Some features of ontologies

- Controlled vocabulary (but oh so much more)
  - Enable machine communication
  - Can be used to annotate data
- Logically defined relationships between terms
  - Enable logical reasoning
  - Expose data to generic query and analysis tools
- Serve as a community representation of knowledge

PHENOSCAPE

- PIs: P. Mabee, T. Vision, M. Westerfield
- J. Balhoff, W. Dahdul, M. Haendel, C. Kothari, S. Lewis, C. Mungall, J. Lundberg, P. Midford
- Dozens of contributors to ontologies and curation

CToL ZFIN DeepFin

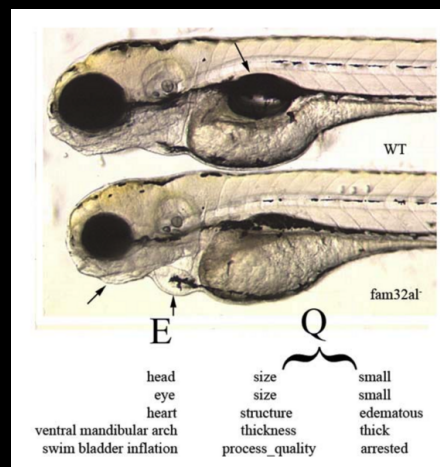THE NATIONAL CENTER FOR BIOMEDICAL ONTOLOGY NESCent NSF

Phenotypes as structured text
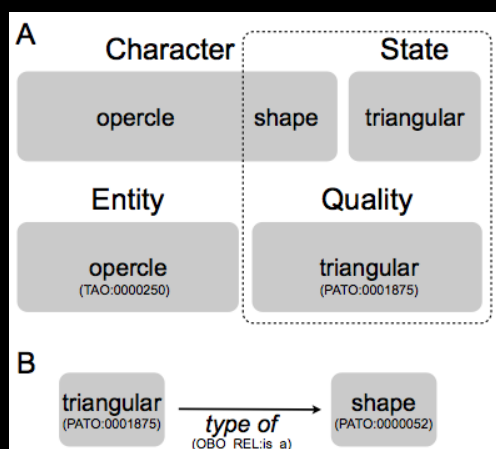
APPENDIX 1. CHARACTER SUMMARY

1. Fifth infraorbital. 0, well developed, without contact between fourth and sixth infraorbitals; 1, greatly reduced, with posteroventral margin of sixth infraorbital in contact with posterodorsal margin of fourth infraorbital.
2. Antorbital-lateral ethmoid contact. 0, no contact; 1, antorbital contacting ventral wing of lateral ethmoid along its entire lateral edge.
3. Antorbital. 0, flat, platelike, without medial process; 1, with a short medial, vertically aligned process at its posterior edge that extends along posterior surface of ventral wing of lateral ethmoid; 2, with enlarged medial, vertically aligned process at its posterior edge that extends along posterior surface of ventral wing of lateral ethmoid.
4. Mesethmoid spine. 0, conical, or with a dif-

    tilaginous surface at posterior portion of main body of vomer.

14. Portion on vomer for articulation of maxilla. 0, not modified in 1; 1, Presence of a shallow depression on its anterolateral surface where anterior tip of maxilla abuts.
15. Ridge on lateral surface of vomer. 0, absent; 1, present.
16. Rhinosphenoid. 0, present; 1, absent.
17. Lateral ethmoid-orbitosphenoid contact. 0, absent; 1, present.
18. Parasphenoid and main portion of orbitosphenoid. 0, well separated; 1, close to each other.
19. Dilatator fossa. 0, not extending anteriorly on dorsal surface of frontal or if so, only to dorsoposterior edge of orbit; 1, highly developed, extending anteriorly on dorsal surface of frontal beyond dorsoposterior edge of orbit.
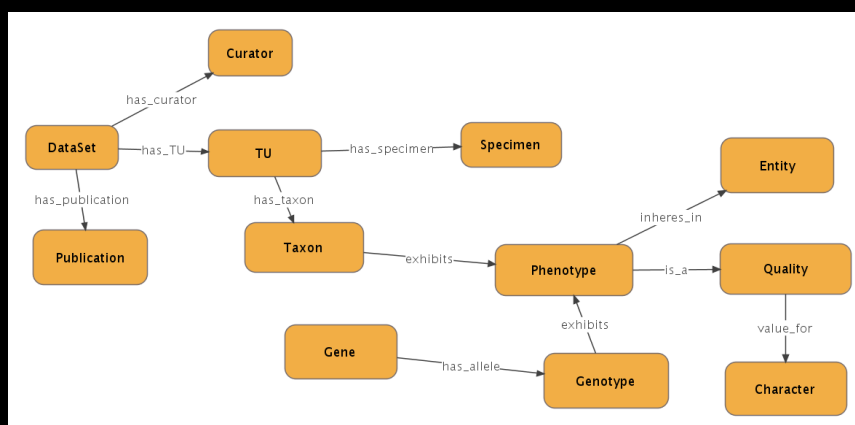
(Toledo-Piza 2000)

