

international
BARCODE
OF LIFE



Building a Reference Library: Sourcing Specimens Natural History Collections

Jeremy deWaard, Sean Prosser, Jayme Sones, Meredith Miller, Masha Kuzmina,
Valerie Levesque-Beaudin and Evgeny Zakharov

Scientific Steering Committee, October 13 & 14, 2018

Major sources of specimens for iBOL-II

1. Fresh collections
2. Living collections and DNA/tissue repositories
3. Natural history museums and herbaria

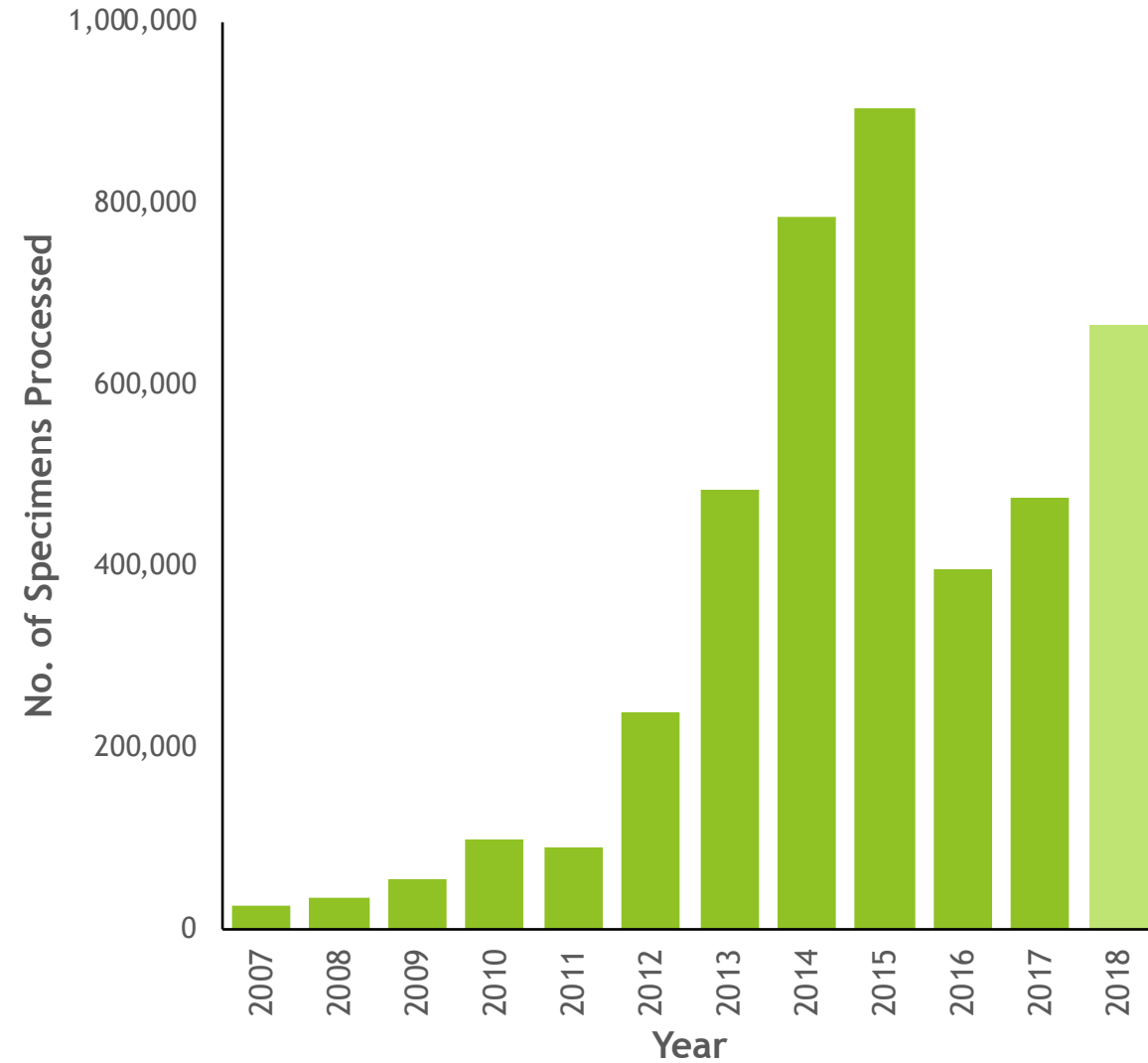
Major sources of specimens for iBOL-II

1. Fresh collections
2. Living collections and DNA/tissue repositories
3. Natural history museums and herbaria



4,023,500

Fresh specimens
sourced, processed
and archived at CBG

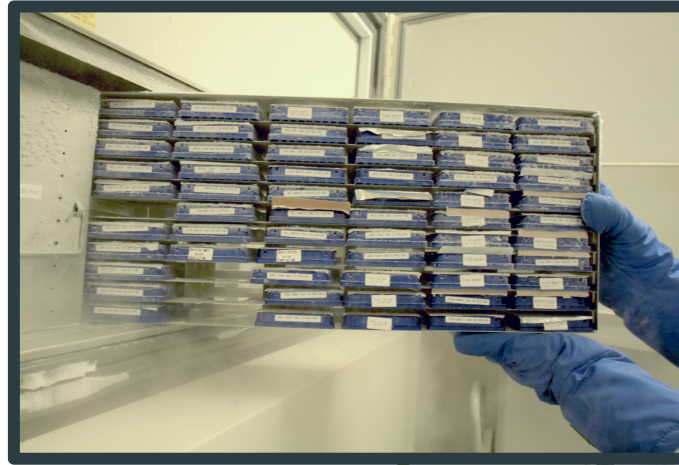


Major sources of specimens for iBOL-II

1. Fresh collections
2. Living collections and DNA/tissue repositories
3. Natural history museums and herbaria



5,760,000
DNA samples
~109K species
without BINs (366K
DNA samples)



Major sources of specimens for iBOL-II

1. Fresh collections
2. Living collections and DNA/tissue repositories
3. Natural history museums and herbaria

Natural history collections for reference library construction

PROS

- ▶ Huge resource of ~3 billion specimens world-wide
- ▶ Name-bearing types and other expertly identified/verified material
- ▶ Rare and extinct taxa often available, including from restrictive countries

