Phenoscape: Using ontologies to link comparative morphology to genes

Paula Mabee
University of South Dakota

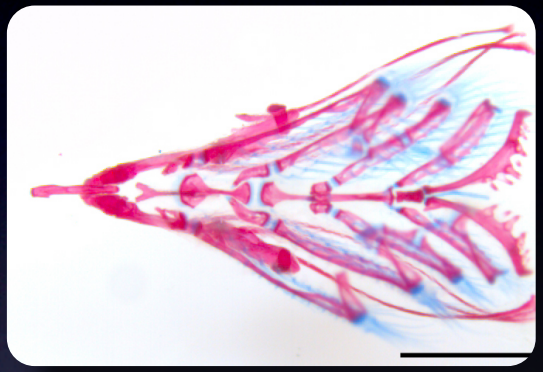
Jim Balhoff, Wasila Dahdul, Cartik Kothari, Hilmar Lapp, John Lundberg, Peter Midford, Todd Vision, Monte Westerfield

Phenoscape Knowledgebase (KB)

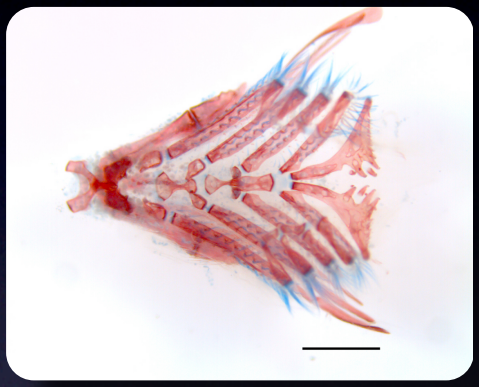
kb.phenoscape.org

Comparative morphology

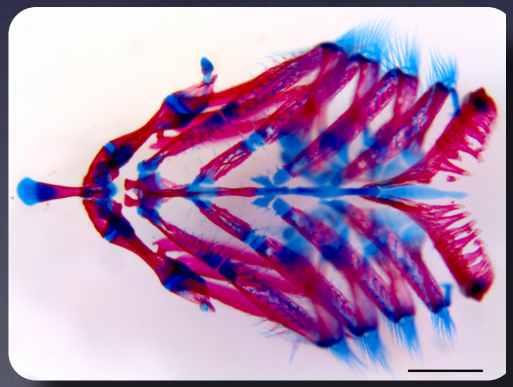
Cyprinus carpio



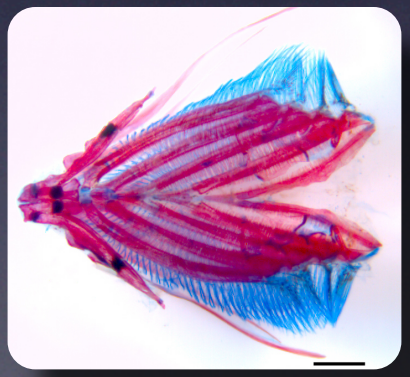
Pangio anguillaris



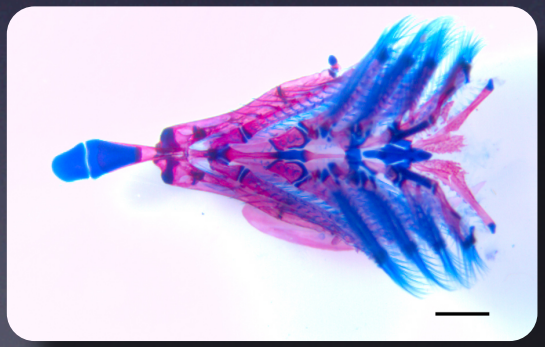
Nemacheilus fasciatus



Catostomus commersoni



