



Informatics Infrastructure at the start of the Second Decade of DNA Barcoding

SUJEEVAN RATNASINGHAM

BIODIVERSITY INSTITUTE OF ONTARIO

UNIVERSITY OF GUELPH



10+ Years



4M+ Records



0.5M+ Species

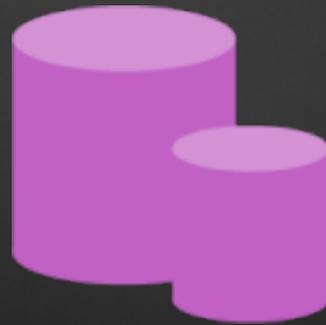


100+ Nations

Decade 1 - Capacity Building



Collaborative Networks

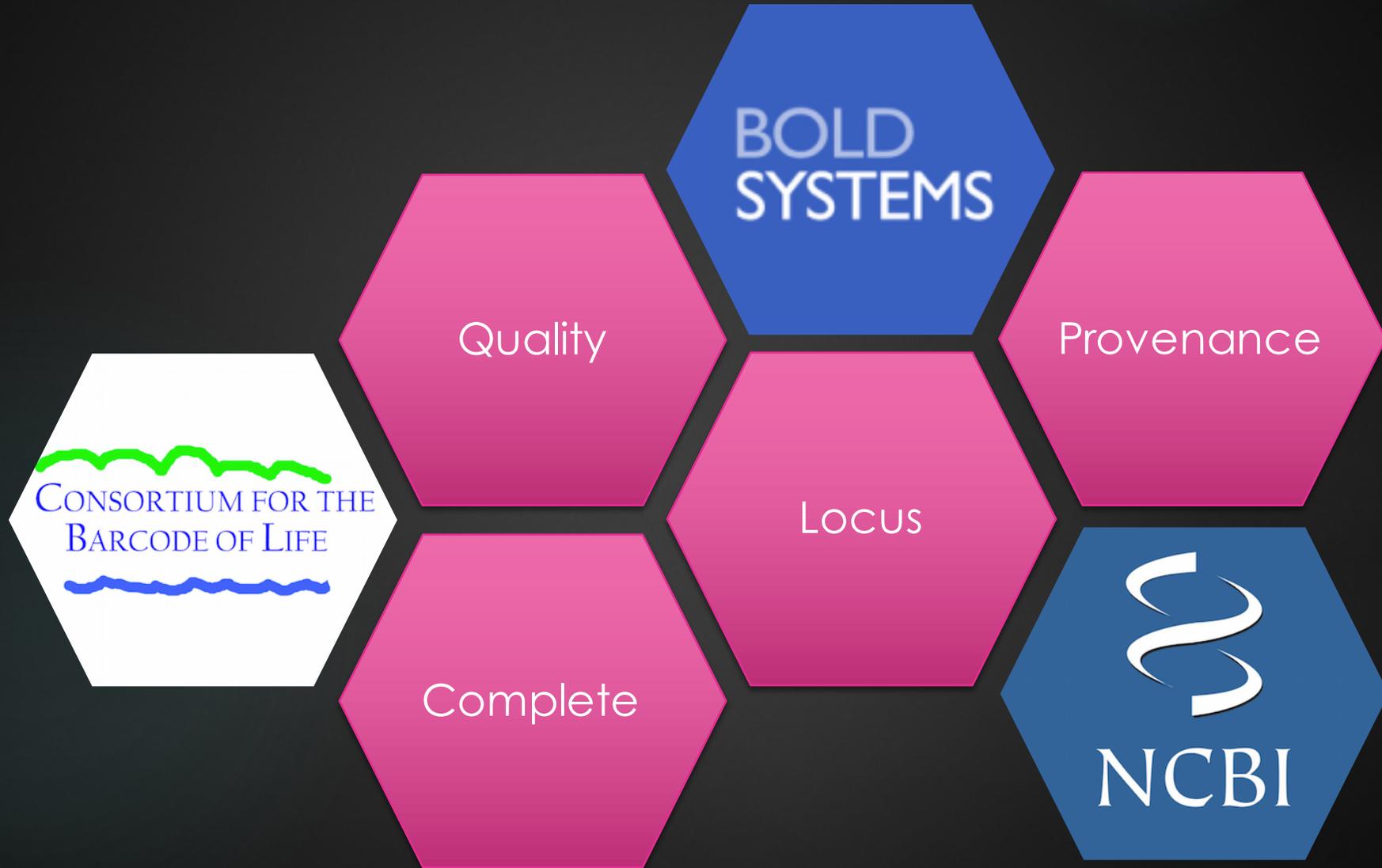


Community Database



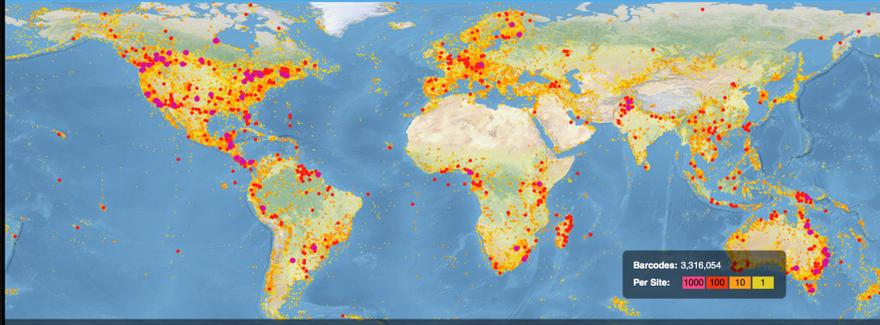
Software Tools

Data Standards



Building the Library

BOLD SYSTEMS | Databases | Taxonomy | Identification | Workbench | Resources



Barcodes: 3,316,054
Per Site: 1000 100 10 1

Taxonomy Search

Public Data Portal:
A data retrieval interface that allows for searching over 1.7M public records in BOLD using multiple search criteria including, but not limited to, geography, taxonomy, and depository.

Barcode Index Numbers:
A searchable database of Barcode Index Numbers (BINs), sequence clusters that closely approximate species.

DNA Barcode Education Portal:
A custom platform for educators and students to explore barcode data and contribute novel barcodes to the BOLD database.

Workbench:
An integrated data collection and analysis environment that securely supports the assembly and validation of DNA barcodes and ancillary sequences.

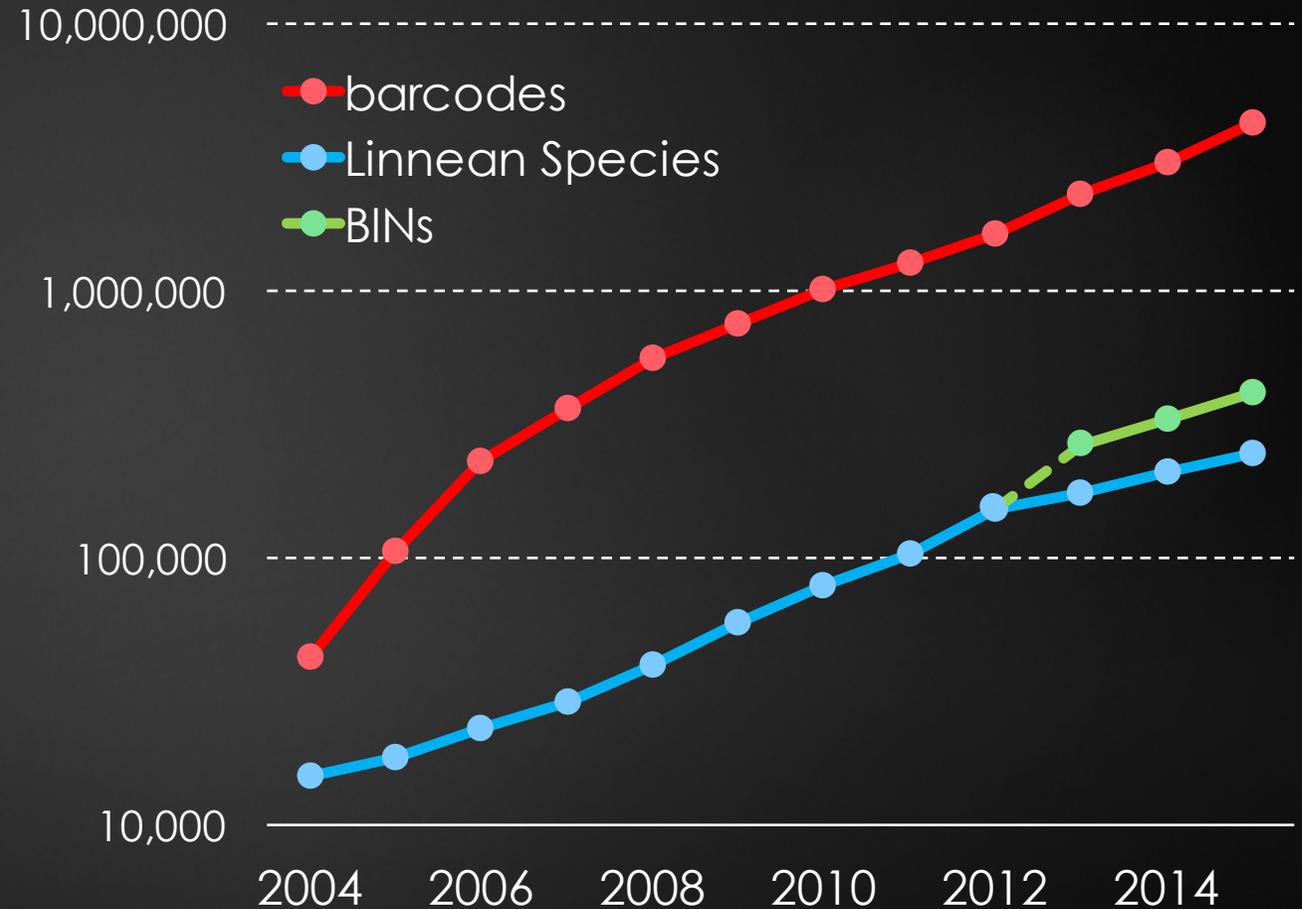
The Barcode of Life Data Systems is designed to support the generation and application of DNA barcode data. The platform consists of four main modules: a data portal, a database of barcode clusters, an educational portal, and a data collection workbench.

Sequence statistics		Species coverage (formally described)	
Barcode clusters for animals (BINs)	350,542	Animals	143,471
All Sequences	3,837,925	Plants	53,498
Barcode Sequences	3,316,054	Fungi & Other Life	16,358