Annotation of mutant phenotypes

E	Q	
head	size	small
eye	size	small
heart	structure	edematous
ventral mandibular arch	thickness	thick
swim bladder inflation	process_quality	arrested

WT

fam32al-

Reducing ontology complexity: use qualities that imply attributes

Phenoscape data model
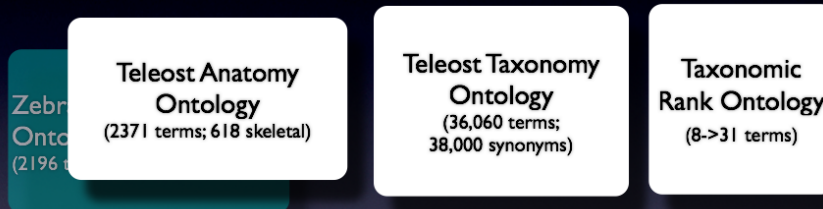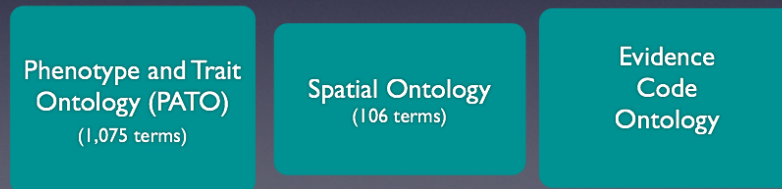
Teleost taxonomy ontology

- http://bioportal.bioontology.org/ontologies/40796
  - Based on the authoritative "Catalog of Fishes"
    - ♦ Bill Eschmeyer, Stan Blum, Peter Midford
  - 36,060 valid taxonomic names and 38,000 synonyms
  - Covers *all* published names encountered in literature curation, even mispellings
  - >400 fossil taxa
  - Can include cross-references to other taxonomic identifiers, more complicated metadata
  - Orthogonal to rank
  - Taxa related by *is_a* relationships, which allow one to reason using transitivity
  - Required ~6 months of curation effort

Teleost Anatomy Ontology

- Seeded from Zebrafish Anatomy Ontology
- Homology assertions are kept separately and attributed to an *authority* with an *evidence code*
- Participation is open
  - Mailing list with occasional jamborees
  - Ontology gatekeeper
- Ontology is built as needed for data annotation

Teleost Anatomy Ontology
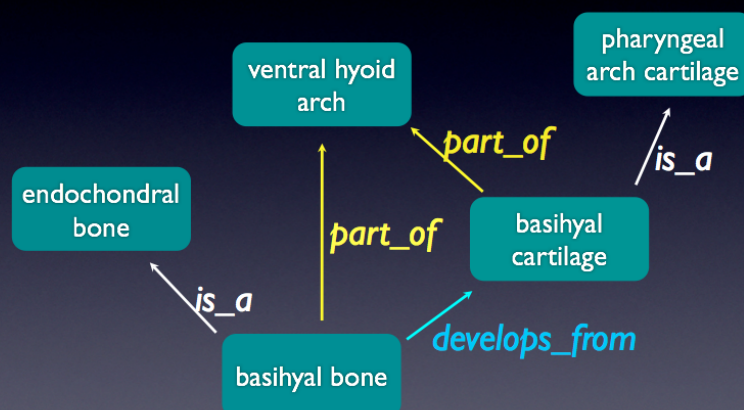
Dahdul, W. M., J. G. Lundberg, P. E. Midford, J. P. Balhoff, H. Lapp, T. J. Vision, M. A. Haendel, M. Westerfield, and P. M. Mabee. *in press*. The Teleost Anatomy Ontology: Anatomical representation for the genomics age. *Systematic Biology*.

