An Introduction to the Use of Ontologies in Linking Evolutionary Phenotypes to Genetics

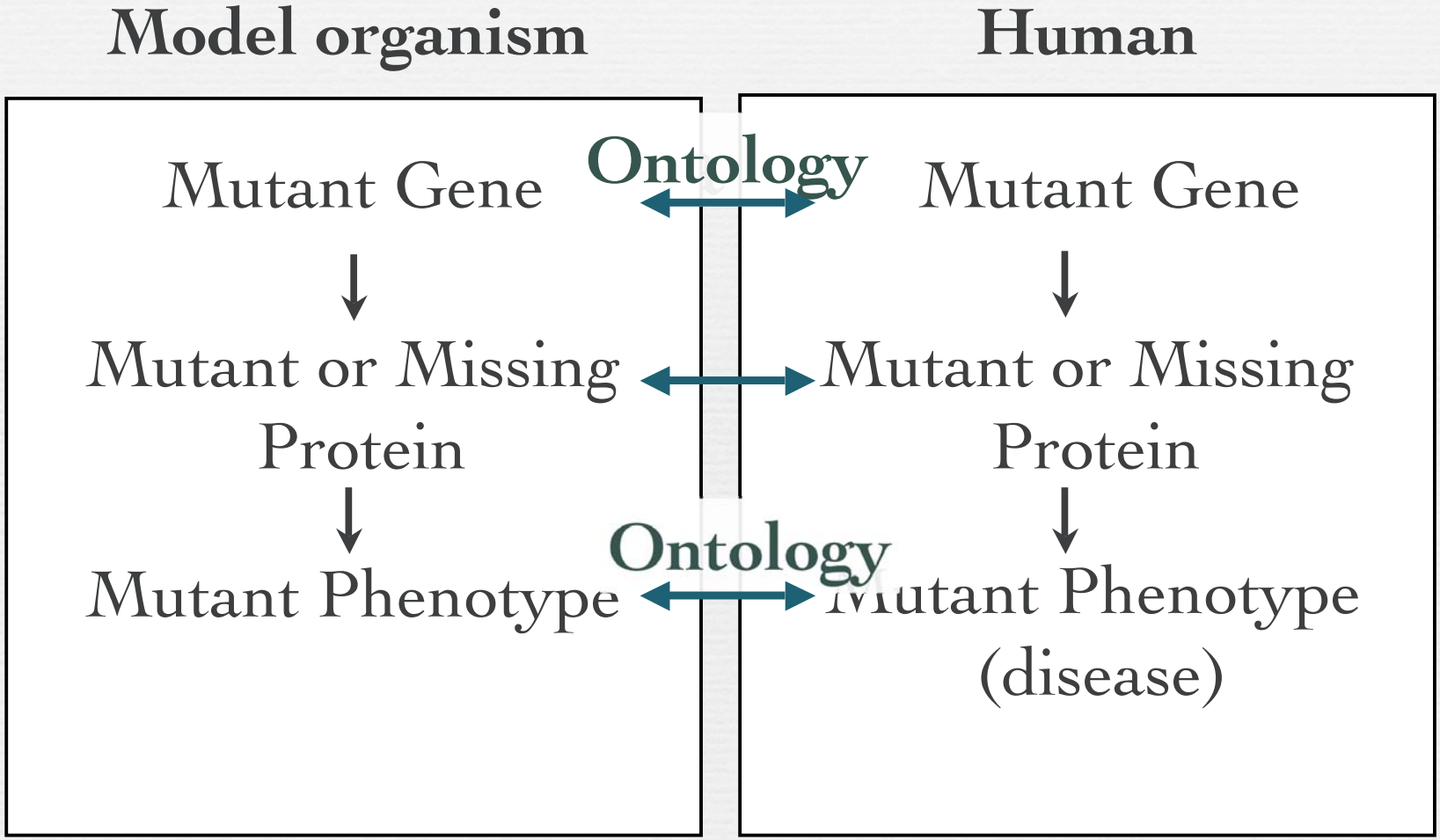
Paula Mabee
University of South Dakota

PHENOSCAPE

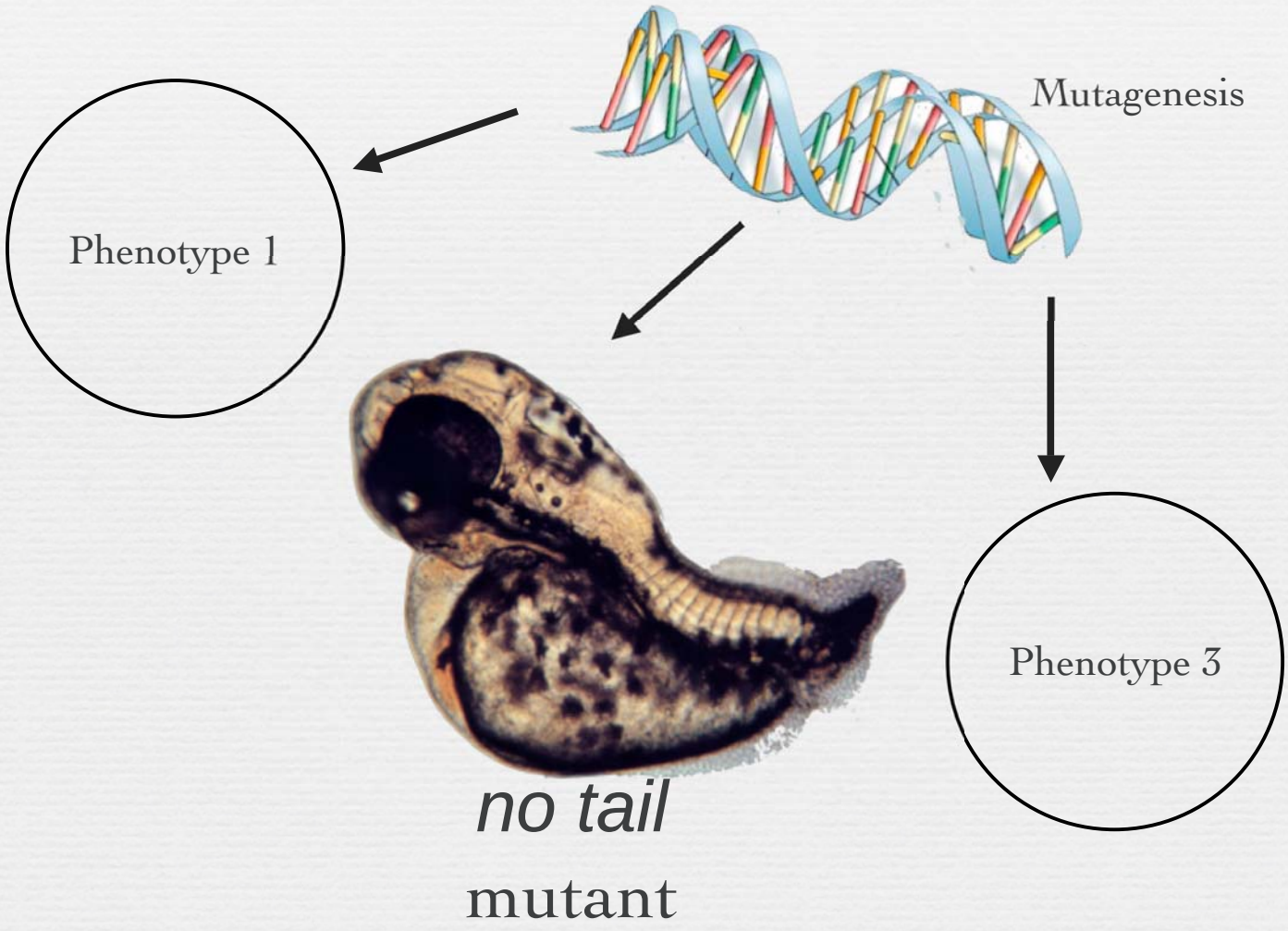
NSF

The University of South Dakota

Shared ontologies & syntax connects models to humans



Zebrafish:
Mutagenesis produces phenotypes



Halpern et al. (1993) *Cell*

Phenotypes mapped to genes

- **Maxilla**: size reduction
  - ➡ *sox9ahi1134*
- **Dentary**: size reduction
  - ➡ *sox9ahi1134*
- **Retroarticular**: loss
  - ➡ *edn1*
- **Opercle**: size reduction; loss
  - ➡ *sox9ahi1134; lockjaw*
- **Ceratohyal**: shape change
  - ➡ *val*
- **Branchiostegals**: number decrease
  - ➡ *edn1*
- **Branchiostegals**: shape change
  - ➡ *she, stu, edn1*-MO

