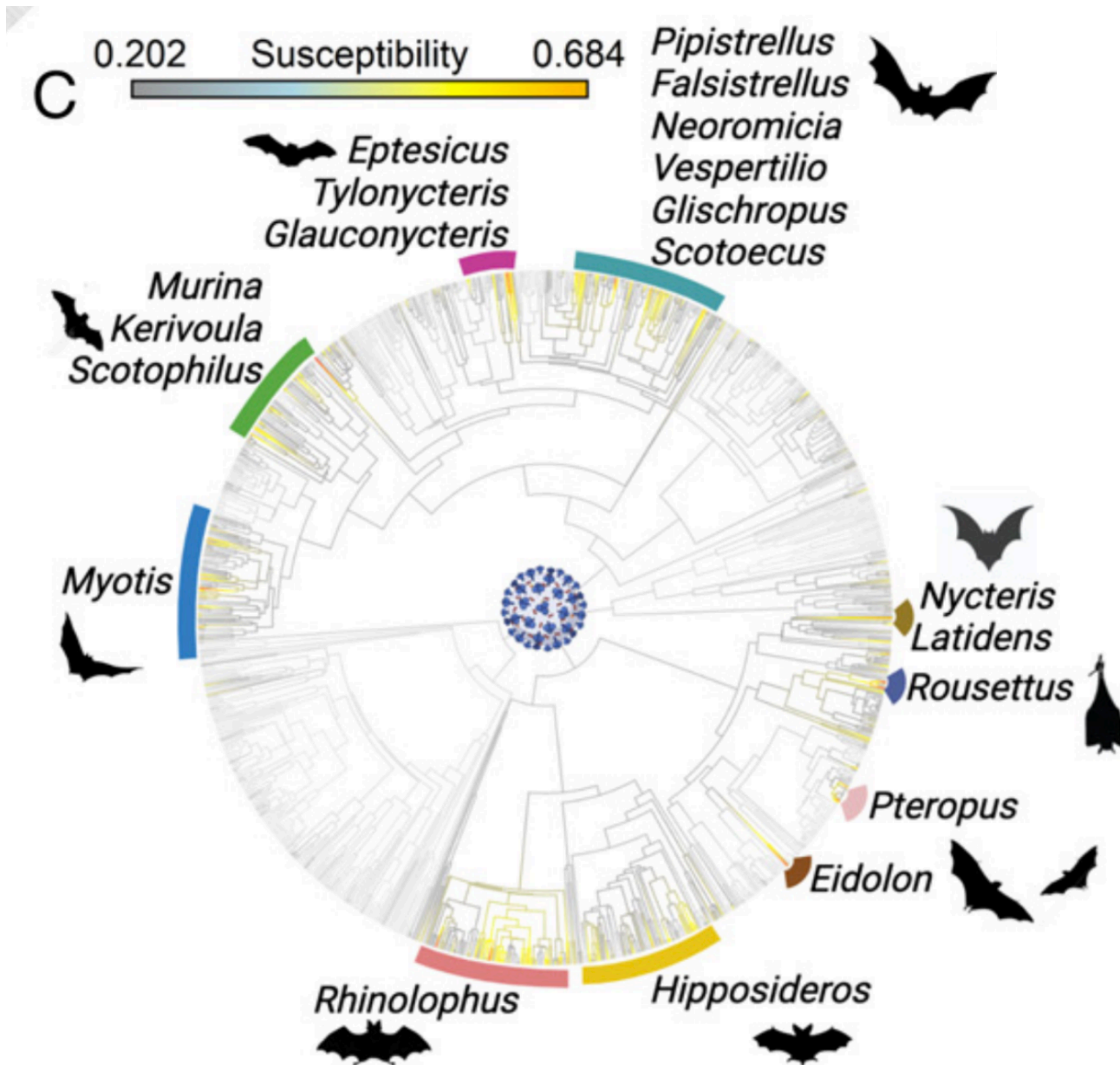


# Spillover Risk Modeling

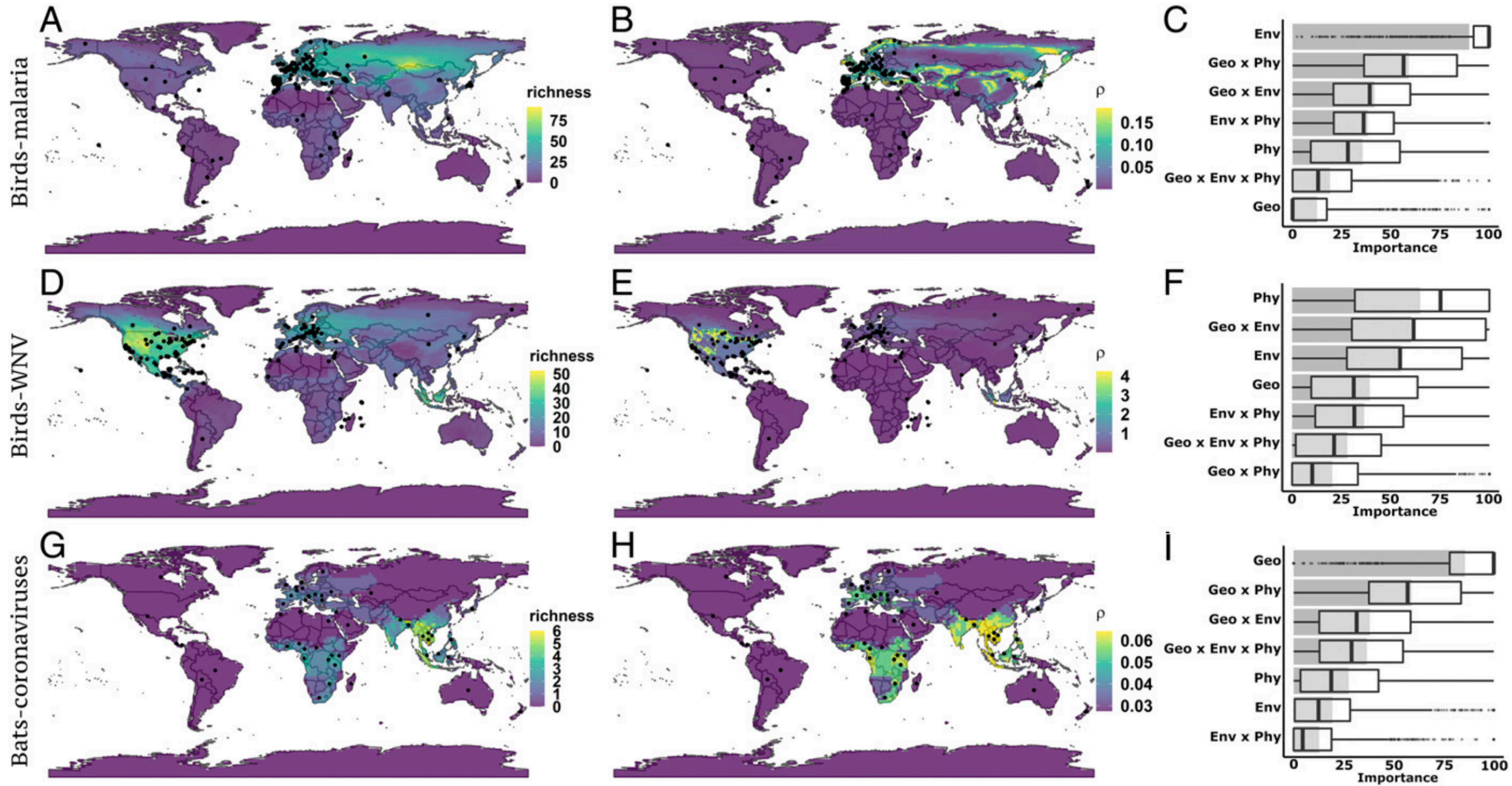




# Spillover Risk Modeling











# Rodentia-Hantavirus work



[Upload](#)[Communities](#)

April 15, 2022

[Journal article](#)[Open Access](#)

