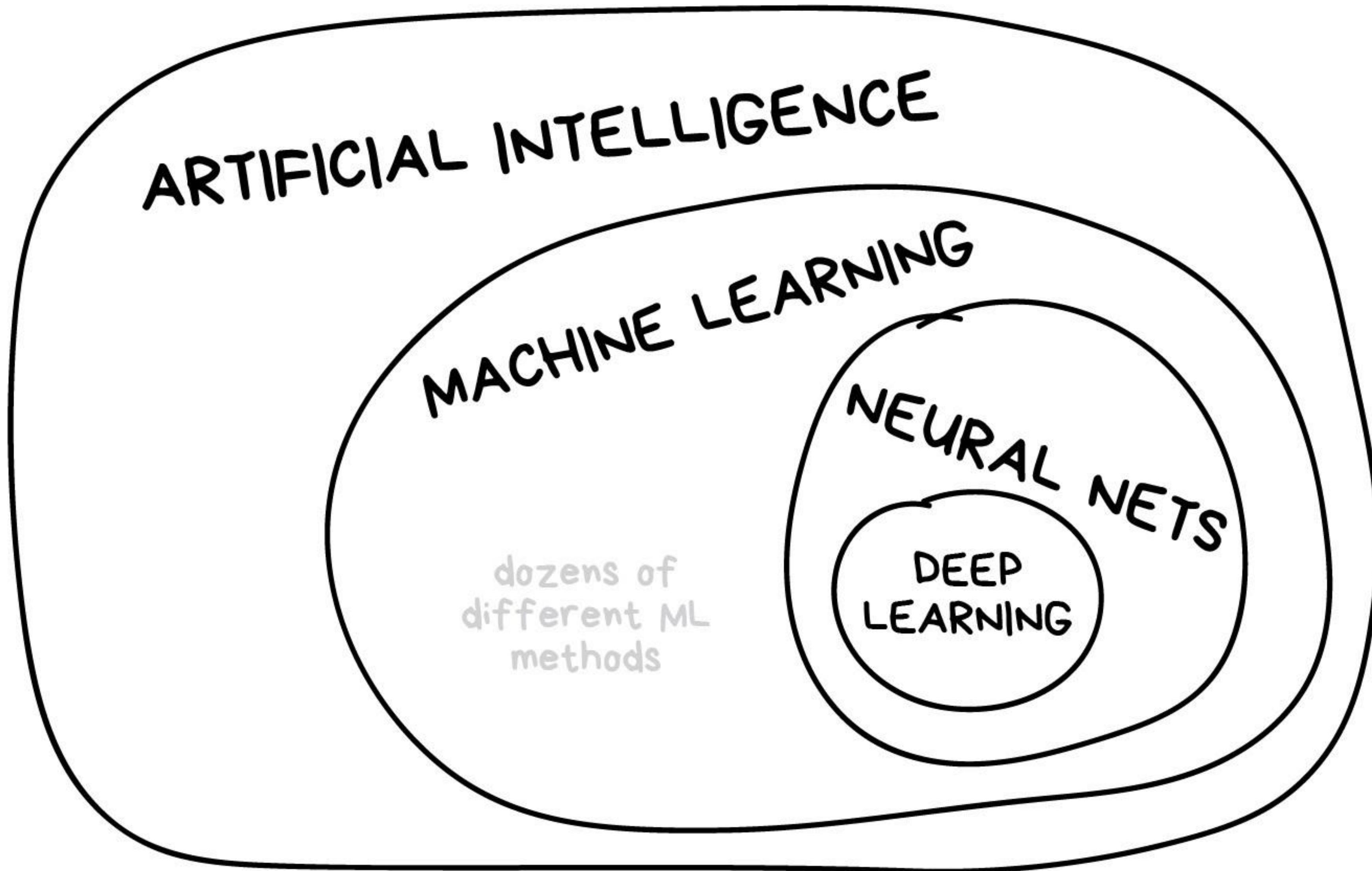


BIO  
DIVERSITY  
NEXT

# Machine Learning & Automation

**Laurence Livermore | Natural History Museum,  
London**





## Key technologies/applications

- Computer Vision (CV)
- Optical Character / Handwritten Text Recognition (OCR/ HTR)
- Natural language processing (NLP)
- Text/Data Mining

\*Not necessarily using ML approaches!

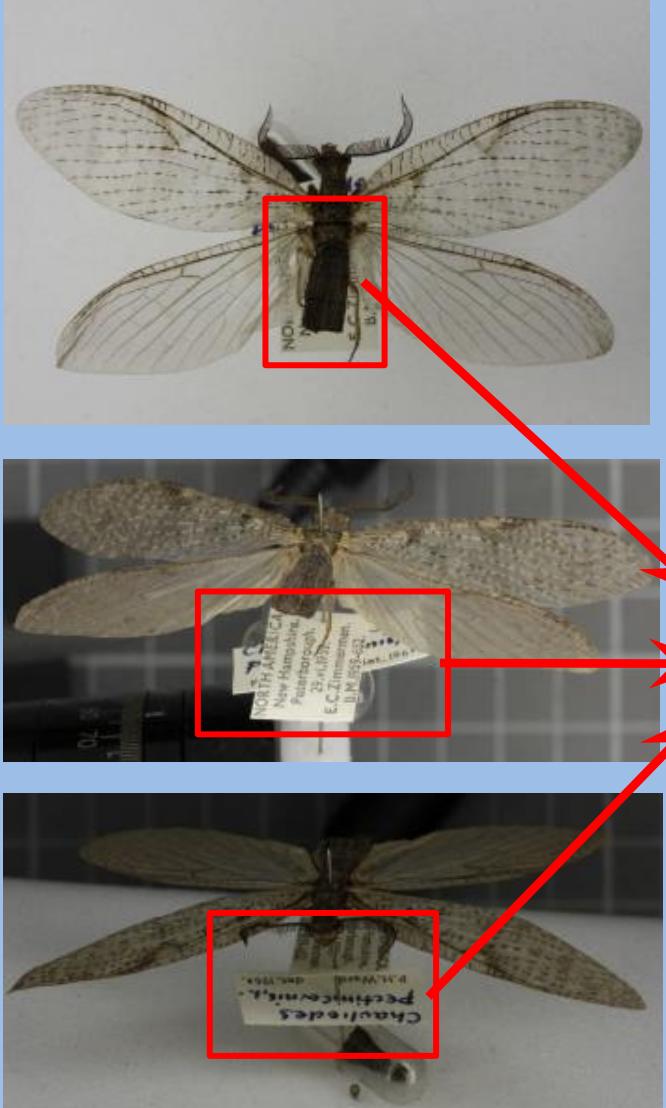


## Why is it relevant to our community?

- Automate parts of digitisation workflows
- Speed up transcription
- Capture new types of data
- Create links to enrich specimens / datasets