Zebrafish ⟷ Human

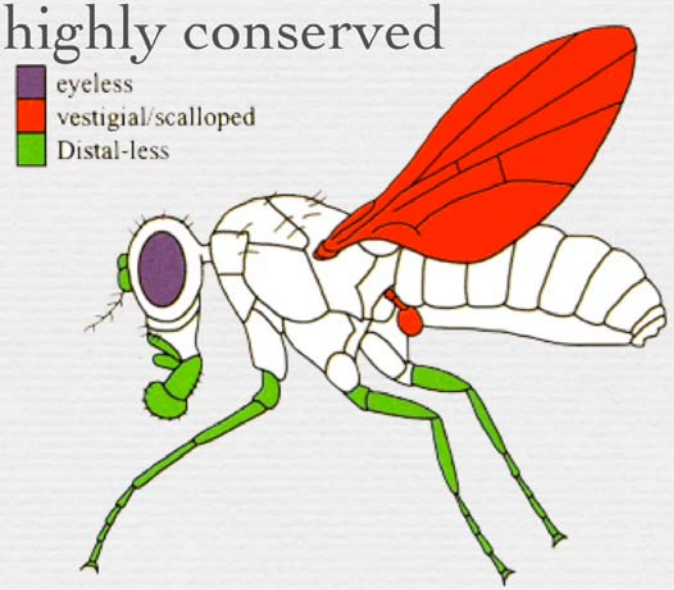
Conservation of gene sequence function

(Lamason et al., 2005)

16 December 2005: The lightly pigmented *golden* zebrafish show a striking resemblance to lighter skinned humans. The zebrafish pigment gene *slc24a5* is functionally conserved across evolution; a single base change in its human ortholog may play a role in pigment variation in human populations

Discovery of genetic 'tool kit'

- Small number of gene families are involved in development
- Transcription factors, signaling proteins, ncRNA
- These genes and pathways are highly conserved



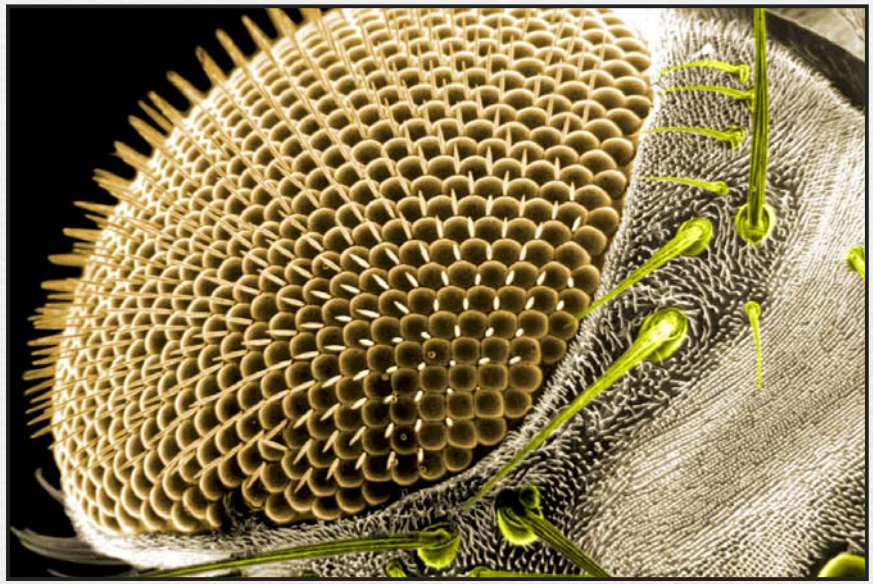
Carroll (2007)

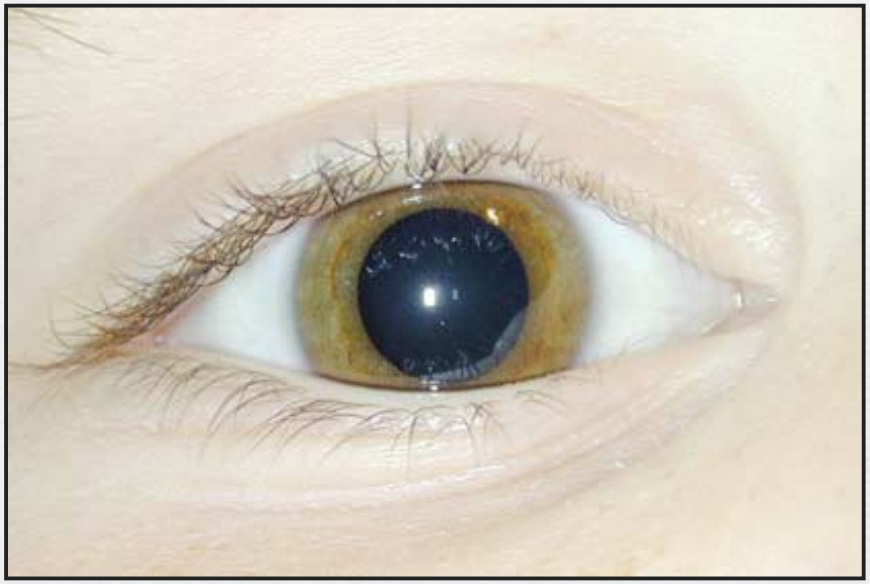
Evolutionarily conserved developmental pathways

- 40+ pathways
- E.g.
  - Limb outgrowth (*Distal-less*)
  - Subdivision of the anterior-posterior axis (*Hox*)
  - Dorsal-ventral polarity (*Bmp*)

http://www.sdbonline.org/fly/aimain/aadevinx.htm

Eye morphogenesis (*eyeless/Pax6*)