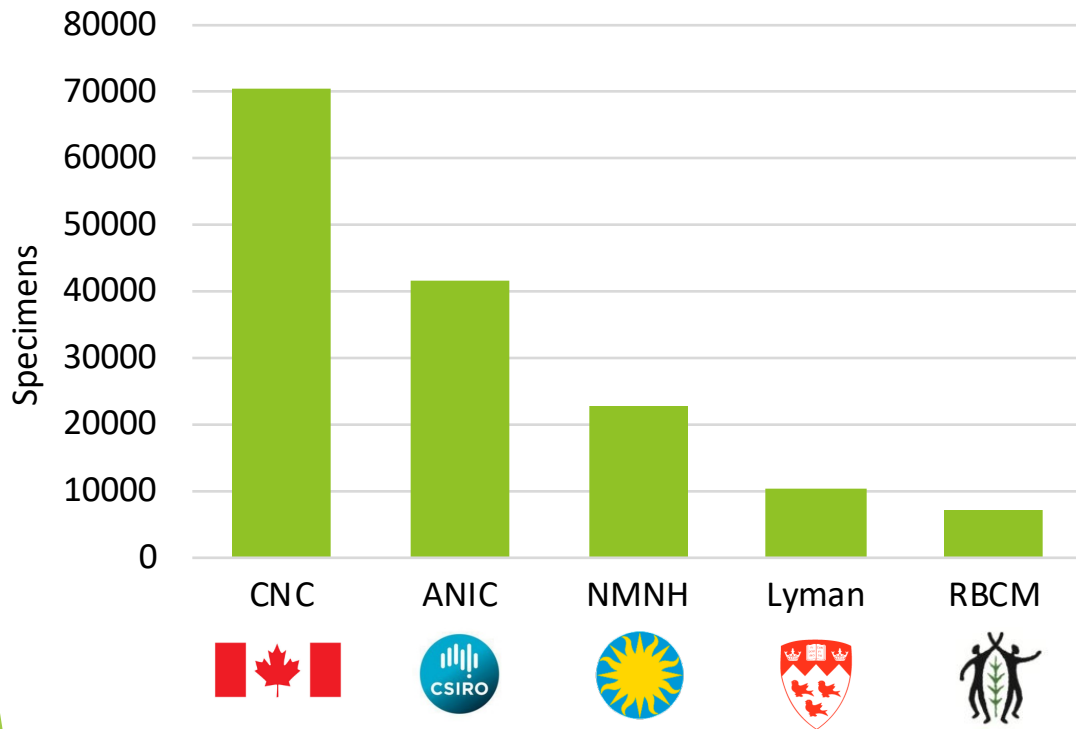
CONS

- ▶ Technical difficulties in recovering DNA due to DNA degradation (= slower, costlier)
- ▶ Restrictions on access, sampling and data release
- ▶ Biases in taxa/regions based on curators



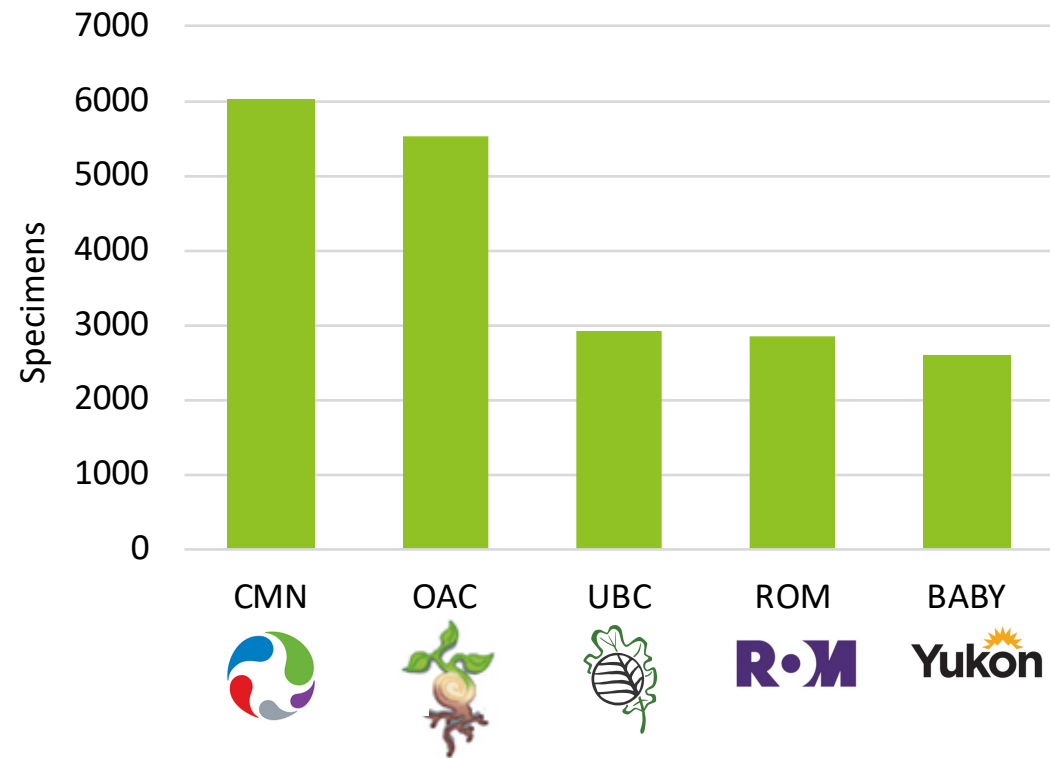
CBG natural history collection harvesting

Museums



~165K specimens from 14 institutions

Herbaria



~15K specimens from 27 institutions

Factors affecting recovery of Sanger sequenced DNA barcodes

OPEN ACCESS Freely available online

PLOS ONE

A DNA 'Barcode Blitz': Rapid Digitization and Sequencing of a Natural History Collection

Hebert et al. 2013 doi: [10.1371/journal.pone.0068535](https://doi.org/10.1371/journal.pone.0068535)

 Applications
in Plant Sciences

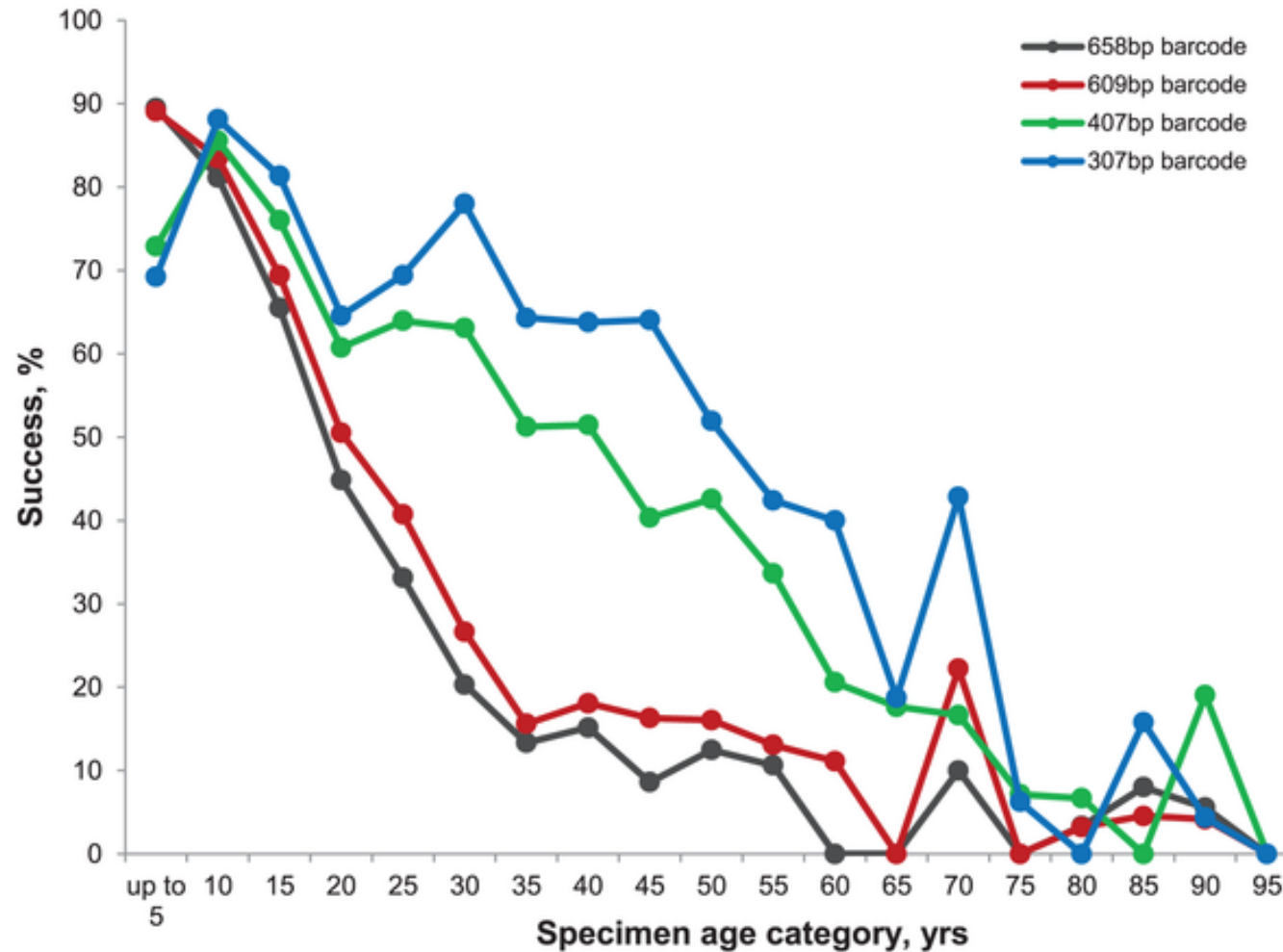
Applications in Plant Sciences 2017 5(12): 1700079