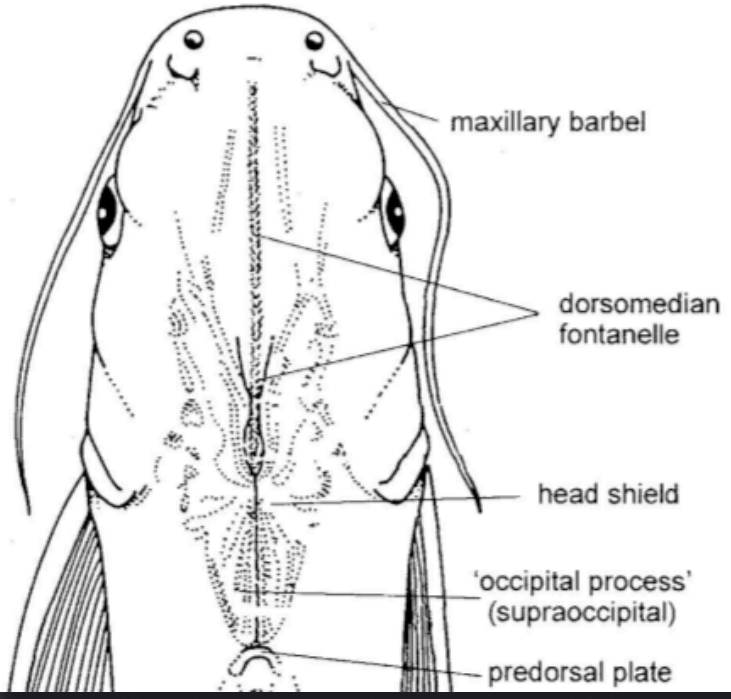
Gyrinocheilus aymonieri



Phenacogrammus interruptus

"Free-text' format not computable across studies

47. *Shape of posterior dorsomedian fontanelle.* An elongate-rectangular posterior fontanelle of most catfishes appears to be the plesiomorphic condition (see also Tilak 1963, 1964, 1965a; Lundberg 1982; Arratia 1987; Grande 1987). An irregularly-shaped or rounded



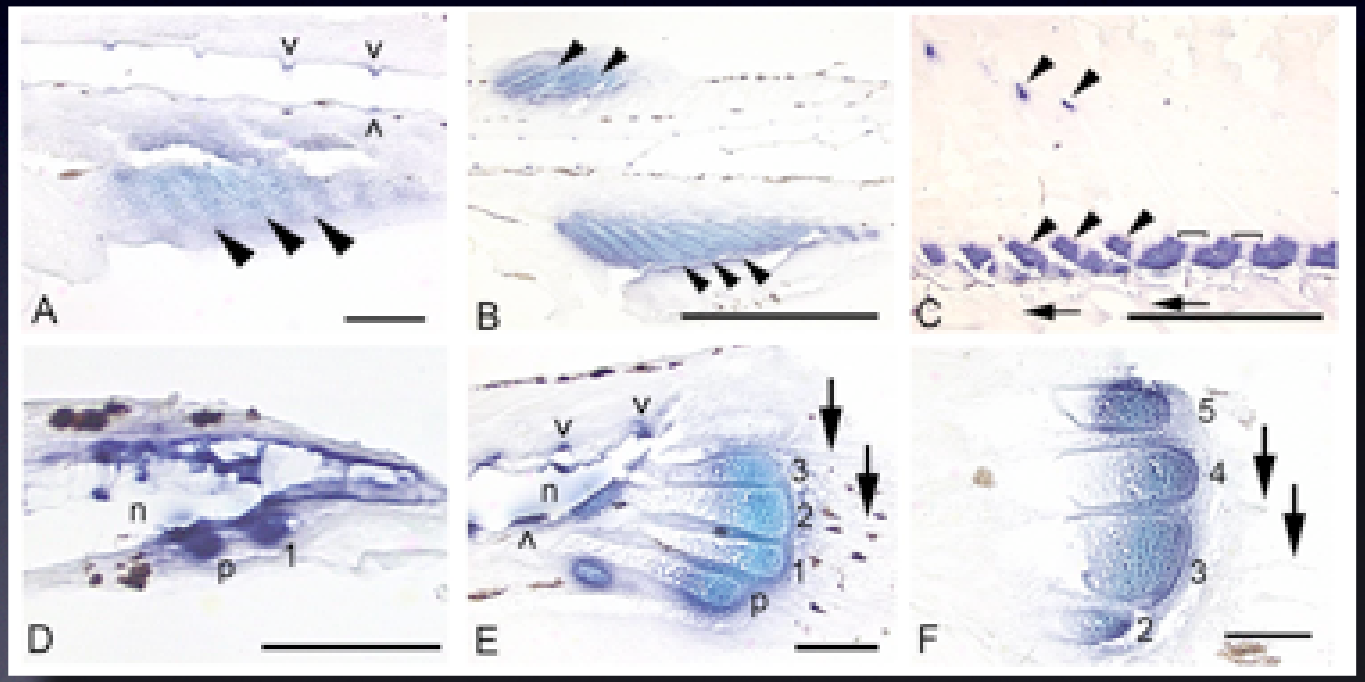
State 0 = frontal broad anteriorly and moderately
narrow p
arms mo
posterio
2 = fron
space en
49. *L*
laminar
(except
extensiv
evidence
the lami
of the tr
I consid
is plesio

