Insect Soups: Enabling access to an underutilised and valuable resource?

Paul Flemons

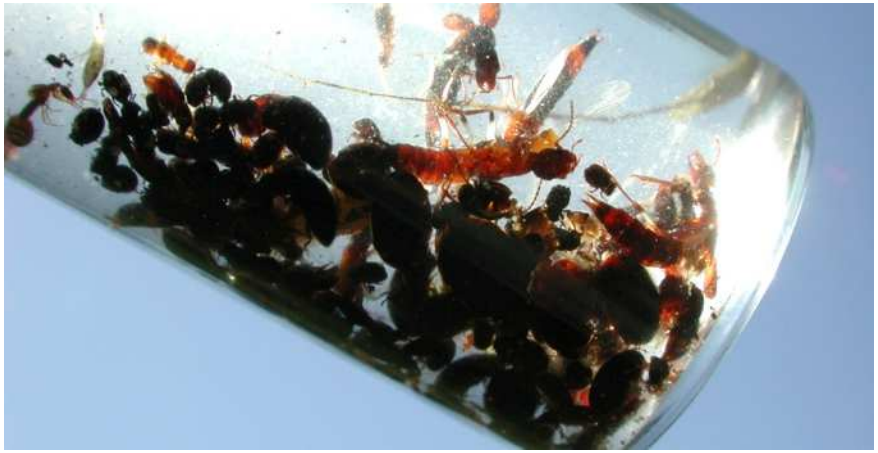
November 2013

nature culture **discover**

Australian **museum**

What is Insect Soup?

- Samples of insects collected through bulk/indiscriminate sampling techniques:
  - Malaise traps, pitfall traps, flight intercept traps, berlese funnels, etc.
- Benefits:
  - Large samples of broad range of taxa from one point over time.
  - Targets specimens not easily captured by other means (active searches, light traps etc.).
  - Passive capture, i.e. can be set and left.



What is Insect Soup?

- Storage:
  - Jars or tubes of ethanol.
  - Some sorted into ordinal groups (e.g. Coleoptera, Diptera etc.).
  - Many are mixed lots.

- What we always know:
  - Where and when the sample was collected.
  - Who collected the sample.
  - Where the vial is physically located in the collection.

What is Insect Soup?

- What we often don't know:
  - What's in them.
    - Taxa?
    - No. of specimens?
  - What was in them, and is now not in them (i.e. specimens removed from sample).

Without this information, it makes it extremely difficult for researchers to evaluate whether or not a particular bulk sample is useful.

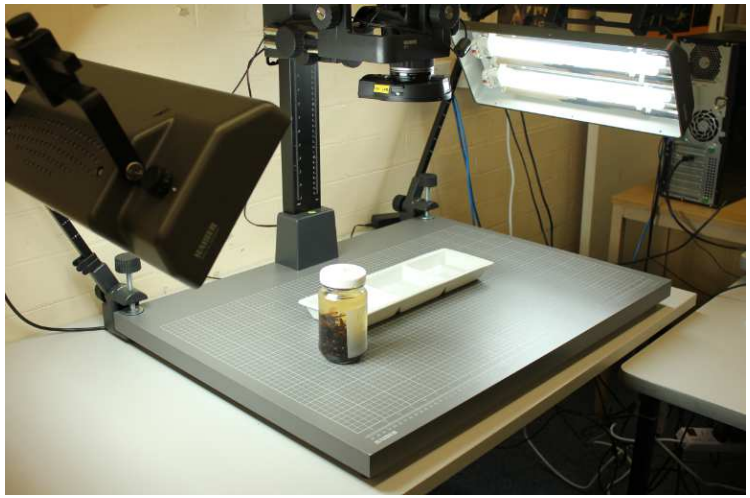
Why is Insect Soup useful?

- Insect soup is a source of real and potential data for:
  - discovering new species,
  - extensions of species distributions /ranges
    - eg pest species
  - statistics on ecological patterns and measures of biodiversity richness and diversity,
  - record of taxa used in eco-genomics analyses
  - time series analysis

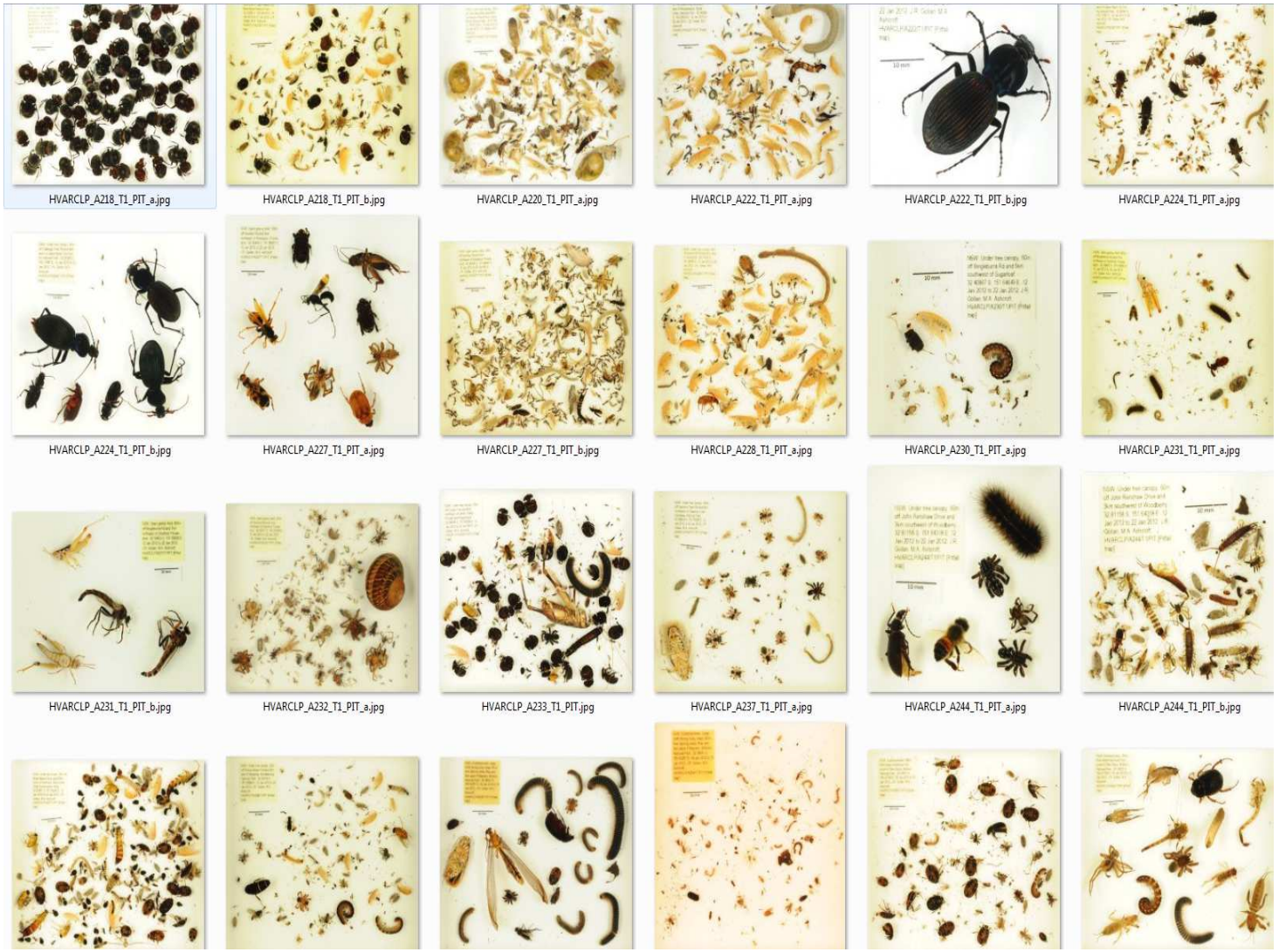
- This data has application in :
  - land use and conservation planning
  - quarantine and pest species management
  - climate change research
  - disease vector analysis
  - etc

Making our Insect Soups Accessible and useful to Science

Step 1: Photograph Soups



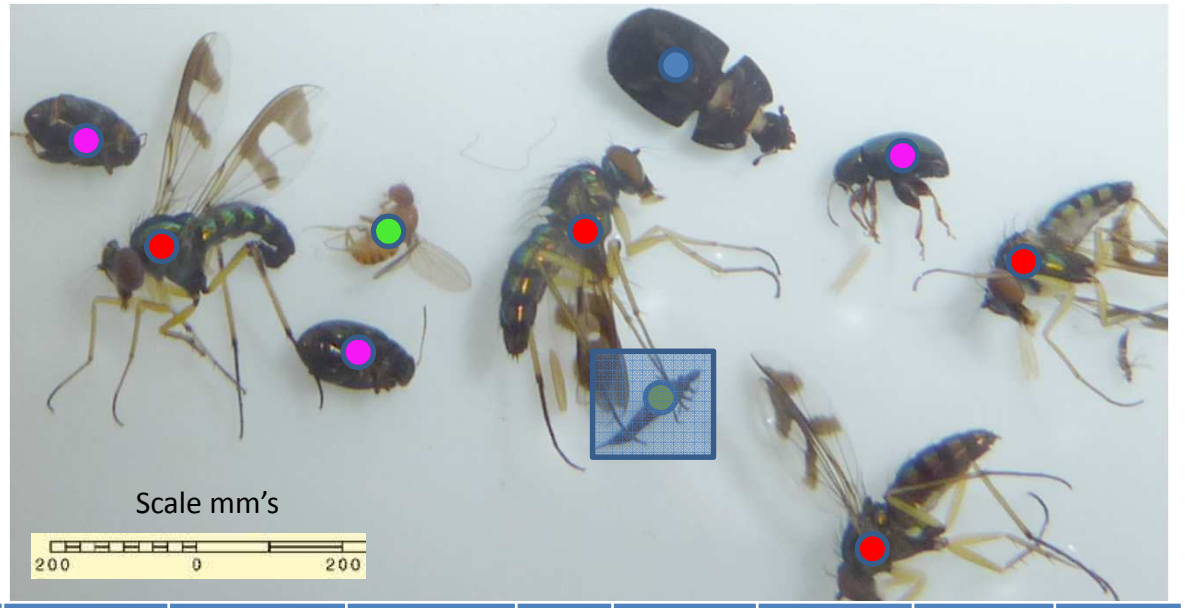
Patterns immediately obvious



Step 2: Build a website where the patterns are captured

- Group “like” insects
- Tag
- Extract image
- Count

Store in database



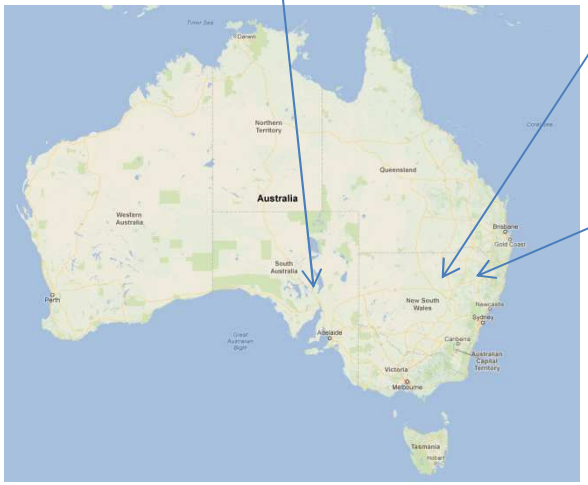
Select	Symbol	Image	Class (Key)	Order (Key)	Sub-Order (Keys for each Order)	Family (Keys for each Orders)	Genus (Keys for each Family)	Count	Body	Wings	Legs	Colour
Edit			Insecta	Coleoptera	unknown	unknown	unknown	1	2 cm Thorax divided from abdomen	unknown	unknown	Black
Edit			Insecta	Diptera	unknown	unknown	unknown	4	3cm, long hairs: 3mm	2cm, clear with brown pattern	Thin,long, 3cm	Metallic green
Edit			Insecta	Coleoptera		unknown		3				
Edit			Insecta	Diptera		unknown		1				
Edit			unknown	unknown		unknown		1				
	Add Species											

Step 3: Build website tools for researchers to explore the images: for new species, extensions of species distributions and the statistics for ecological patterns, and taxa used in eco-genomics analyses.

For images similar to:

Tags:
Wasp
Orange
Four wings
Six legs
Black and orange

Latitude:
Longitude:
Insects = 64/18
Coleoptera (beetles) = 8/3
Hemiptera (bugs) = 12/2
Hymenoptera (ants/wasps)



Result:

Cost effective way of researching insect richness and diversity by making the most of our existing collections

Examples of science making use of data on insect abundance and diversity

Nature **448**, 692-695 (9 August 2007) | doi:10.1038/nature06021; Received 20 April 2007; Accepted 14 June 2007

Low beta diversity of herbivorous insects in tropical forests

Vojtech Novotny[1], Scott E. Miller[2], Jiri Hulcr[1,3], Richard A. I. Drew[4], Yves Basset[5], Milan Janda[1], Gregory P. [illegible], John Auga[8], Brus Isua[8], Kenneth [illegible], Martin Mogia[8] & George D. Weible

Journal of Insect Conservation
1997, Volume 1, Issue 1, pp 25-42

Diversity, disturbance, and sustainable use of Neotropical forests: insects as indicators for conservation monitoring

Keith S. Brown

Blair, Robert B. 1999. BIRDS AND BUTTERFLIES ALONG AN URBAN GRADIENT: SURROGATE TAXA FOR ASSESSING BIODIVERSITY? Ecological Applications 9:164–170. http://dx.doi.org/10.1890/1051-0761(1999)009[0164:BABAAU]2.0.CO;2

Articles

BIRDS AND BUTTERFLIES ALONG AN URBAN GRADIENT: SURROGATE TAXA FOR ASSESSING BIODIVERSITY?

Robert B. Blair[1,2]

Ann. Zool. Fennici 31:71–81
Helsinki 31 January 1994
ISSN 0003-455X
© 1994 Finnish Zoological Publishing Board

PROCEEDINGS OF THE ROYAL SOCIETY B
ome | Current issue | Past issues | Submit | Subscribe | Alerts

LETTERS TO NATURE

Insect species diversity, abundance and body size relationships

Evan Siemann, David Tilman & John Haarstad

Department of Ecology, Evolution and Behavior, University of Minnesota, St Paul, Minnesota 55108, USA

Hot-spots of insect diversity in northern Europe

Conservation of Insect Diversity: a Habitat Approach

Jennifer B. Hughes, Gretchen C. Daily, Paul R. Ehrlich

Article first published online: 7 JUL 2008
DOI: 10.1111/j.1523-1739.2000.99187.x

Issue

Conservation Biology
Volume 14, Issue 6, pages 1788–1797, December 2000

CrossMark

Self-similar patterns of nature: insect diversity at local to global scales

Ecological Applications, 6(2), 1996, pp. 594–607
© 1996 by the Ecological Society of America

DESIGNING A COST-EFFECTIVE INVERTEBRATE SURVEY: A TEST OF METHODS FOR RAPID ASSESSMENT OF BIODIVERSITY[1]

IAN OLIVER AND ANDREW J. BEATTIE
School of Biological Sciences, Macquarie University, Sydney 2109 Australia

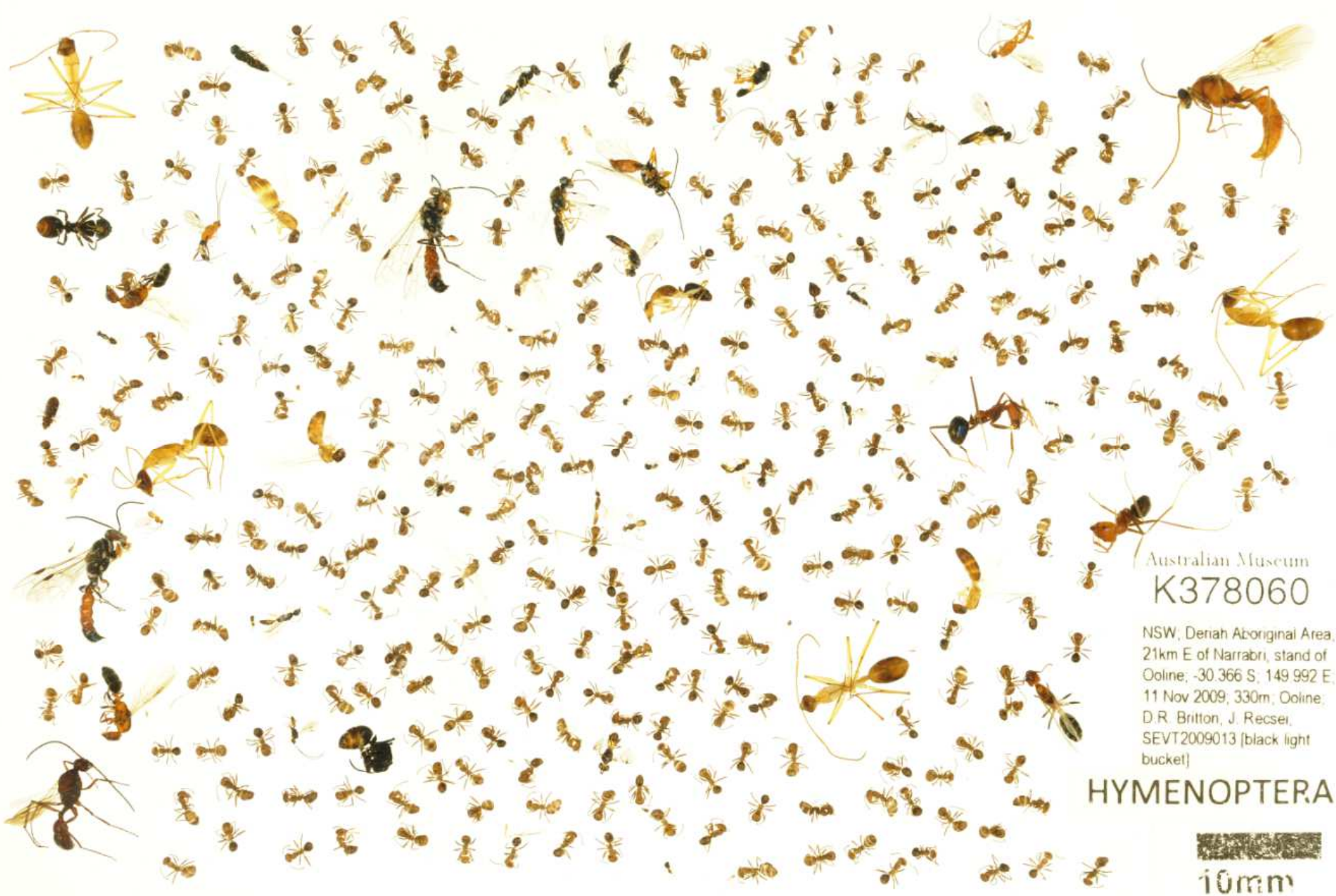
Forest Ecology and Management
Volume 113, Issue 1, 4 January 1999, Pages 11–21