APPLICATION ARTICLE

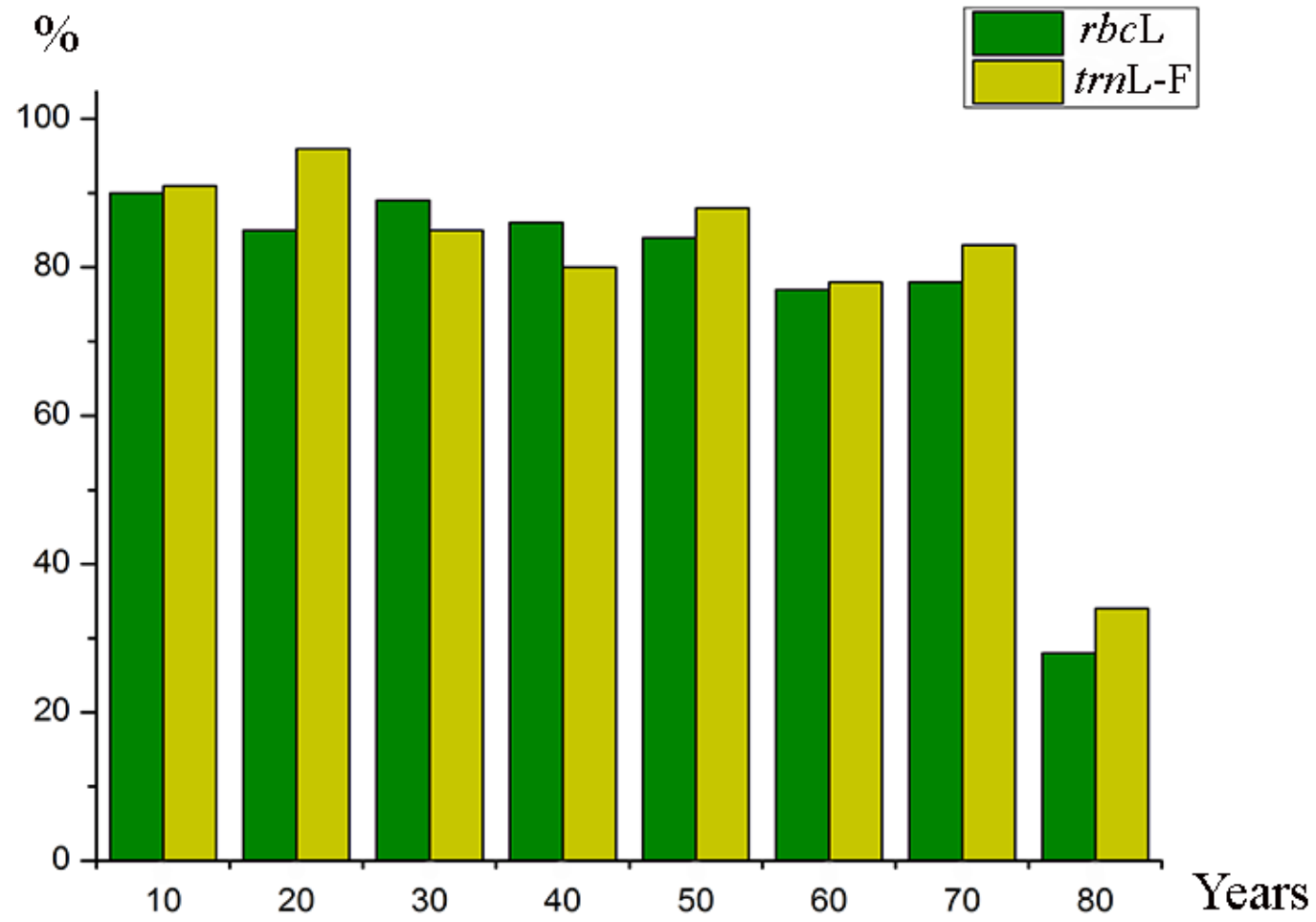
USING HERBARIUM-DERIVED DNAs TO ASSEMBLE A LARGE-SCALE DNA BARCODE LIBRARY FOR THE VASCULAR PLANTS OF CANADA¹

Kuzmina et al. 2017 doi: [10.3732/apps.1700079](https://doi.org/10.3732/apps.1700079)

