Software and database resources for curation and management of evolutionary phenotypes

Jim Balhoff, Cartik Kothari, Hilmar Lapp, Todd Vision
US National Evolutionary Synthesis Center
Durham NC
Questions for a comparative anatomical knowledgebase

- What do we know about phenotypic variation for a given anatomical feature?

- What do we know about the phenotype of a given taxon?

- What sister taxa vary from each other in same way that a given genetic mutant varies from wild type?
Toward computable phenotypes

• Hard-code detailed semantics into a database application
• Capture expert knowledge in ontologies and let the computer do generic logical reasoning
Teleost taxonomy

Tree View
Tree view constructed based on is_a hierarchy

- Chordata
  - Craniata
    - Myxinomorphi
    - Vertebrata
      - Gnathostomata
        - Chondrichthyes
        - Osteichthyes
          - Actinopterygii
            - Chondrostei
            - Cladista
          - Neopterygii
            - Halecostomi
              - Amiiformes
              - Teleostei
                - Elopomorpha
                - Euteleostei
      - Ostariophys
        - Ostarioclupeomorpha
        - Osteoglossomorpha

Class/Type Details

General
Class/Type Name: Euteleostei
Id: TTO:254

Graph View
Graph Type: Local Neighborhood

Diagram showing the relationships between different classes and subclasses within the Teleost taxonomy, with Euteleostei at the top and various other classes branching out below it.
Informatics requirements for application to clade X

- **Ontologies**
  - Entities: X Anatomy Ontology
  - Qualities: Phenotype and Trait Ontology (PATO)
  - Taxonomy: X taxonomy ontology

- **Curation tool:** Phenex
  - Mutant phenotypes from zebrafish (14K in ZFIN)
  - Cypriniform “evolutionary phenotypes” from the literature

- **Ontology-driven database:** OBD
  - With a user interface tailored for evolutionary queries
Character Matrix and Entity-Quality syntax

- Character: Entity (dorsal fin) Attribute: shape Value: round
- Character State: Value: round
- Quality (PATO)

Entity (TAO)
Phenex data curation software

- Separates ontology curation from data curation
- User interface
  - Displays original character matrix-style assignments
  - Guides user to select appropriate ontology terms for EQ annotations
  - Incorporates references to literature and specimens
- Technical features
  - Loads current ontologies at startup
  - Generates NeXML files that can be easily loaded into the database
  - Platform independent
  - Open-source code base
**Term Info: ectopterygoid**

- **Basic Info**
  - Term: ectopterygoid
  - ID: TAO:0000656
  - Ontology: teleost_anatomy
  - Definition: The ectopterygoid is a dermal bone located at the anterior part of the palatoquadrate. It is first mentioned as a projection in the ventrolateral border of the vomer by means of a sheath of connective tissue.

- **Links (4)**
  - Parents
    - is_a: dermal bone
    - part_of: dermocranium, palatoquadrate_arch
  - Children
    - part_of: ectopterygoid_tooth

- **DBxrefs (2)**
  - ZFA:0000656
  - ZFIN:ZDB-ANAT-011113-581

**Phenotypes for State: 0 - ectopterygoid attached to the ventrolateral border of vomer by means of a sheath of connective tissue**

<table>
<thead>
<tr>
<th>Entity</th>
<th>Quality</th>
<th>Related Entity</th>
<th>Count</th>
<th>Measurement</th>
<th>Unit</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>ectopterygoid</td>
<td>attached to</td>
<td>vomer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ectopterygoid attached to the ventrolateral border of vomer by means of a sheath of connective tissue
The Phenoscape database

- Based on the Open Biomedical Database (OBD) model
  - Developed by Chris Mungall
- Generic (ie organism-agnostic)
- Represents both the ontologies and data annotations
- Exploits logical structure of the ontologies
  - Querying annotations of “dermal bone” will return annotations of “ectopterygoid”
Nodes

<table>
<thead>
<tr>
<th>node_id</th>
<th>ontology term</th>
<th>label</th>
</tr>
</thead>
<tbody>
<tr>
<td>548</td>
<td>PATO:0000462</td>
<td>&quot;absent from organism&quot;</td>
</tr>
<tr>
<td>10373</td>
<td>TTO:1001979</td>
<td>&quot;Danio rerio&quot;</td>
</tr>
<tr>
<td>16249</td>
<td>TTO:101040</td>
<td>&quot;Danio&quot;</td>
</tr>
<tr>
<td>53</td>
<td>OBO_REL:part_of</td>
<td>&quot;part_of&quot;</td>
</tr>
<tr>
<td>77</td>
<td>OBO_REL:is_a</td>
<td>&quot;is_a&quot;</td>
</tr>
<tr>
<td>48988</td>
<td>PHENOSCAPE:exhibits</td>
<td>&quot;exhibits&quot;</td>
</tr>
</tbody>
</table>

Links

<table>
<thead>
<tr>
<th>node_id</th>
<th>predicate</th>
<th>node_id</th>
</tr>
</thead>
<tbody>
<tr>
<td>10373</td>
<td>77</td>
<td>16249</td>
</tr>
<tr>
<td></td>
<td>--&gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&quot;Danio rerio is_a Danio&quot;</td>
<td></td>
</tr>
<tr>
<td>10373</td>
<td>48988</td>
<td>109224</td>
</tr>
<tr>
<td></td>
<td>--&gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&quot;Danio rerio exhibits absence in a process which is part of infraorbital 1&quot;</td>
<td></td>
</tr>
</tbody>
</table>
User interface

• Search for anatomical entity
  ▪ What do we know about phenotypic variation for a given anatomical feature?

• Search for taxon
  ▪ What do we know about the phenotype of a given taxon?

• Search for gene
  ▪ What sister taxa vary from each other in same way that a given genetic mutant varies from wild type?
Informatics requirements for application to clade X

- **Ontologies**
  - Entities: **X Anatomy Ontology**
  - Qualities: **Phenotype and Trait Ontology (PATO)**
  - Taxonomy: **X taxonomy ontology**

- **Curation tool:** **Phenex**
  - Mutant phenotypes from zebrafish (14K in ZFIN)
  - Cypriniform “evolutionary phenotypes” from the literature

- **Ontology-driven database:** **OBD**
  - With a user interface tailored for evolutionary queries
Acknowledgements

- **Phenoscapers**
  - Wasila Dahdul, John Lundberg Paula Mabee, Peter Midford, Monte Westerfield
  - Our data curators, workshop participants and software testers

- **BBOP & ZFIN**
  - Suzi Lewis, Mark Gibson, Melissa Haendel, Chris Mungall, Nicole Washington

- **Funding from the National Science Foundation**