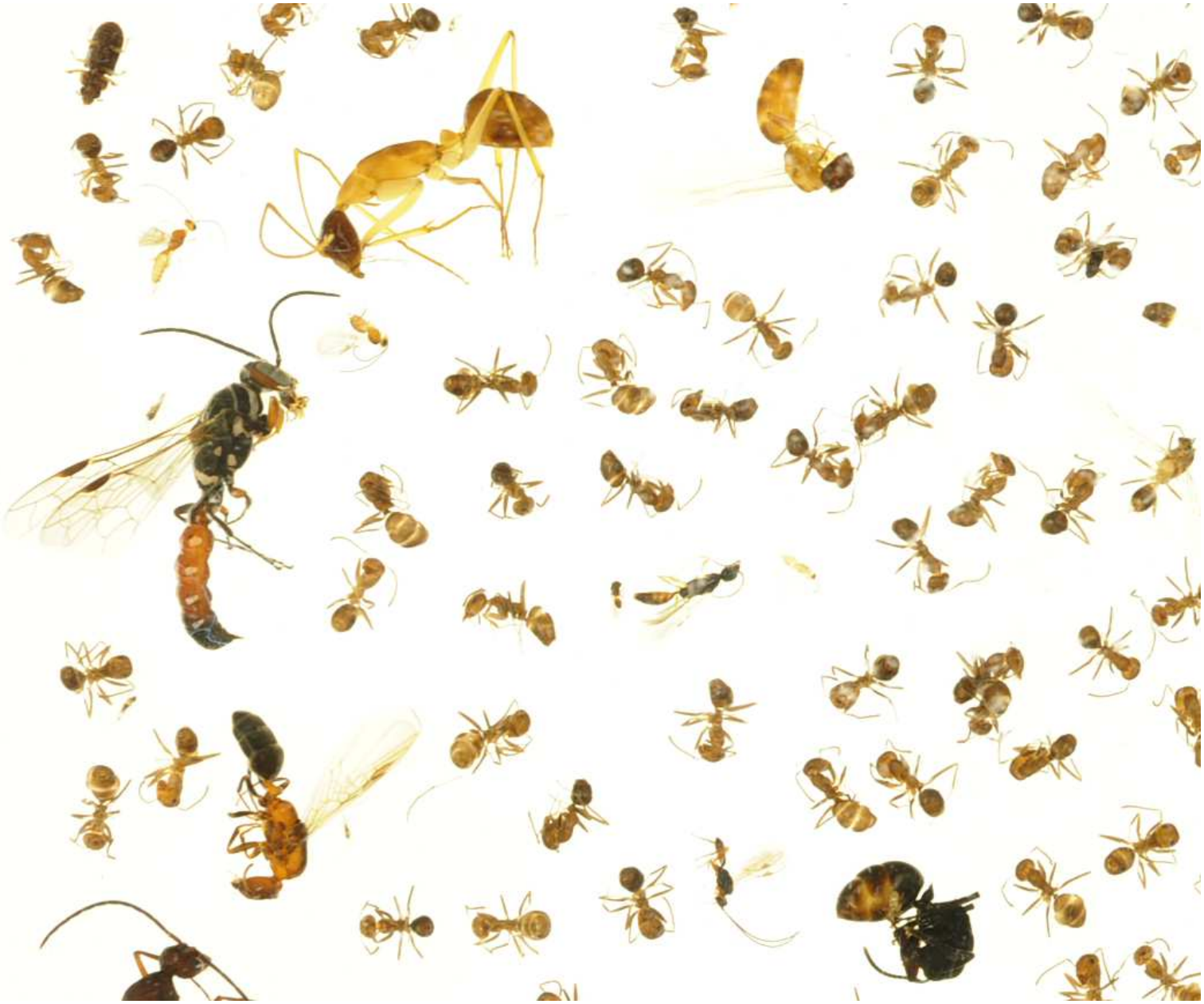
Relationships between insect diversity and habitat characteristics in plantation forests

10mm
MT. GLORIOUS QLD
13.ii.1998
TRAP 2
16

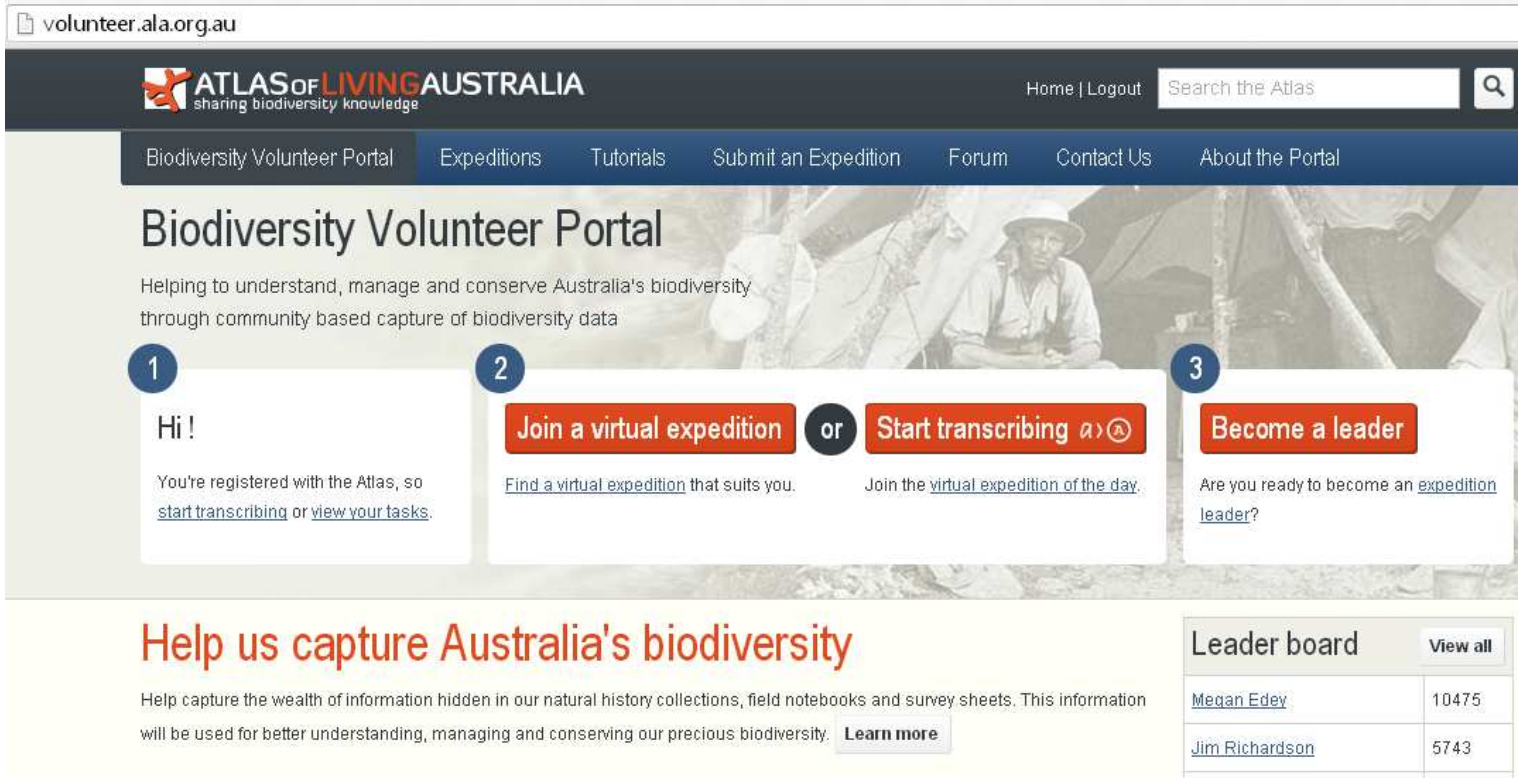
10mm
MT. GLORIOUS QLD
13.ii.1998 TRAP 2
COLEOPTERA

NSW; Deriah Aboriginal Area, 20.9km E of Narrabri, stand of Ooline; -30.368 S; 149.991 E; 11 to 19 Nov 2009; 340m; D.R. Britton, J. Recsei; SEVT2009014 [Malaise]
Australian Museum
K378075
DIPTERA
10mm


Australian Museum
K378060
NSW, Deriah Aboriginal Area, 21km E of Narrabri, stand of Ooline; -30.366 S, 149.992 E; 11 Nov 2009; 330m; Ooline; D.R. Britton, J. Recsei; SEVT2009013 [black light bucket]
HYMENOPTERA
10mm



Using Crowdsourcing: Manual segmentation and classification



Use Case 1: Non-expert

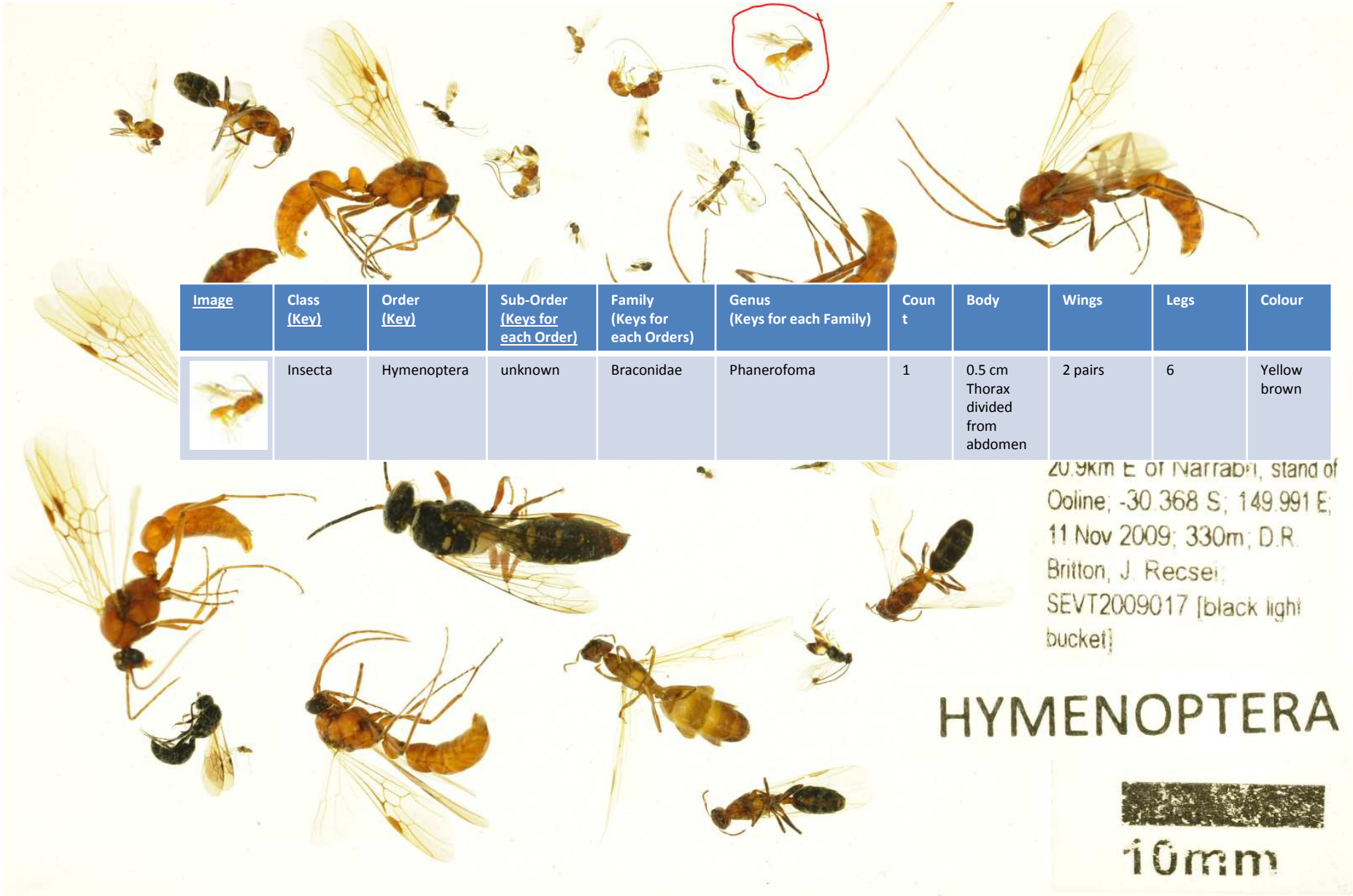


Image	Class (Key)	Order (Key)	Sub-Order (Keys for each Order)	Family (Keys for each Orders)	Genus (Keys for each Family)	Count	Body	Wings	Legs	Colour
	Insecta	Hymenoptera	unknown	Braconidae	Phanerofoma	1	0.5 cm Thorax divided from abdomen	2 pairs	6	Yellow brown

Australian Museum

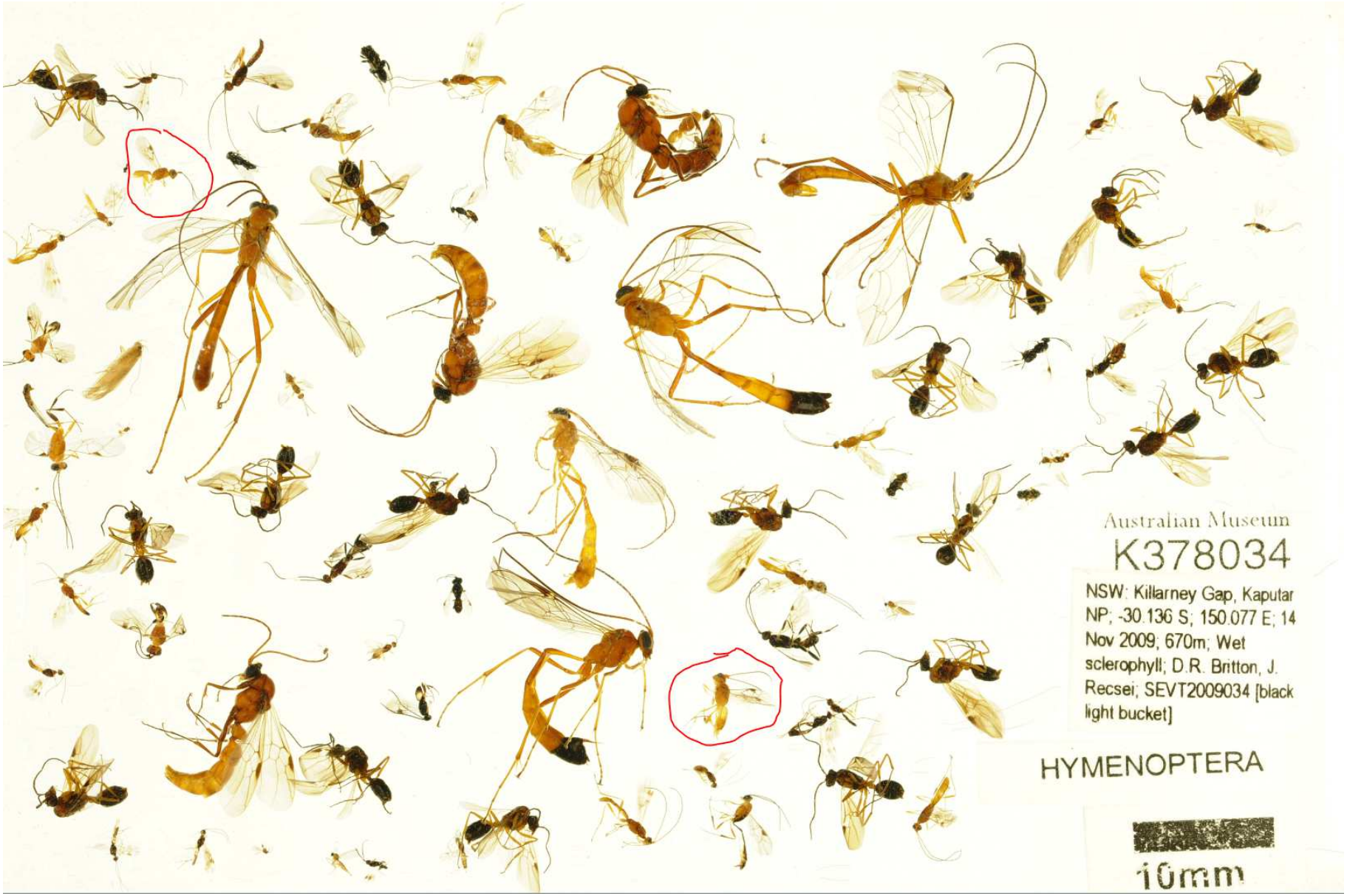
K378033

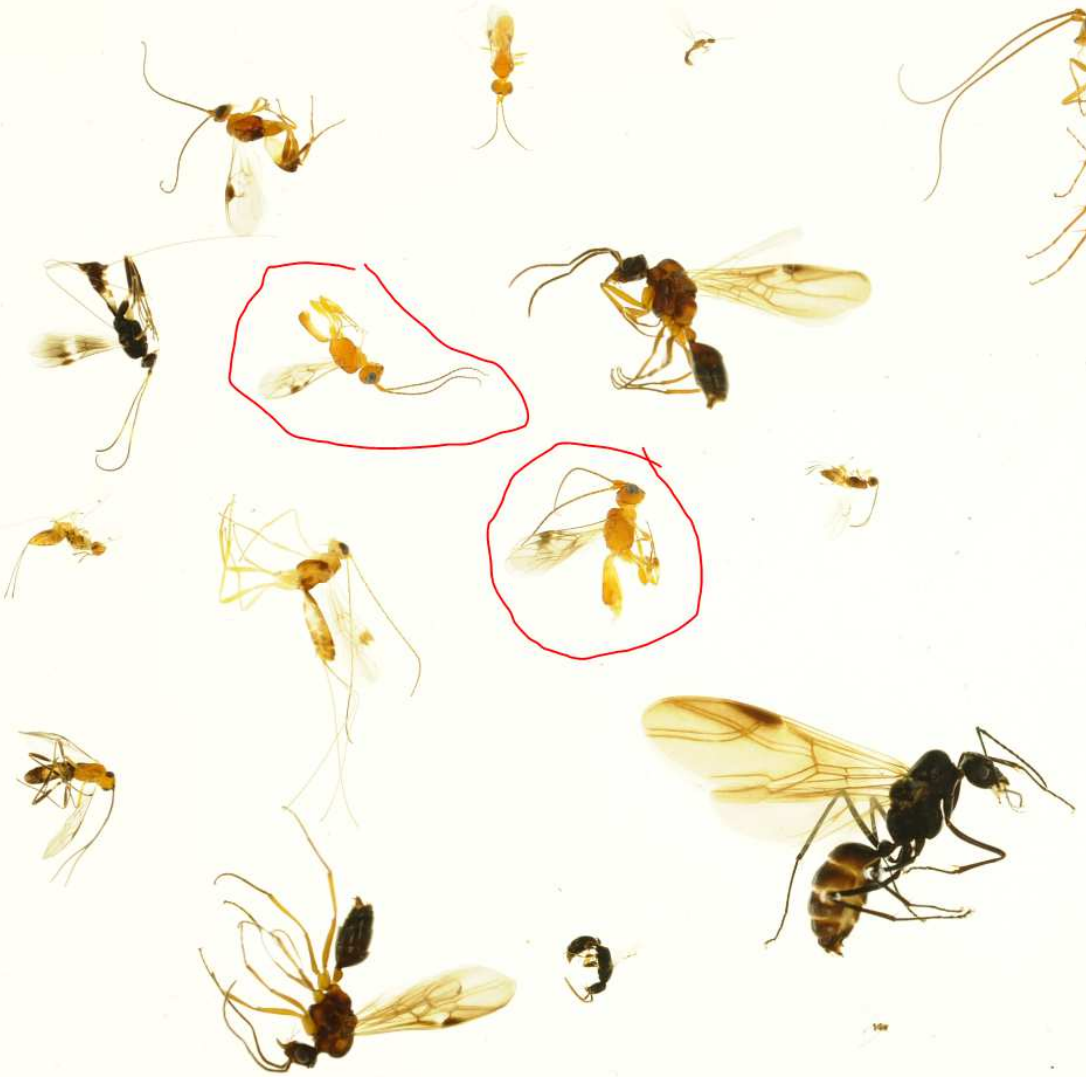
NSW; Travelling Stock Reserve, 46.2km NE of Narrabri; -30.002 S; 150.042 E; 13 Nov 2009; 455m; Semi-evergreen vine thicket; D.R. Britton, J. Recsei; SEVT2009028 [black light bucket]