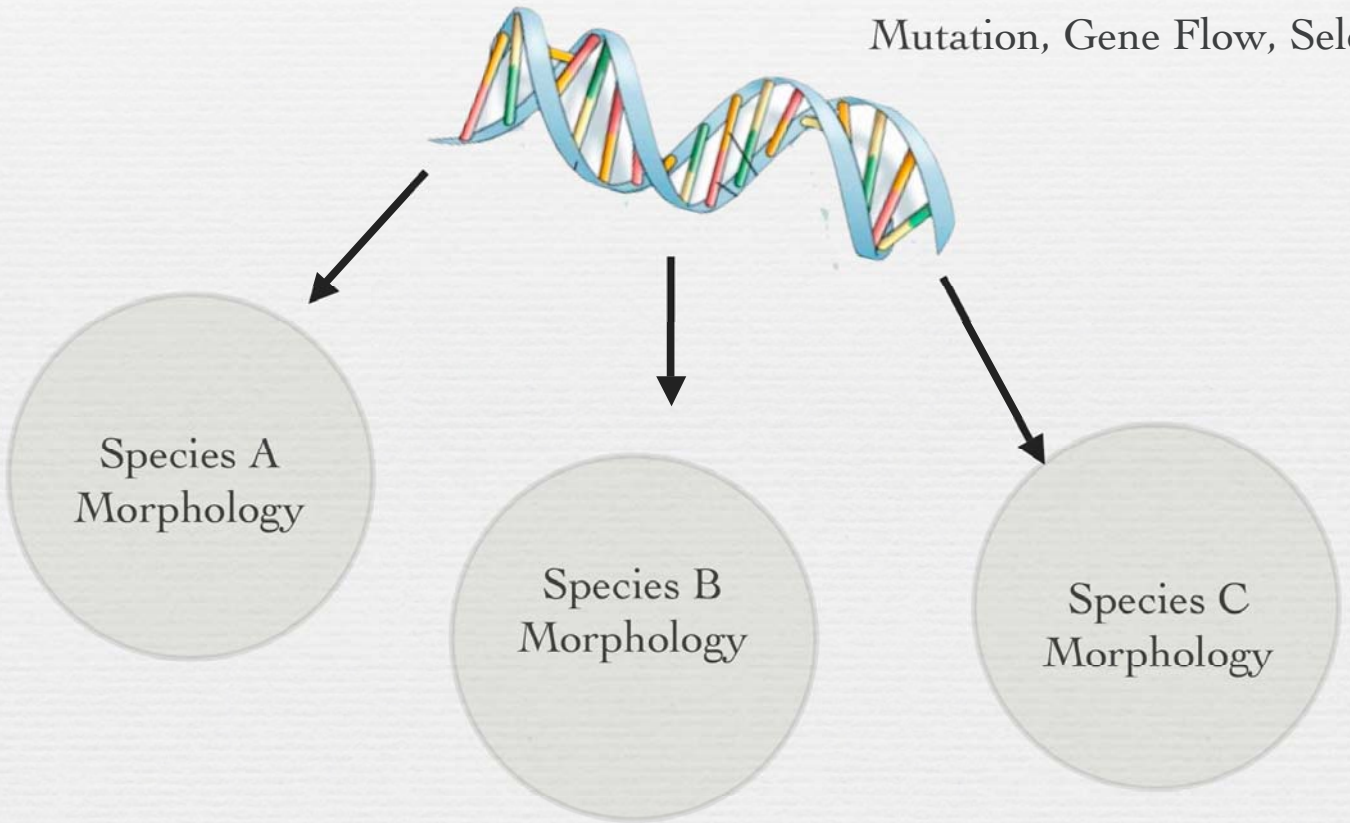




Used with permission from Micromundi: http://remf.dartmouth.edu/images/MicromondiImages/16chrysomelidae220.tif

Courtesy of Kevin Helenurm

Evolutionary phenotypes



Genetic bases of morphology unknown

Mutation, Gene Flow, Selection, Drift

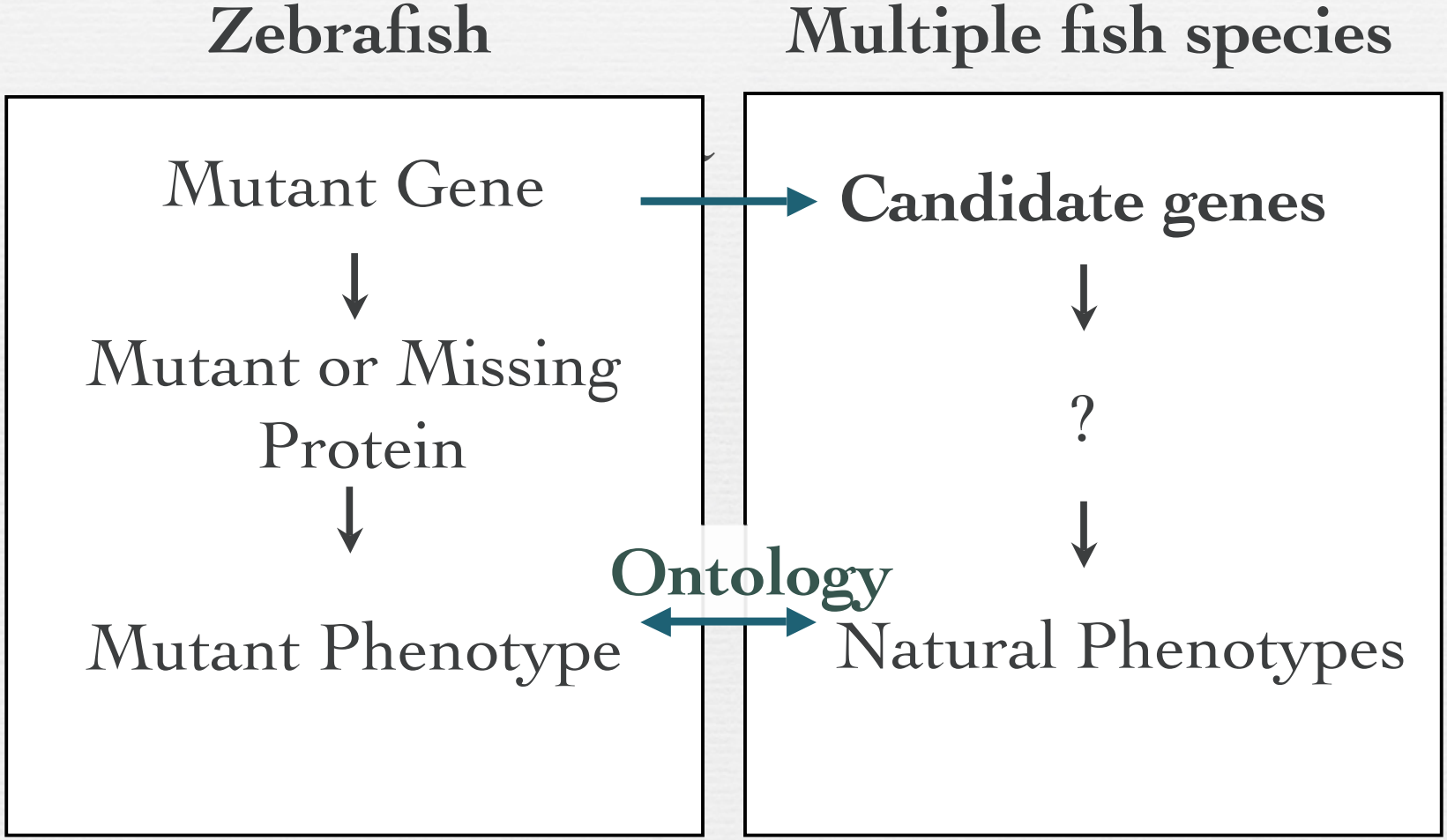
?