# Applications - IEFI

### Label Detection



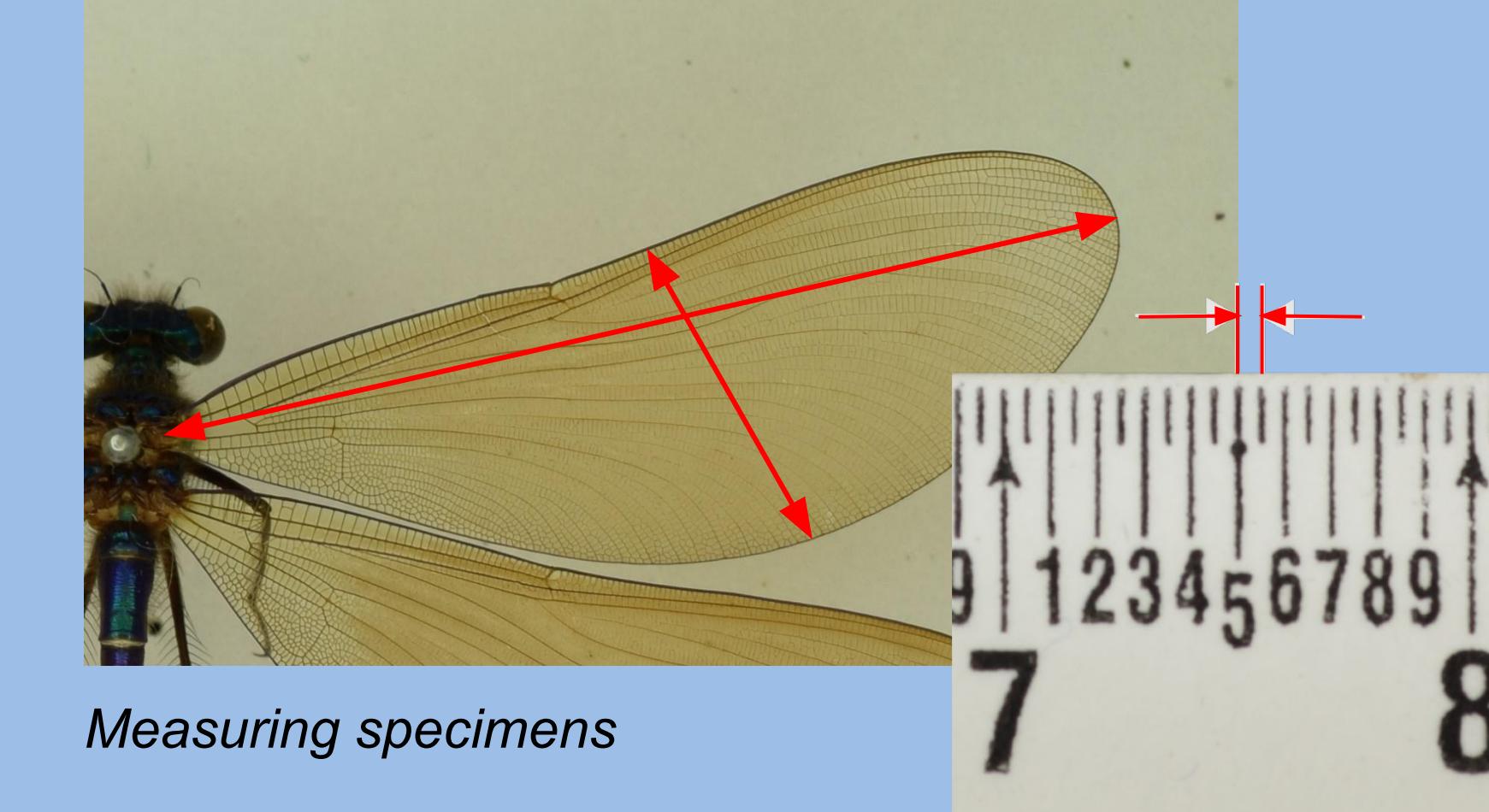
**Source**

**Reconstruction**

*Chauliodes pectinicornis, L.*  
P.H.Ward det. 1961.

NORTH AMERICA  
New Hampshire.  
Peterborough.  
29.vi.1959.  
E.C.Zimmerman.  
B.M.1959-652.