HYMENOPTERA

[redacted]

10mm


Australian Museum
K378034
NSW: Killarney Gap, Kaputar NP; -30.136 S; 150.077 E; 14 Nov 2009; 670m; Wet sclerophyll; D.R. Britton, J. Recsei; SEVT2009034 [black light bucket]
HYMENOPTERA
10mm



Australian Museum

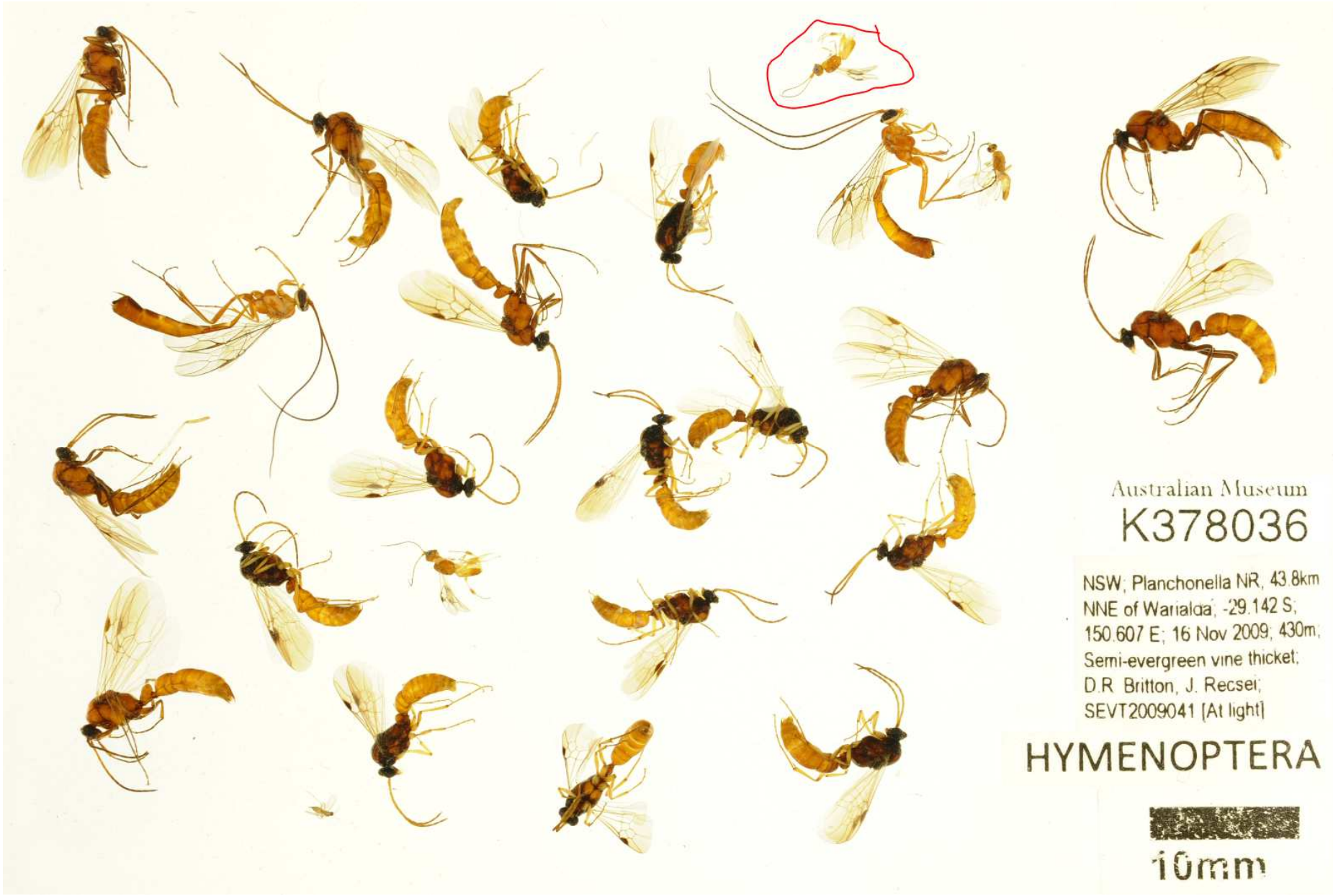
K378035

NSW; Killarney Gap, Kaputar NP; -30 144 S; 150.060 E; 14 Nov 2009; 615m; grassy sclerophyll forest; D.R. Britton, J. Recsei; SEVT2009037 [black light bucket]

HYMENOPTERA

[redacted]

10mm



Australian Museum

K378036

NSW; Planchonella NR, 43.8km NNE of Warialda; -29.142 S; 150.607 E; 16 Nov 2009; 430m; Semi-evergreen vine thicket; D.R Britton, J. Recsei; SEVT2009041 [At light]

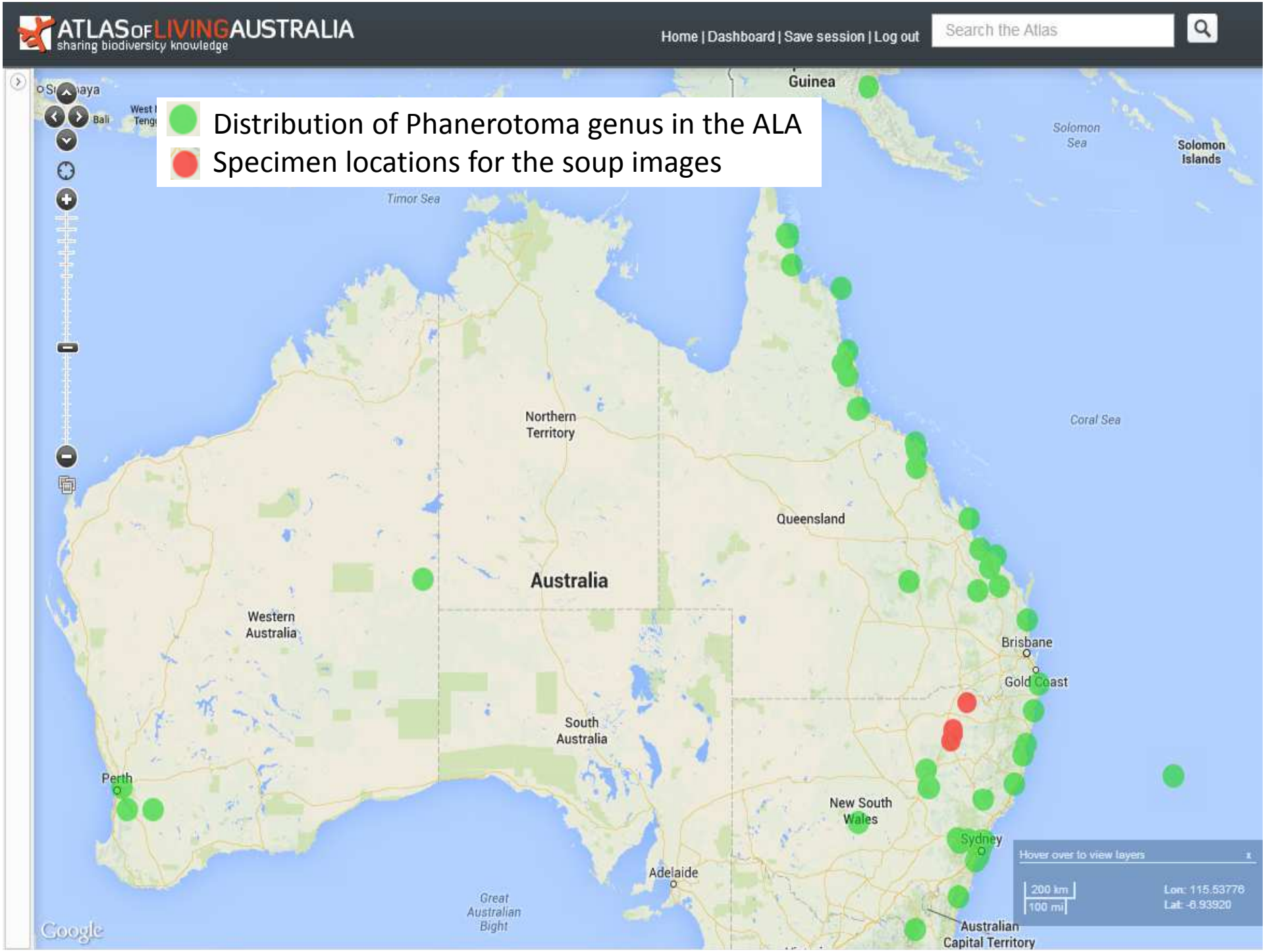
HYMENOPTERA

[redacted]

10mm

Tags

Yellow
Dark eyes
wing venation
Long antennae
Thickened wing venation
Yellow legs
0.5 to 1cm
Hymenoptera
Geographic location


ATLAS OF LIVING AUSTRALIA
sharing biodiversity knowledge
Home | Dashboard | Save session | Log out
Search the Atlas
Distribution of Phanerotoma genus in the ALA
Specimen locations for the soup images
Guinea
Solomon Sea
Solomon Islands
Timor Sea
Coral Sea
Northern Territory
Queensland
Australia
Western Australia
South Australia
New South Wales
Brisbane
Gold Coast
Sydney
Perth
Adelaide
Great Australian Bight
Australian Capital Territory
Hover over to view layers
200 km
100 mi
Lon: 115.53776
Lat: -8.93920
Google

ATLAS OF LIVING AUSTRALIA
sharing biodiversity knowledge

My profile Log out Search the Atlas

Species Locations Collections Mapping & analysis Data sets Blogs Get involved About the Atlas

Advanced search

Quick search

Occurrence search results

5 results for **Data resource: InsectSoupsWithImages**

Refine results

Sub Family
Cheloninae (5)

Scientific name
Phanerotoma (5)

Family
BRACONIDAE (5)

Lifeform
Animals (5)
Arthropods (5)
Insects (5)
choose more...

Local Gov. Area
Gwydir (A) (2)
Narrabri (A) (2)
Moree Plains (A) (1)
choose more...

State/Territory
New South Wales (5)

Records Map Charts Species images Record images

Downloads Alerts per page: 20 sort: Date added order: Descending

Genus: *Phanerotoma* State: **NSW**
Catalog number: :K378033 View record

Genus: *Phanerotoma* State: **NSW**
Catalog number: :K378036 View record

Genus: *Phanerotoma* State: **NSW**
Catalog number: :K378034 View record

Genus: *Phanerotoma* State: **NSW**
Catalog number: :K378032 View record

Genus: *Phanerotoma* State: **NSW**
Catalog number: :K378035 View record

1

Advanced search

Occurrence search results

Quick search

5 results for **Data resource: InsectSoupsWithImages**

Refine results

Sub Family
Cheloninae (5)

Scientific name
Phanerotoma (5)

Family
BRACONIDAE (5)

Lifeform
Animals (5)
Arthropods (5)

Records Map Charts Species images Record images

Images from occurrence records

ATLAS OF LIVING AUSTRALIA
sharing biodiversity knowledge

My profile Log out

Search the Atlas

Species | Locations | Collections | Mapping & analysis | Data sets | Blogs | Get involved | About the Atlas

Advanced search

Quick search

Occurrence search results

5 results for **Data resource: InsectSoupsWithImages**

Records | Map | Charts | Species images | Record images

Refine results

Sub Family
Cheloninae (5)

Scientific name
Phanerotoma (5)

Family
BRACONIDAE (5)

Lifeform
Animals (5)
Arthropods (5)
Insects (5)
choose more...

Local Gov. Area
Gwydir (A) (2)
Narrabri (A) (2)
Moree Plains (A) (1)
choose more...

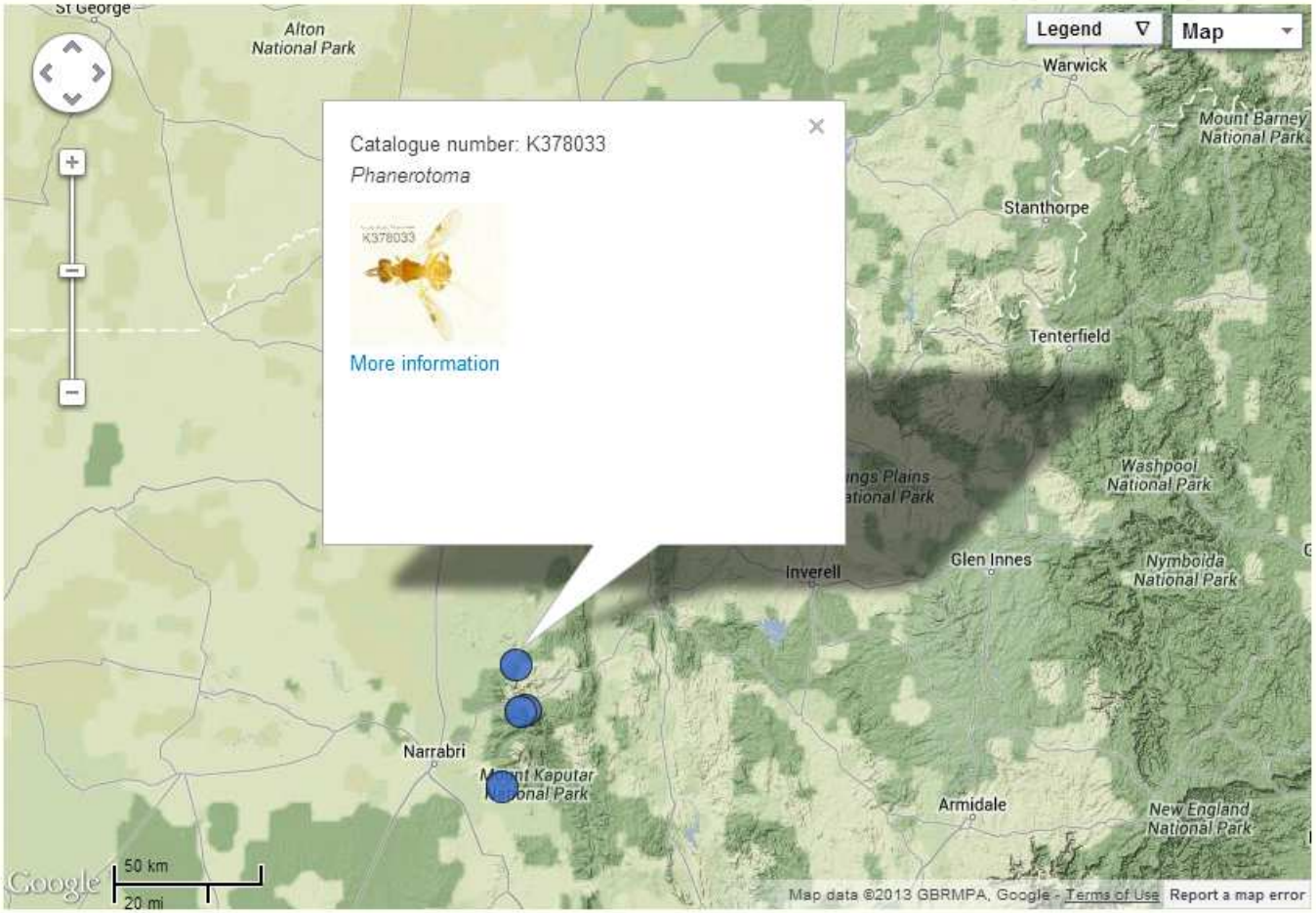
State/Territory
New South Wales (5)

Country
Australia (5)

IBRA
Nandewar (3)
Brigalow Belt South (2)
choose more...

Vegetation types - native
Eucalypt open forests (4)
Mallee woodlands and shrublands (1)
choose more...

