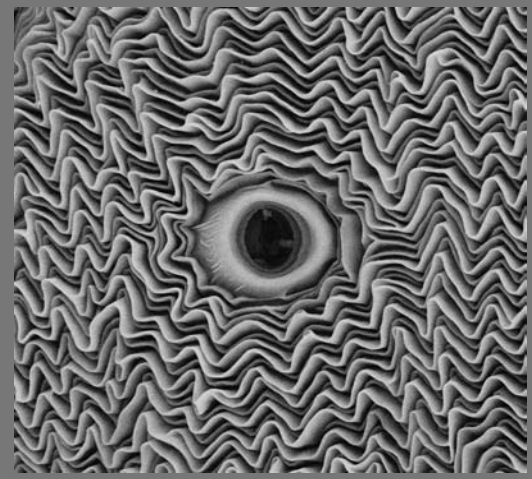


Ontologies, Image Databases, and Evolutionary Biology

Martín J. Ramírez
Museo Argentino de Ciencias Naturales – CONICET
Buenos Aires, Argentina

Evolutionary Biology & Ontologies Workshop
Evolution 2008, Minneapolis, 20 June 2008

Main topics

1. Use of an ontology to link images to a phylogenetic dataset: The Spider AToL

2. Image annotation using the anatomical ontology

3. Notes on multi-species ontology design

4. Ontologies bridging communities of systematics and model organisms

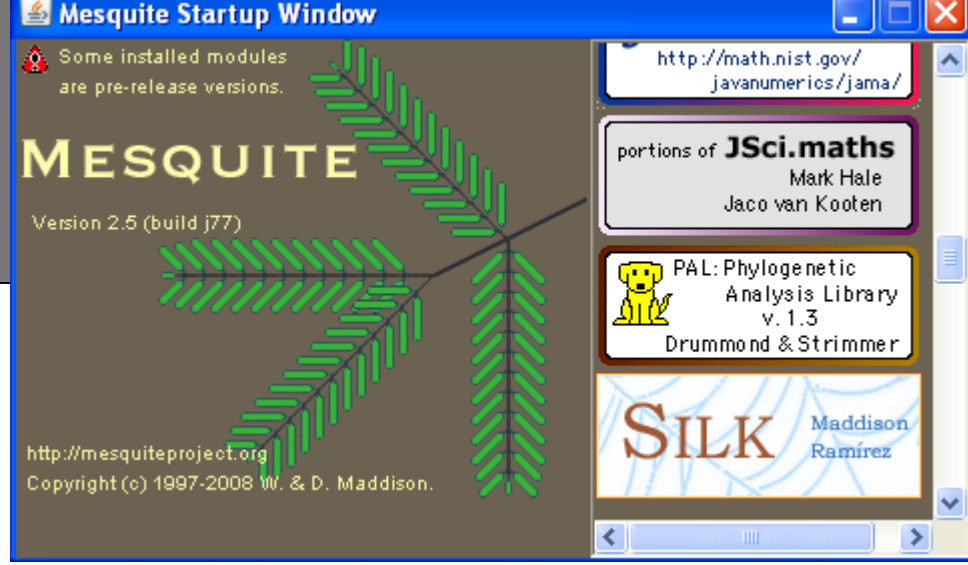
Who I am

Spider AToL curator of image collection. Atlas of anatomy & ontology.

Links to and from phylogenetic dataset.

Developed ontology (independently) to solve treatment of images

SILK



Syst. Biol. 56(2):283–294, 2007

ISSN: 1063-5157 print / 1076-836X online
DOI: 10.1080/10635150701313848

Linking of Digital Images to Phylogenetic Data Matrices Using a Morphological Ontology

MARTÍN J. RAMÍREZ,[1] JONATHAN A. CODDINGTON,[2] WAYNE P. MADDISON,[3,4] PETER E. MIDFORD,[3] LORENZO PRENDINI,[5] JEREMY MILLER,[6] CHARLES E. GRISWOLD,[6] GUSTAVO HORMIGA,[7] PETRA SIERWALD,[8] NIKOLAJ SCHARFF,[9] SURESH P. BENJAMIN,[7] AND WARD C. WHEELER[5]

[1]*Museo Argentino de Ciencias Naturales "Bernardino Rivadavia"—CONICET. Avenida Angel Gallardo 470, C1405DJR, Buenos Aires, Argentina; E-mail: ramirez@macn.gov.ar*
[2]*Entomology, Smithsonian Institution, PO Box 37012, NMNH E529, NHB-105, Washington, DC 20013-7012, USA*
Departments of [3]*Zoology and* [4]*Botany, 6270 University Boulevard, University of British Columbia, Vancouver, BC V6T 1Z4, Canada*

Syst. Biol. 2007

SILK: Simple Image LinKing (Maddison & Ramírez 2006)
A Mesquite package for associating images with character matrices.
Beta test version available online at http://mesquiteproject.org/SILK

How to link images to dataset cells?

ATOL Spiders. Phylogenetic dataset with

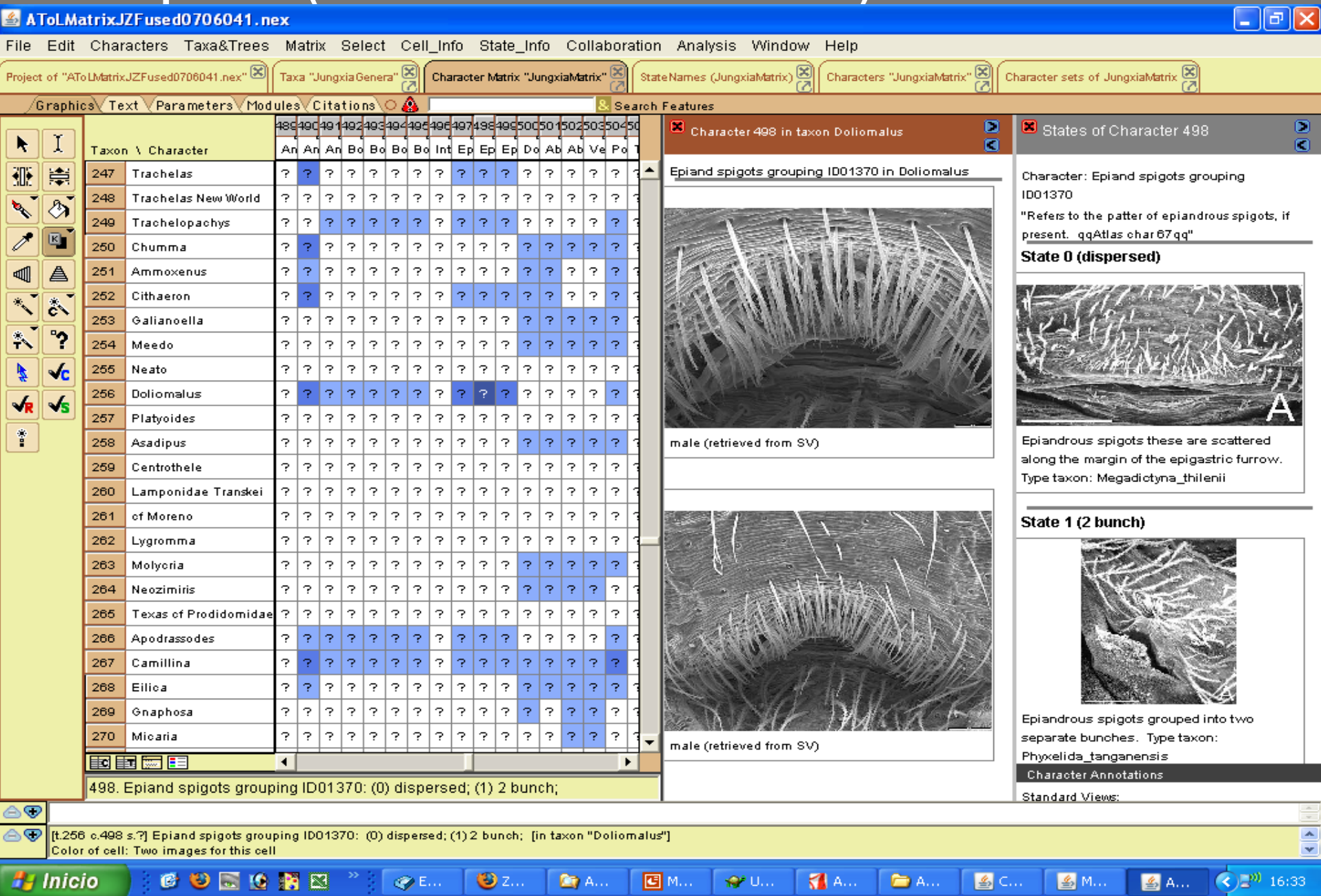
500 terminals

1000 to **2000** characters

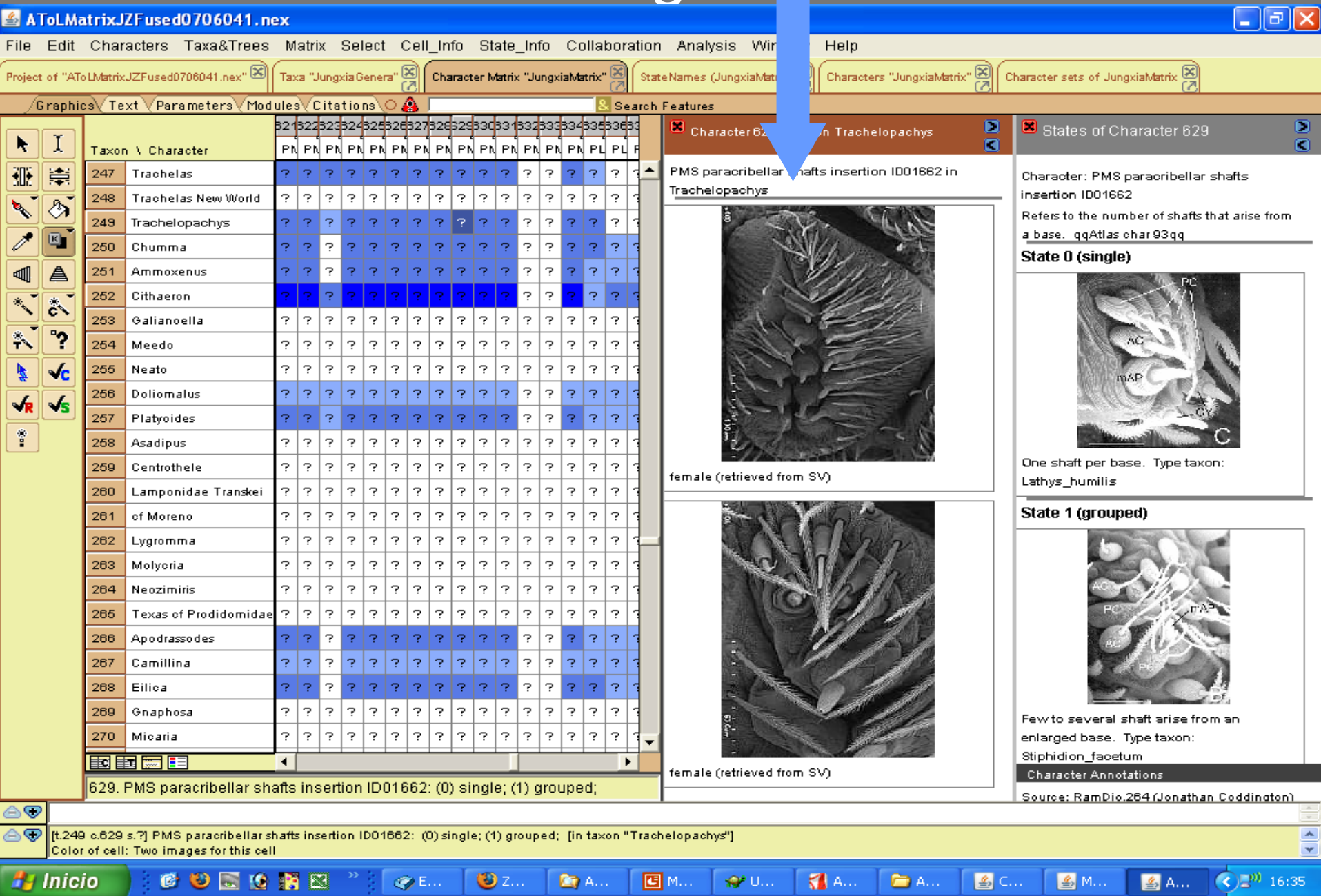
= **500,000** to **1,000,000** cells

1. It is impossible to link images one by one. Documentation should not be such a heavy burden
2. Ad-hoc links are difficult to maintain (characters and terminals revised, …)
3. Further images can be produced after the scoring. Workflow for images and dataset may be independent
4. Images are very useful before scoring cells (how a structure looks in other terminals not yet scored?)