Species A Morphology

Species B Morphology

Species C Morphology

Shared ontologies & syntax connects models to other species

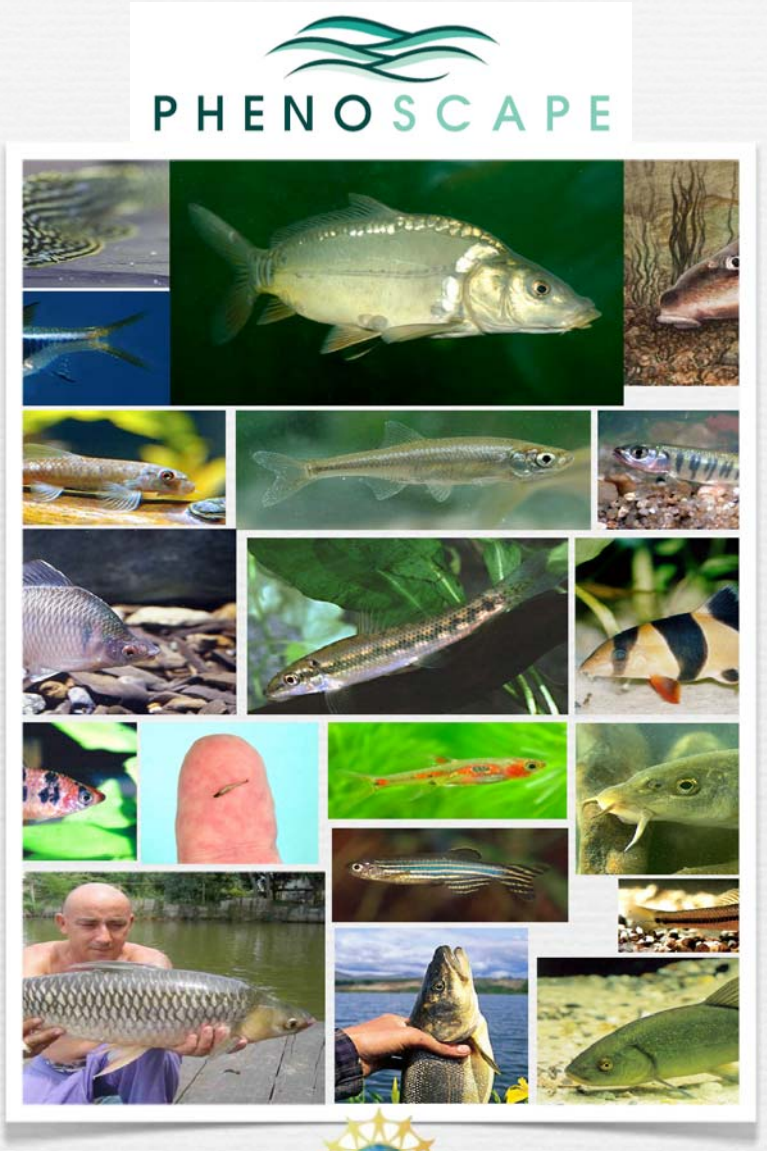


Phenoscape.org

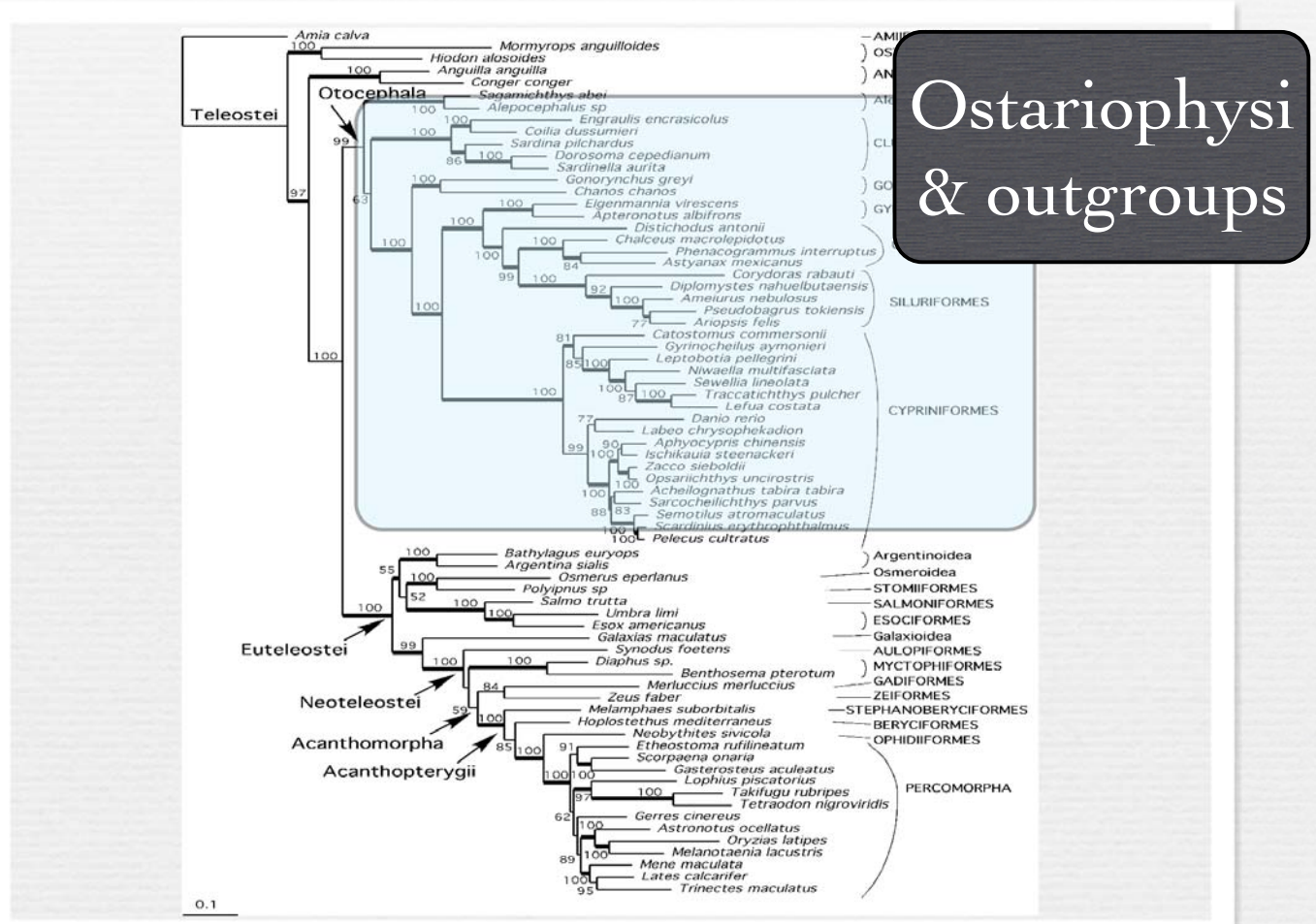
- **History**: Communication between zebrafish model organism community and Cypriniformes Tree of Life group through NESCent workshops (Mabee-Westerfield)

- **Goal**: Create curated, ontology-based evolutionary phenotype database that maps to genetic databases

- **Generalizable system:** Prototype with ostariophysan fishes



Prototype with Ostariophysan fishes (zebrafish)



Ostariophysi & outgroups

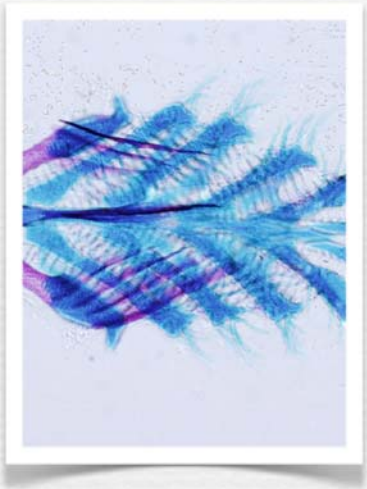
Mayden et al. 2008 (unpublished)

Needs analysis use cases:

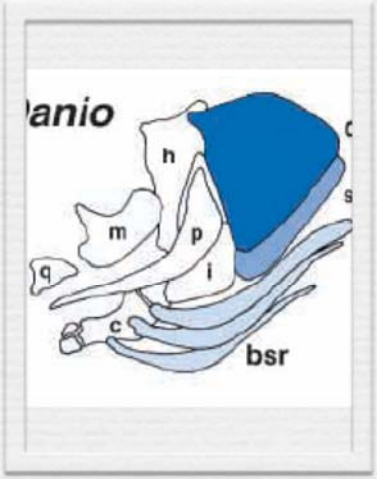
- **Phenoscape designed to meet top-priority questions/needs of the community concerning development and evolution of morphology, e.g.**
- Find genes underlying morphological characters (Which ones? How many?)
- Discover patterns of correlation across genes and morphology
- Formulate models of morphological evolution; data mining and discovery
- Phenotypic BLAST to discover similar phenotypes and taxa

Example 1: Candidate genes?