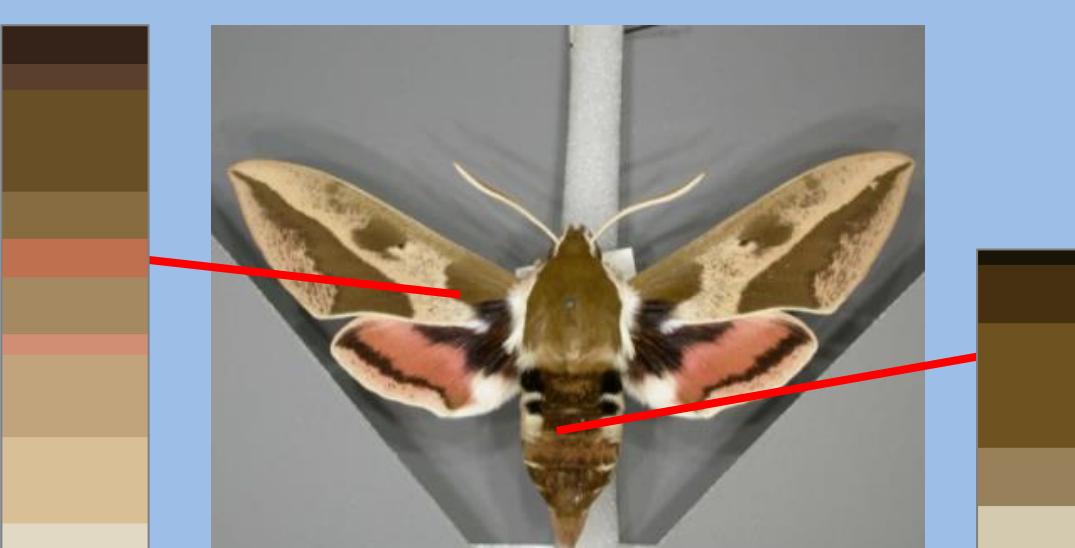
### Morphometrics



**Measuring specimens**

**Measuring rulers**

### Dominant Colour