Factor: Age

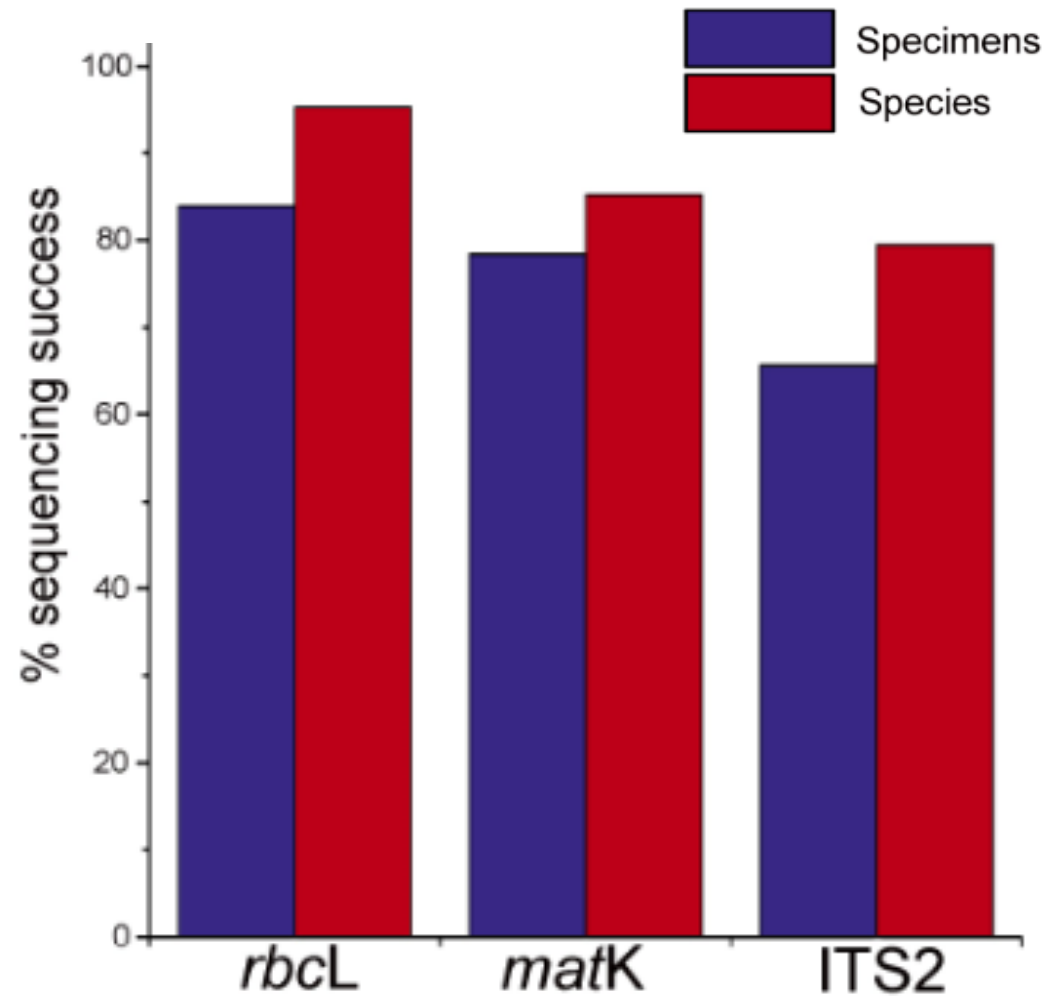


Factor: Age

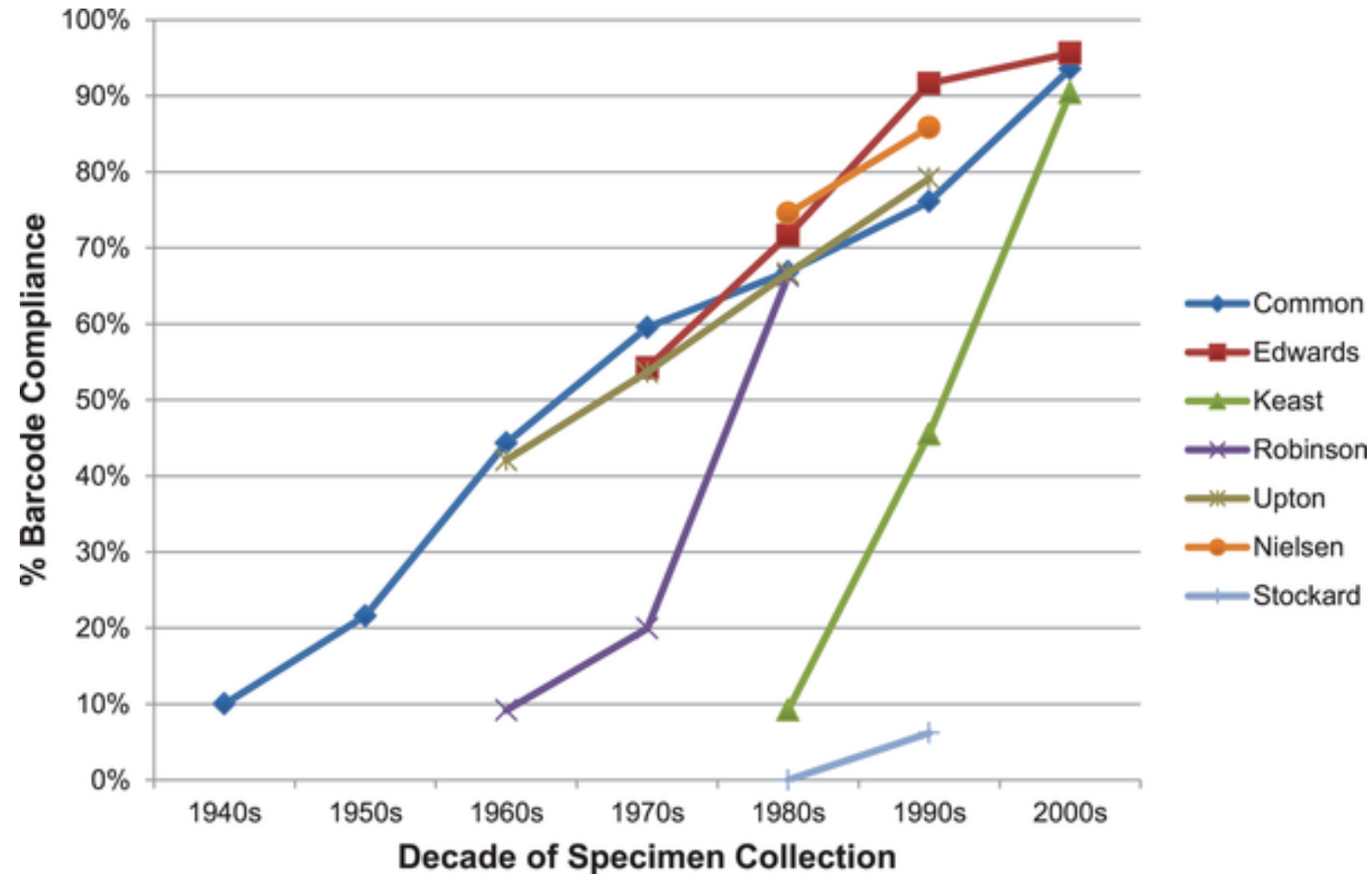


Canadian Museum Nature mosses 2016-2017 (N = 1981)

