

# Phenoscape: Use Cases and Anatomy Ontology Requirements for Linking Evolutionary and Model Organism Phenotypes

Wasila Dahdul<sup>1,2</sup>, James Balhoff<sup>2,3</sup>, Hilmar Lapp<sup>2</sup>, Peter Midford<sup>2</sup>,  
Todd Vision<sup>2,3</sup>, Monte Westerfield<sup>4</sup>, Paula Mabee<sup>1</sup>

<sup>1</sup>University of South Dakota, Vermillion, SD, USA; <sup>2</sup>National Evolutionary Synthesis Center, Durham, USA;

<sup>3</sup>University of North Carolina, Chapel Hill, USA; <sup>4</sup>University of Oregon, Eugene, USA

The naturally occurring phenotypes documented for groups of species in the systematics literature are recorded in free text which, although usefully precise and expressive, is not amenable to computational processing. This prevents its use for large-scale analysis or integration with the genetics knowledge that is available for single species in model organism databases [1,2]. Using ontologies, Phenoscape developed a knowledgebase (PhenoscapeKB; kb.phenoscape.org) that connects phenotypes for fish species with genetically characterized phenotypes for the zebrafish, *Danio rerio*, from the ZFIN database (zfin.org). Through ontology-based reasoning over expert knowledge in taxonomy, comparative anatomy and developmental genetics, the PhenoscapeKB is designed to enable the discovery of candidate genes for the natural diversity of phenotypes across taxa, and the aggregation of phenotypes across systematics studies to enable a global view of phenotype data available in large clades. These use cases necessitated the development of two multispecies anatomy ontologies: one for fishes (~30,000 species) and recently one for all vertebrates (~50,000 species), which will facilitate the expansion of Phenoscape to all vertebrates by connecting to other existing vertebrate ontologies and databases (amphibian, mouse, frog). These ontologies are required to fully represent the diversity of structures present in these extinct and extant species. It also required bridging across multiple scales of biological organization, from cells to anatomical systems. We discuss the implications that these requirements have for ontology design.

We developed the multispecies Teleost Anatomy Ontology (TAO) [3], with feedback from the ichthyological community, to represent the diversity of structures in teleost fishes and

for the phenotype curation of the fish systematics literature [4]. TAO was derived from, and is kept synchronized with, the species specific Zebrafish Anatomical Ontology (ZFA), which is used to describe phenotypes for the model organism *Danio rerio* in ZFIN. In expanding a single species ontology to one that is applicable to many species, existing term definitions were broadened to be universally applicable to all teleosts. New terms were added for structures present in teleosts but not in zebrafish, and this required the addition of grouping terms to TAO, which facilitate queries on similar structures for all fishes. Some of these grouping terms were unnecessary for ZFA (e.g., ‘tooth’ as a grouping term was not required in ZFA because zebrafish only have one type of tooth, represented by the term ‘ceratobranchial 5 tooth’). Some relationships between terms inherited from ZFA were removed from TAO because they were not applicable to all teleosts (e.g., the assertion ‘vertebra 1’ is\_a ‘Weberian vertebra’ is valid for ZFA but not for TAO because not all teleost vertebra 1 are subtypes of Weberian vertebra). These taxonomically variable relationships have implications for deriving a single species ontology from a multispecies ontology, and vice versa, because relationships that might be required for a single species may be invalid in a multispecies context. We are exploring how to represent these relationships, possibly as annotations with *in\_taxon* relationships to terms from a taxonomy ontology.

To initiate the expansion of Phenoscape to all vertebrates, we invited experts to a workshop to reevaluate and redefine existing skeletal terms (cells, tissues, development, anatomical structures) for their applicability across vertebrates, and to create terms for concepts not yet represented in existing

ontologies. This work resulted in the Vertebrate Anatomy Ontology (VAO), which will serve as a reference ontology for new and existing vertebrate subontologies. VAO accommodates the various ways that biologists classify bones and cartilages, e.g., as distinct elements and tissue types and based on developmental and locational criteria. Textual definitions are included for all terms, and an effort is currently underway to translate these text definitions into computable logical definitions. Logical definitions assist in ontology maintenance and error checking because a reasoner is used to automatically classify terms. Logically defined terms take the form of genus-differentia definitions, in which “X is a G that D” where X is the term being defined, G is the genus, and D is the differentia. In the example below, ‘replacement element’ (X) is a type of skeletal element (G) that participates in ‘replacement ossification’ (D; this term will be requested from the Gene Ontology and imported into VAO):

```
[Term]
id: VAO:0000135
name: replacement element
def: "Skeletal element that forms
    as a replacement or substitution
    of another element or tissue."
[VAO:curator]
intersection_of:
    VAO:0000128 ! skeletal element
intersection_of: participates_in
    VAO:0000140 ! replacement
    ossification
```

Synthesis and discovery in combined genetic, developmental and evolutionary data requires anatomy terms to be related across

scales of biological organization. The VAO enables this discovery by relating tissues to cells (e.g., ‘osteocyte’ part\_of ‘cellular bone tissue’) and structures to processes (e.g., ‘replacement element’ participates\_in ‘replacement ossification’). Putting in place these relationships to biological data at different scales will significantly increase the potential for discovery of candidate genes and taxa from queries on anatomical structure.

## References

1. Mabee, P. M., G. Arratia, M. Coburn, M. Haendel, E. J. Hilton, J. G. Lundberg, R. L. Mayden, N. Rios, and M. Westerfield. 2007. Connecting evolutionary morphology to genomics using ontologies: A case study from Cypriniformes including zebrafish. *Journal of Experimental Zoology Part B-Molecular and Developmental Evolution* 308B:655-668.
2. Mabee, P., M. Ashburner, Q. Cronk, G. Gkoutos, M. Haendel, E. Segerdell, C. Mungall, and M. Westerfield. 2007. Phenotype ontologies: the bridge between genomics and evolution. *Trends in Ecology & Evolution* 22:345-350.
3. Dahdul, W. M., J. G. Lundberg, P. E. Midford, J. P. Balhoff, H. Lapp, T. J. Vision, M. A. Haendel, M. Westerfield, and P. M. Mabee. 2010. The Teleost Anatomy Ontology: anatomical representation for the genomics age. *Systematic Biology* 59:369-383.
4. Dahdul, W. M., J. P. Balhoff, J. Engeman, T. Grande, E. J. Hilton, C. Kothari, H. Lapp, J. G. Lundberg, P. E. Midford, T. J. Vision, M. Westerfield, and P. M. Mabee. 2010. Evolutionary characters, phenotypes and ontologies: curating data from the systematic biology literature. *PLoS ONE* 5:e10708.