# PHENOSCAPE: Curation of evolutionary phenotypes using the MOD model



Paula Mabee<sup>1</sup>, Wasila Dahdul<sup>1,2</sup>, John Lundberg<sup>3</sup>, Peter Midford<sup>4</sup>, Todd Vision<sup>2,5</sup>, Hilmar Lapp<sup>2</sup>, Jim Balhoff<sup>2</sup>, Cartik Kothari<sup>2</sup>, Monte Westerfield<sup>6</sup>







<sup>2</sup>The University of South Dakota; <sup>2</sup>National Evolutionary Synthesis Center (NESCent); <sup>3</sup>Academy of Natural Sciences, Dept. Ichthyology; <sup>4</sup>University of Kansas; <sup>5</sup>University of North Carolina at Chapel Hill; <sup>6</sup>University of Oregon, Zebrafish Information Network (ZFIN)

# 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 evolutionary biology and the genomic, genetic, and developmental data curated by model organism databases.



# 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 9,000 other species.

#### **Evolutionary phenotypes** (Characters and states)

#### Phenotype to taxon mapping (matrix)

	ta matrix of Taxa Se									
	ta enarest or 1202 to L'apomorphic states									
	rationán hant. L. h									
_										
		1-5	6-13	11-15	36-23	21-25	26-38	71-35	36-40	41-4
	1/48/things	11322	27500	00077	27997	00011	64200	00000	19067	.0001
2	Armie cultur	00127	00000	00000	00000	00000	00330	00003	20000	0000
	1.Amia pallorioni	00000	00000	11900	50050	20000	00330	00000	22007	0000
	*Anashalion	00000	20000	21909	27910	11000	C4118	00000	10000	0000
							06707		79777	0000
				69977	20070					0000
			00917	00077	20010		00000		10007	0000
۸.	Chance	00011	21100	COULT	21033	00011	04200	00000	10011	1000
		77000	00911	11002	77770	00000	00338	00000	10077	0000

#### Taxa (extant and extinct)

Arratia, 1999 Zoological Journal of the Linnean Society 151:691-757

# **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 September 2007. Development of the TAO is currently centered on the skeletal system, since it is often the focus of evolutionary studies in ichthyology as well as genetic studies in zebrafish. See related poster for more information.

# Phenotype and trait ontology

The Phenotype and Trait Ontology (PATO) is an ontology of qualities 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. This ontology is generated from a database dump from the Catalog of Fishes (CoF), an expert database. We have also added, via manual curation, taxonomic updates from several area experts in Siluriformes, Characiformes, and Cypriniformes as well as several fossil taxa. This ontology currently contains 30,800 species and over 35,000 tota

# Phenex for evolutionary data curation



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 OBO 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 stateoriented Phenex interface

Browse the list of characters, character states and taxa directly

 Compose EQ phenotype descriptions - ontological descriptions of character states - using anatomy and phenotype ontologies

• Choose taxa directly from the Teleost Taxonomy Ontology (TTO)

## Embedding EQ into traditional character data via NeXML Systematic data are commonly stored within the familiar NEXUS file format for

NeXML is a new XML version of NEXUS - it is more robust, and easier to process

and integrate custom annotations, such as ontology-based EQ phenotypes
• EQ data are stored using the PhenoXML format, embedded within NeXML Taxon TTO identifiers are stored as custom annotations to taxa in the taxa 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





## Phenoscape knowledge-base: We are developing a prototype database and web interface for enabling cross-domain gueries to the zebrafish (ZFIN) database

Taxon	# Papers	# Species	# Characters		
Cypriniformes	10	930	725		
Siluriformes	44	2116	2,582		
Characiformes	20	595	1,661		
Gymnotiformes	1	41	250		
Gonorynchiformes	3	69	187		
Clupeiformes	6	149	430		
TOTAL	81	4.000	5.960		

Curation of evolutionary characters will result in >8 million phenotype annotations



Sharing standards across domains is critical in order to meet the needs of cross-domain gueries 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 Web services annotations into a unified set of statements. Semantics in OBD Java API the ontologies allow the OBD reasoner to generate additional implied annotations.

The web user interface

OBD database

# 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.

Phenoscape Links

Homepage: http://www.phenoscape.org Blog: http://blog.phenoscape.org Email discussion: phenoscape-discuss@lists.sourceforge.net

Acknowledgements We thank NSF DBI 0641025; NIH HG002659 and the National Evolutionary Synthesis Center (NESCent) NSF #EF-0423641 for funding.