Mesquite (W. & D. Maddison) & SILK



Cell images


AToLMatrixJZFused0706041.nex
File Edit Characters Taxa&Trees Matrix Select Cell_Info State_Info Collaboration Analysis Help
Project of "AToLMatrixJZFused0706041.nex"
Taxa "JungxiaGenera"
Character Matrix "JungxiaMatrix"
Characters "JungxiaMatrix"
Character sets of JungxiaMatrix
Graphics Text Parameters Modules Citations
Search Features
Taxon \ Character
247 Trachelas
248 Trachelas New World
249 Trachelopachys
250 Chumma
251 Ammoxenus
252 Cithaeron
253 Galianoella
254 Meedo
255 Neato
256 Doliomalus
257 Platyoides
258 Asadipus
259 Centrothele
260 Lamponidae Transkei
261 cf Moreno
262 Lygromma
263 Molycria
264 Neozimiris
265 Texas cf Prodidomidae
266 Apodrassodes
267 Camillina
268 Eilica
269 Gnaphosa
270 Micaria
629. PMS paracribellar shafts insertion ID01662: (0) single; (1) grouped;
Character 62 n Trachelopachys
PMS paracribellar shafts insertion ID01662 in Trachelopachys
female (retrieved from SV)
female (retrieved from SV)
States of Character 629
Character: PMS paracribellar shafts insertion ID01662
Refers to the number of shafts that arise from a base. qqAtlas char 93qq
State 0 (single)
PC
AC
mAP
C
One shaft per base. Type taxon: Lathys_humilis
State 1 (grouped)
Few to several shaft arise from an enlarged base. Type taxon: Stiphidion_facetum
Character Annotations
Source: RamDio.264 (Jonathan Coddington)
[t.249 c.629 s.?] PMS paracribellar shafts insertion ID01662: (0) single; (1) grouped; [in taxon "Trachelopachys"]
Color of cell: Two images for this cell
Inicio
16:35

Character state images & text

AToLMatrixJZFused0706041.nex

File Edit Characters Taxa&Trees Matrix Select Cell_Info State_Info Collaboration Analysis Window Help

Project of "AToLMatrixJZFused0706041.nex" | Taxa "JungxiaGenera" | Character Matrix "JungxiaMatrix" | State Names (JungxiaMatrix) | Characters "JungxiaMatrix" | Character sets of Jungxi

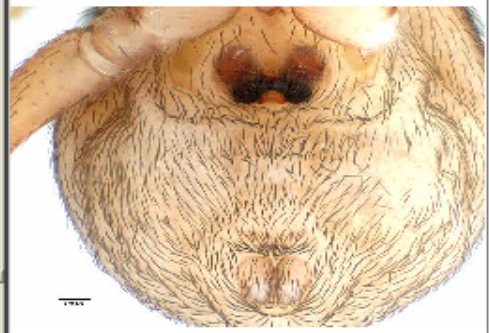
Graphics Text Parameters Modules Citations Search Features

Taxon \ Character		681	682	683	684	685	686	687	688	689	690	691	692	693	694	695	696
		PL	PL	An	Su	An	Abdc	He	He	Fe	Ep	Ep	Sc	Ep	Co	Co	Ep
247	Trachelas	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
248	Trachelas New World	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
249	Trachelopachys	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
250	Chumma	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
251	Ammoxenus	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
252	Cithaeron	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
253	Galianoella	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
254	Meedo	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
255	Neato	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
256	Doliomalus	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
257	Platyoides	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
258	Asadipus	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
259	Centrothele	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
260	Lamponidae Transkei	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
261	cf Moreno	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
262	Lygromma	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
263	Molycria	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
264	Neozimiris	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
265	Texas cf Prodidomidae	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
266	Apodrassodes	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
267	Camillina	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
268	Eilica	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
269	Gnaphosa	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
270	Micaria	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?

690. Epigynum ID02694: (0) abs; (1) pres;

Character 690 in taxon Chumma

Epigynum ID02694 in Chumma



female (retrieved from SV)

States of Ch...690

Character: Epigynum ID02694

Any sclerotized modification of the cuticle around the female genital region. qqAtlas char 131qq

State 0 (abs)

Cuticle around genital region similar to surrounding. Type taxon: Archoleptoneta

State 1 (pres)

"Cuticle around genital region sclerotized, contrasts to surrounding, may or may not have elaborate processes. Type taxon:

Character Annotations

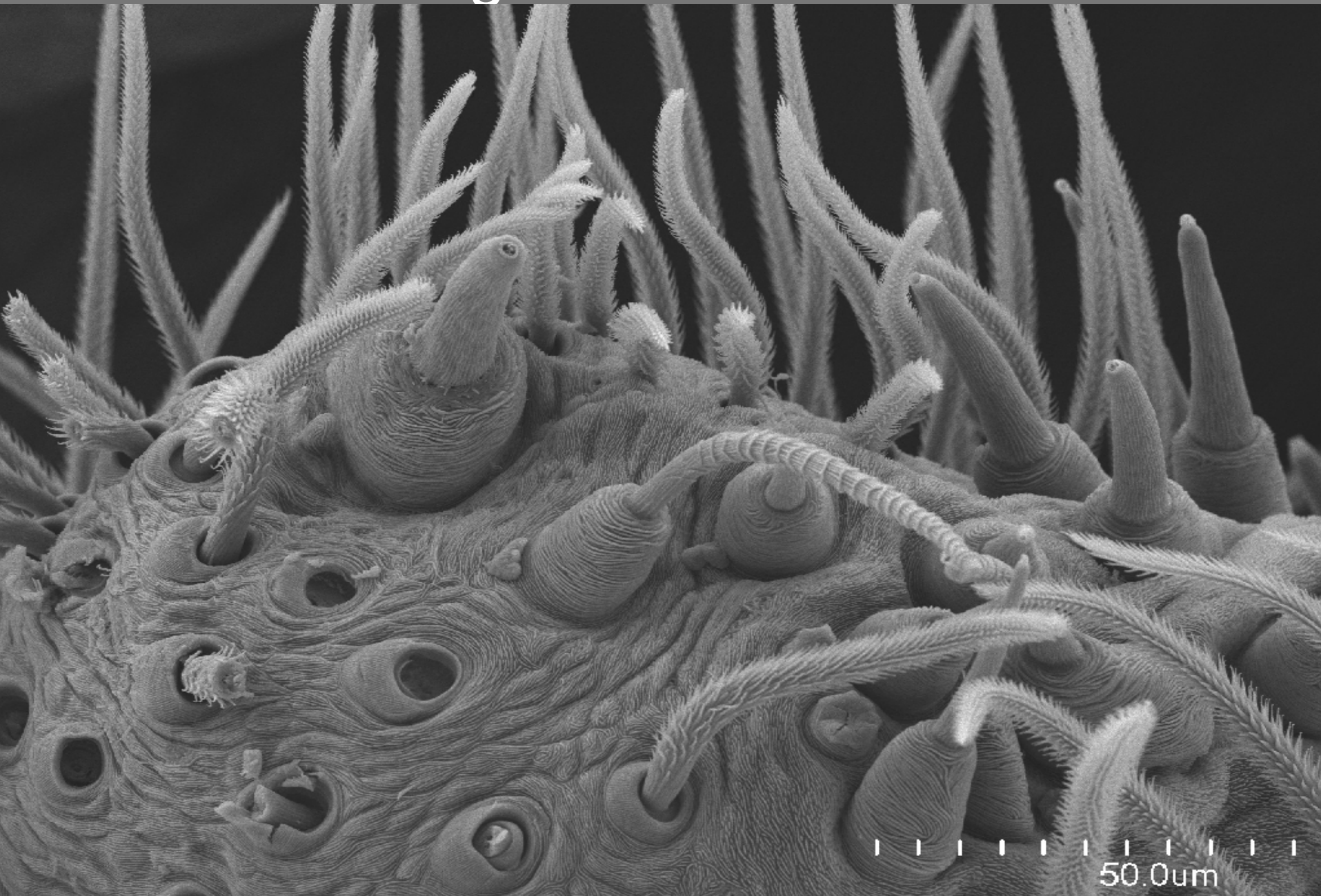
Source: n67.040 c55.115 m72.026

[t.250 c.690 s.?] Epigynum ID02694: (0) abs; (1) pres; [in taxon "Chumma"]

Color of cell: One image for this cell

Inicio 16:35

Double click: High resolution



Zoom in, move around, etc.