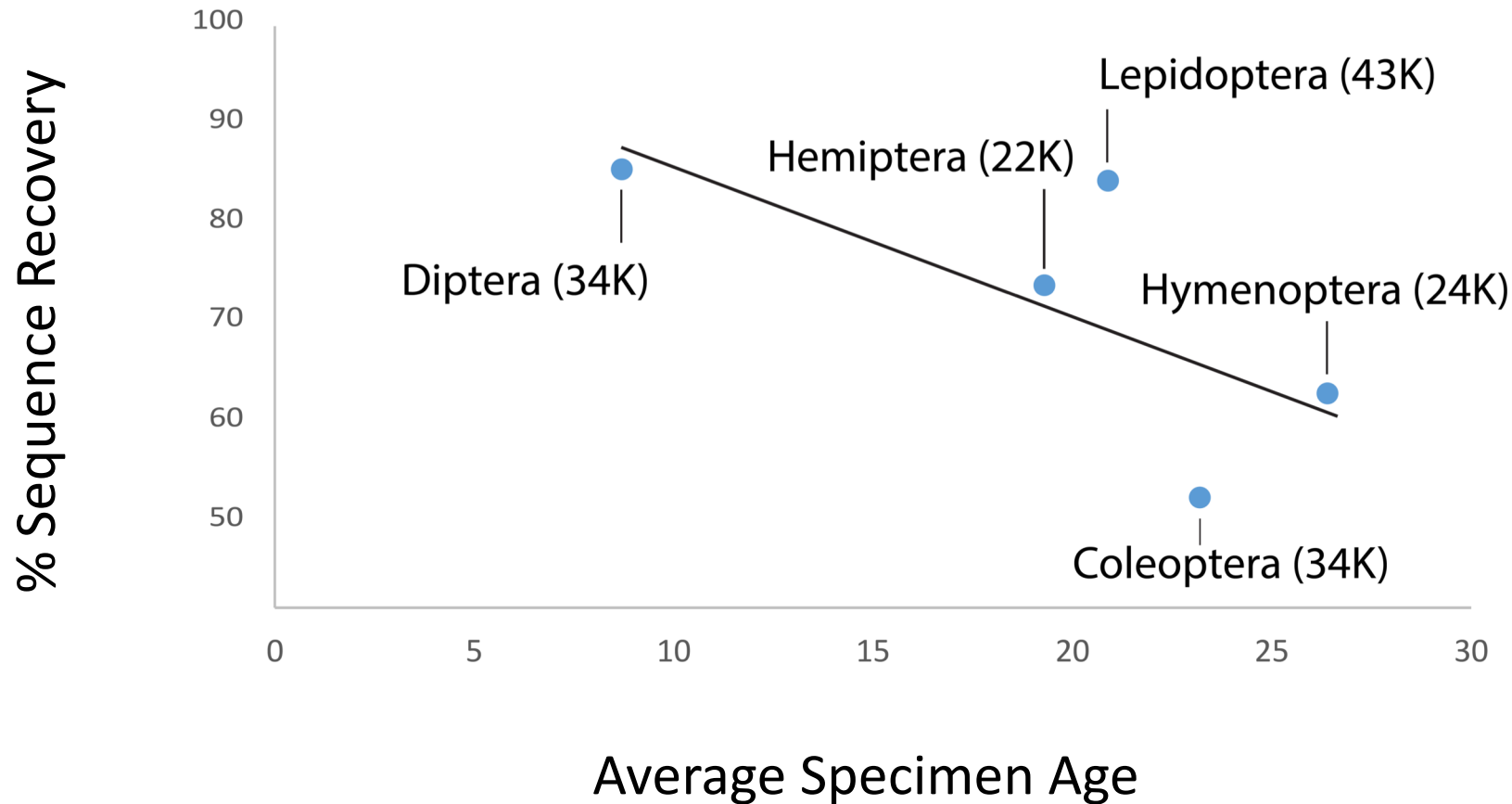
Factor: Marker



Factor: Collector

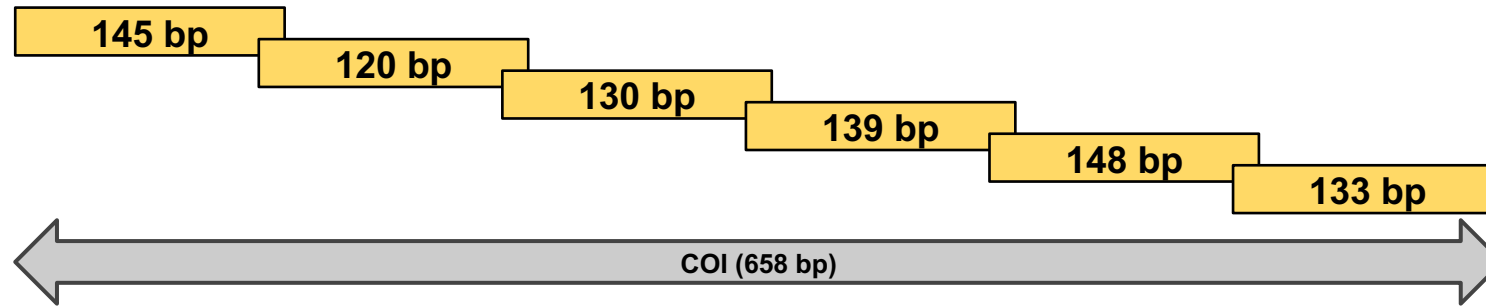


Factor: Interplay of multiple variables



Canadian National Collection insects 2008-2014 (N = 157K)

Next generation sequencing solutions



MOLECULAR ECOLOGY RESOURCES

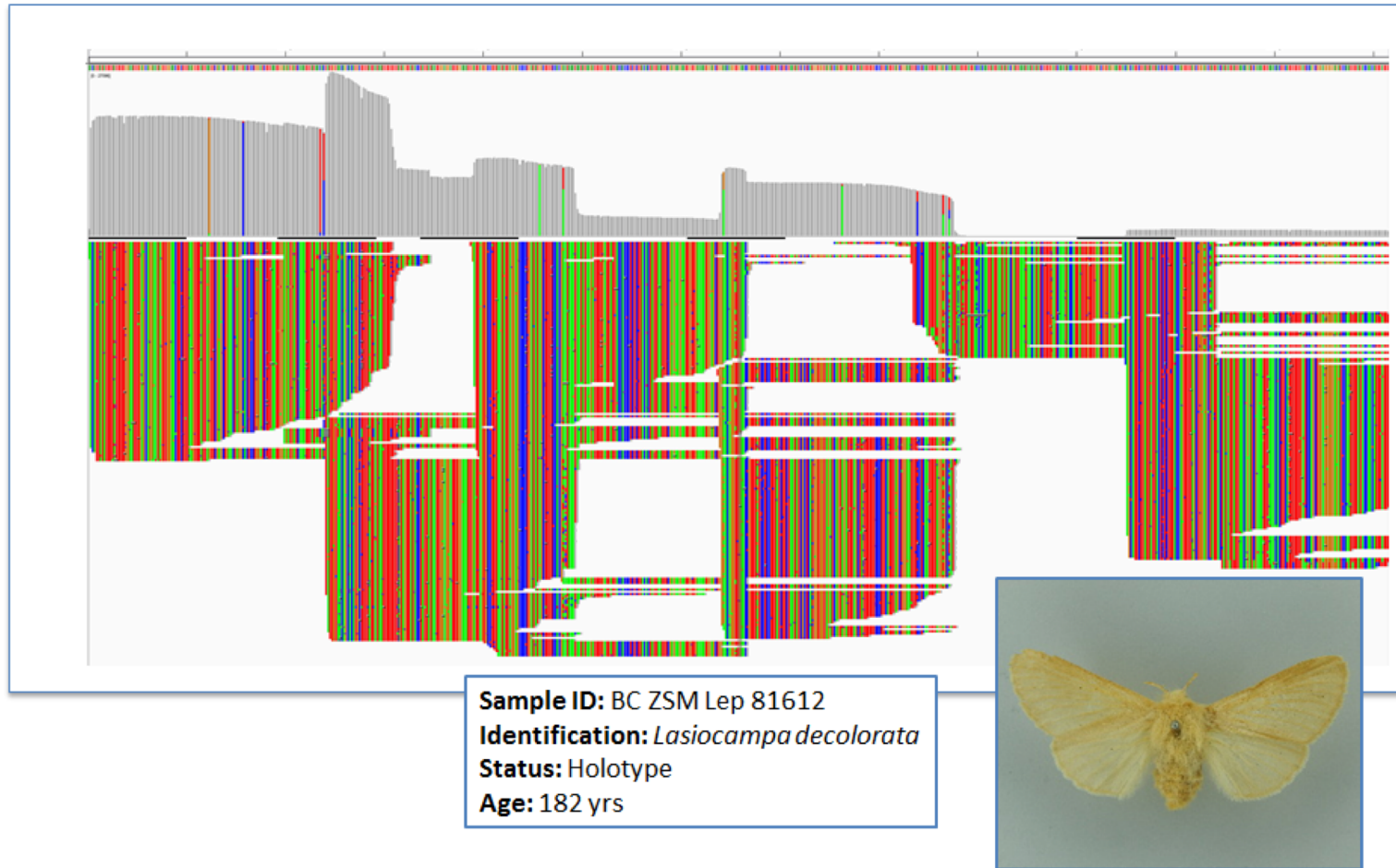
Molecular Ecology Resources (2016) 16, 487–497

doi: 10.1111/1755-0998.12474

**DNA barcodes from century-old type specimens using
next-generation sequencing**

Prosser et al. 2016 doi: [10.1111/1755-0998.12474](https://doi.org/10.1111/1755-0998.12474)

Degraded DNA protocol for animals



- Reads assembled into full-length barcode

e.g. 3000 types in Hausmann et al. 2016 doi: [10.1139/gen-2015-0197](https://doi.org/10.1139/gen-2015-0197)