## Modeling Rodentia-Hantaviridae interactions from different biodiversity dimensions through machine learning

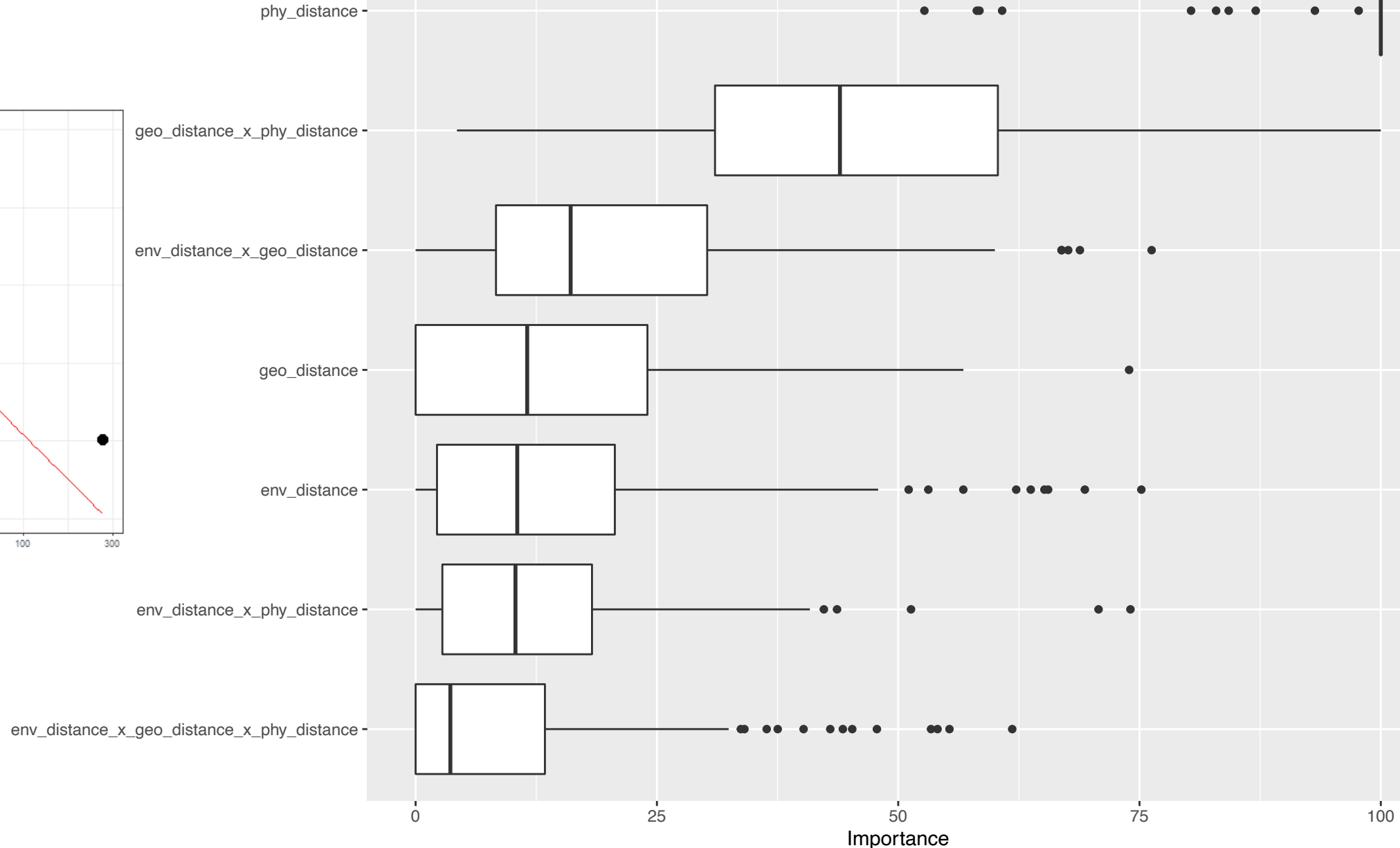
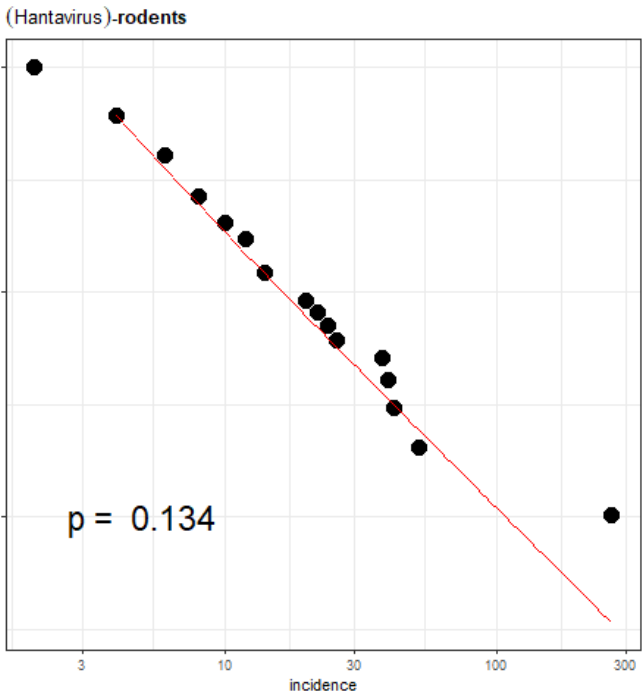
Angel Luis Robles Fernandez; Nathan Upham

Understanding the eco-geographic ‘rules’ of host-pathogen interactions is needed for developing fine-grained tools for predicting the risk of different wildlife pathogens being transmitted to humans (i.e., spillover risk). Biodiversity conservation efforts similarly benefit from understanding these general rules of community assembly, which predict expected host-host, host-environment, and host-pathogen interactions under different amounts of human disturbance.

We use machine learning to analyze how these variables influence the incidence of pathogens for the Rodentia

