

DNA Barcoding and Molecular Specimen Identification



Monica R. Young

myoung02@uoguelph.ca

Outline of talk

- Brief overview of DNA barcoding
- Results of barcoding mite communities associated with poultry houses
- Live demonstration of the Barcode of Life Datasystems (BOLD)

Specimen identification

- traditional morphological identifications
 - keys require expertise
 - keys are often life-stage or gender specific
 - extreme phenotypic variation



Above: Eupodes spp.; three species are shown

DNA barcoding (DNAB)

- Hebert et al 2003: UPC analogy
 - unique product tags = short, standard fragments of DNA
 - products = species
 - fast & automated
 - automated biodiversity assessment



(Stoeckle & Hebert 2008)

Centre for Biodiversity Genomics

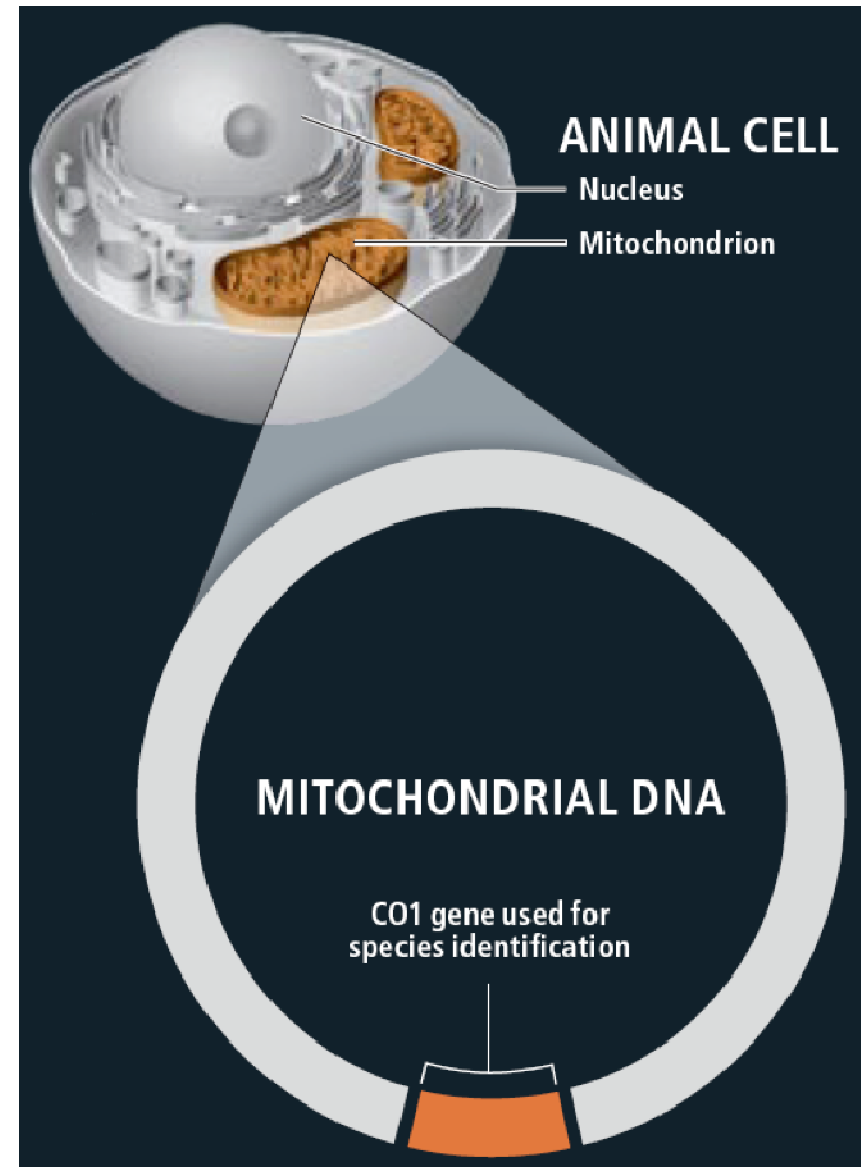


Biodiversity Institute
of Ontario

<http://biodiversitygenomics.net/>

DNAB: the basics

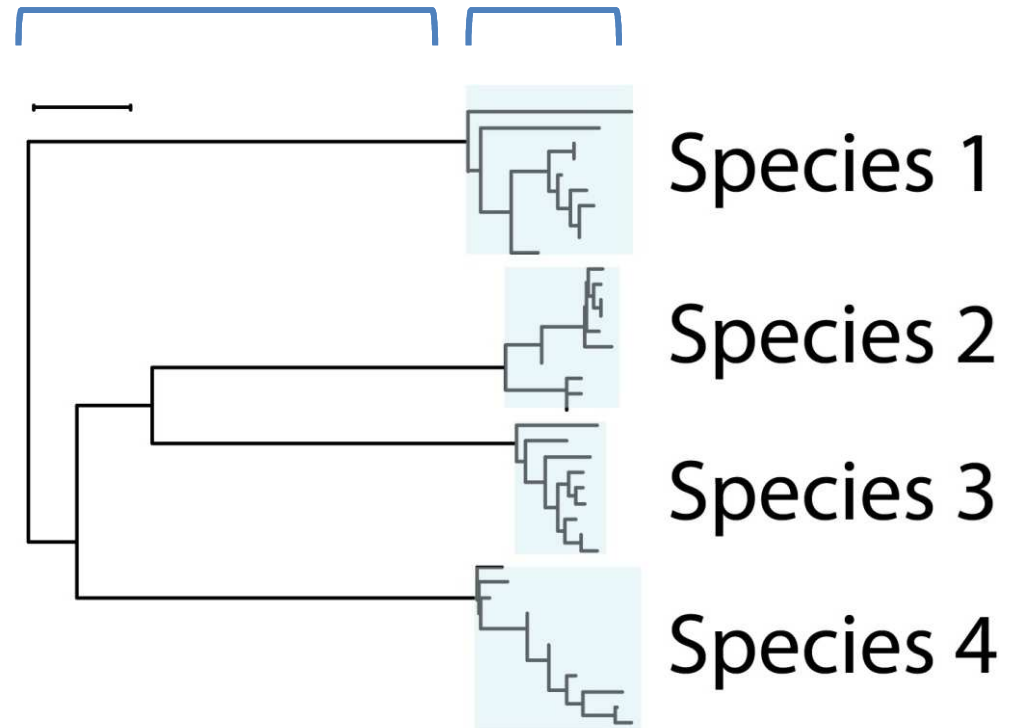
- new:
 1. standardization
 2. specimen vouchering
 3. curation of database
- animal life:
cytochrome c oxidase subunit I (COI) gene



(Stoeckle & Hebert 2008)

DNAB: the basics

- relies on 'barcode gap' of sequence divergence
 - high interspecific divergence
 - low intraspecific divergence

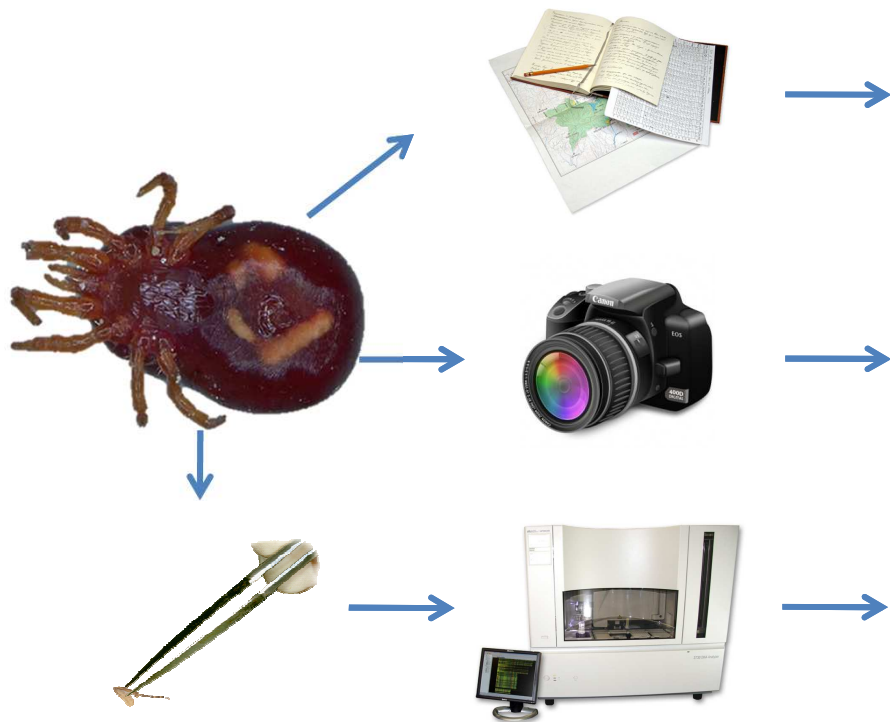


DNAB: the procedure

- two phases:
 1. library construction
 2. identification of unknowns

DNAB: the procedure

- two phases:
 - library construction



BOLD SYSTEMS

Specimen - Soil mites in association with the red poultry mite [SMRPM] Print

[Edit Specimen](#)

