Phenoscape: Extending model organism ontologies for devo-evo studies of evolutionary phenotypes

Paula Mabee
University of South Dakota
**Goal:** Connect these data using ontologies

- Model organisms (biomed)
  - Phenotypes
  - Genetics

- Comparative Evolutionary Biology
  - Morphology
  - Phylogeny
  - Homology
  - Taxa

**zebrafish** vs. **ostariophysan fishes**
Ontologies can connect human disease to candidate genes

**Animal models**

- Mutant gene
- Mutant or Missing Protein
- Mutant phenotype (disease model)

**Humans**

- Mutant gene
- Mutant or Missing Protein
- Mutant phenotype (disease)

BLAST

ONTLOGIES
Zebrafish → Human
Conservation of gene sequence & function

SLC24A5 involved in pigmentation similarity between fish and humans (Lamason et al., 2005)
Fly  Human
Conservation of developmental genes & toolkit

Pax6/eye/less gene directs development of fly and human eyes

Used with permission from Micromundi: http://remf.dartmouth.edu/images/
MicromundiImages/16chrysomelidae220.tif

Courtesy of Kevin Helenurm
Problem: Lack of genetic data for most species

- Mutant gene
- Mutant or Missing Protein
- Mutant phenotype (disease model)

Animal models

Evolutionary Species

? “Natural” phenotype
Ontologies can connect species variation to candidate genes

**Animal models**

- Mutant gene
- Mutant or Missing Protein
- Mutant phenotype (disease model)

**Evolutionary Species**

- Candidate gene
- “Natural” phenotype

**ONTOLOGIES**
Goals:

- identify candidate genes for evolutionary phenotypes
- mapping and identification of taxa in which specific features vary
- discover similar phenotypes and similar OTUs (phenotypic BLAST)
- correlation matrices of traits & genes
Phenoscape.org

- **History**: Communication between zebrafish model organism community and Cypriniformes Tree of Life group through NESCent workshops (Mabee-Westerfield)

- **Goal**: Create curated, ontology-based evolutionary phenotype database that maps to genetic databases

- **Generalizable system**: Prototype with ostariophysan fishes
Ostariophysi: Diverse, speciose, freshwater fishes including zebrafish
Use case: Basihyal lost in catfishes
(Basihyal primitively present)

Sample query: What are all the possible genes underlying evolutionary loss of the basihyal?
### Results:

<table>
<thead>
<tr>
<th>Genotype (Background)</th>
<th>Affected Genes</th>
<th>Phenotype</th>
<th>Figures</th>
</tr>
</thead>
<tbody>
<tr>
<td>brpf1 b936/b943</td>
<td>brpf1</td>
<td>absent, hypoplastic</td>
<td>2 figures from 1 publication</td>
</tr>
<tr>
<td>brpf1 b20x102/b200x02</td>
<td>brpf1</td>
<td>present</td>
<td>4 figures from 1 publication</td>
</tr>
<tr>
<td>brpf1 b25114/b25114</td>
<td>brpf1</td>
<td>absent, hypoplastic</td>
<td>2 figures from 1 publication</td>
</tr>
<tr>
<td>sox9a bw37/+</td>
<td>sox9a</td>
<td>aplastic</td>
<td>1 figure from 1 publication</td>
</tr>
<tr>
<td>sox9a bw37bw37</td>
<td>sox9a</td>
<td>aplastic</td>
<td>1 figure from 1 publication</td>
</tr>
</tbody>
</table>
Basihyal loss in zebrafish due to mutation in brpfl gene

Laue et al., 2008 - The multidomain protein Brpf1 binds histones and is required for Hox gene expression and segmental identity. Development 135(11):1935-1946 - Full text @ Development
New hypothesis:
-Evolutionary loss of basihyal in catfishes due to changes in brpfl or Hox gene

-Test by looking at predicted gene expression in catfish

In the brpfl zebrafish mutant, bapx1 is expressed in the 2nd pharyngeal arch
Requirements:

- Develop ontologies
- **Annotate evolutionary phenotypes**
  - using ontologies
  - using a syntax shared with model organism databases
- Develop curation software
Traditional comparative study: survey morphological variation

*Cyprinus carpio*  
*Pangio anguillaris*  
*Nemacheilus fasciatus*  
*Catostomus commersoni*  
*Gyrinocheilus aymonieri*  
*Phenacogrammus interruptus*
Morphological data in free text format

- not a computable format
- cannot be compared among species
- cannot be linked to underlying genetics
- cannot be reasoned across

APPENDIX 1: CHARACTERS USED FOR PHYLOGENETIC ANALYSIS

Unless otherwise indicated, terminology follows Ronquist (1995a) and Ronquist and Nordlander (1989). Transformation series hypotheses are given for multistate characters. Following each character is the character's consistency index and retention index on the preferred tree (Fig. 4). Observed character states are given in Table 3.

General Body Sculpture

1. Microsculpture on vertex, lateral surface of pronotum and mesoscutum: (0) absent, surface not dull (Figs. 9A–9D and 10A–10C); (1) present, linear, making the surface dull (not illustrated); (CI =1.00, RI = 1.00, goodness of fit (G-fit) = 10).

Head

2. Shape of head in anterior view: (0) rounded, approximately as high as broad (Figs. 8A, 8B, and 9C); (1) elongate, higher than broad (Figs. 8C, 8D, 9A, and 9B); (CI =0.25, RI = 0.82, G-fit = 5).

3. Relative position of eye: (0) close to ocelli, ratio of distance between compound eye and posterior mandibular articulation to distance between posterior ocellus and compound eye > 1.18 (Figs. 8B and 8C); (1) removed from ocellus, ratio < 1.13 (not illustrated); (CI = 0.20, RI = 0.50, G-fit = 4.3).

4. Size of ocelli: (0) small, ratio of maximum diameter of a lateral ocellus to shortest distance between lateral ocelli 0.22–0.40 (not illustrated); (1) large, ratio 0.44–0.65 (Figs. 8B and 8D); (CI = 0.11, RI = 0.62, G-fit = 2.7).

7. Shape of compound eyes in dorsal view: (0) rounded, distinctly protruding from the surface of the head, particularly anteriorly (Figs. 8B and 8D); (1) less rounded, not distinctly protruding from the surface of the head (not illustrated); (CI = 0.25, RI = 0.73, G-fit = 5).

8. Lateral frontal carina: (0) absent (Fig. 8D); (1) present (Fig. 8B, more easily seen in dorsal view); (CI = 0.50, RI = 0.50, G-fit = 7.5).

9. Hair punctures on lateral part of vertex: (0) indistinct or absent (Figs. 8B and 8D); (1) present, distinctly enlarged (not illustrated); (CI = 0.33, RI = 0.60, G-fit = 6).

10. Sculpture on posterior part of vertex (seen in dorsal view, not illustrated); (0) smooth or punctate, without linear component; (1) with parallel or slightly radiating, transverse striae (CI = 0.50, RI = 0.75, G-fit = 7.5).

11. Relative position of antennal sockets: (0) close to ocelli, ratio of vertical distance between inner margin of antennal foramen and ventral margin of clypeus to vertical distance between anterior ocellus and antennal rim < 2.6 (not illustrated); (1) intermediate, ratio 2.25–4.1 (Figs. 8B and 8D); (2) far from ocelli, ratio > 4.4 (not illustrated). Ordered 0.12; (CI = 0.08, RI = 0.33, G-fit = 1.2).

12. Vertical carinae adjacent to ventral margin of antennal socket: (0) absent (Fig. 8B); (1) present (Fig. 8D); (CI = 0.50, RI = 0.86, G-fit = 7.5).

13. Vertical delineations on lower face: (0) absent (Figs. 8B and 8D); (1) single carina or ledge (not illustrated); (2) several parallel or subparallel carinae (not illustrated). Unordered; (CI = 0.29, RI = 0.54, G-fit = 3.8).