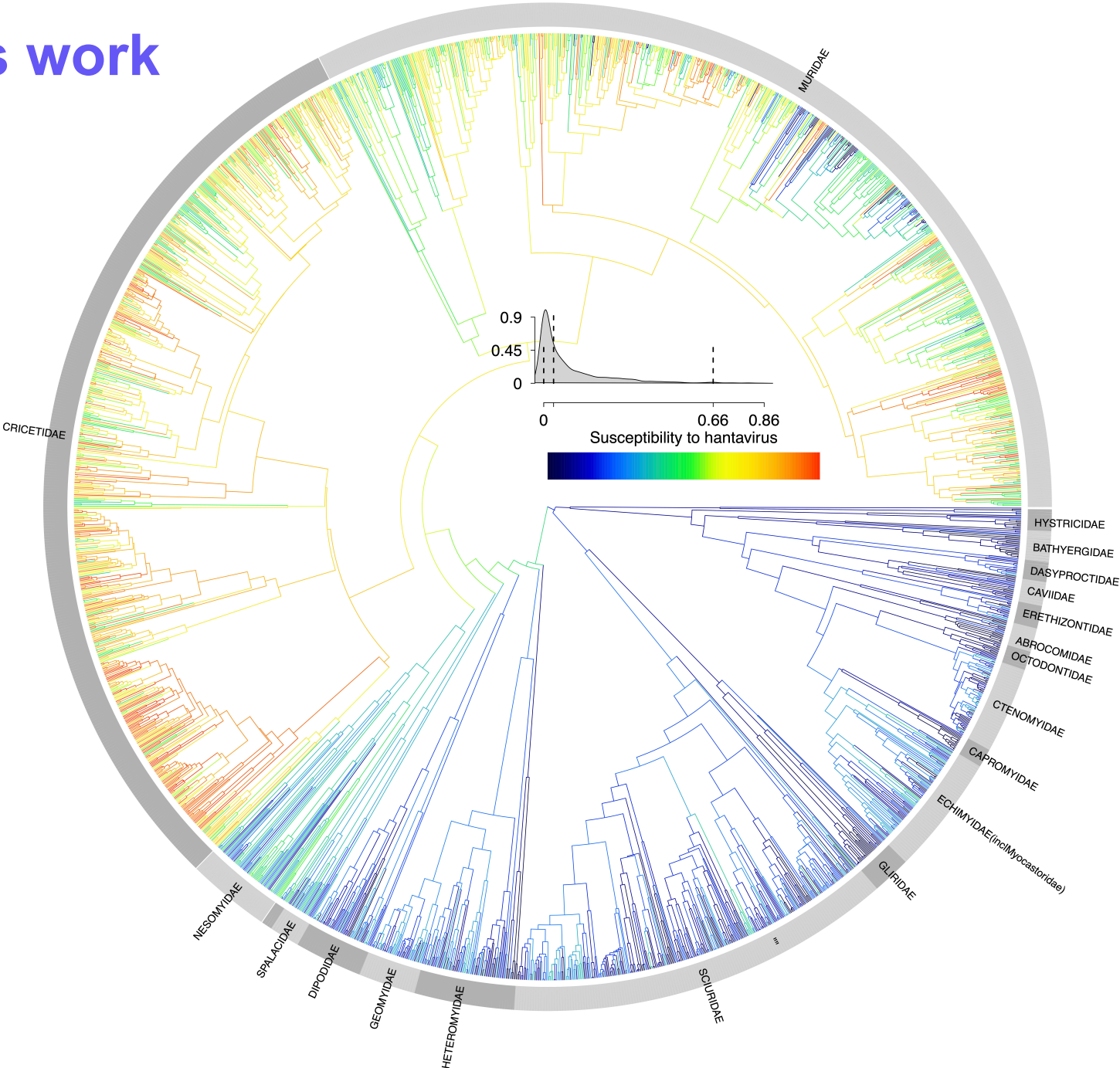


# Rodentia-Hantavirus work





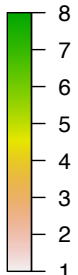
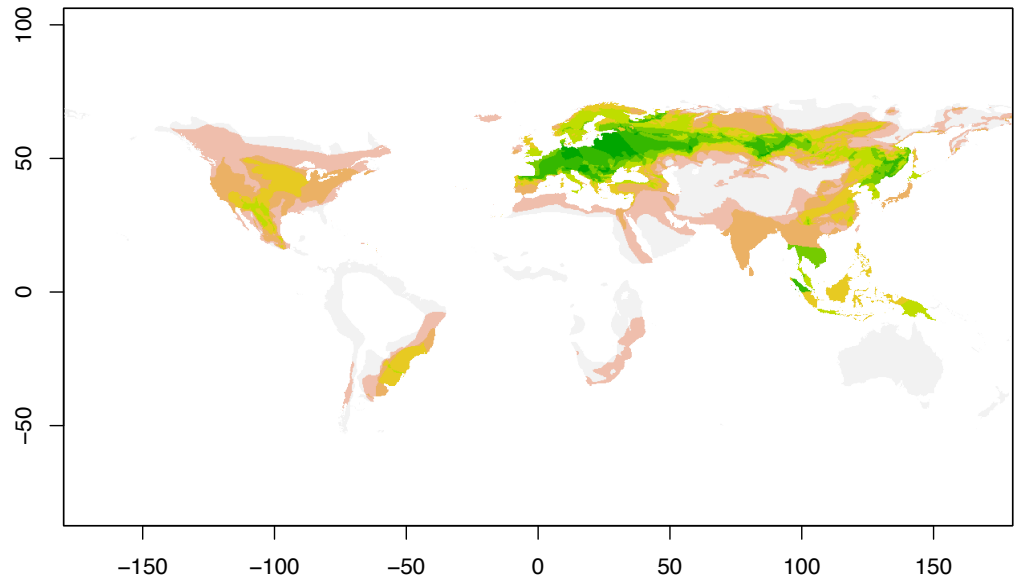
# Rodentia-Hantavirus work