- **Evolution**: Ceratobranchial 5 varies in size and shape within cypriniform fishes



- **Evolution**: Branchiostegal rays vary in number in ostariophysan fishes



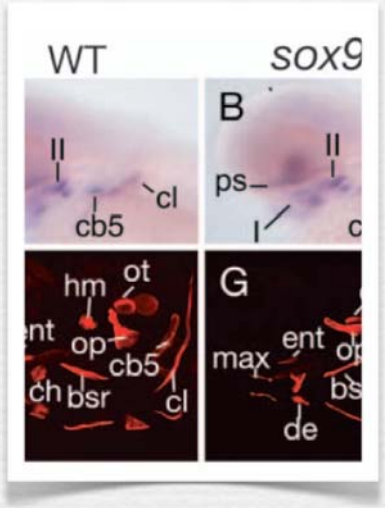
Results: Phenotype and genes

Mutant: *sox9ahi1134*

Gene(s): sox9a

Expression: 1 Fig.

Reference(s): Yan et al. 2005

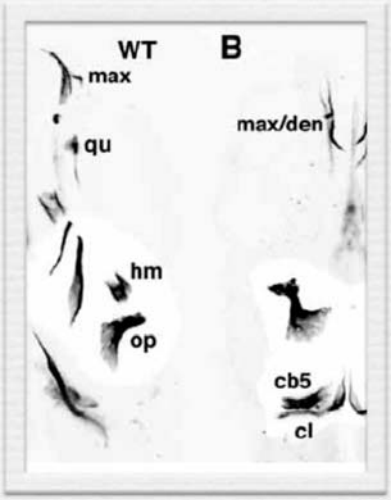


Mutant: *sucker*

Gene(s): endothelin1

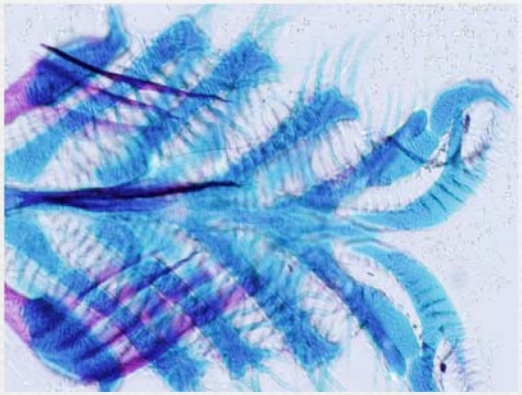
Expression: 2 Figs.

Reference(s): Kimmel et al. 03



Example 2: Candidate taxa?

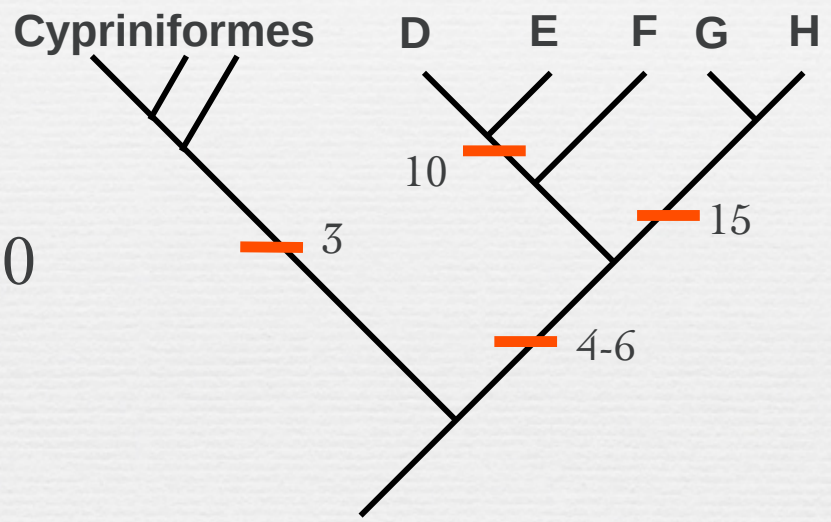
- **Mutant**: Branchiostegal ray number is reduced in a particular zebrafish mutant(s).



- **Query**: What fish species show reduction in branchiostegal ray number? What is the pattern of evolution of this character?

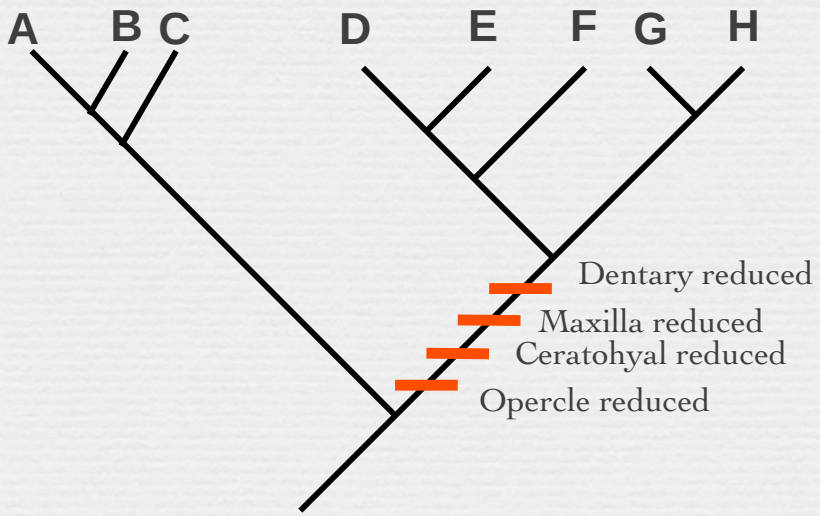
Results:

Taxa	# Bsr
Cypriniformes	3
Characiforms	3–5
Siluriformes	3–20
Gonorynchiformes	3–4
Ghost pipefishes	1
Pipefishes, seahorses	1–3

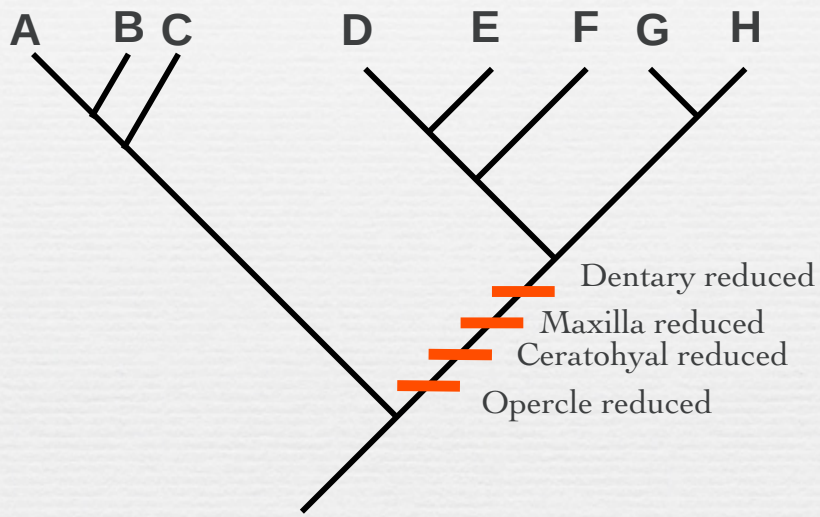


Example 3: Correlated characters

- **Evolution**: Reduction in size of the dentary, maxilla, ceratohyal, and opercle support the monophyly of Clade 1.
- **Query**: 4 characters? Correlated?