14. (Subdivision of 13:1) Shape of single vertical delineation of lower face (not illustrated); (0) rounded, divergent ledges running from antennal sockets to dorso...
Phenoscape ontologies

New:

Teleost Anatomy Ontology
(2371 terms; 395 skeletal)

Teleost Taxonomy Ontology
(36,060 terms; 38,000 synonyms)

Taxonomic Rank Ontology
(8->31 terms)

Existing:

Phenotype and Trait Ontology
(1,075 terms)

Spatial Ontology
(106 terms)

Evidence Code Ontology

Existing:

Teleost Taxonomy Ontology
(36,060 terms; 38,000 synonyms)

Taxonomic Rank Ontology
(8->31 terms)
Requirements:

- Develop ontologies
- Annotate evolutionary phenotypes
  - using ontologies
  - using a syntax shared with model organism databases
- Develop curation software
Annotation

- Tagging or marking up something (image, text) with terms from one or more ontologies
Image annotated with ontology term

Teleost Anatomy
Ontology
TAO:0000316

Basihyal
Phenotypes annotated with ontology terms and EQ syntax

Entity
Caudal fin
ZFA: 0001058

Quality
Decreased size
PATO: 0000587

Teleost Anatomy Ontology: TAO
Phenotype and Trait Ontology: PATO

no tail mutant
Annotation of evolutionary phenotypes

**Free-text character:**

“Character 1: Parietal and supraoccipital fused (state 1) or separated (state 2)"

**Annotations:**

- Parietal fused_with supraoccipital
  - Parietal bone TAO:0000486
  - Fused with PATO:0000642
  - Supraoccipital bone TAO:0000595

- Parietal separated_from supraoccipital
  - Parietal bone TAO:0000486
  - Separated from PATO:0001505
  - Supraoccipital bone TAO:0000595
Requirements:

• Develop ontologies

• Annotate evolutionary phenotypes
  • using ontologies
  • using a syntax shared with model organism databases

• Develop curation software
Phenex: software for curation of evolutionary phenotypes
## Curation of phenotypic data

<table>
<thead>
<tr>
<th>Taxon</th>
<th>#Pubs</th>
<th>#Characters</th>
<th>#Species</th>
<th>Annotations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cypriniformes</td>
<td>10</td>
<td>725</td>
<td>930 (3,268)</td>
<td>674,250</td>
</tr>
<tr>
<td>Siluriformes</td>
<td>44</td>
<td>2582</td>
<td>2116 (2,867)</td>
<td>546,312</td>
</tr>
<tr>
<td>Characiformes</td>
<td>20</td>
<td>1661</td>
<td>595 (1,674)</td>
<td>988,295</td>
</tr>
<tr>
<td>Gymnotiformes</td>
<td>1</td>
<td>250</td>
<td>41 (134)</td>
<td>10,250</td>
</tr>
<tr>
<td>Gonorynchiformes</td>
<td>3</td>
<td>187</td>
<td>69 (37)</td>
<td>12,903</td>
</tr>
<tr>
<td>Clupeiformes</td>
<td>6</td>
<td>430</td>
<td>149 (364)</td>
<td>64,070</td>
</tr>
<tr>
<td>TOTAL</td>
<td>81</td>
<td>5960</td>
<td>4,000 (8,344)</td>
<td>7,225,780</td>
</tr>
</tbody>
</table>
Summary

• Evolutionary morphology and genetic data can be linked using ontologies

• **Phenoscape**: prototype of a generalizable system for making this connection

• Crosses disciplines, facilitates new discoveries through data mining
Acknowledgements

- NSF DBI0641025 for Phenoscape funding (Mabee, Vision, Westerfield); Lapp, Dahdul, Kothari, Balhoff, Midford, Lundberg

- National Evolutionary Synthesis Center NSF EF-0423641

- NIH HG002659 (to Monte Westerfield)

- Cypriniformes Tree of Life (NSF 0431290), colleagues and students (Mayden, Miya, Saitoh, He, Coburn, Arratia, Simon, Conway, Grey, Engeman, Bogutskya, Hilton, Aspinwall)

- Suzanenna Lewis, Chris Mungall (Lawrence Berkeley National labs)