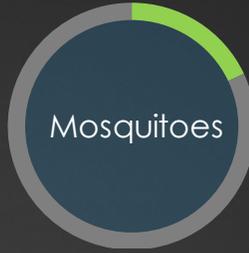






Community Benchmarks



Taxonomic

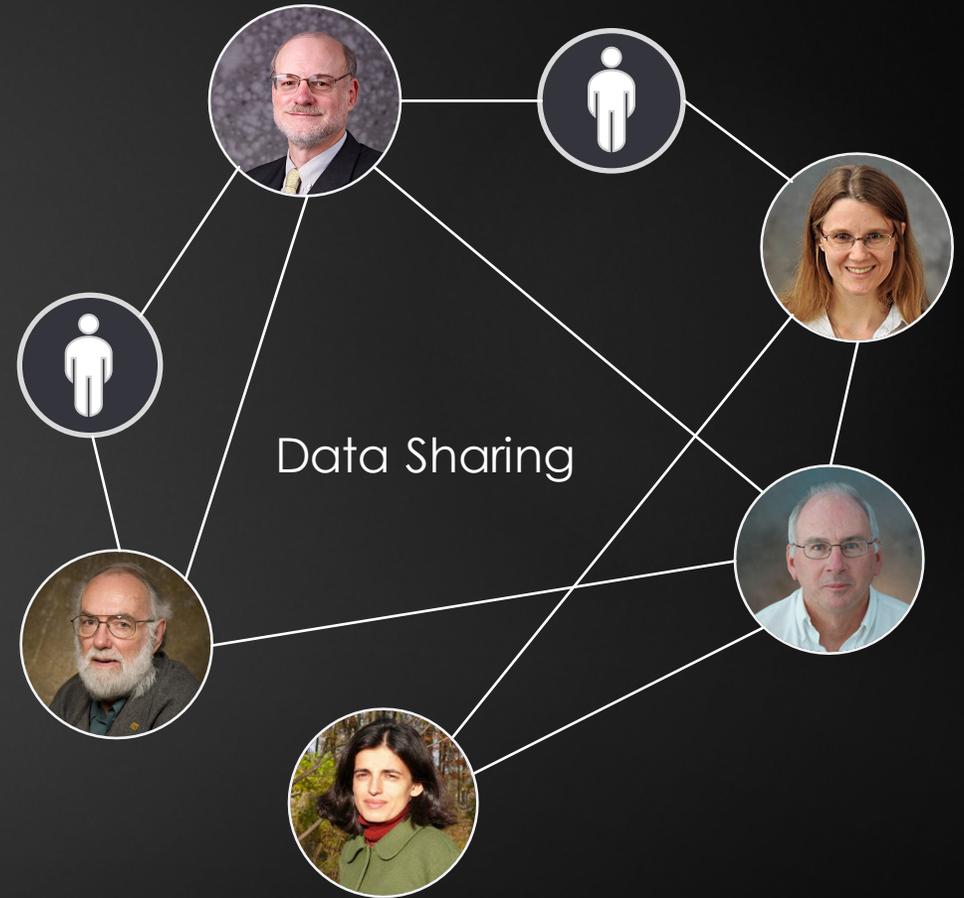
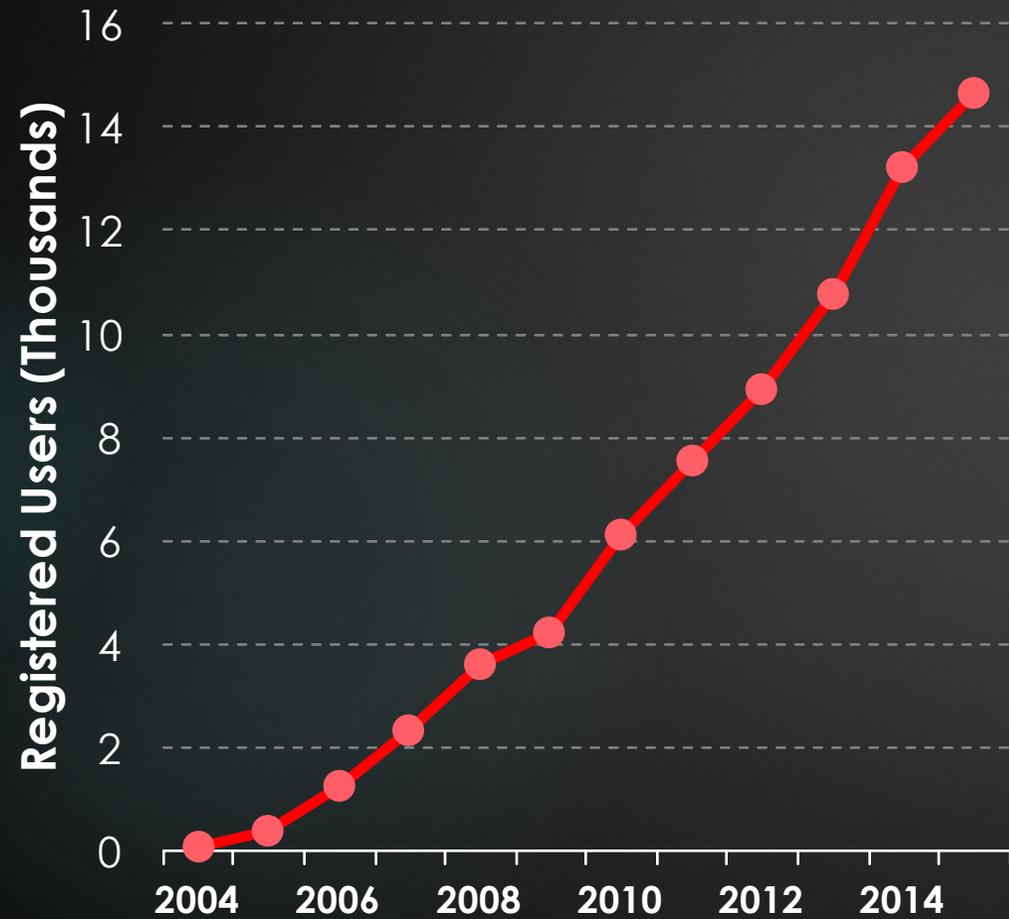
Geographic