Clustering



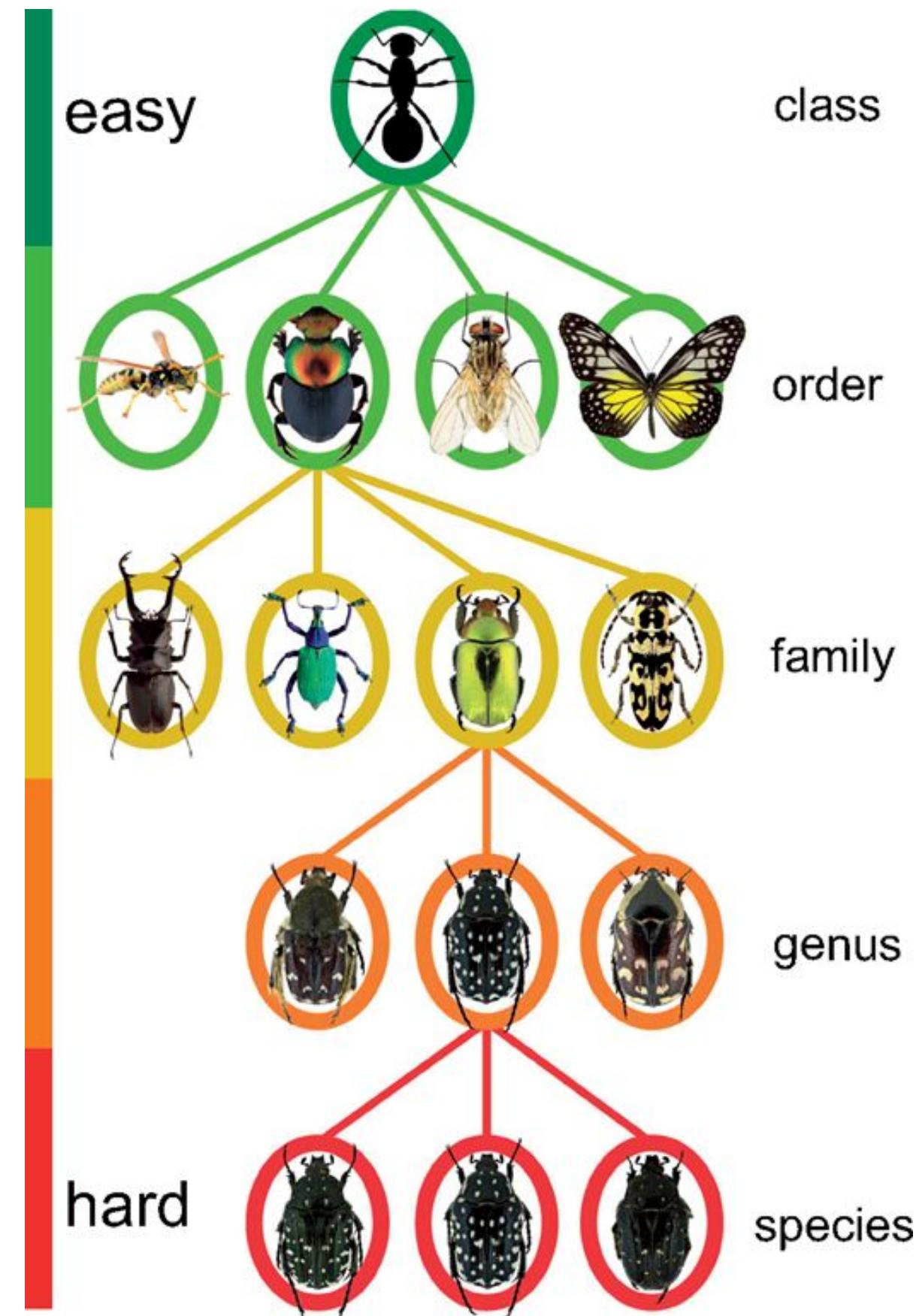
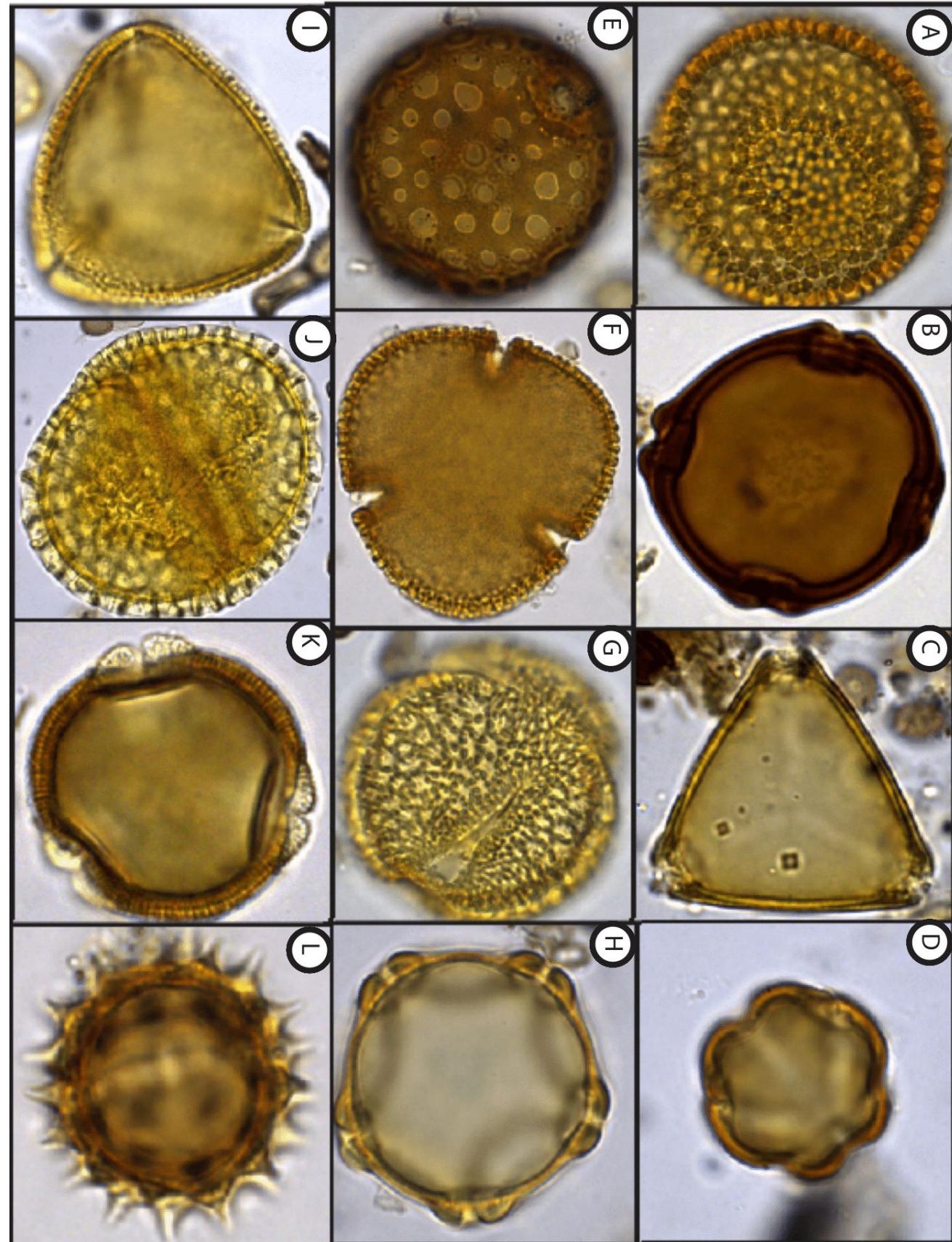
**Understanding hybridisation**