OBO Relations Ontology

- Foundational
  - *is_a, part_of*
- Spatial
  - *located_in, contained_in, adjacent_to*
- Temporal
  - *transformation_of, derives_from, preceded_by*
- Participation
  - *has_participant, has_agent*
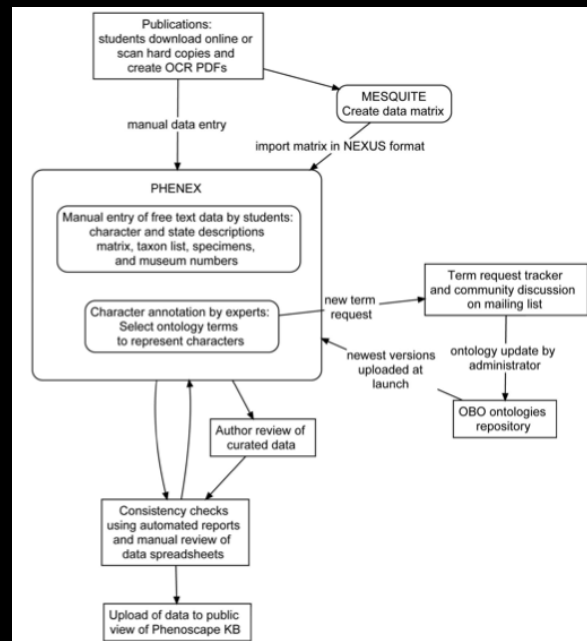
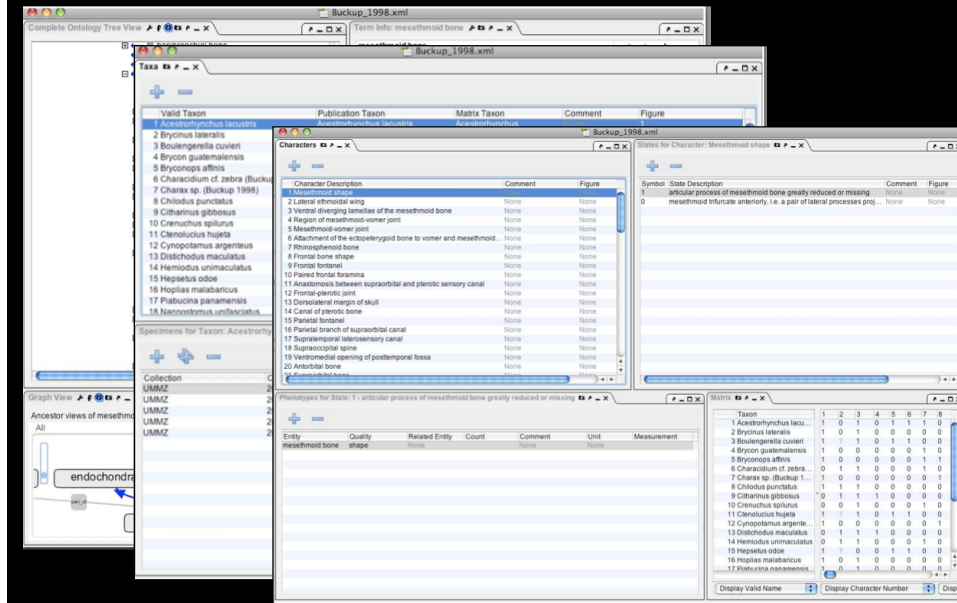# Phenotype and Trait Ontology (PATO)
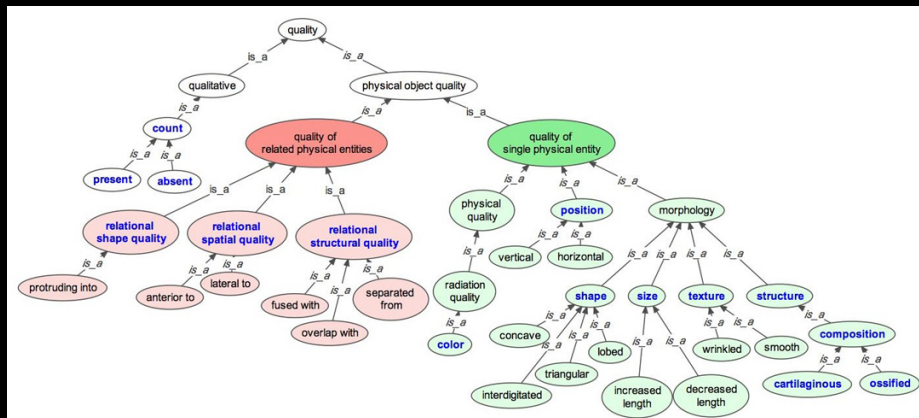
Curation Workflow