Degraded DNA NGS protocol for animals

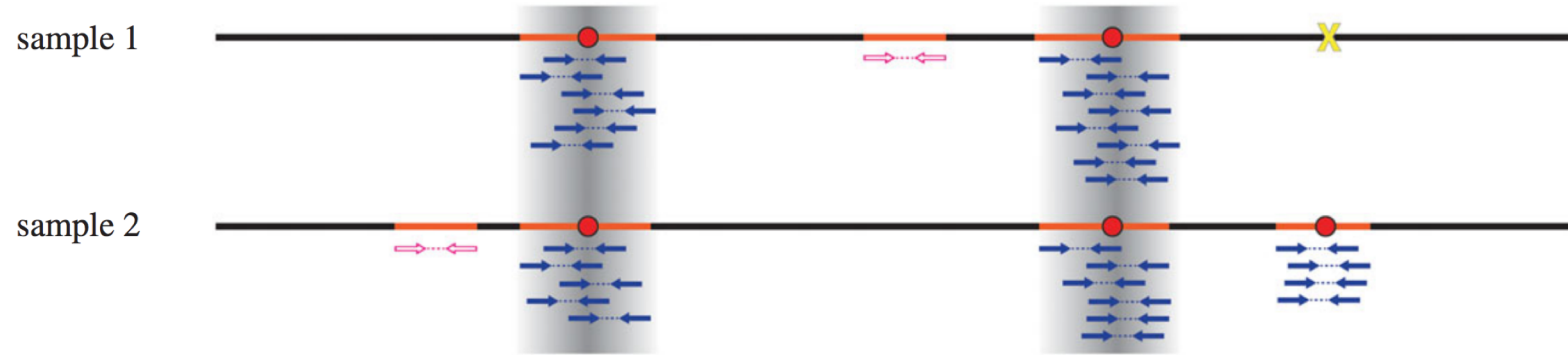
- ▶ Multiplex 96 samples on one run
- ▶ SMRT sequencing on PacBio Sequel
- ▶ Validation important - higher likelihood of chimeras & pseudogenes
- ▶ 2-4X the cost of standard Sanger-based museum protocols
- ▶ 70.4% mean sequence recovery (on hardest samples); 5-20X increase over Sanger; weak association with age ($R^2=0.04$)



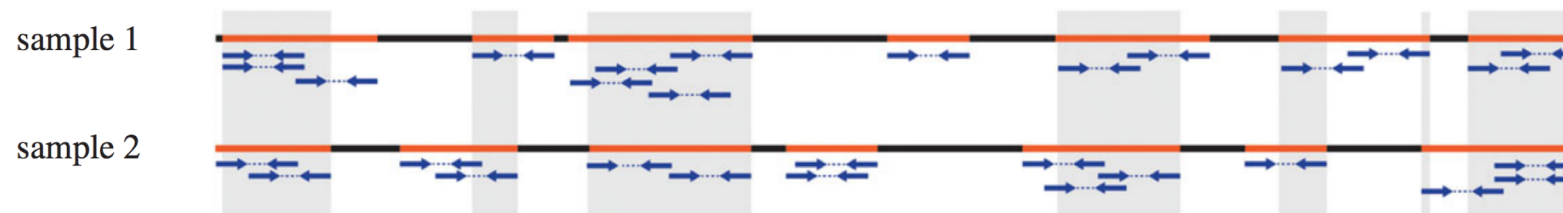
- ▶ Service now offered at  **CCDB**
Canadian Centre for
DNA Barcoding

‘Upgraded’ plant barcoding

(a) target enrichment



(b) genome skimming

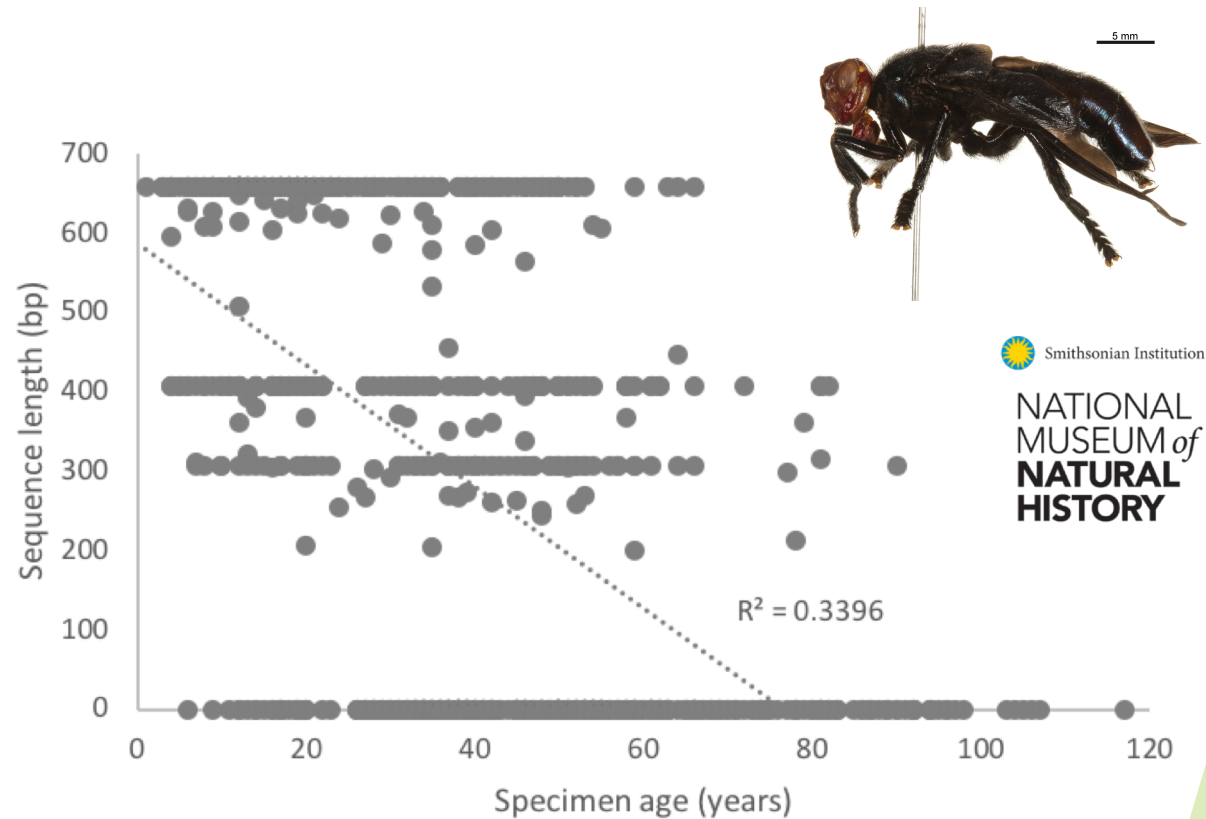


Hollingsworth et al. 2016 doi: [10.1098/rstb.2015.0338](https://doi.org/10.1098/rstb.2015.0338)

- Some groups moving to complete chloroplast genomes, e.g. Zhang et al. 2018 doi: [10.1038/s41598-017-00321-6](https://doi.org/10.1038/s41598-017-00321-6)