### Trait detection



**Sex brands in *Hesperia comma***

# Classifiers



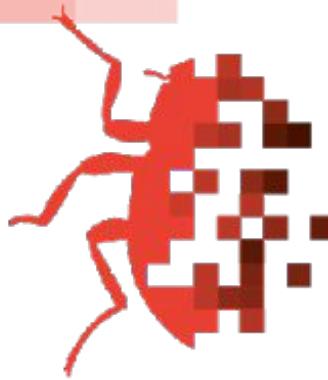
Derek Haselhorst, Program in Ecology, Evolution and Conservation Biology, University of Illinois, iDigBio Research Spotlight: September 2017

Miroslav Valan, Karoly Makonyi, Atsuto Maki, Dominik Vondráček, Fredrik Ronquist, Automated Taxonomic Identification of Insects with Expert-Level Accuracy Using Effective Feature Transfer from Convolutional Networks, *Systematic Biology*, , syz014, <https://doi.org/10.1093/sysbio/syz014>

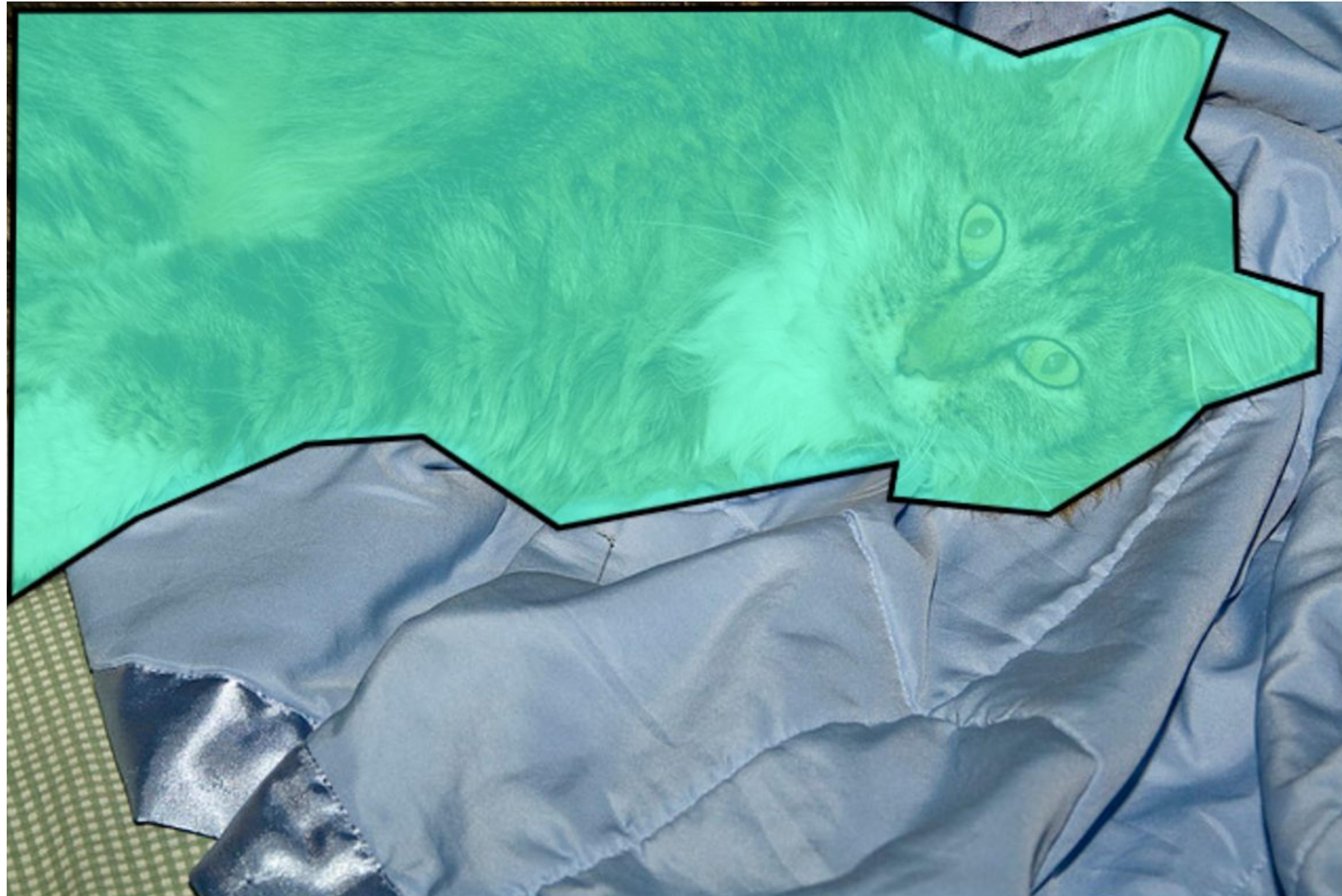


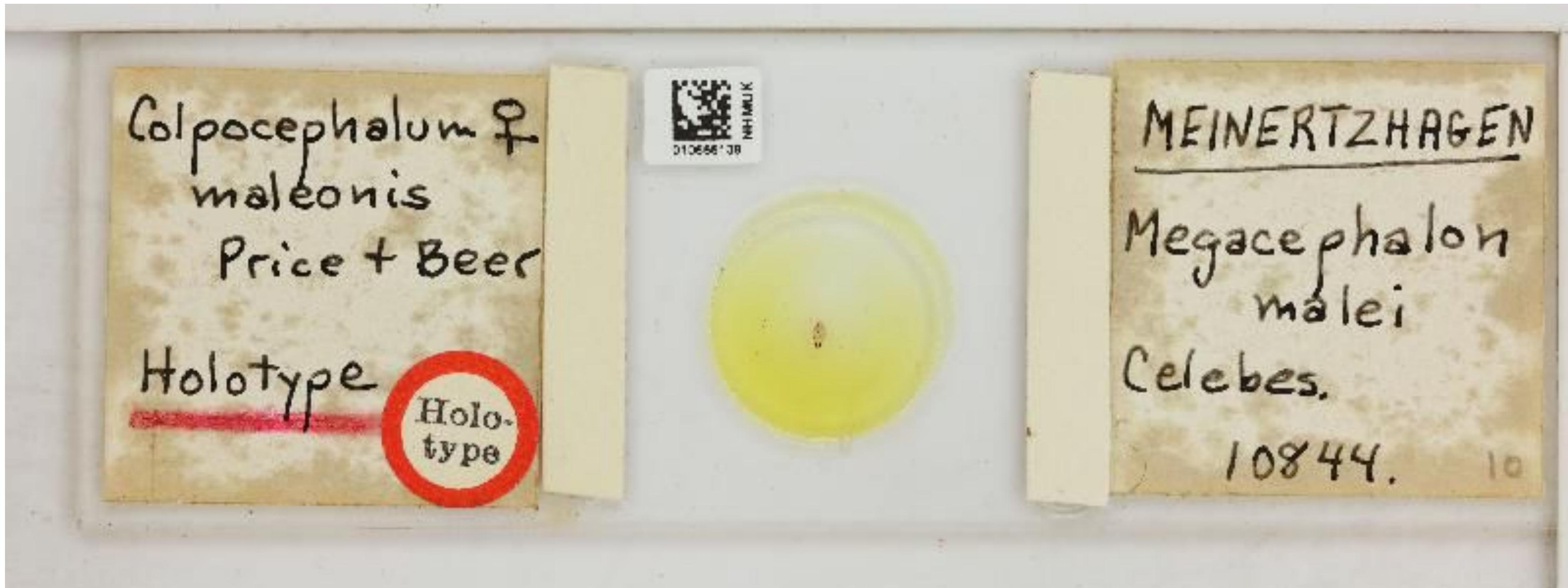
Carranza-Rojas J, Goeau H, Bonnet P, Mata-Montero E, Joly A (2017) Going deeper in the automated identification of Herbarium specimens. *BMC Evolutionary Biology* 17:181. doi: 10.1186/s12862-017-1014-z