Colour by: None Size: 10 View in spatial portal Download map



Occurrence record: 71962a09-4cfc-48ba-acae-47428def8acc

Phanerotoma

Back to search results

Data quality tests

Failed quality tests: 1
Data warnings: 2
Passed quality tests: 19
Missing record properties: 14
Test not ran (lack of data): 41

View full data quality report

Flag an issue

Original vs Processed

Images

Original image (12 KB)

Original image (12 KB)

Original image (13 KB)

Original image (12 KB)

Dataset

Data resource	
Institution	*Supplied institution code "Australian Museum"*
Catalogue number	K378033
Basis of record	Not supplied
Field number	SEVT2009028
Collector/Observer	Britton, D.R.|Recsei, J. *Supplied as "D.R. Britton, J Recsei"*
Record date	*Supplied date "13.11.2009"*
Sampling protocol	black light bucket

Taxonomy

Scientific name	*Phanerotoma*
Taxon rank	Genus
Kingdom	ANIMALIA
Phylum	ARTHROPODA
Class	INSECTA
Order	HYMENOPTERA
Family	BRACONIDAE
Genus	*Phanerotoma*
Taxonomic issues	No issues
Name match metric	Exact match The supplied name matched the name exactly.
Taxonomic issue	["noIssue"]
Name match metric	Exact match
Name parse type	Well formed scientific name according to present nomenclatural rules

Geospatial

Country	Australia
State or territory	New South Wales *Supplied as: "NSW"*
Local government area	Moree Plains
Locality	Travelling Stock Reserve, 46.2km NE of Narrabri
Biogeographic region	Nandewar
Terrestrial/Marine	Terrestrial
Latitude	-30.002
Longitude	150.042
Geodetic datum	EPSG:4326
Coordinate precision	Unknown
Coordinate uncertainty in metres	Unknown
Verbatim elevation	455.0 *Supplied as 455*

Additional properties

Sub_family	Cheloninae

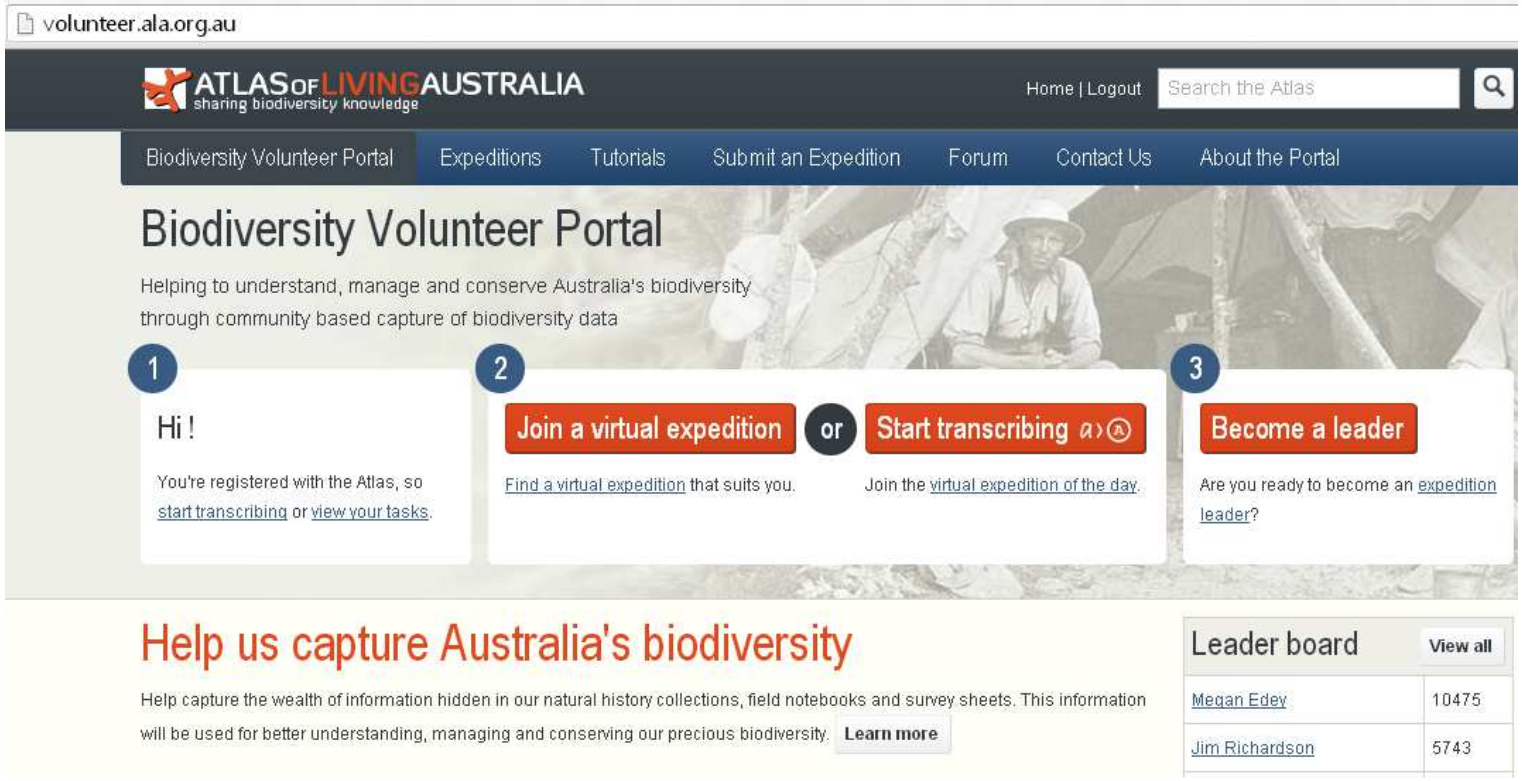
Additional political boundaries information

Area Management	
Land use	Modified pastures
Biodiversity	
IBRA 6 Regions	Nandewar
Marine	
States including coastal waters	New South Wales (including Coastal Waters)
Political	
Australian States and Territories	New South Wales
LGA Boundaries (deprecated)	Moree Plains
Local Government Areas	Moree Plains (A)
Vegetation	
Vegetation types - native	Eucalypt open forests
Vegetation types - present	Eucalyptus open forest

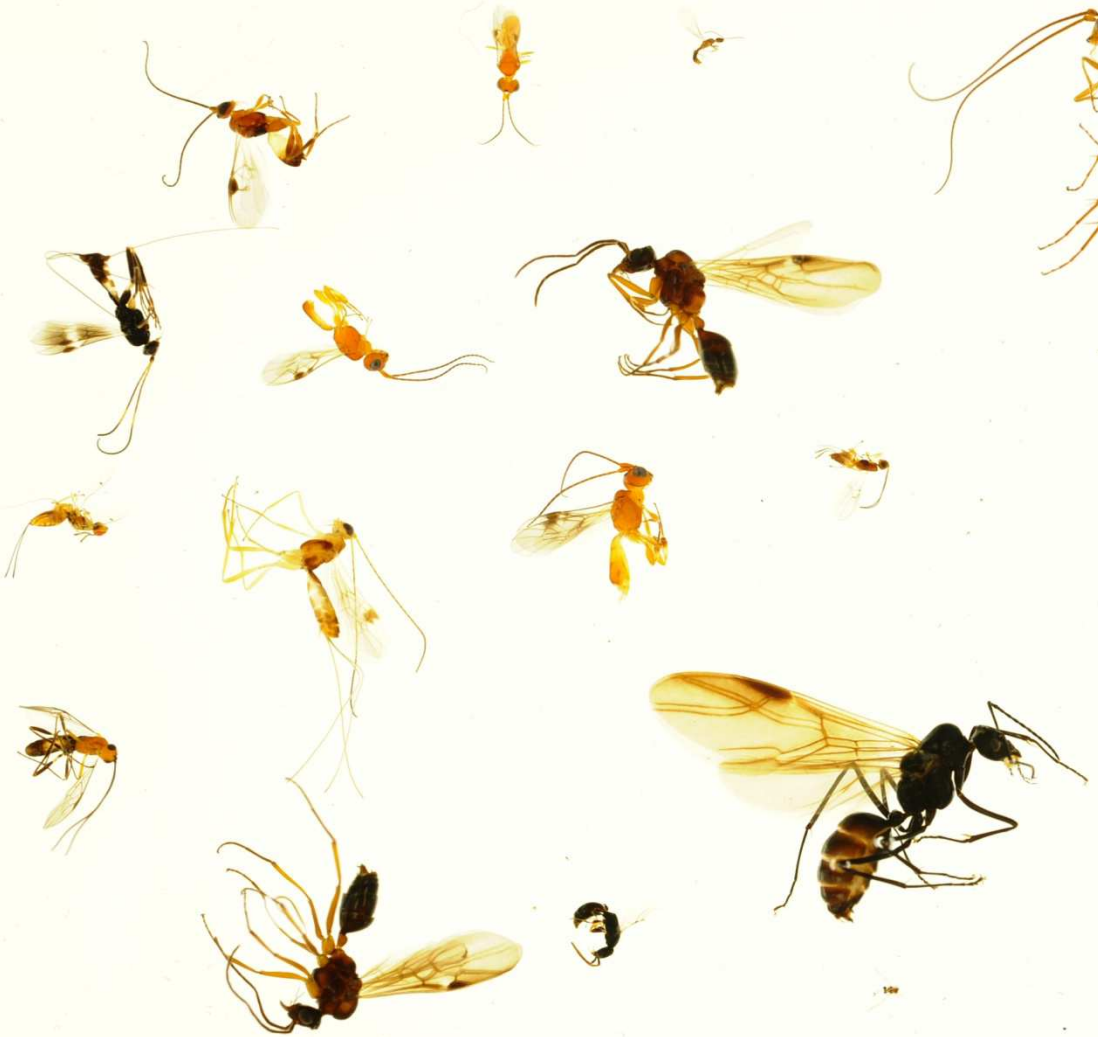
Environmental sampling for this location

Climate	
Precipitation - driest quarter (Bio17)	156.0 mm
Precipitation - seasonality (Bio15)	28.0 mm
Radiation - seasonality (Bio23)	29.0
Radiation - warmest quarter (Bio26)	23.5 MJ/m2/day
Substrate	
Moisture Index - highest quarter mean (Bio32)	0.74

Using Crowdsourcing: Manual segmentation and classification



Use Case 2: expert



Australian Museum

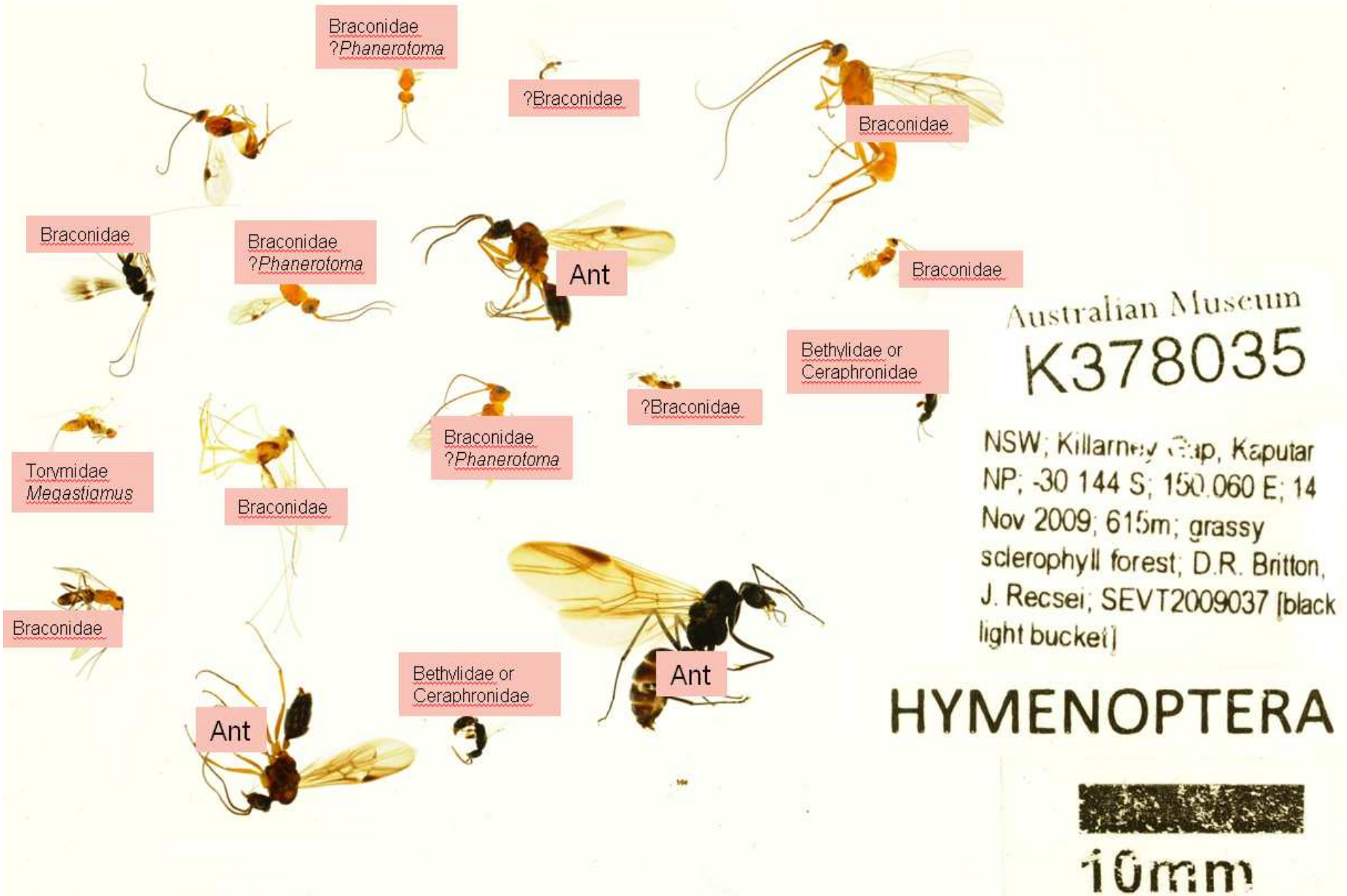
K378035

NSW; Killarney Gap, Kaputar NP; -30 144 S; 150.060 E; 14 Nov 2009; 615m; grassy sclerophyll forest; D.R. Britton, J. Recsei; SEVT2009037 [black light bucket]

HYMENOPTERA

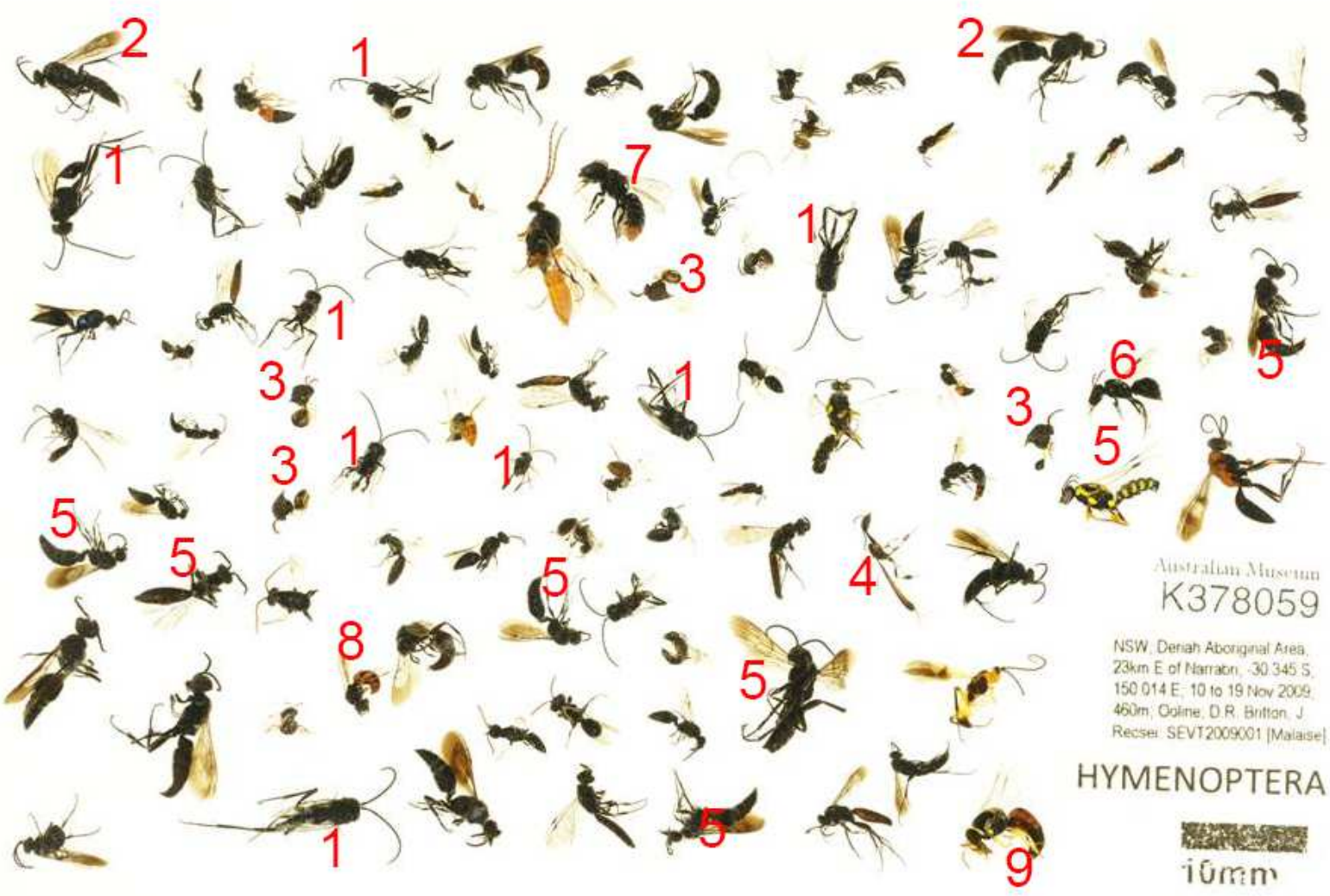
[redacted]

10mm

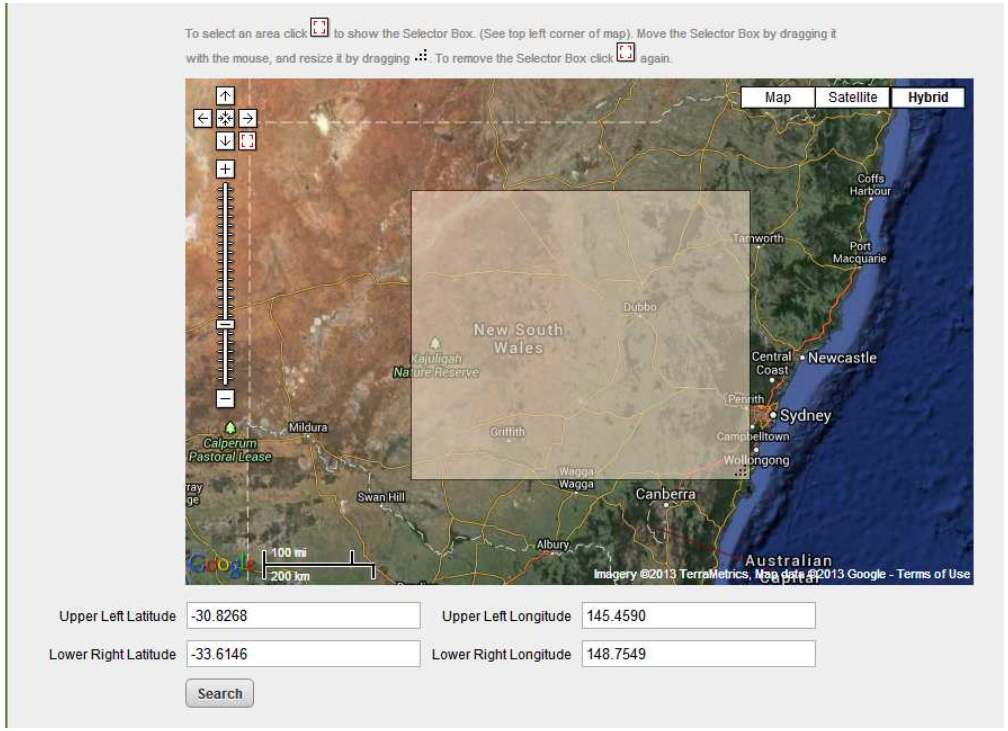

Braconidae
?Phanerotoma
?Braconidae
Braconidae
Braconidae
Braconidae
?Phanerotoma
Ant
Braconidae
Australian Museum
K378035
Bethylidae or
Ceraphronidae
?Braconidae
Braconidae
?Phanerotoma
Torymidae
Megastigmus
Braconidae
NSW; Killarney Gap, Kaputar
NP; -30 144 S; 150.060 E; 14
Nov 2009; 615m; grassy
sclerophyll forest; D.R. Britton,
J. Recsei; SEVT2009037 [black
light bucket]
Braconidae
Ant
Bethylidae or
Ceraphronidae
Ant
HYMENOPTERA
10mm

I have just spent less than 10 minutes putting family and other names (some down to species) to some in the soup.