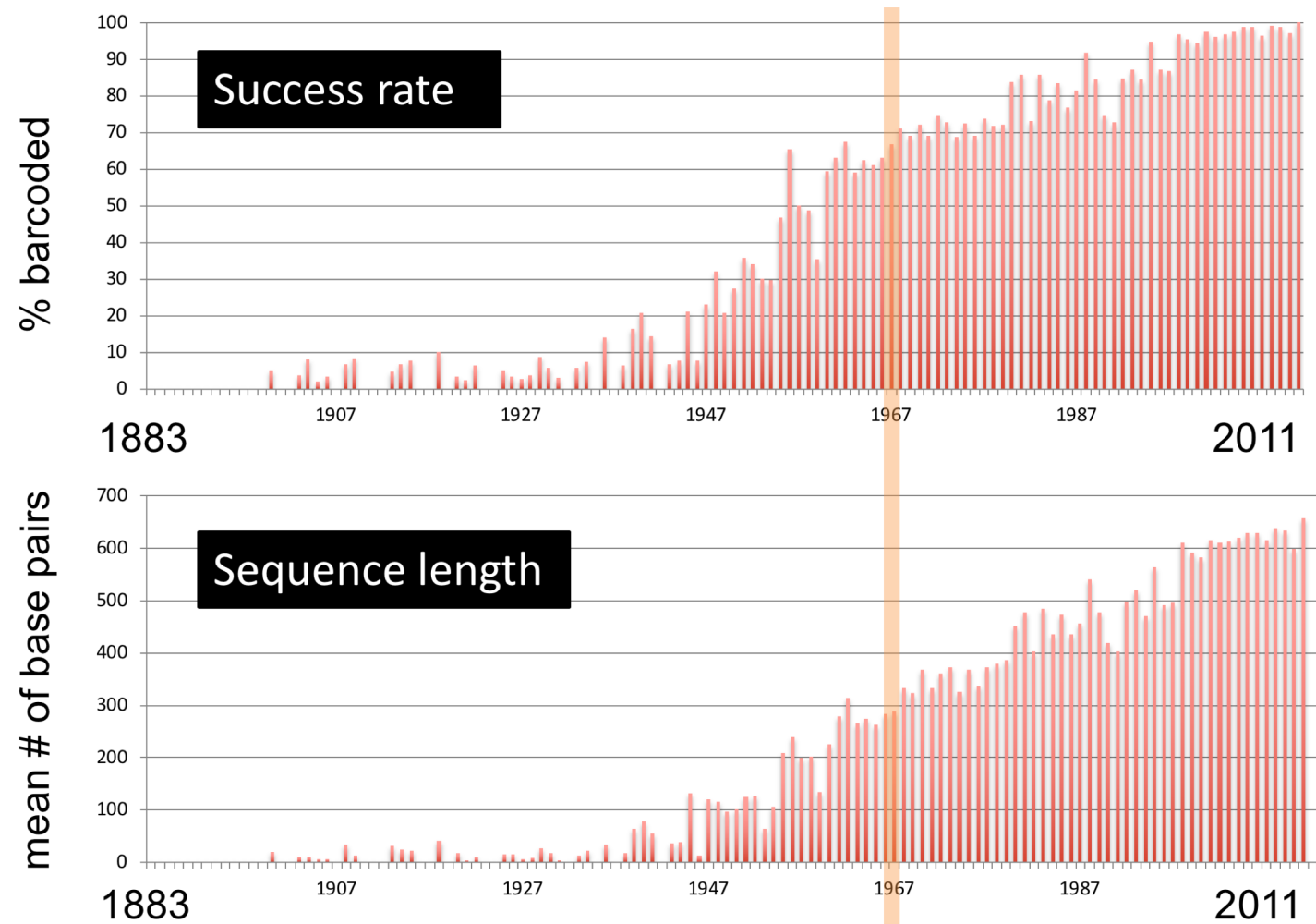
Mixed Sanger/NGS strategy for museum harvesting

- ▶ Pilot: SI-NMNH Diptera
 - ▶ Selected/analyzed 950 specimens, mostly 1950+
 - ▶ Mean age = 38.3 years
 - ▶ Sanger-based approach first (307 bp and 407bp fragments)
 - ▶ 53.8% barcode recovery
 - ▶ Degraded DNA NGS protocol on failures
 - ▶ 88.4% barcode recovery



General harvesting strategies

- ▶ Generate pick lists before visit
- ▶ Set thresholds / expectations based on factors and \$
- ▶ Small sequences fine => often match freshly collected material (35.4% for SI-NMNH material)



Resources



Browser address bar: iBOL 2017 | Centre for Biodiversity Genomics | biodiversitygenomics.net/resources/ibol-2017/

Centre for Biodiversity Genomics

Home About Facilities Units Services Projects Resources Funders & Partners

iBOL 2017

Pre-Conference Workshop Resources

Workshop presentation

1. Welcome and introduction - Natural History Collection Harvesting Strategies & Logistics (PDF)
2. Sample Harvesting in Natural History Collections (PDF)
3. Specimen Digitization (PDF)
4. Pre-lab Processing (PDF)
5. Standard Protocols for Natural History Collection Barcoding (PDF)
6. Data Analysis and Archiving (PDF)
7. New Developments in the lab (PDF)

Files provided

CCDB and BOLD Templates

- **BOLD ImageData Submission (XLS)**: Spreadsheet used to upload image data to BOLD.
- **BOLD SpecimenData v3 Submission (XLS)**: Spreadsheet used to upload specimen data to BOLD.
- **CCDB-00000 Record (XLS)**: Used to submit plates to CCDB or can be used to record sample ID for arrays (including map).

Museum Templates

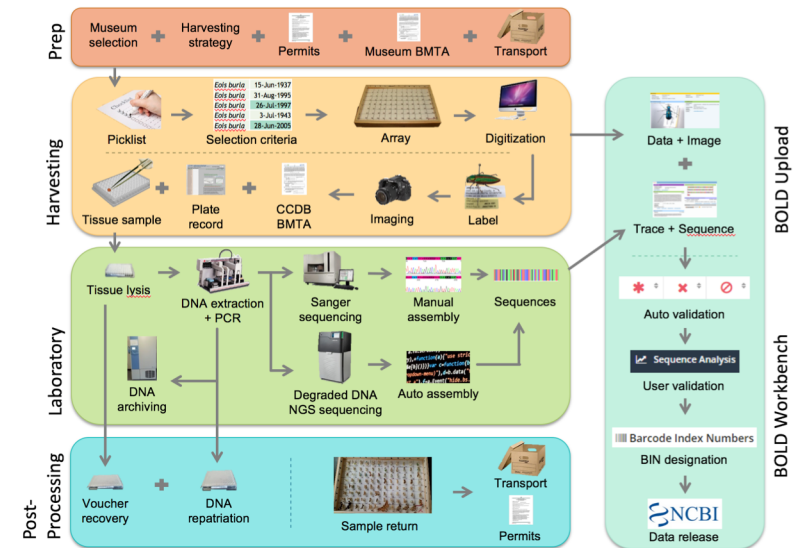
- **Specimen Removal Labels (XLS)**: Create label place holders for specimen removal from a museum collection. *Need to change the institution on loan to and the plate numbers accordingly.
- **DNA Barcode Labels (XLS)**: Labels to add to specimens that are tissue sampled on site to mark that they have been sampled for sequencing. *Need to update the sample ID accordingly.
- **Museum Data Entry (XLS)**: Spreadsheet to enter data while at the museum. The information will also be used by the museum for the loan invoice.

Documentation

- **iBOL-BIO BMTA and Data Policy Agreement (DOC)**: iBOL Biological Material Transfer and Data Policy Agreement.
- **CBG Loan Conditions (PDF)**: Our loan conditions if you are interested in borrowing specimens.

Shipping

- **Protocol for Shipping Ethanol (PDF)**: International Air Transport Association (IATA) requirements for shipping ethanol.
- **NZNNH Shipping Dried Herbarium Specimens (PDF)**: New Zealand National Herbarium Network - Standard for shipping of dried herbarium specimens.