IDENTIFIERS	
Sample ID:	HUJINVACA2
Process ID:	SMRPM002-16
Institution Storing:	Hebrew University of Jerusalem, National Collections of Natural History
Field ID:	
Museum ID:	HUJ
Collection Code:	INVACA

TAXONOMY	
Identification:	Dermanyssus gallinae (De Geer, 1778, 1778)
Rank:	Species
Identifier:	Maria L. Moraza
Identification Method:	Morphological
Identifier Institution:	Universidad de Navarra
Identifier Email:	mlmoraza@unaves
Taxonomy Note:	
Rank:	Current Record (HUJINVACA2)
Phylum:	Arthropoda
Class:	Arachnida
Order:	Mesostigmata
Family:	Dermanyssidae
Subfamily:	
Genus:	Dermanyssus
Species:	Dermanyssus gallinae
Subspecies:	

SPECIMEN DETAILS	
Voucher Status:	
Tissue Descriptor:	
Sex:	Female
Reproduction:	
Life Stage:	Adult

PHOTOGRAPHS

License: No Rights Reserved (2016)
License Holder: Eric Palevsky, Agricultural Research Organization, Israel

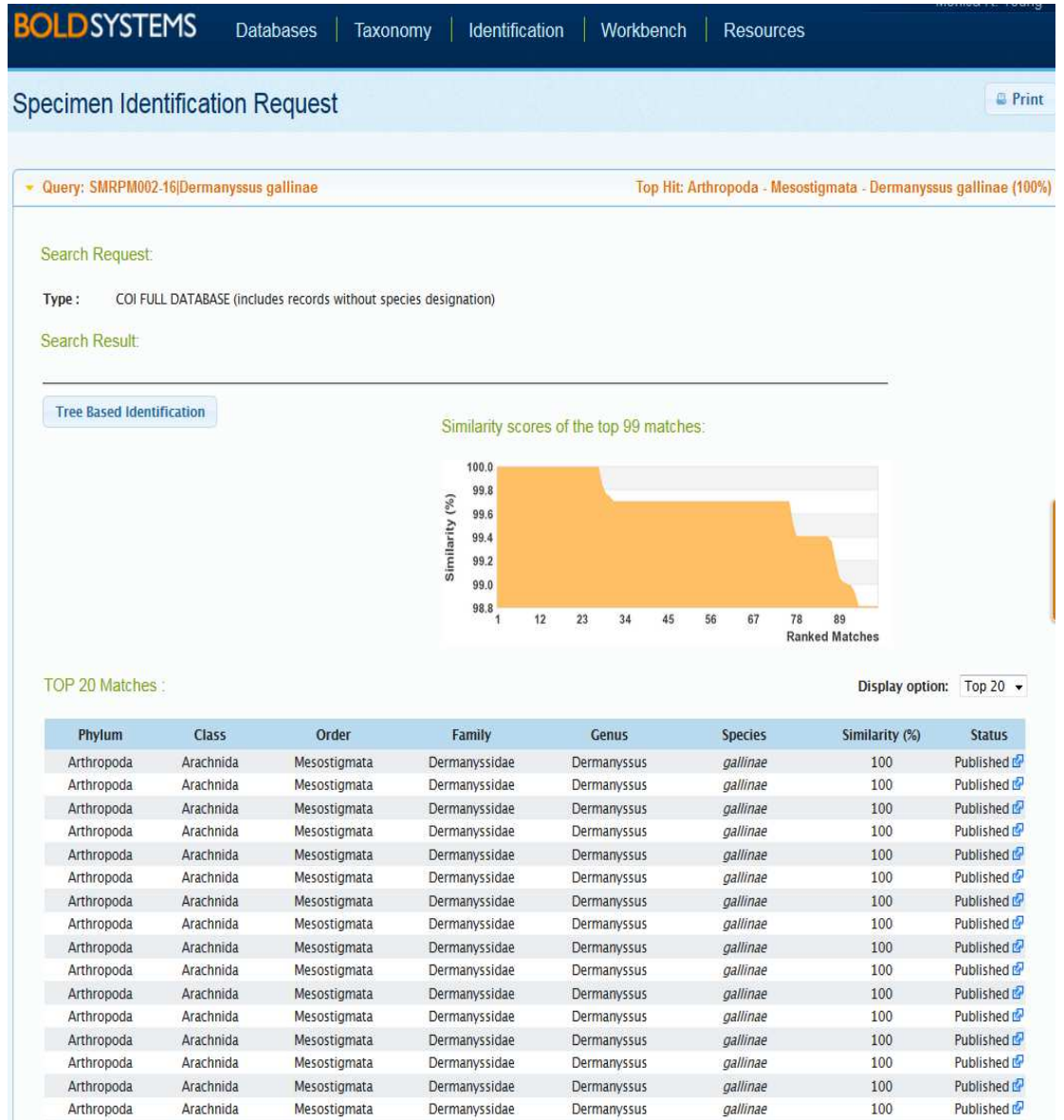
[Add Tags & Comments](#) Comments: 0 Associated Tags: No Tags

GEOGRAPHY

Country: Israel
Province/State:

Introduction

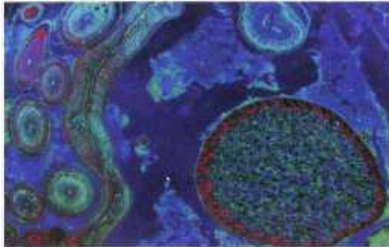
-

[illegible]

DNAB: applications

WHEN SPECIES IDENTIFICATION MATTERS

Fighting
Disease



Monitoring
the Environment



Controlling
Pests



Protecting
Endangered
Species



Exposing
Market Fraud



Making every species count

DNAB: empirical tests

- >>200 studies

Table 1. Design details and major results from selected studies that utilised part of the cytochrome c oxidase subunit 1 gene to provide a DNA barcode for identifying animal taxa

