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

Links to and from phylogenetic dataset.
Developed ontology (independently) to solve treatment of images
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

1Museo Argentino de Ciencias Naturales “Bernardino Rivadavia”—CONICET, Avenida Angel Gallardo 470, C1405DJR, Buenos Aires, Argentina; E-mail: ramirez@macn.gov.ar
2Entomology, Smithsonian Institution, PO Box 37012, NMNH E529, NHB-105, Washington, DC 20013-7012, USA
3Departments of 4Zoology and 5Botany, 6270 University Boulevard, University of British Columbia, Vancouver, BC V6T 1Z4, Canada
4Division of Environmental Biology, Museum of Natural History, Central Park West at 79th Street, New York, NY 10024-5599

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?)
Character: Epigrum 1002694

Any sclerotized modification of the cuticle around the female genital region, e.g., facial chitin.

State 0 (abs)
Cuticle around genital region similar to surrounding. Type taxon: Archisphincta

State 1 (pres)
Cuticle around genital region sclerotized, contrasts to surrounding, may or may not have elaborate processes. Type taxon: Archisphincta

Source: n70 040 c55.115 n72.028
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

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

SV70
SPD:0000019
Images on the anatomical ontology
### Characters annotated with Anatomical IDs

#### Char ID – Anatomical IDs

<table>
<thead>
<tr>
<th>Character Description</th>
<th>Char ID</th>
<th>Group</th>
<th>Probability Model</th>
<th>OBO term ID</th>
<th>Standard View</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endite medial surface gland</td>
<td>ID000672</td>
<td>Endite</td>
<td>Mk1 (est)</td>
<td>ID000572</td>
<td>SV77</td>
</tr>
<tr>
<td>Maxillary gland position</td>
<td>ID003072</td>
<td>Endite</td>
<td>Mk1 (est)</td>
<td>ID000572</td>
<td>SV77</td>
</tr>
<tr>
<td>Palpal claw hairs</td>
<td>ID000730</td>
<td>Endite</td>
<td>Mk1 (est)</td>
<td>ID000730</td>
<td>SV242</td>
</tr>
<tr>
<td>Cusuples extending onto heel</td>
<td>ID000800</td>
<td>Endite</td>
<td>Mk1 (est)</td>
<td>ID000800</td>
<td>SV78</td>
</tr>
<tr>
<td>Endite ventral surface</td>
<td>ID000874</td>
<td>Endite</td>
<td>Mk1 (est)</td>
<td>ID000574</td>
<td>SV78</td>
</tr>
<tr>
<td>Palpal femoral prosternal area</td>
<td>ID000598</td>
<td>Paip</td>
<td>Mk1 (est)</td>
<td>ID000590</td>
<td>SV301</td>
</tr>
<tr>
<td>Palpal femoral pro-lateral surface</td>
<td>ID000590</td>
<td>Paip</td>
<td>Mk1 (est)</td>
<td>ID000590</td>
<td>SV301, SV00</td>
</tr>
<tr>
<td>Shunt medially bisected female palpal tarsus</td>
<td>ID000617</td>
<td>Paip</td>
<td>Mk1 (est)</td>
<td>ID000617</td>
<td>SV09</td>
</tr>
<tr>
<td>Female palpal tarsus chemosensory scopula on apical bunset</td>
<td>ID000922</td>
<td>Paip</td>
<td>Mk1 (est)</td>
<td>ID000922</td>
<td>SV103, SV37</td>
</tr>
<tr>
<td>Palpal tarsus apical setae</td>
<td>ID000922</td>
<td>Paip</td>
<td>Mk1 (est)</td>
<td>ID000922</td>
<td>SV103</td>
</tr>
<tr>
<td>Female palpal tarsus venal lateral setae</td>
<td>ID000623</td>
<td>Paip</td>
<td>Mk1 (est)</td>
<td>ID000623</td>
<td>SV103</td>
</tr>
<tr>
<td>Blunt seta at side of claw</td>
<td>ID000677</td>
<td>Paip</td>
<td>Mk1 (est)</td>
<td>ID000677</td>
<td>SV103</td>
</tr>
<tr>
<td>Female palpal tarsus dorsal chemosensory scopula</td>
<td>ID000619</td>
<td>Paip</td>
<td>Mk1 (est)</td>
<td>ID000619</td>
<td>SV103</td>
</tr>
<tr>
<td>Palpal tarsal tip</td>
<td>ID000642</td>
<td>Paip</td>
<td>Mk1 (est)</td>
<td>ID000642</td>
<td>SV103</td>
</tr>
<tr>
<td>Palpal claw size</td>
<td>ID000644</td>
<td>Paip</td>
<td>Mk1 (est)</td>
<td>ID000644</td>
<td>SV103</td>
</tr>
<tr>
<td>Shape of palpal claw apico truncate</td>
<td>ID000846</td>
<td>Paip</td>
<td>Mk1 (est)</td>
<td>ID000846</td>
<td>SV103</td>
</tr>
<tr>
<td>Palpal claw form</td>
<td>ID000810</td>
<td>Paip</td>
<td>Mk1 (est)</td>
<td>ID000810</td>
<td>SV103</td>
</tr>
<tr>
<td>Palpal claw reduced to stub</td>
<td>ID14101</td>
<td>Paip</td>
<td>Mk1 (est)</td>
<td>ID14101</td>
<td>SV103</td>
</tr>
<tr>
<td>Palpal claw both count</td>
<td>ID14102</td>
<td>Paip</td>
<td>Mk1 (est)</td>
<td>ID14102</td>
<td>SV103</td>
</tr>
<tr>
<td>Palpal tarsus muscle M28</td>
<td>ID003110</td>
<td>Paip</td>
<td>Mk1 (est)</td>
<td>ID003110</td>
<td>SV103</td>
</tr>
<tr>
<td>Endite sexual dimorphism (placeholder)</td>
<td>ID000646</td>
<td>Paip</td>
<td>Mk1 (est)</td>
<td>ID000646</td>
<td>SV78</td>
</tr>
<tr>
<td>Palpal tibia apophysis surface (MM)</td>
<td>ID002731</td>
<td>Paip</td>
<td>Mk1 (est)</td>
<td>ID002731</td>
<td>SV20, SV23</td>
</tr>
<tr>
<td>Palpal femoral retrolateral surface (MM)</td>
<td>ID001709</td>
<td>Paip</td>
<td>Mk1 (est)</td>
<td>ID001709</td>
<td>SV304, SV315</td>
</tr>
<tr>
<td>Palpal femoral pro-lateral surface (MM)</td>
<td>ID001710</td>
<td>Paip</td>
<td>Mk1 (est)</td>
<td>ID001710</td>
<td>SV304, SV315</td>
</tr>
<tr>
<td>Palpal tarsus apical seta</td>
<td>ID001711</td>
<td>Paip</td>
<td>Mk1 (est)</td>
<td>ID001711</td>
<td>SV304, SV315</td>
</tr>
<tr>
<td>Palpal femoral prosternal surface (MM)</td>
<td>ID001712</td>
<td>Paip</td>
<td>Mk1 (est)</td>
<td>ID001712</td>
<td>SV304, SV315</td>
</tr>
</tbody>
</table>