Thematic



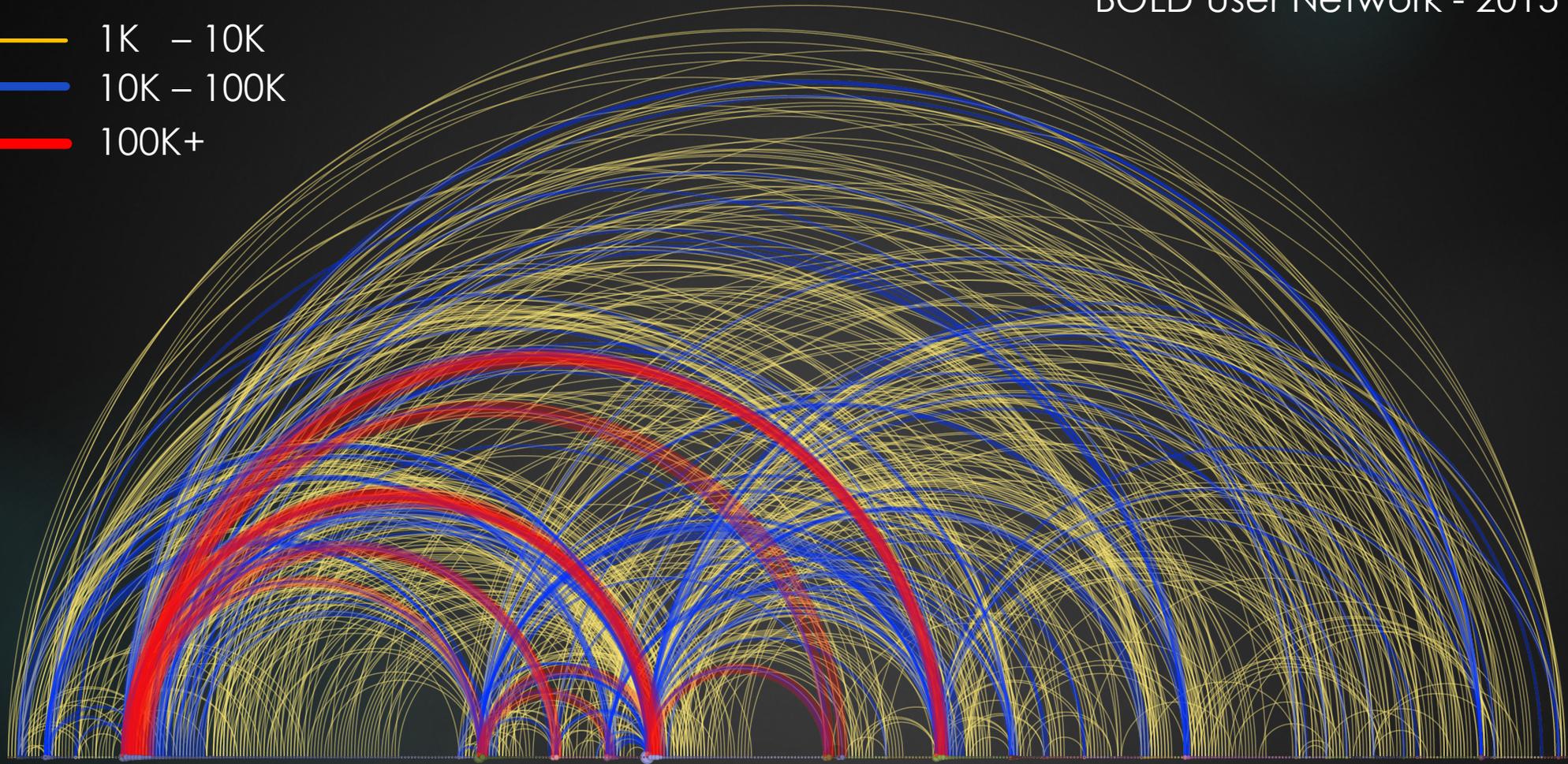
Collaborative Networks



2005 – 102 users from 30 institutions

BOLD User Network - 2015

- 1K – 10K
- 10K – 100K
- 100K+



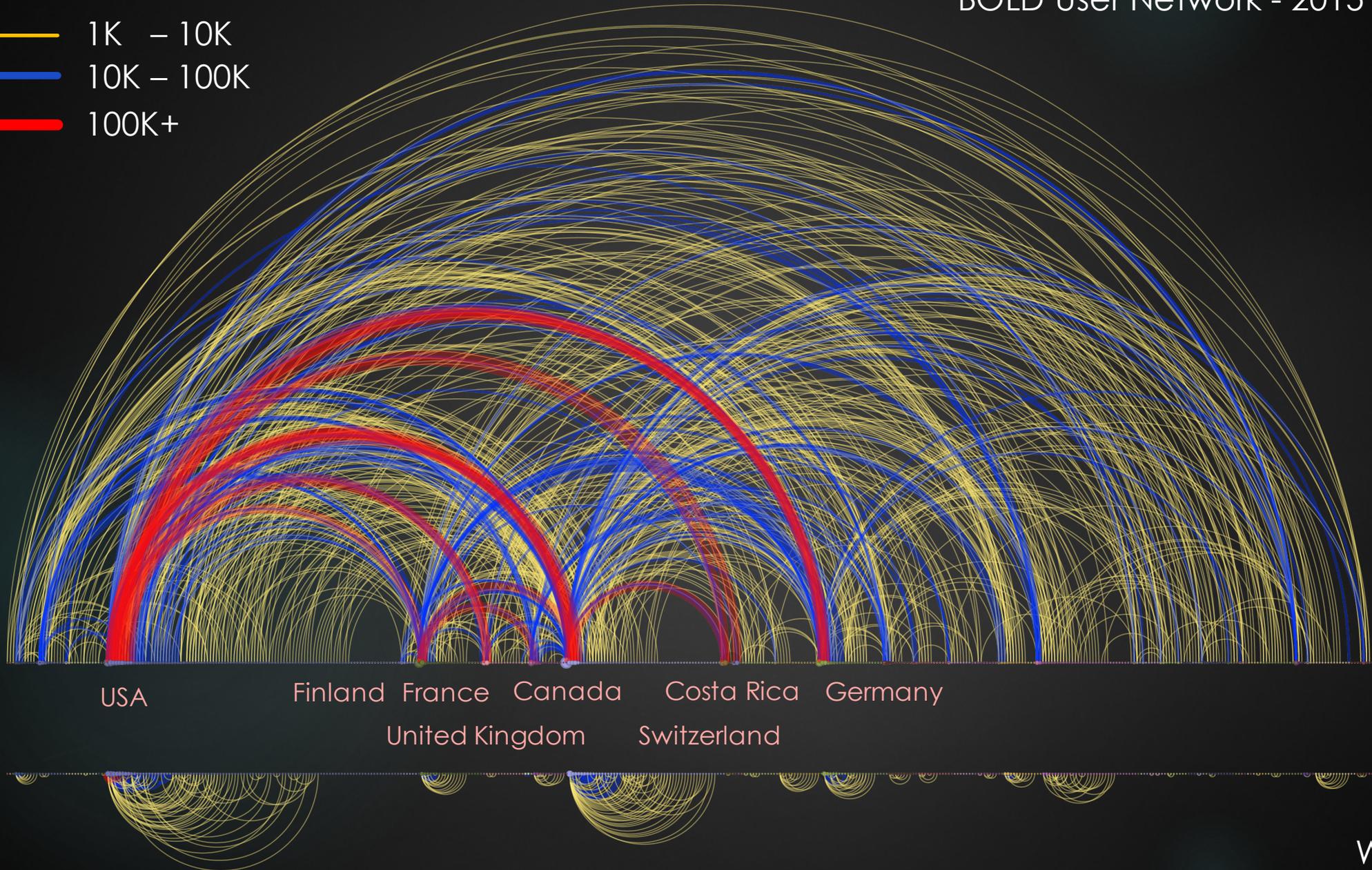
Across Nations

1000+ Institutions from 94 countries sharing data on BOLD

Within Nations

BOLD User Network - 2015

- 1K – 10K
- 10K – 100K
- 100K+

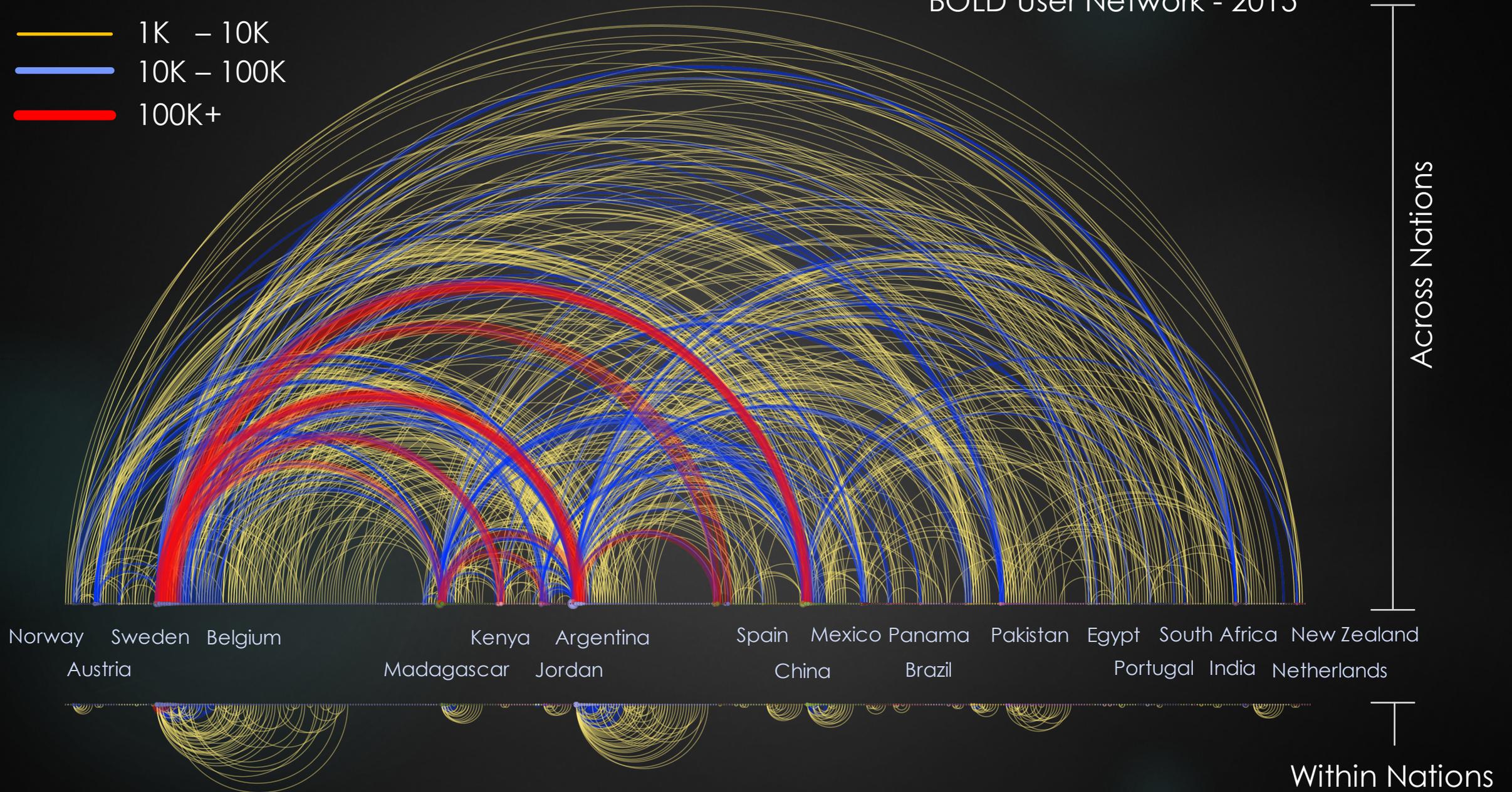


Across Nations

Within Nations

BOLD User Network - 2015

- 1K – 10K
- 10K – 100K
- 100K+



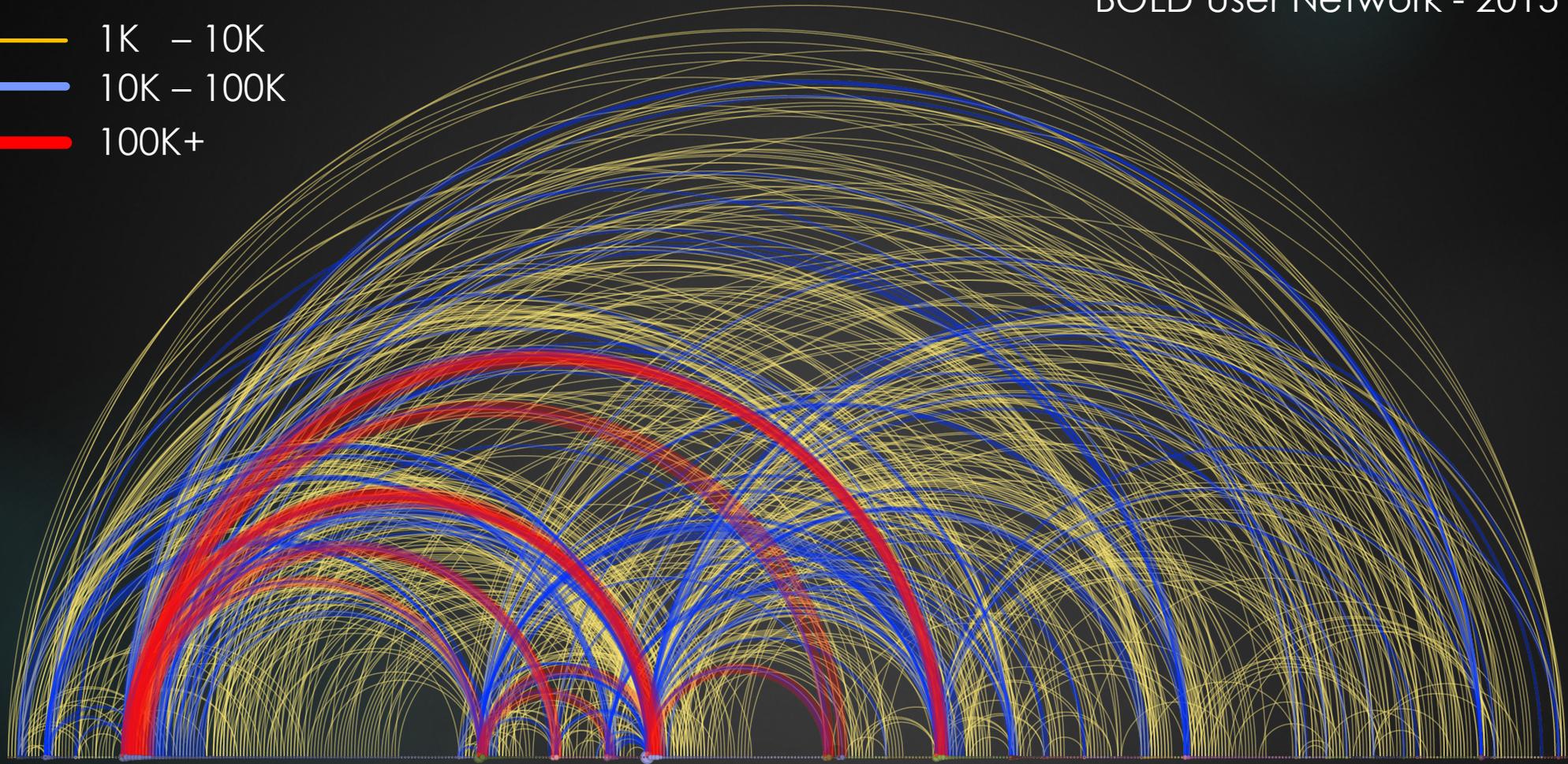
Norway Sweden Belgium Kenya Argentina Spain Mexico Panama Pakistan Egypt South Africa New Zealand
Austria Madagascar Jordan China Brazil Portugal India Netherlands

Across Nations

Within Nations

BOLD User Network - 2015

- 1K - 10K
- 10K - 100K
- 100K+



Across Nations

66 Other Countries from Every Continent

Within Nations

Testing the Library Depth



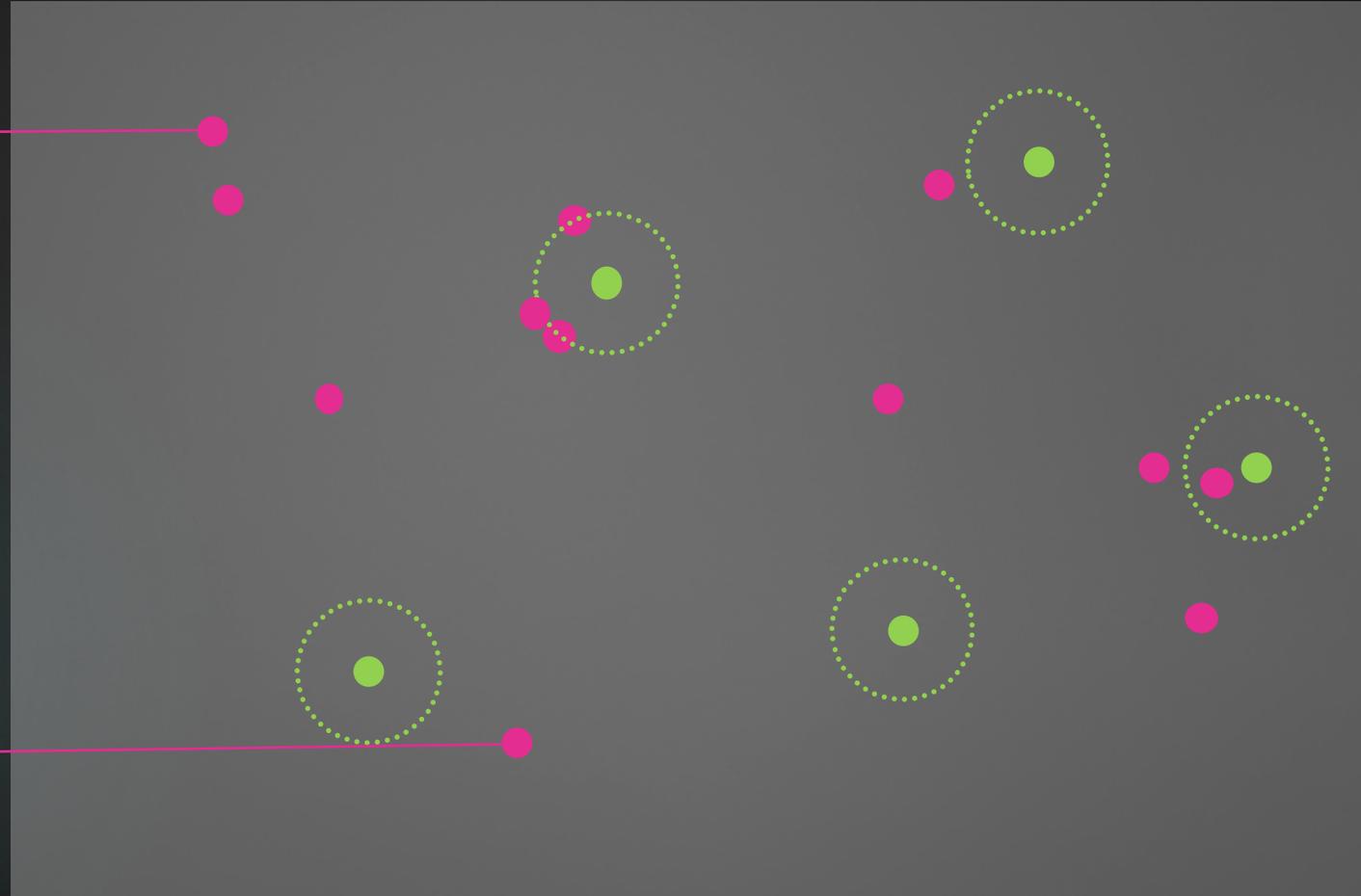
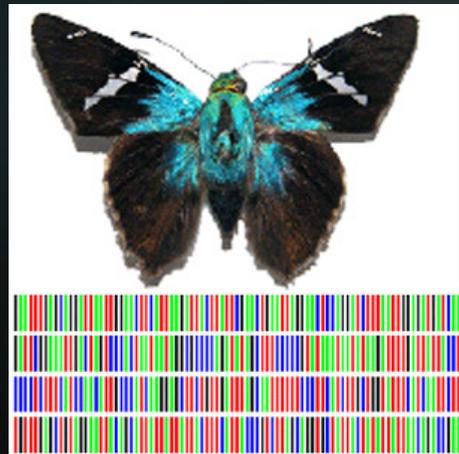
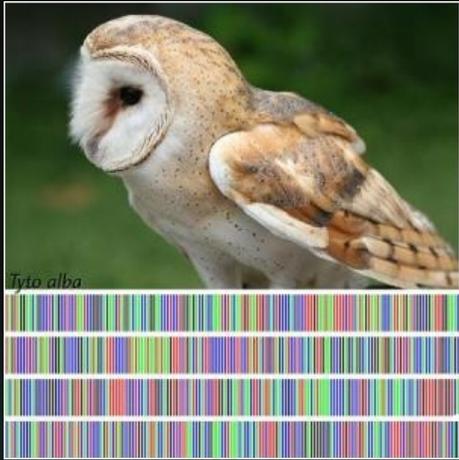
Animals

400K+
163K+
80%
500+

Species
with Full Taxonomy
of the BOLD Library
Orders

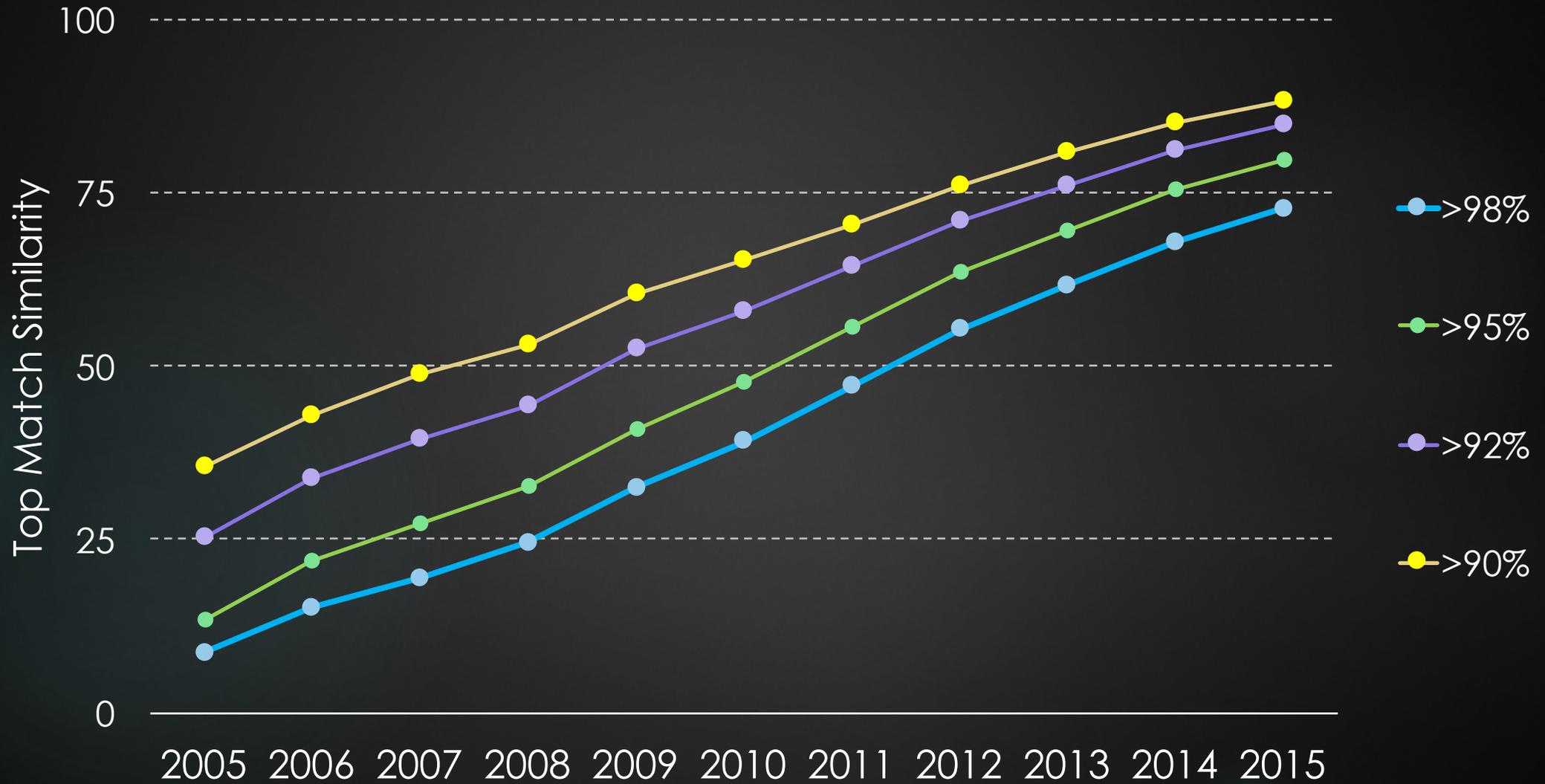
Test Data:
4000 species from
200 orders,
20 per order