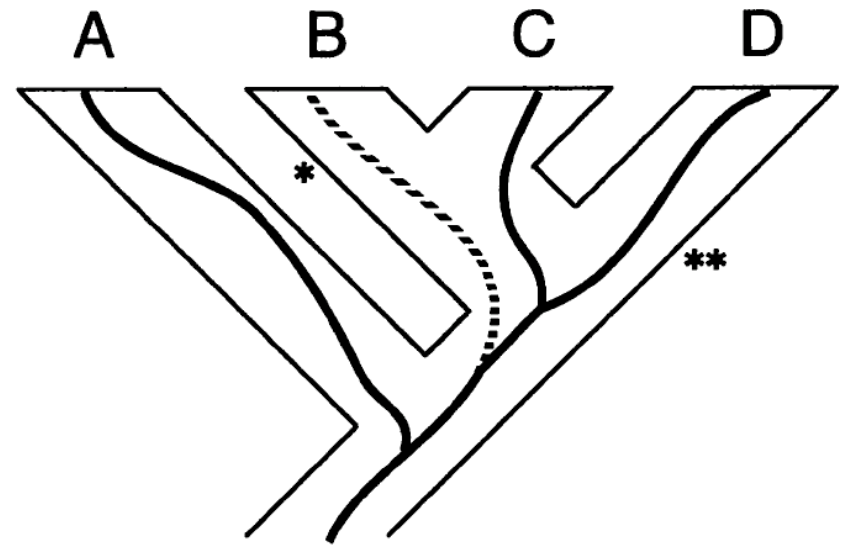
Reference	Sequence length	No. of taxa	% identified	K2P % intra vs interspecific COI variation	Comment
Barrett et al. 2005 ⁽²³⁾	600bp	203 arachnid spp.	100	1.4 vs 16.4	Mean intra- and interspecific nucleotide divergences did not overlap except in the case of probable cryptic species. COI was used as a taxonomic descriptor for a new species.
Brown et al. 2003 ⁽⁴¹⁾	648bp	1 Lepidopteran sp. <i>Gnetom gnemon</i>	100		
Greenstone et al. 2005 ⁽²⁶⁾	439bp	32 Carabidae spp. and 39 Araneae spp.			COI used to identify species and link different life stages.
Hajibabaei et al. 2006 ⁽²¹⁾	311–612bp	521 Lepidopteran spp.	97.9%	0.17–0.46 vs 4–6	Morphologically distinct sympatric species from three families identified.
Hebert et al. 2003a ⁽⁷⁾	658bp	7 animal phyla	96.4		The efficacy of COI at identifying species, orders and phyla assessed.
Hebert et al. 2003b ⁽²⁸⁾	>400bp	8 insect orders	100	Lepidoptera 0.25 vs 6.84 Overall, usually <2 vs 11.3	The efficacy of COI at identifying species from eight major and several minor phyla plus a variety of arthropod classes was assessed. Cnidarians showed less COI variation between species than all other taxonomic groups, 94.1% vs 1.9% showing <2% K2P between spp. ($p < 0.0001$).
		200 Lepidopteran spp.	100		
		2238 Annelida, Arthropoda, Chordata, Cnidaria, Echinodermata, Mollusca, Nematoda, Platyhelminthes, and minor phyla	>98		
Hebert et al. 2004a ⁽³⁴⁾	648bp	10 Lepidopteran spp.	100		Ten new taxa identified. Different life stages matched for species.
Hebert et al. 2004b ⁽²⁴⁾	648bp	260 Avian spp.	100	0.43 vs 7.93	Four possible cryptic species identified.
Hogg et al. 2004 ⁽²⁵⁾	710bp	19 Collembola spp.	100	0.76 vs 19.0	Produces high resolution in Collembola species. Possible cryptic species identified.
Hu et al. 2002 ⁽³⁶⁾	<530bp	7 Hookworm spp.	—		Three of the seven species appeared to be possible species complexes based on intraspecific COI variation.
Hu et al. 2005 ⁽³⁸⁾	≈450bp	3 <i>Progamotenia</i> spp. (Platyhelminthes)			Variation at COI suggests that all three species are species complexes.
Lambert et al. 2005 ⁽¹³⁾	596bp	10 Mos spp.	100		COI used to identify extinct species. Possible species synonymy highlighted.
Lorenz et al. 2005 ⁽²⁷⁾	727bp	56 primates		0.011 vs -	Problems with taxon specific patterns of 'universal primer' failure. Taxon specific primers developed.
Ngamamopiratt et al. 2005 ⁽³⁷⁾	450bp	1 <i>Gnathostoma spinigerum</i> (hookworm)	—		COI barcode variation did not match the morphological variation observed in 3 rd larval stage of this hookworm.
32 Paquin et al. 2004 ⁽³⁹⁾	≈1000bp	23 <i>Clarinia</i> spp. (Arachnida)	≈100	1.09 vs 7.12	Immature specimens identified to species level. Possible species synonymy identified.
Penton et al. 2004 ⁽²⁹⁾	709bp	2 <i>Daphnia</i> spp. (Crustacea)	100		Identification of morphologically cryptic species with overlapping distribution.
Remigio et al. 2003 ⁽²⁸⁾	672bp	70 Gastropod spp.	—		COI used to identify species and higher taxonomic relationships. Insertion or deletions more common in COI in this taxonomic group.
Smith et al. 2006 ⁽³⁰⁾	658bp	32 Dipteran spp.	100	0.17 vs 5.78	Fifteen cryptic species found using COI.
Vences et al. 2005 ⁽³³⁾	550–650bp	9 Mantellid frog spp. 4 <i>Aneides</i> spp. (salamanders)		5.4 (mantellid frogs) and 4.3 (salamanders) vs 20.7 and 13.5	Found high intraspecific variability (7–18%). The use of mitochondrial 16S rRNA gene to supplement COI suggested.
Ward et al. 2005 ⁽²⁹⁾	655bp	207 fish	100	0.39 vs 9.93	Efficacy of COI at identifying species and higher taxonomic relationships assessed. Possible cryptic species identified.
Whiteman et al. 2004 ⁽⁴⁰⁾	379bp	2 Lice spp. (Insecta)	100		Study of bird parasites with similar morphology that can be identified using COI.

Waugh 2007 Table 1

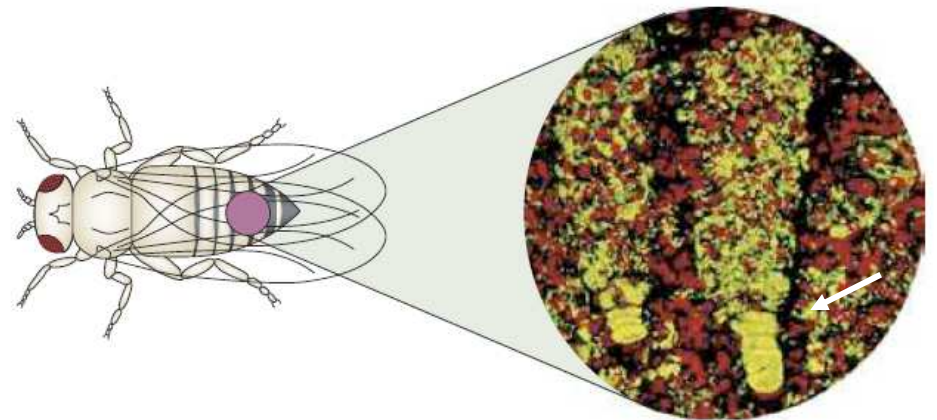
No. of taxa	% identified
203 arachnid spp.	100
1 Lepidopteran sp. <i>Gnetom gnemon</i>	100
32 Carabidae spp. and 39 Araneae spp.	
521 Lepidopteran spp.	97.9%
7 animal phyla	96.4
8 insect orders	100
200 Lepidopteran spp.	100
2238 Annelida, Arthropoda, Chordata, Cnidaria, Echinodermata, Mollusca, Nematoda, Platyhelminthes, and minor phyla	>98
10 Lepidopteran spp.	100
260 Avian spp.	100
19 Collembola spp.	100

DNAB: the limitations

- recently diverged species
 - incomplete lineage sorting
- introgressive hybridization
- *Wolbachia* infection



incomplete lineage sorting (Maddison 1997)



Wolbachia infection (Werren et al 2008)