Dentary, opercle, maxilla size co-regulated [in part] by *sox9a*

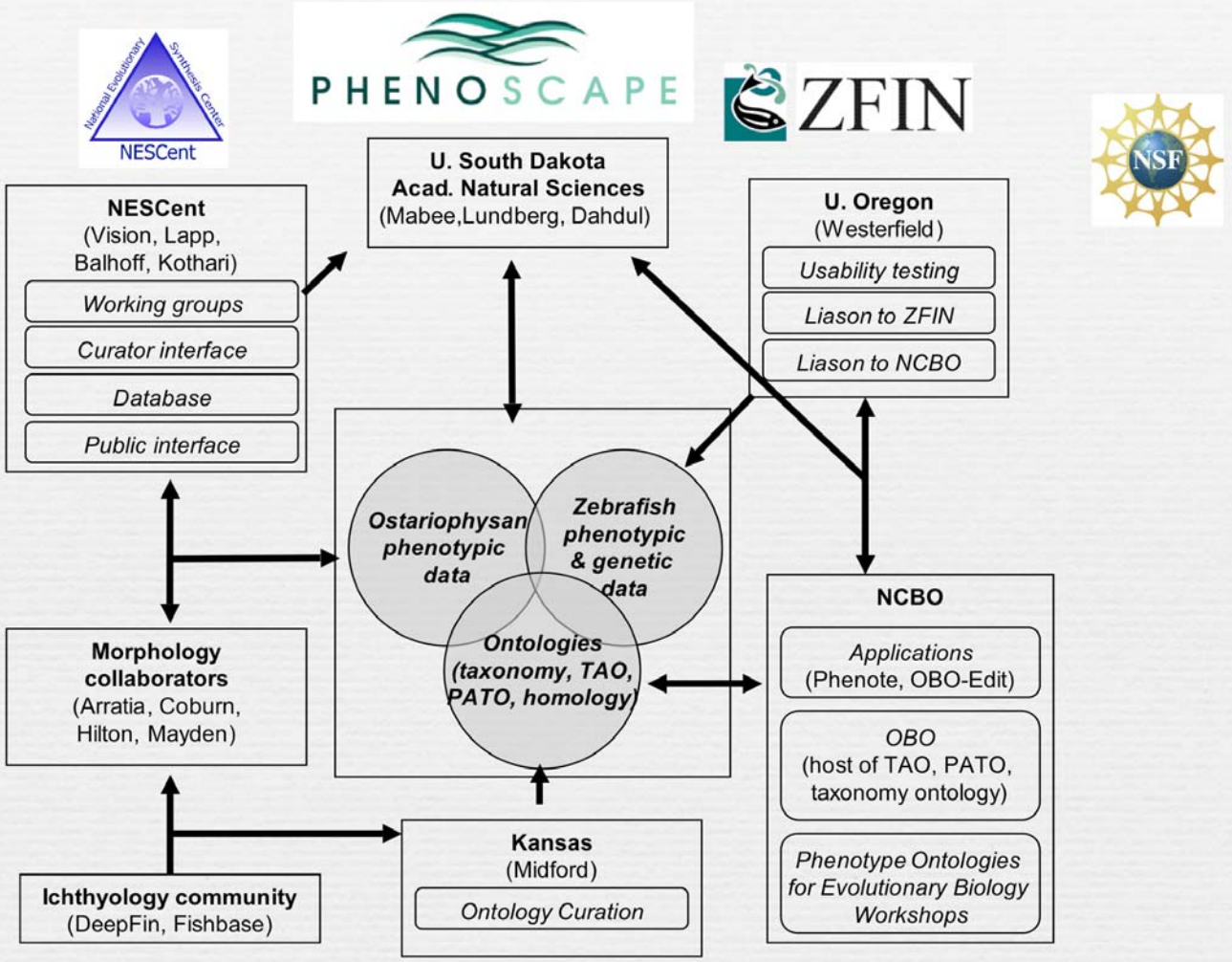


- Dentary, opercle, maxilla size co-regulated [in part] by *sox9a* (or pathway)
- Support for monophyly of clade is not as strong as it appears (2 vs. 4 characters)

Phenoscape priorities:

- Establish and maintain **communication**
- Develop **ontologies** for evolutionary work
- Refine **syntax** for evolutionary characters
- Develop curation **tools**
- **Curate** phenotypes (characters)

Communication



DeepFin

FishBase

THE NATIONAL CENTER FOR BIOMEDICAL ONTOLOGY

Phenoscape ontologies

Cloned:

Zebr
Anat
Onto
(2196 terms; 310 skeletal)

Teleost Anatomy Ontology
(2233 terms; 387 skeletal)

New:

Teleost Taxonomy Ontology
(36,060 terms; 38,000 synonyms)

Taxonomic Rank Ontology
(8->31 terms)

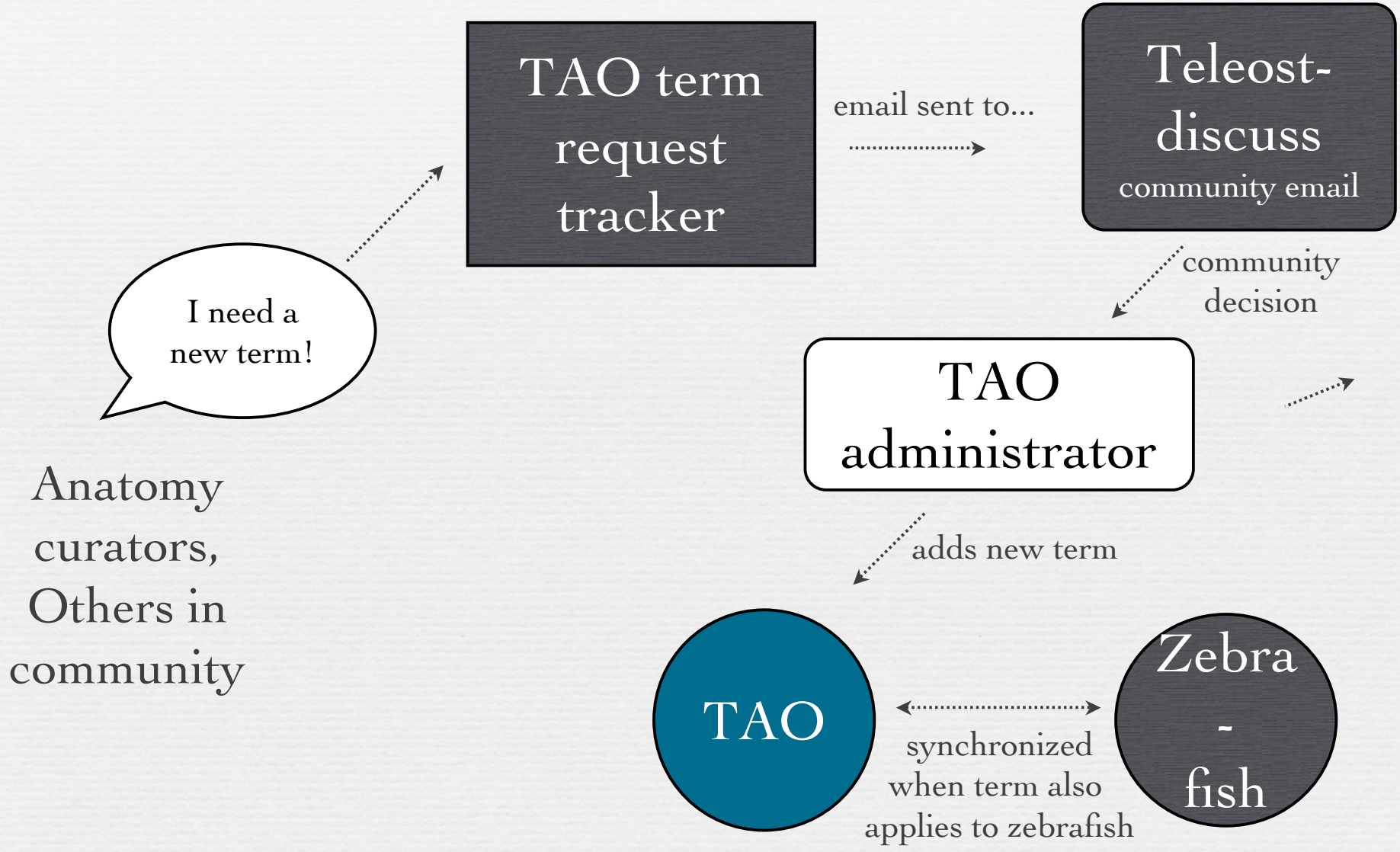
Existing:

Phenotype and Trait Ontology
(1,075 terms)

Spatial Ontology
(106 terms)

Evidence Code Ontology

Adding new terms to Teleost Anatomy Ontology (TAO)