Testing the Library Depth

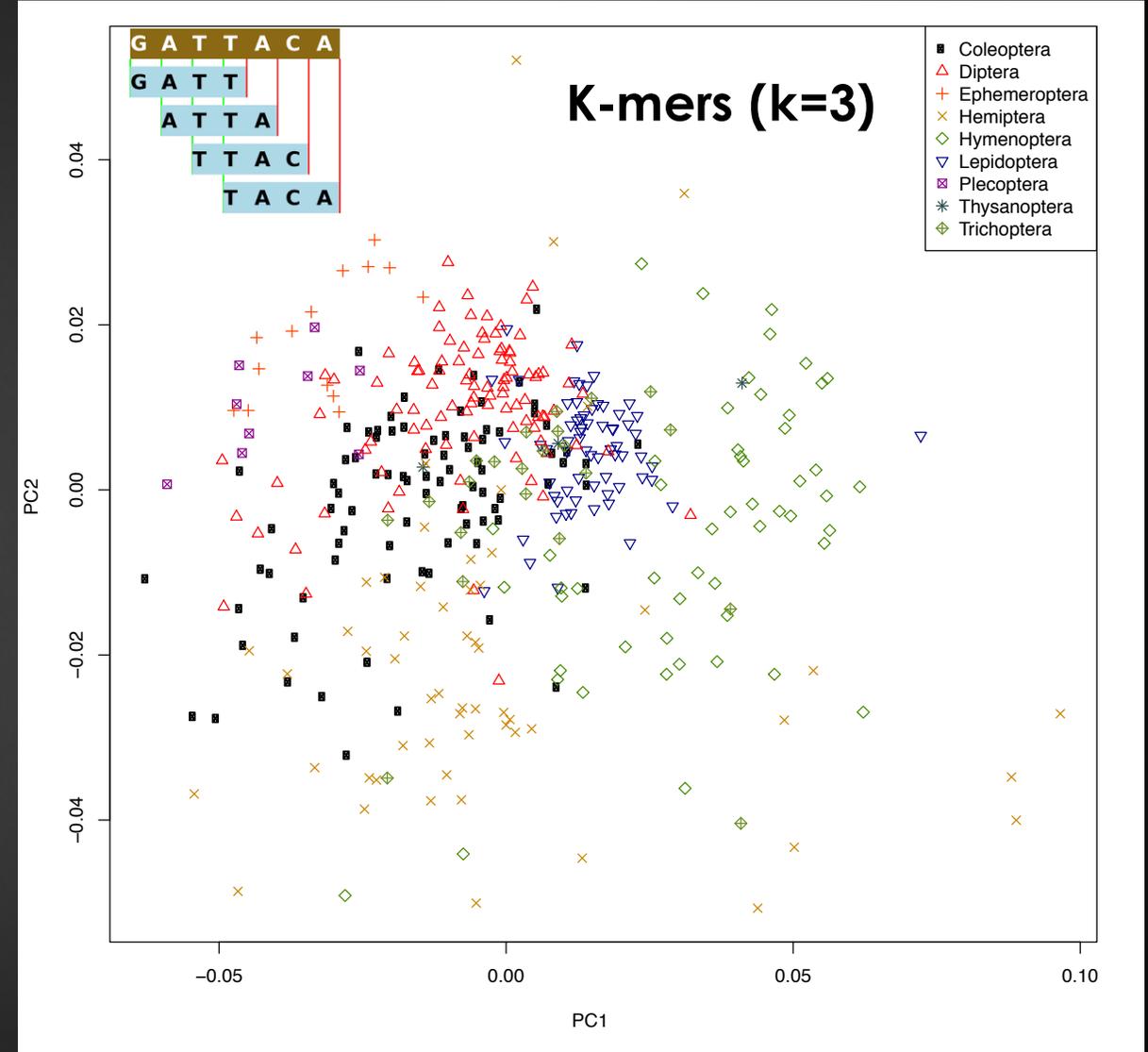
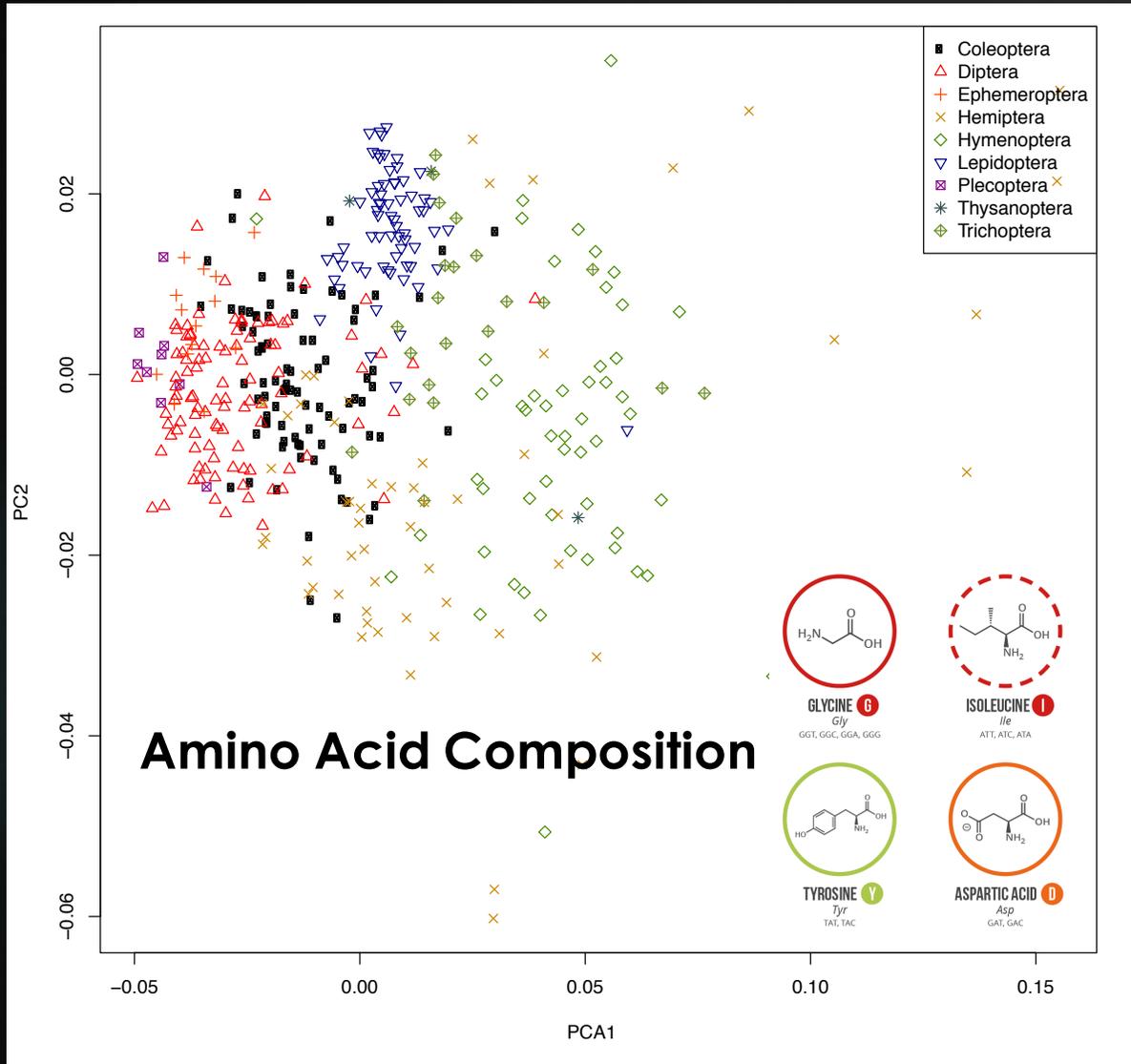


- Reference
- Queries

Testing the Library Depth



Testing the Library Depth



Barcode Index Number (BIN)

The screenshot shows the PLOS ONE article page for the paper 'A DNA-Based Registry for All Animal Species: The Barcode Index Number (BIN) System' by Sujeewan Ratnasingham and Paul D. N. Hebert. The page features a navigation bar with 'PLOS ONE', 'Subject Areas', 'For Authors', and 'About Us'. A search bar is located in the top right. Below the navigation, there are statistics: 9,111 views, 35 citations, 92 saves, and 44 shares. The article title is prominently displayed, followed by the authors' names and the publication date (July 08, 2013). The abstract is visible, starting with 'Because many animal species are undescribed...'. A sidebar on the left contains a table of contents with links to 'Abstract', 'Introduction', 'Materials and Methods', 'Results', 'Discussion', 'Supporting Information', 'Acknowledgments', 'Author Contributions', and 'References'. On the right, there are buttons for 'Download PDF', 'Print', and 'Share', along with a 'Subject Areas' section listing 'Biodiversity', 'DNA sequence anal...', 'Graphs', 'Moths and butterflies', 'Multiple alignment c...', 'Sequence alignment', 'Sequence analysis', and 'Taxonomy'. At the bottom, there is a 'Citation' box with the full citation text, an 'Editor' box with the name 'Diego Fontaneto', and a 'Received' box with dates.

Algorithm

- Tuned to the marker (COX1)
- Fixed parameters for balanced OTU generation
- Uses prior threshold but refines for each group

Registry

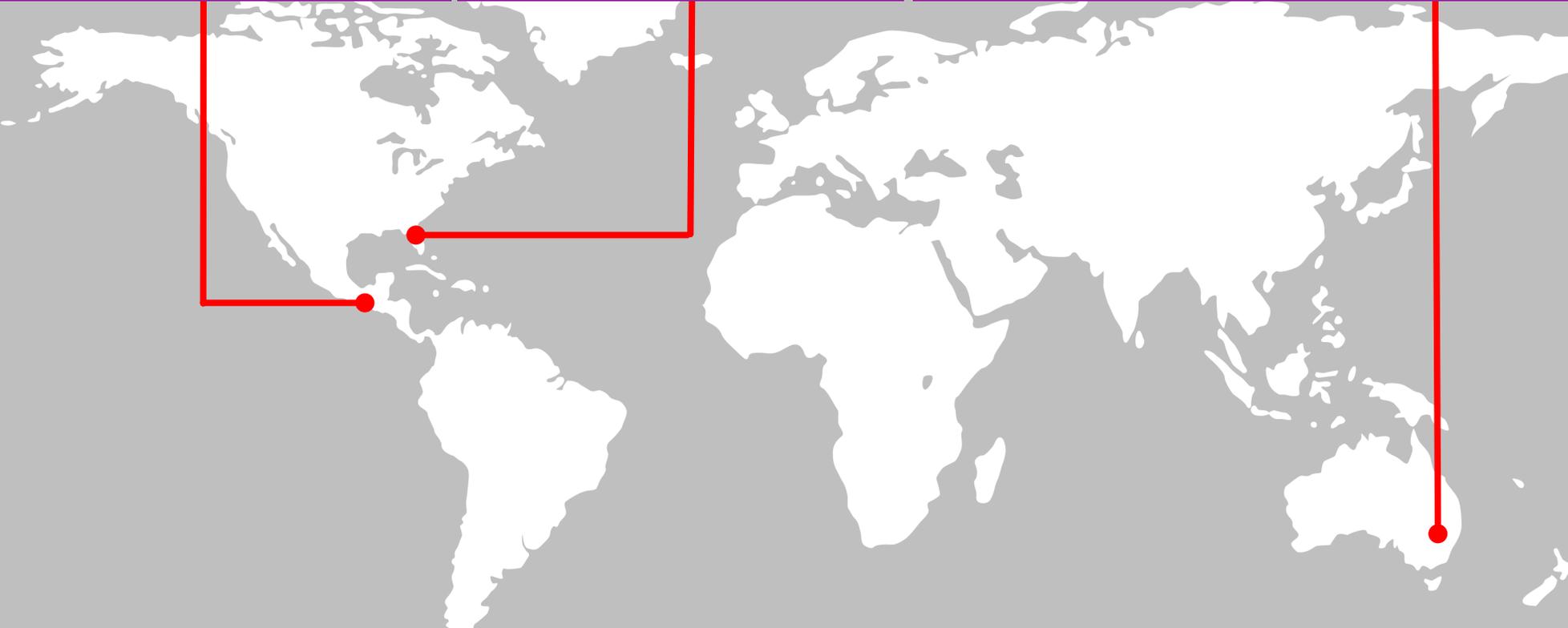
- Occurrence of DNA Barcode (place and time)
- Aggregation of all associated metadata
- Reusable - works across studies