DNAB: case study

- **Soil mites in association with the red poultry mite**

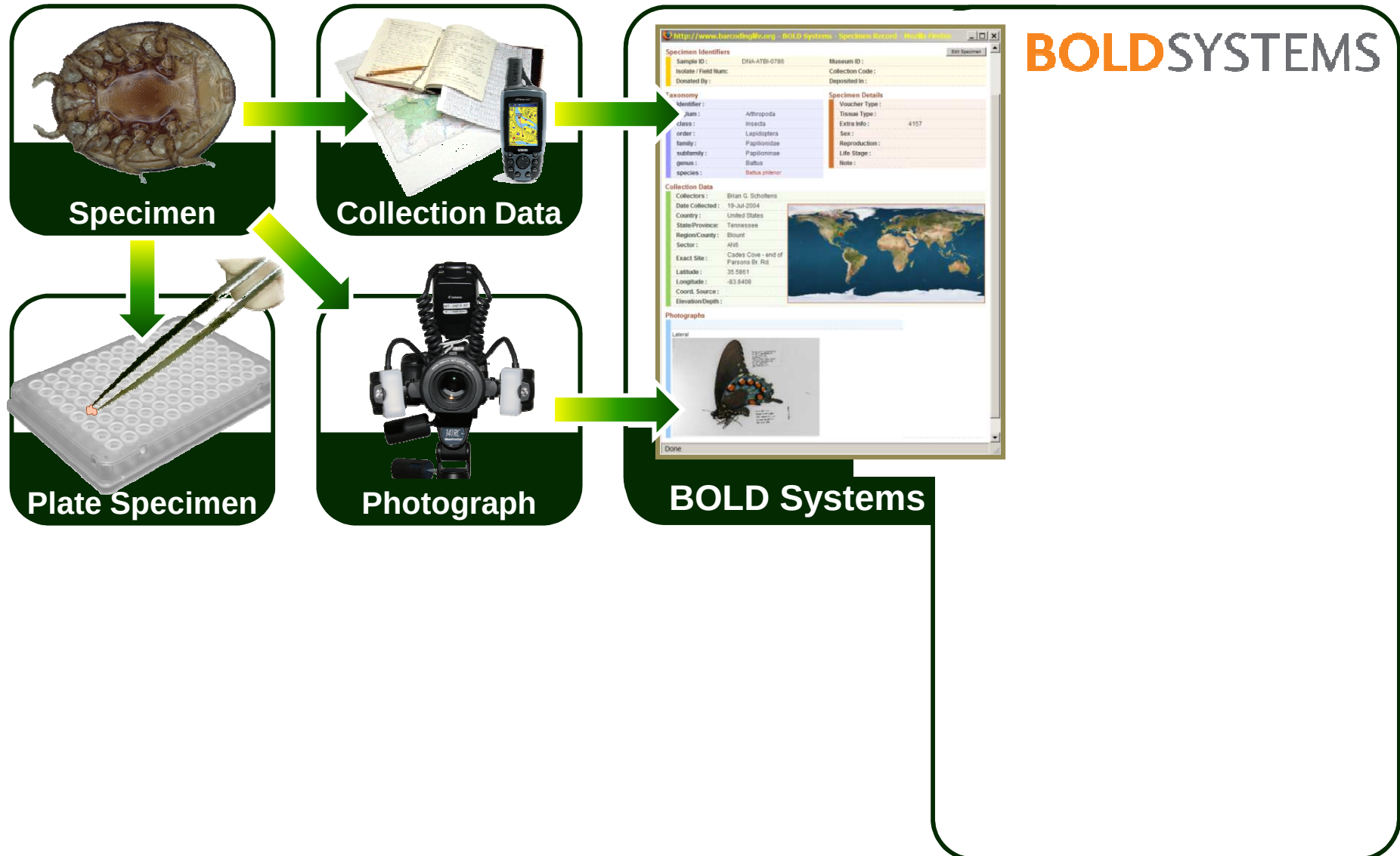
1. library construction

- Collect samples

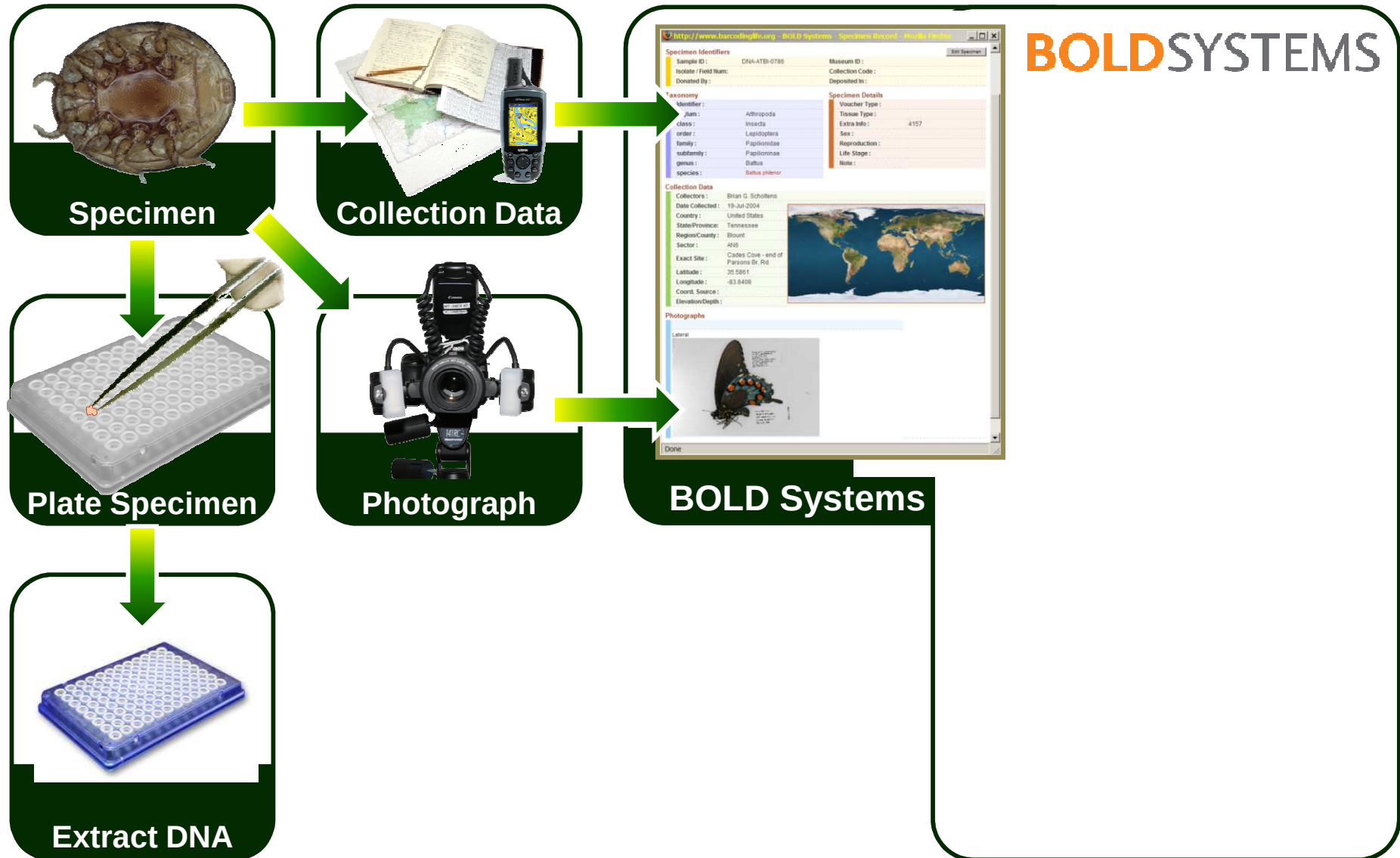
- Belgium
- Croatia
- France
- Israel
- Poland