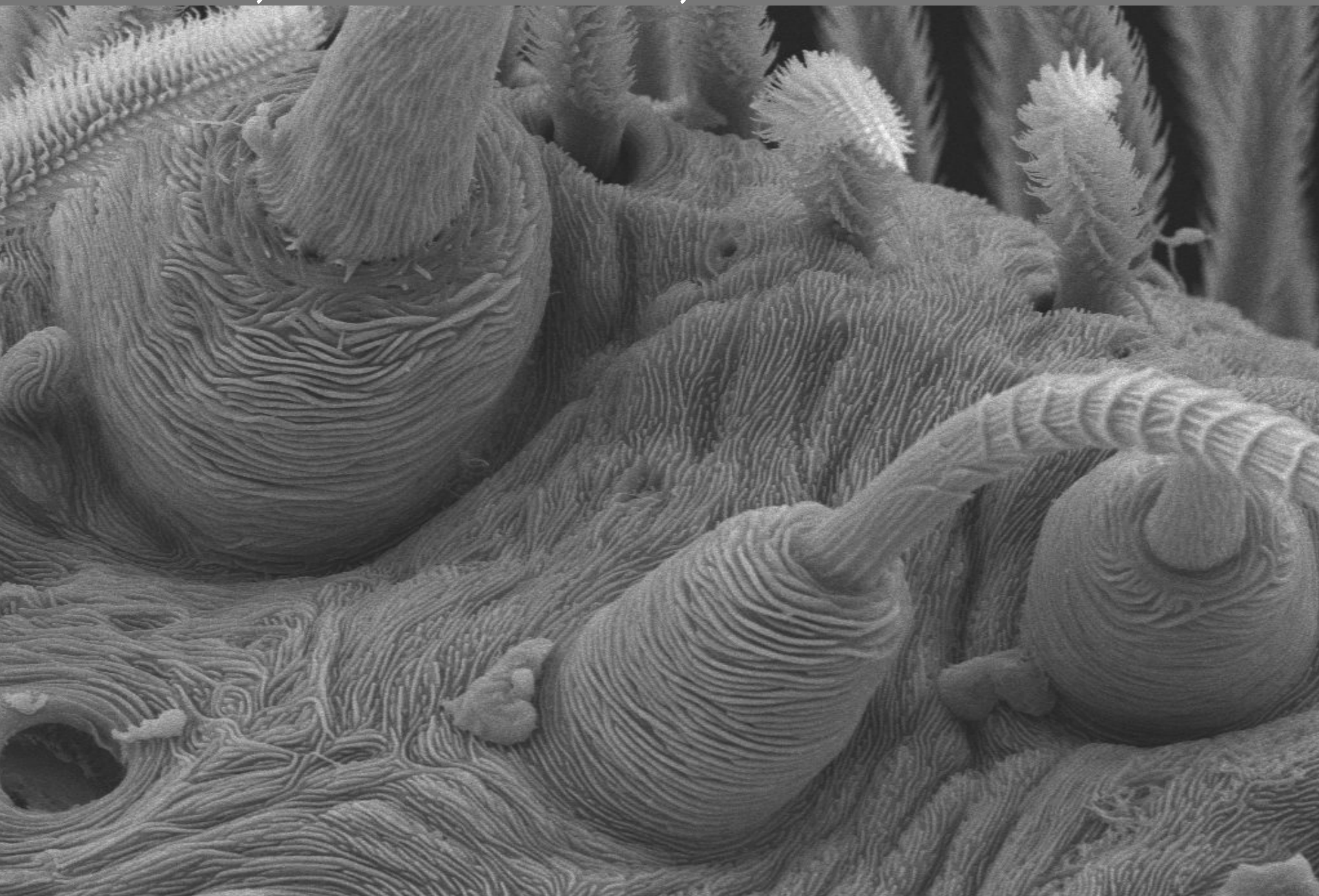


Image database: Metadata

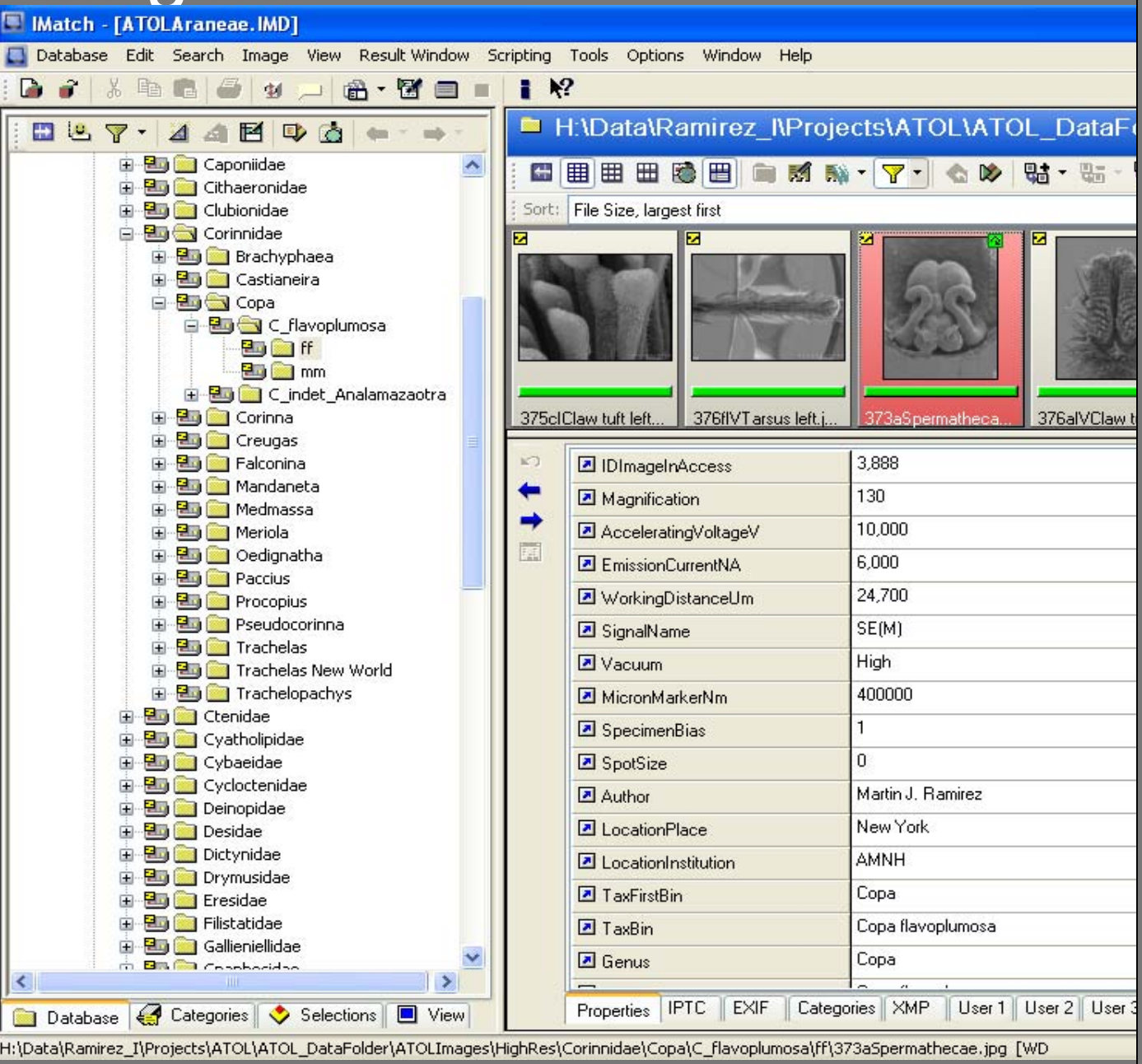


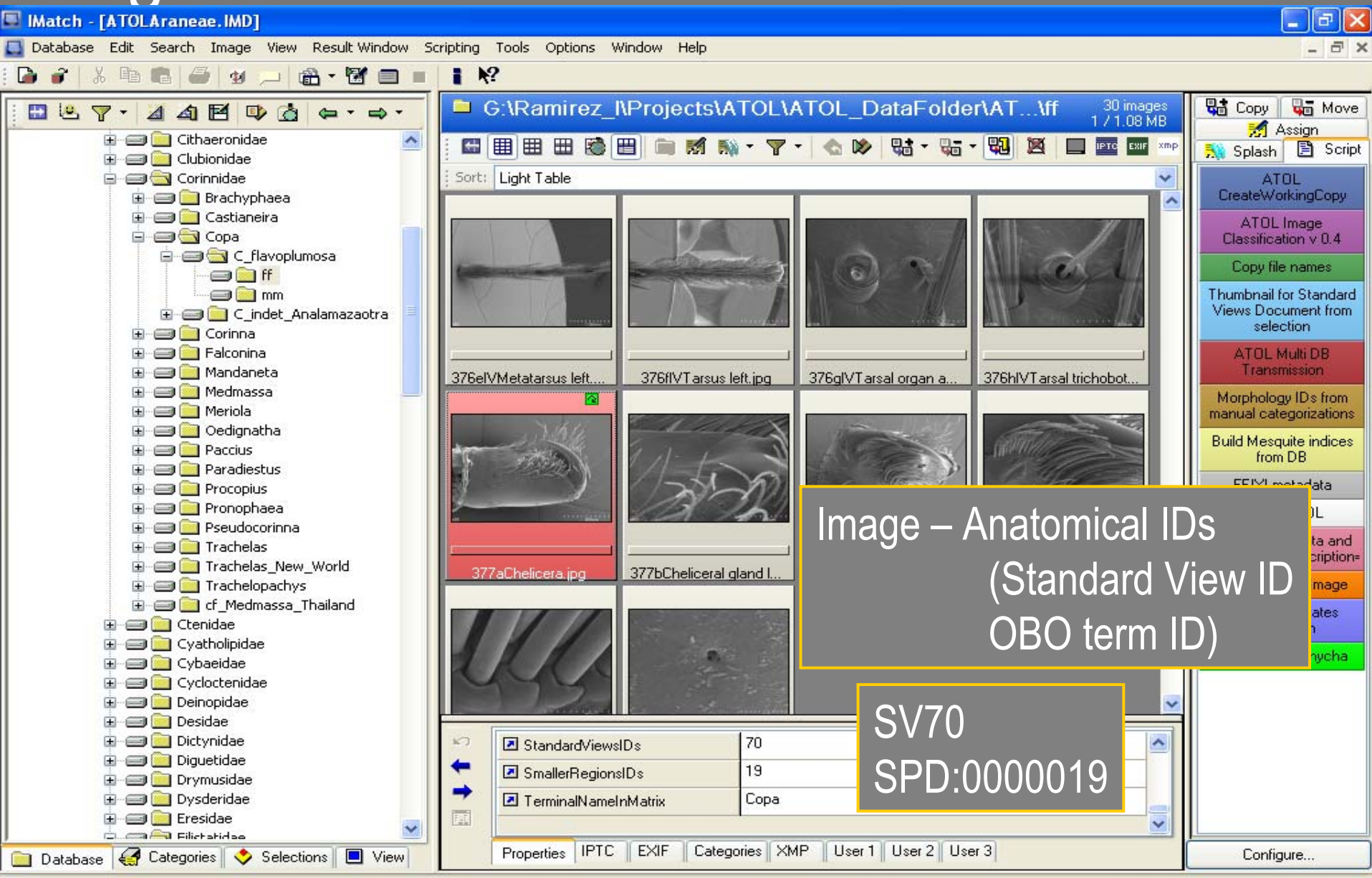
Image
Resolution
Microscope settings
Author …