Oraesia Janzen 03

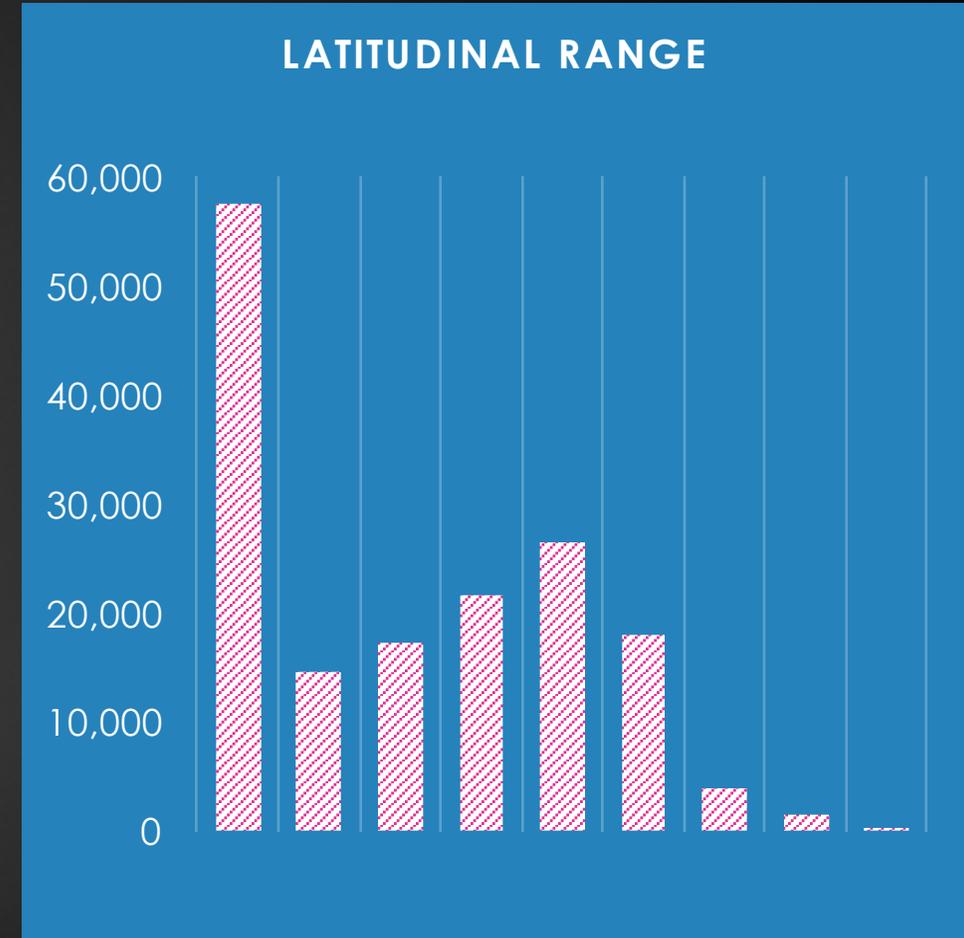
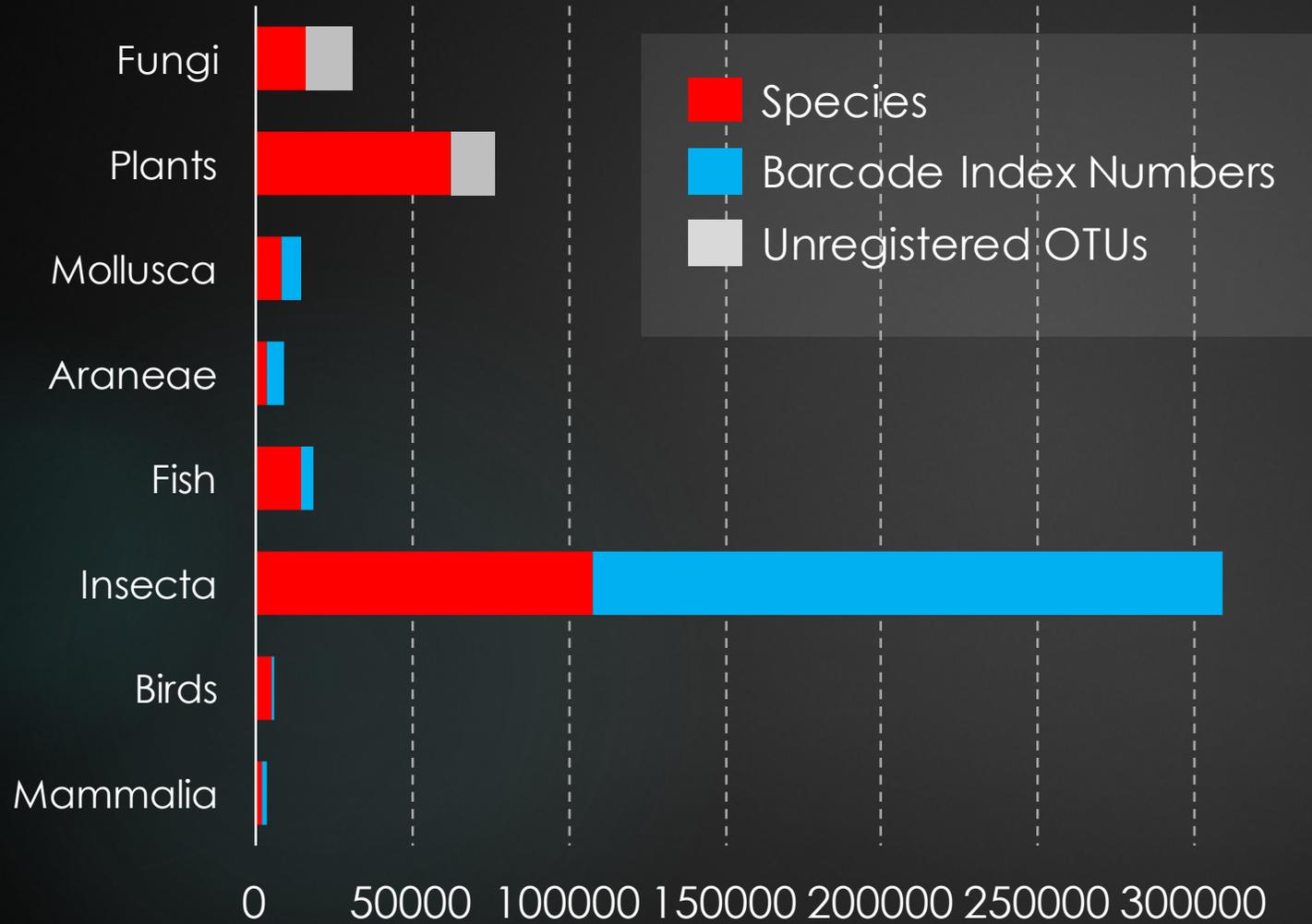
Notioplusia illustratata

Ctenoplusia sp. ANIC1



BOLD:AAF2716

Importance of registering OTUs

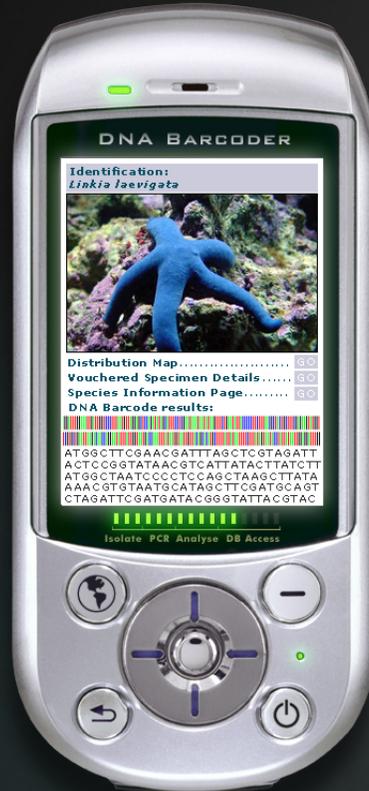




goinggoing

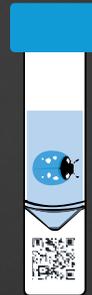
The technology that links taxonomy and Star Trek

*What made Star Trek's original tricorder a great piece of fictional technology, writes **Maggie Koerth-Baker**, wasn't its sci-fi looks. It was what it did.*



Allow anyone, anywhere, to identify any organism

lifescanner





CHAETODON SPECULUM
rare


Chaetodon, Chaetodontidae,
Perciformes, Actinopterygii,
Chordata

STATISTICS

12 records available

8% are public

1 species represented



CHAETODON SPECULUM
rare


Chaetodon, Chaetodontidae,
Perciformes, Actinopterygii,
Chordata

DETAILS

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Fusce non ultricies ante. Tristique facilisis quam lacus, eget venenatis fella praetium quis. In hac habitasse platea dictumst. Ut eu odio non odio laoreet sagittis. Vestibulum eget metus laoreet, pretium tortor vitae, volutpat ante. Pellentesque luctus nisl ut erat viverra, nec aliquet mauris suscipit. Vivamus placerat volutpat odio id aliquet. Donec fringilla pretium elit et interdum. Nullam dictum orci in consequat gravida.

Sed nunc lorum, vulputate quis diam eget, pharetra pulvinar sem. Nulla facilis. Nullam pellentesque venenatis facilisis. Praesent vulputate quis tortor vitae fringilla. Quisque vitae purus libero. Nullam volutpat varius dignissim. Nam nulla lacus sit amet dui vestibulum posuere. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Morbi varius metus tellus, sit amet imperdiet neque ultricies ut. Nullam sem diam, luctus et erat in, blandit fermentum mi. Cras quis tortor pretium, volutpat nisl sed, aliquam tortor.

LOCATION

neque ultricies ut. Nullam sem diam, luctus et erat in, blandit fermentum mi. Cras quis tortor pretium, volutpat nisl sed, aliquam tortor.

LOCATION

We don't have a location for this.

Where on Earth?