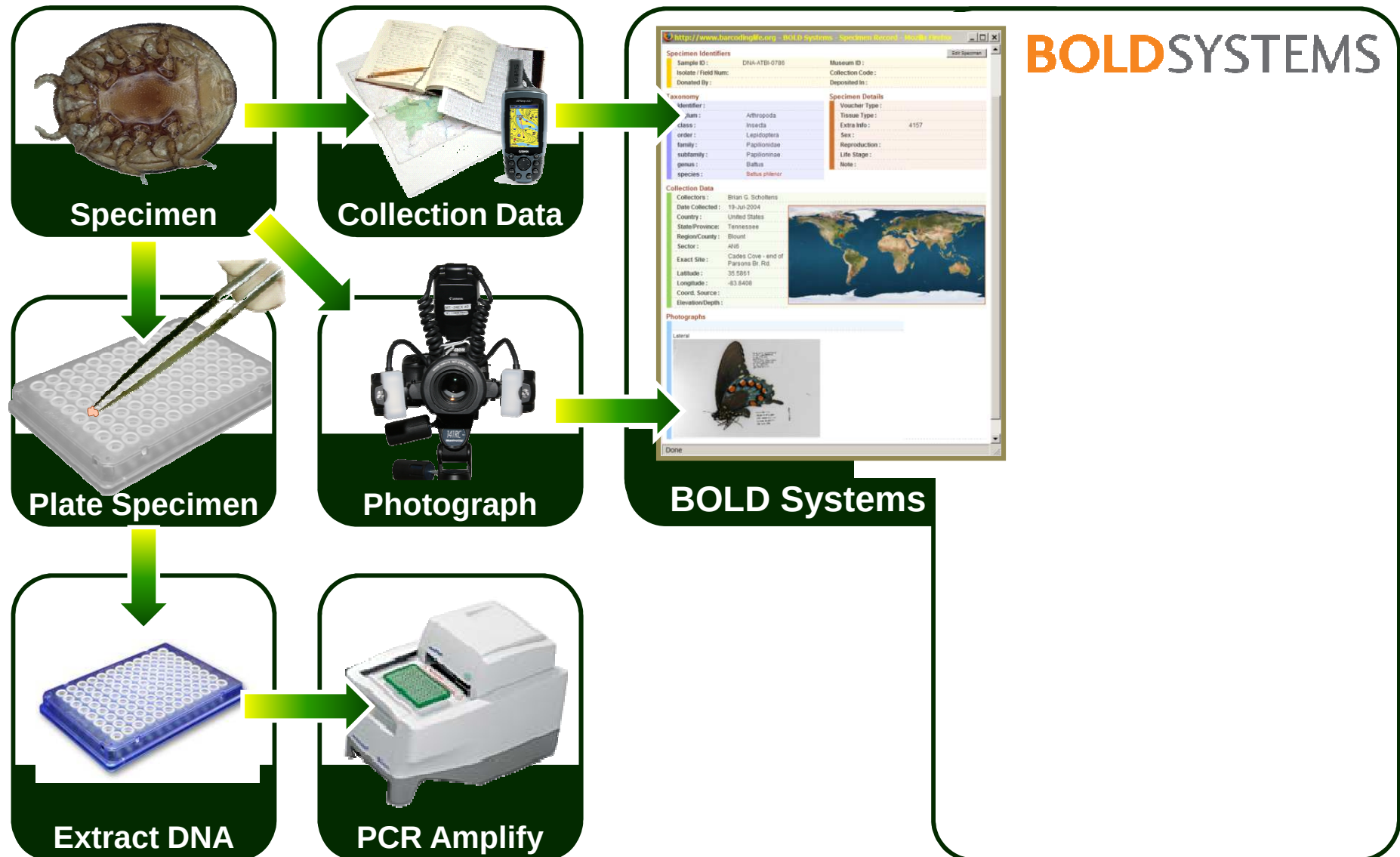
Case study: specimen processing



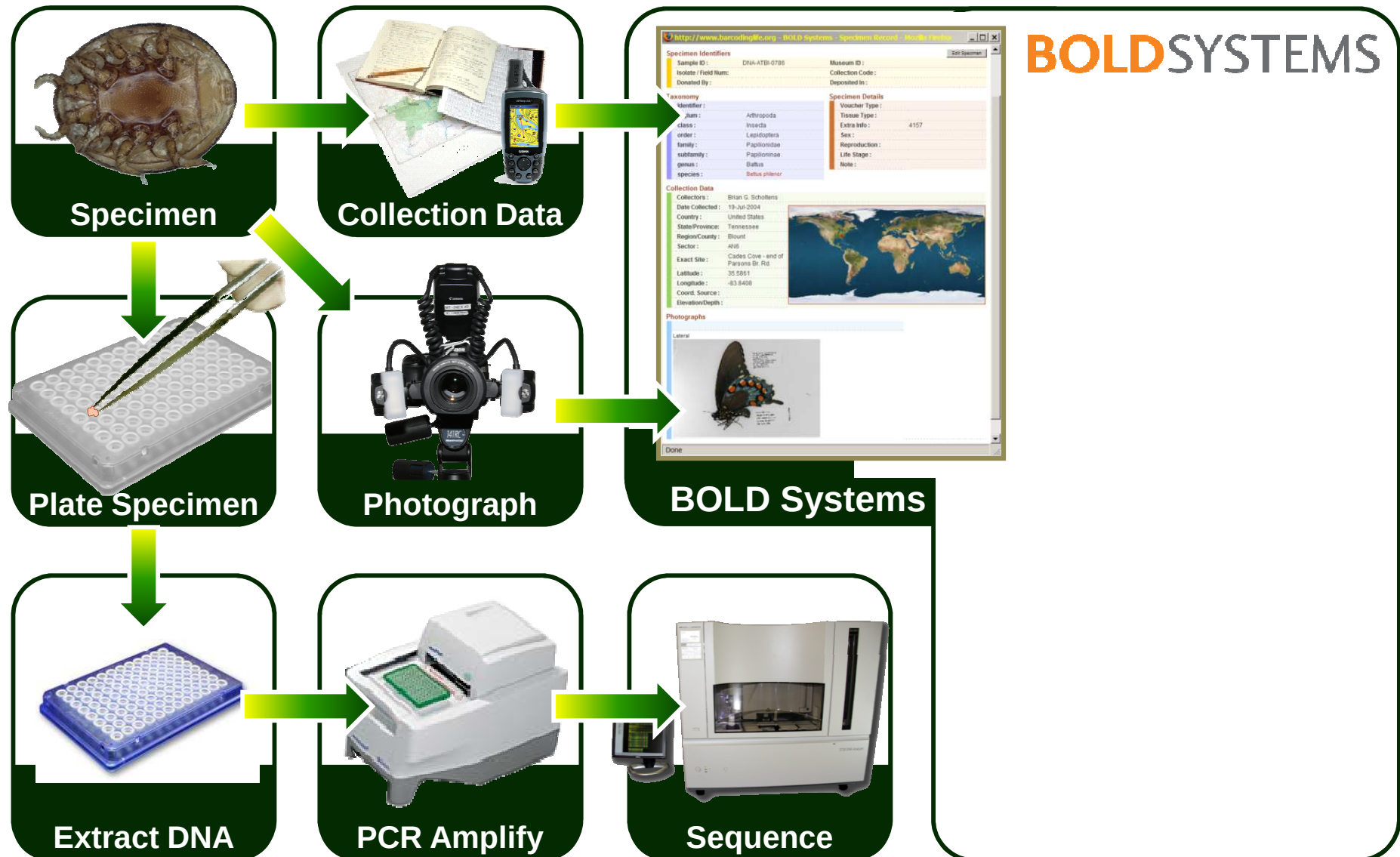
Case study: extraction



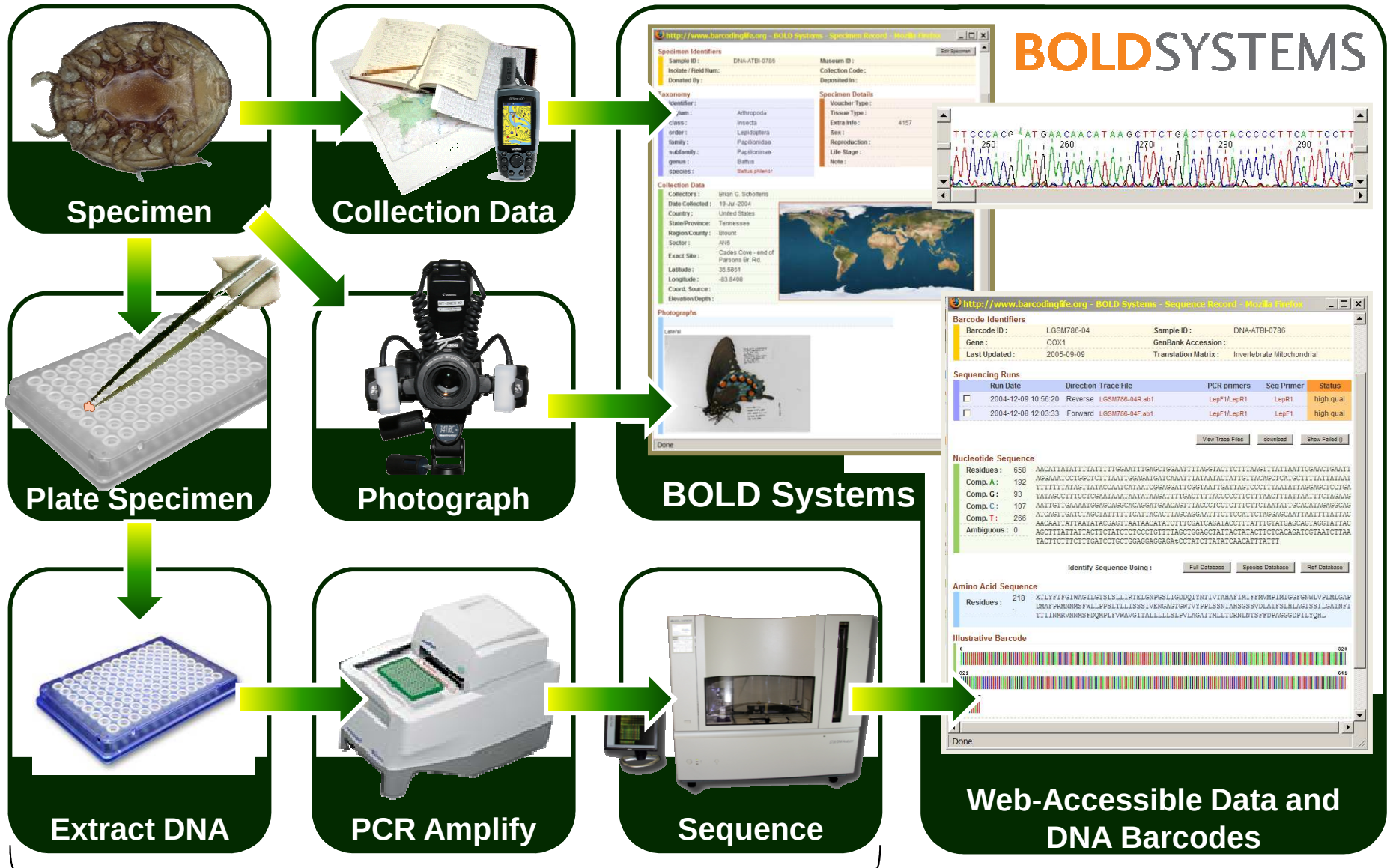
Case study: amplification



Case study: sequencing