~5 person-years, despite only annotating at a coarse level

Phenex data curation software

From character state to EQ

- One character state
  - ♦ "Form and area of attachment of primordial ligament: (0) ligament relatively narrow and attaching to posteromedial portion of ascending process of maxilla..... (Zanata & Vari, 2005)

- Corresponds to multiple EQ Phenotypes
  - ♦ E: primordial ligament; Q: size, narrow
  - ♦ E1: primordial ligament; Q: attached to; E2: maxilla ascending process
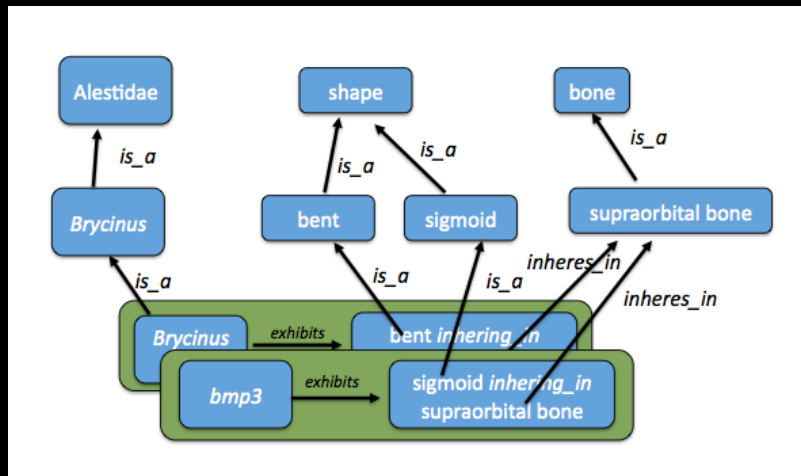
NeXML output from Phenex

- Original character and state definitions
- Taxa
  - Including specimen and collection IDs
- Character matrix
- Entity-quality phenotype assignments to taxa

An example of reasoning
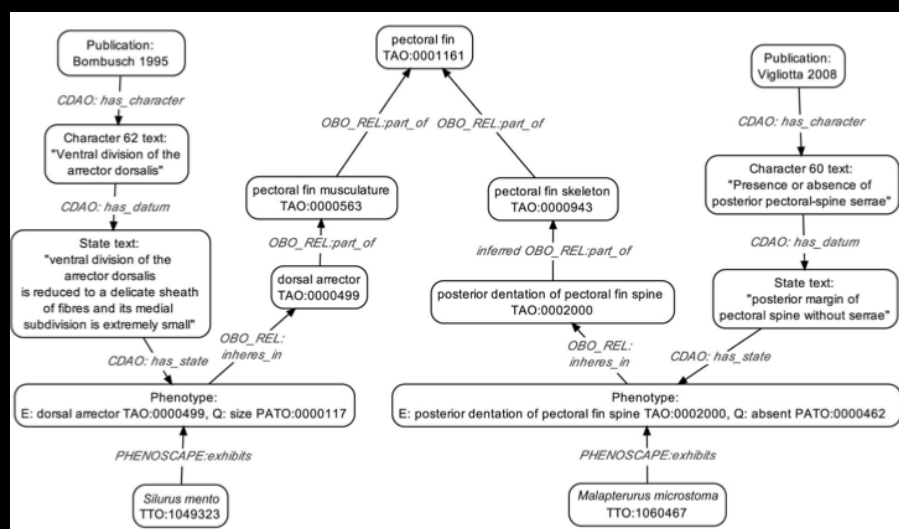
Linking genotype to phenotype