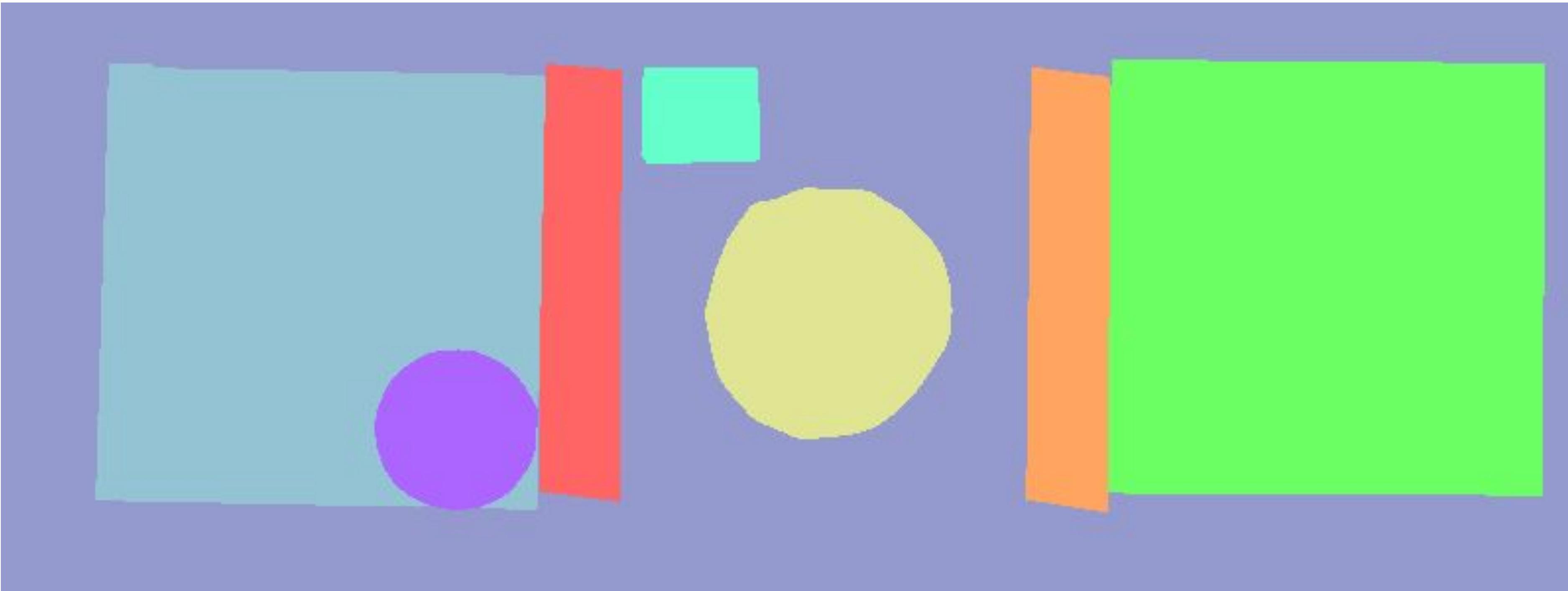




BIO  
DIVERSITY  
NEXT

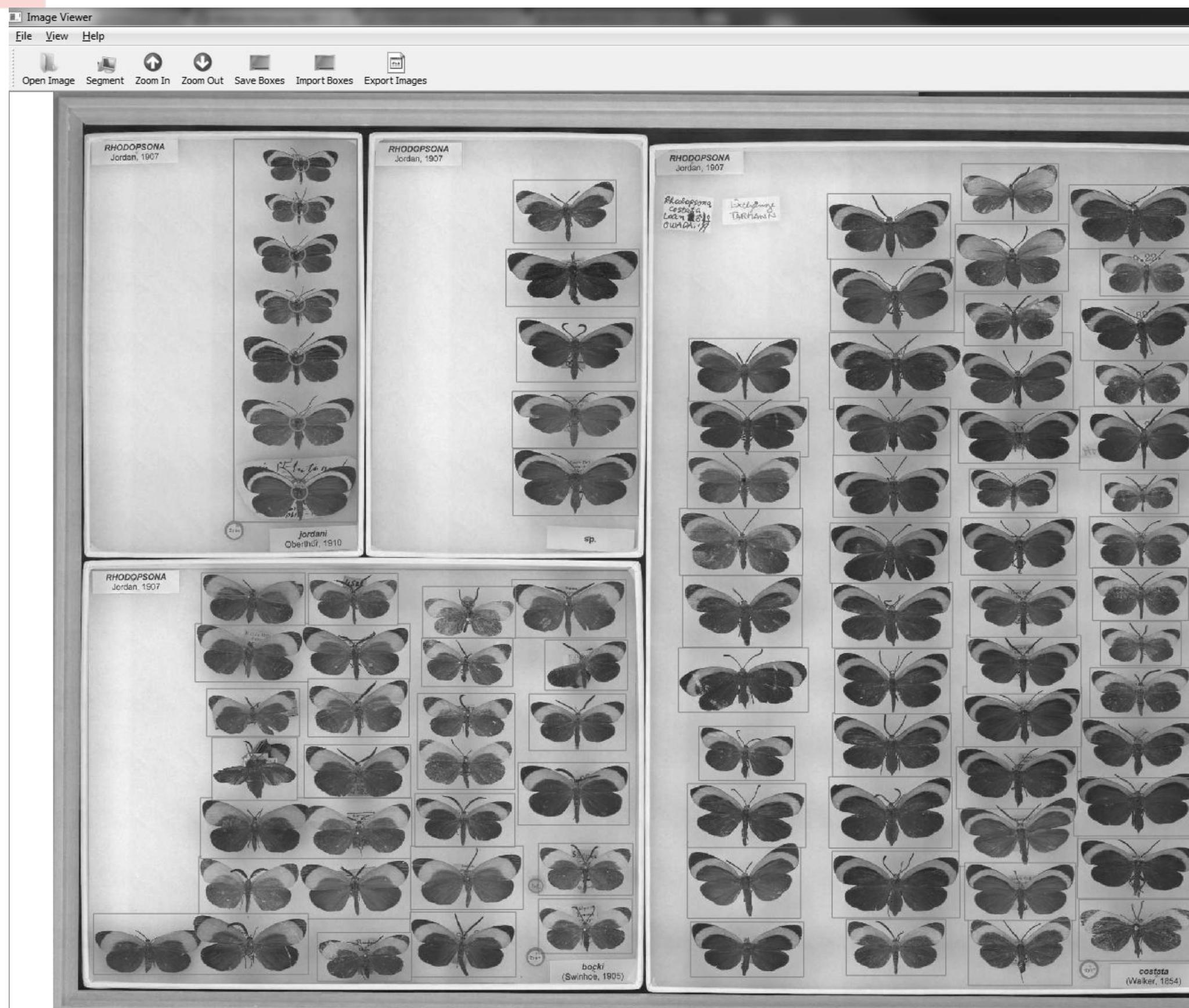








# Inselect



**Whole drawer image**

Auto. detects specimens, crops & exports images, rapid annotation, barcode reading, data export (JSON), open source, Python based, Windows, OSX & Linux