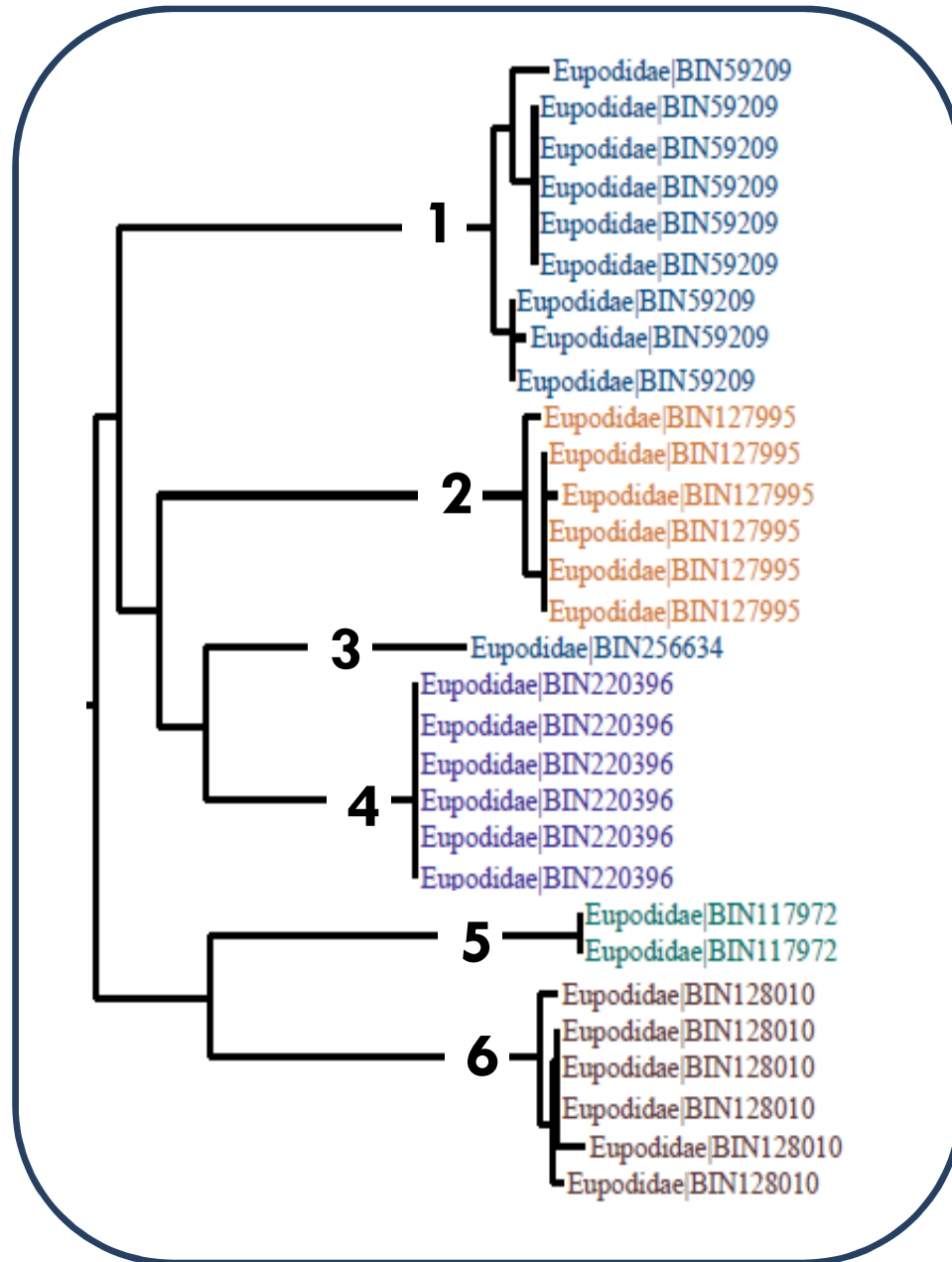


Case study: data integration

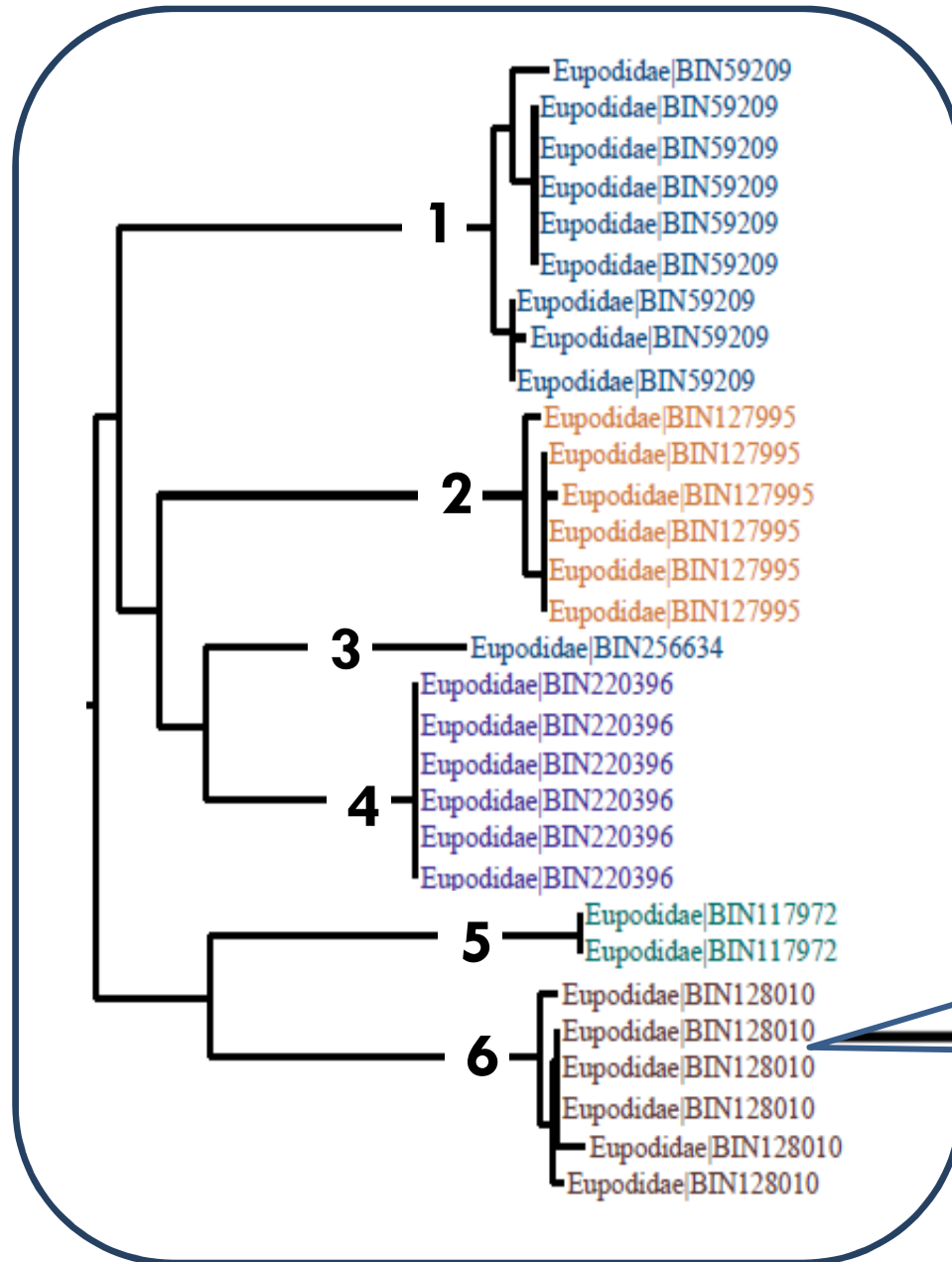


<http://ccdb.ca/>

Case study: delineating species



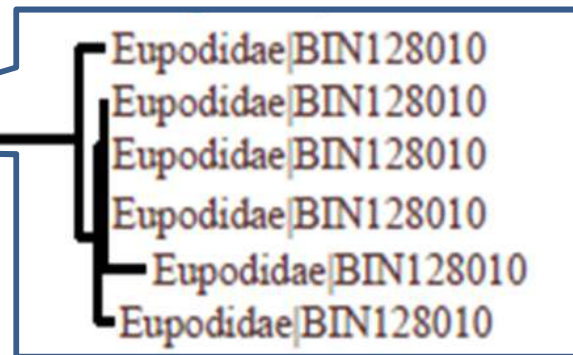