Preparation
Critical point dried
Coated …

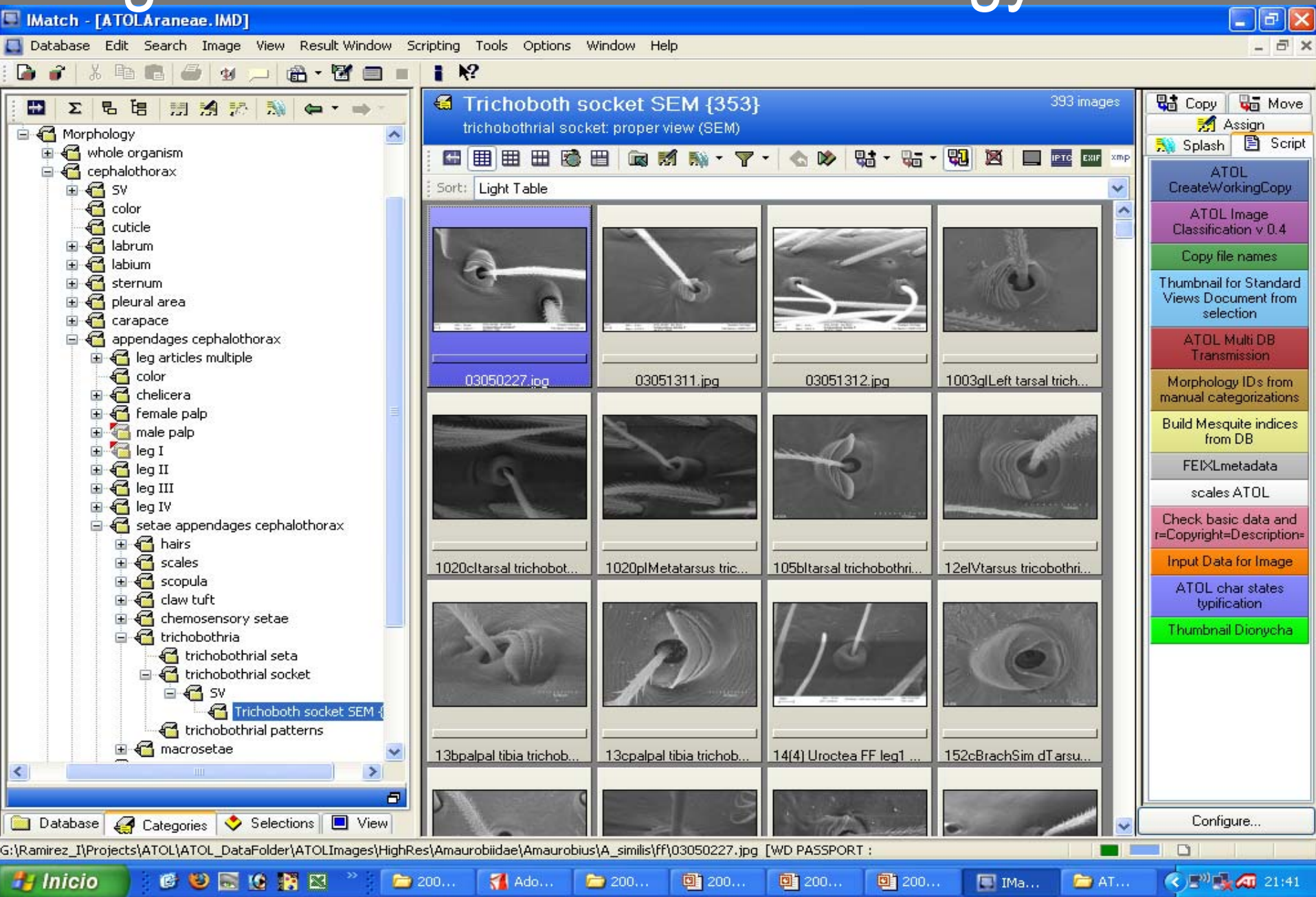
Voucher
Locality
Museum ID …

Taxonomy
Species etc.
Identified by / date …

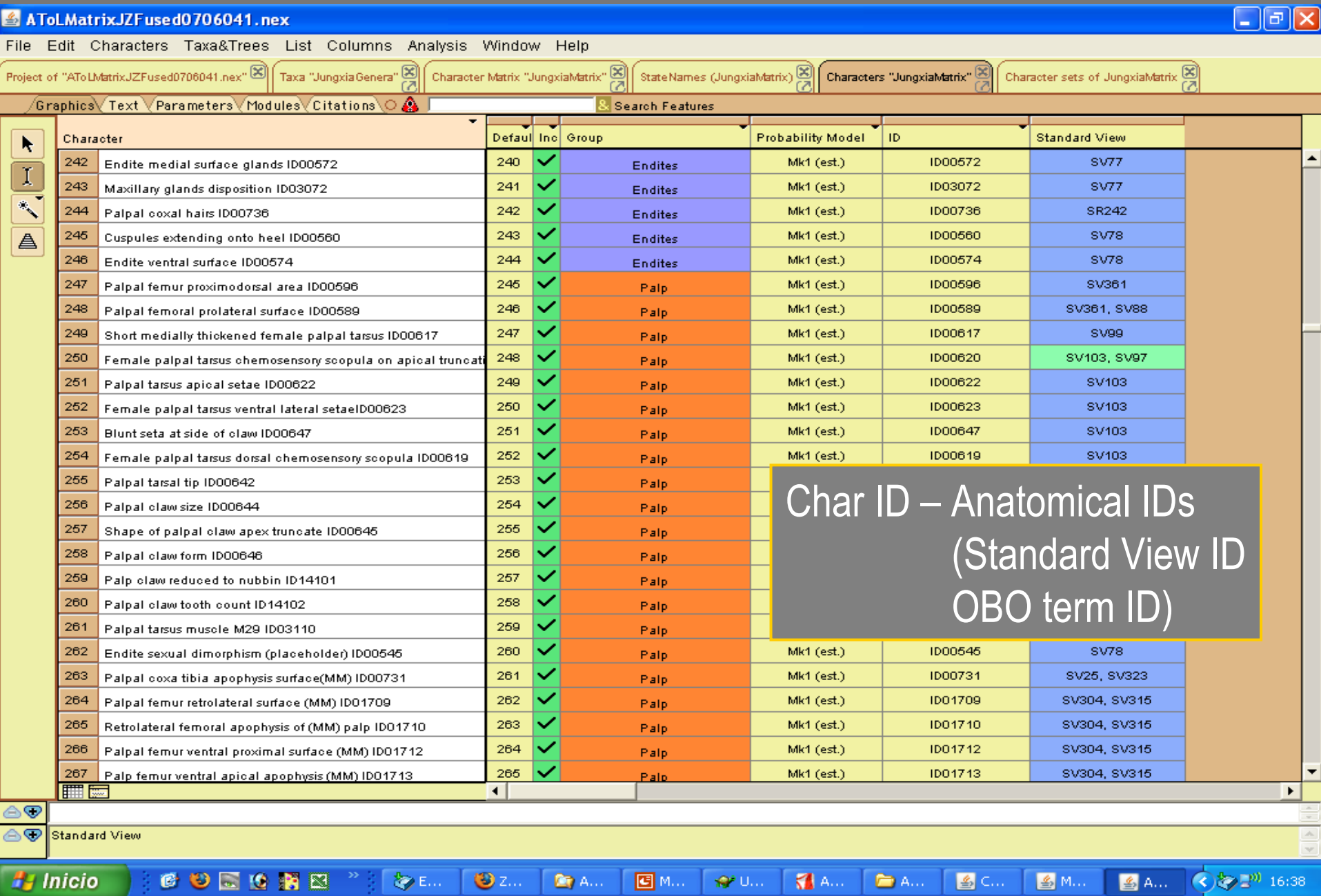
Image metadata: Anatomical IDs



Images on the anatomical ontology

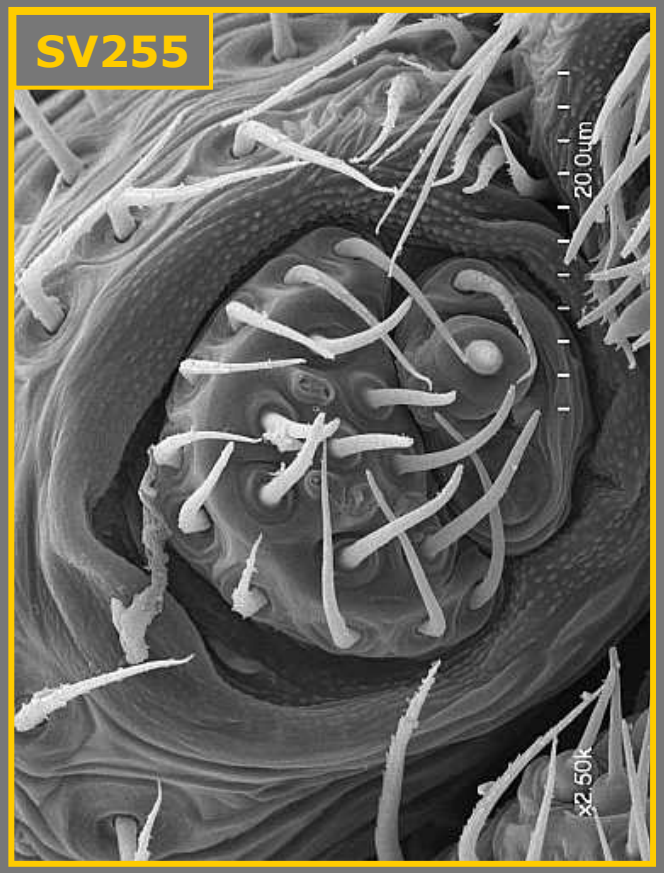


Characters annotated with Anatomical IDs



Char ID – Anatomical IDs
(Standard View ID
OBO term ID)

Standard Views



A combination of specific:

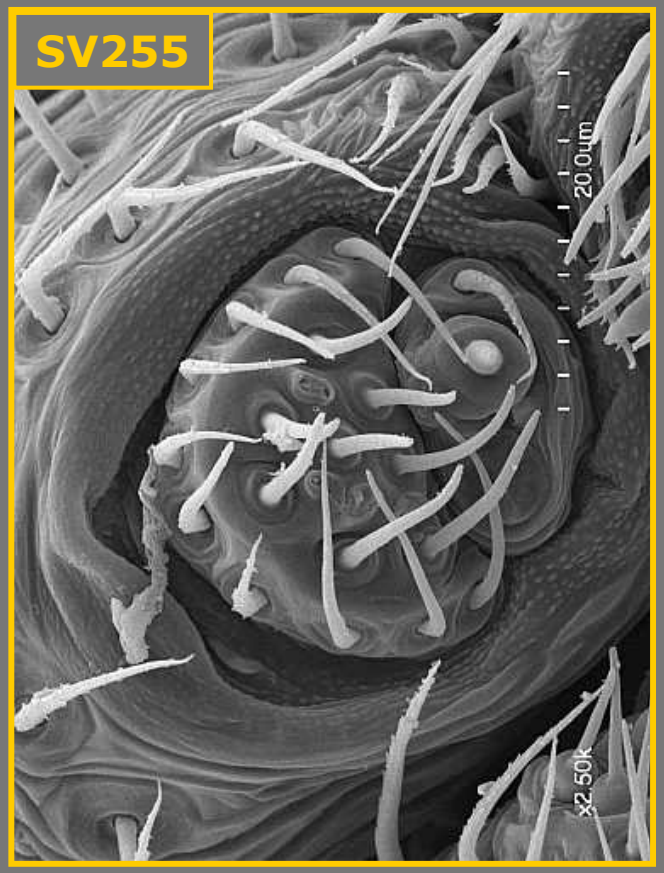
- Ontology term
- Sex and life stage
- Orientation
- Device and preparation technique

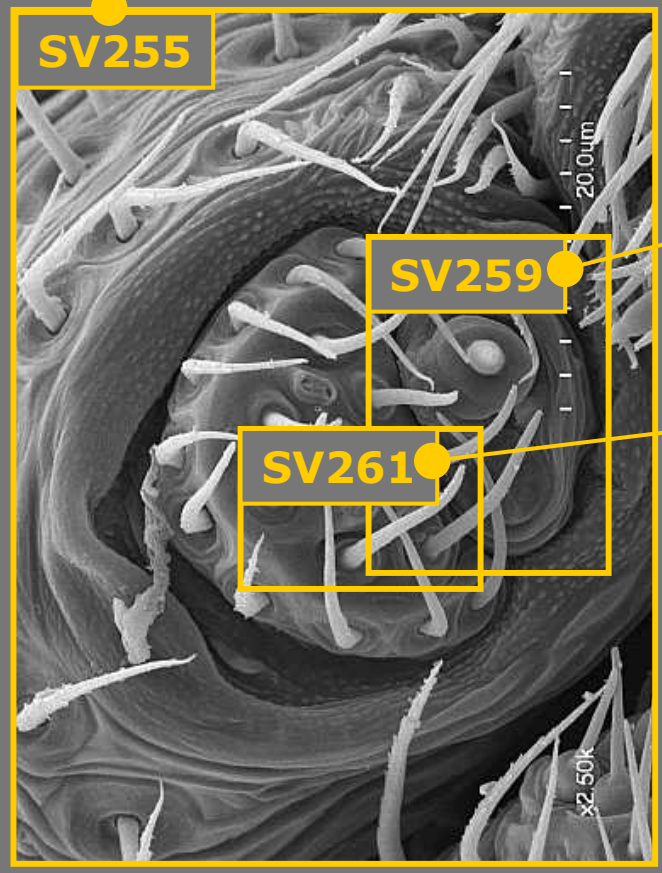
SV255: Male ALS spinning field SEM:

- Anterior lateral spinneret (SPD:0000125)
- Adult male
- Ventral
- SEM

Images are annotated as they are produced

Multiple SVs in one image





A **high resolution** image may serve multiple standard views (by zooming in)