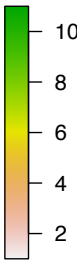
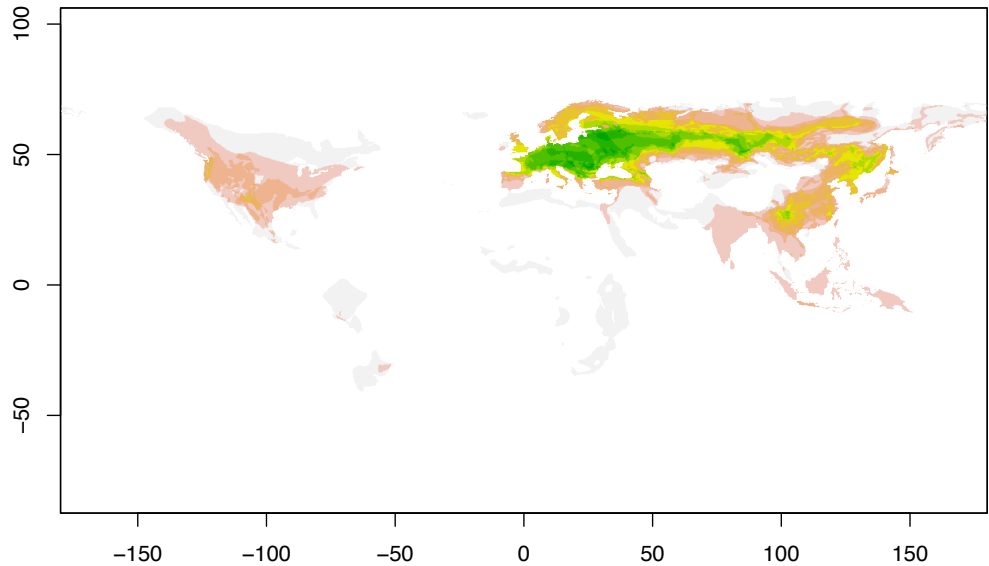