Case study: delineating species



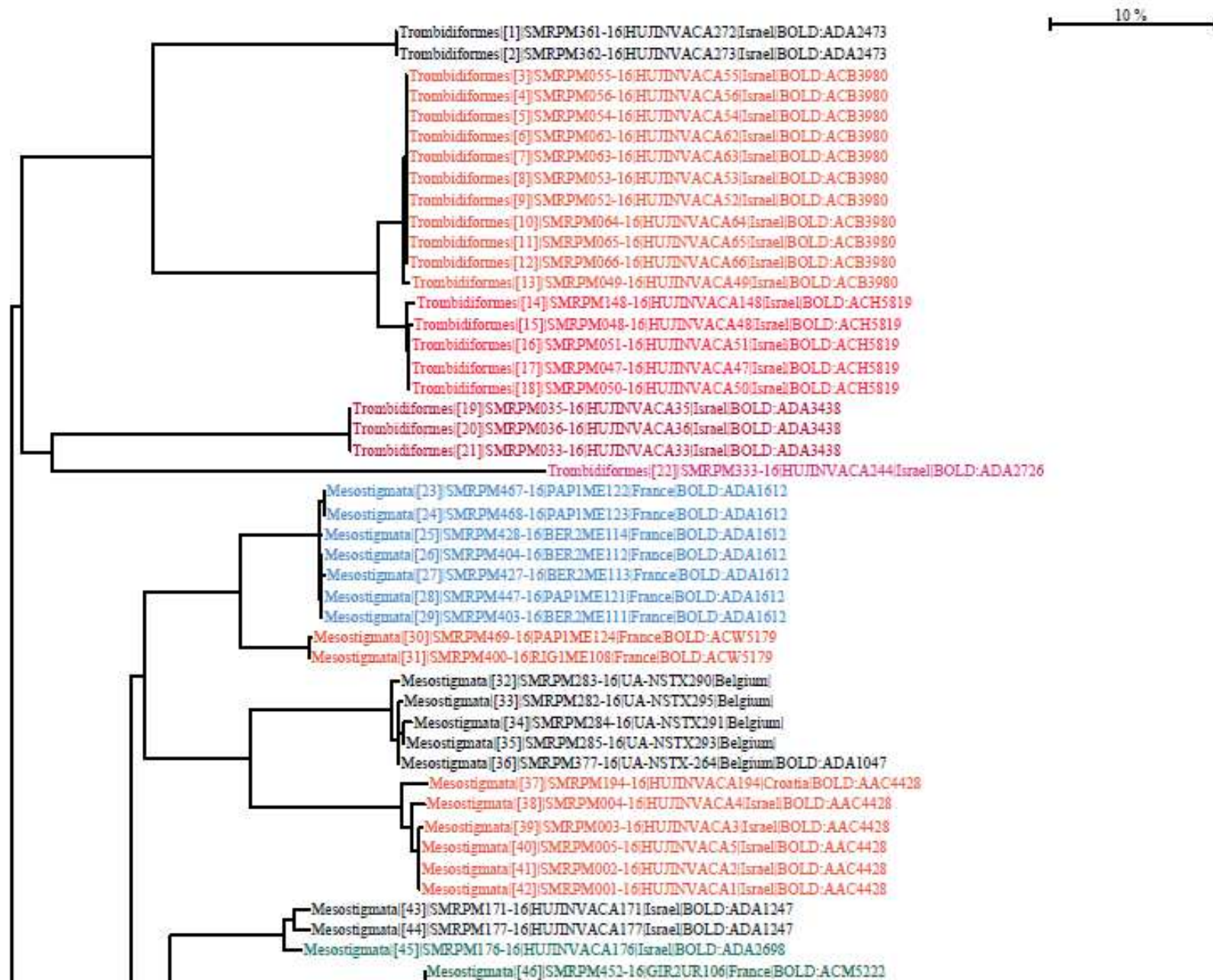
Barcode Index Number (BIN) system

(Ratnasingham and Hebert 2013)

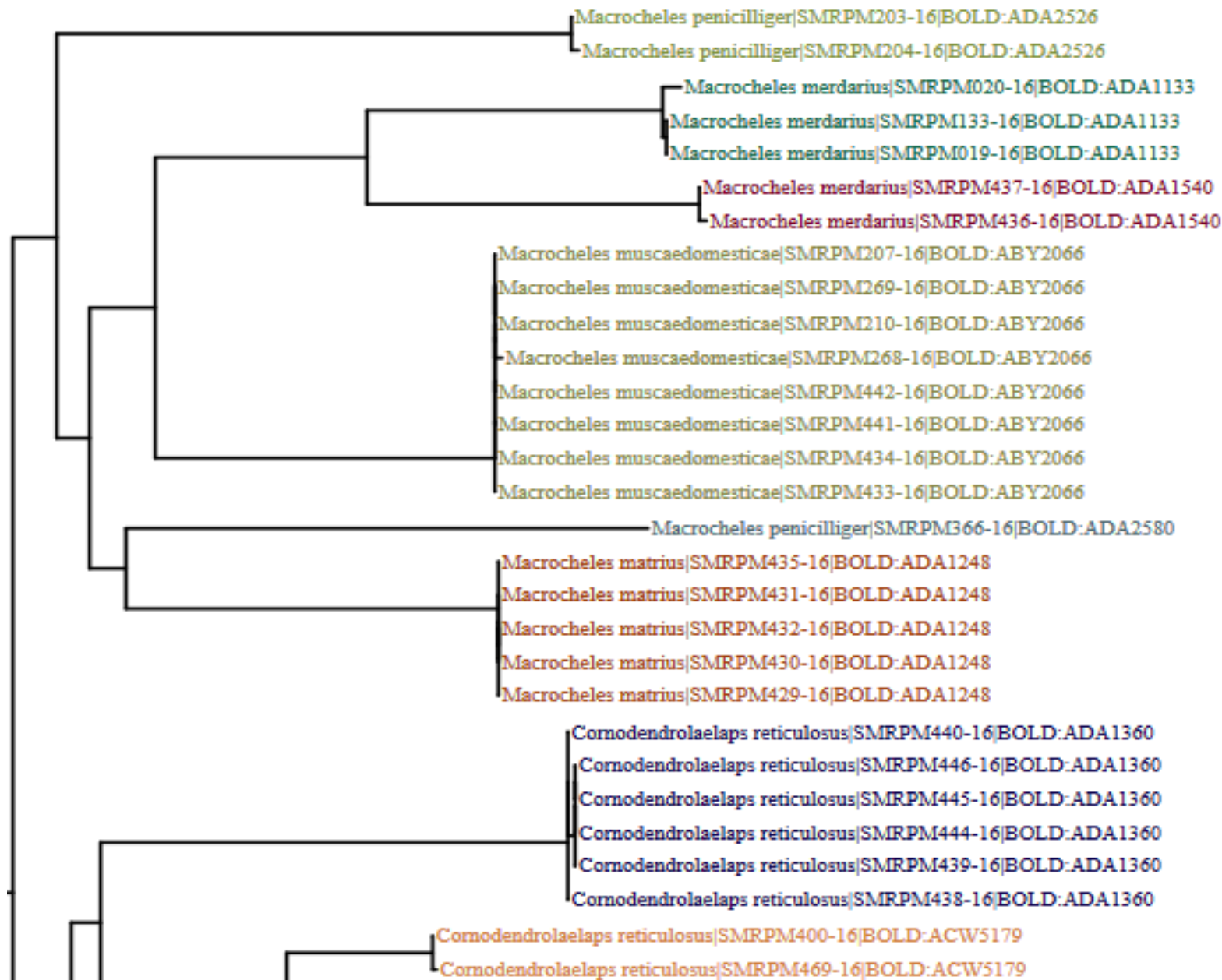
- Clustering algorithm
- Defines molecular taxonomic units