Simple query to populate cells with images

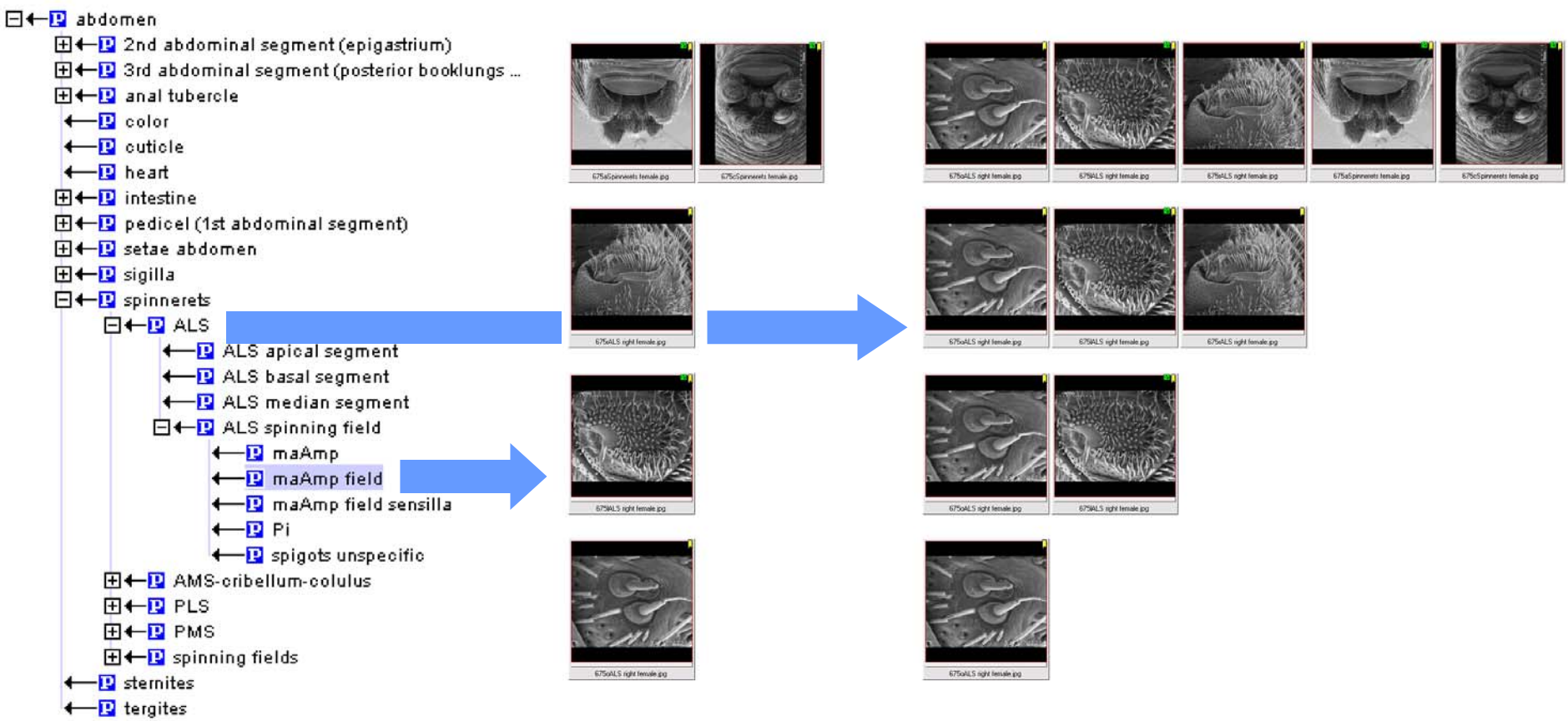
Cell = Character X, Terminal Y

For Character X – Retrieve Anatomical IDs W (SVs, OBO IDs)

Cell images =
For Terminal Y – Retrieve images tagged with Anatomical IDs W

Next: Intelligent algorithms to fetch images

1. The basic:
   - Fetch parent terms for broader selection
   - Fetch parent term if child does not retrieve images



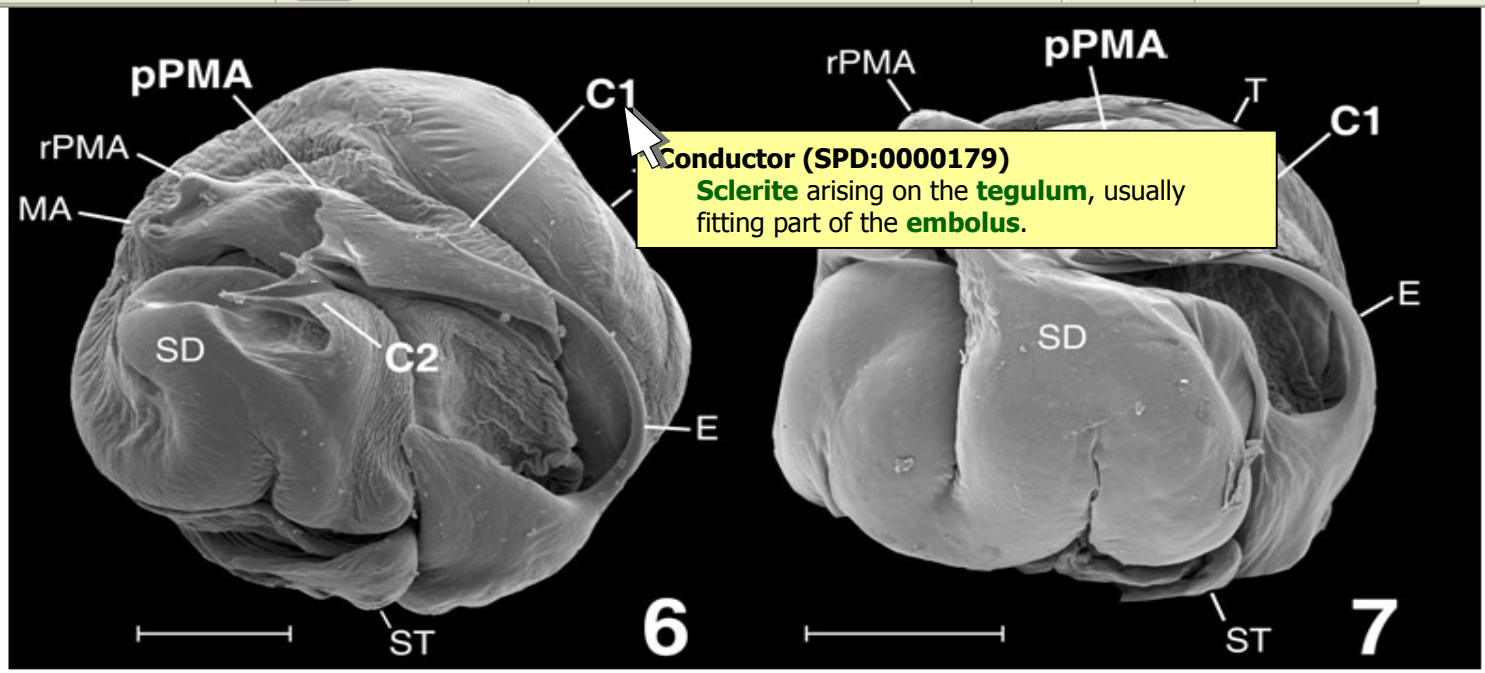
2. Process feedback from users …

Future: Links to OBO from PDF, HTML, …

Morphster, working on labels. HymAToL on ontology-driven text recognition

File Edit View Document Tools Window Help

Save a Copy Search Select 125% Help Search Web



Conductor (SPD:0000179)
Sclerite arising on the **tegulum**, usually fitting part of the **embolus**.

Figs 4–7. Left copulatory bulbs of Amaurobioidini. 4. *Gamakia hirsuta* Ramírez. 5. *Coptoprepes flavopilosus* Simon. 6. *Ferrieria echinata* Simon. 7. *Aysenoides colecole* Ramírez. (C1 = primary conductor; C2 = secondary conductor; E = embolus; MA = median apophysis; pPMA = prolateral cusp of paramedian apophysis; rPMA = retrolateral cusp of the paramedian apophysis; SD = sperm duct on distal tegulum; ST = subtegulum; T = tegulum.) Scale bars = 0.1 mm.

some sclerites are missing in certain species. Between major groups, however, the correspondences are unclear. In Ramírez (2003) I made a thorough morphological examination of many representatives using an scanning electron microscope, expansions and dissec-

studies, especially the positional criterion, admitting that the election was to some extent arbitrary. Under an alternative schema of correspondences (Ramírez, 2003, p. 50), the C1 as identified in Gayennini could be homologous to the pPMA as identified in Amaurobioidini.

Pages Attachments Comments

8.27 x 10.87 in

8 of 25

AToL Spider project, as of today

31,588 images, **62,121** anatomical annotations

910 characters (processed 4395 chars. from 67 datasets)

450 ontology terms (to accommodate characters). Not only anatomy.