# Rodentia-Hantavirus work

Richness map of top 30 species  
with incidence to hantavirus



Richness map of susceptible host to hantavirus





# CONCLUSIONS

1. Knowing *which species are present* is the first step to knowing many other things...

Phylogeny

Ecology

Taxonomy





# CONCLUSIONS

**Phylogeny**

**Ecology**

**Taxonomy**

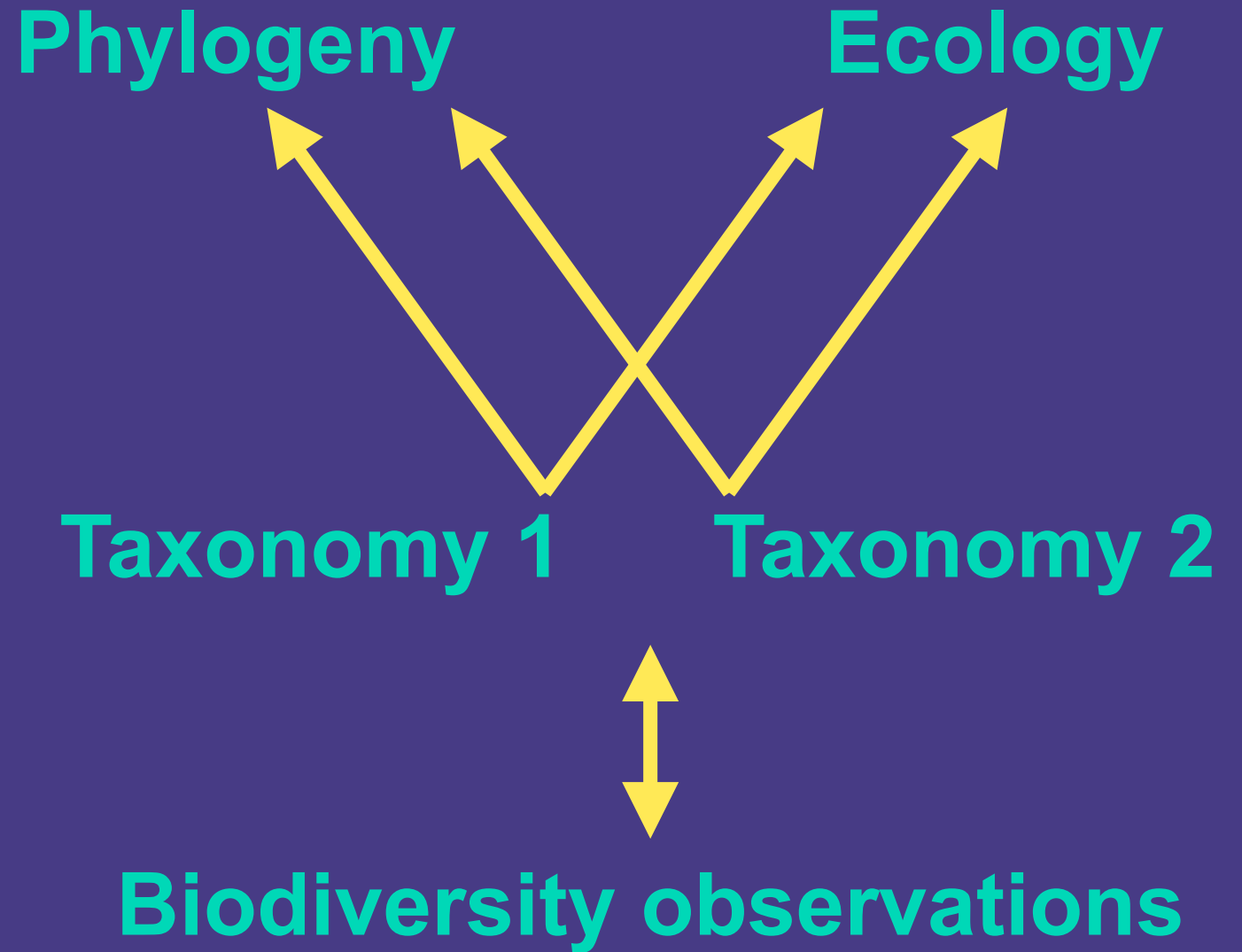
**Biodiversity observations**





# CONCLUSIONS

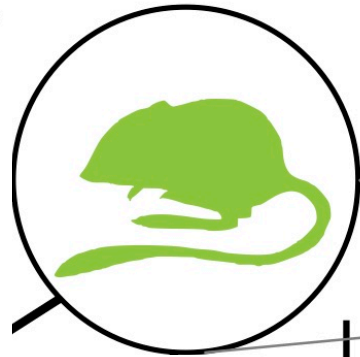
2. The real goal is to **treat taxonomy as a variable**, test its impact on results and policy decisions