### Additional Information

- **Char ID** refers to the unique identifier for each anatomical feature.
- **Anatomical IDs** are linked to the Standard View ID and OBO term ID for further reference.
- **Standard View** provides a reference for additional context or documentation.
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)
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 …
Conductor (SPD:0000179) Sclerite arising on the tegulum, usually fitting part of the embolus.

Figs 4-7. Left copulatory bulbs of Amaurobioidini. 4. *Ganotea bitrata* Ramirez. 5. *Coptotrepes flavopilosus* Simon. 6. *Ferrierta echinata* Simon. 7. *Aytenoides cococo* Ramirez. (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 Ramirez (2003) I made a thorough morphological examination of many representatives using an scanning electron microscope. Expansions and dissec-

...
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

Species 1

Species 2

Aligned by homology relations

Contrasting with single-species ontologies, homology is mandatory

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

Species 1

Species 2

ALS maAmp shaft transformed
Gains and losses

Species 1

Species 2

ALS absent
Gains and losses

Species 1

Species 2

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

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

Diversity
Distinctiveness
Unique historical events

vs. generalization

www.tolweb.org
A privileged access to diversity

Voucher localities for Dionycha study
The atomization of the organism in characters

- Cribellum
- Cribellum divided
- Cribellar spigots
- Clumps of cribellate spigots
- Cribellum spigot morphology
- CR spigot surrounding cuticle
- Collulus

- ALS
- ALS advanced Molycrininae
- Specialized setae on ALS
- ALS segments
- ALS distal segment
- ALS cyl. large pillars on collapsable mound

- ALS Club graph

- Spin cuti (ALS if different)
- Tartigons
- Spigot shaft surface
- ALS Pi spigot base cuticle (Cteno131)
- Setae among spigots

- ALS Pinforms bases reduced
- female pinform larger than maAmp
- Pi shaft continuous with base
- separate maAmp field with some Pi
- ALS pinforms number and other details
- pinforms w. elongate bases acc. by plumose setae

- ALS MmaAmp
- Ampullate spigot shafts papillate
- ALS MmaAmp in Pi field
- ALS maAmp general #
- ALS MmaAmp # female
- female maAmp shaft sizes and details
- ALS female MmaAmp cluster position
- Position of MAP on ALS
- ALS MmaAmp # male
- maAmp field separate by furrow
- ALS maAmp on separate ‘segment’

- PLS Cy number
- PLS cyl. Mimetiidae indented

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

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

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

- female PMS miAmp
- male PMS miAmp
- miAmp on posterior median margin

- PMS miAmp (c55 069)

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

Plus: Homology annotations & combinatorial

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