IMAGES



STATISTICS

12 records available

8% are public



Panthera
rare

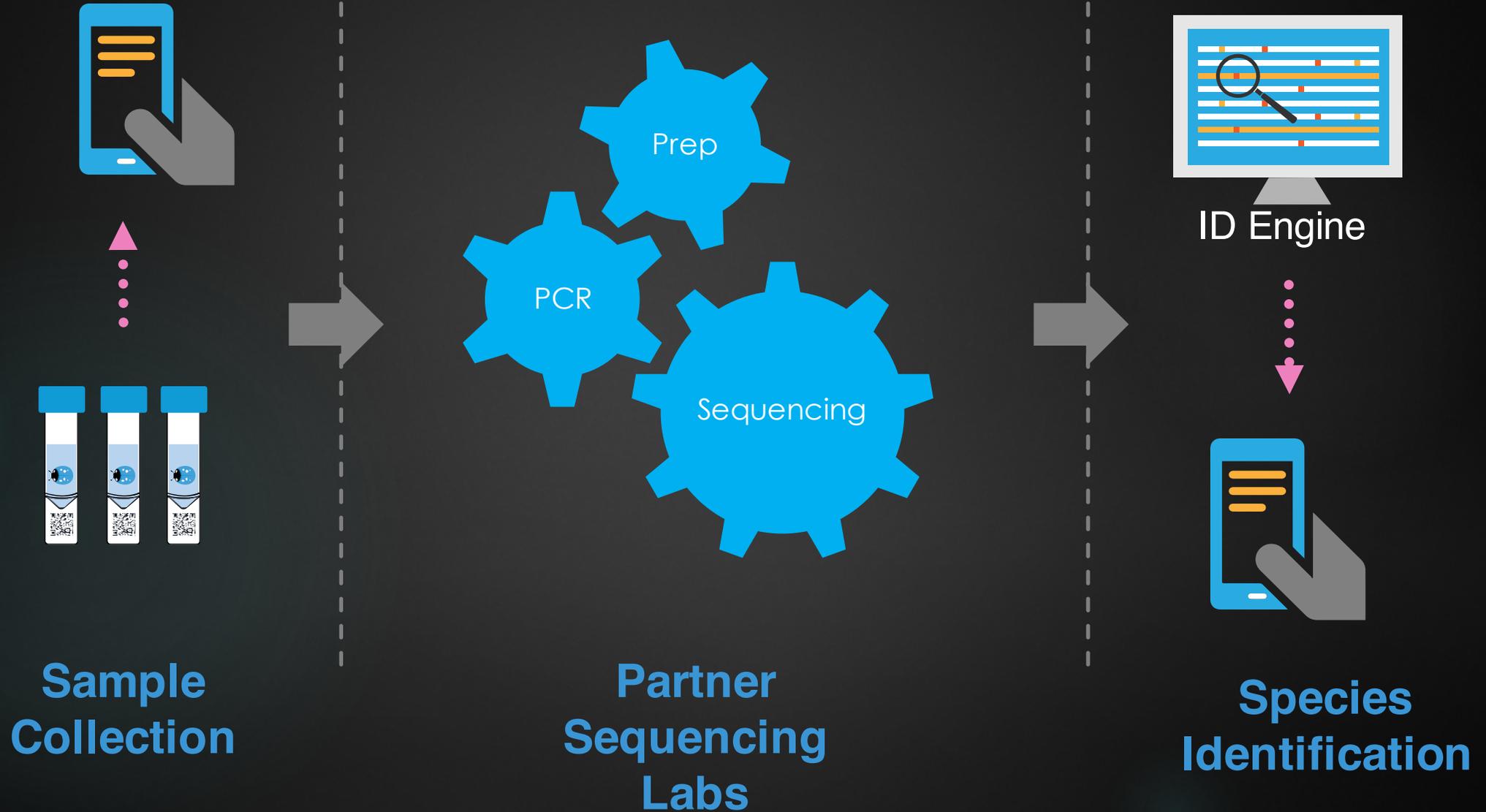

Chaetodon, Chaetodontidae,
Perciformes, Actinopterygii,
Chordata

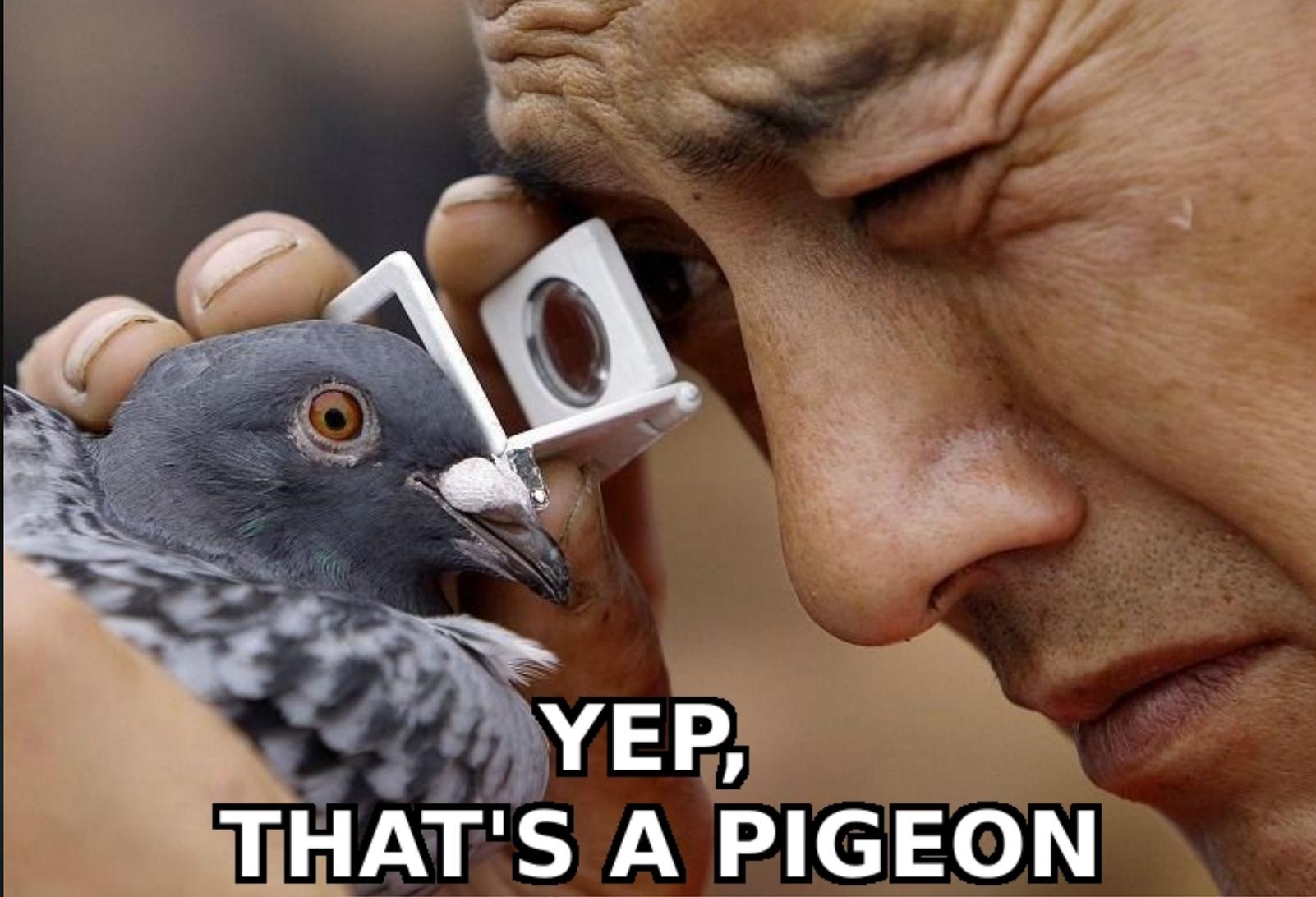
SPECIES	EXPLORE MORE
GENUS	Panthera
SUB-FAMILY	Pantherinae
FAMILY	Felidae
ORDER	Carnivora
CLASS	Mammalia
PHYLUM	Chordata

Living things are organized hierarchically

[Learn More Here.](#)

LifeScanner Solution Overview





**YEP,
THAT'S A PIGEON**

nature

Rise of the citizen scientist

From the oceans to the soil, technology is changing the part that amateurs can play in research. But this greater involvement raises concerns that must be addressed.

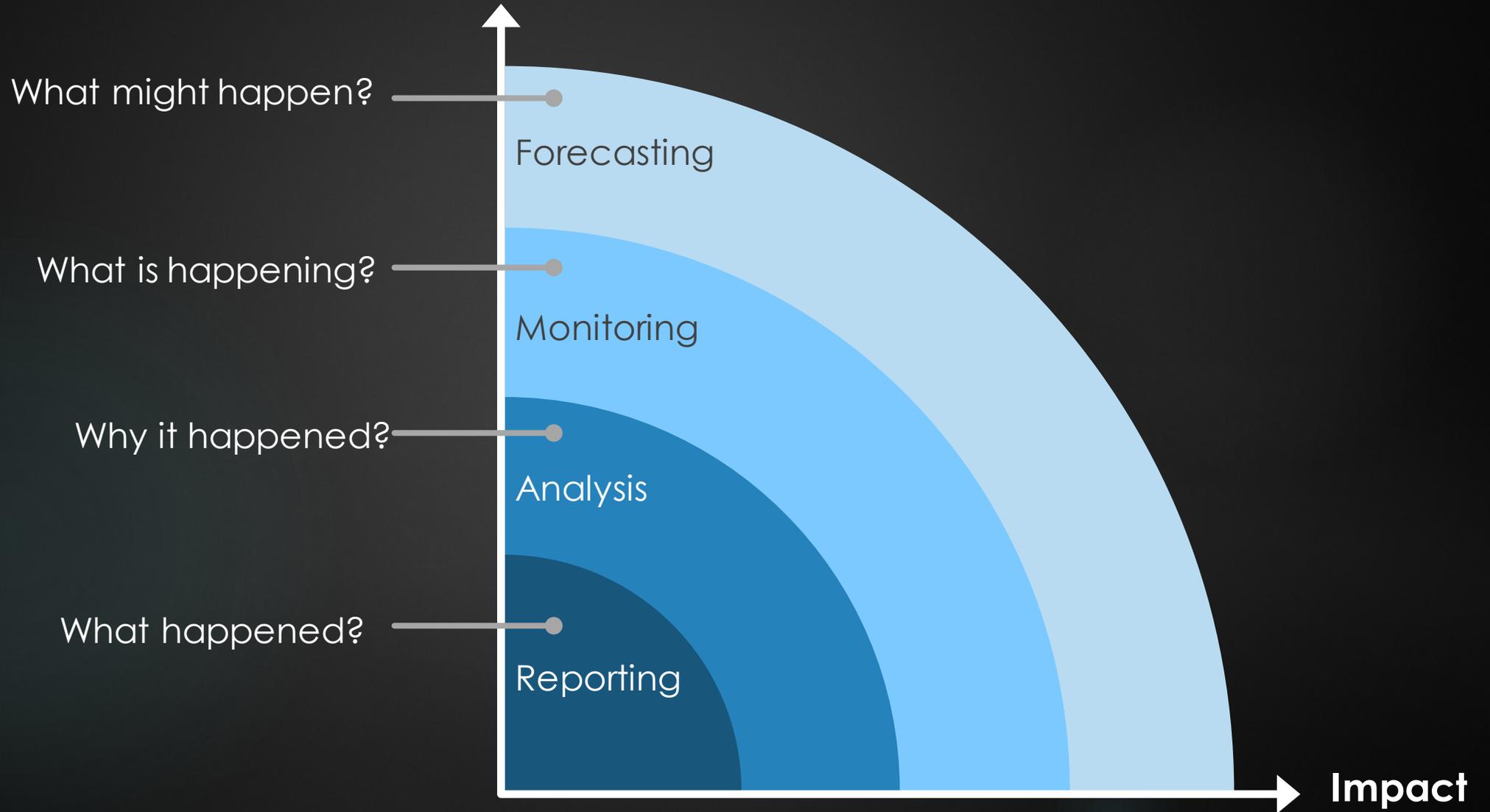
lifescanner



Moving into decade 2

Embrace Big Data

Complexity (Data volume & Dimensionality)



Support for multiple scales

10^2



10^3



10^4



Support for NGS based Barcoding



BOLD4

Twitter Facebook Google+ Instagram YouTube

BOLD SYSTEMS

DATABASES IDENTIFICATION TAXONOMY WORKBENCH RESOURCES

LOGIN

BARCODE OF LIFE DATA SYSTEMS

A BIOINFORMATICS PLATFORM FOR DNA BARCODING OF ANIMAL, PLANT, AND FUNGAL SPECIES.

WHERE TO START EXPLORE THE DATA

Search

DESIGNED TO SUPPORT THE GENERATION & APPLICATION OF DNA BARCODES

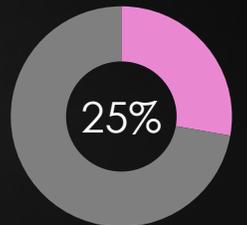
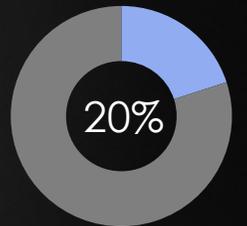
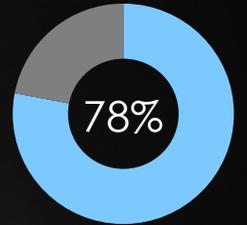
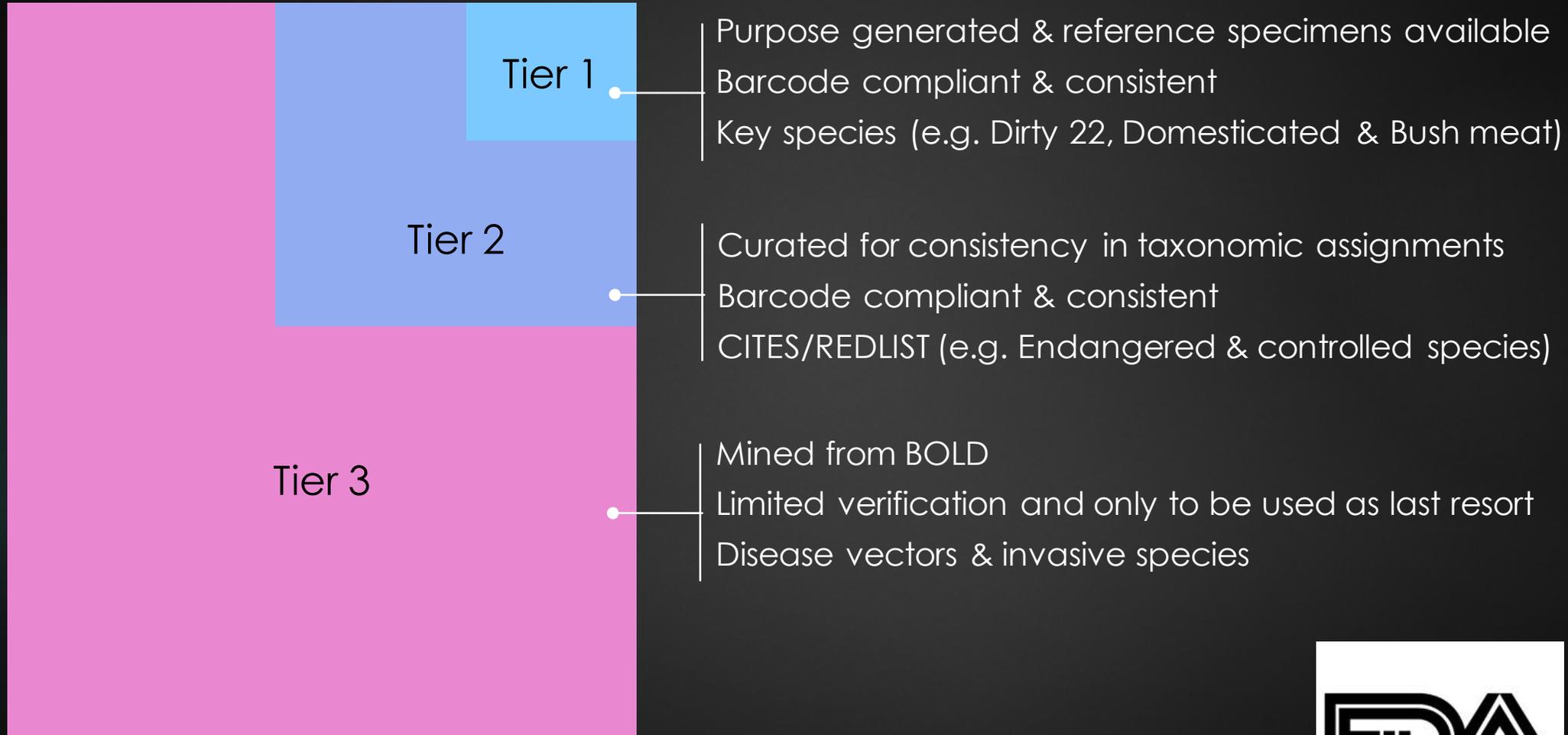
- PUBLIC DATA PORTAL**
A DATA RETRIEVAL INTERFACE THAT ALLOWS FOR SEARCHING OVER 1.7M PUBLIC RECORDS IN BOLD USING MULTIPLE SEARCH CRITERIA INCLUDING, BUT NOT LIMITED TO, GEOGRAPHY, TAXONOMY, AND DEPOSITORY.
- DNA BARCODE EDUCATION PORTAL**
A CUSTOM PLATFORM FOR EDUCATORS AND STUDENTS TO EXPLORE BARCODE DATA AND CONTRIBUTE NOVEL BARCODES TO THE BOLD DATABASE.
- BARCODE INDEX NUMBERS**
A SEARCHABLE DATABASE OF BARCODE INDEX NUMBERS (BINS), SEQUENCE CLUSTERS THAT CLOSELY APPROXIMATE SPECIES.
- WORKBENCH**
AN INTEGRATED DATA COLLECTION AND ANALYSIS ENVIRONMENT THAT SECURELY SUPPORTS THE ASSEMBLY AND VALIDATION OF DNA BARCODES AND ANCILLARY SEQUENCES.