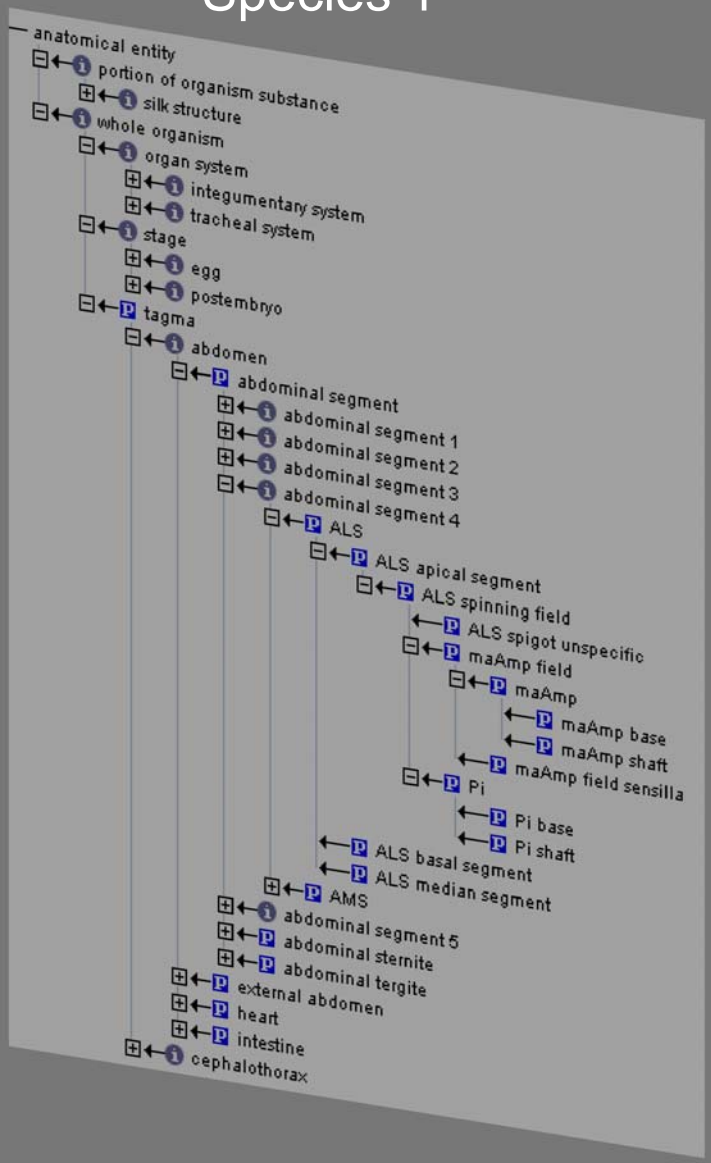
growing quickly…

Preparing migration to MorphBank

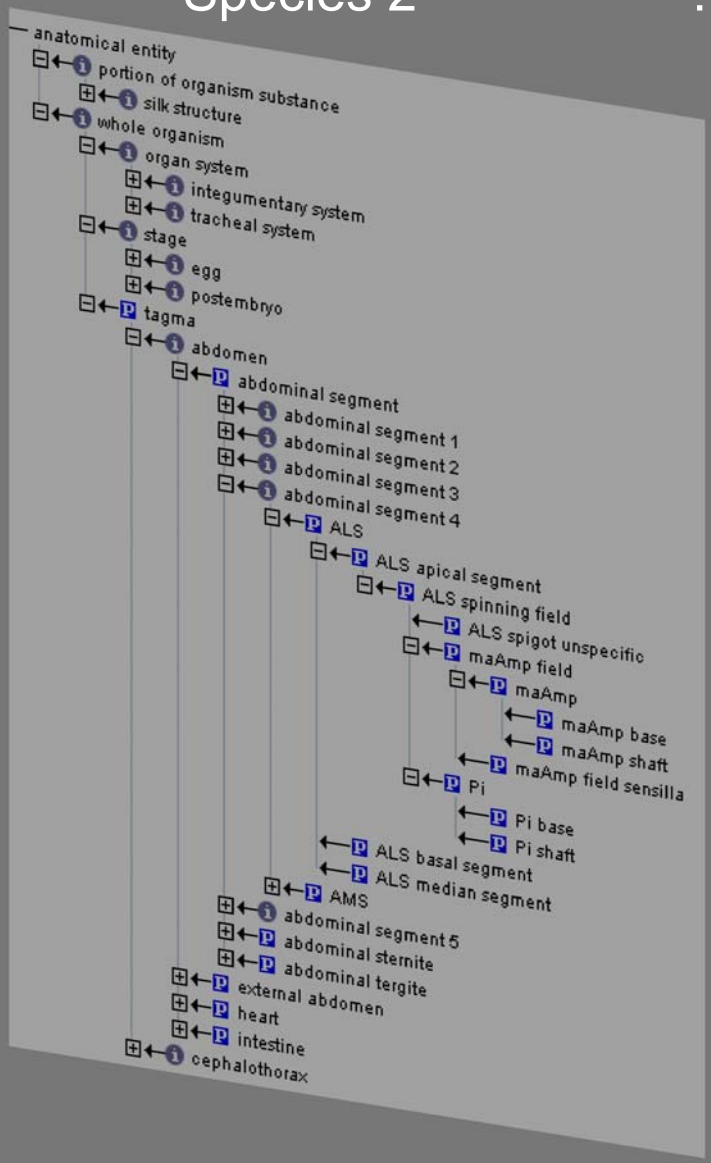
Documentation – Accumulation – Efficiency

Multi-species ontology

Species 1



Species 2



… Species n

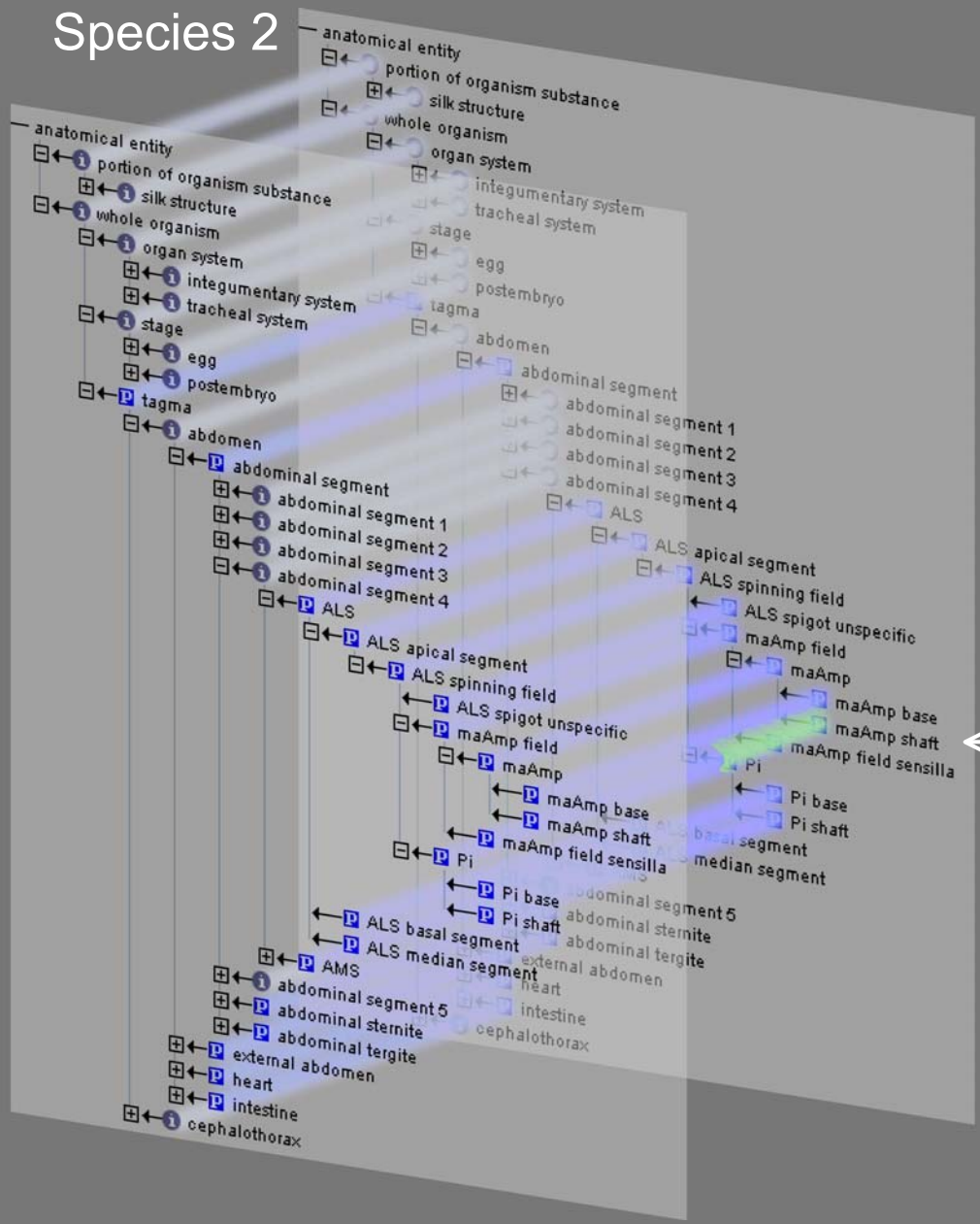
Aligned ontologies

Aligned by homology relations

Contrasting with single-species ontologies, homology is mandatory

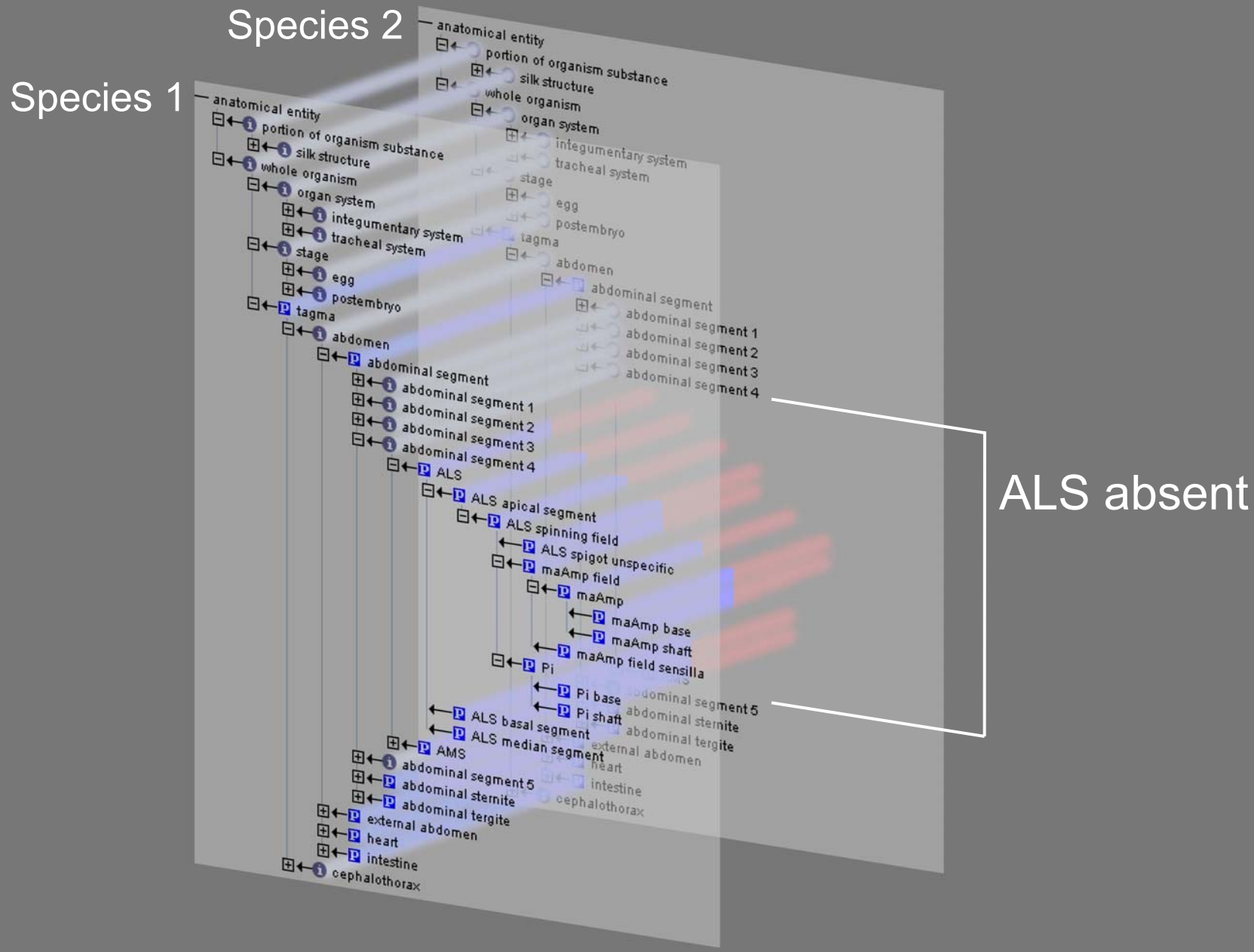
Otherwise, every term repeated for each species (or specimen!)