Case study: delineating species

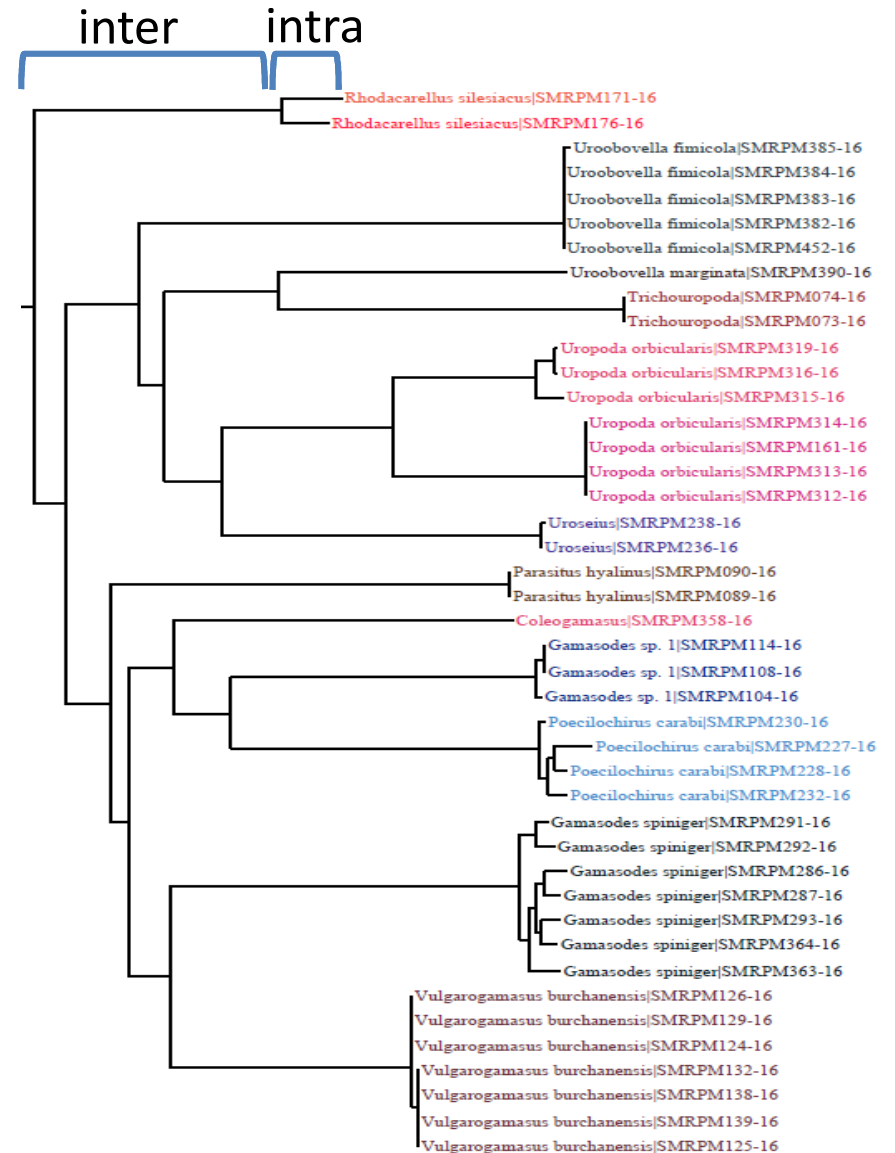


Case study: specimen identification



Case study: results

- Extremely successful!!!
- High interspecies distances & low interspecies distances = **good barcode resolution**
- However.....



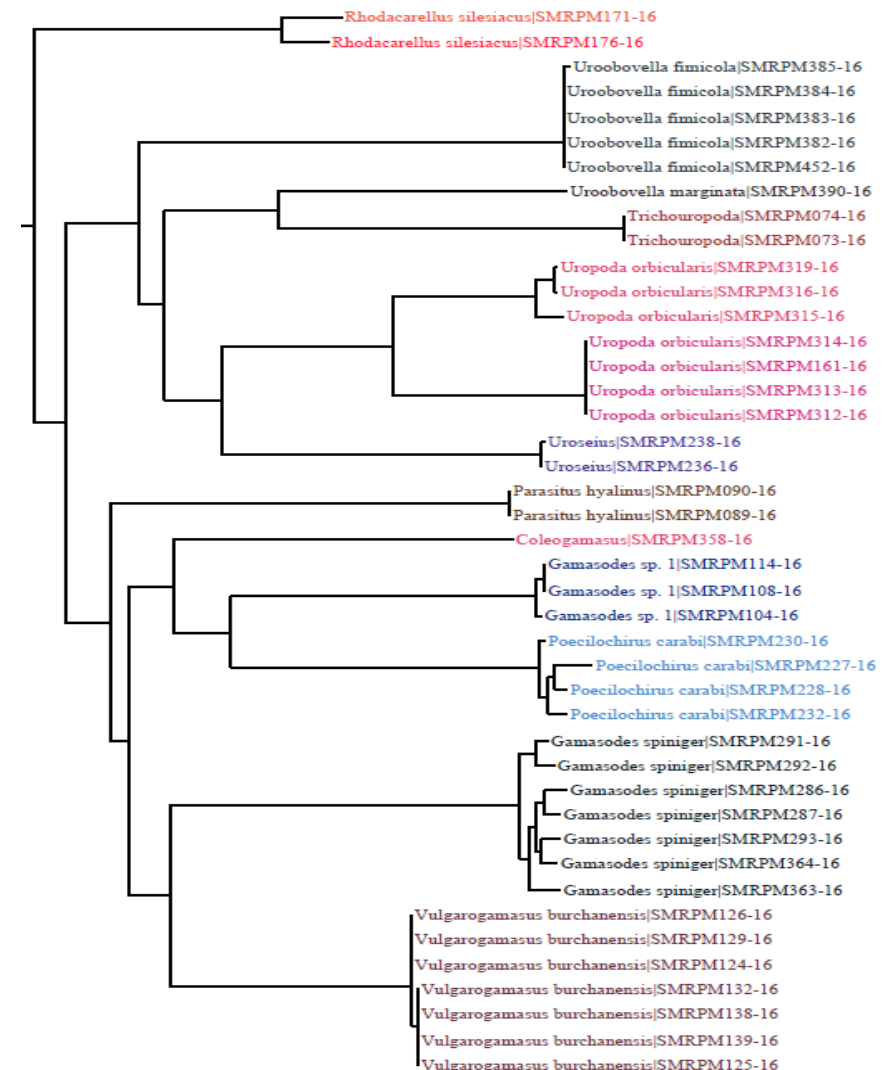
Case study: results

- Extremely successful!!!
- High interspecies distances & low intraspecies distances = **good barcode resolution**
- However.....
- Some cases of species splitting



Case study: building the key

- Based on these identifications
 - Instructors compiled a species list
 - Constructed a key to the taxa
 - Prepared slides and SEM photos



Case study: next steps

- **Soil mites in association with the red poultry mite**

1. library construction
2. monitoring/identification of unknowns



Case study: next steps

- **Soil mites in association with the red poultry mite**

1. library construction
2. monitoring/identification of unknowns



**High
Throughput
Sequencing**



Next Generation Sequencing Methods



High Throughput Sequencing



Species identifications from 'slurries'

Barcode of Life Datasystems (BOLD)



DESIGNED TO SUPPORT THE GENERATION & APPLICATION OF DNA BARCODE DATA

BOLD is a cloud-based data storage and analysis platform developed at the Centre for Biodiversity Genomics in Canada. It consists of four main modules, a data portal, an educational portal, a registry of BINs (putative species), and a data collection and analysis workbench.

Please note that this version of BOLD is in beta and will contain bugs. Users can help address these bugs by testing the system and reporting issues to support@boldsystems.org. This version is very different from the prior one but has access to all the same data.



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.



EDUCATION PORTAL

A custom platform for educators and students to explore barcode data and contribute novel barcodes to the BOLD database.



BIN DATABASE

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



WORKBENCH

A data collection and analysis environment that supports the assembly and validation of DNA barcodes and other sequences.

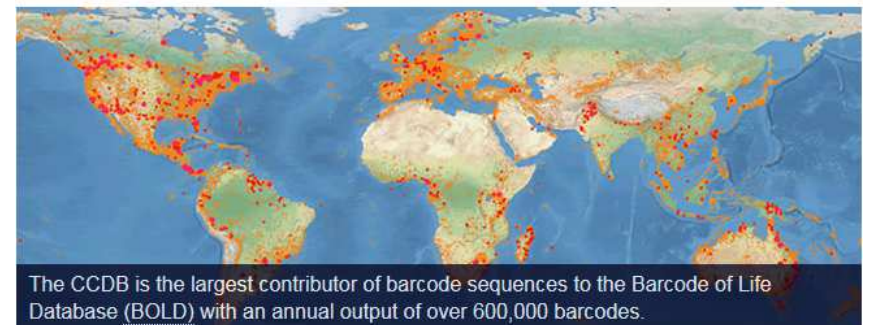
BOLD: barcode repository



- Data sources
 - CCDB (Canadian Centre for DNA Barcoding, Guelph, Canada)
 - User generated
 - GenBank mined
- Markers
 - COI primarily (barcode region)
 - ANY and ALL!



[Home](#) | [Services](#) | [Research](#) | [BOLD](#) |
[About Us](#) | [Contact Us](#)



BOLD: barcode repository



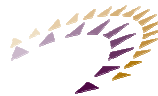
- Data integrity
 - Active, ongoing, and real-time curation of taxonomy and sequence data!
 - Combines taxonomy, photographs, sequences, trace files, collection/field notes

BOLD: live demo



GenomeCanada

international
BARCODE
OF LIFE



Canada Foundation for Innovation
Fondation canadienne pour l'innovation



NSERC
CRSNG



Ontario**Genomics**Institute



Ontario
MINISTRY OF
RESEARCH AND INNOVATION

CN tower 'barcoded' with Canadian
beaver COI for the launch of the
International Barcode of Life
Project in Sept 2010



myoung02@uoguelph.ca