BOLD4

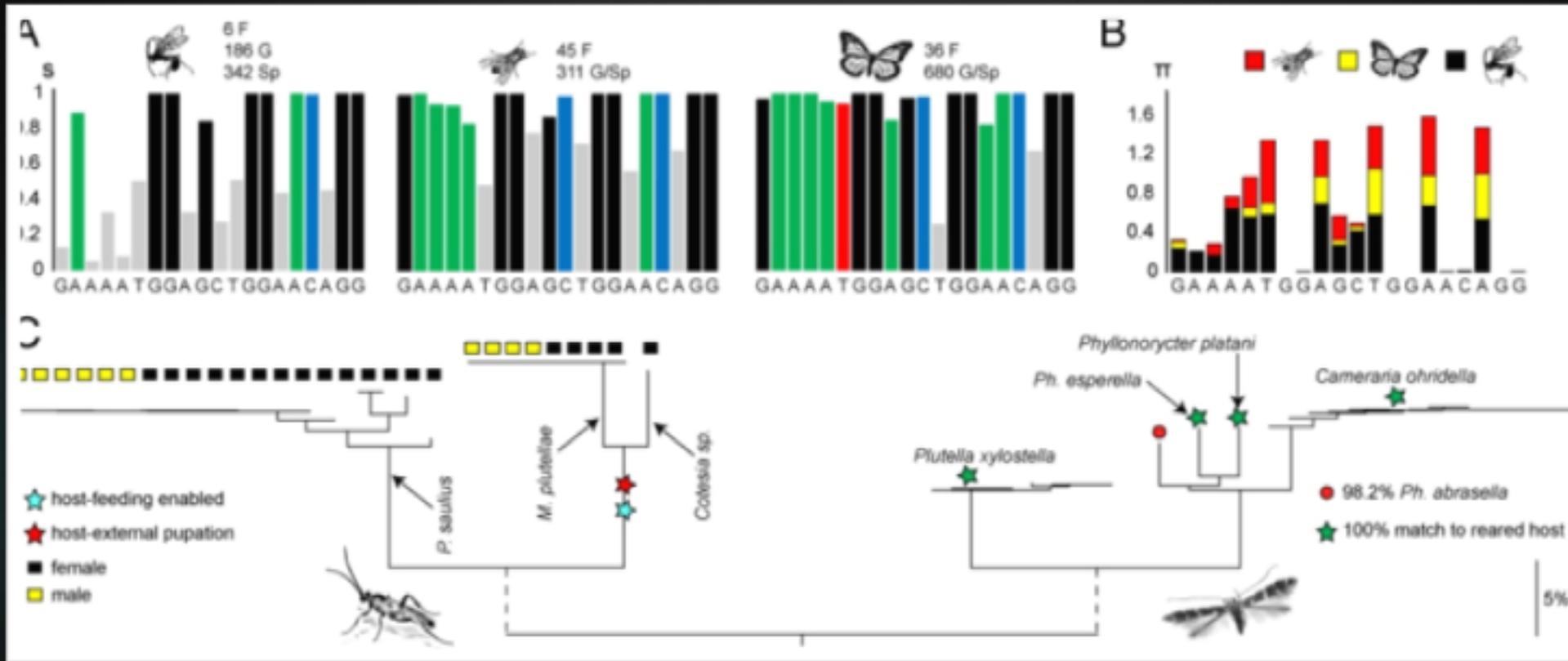
Launch of Beta on Sept 29, 2015

Advancing the Automation of
Species Identification

Community Curated Libraries

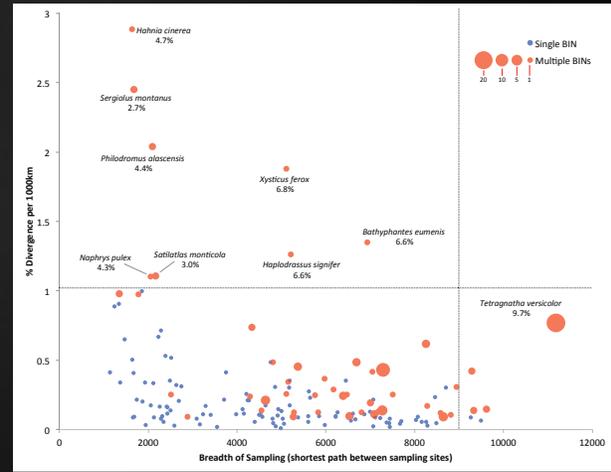
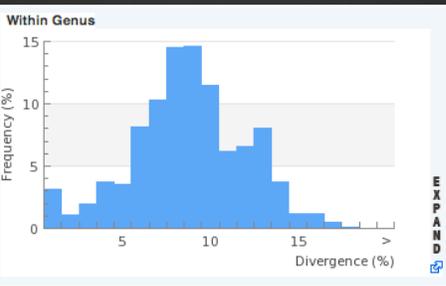
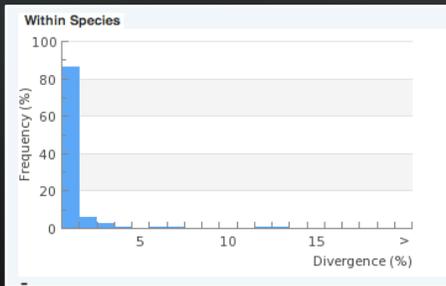
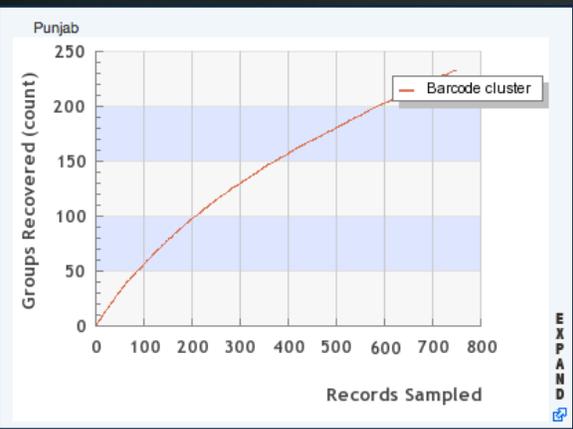
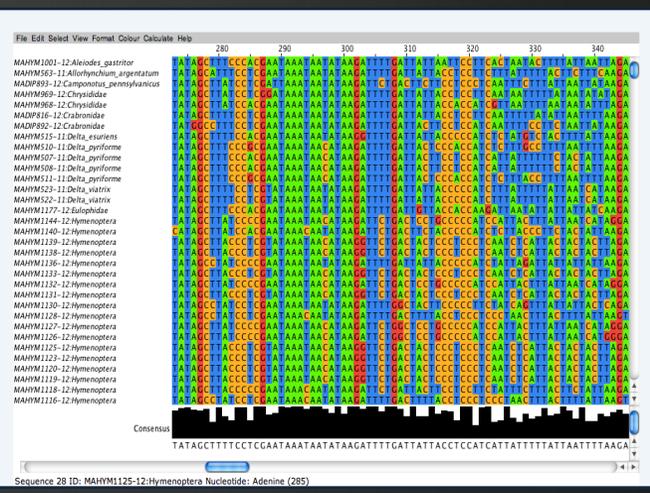
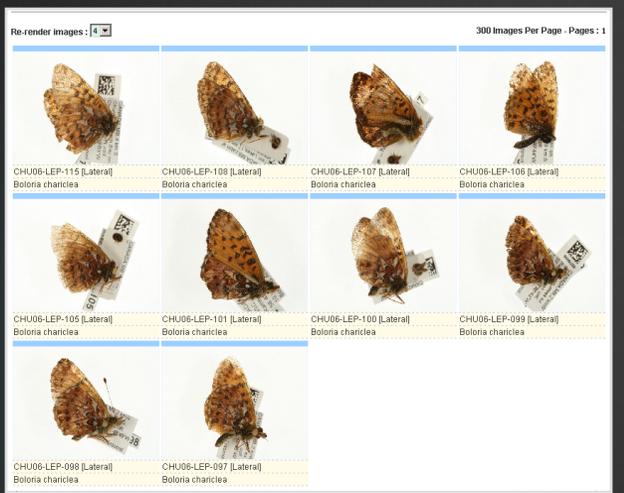
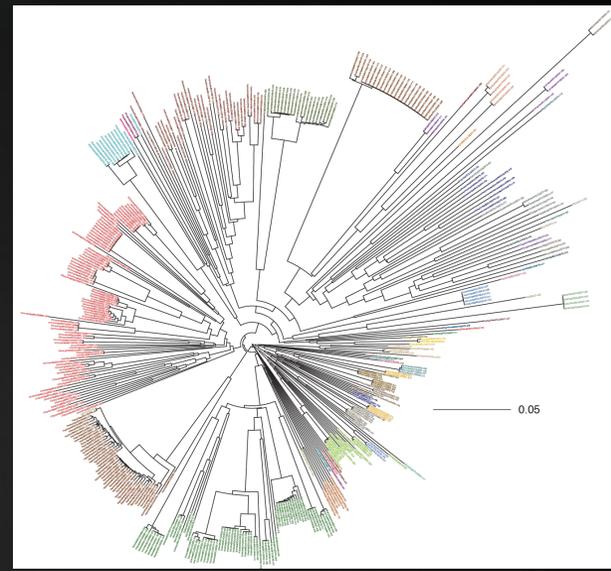
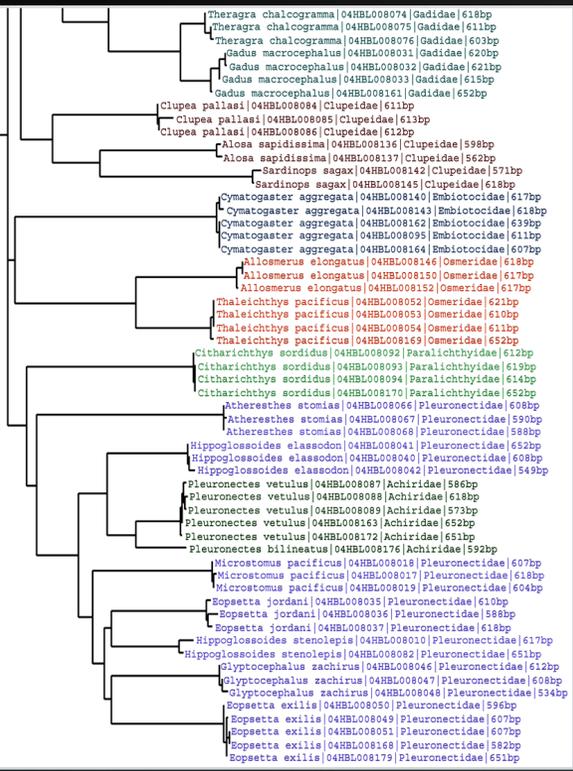


Community Defined Metadata Extensions



Rougerie R, Smith AM, Fernandez-Triana J, Lopez-Vaamonde C, Ratnasingham S, Hebert PDN. 2011. Molecular analysis of parasitoid linkages (MAPL): gut contents of adult parasitoid wasps reveal larval host. *Mol Ecol* 20:179-186.