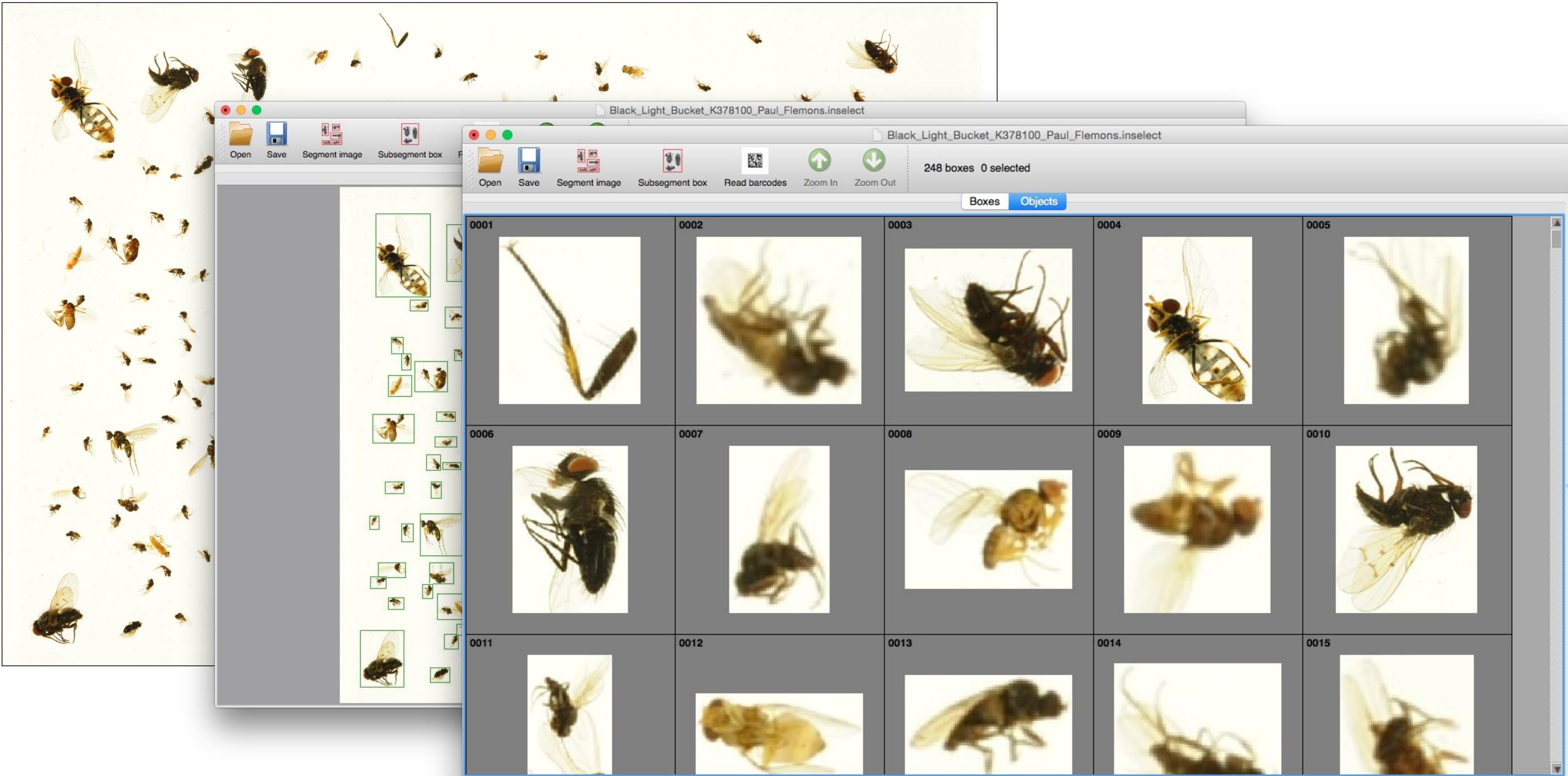
**Auto-segmented**

The screenshot shows the Inselect project page on GitHub. It includes a 'View project on GitHub' button, download links for '.zip' and '.tar.gz' files, and credits to NaturalHistoryMuseum. The main content area features a welcome message, authors and contributors (with Stefan van der Walt as the lead developer), and a note about SYNTHESYS funding. Below this is a 'Support or Contact' section and a detailed view of a specimen crop labeled 'Crop number 002'. The crop view shows a grid of 12 small images of insect specimens, each with a checkbox and a 'Delete' icon. To the right of the grid, there are fields for 'Specimen number' (BM001128339), 'Current taxon name', and 'Location in collection'.

**Annotation (vocab. import)**



# Inselect



1933] WOODSON — STUDIES IN THE APOCYNACEAE. IV 693 39. **Mandevilla subcordata** TAXON Rusby, Bull. N. Y. Bot. Gard. 4: 315. 1907. Suffruticose TRAIT lianas TRAIT ;

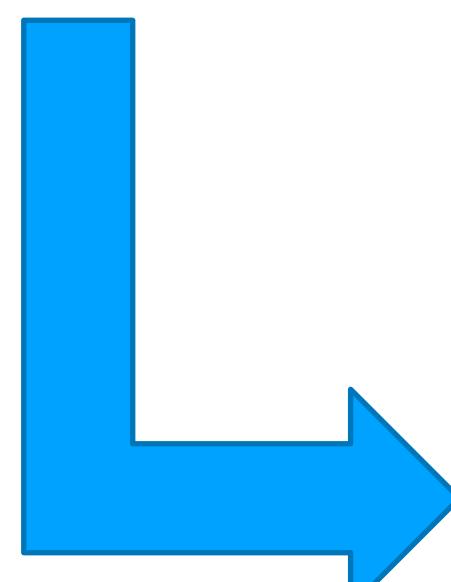
stems terete, relatively stout, puberulent TRAIT to glabrate TRAIT ; leaves opposite TRAIT , petiolate, broadly ovate to oblongelliptic TRAIT , apex abruptly acuminate, base broadly cordate TRAIT , rarely obtuse TRAIT or rounded TRAIT , 4-10 cm. long, 2.5-5.0 cm. broad, membranaceous, above glabrous TRAIT or very minutely puberulent when young TRAIT , glandular TRAIT at the base of the midrib, beneath densely and minutely tomentulose TRAIT to glabrate TRAIT ; petiole 1.0-2.5 cm. long; nodal appendages relatively inconspicuous; inflorescence lateral, alternate, simply racemose, about twice as long as the subtending leaves TRAIT , bearing 5-20 lax, greenishwhite or cream-colored flowers; pedicels 1.0-1.25 cm. long; bracts minutely ovate TRAIT , scarious; calyx-lobes ovate-lanceolate, 0.5-0.7 cm. long, scarious, minutely puberulent TRAIT to glabrate TRAIT , the squamellae in alternate groups of 4-6; corolla infundibuliform TRAIT , glabrous TRAIT without, the proper-tube straight, 1.5-2.0 cm. long, about 0.15 cm. in diameter at the base, the throat rather narrowly conical, 2.0-2.25 cm. long, about 1 cm. in diameter at the orifice, the lobes obliquely obovate TRAIT , shortly acuminate, 1.5 cm. long, spreading; anthers TRAIT auriculate, 0.5 cm. long; ovary TRAIT ovoid-oblongoid, 0.2 cm. long, glabrous TRAIT ; nectaries 5, compressed-ovoid, truncate TRAIT or obscurely emarginate TRAIT , about 0.1 cm. long; follicles relatively stout TRAIT , continuous, 20-30 cm. long, glabrous TRAIT ; seeds TRAIT about 0.75 cm. long, the pale TRAIT tawny coma about 2 cm. long.