Phenoscape Knowledgebase

- http://kb.phenoscape.org
  - ♦ 333,987 phenotype statements about 2310 taxa from 51 publications
  - ♦ 11,267 phenotype statements about 2953 genes from ZFIN.

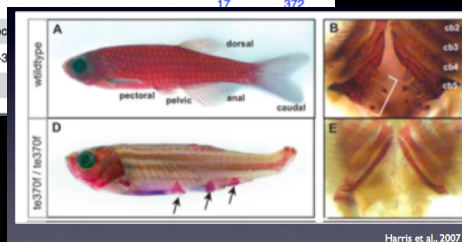# From character states to EQ statements

- Complexity of character definitions
  - 1 EQ
  - 2 EQ: 14%
  - 3 or more EQ: 2%
- Relational: 5%
- Binary: 69%
- Presence-absence: 28%
- Post-compositional: 35%

Search for ZFIN mutants affecting scale development

Phenotypes

46 genes

Anatomy	Quality	Zebrafish Genes	Taxa
scale	relational spatial quality : angular placement, relational spatial quality towards process of parietal bone, relational spatial quality towards supraoccipital crest	3	19
scale	texture : texture of	0	91
scale	structure	0	95
scale	position : inverted, spatial pattern	10	79
scale	shape : round	17	372
scale	quality : discontinuous, malformed, oc		
scale	count : absent, count of 0, count of 0-3		
scale	size : decreased size		

Harris et al., 2007

finless (*eda*)

Loss of scales in fish evolution

Gasterosteus aculeatus

Linking developmental genetics to evolutionary variation
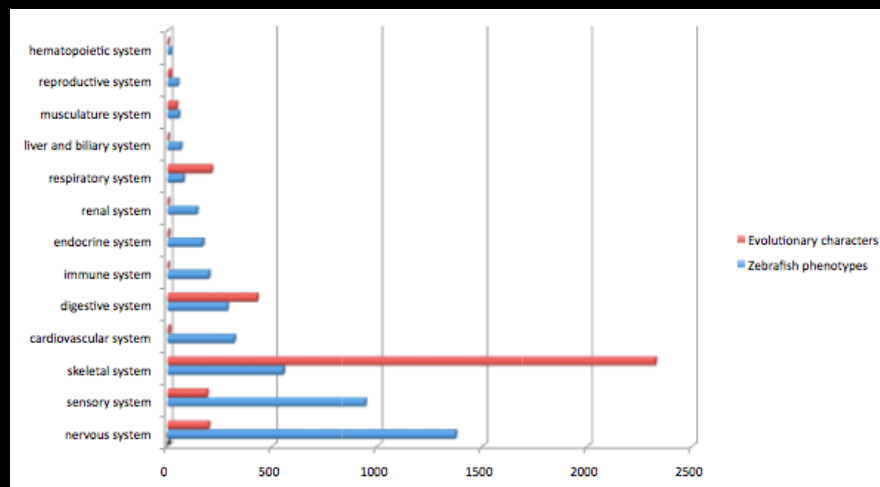
Application to plants?

- A large (but maybe not *so* large) legacy literature
- Abundance of mutant data
  - Some of which is now being annotated in EQ?
- A good foundational Plant Ontology
- Potential for linking G2P and iPToL grand challenge projects

- But this is not something that iPlant can do without leadership from morphologists
  - The work that needs to be done is almost 100% data curation and ontology refinement