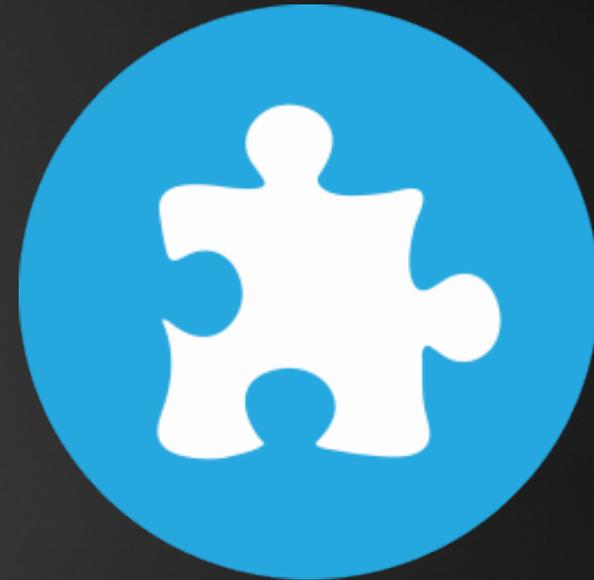
More Analytical Tools



Community Developed Analytical Tools



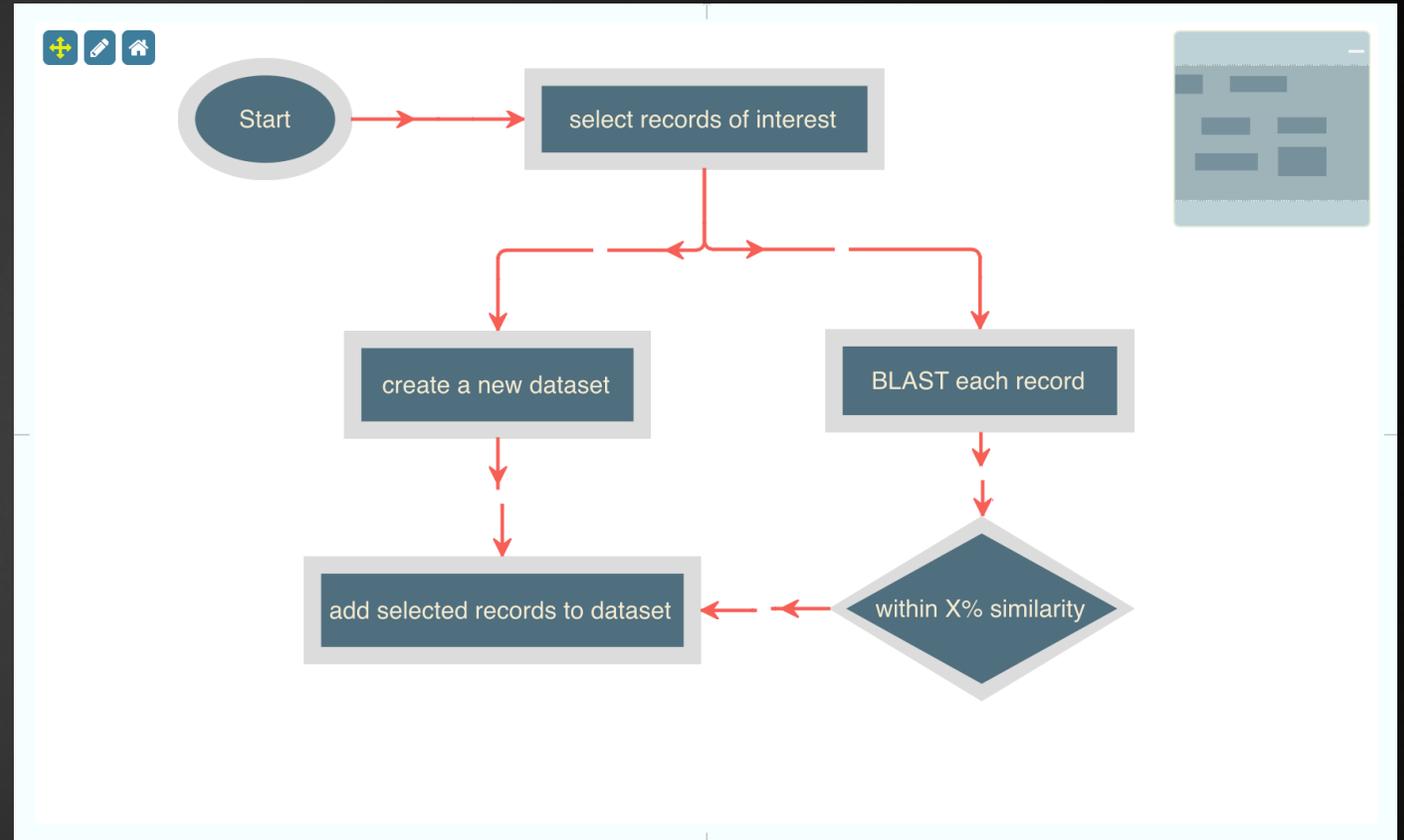
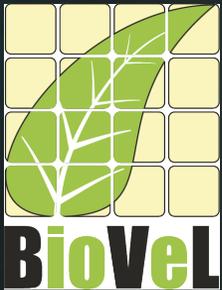
Plug-in Framework



6000+ Analytical Packages in CRAN

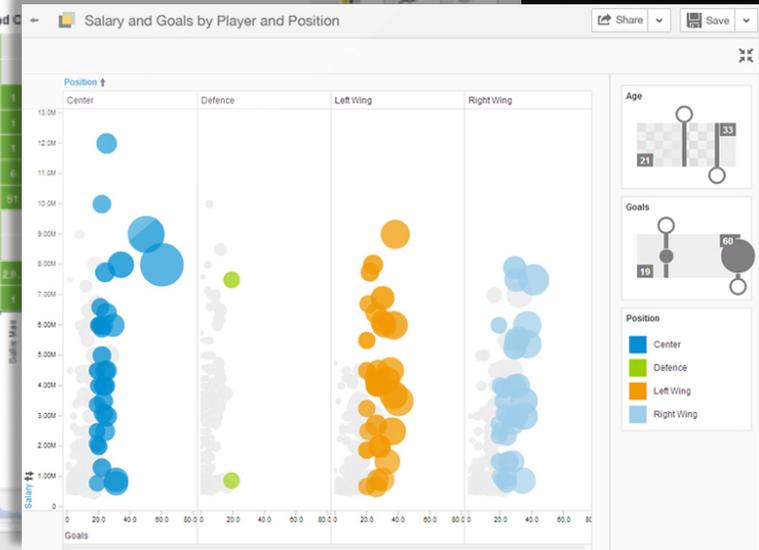
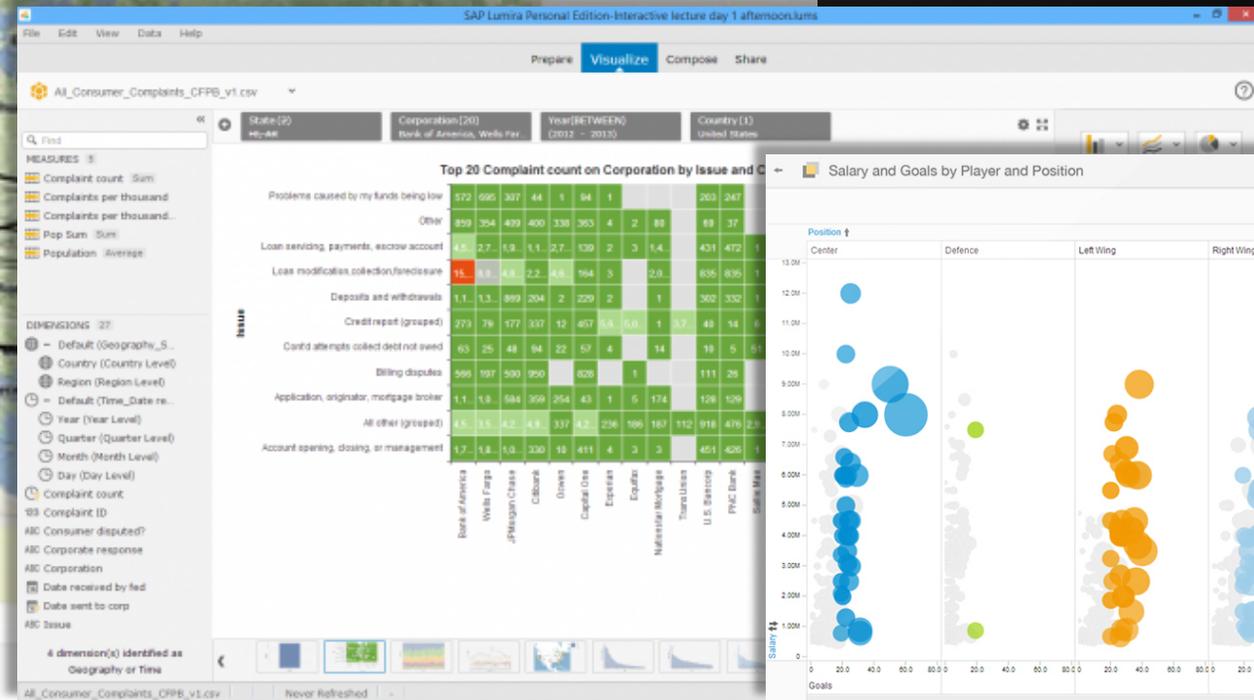
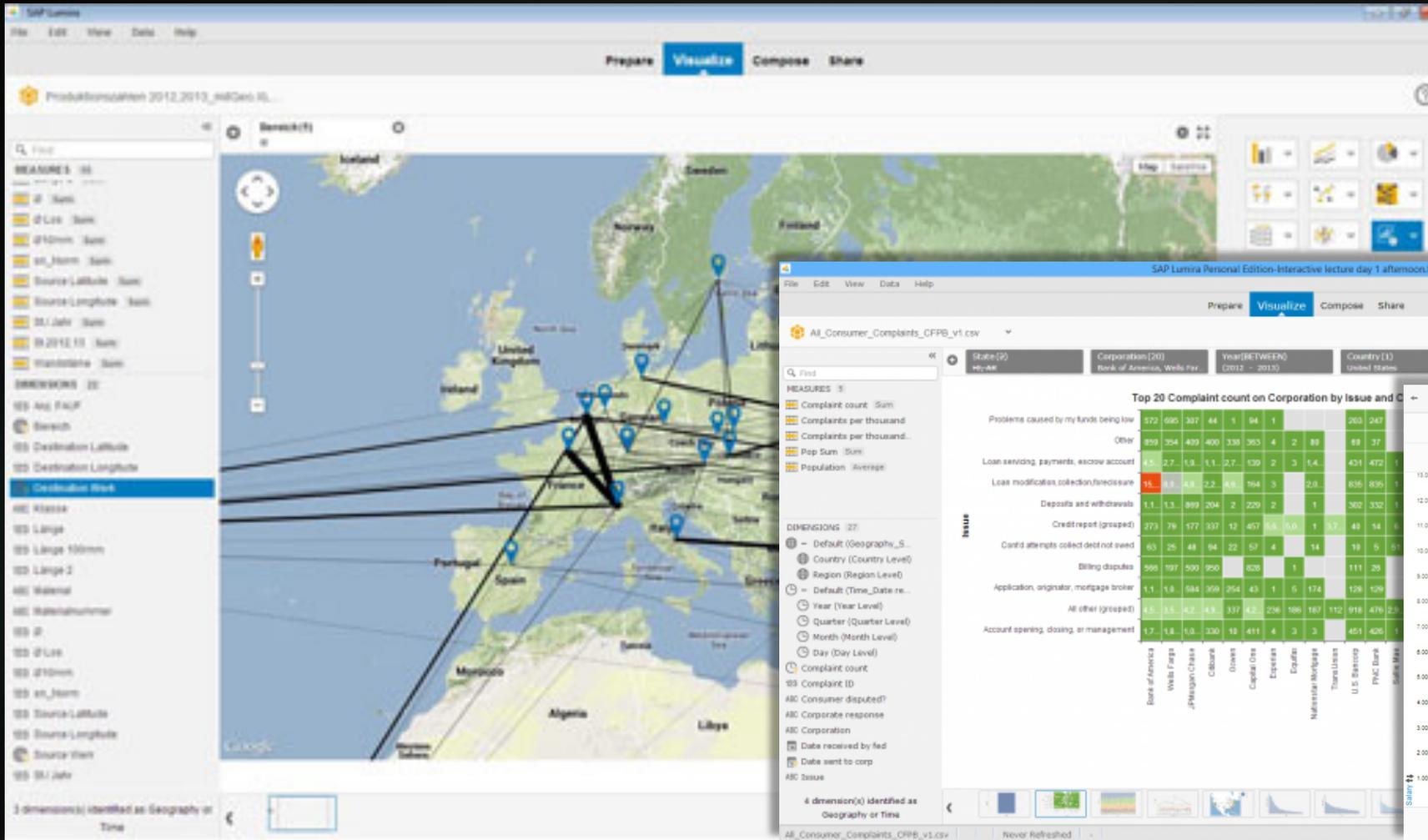
Analytical Pipelines

BOLD



Simple Pipeline

BOLD 4 + SAP Lumira



BOLD4 – Some other features

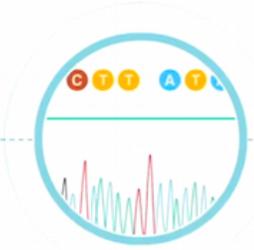


- Checklist Support (synonyms, progress, shopping lists)
- Data portal for core facilities
- Complete record histories
- RESL algorithm on your own datasets
- Storage and analysis of pre-clustered NGS data

Support for Metabarcoding

Support for the application of Next Generation Sequencing in DNA Barcoding Initiatives

from the Instrument to Interpretation



Storage of NGS Data

Support for storage of instrument data and experiment metadata in the cloud along with a publication pipeline that handles submission to GenBank SRA, & Dryad. Metadata templates are available for biodiversity surveys, ecological studies, and longitudinal studies. Metadata data standards are enforced in compliance with the Genomics Standards Consortium.

Data Preprocessing

Submitted NGS data files are preprocessed in the cloud based on preset project parameters. Stream processing is employed with software that can be run on the desktop employing compression and bittorrent to transfer data to BRAVE with minimal bandwidth impact. Computational tools employ FPGAs to achieve quick turnaround on analytical requests.



mBRAVE

Metabarcoding Research And Visualization Environment



Linkages & Partners

Acknowledgements



Paul Hebert
Dan Janzen & Winnie Hallwachs
Scott Miller
Axel Hausmann
Stefan Kremer
Many others!

Biodiversity Institute Collections Team
Canadian Center for DNA Barcoding Team

BOLD Users

INNOVATION.CA
CANADA FOUNDATION FOR INNOVATION | FONDATION CANADIENNE POUR L'INNOVATION



BIO Biodiversity Institute of Ontario

Ontario **Genomics** Institute



GenomeCanada

Ontario
MINISTRY OF
RESEARCH AND INNOVATION