**Curation of evolutionary literature**

Comparative anatomical studies form the historical basis for evolutionary hypotheses among organisms, yet the data have remained in free-text format in systematic monographs and journal articles. The Phenoscape project is modeling its curation effort after the model organism community by using ontologies and adopting entity-quality syntax to annotate the evolutionary phenotypes described in the literature. This ensures interoperability and enables computational analysis among evolutionary studies and between computational analysis among the ontologies and adopting entity-quality model organism community.

**Prototyping with fishes:**

We are prototyping the approach by curating the evolutionary phenotypes described for Ostariophysi, a clade of teleost fishes containing zebrafish and approximately 5,000 other species.

**Phenoscape knowledge-base:**

We are developing a prototype database and web interface for enabling cross-domain queries to the zebrafish (ZFIn) database.

**Candidate genes for evo-devo**

Sharing standards across domains is critical in order to meet the needs of cross-domain queries such as those from "evo-devo," e.g. identifying candidate genes for evolutionary phenotypes, candidate taxa for mutant phenotypes, and correlations across data types.

**Phenoscape database and web services**

The knowledge-base is stored in a relational database using the OBD schema – which combines the ontologies and annotations into a unified set of statements. Semantics in the ontologies allow the OBD reasoner to generate additional implied annotations.

**Phenoscape web user interface**

The web user interface uses publicly available web services to access the knowledge-base.

**Embedding EQ into traditional character data via NeXML**

• Systematic data are commonly stored within the familiar NEXUS file format for character matrices.
• NeXML is a new XML version of NEXUS – it is more robust, and easier to process and integrate custom annotations, such as ontologies and EQ phenotypes.
• EQ data are stored using the PhenexXML format, embedded within NeXML.
• Trait TTO identifiers are stored as custom annotations in taxa and the taxon block.
• Ontology annotations should be safely round-tripped within NeXML by upcoming versions of Mesquite, while free-text data can be edited at any time within either Phenex or Mesquite.

**The unique challenges faced in annotating these evolutionary data have required the development of new resources, tools and extensions to existing ones. Phenex is a software application developed by Phenoscape and hosted by the OBD Project for the curation of evolutionary data using ontologies. Data annotated with Phenex is based on the Entity! Quality (EQ) model for representing phenotypes. Phenex provides facilities for loading ontologies and easily selecting terms.**

**Character/character state-oriented Phenex interface**

• Browse the list of taxa, character states and taxa directly from a character matrix.
• Computed EQ phenotypes descriptions – ontological descriptions of character states – using anatomy and phenotype ontologies.
• Choose taxa directly from the Teleost Anatomy Ontology (TAO).

**Ontologies**

**Teleost anatomy ontology**

The Teleost Anatomy Ontology (TAO) is a multi-species anatomy ontology for teleost fishes that was developed by our group in recent years; it includes about 1000 terms describing the anatomical structure of fishes, such as those on the skeleton system, since it is the focus of evolutionary studies in osteichthyan fishes as well as genetic studies of zebrafish. See related poster for more information.

**Phenotype to taxon mapping (matrix)**

**Phenotype and trait ontology**

The Teleost Anatomy Trait Ontology (TATO) is an ontology of anatomy, that is used by multiple communities to annotate phenotypes.

**Teleost taxonomy ontology**

The Teleost Taxonomy Ontology (TTO) is a tree of taxonomic terms, connected by the relation is_a, and a set of rank terms (e.g., order, family, etc.) which are associated with taxonomic terms with a special has_rank relation. TTO is generated from a database dump from the Catalog of Fishes (CoF), an expert database. We have also added, via manual curation, anatomical queries from several area experts in Sturgeons, Chondrichthyes, and Gymnotiformes as well as several fossil taxa. This ontology currently contains 5000 species and over 15000 taxonomic terms.

**Phenoscape database**

The web user interface can be used to store additional implied annotations.

**Phenoscape web user interface**

The web user interface will allow users to search and browse annotations in the knowledge-base, via anatomical part, taxon, gene, or publication.

**Corresponding evolutionary and model organism phenotypes are linked due to use of related anatomical and quality terms.**