Resources

Strategic Priorities for DNA Barcoding Natural History Collections



Authors: Peter M. Hollingsworth^{1*}; David Harris¹; Michelle Hart¹; Gabi Dröge²; Isabel Ray³; Inés Álvarez Fernández³; Javier Fuertes Aguilar³; Beatriz Alvarez Dorda³; Tim Fulcher⁴; Stefanie Krause⁵; Thomas von Rintelen⁵; Michel Guiraud⁶; Carole Paleco⁷; Thierry Backeljau⁷; Patrick Grootaert⁷; Barbara Gravendeel⁸; René Dekker⁸; Elisabeth Haring⁹; Karin Wiltshcke-Schrotta⁹; Nikolaus Szucsich⁹; Luise Kruckenhauser⁹; Vacek František¹⁰; Jiří Kvaček¹⁰; Jackie Mackenzie-Dodds¹¹; Stephen Russell¹¹; Garin Cael¹²; Danny Meirte¹²; Sean Prosser¹³; Paul Hebert¹³

Hollingsworth et al. 2016: <http://synthesys3.myspecies.info/node/616>

Acknowledgements

Co-authors

CBG and iBOL colleagues

S. Miller, A. Hausmann & R. Rougerie

GGI / SIBN / SBF funding



PDF (with links and papers) posted at: <http://biodiversitygenomics.net/resources/ibol-2017/>



**CANADA
FIRST**
RESEARCH
EXCELLENCE
FUND

 **Ontario**
MINISTRY OF
RESEARCH AND INNOVATION

**UNIVERSITY
of GUELPH**

Resources

The screenshot shows the GGI Gap Analysis Tool web interface. The browser address bar displays the URL <https://ggidata.shinyapps.io/gapanalysis/>. The page has a navigation bar with tabs: "Input", "Results Summary", "Results Table", and "Download Results". The "Input" tab is active.

Input Names
Names can be uploaded in a file or pasted directly into the text box.
Only names at the kingdom, phylum/division, class, order, family, or genus rank can be analyzed. **This tool does not work on species names.**

Upload name file
Upload a .txt file containing one name per line.
Names file (.txt)
Browse... No file selected

Paste list of names
Paste a list of names directly into this text box. Please enter one name per row.
List of names
One name per row

Run Analysis
Run Analysis
After clicking Run Analysis, manually click on Results Summary or Results Table to view results.

Name Options
These options can be specified for the list of input names.

Taxonomic rank of submitted names
Select the taxonomic rank of the names submitted. E.g. If submitting a list of all families, select **Family** as the **Taxonomic rank of submitted names**
Taxonomic rank of submitted names
Not Specified

Dataset label
This label is used to name the download files.
Dataset label
User Data

Output Filtering
These options filter the output of the gap analysis.

Filter by names status
Filter results based on the GBIF name status. E.g. To receive only the results matching accepted GBIF names, select **accepted** from **Name status**
Name status
Not Specified

Filter by taxon
These options allow for filtering by a specified taxonomic groups. e.g. If submitting a list of mammal families, select **Class** from **Select filter rank**, then select **Mammalia** from **Select filter taxonomic name**
Select filter rank
Not Specified
Select filter taxonomic name
None Selected

GGI Gap Analysis Tool: <https://ggidata.shinyapps.io/gapanalysis/>

Resources

B

CL-NALEP (Lepidoptera of North America, North o... Checklist

Code: CL-NALEP

Title: Lepidoptera of North America, North of Mexico

Citation: Pohl, GR, Patterson, B, and Pelham JP, 2016 [unpublished]. Annotated taxonomic checklist of the Lepidoptera of North America, North of Mexico. EXCEL version 1.1. (15 June 2016).

URL: https://www.researchgate.net/publication/302570819_Annotated_taxonomic_checklist_of_the_Lepidoptera_of_North_America_North_of_Mexico

Description: This file is an updated EXCEL version of the P3 working checklist by Pohl, Patterson, and Pelham (version 1.0) made available in PDF form online (Researchgate). It is a non-reviewed, provisional checklist, made available by the authors as a working paper. Please send corrections and additions to GR Pohl at micromothman@gmail.com.

Tax Breakdown:

Phyla: 1
Classes: 1
Orders: 1
Superfamily: 34
Families: 92
Subfamilies: 237
Tribe: 247
Subtribe: 70
Genera: 2504
Subgenus: 79
Species: 12763
Subspecies: 2911

B

Checklist Progress Report for CL-NALEP

Download Progress

Progress Report Results

Progress Report provides results of the comparison between checklist **CL-NALEP** and all data on BOLD.

Summary of Results

Taxon Level	Total	%Sampled	%Sequenced	%Barcoded
Subspecies	2911	13.57%	12.37%	10.89%
Species	12763	89.20%	83.56%	79.84%
Subgenus	79	0.00%	0.00%	0.00%
Genus	2504	95.53%	93.77%	92.29%
Subtribe	70	7.14%	7.14%	7.14%
Tribe	247	30.77%	30.77%	30.77%
Subfamily	237	78.06%	78.06%	78.06%
Family	92	95.65%	95.65%	95.65%
Superfamily	34	2.94%	2.94%	2.94%
Order	1	100.00%	100.00%	100.00%

Search:

family	Subfamily Reference	Tribe	Tribe Reference	Subtribe	Subtribe Reference	Genus	Genus Reference	Subgenus	Subgenus Reference	Species	Species Reference
tuinae	Latreille 1809	Xylenini	Guenée 1837	Cosmiina	Guenée 1852	Zotheca	Grote 1874			Zotheca tranquilla	Grote 1874
tuinae	Latreille 1809	Eriopygini	Fibiger & Lafontaine 2005			Zosteropoda	Grote 1874			Zosteropoda hirtipes	Grote 1874
tuinae	Latreille 1809	Eriopygini	Fibiger & Lafontaine 2005			Zosteropoda	Grote 1874			Zosteropoda clementei	Meadows 1942
itinae	Zeller 1839	Phycitini	Zeller 1839			Zophodia	Hübner 1825			Zophodia multistriatella	(Blanchard & Knudson 1982)

BOLD Checklists: http://boldsystems.org/index.php/Checklists_Management/

Resources

BOLD

SYSTEMS

DATABASES

IDENTIFICATION

TAXONOMY

WORKBENCH

RESOURCES

TAXONOMY BROWSER: Asilidae

Family : Asilidae

Arthropoda / Insecta / Diptera / Asilidae

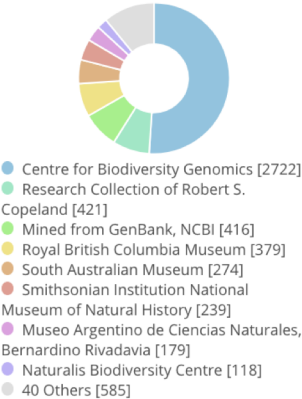


CC BY-NC-SA CBG Photography Group
2009
Image of Asilidae

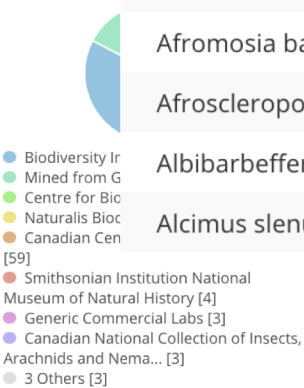
Statistics

Specimen Records:	5,606
Specimens with Sequences:	4,676
Specimens with Barcodes:	4,316
Species:	839
Species With Barcodes:	504
Public Records:	2,649
Public Species:	492
Public BINs:	712
SPECIES LIST PUBLIC DATA	

Specimen Depositories



Sequencing



Species	Specimens	Sequences	Barcodes >500bp
Ablautus coleii	5	3	0
Ablautus rufotibialis	2	2	1
Ablautus sp.	2	2	2
Acasilus tigrimontis	1	1	1
Acnephalomyia andrenoides	1	1	1
Acnephalomyia dorsalis	4	4	4
Acronyches sp.	1	1	1
Afroestricus chiasmoneurus	3	3	3
Afroestricus variabilis	1	0	0
Afroholopogon mauros	1	1	1
Afroholopogon sp.	3	1	1
Afromochtherus mendax	1	1	1
Afromosia barkemeyeri	1	0	0
Afroscleropogon dilutus	1	1	1
Albibarbefferia albibarbis	1	1	1
Alcimus slenurus	3	0	0

Factor: Body size