Bolivia: la paz: near snow-line, Mt. Tunari, 1891, Bang 1120 (C, FM, NY, type, US, MBG, photograph and analytical drawings); Yungas, alt. 4000 ft., 1885, Rusby 2394 (NY). 40. **Mandevilla**

**Bridgesii** TAXON (Muell.-Arg.) Woodson, Ann. Mo. Bot. Gard. 19:67. 1932. Ambly anther a Bridgesii Muell.-Arg. Linnaea 30: 420. 1860; Miers, Apoc. So. Am. 189. 1878. Mandevilla Mandoni

Britton, Bull. Torrey Bot. Club 25: 496. 1898. **Mandevilla Bangii** Rusby, Bull. N. Y. Bot. Gard. 4: 315. 1907. Suffruticose TRAIT lianas TRAIT ; stems terete, relatively stout, glabrous

TRAIT or essentially so; leaves opposite TRAIT , petiolate, ovate TRAIT to broadly (89)



	Taxon	Traits
0	<b>Mandevilla subcordata</b>	{lianas, very minutely puberulent when young, glabrous, glabrate, infundibuliform, minutely ovate, rarely obtuse , anthers , rounded, follicles relatively stout, minutely tomentulose , glabrous , puberulent , leaves opposite, obscurely emarginate, seeds , leaves, Suffruticose , truncate , broadly cordate, glandular , minutely puberulent , oblongelliptic, ovary , pale , obliquely obovate}
1	<b>Mandevilla Bridgesii</b>	{lianas, glabrous , leaves opposite, ovate , Suffruticose }

ncbi.nlm.nih.gov/nuccore/MK810076.1

Apps 3D Museum Barcoding Manuscri... Blogs Boat Building Other bookmarks

NCBI Resources How To

Nucleotide Nucleotide Advanced

GenBank Send to:

## Sternaspis cf. scutata NHMUK ANEA 2019 voucher NHMUK ANEA 2019.7274 16S ribosomal RNA gene, partial sequence; mitochondrial

GenBank: MK810076.1

[FASTA](#) [Graphics](#) [PopSet](#)

Go to:

LOCUS MK810076 441 bp DNA linear INV 11-SEP-2019

DEFINITION Sternaspis cf. scutata NHMUK ANEA 2019 voucher NHMUK ANEA 2019.7274 16S ribosomal RNA gene, partial sequence; mitochondrial.

ACCESSION MK810076

VERSION MK810076.1

KEYWORDS .

SOURCE mitochondrion Sternaspis cf. scutata NHMUK ANEA 2019

ORGANISM [Sternaspis cf. scutata NHMUK ANEA 2019](#)  
Eukaryota; Metazoa; Lophotrochozoa; Annelida; Polychaeta;  
Scolecida; Terebellida; Sternaspidae; Sternaspis.

REFERENCE 1 (bases 1 to 441)

AUTHORS Drennan,R., Wiklund,H., Rouse,G.W., Georgieva,M.N., Wu,X.,  
Kobayashi,G., Yoshino,K. and Glover,A.G.

TITLE Taxonomy and phylogeny of mud owls (Annelida: Sternaspidae),  
including a new synonymy and new records from the Southern Ocean,  
North East Atlantic Ocean and Pacific Ocean: challenges in  
morphological delimitation

JOURNAL Mar Biodivers (2019) In press

REMARK Publication Status: Available-Online prior to print

REFERENCE 2 (bases 1 to 441)

AUTHORS Drennan,R., Wiklund,H., Rouse,G.W., Georgieva,M.N., Wu,X.,  
Kobayashi,G., Yoshino,K. and Glover,A.G.

TITLE Direct Submission

JOURNAL Submitted (14-APR-2019) Life Sciences Department, Natural History  
Museum London, Cromwell Rd., London SW7 5BD, UK

data.nhm.ac.uk/dataset/56e711e6-c847-4f99-915...

Apps 3D Museum Barcoding Manuscri... Blogs Other bookmarks

BETA NATURAL HISTORY MUSEUM Data Portal HOME DATA ABOUT Log in Contact

Home > Data > Collection specimens > Specimens > 2019.7274

Normal view Darwin Core view Back to resource Download Contact record curator

## 2019.7274

GBIF quality indicators: UNKNOWN

**Classification**

Scientific name: Sternaspis cf. scutata (Ranzani, 1817)  
Author: (Ranzani, 1817)  
Kingdom: Animalia  
Phylum: Annelida  
Class: Polychaeta  
Order: Terebellida  
Family: Sternaspidae  
Genus: Sternaspis  
Higher classification: Animalia; Annelida; Polychaeta; Terebellida; Sternaspidae

**Location**

Locality: Mayflower Marina  
State province: England  
Country: United Kingdom  
Continent: Europe  
Higher geography: Europe; United Kingdom; England; Plymouth  
Decimal latitude: 50.3630278  
Decimal longitude: -4.1672111  
Verbatim latitude: 50 21 46.900 N  
Verbatim longitude: 004 10 01.960 W  
Minimum depth(m): 18

**Specimen**

Catalogue number: 2019.7274  
Collection code: ZOO (Zoology)  
Sub department: Annelida  
Other catalog numbers: NHMUK:ecatalogue:8889822  
Kind of object: wet  
Preservative: 80% ethanol  
Donor name: Dr Adrian Glover  
Individual count: 0  
Sex: unknown  
Life stage: juvenile

Regan Drennan  
RV Callista  
Van Veen Grab Sample  
26 / 6 / 2017

Regan Drennan  
Names Types Filed As  
Sternaspis cf. scutata (Ranzani, 1817)



## What else can you use to automate workflows?

ML not only way to automate your workflows with software

Many options for (semi) automation:

- File processing
- Data capture (OCR)
- Bulk record creation / import
- Quality control / assurance (images & data)
- Metrics / reporting



**(Come find me later for Trainer Code!)**

## Resources and References

- Machine Learning by Vasily Zubarev ([https://vas3k.com/blog/machine\\_learning/](https://vas3k.com/blog/machine_learning/))
- Inselect <https://naturalhistorymuseum.github.io/inselect/>
- Short article with code on dominant colour analysis  
<https://buzzrobot.com/dominant-colors-in-an-image-using-k-means-clustering-3c7af4622036>
- Carranza-Rojas J, Goeau H, Bonnet P, Mata-Montero E, Joly A (2017) Going deeper in the automated identification of Herbarium specimens. BMC Evolutionary Biology 17:181. doi: 10.1186/s12862-017-1014-z
- Miroslav Valan, Karoly Makonyi, Atsuto Maki, Dominik Vondráček, Fredrik Ronquist, Automated Taxonomic Identification of Insects with Expert-Level Accuracy Using Effective Feature Transfer from Convolutional Networks, Systematic Biology, , syz014, <https://doi.org/10.1093/sysbio/syz014>

# Artificial Intelligence

Machine Learning

Neural Networks

Deep Learning

Useful  
Digitisation  
Automation  
Applications