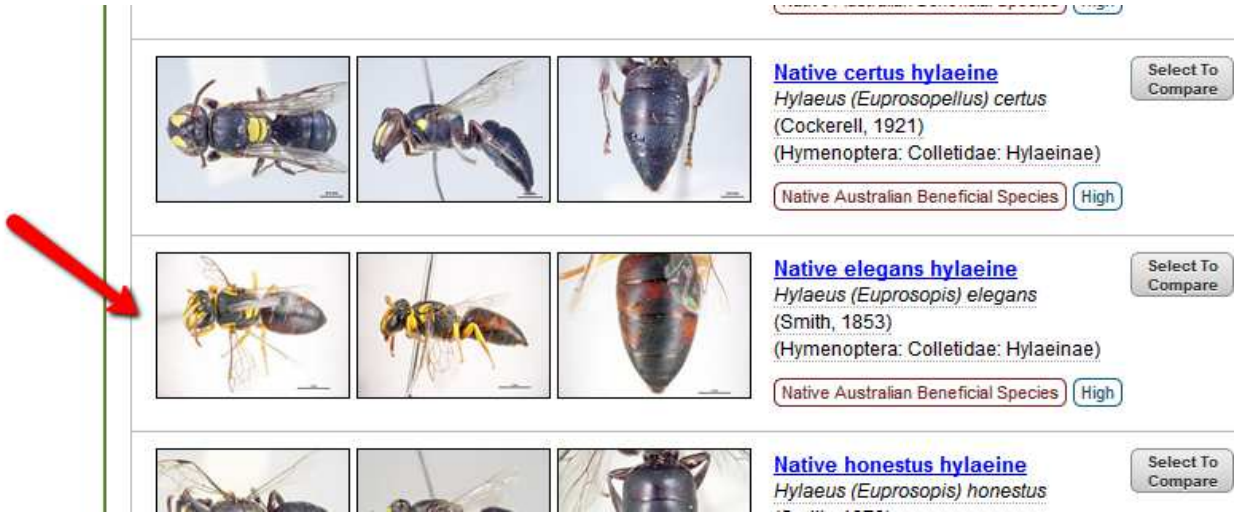
1. Evaniidae
2. Sphecidae
3. Cynipidae
4. Gastrupiidae
5. Tiphiidae
6. Apidae – *Braunsapis*
7. Megachilidae – *Megachile ferox*
8. Halictidae – Lasioglossum (Chilalictus)
9. Colletidae – *Hylaeus (Euprosopis elegans* (See below)



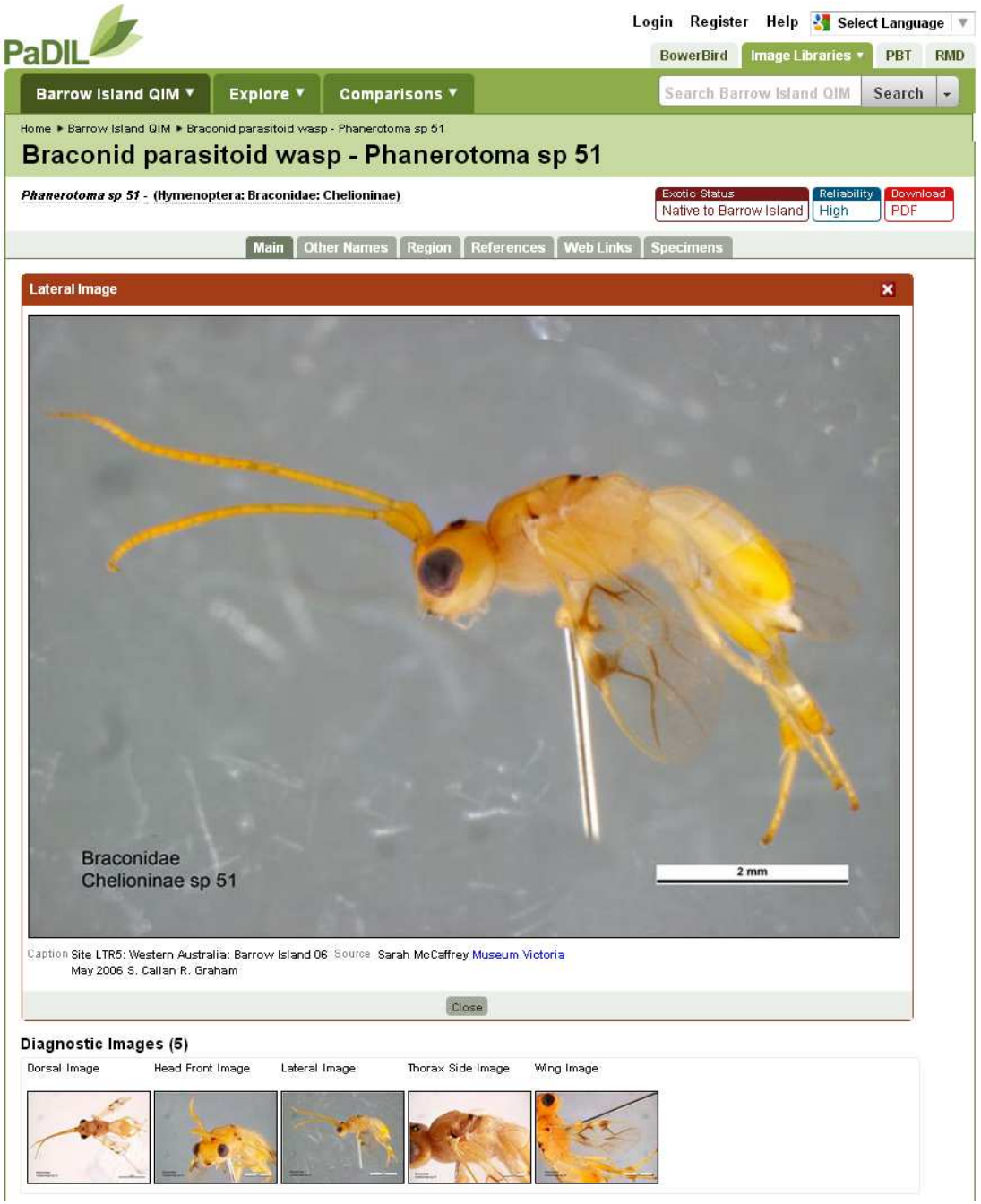
http://www.padil.gov.au/pollinators/Search?sortType=ScientificName&viewType=Details&pageSize=10&upperLeftLatitude=-30.1451&upperLeftLongitude=144.1406&lowerRightLatitude=-34.8859&lowerRightLongitude=150.7324&queryText1=Hylaeus&queryType1=all

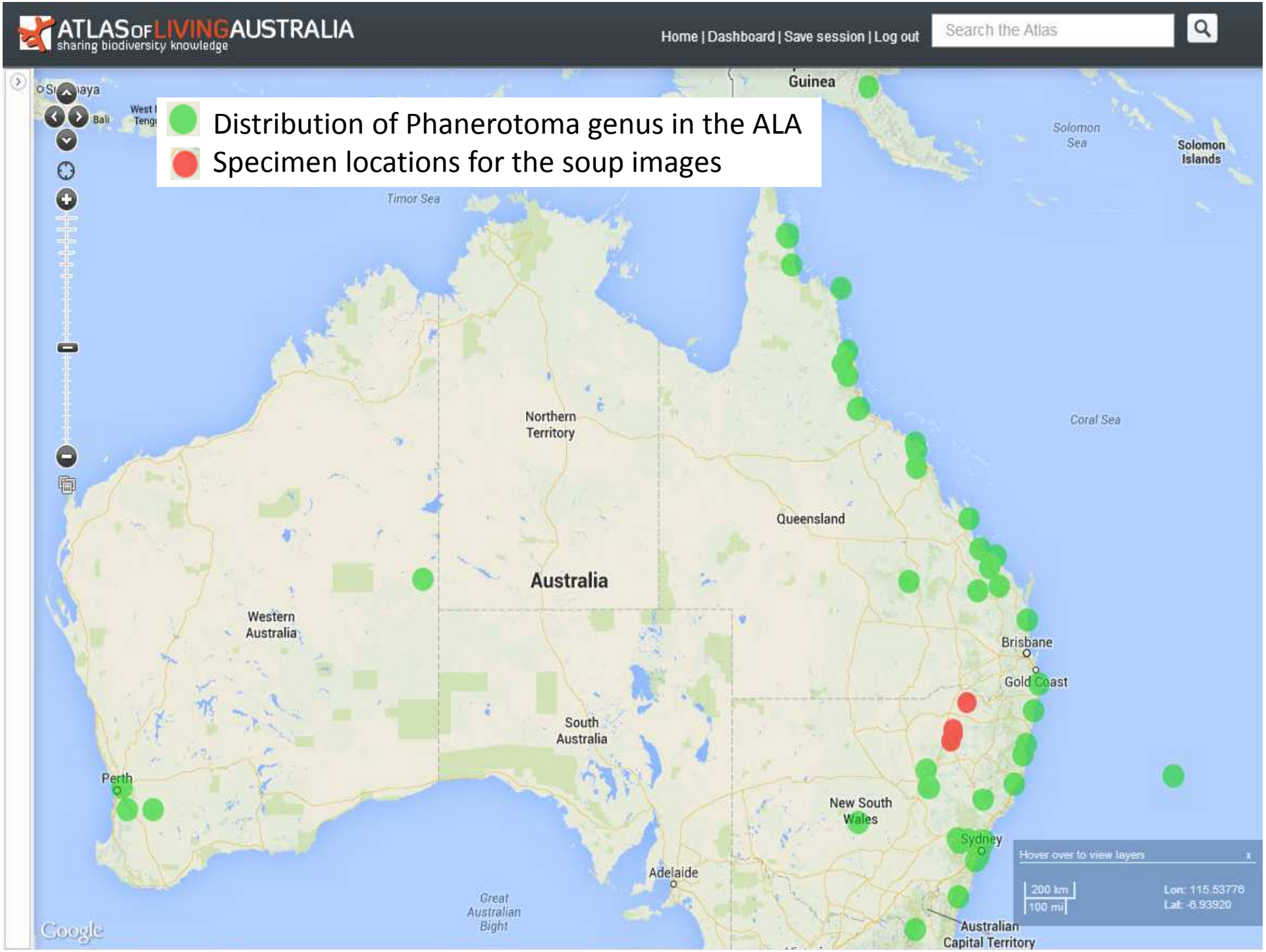


Search in PaDIL on Hylaeus in the area of interest returns a number of possible species:



http://www.padil.gov.au/barrow-island/Pest/Main/137524/12301




ATLAS OF LIVING AUSTRALIA
sharing biodiversity knowledge
Home | Dashboard | Save session | Log out
Search the Atlas
Distribution of Phanerotoma genus in the ALA
Specimen locations for the soup images
Guinea
Solomon Sea
Solomon Islands
Timor Sea
Coral Sea
Northern Territory
Queensland
Australia
Western Australia
South Australia
New South Wales
Brisbane
Gold Coast
Sydney
Perth
Adelaide
Great Australian Bight
Australian Capital Territory
Hover over to view layers
200 km
100 mi
Lon: 115.53776
Lat: -8.93920
Google

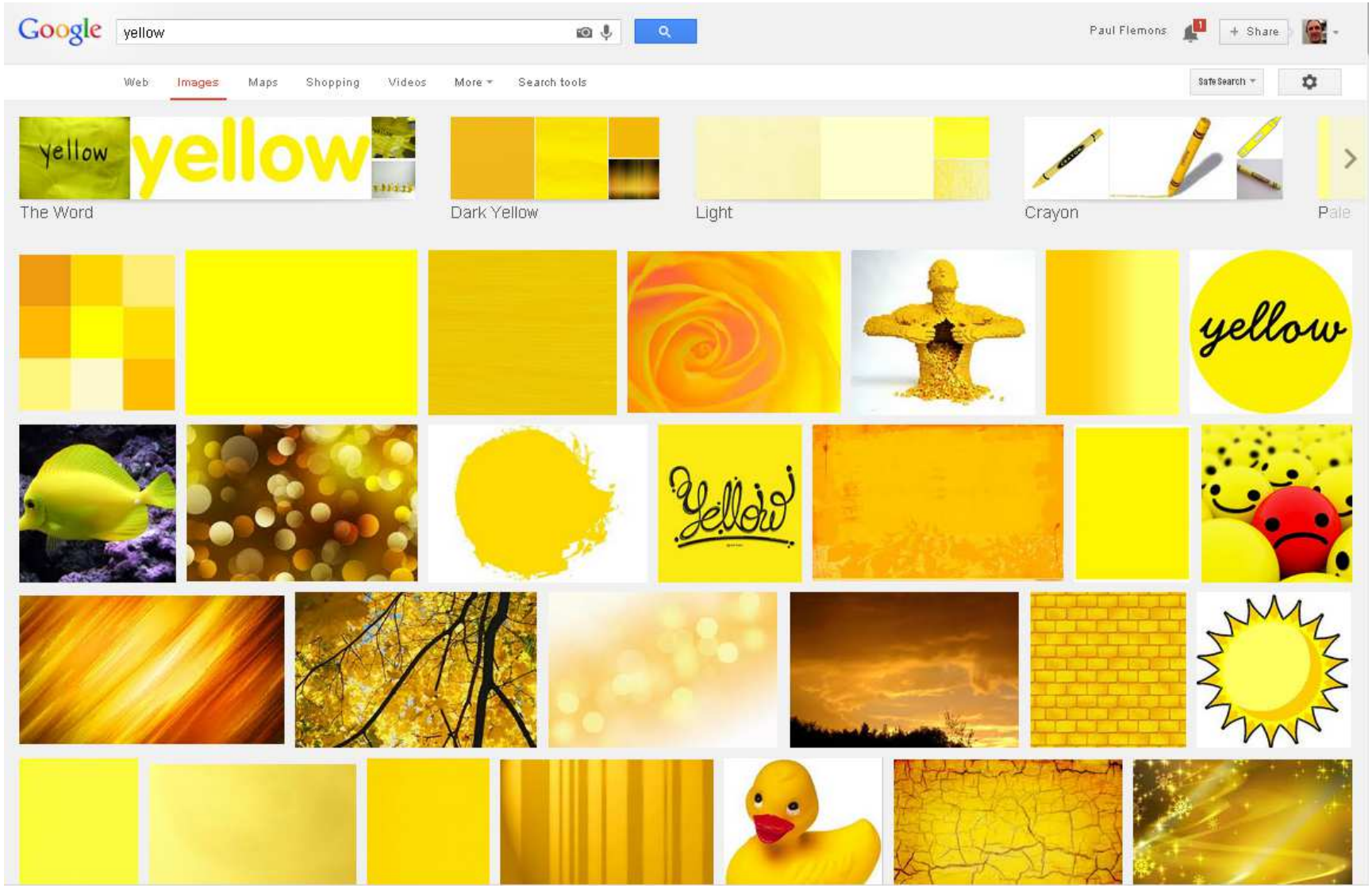
Online image library

Tags

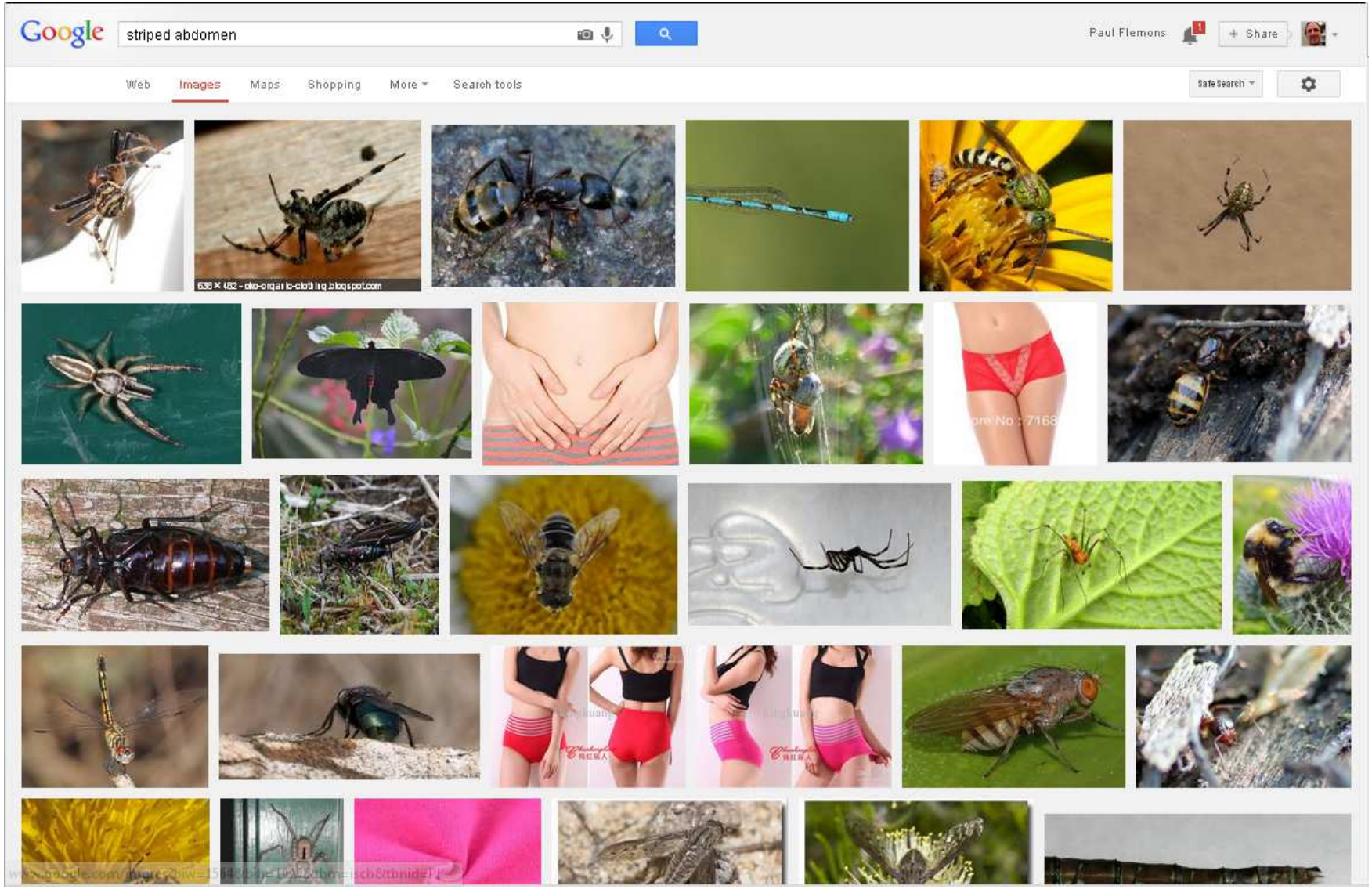
Yellow
Black
wing venation
striped abdomen
Short antennae
Thickened wing venation
Yellow legs
0.5 to 1cm
Hymenoptera
Tiphiidae

Online image library:
Image recognition and data mining:

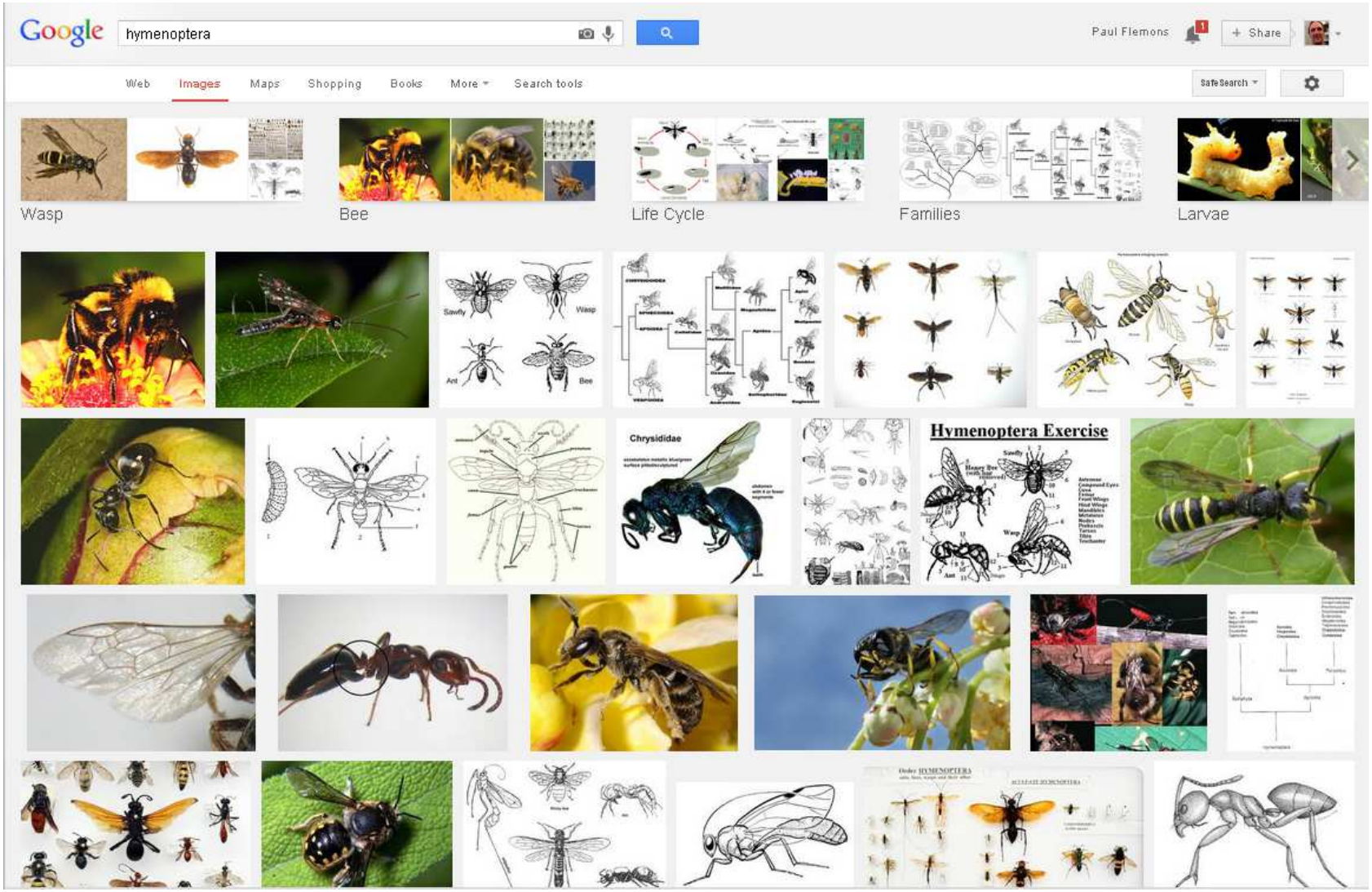
Yellow


Google
yellow
Paul Flemons
+ Share
Web
Images
Maps
Shopping
Videos
More
Search tools
SafeSearch
The Word
Dark Yellow
Light
Crayon
Pale