Transformation

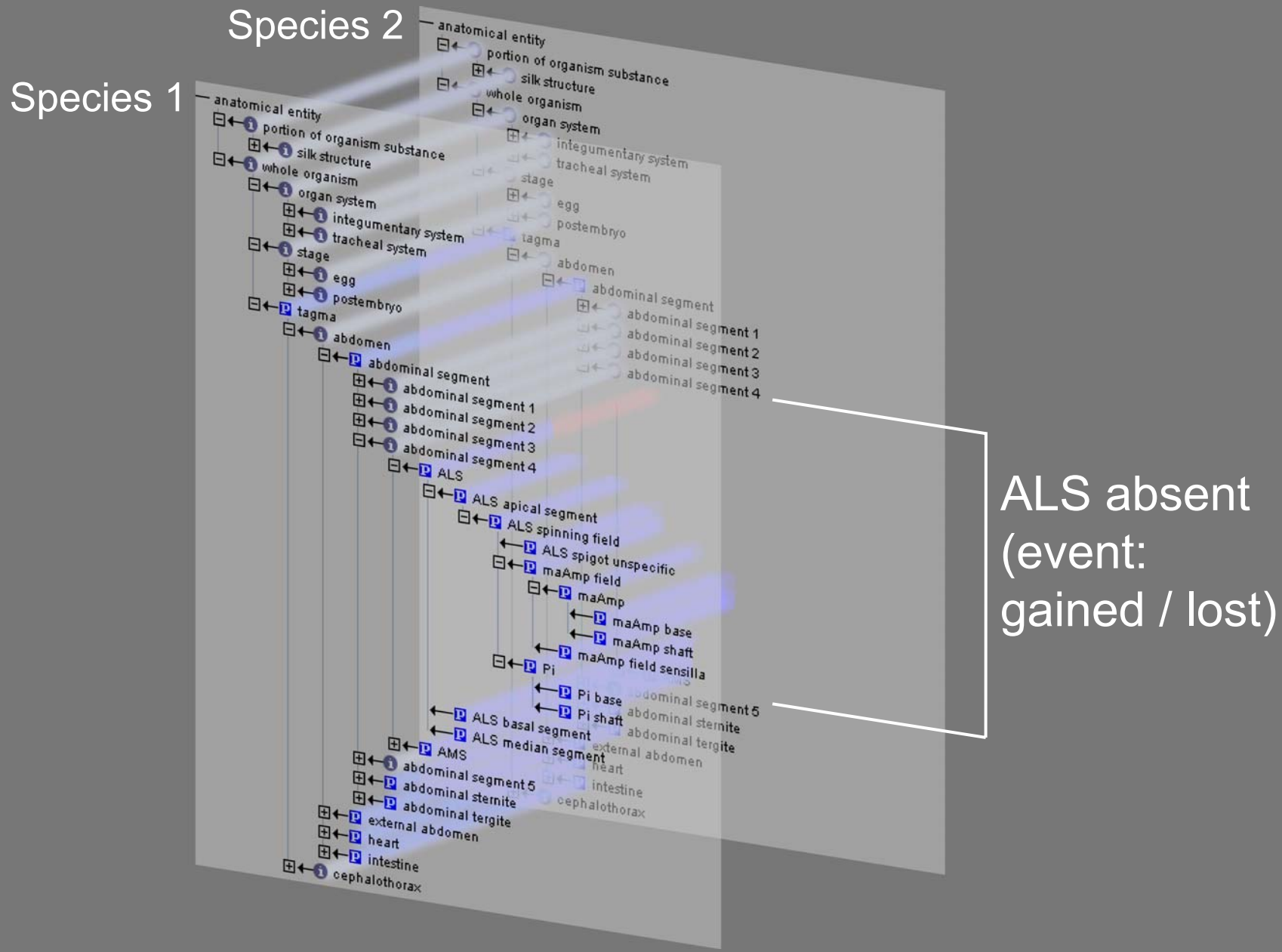


← ALS maAmp shaft transformed

Gains and losses



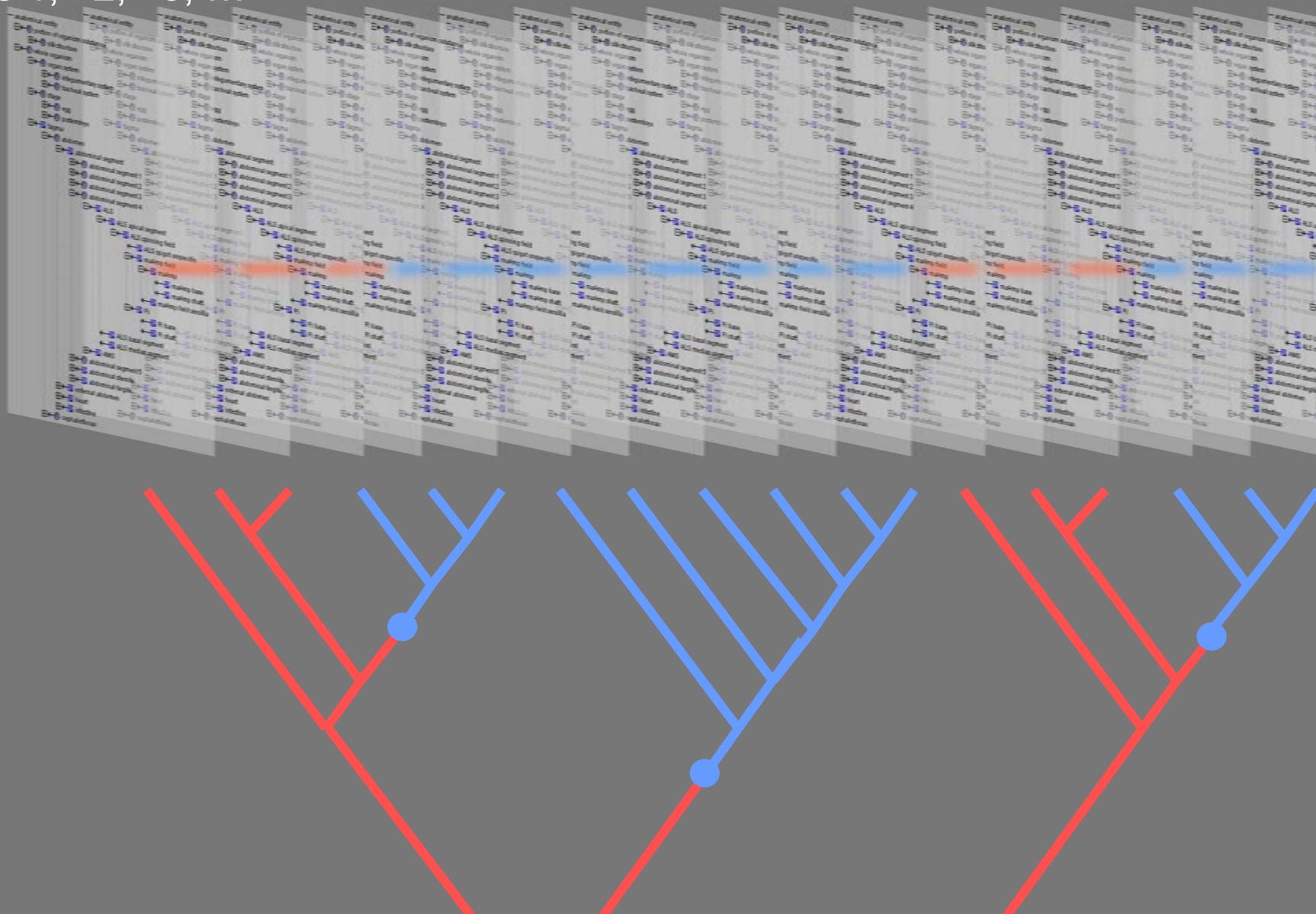
Gains and losses



ALS absent (event: gained / lost)