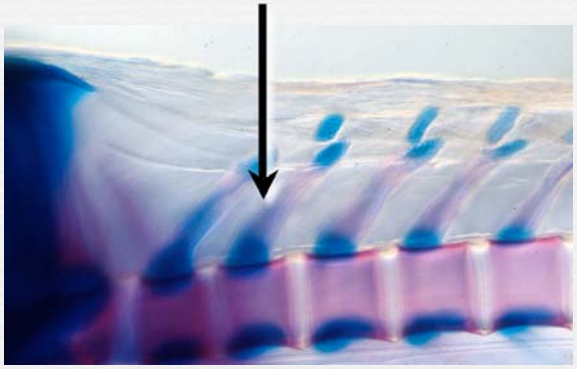
Challenges in developing a multi-species ontology

- Representing serial homologues
- Representing multiple developmental pathways
- Representing comparative standard (wt/outgroup)

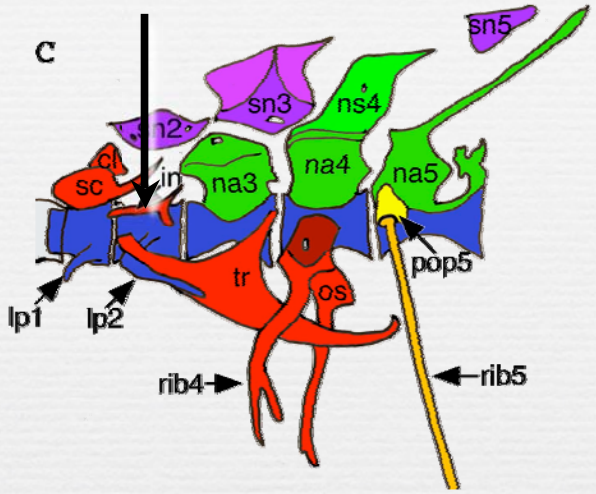
***Representation affects outcome of searches**

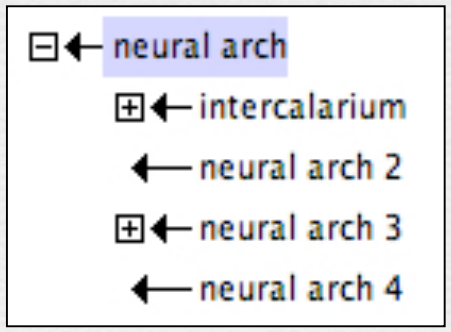
Representing serial homologues

Neural arch 2



Intercalarium





Teleost Anatomy Ontology

Neural arch 2 and intercalarium
children of neural arch

Homology annotation:

Publication	Entity 1	Taxon 1	Entity 2	Taxon 2	Evidence
Fink and Fink 1981	neural arch 2	Teleostei	intercalarium	Otophysi	inferred from positional similarity

Phenoscape priorities:

- Establish and maintain **communication**
- Develop **ontologies** for evolutionary work
- Refine **syntax** for evolutionary characters
- Develop curation **tools**
- **Curate** phenotypes (characters)

Entity-Quality (EQ) syntax

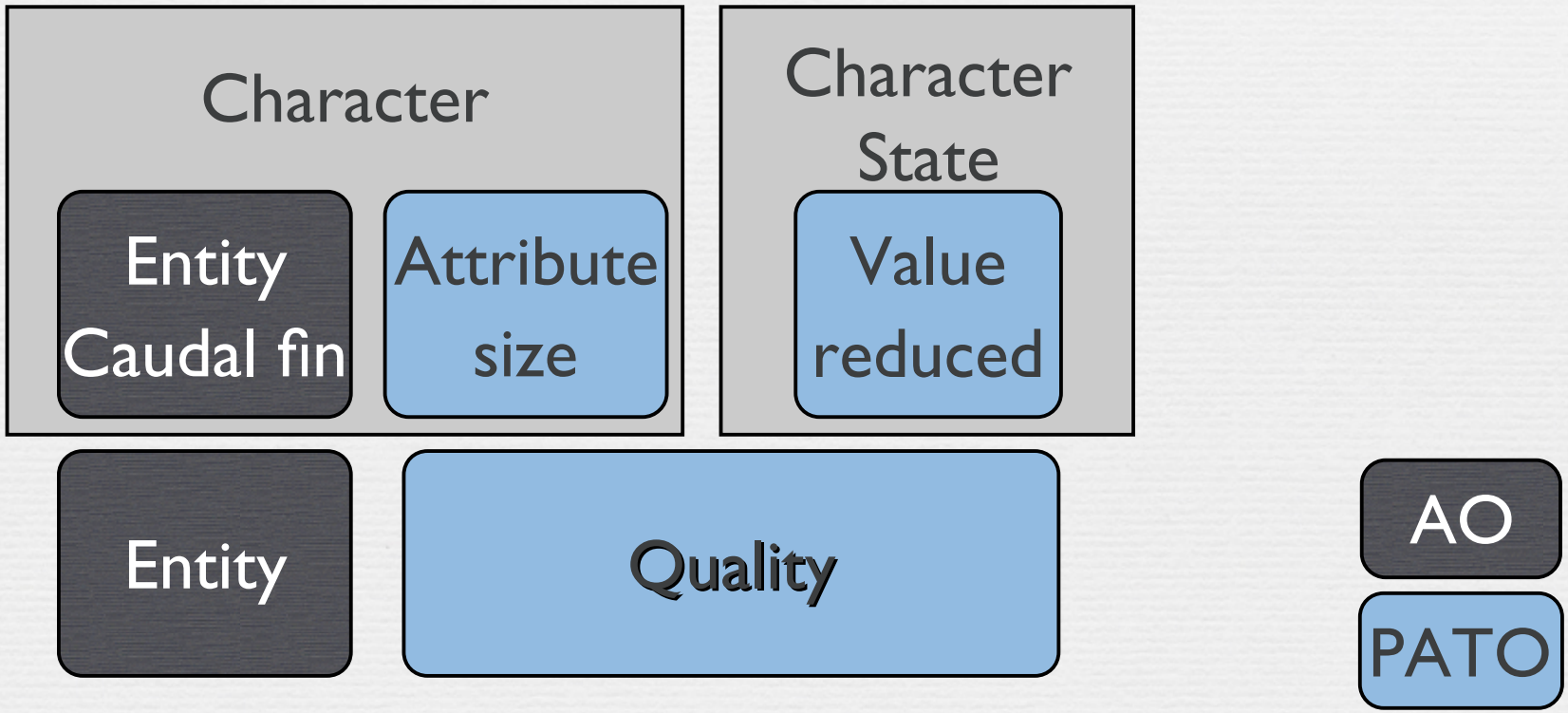
Entity	Quality
Caudal fin	Reduced (size)

no tail mutant

Anatomy Ontology: AO

Quality Ontology: PATO

Systematic characters can also be described using Entity-Quality (EQ) syntax



Expanding EQ syntax:

Free-text character:

- Parietal and supraoccipital fused

Ontology EQ character:

- Parietal fused with supraoccipital

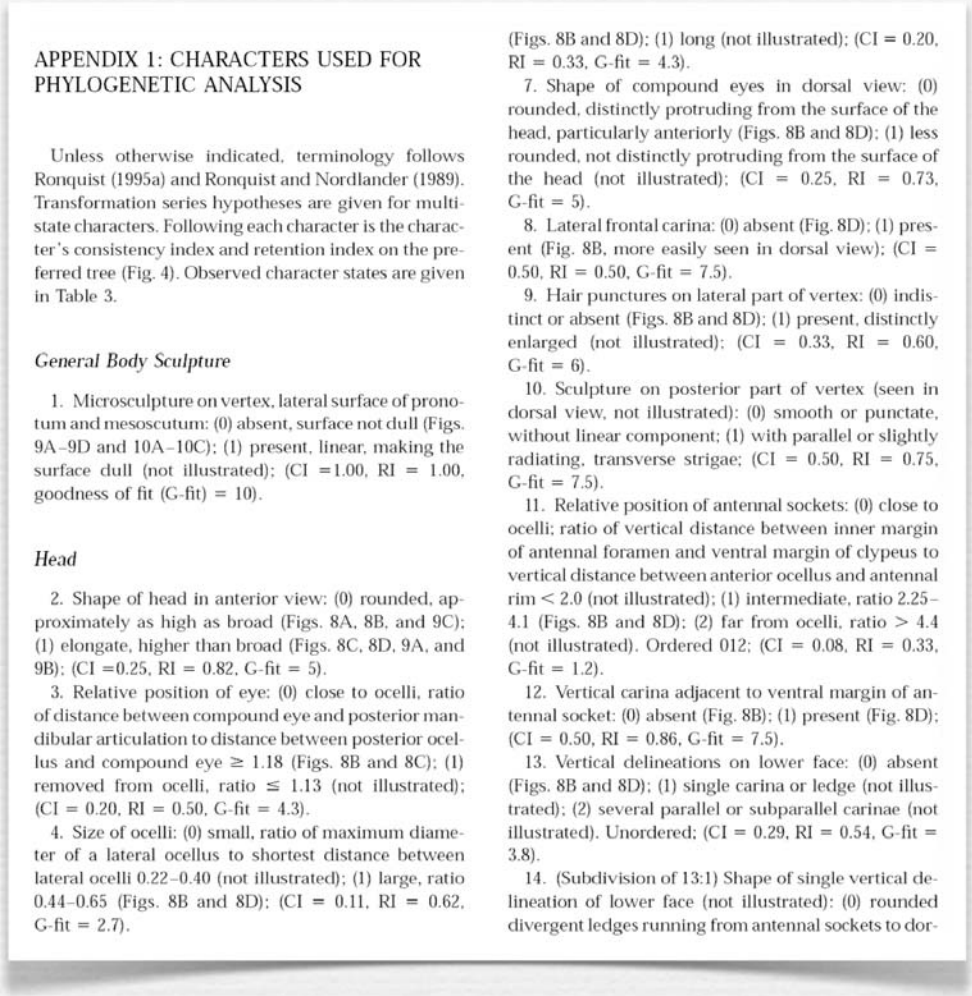
E — Relational Quality — E

Phenoscape priorities (yr 1):

- Establish and maintain **communication**
- Develop **ontologies** for evolutionary work
- Refine **syntax** for evolutionary characters
- Develop curation **tools** (Phenote)
- **Curate** phenotypes (characters)

Curate evolutionary phenotypes from free-text

- This is not a computable format
- These data cannot be easily compared across taxa
- These data cannot be linked to developmental genetics
- These data cannot be reasoned across



APPENDIX 1: CHARACTERS USED FOR PHYLOGENETIC ANALYSIS

Unless otherwise indicated, terminology follows Ronquist (1995a) and Ronquist and Nordlander (1989). Transformation series hypotheses are given for multistate characters. Following each character is the character's consistency index and retention index on the preferred tree (Fig. 4). Observed character states are given in Table 3.

General Body Sculpture

1. Microsculpture on vertex, lateral surface of pronotum and mesoscutum: (0) absent, surface not dull (Figs. 9A–9D and 10A–10C); (1) present, linear, making the surface dull (not illustrated); (CI =1.00, RI = 1.00, goodness of fit (G-fit) = 10).

Head

2. Shape of head in anterior view: (0) rounded, approximately as high as broad (Figs. 8A, 8B, and 9C); (1) elongate, higher than broad (Figs. 8C, 8D, 9A, and 9B); (CI =0.25, RI = 0.82, G-fit = 5).

3. Relative position of eye: (0) close to ocelli, ratio of distance between compound eye and posterior mandibular articulation to distance between posterior ocellus and compound eye ≥ 1.18 (Figs. 8B and 8C); (1) removed from ocelli, ratio ≤ 1.13 (not illustrated); (CI = 0.20, RI = 0.50, G-fit = 4.3).

4. Size of ocelli: (0) small, ratio of maximum diameter of a lateral ocellus to shortest distance between lateral ocelli 0.22–0.40 (not illustrated); (1) large, ratio 0.44–0.65 (Figs. 8B and 8D); (CI = 0.11, RI = 0.62, G-fit = 2.7).

(Figs. 8B and 8D); (1) long (not illustrated); (CI = 0.20, RI = 0.33, G-fit = 4.3).

7. Shape of compound eyes in dorsal view: (0) rounded, distinctly protruding from the surface of the head, particularly anteriorly (Figs. 8B and 8D); (1) less rounded, not distinctly protruding from the surface of the head (not illustrated); (CI = 0.25, RI = 0.73, G-fit = 5).

8. Lateral frontal carina: (0) absent (Fig. 8D); (1) present (Fig. 8B, more easily seen in dorsal view); (CI = 0.50, RI = 0.50, G-fit = 7.5).

9. Hair punctures on lateral part of vertex: (0) indistinct or absent (Figs. 8B and 8D); (1) present, distinctly enlarged (not illustrated); (CI = 0.33, RI = 0.60, G-fit = 6).

10. Sculpture on posterior part of vertex (seen in dorsal view, not illustrated): (0) smooth or punctate, without linear component; (1) with parallel or slightly radiating, transverse strigae; (CI = 0.50, RI = 0.75, G-fit = 7.5).

11. Relative position of antennal sockets: (0) close to ocelli; ratio of vertical distance between inner margin of antennal foramen and ventral margin of clypeus to vertical distance between anterior ocellus and antennal rim < 2.0 (not illustrated); (1) intermediate, ratio 2.25–4.1 (Figs. 8B and 8D); (2) far from ocelli, ratio > 4.4 (not illustrated). Ordered 012; (CI = 0.08, RI = 0.33, G-fit = 1.2).

12. Vertical carina adjacent to ventral margin of antennal socket: (0) absent (Fig. 8B); (1) present (Fig. 8D); (CI = 0.50, RI = 0.86, G-fit = 7.5).

13. Vertical delineations on lower face: (0) absent (Figs. 8B and 8D); (1) single carina or ledge (not illustrated); (2) several parallel or subparallel carinae (not illustrated). Unordered; (CI = 0.29, RI = 0.54, G-fit = 3.8).

14. (Subdivision of 13:1) Shape of single vertical delineation of lower face (not illustrated): (0) rounded divergent ledges running from antennal sockets to dor-

Curation of ichthyological data

Taxon	# Species	# Papers	# Phenotypes
Cypriniformes (Mayden; Coburn)	3,268	70	**1125**
Siluriformes (Lundberg)	2,867	87	**1200**
Characiformes (Dahdul)	1,674	124	**800**
Gymnotiformes (Arratia)	134	2	**200**
Gonorynchiformes (Arratia)	37	80	**75**
Clupeiformes (Hilton)	364	60	**380**
TOTAL	**8,344**	**423**	**3,780**

Summary

- Evolutionary variation in morphology can be connected to developmental genes using ontologies
- Phenoscape: prototype of a generalizable system for making this connection
- Allows morphology to flourish
  - new questions
  - new discoveries
  - new visualizations

Acknowledgements

- NSF for Phenoscape funding (Mabee, Vision, Westerfield) NSF-DBI0641025
- National Evolutionary Synthesis Center NSF EF-0423641
- NIH HG002659 (to Monte Westerfield)
- Phenoscape project: Hilmar Lapp, Wasila Dahdul, Peter Midford, Jim Balhoff,
- ZFIN: Monte Westerfield, Melissa Haendel
- Cypriniformes Tree of Life (NSF 0431290), colleagues and students (Mayden, Miya, Saitoh, He, Coburn, Arratia, Simon, Conway, Grey, Engeman, Bogutskya, Hilton, Aspinwall); Deep Fin RCN
- Suzanna Lewis, Chris Mungall (Lawrence Berkeley National labs)
- National Center for Biomedical Ontology