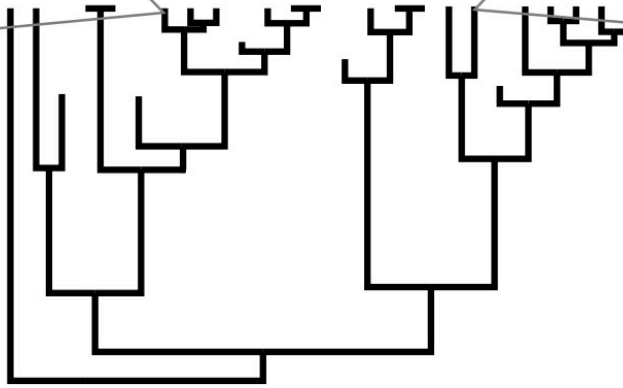


# Summary

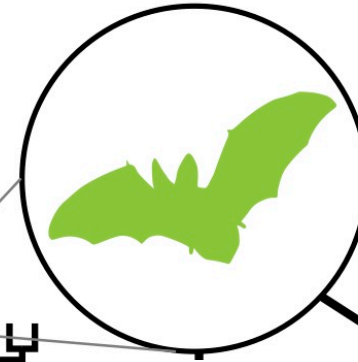
**Taxonomy**



**Phylogeny**



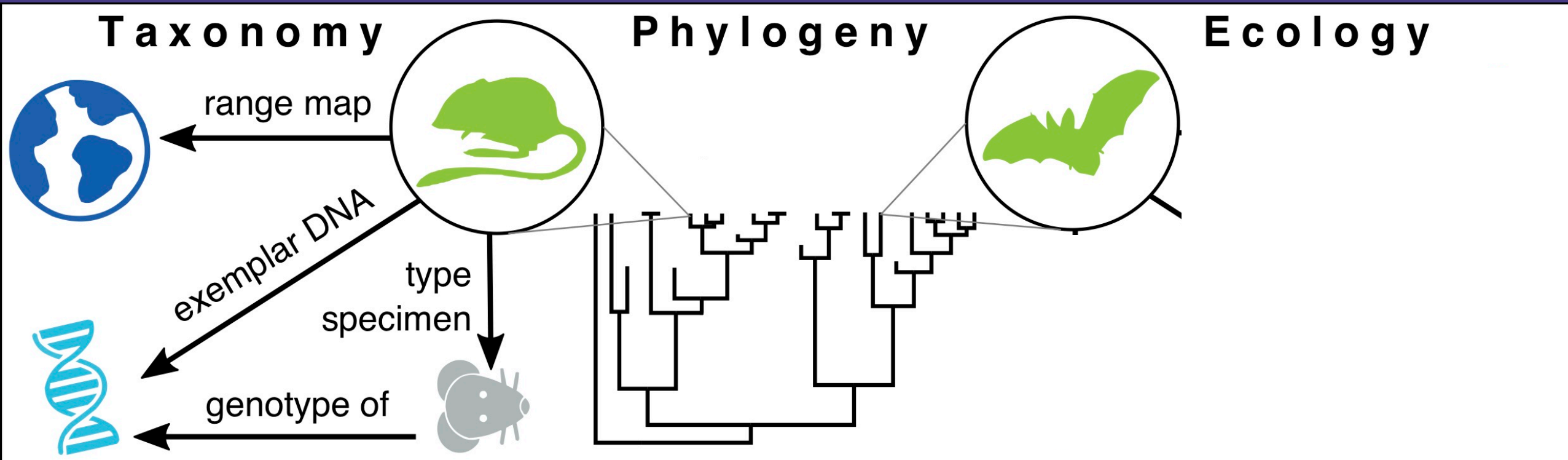
**Ecology**





# Summary

## 1. What are the mammal species?

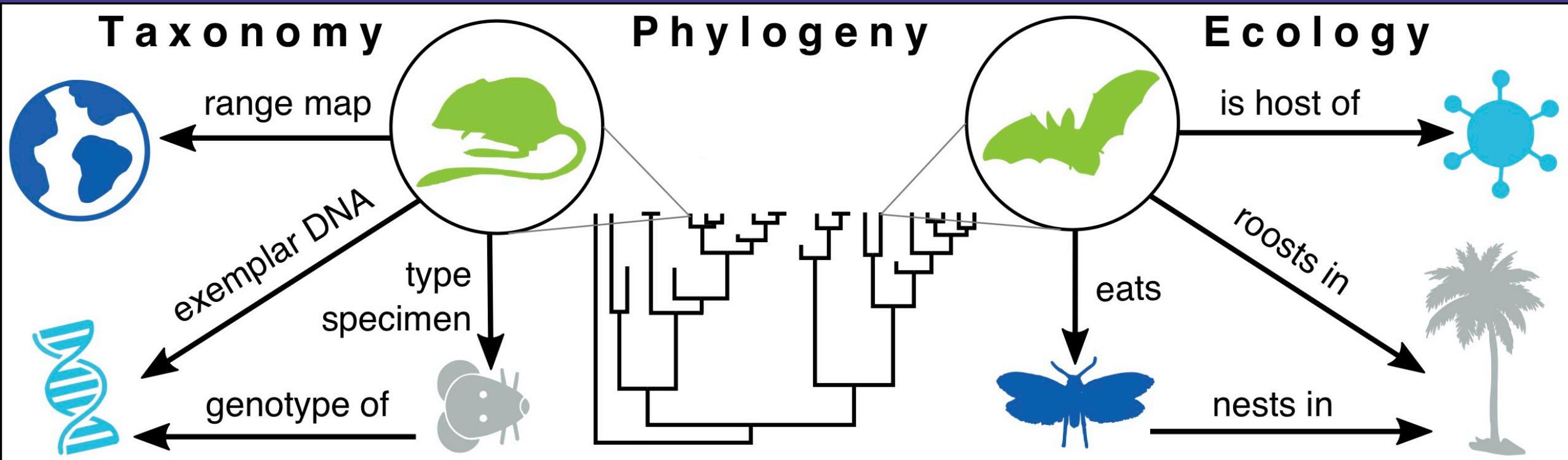




# Summary

## 1. What are the mammal species?

## 2. Which interactions?

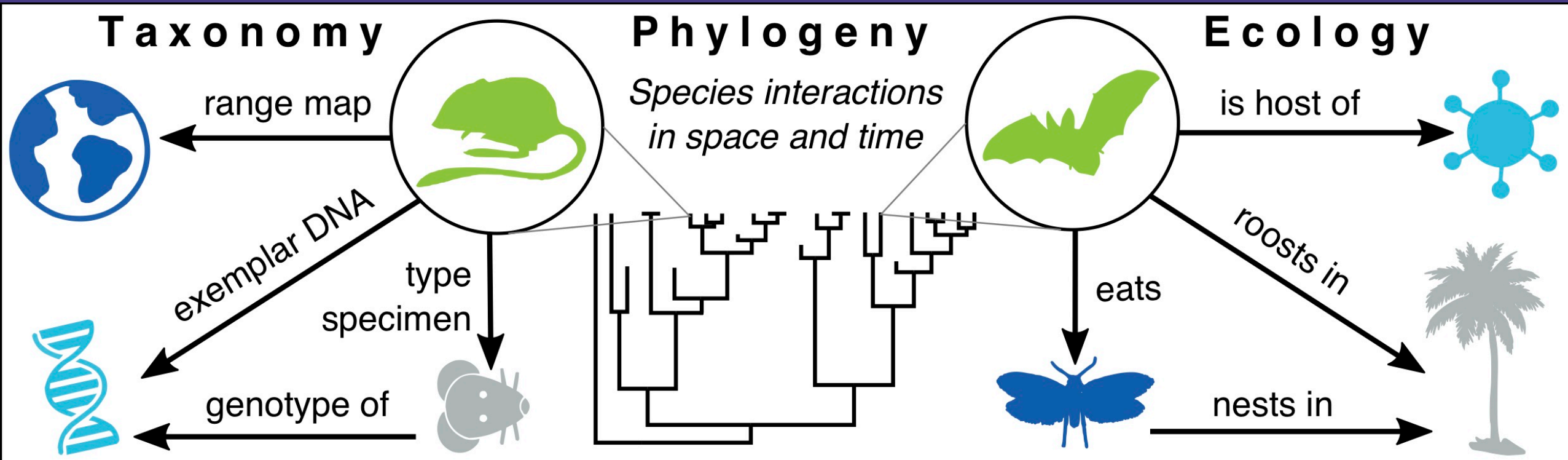