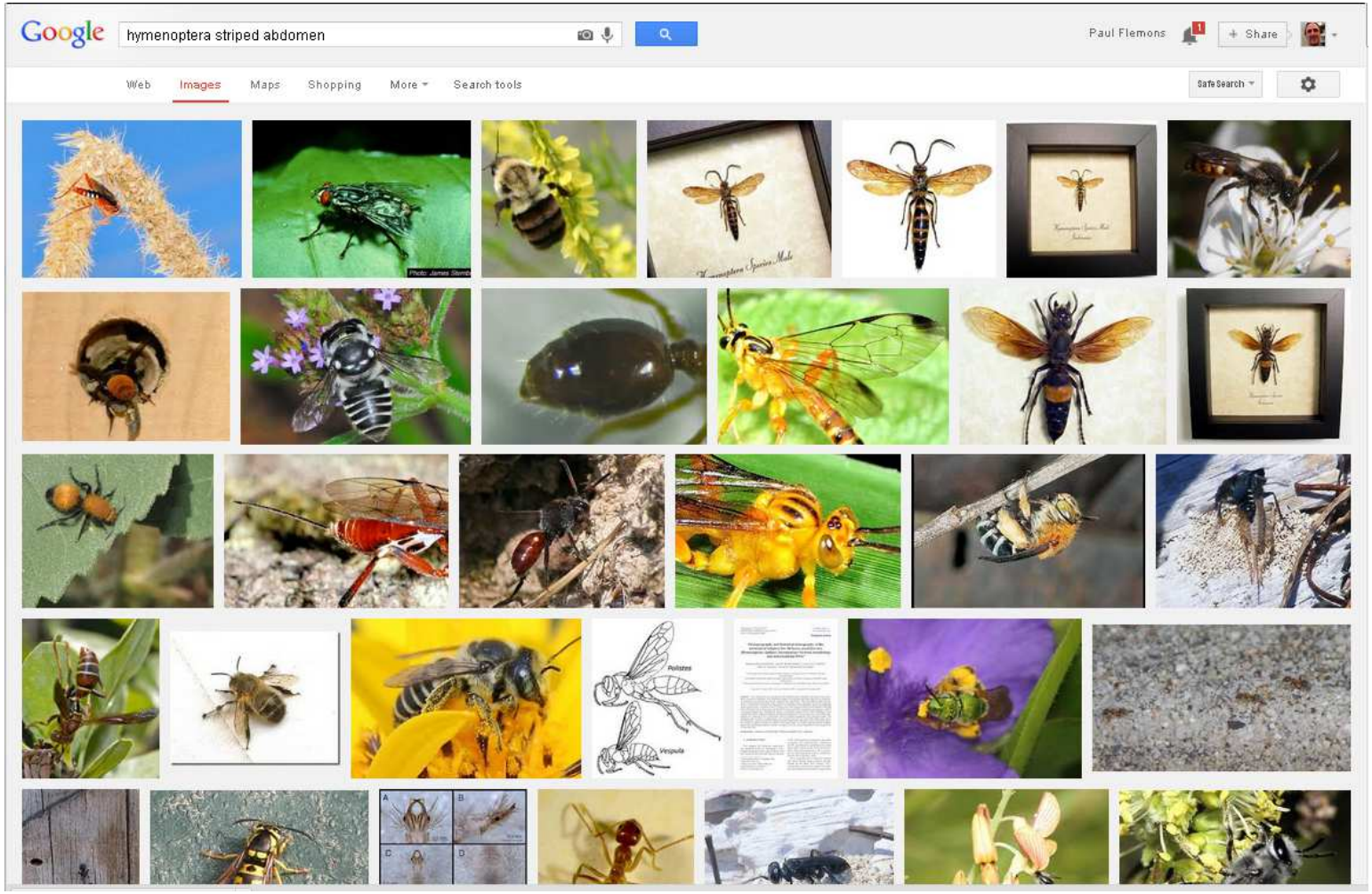
Striped abdomen



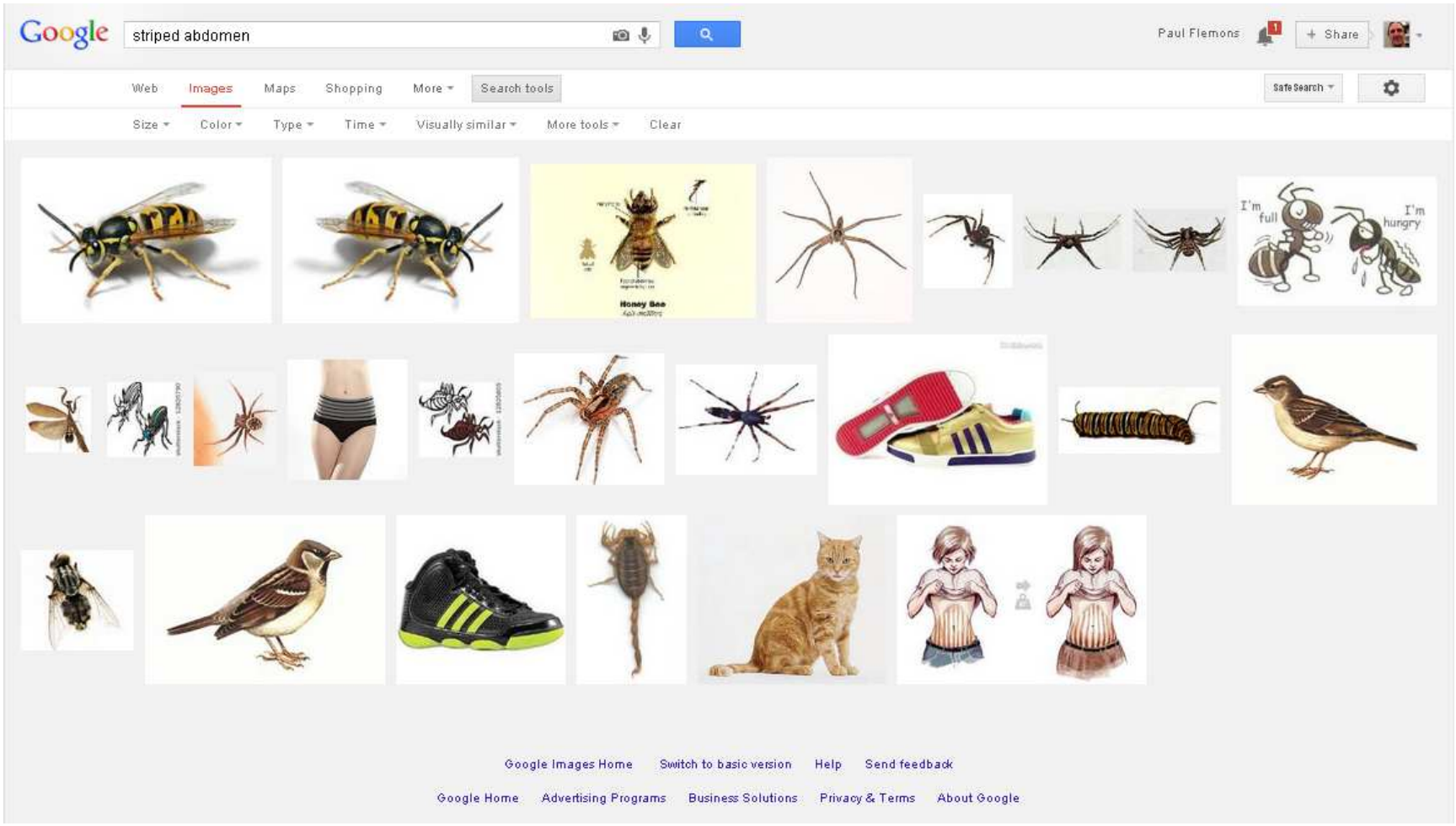
hymenoptera



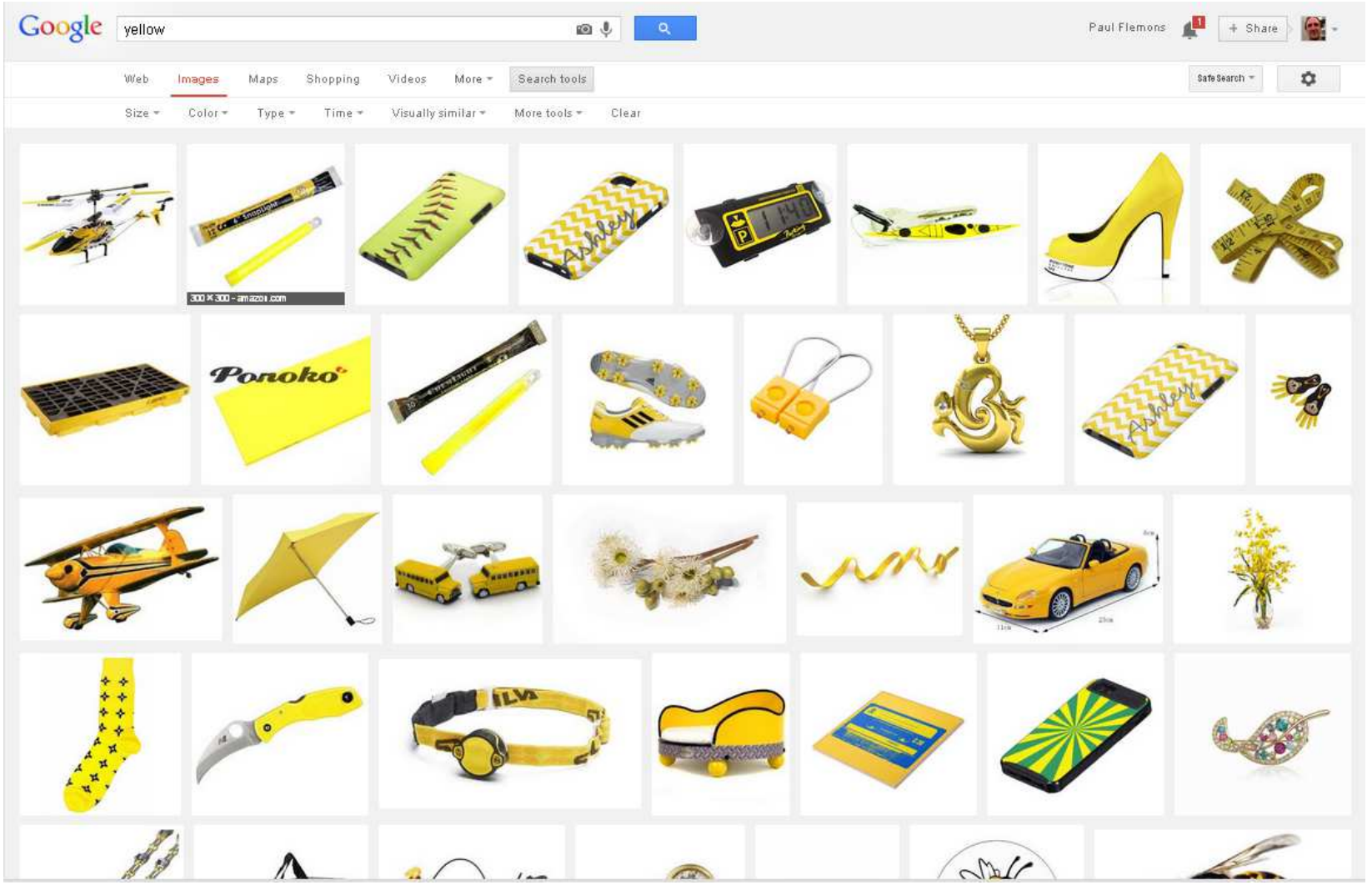
Hymenoptera + striped abdomen



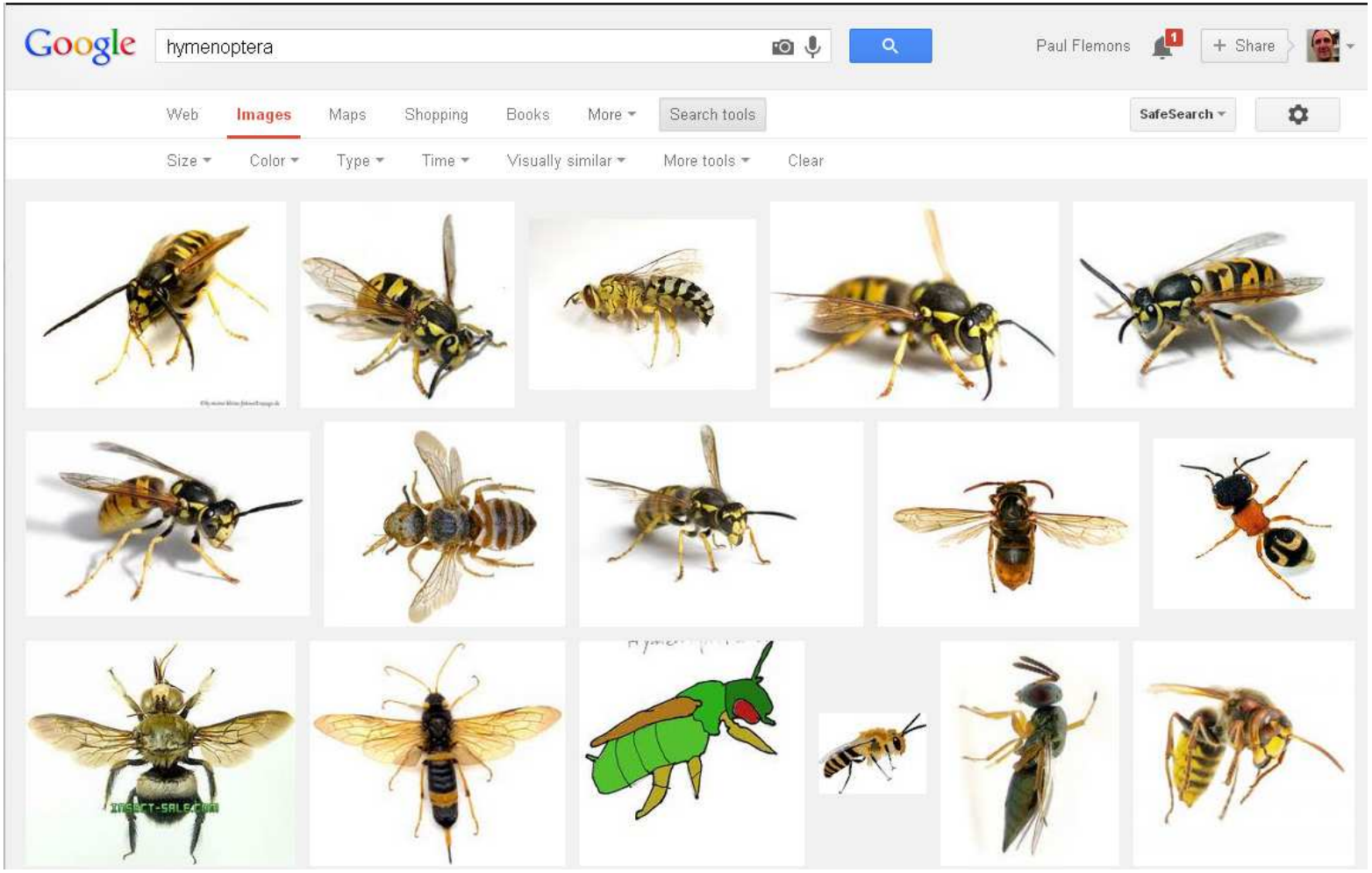
+ Striped abdomen



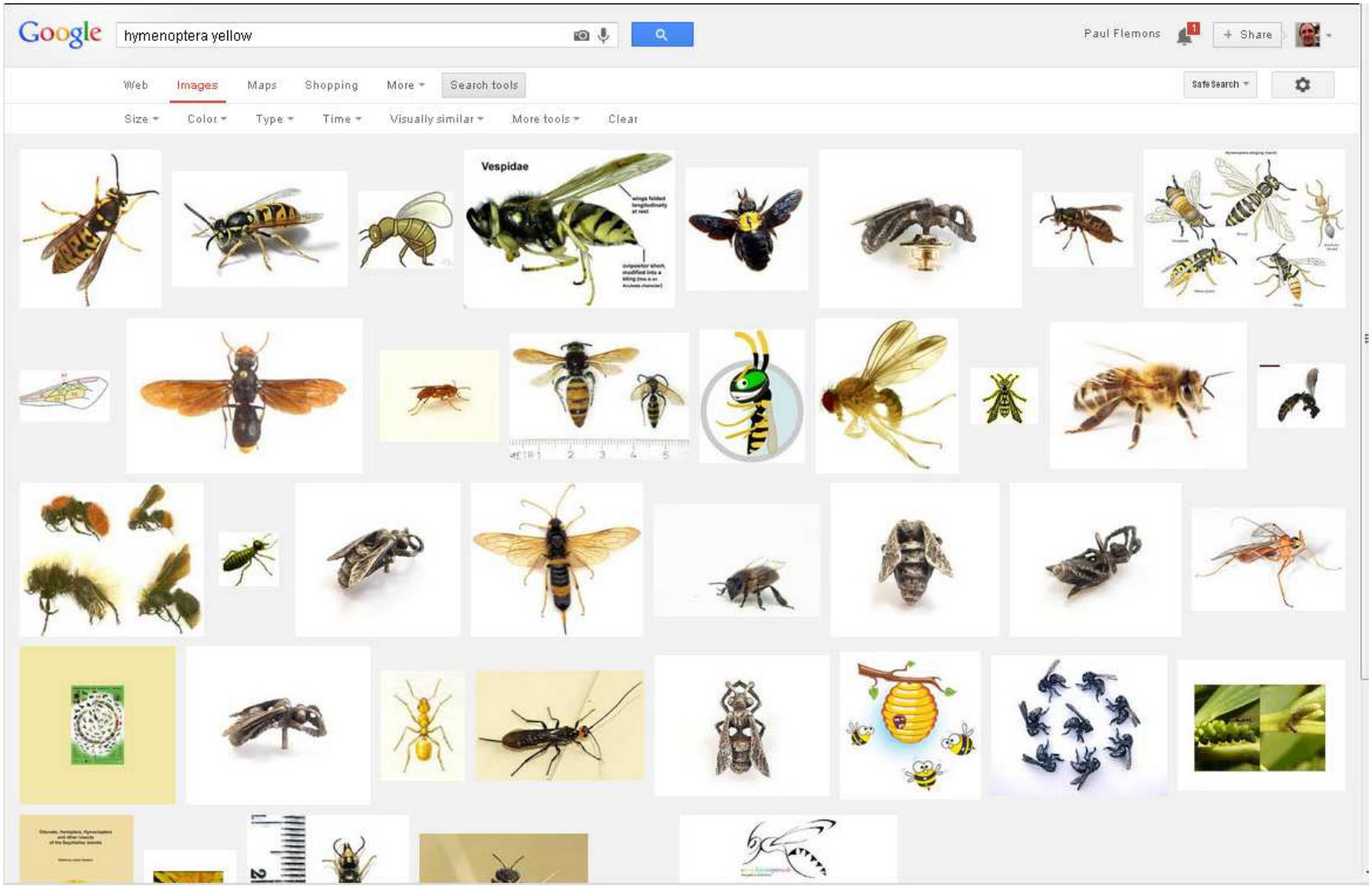
+ yellow


Google
yellow
Paul Flemons
+ Share
Web
Images
Maps
Shopping
Videos
More
Search tools
Safe Search
Size
Color
Type
Time
Visually similar
More tools
Clear
Ponoko
Ashley

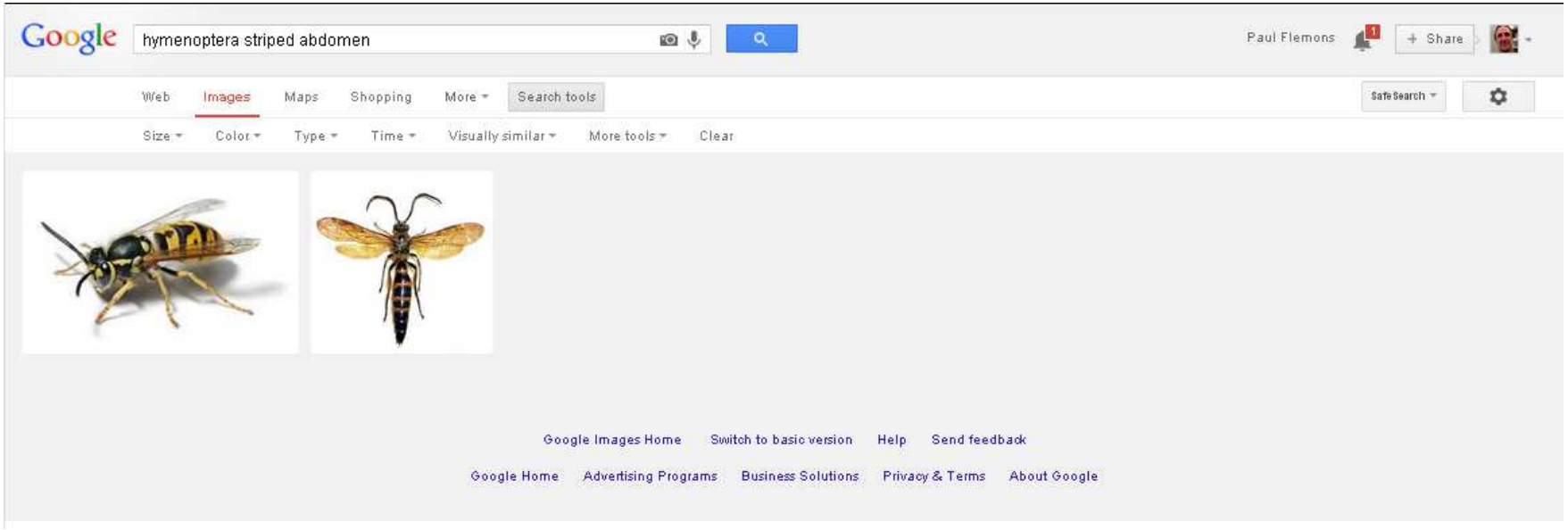
+ hymenoptera



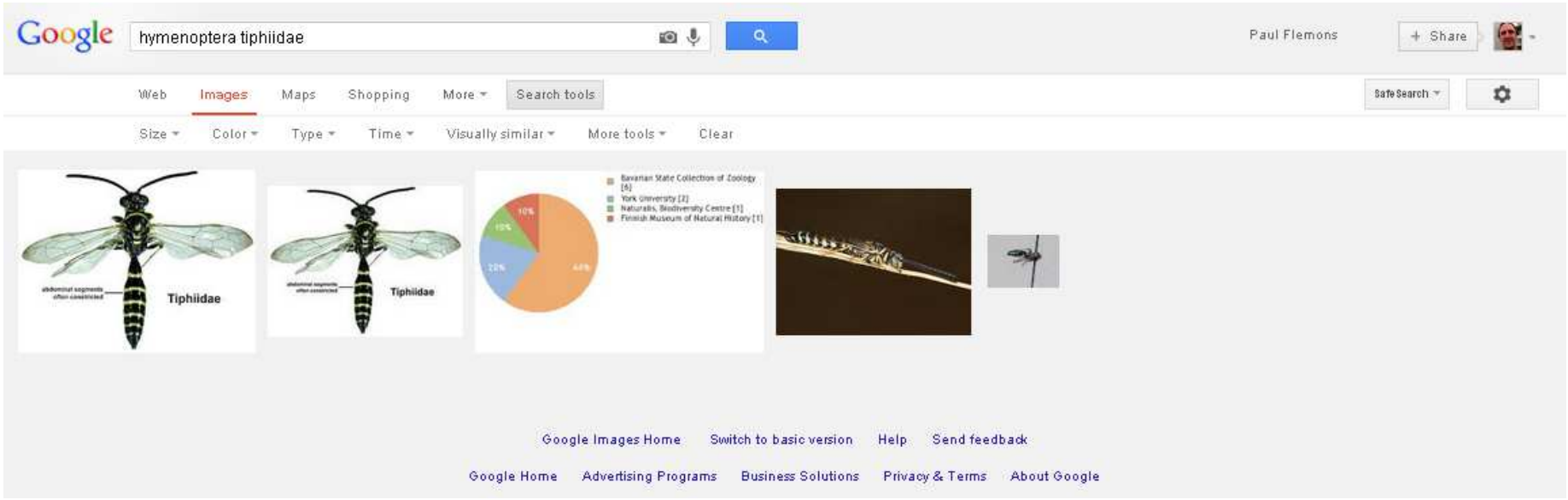
+ Hymenoptera + yellow



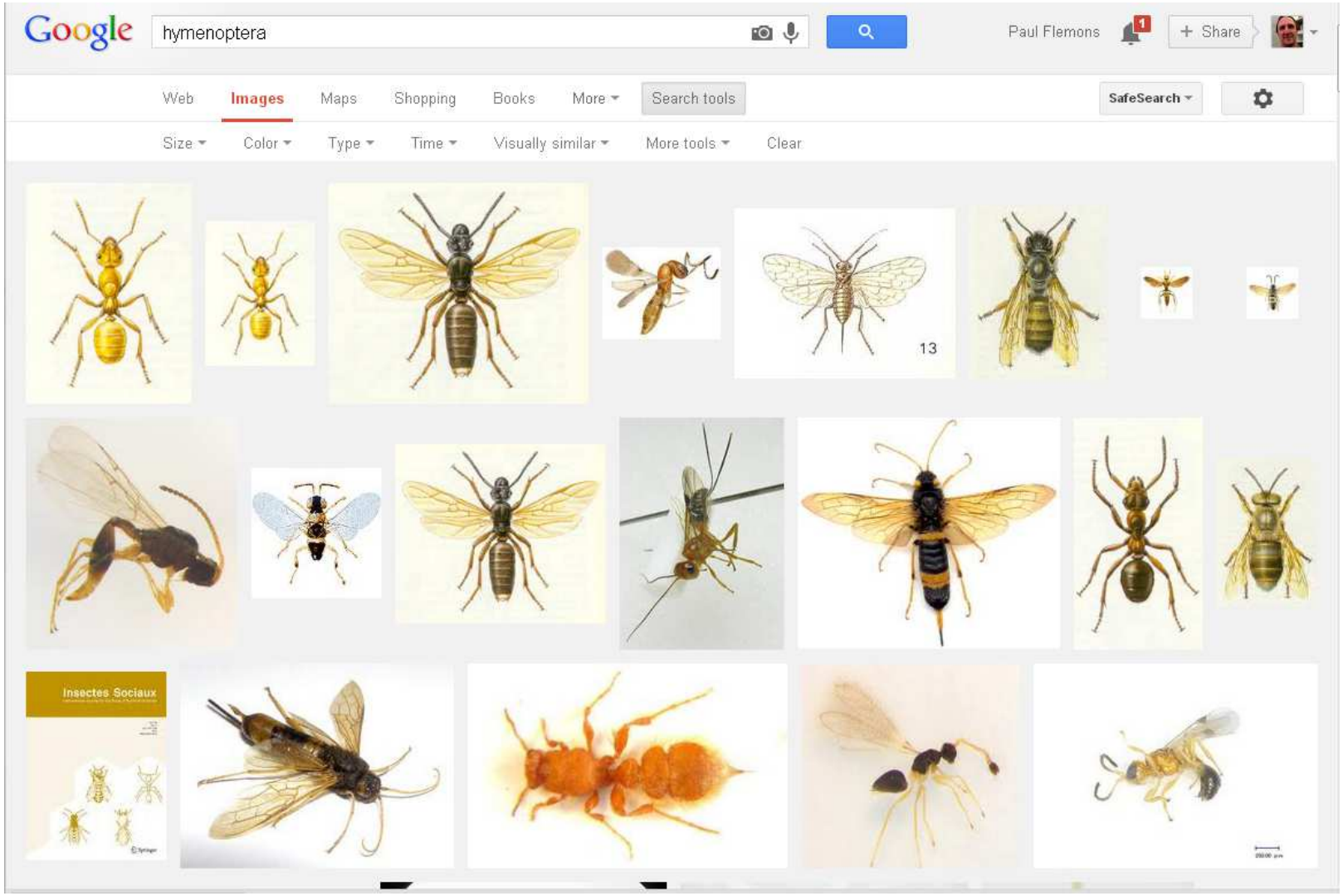
+ Hymenoptera + Striped abdomen



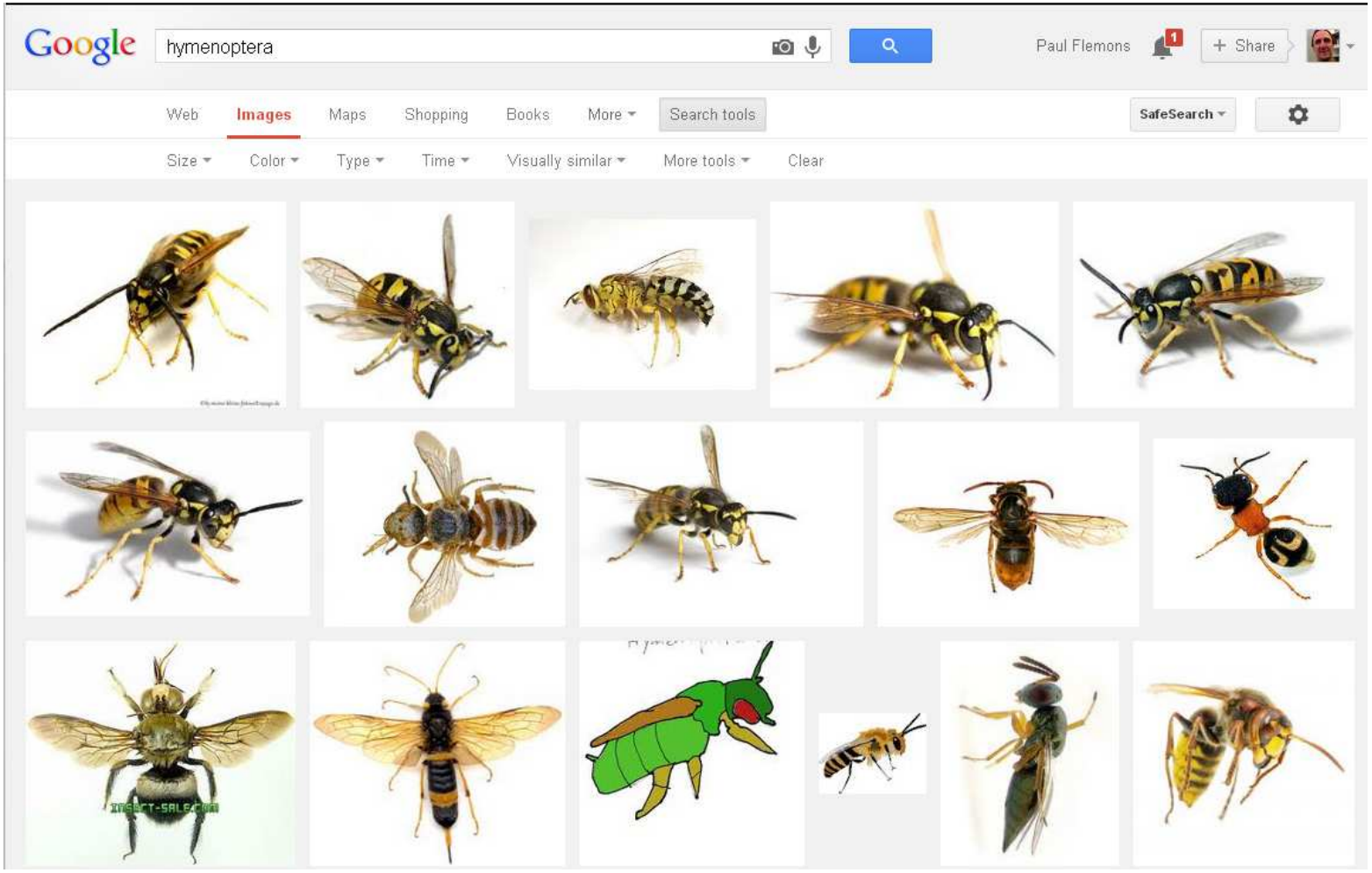
+ Hymenoptera + Tiphiidae



+ hymenoptera



+ hymenoptera



How might computer vision help?

1. Segmentation – extracting individual images from the soup images

How might computer vision help?

Use crowdsourcing apps and games to add more tags :

Tags

Yellow
Black
wing venation
striped abdomen
Short antennae
Thickened wing venation
Yellow legs
0.5 to 1cm
Hymenoptera
Tiphiidae

How computer vision help?

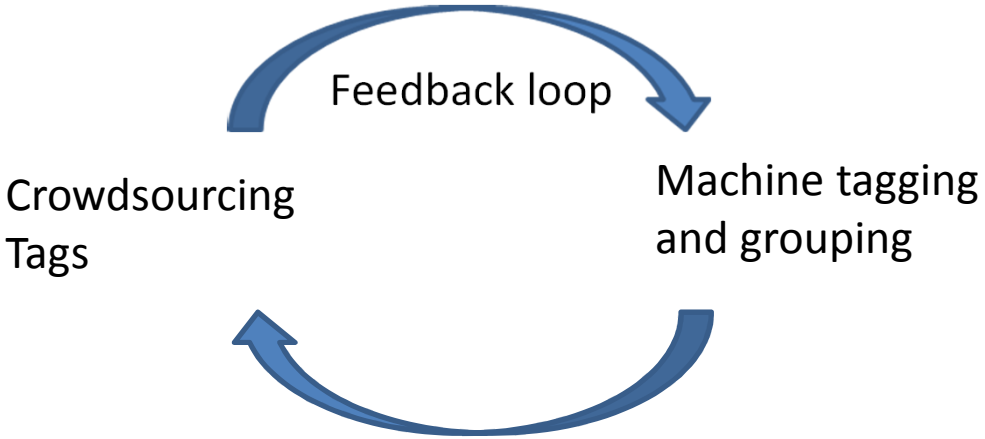
2. Classification – apply tags to each extracted image:

Tags

Yellow
Black
0.5 to 1cm
ratio: length to width

How might computer vision help?

3. Group and compare images:



How might computer vision help?