Many options for the “same” term…

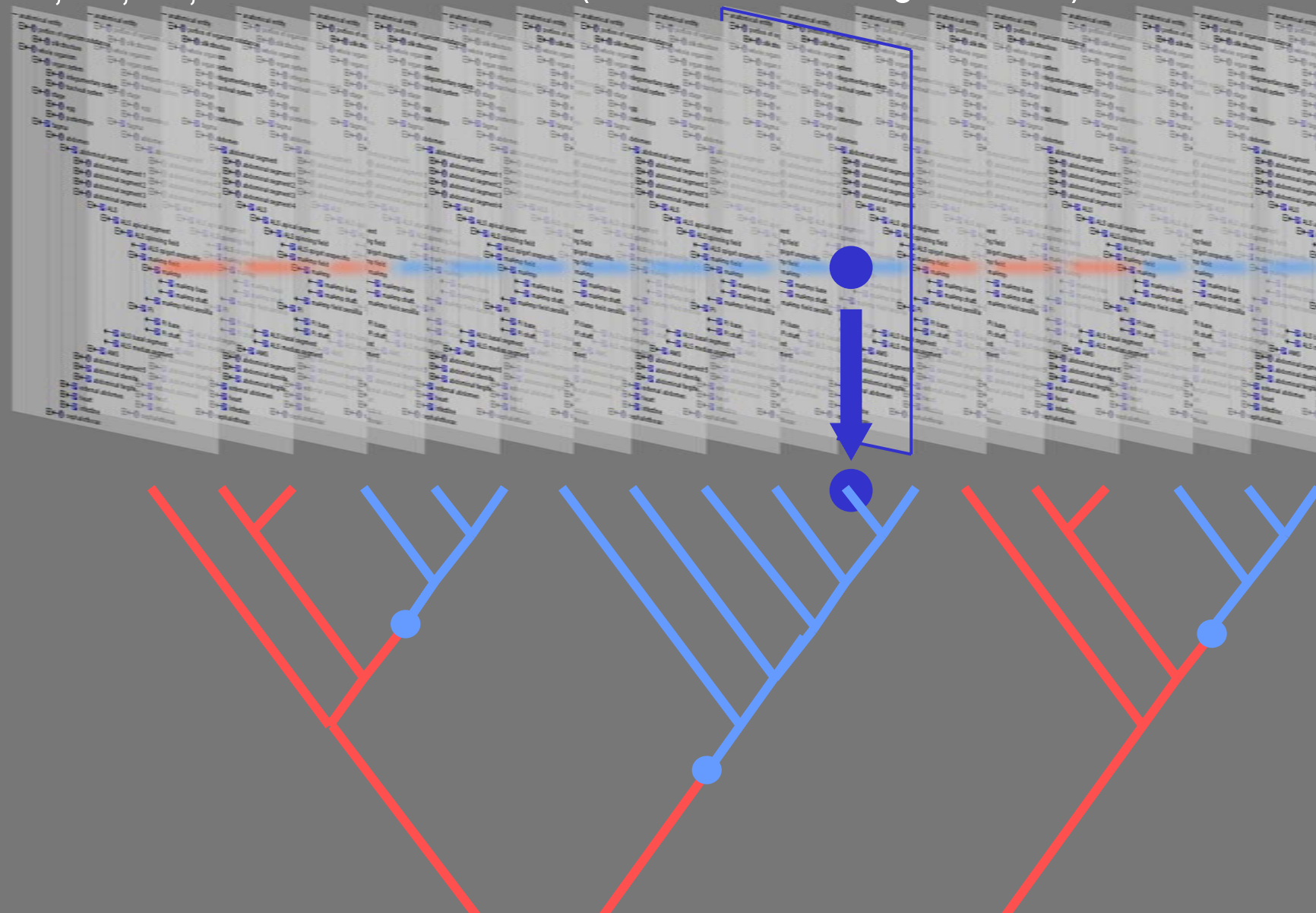
Species 1, 2, 3, …



Typification

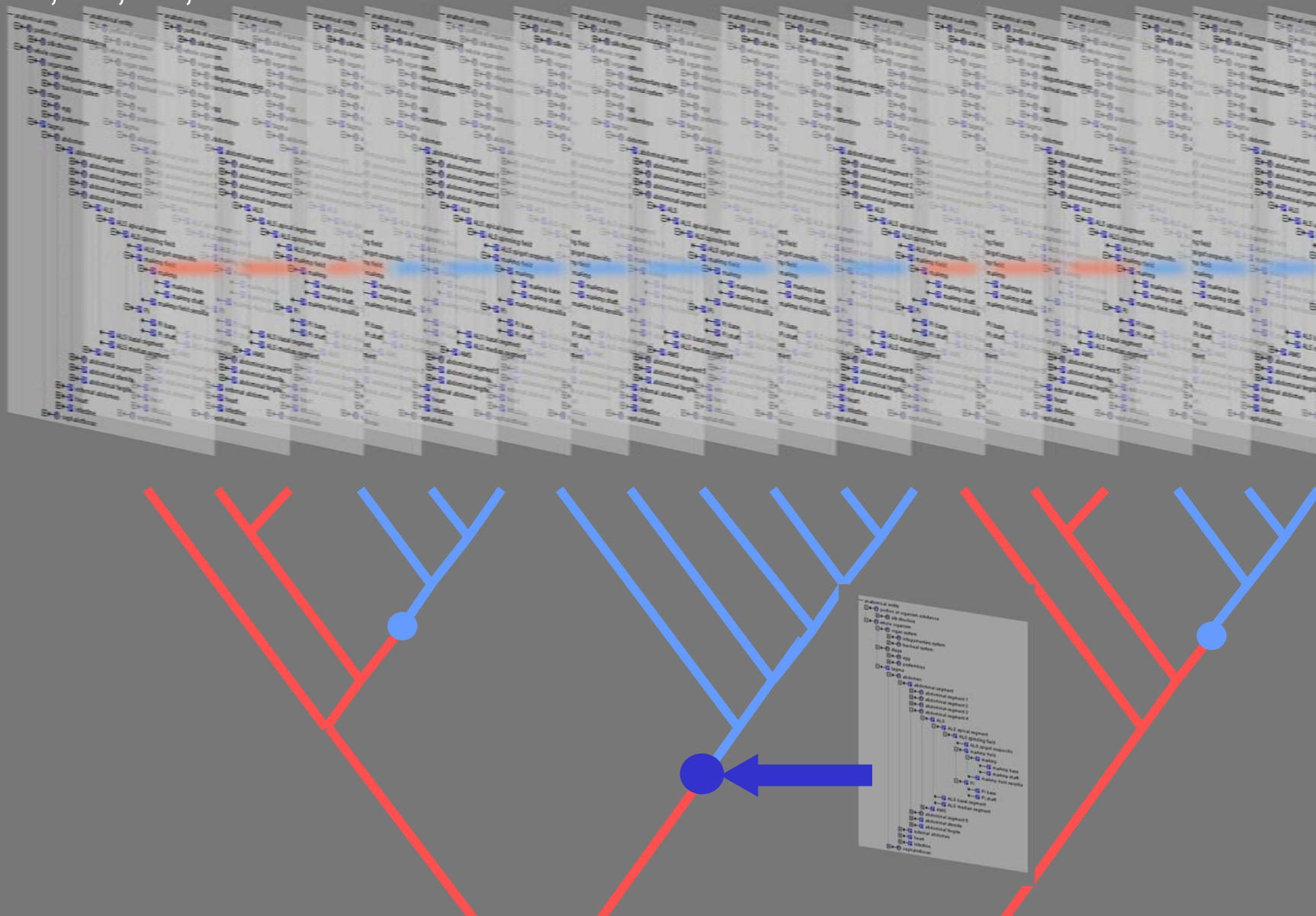
Species 1, 2, 3, …

Fix meaning of terms by choosing one species as reference (Scharff & Coddington 1997)

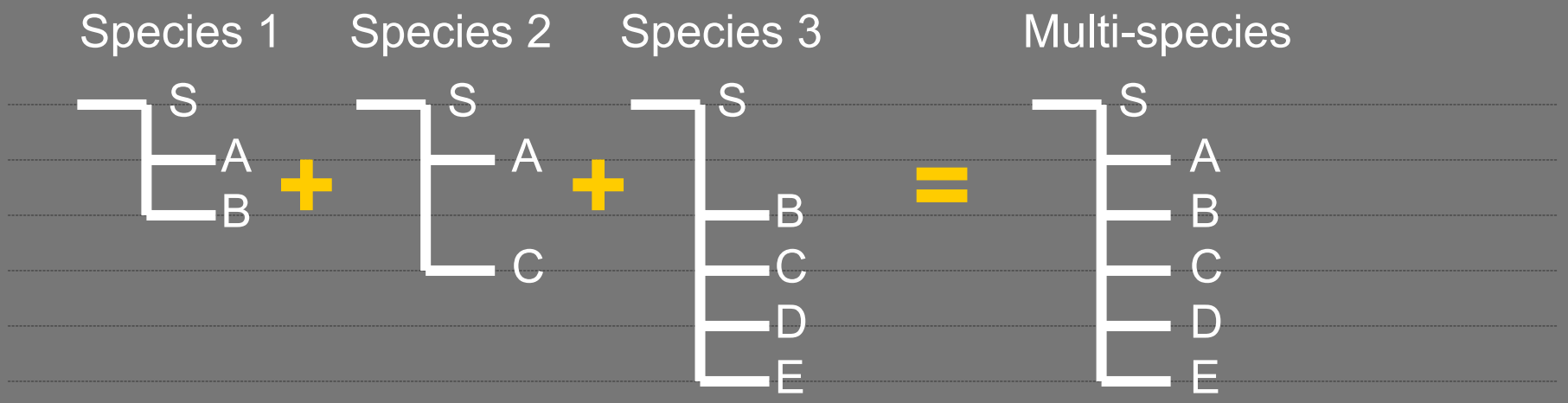


Future: Ancestor ontology

Species 1, 2, 3, …



Combinable components



Homology grid

No real organism has all this

Silk spigot types:
Pkf, Agg, Ac, Pi, maAmp, miAmp, Fg, Psg, MS, …

Open question: What if they cannot be combined?

Pragmatism

Many conceptual issues left unsolved. For example, are we using several different meanings for "is_a"?

Leg IV is_a Leg
Leg IV is_a (serial homolog of a) Leg

Macroseta is_a Seta
Macroseta is_a (modular homolog of a) Seta

Tracheal system is_a Organ system
Tracheal system is_a (granular level) Organ system

part_of… according to external vs. internal anatomy?
part_of… according to body segment, or to (segmented) organ?

Ontologies: What we systematists can get?

- Document and share our data
  - Image a second species, not one species for the second time
  - Community feedback
- Curation of our data
- Permanent repository
- Accumulation of knowledge
- Expert contribution from research on model organisms

What we systematists can contribute?

A science of diversity

www.tolweb.org

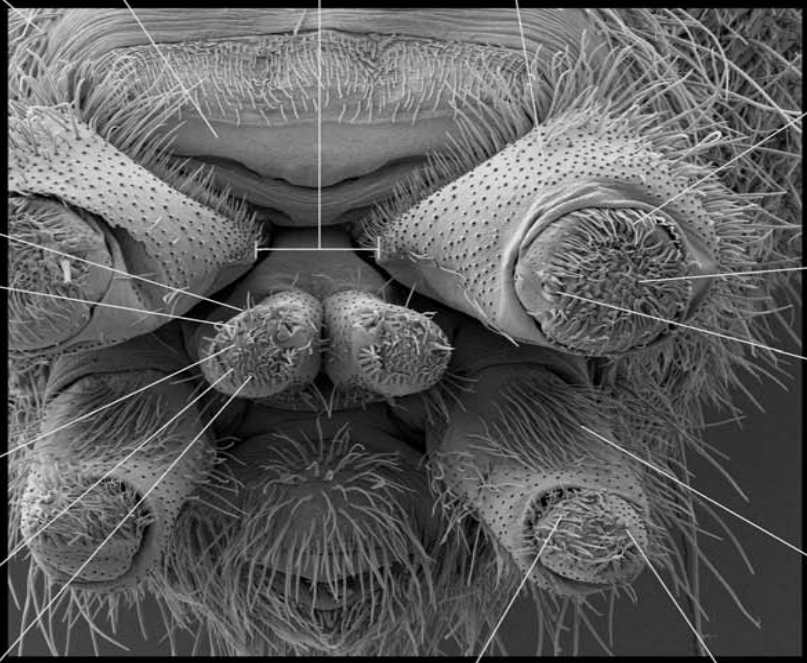
Diversity
Distinctiveness
Unique historical events

vs. generalization

A privileged access to diversity

Voucher localities for Dionycha study

The atomization of the organism in characters



spinnerets on abdominal tube

Cribellum
Cribellum divided
Cribellar spigots
clumps of cribellate spigots
Cribellum spigot morphology
CR spigot sorrounding cuticle
Colulus

ALS separation

ALS
ALS advanced Molycriinae
Specialized setae on ALS
ALS segments
ALS distal segment
ALS cyl large piriforms on collapsable mound ALS Club gnaph

Spin cuti (ALS if different)
tartipores
Spigot shaft surface
ALS Pi spigot base cuticle (Cteno131)
setae among spigots

ALS Piriform bases reduced
female piriforms larger than maAmp
Pi shaft continuous with base
separate maAmp field with some Pi
ALS piriforms number and other details
piriforms w elongate bases acc by plumose setae

female PMS spigots
PMS advanced Ammoxenidae

PMS paracribellar spigots
PMS paracribellar distribution
PMS paracribellar
PMS paracribellar shafts
PMS paracribellar bases
PMS paracribellar
PMS 'paracribelar'

ALS MAmp
Ampullate spigot shafts papillate
ALS MAmp in Pi field
ALS maAmp general #
ALS MAmp # female
female maAmp shaft sizes and details
ALS female Mamp cluster position
Position of MAP on ALS
ALS MAmp # male
maAmp field separate by furrow
ALS ma Amp on separate 'segment'

CY gland spigots
PMS CY three in one row large short shaft
PMS CY #
PMS cylindrical spigot field

female PMS Ac number
male PMS Ac number
PMS Ac shafts two size classes

PMS miAmp (c55 069)
female PMS miAmp #
male PMS miAmp #
miAmp on posterior median margin

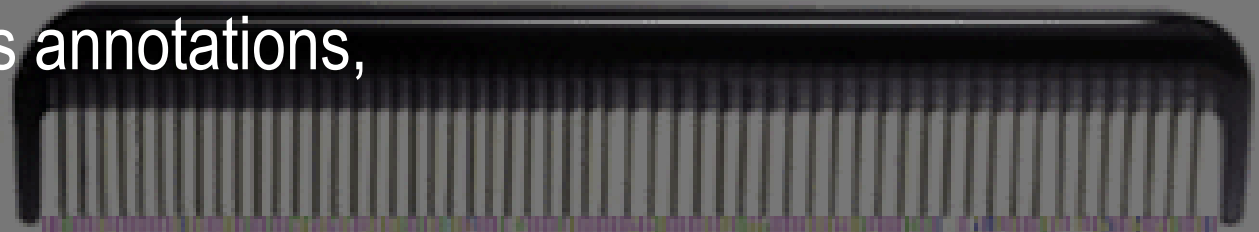
Spinneret claviform setae
PLS row of claviform setae
female PLS spigots
PLS short
PLS spigot base basally annulate flexible

PLS Cy number
PLS cyl Mimetidae indented

PLS modified spigot
PLS MS in male
PLS modified spigot position
PLS triad one shaft fused with modified
PLS triad nubbins in female
PLS MS accompanying spigots
types of PLS MS accompanying spigots

Combing across diversity

If scorings in phylogenetic datasets were processed as annotations,



Modest dataset
50 species, 100 characters
→ 5,000 annotations

Large datasets
150 species, 400 characters
→ 60,000 annotations

500 species, 2000 characters
→1,000,000 annotations

Future: Provide an environment for the sytematics community to map their data to ontologies

Plus: Homology annotations & combinatory

Summary

The ontology is a powerful schema to
- Annotate images
- Retrieve relevant images,
- open to more elaborate algorithms
- Organize workflow

Ontologies can bridge the communities of model organisms and of diversity

A common problem: The mechanistic causes of the diversity of morphology

Acknowledgments

Paula Mabee, Todd Vision, Monte Westerfield, Barry Smith

ATOL Spiders team, Chris Mungall, Michael Ashburner, Jim Balhoff, MorphBank and Morphster teams

CONICET, ANPCYT, NSF, NESCENT

MACN CONICET

NATIONAL CENTER FOR BIOMEDICAL ONTOLOGY

NESCent

NSF

Evolution 2008 University of Minnesota June 19 - 24