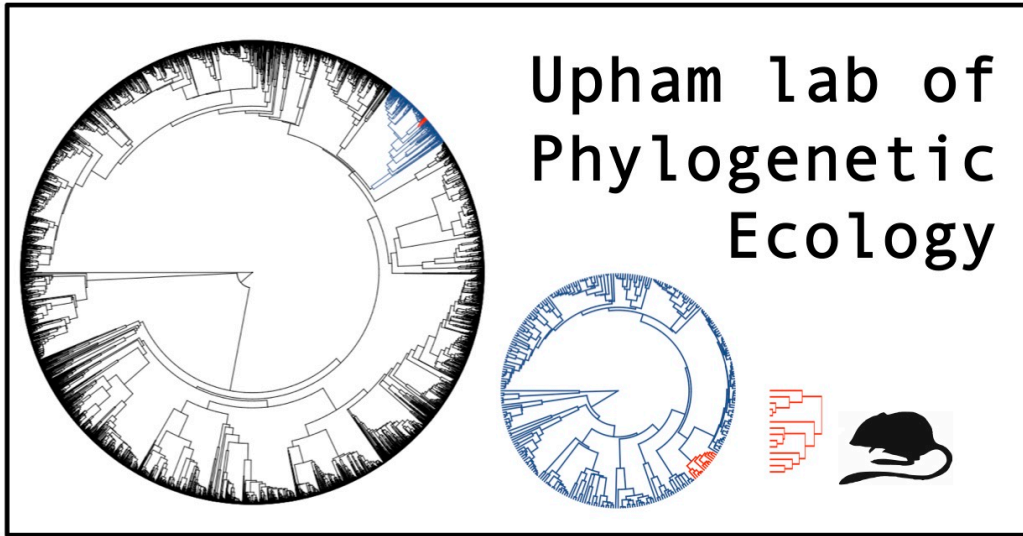
# Summary

## 1. What are the mammal species?

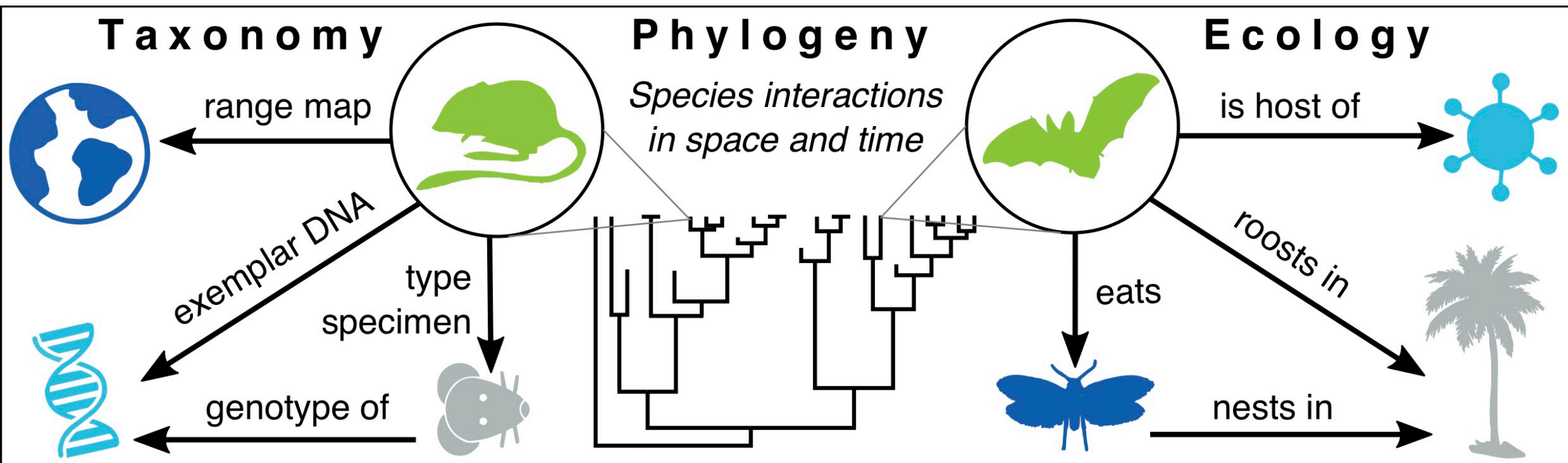
## 2. Which interactions?







**ASU**® Biodiversity Knowledge  
Integration Center  
Arizona State University





# Acknowledgments — thank you!



## **Arizona State Univ:**

Nico Franz, Beckett Sterner, Atriya Sen, Prashant Gupta, Nicole Veeder, Dakota Rowsey, Laura Steger, Jonathan Rees, Andrew Johnston, Laura Prado, Ángel Robles, Caleb Powell, Sharon Hall



## **CETAF Taskforce:**

Donat Agosti, Gábor Csorba, Mariya Dimitrova, Quentin Groom, Joe Miller, Deb Paul, Lyubomir Penev, Jorrit Poelen, DeeAnn Reeder, Nancy Simmons