4. Count and calculate :

Measures of

- richness
- abundance

The Insect Soup Challenge

Goals

Computer vision techniques are to be used to extract meaning from the imagery by categorising and reporting on specimen qualities such as:

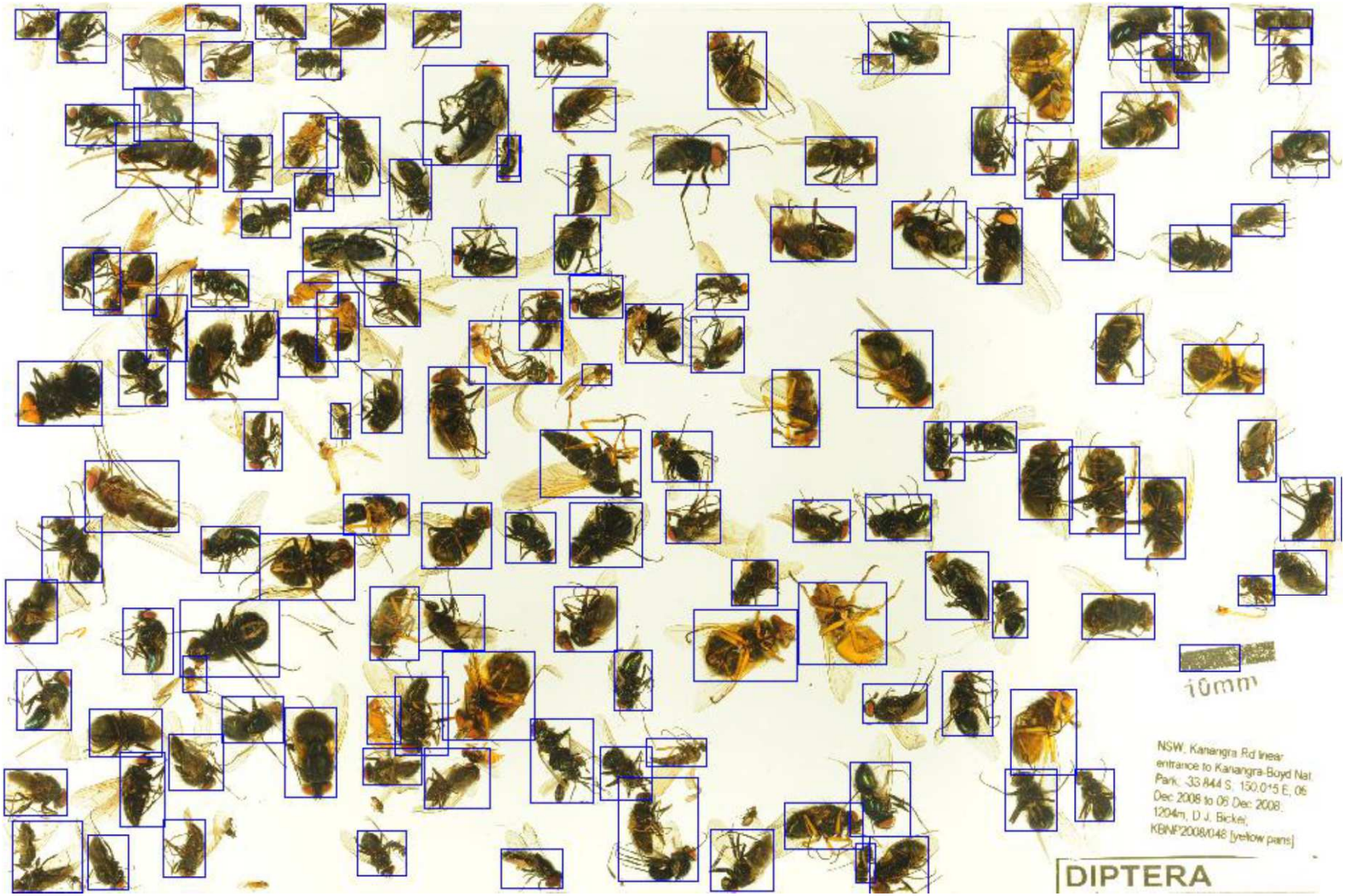
- Colour, Size, Shape, and/or Wing Venation
- Location (via bounding box or colour mask)
- Grouping of specimens within an image into like groups approximating:
  - Taxonomic Orders eg Hymenoptera, Diptera
  - Morphospecies
    [See Bioscience Resources section below for further information]
- Counts of:
  - Overall number of specimens in an image
  - Diversity of groups in an image, for example:
    - Of Orders – number of different Orders in an image
    - Of morphospecies – number of different morphospecies in an image
- Abundance within each group in an image, for example:
  - Number of individual specimens in each order in an image
  - Number of individual specimens in each morphospecies in an image
- Similarity measures within groups
- Outliers

How might computer vision help?

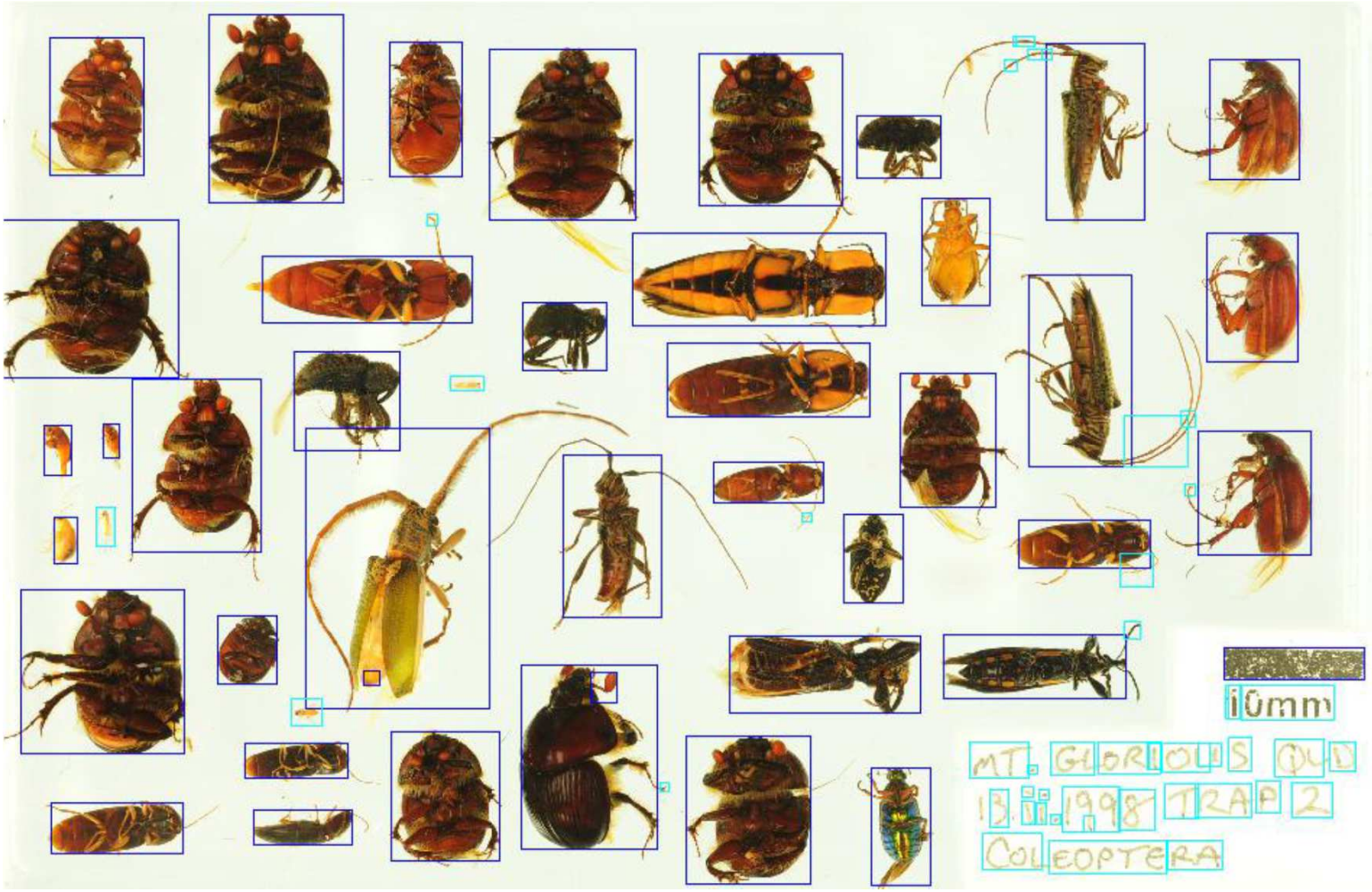
Insect Soup Challenge: Segmentation, Counting, and Simple Classification

Katarina Mele
Computational Informatics, CSIRO
Riverside Life Sciences Centre, 11 Julius Avenue, North Ryde, NSW, 2113
Katarina.Mele@csiro.au

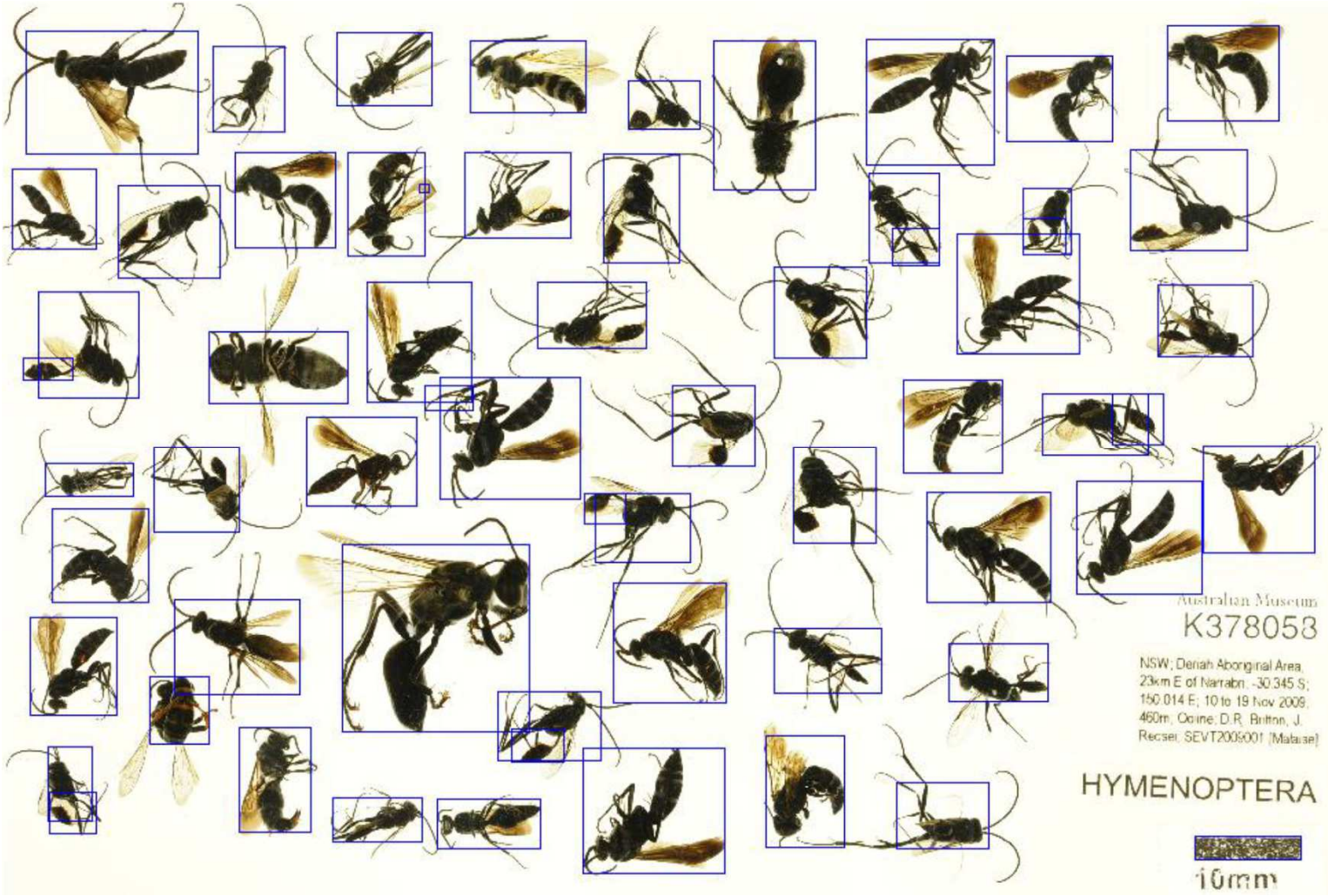
How might computer vision help?



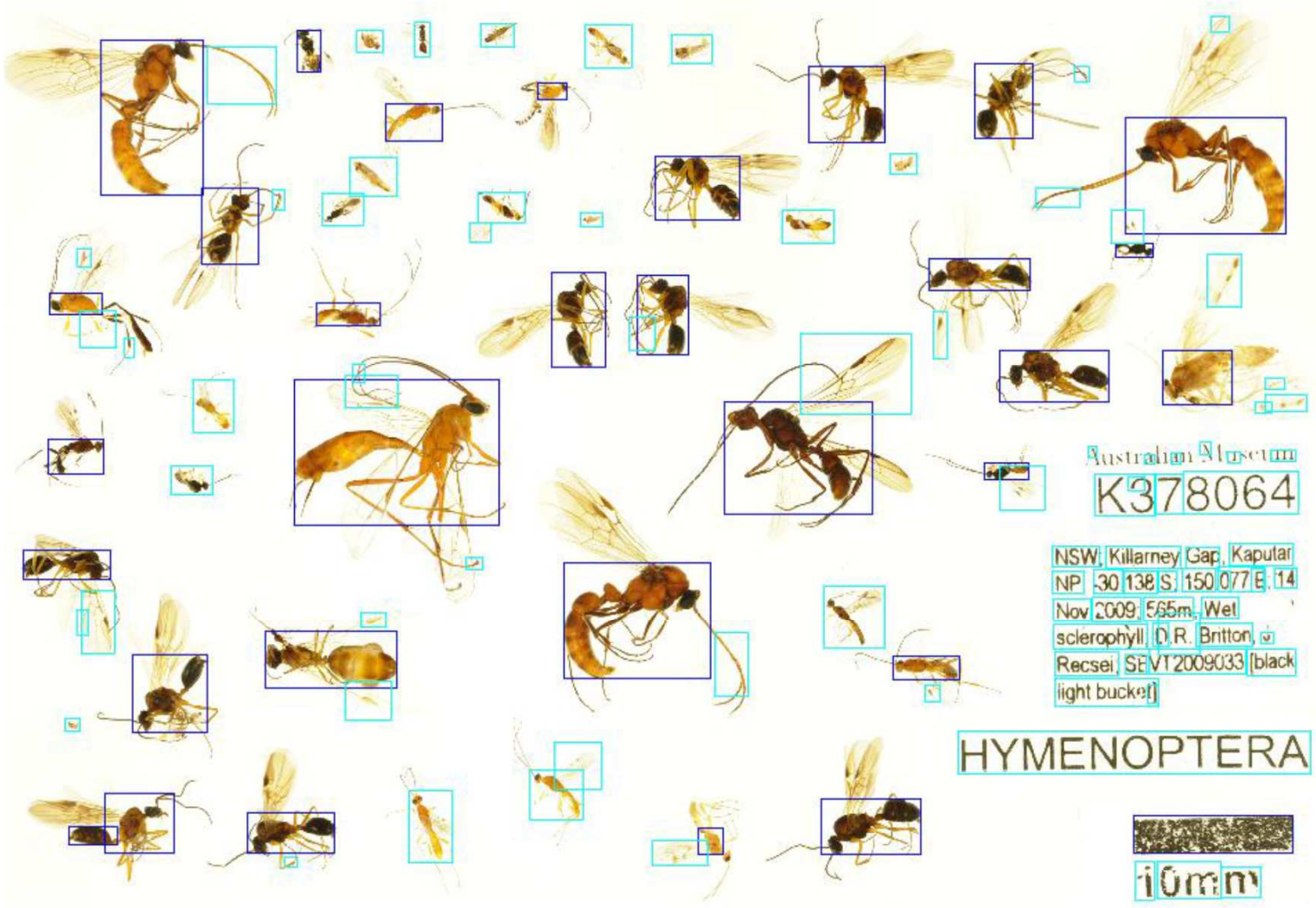
How might computer vision help?



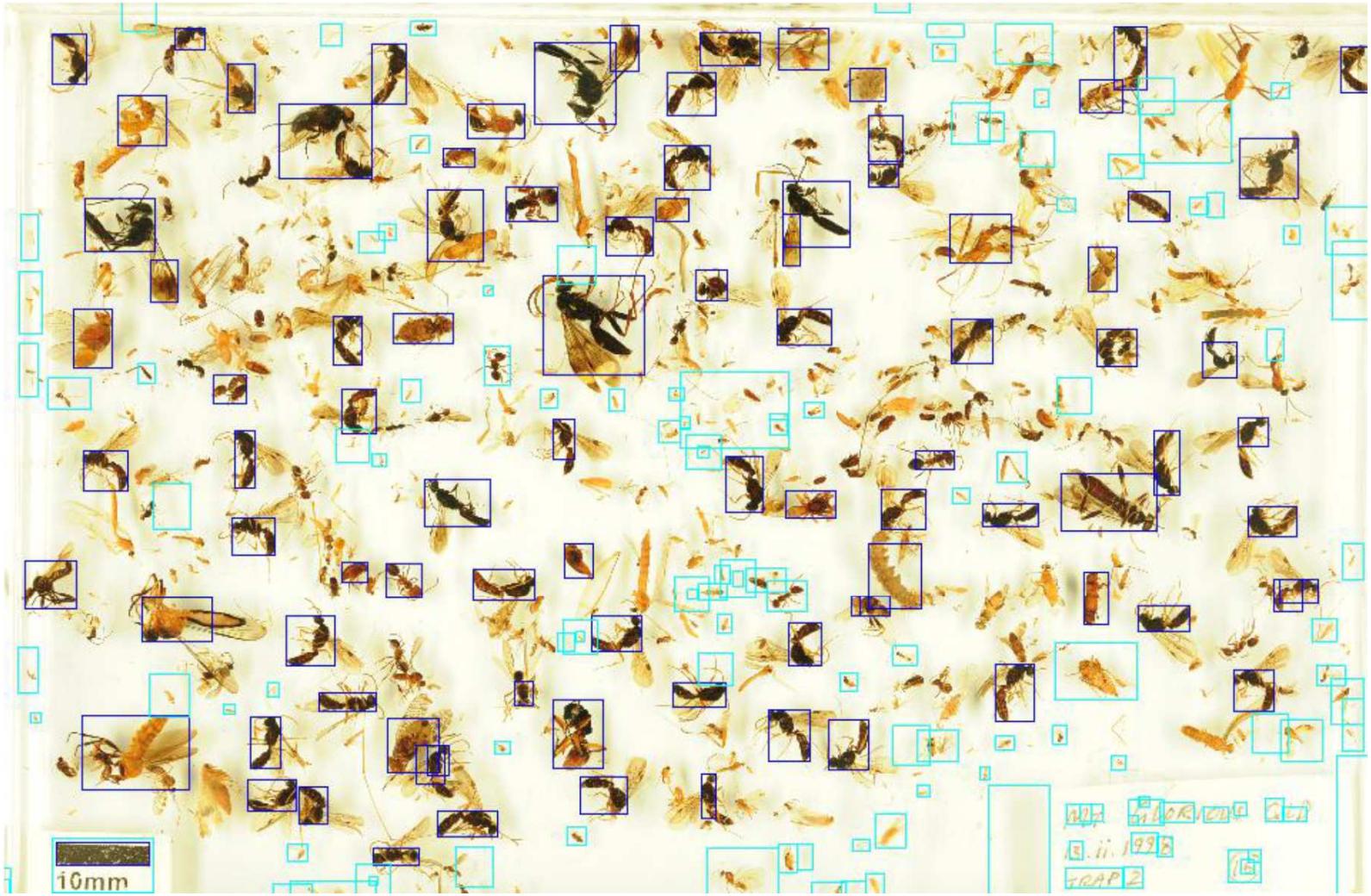
How might computer vision help?



How might computer vision help?



How might computer vision help?



Thank you

www.australianmuseum.net.au

nature culture **discover**

Australian **museum**