?	0	0	?	3	0	?	0	?	0	0	?	1	0	?
1	?	?	?	0	?	?	0	1	0	0	2	0	2	0
0	?	?	?	0	?	?	0	?	0	0	2	1	2	1
0	?	?	?	0	?	?	0	1	?	?	2	1	2	0
1	?	0	?	2	0	?	0	1	0	?	2	1	2	?
1	?	?	?	0	0	?	0	?	0	?	2	1	1	1
3	?	?	?	0	?	?	0	1	0	1	2	1	0	2
3	?	?	?	0	?	?	0	1	0	?	2	1	0	1

cover' as exposed transverse process bases and a deep median excavation on the ventral surface. The laminar bone in ariids extends over four to eight vertebra centra. Some ariids possess apomorphic modifications in the laminar shelf, such as depressions (e.g., *Guiritinga barbus, Cinetodus froggatti*) or median single keel (e.g., high and acute in *Batrachocephalus, Nemapteryx armiger*) or double keel (e.g., *Bagre marinus*).

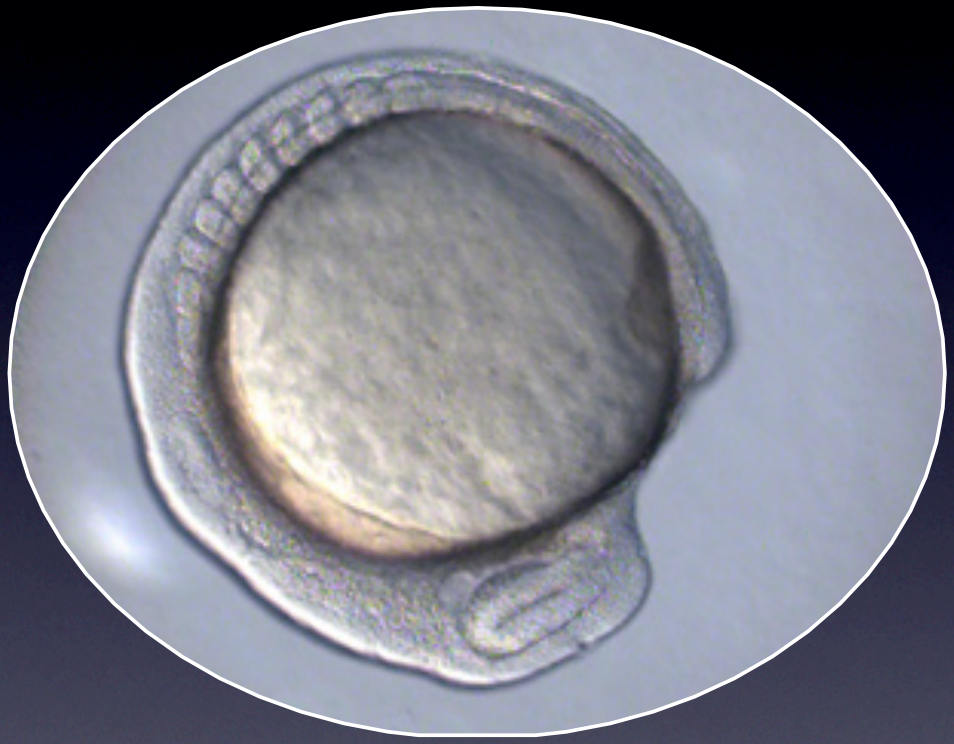
(Kailola 2004)

Cannot be synthesized with developmental and genetic data



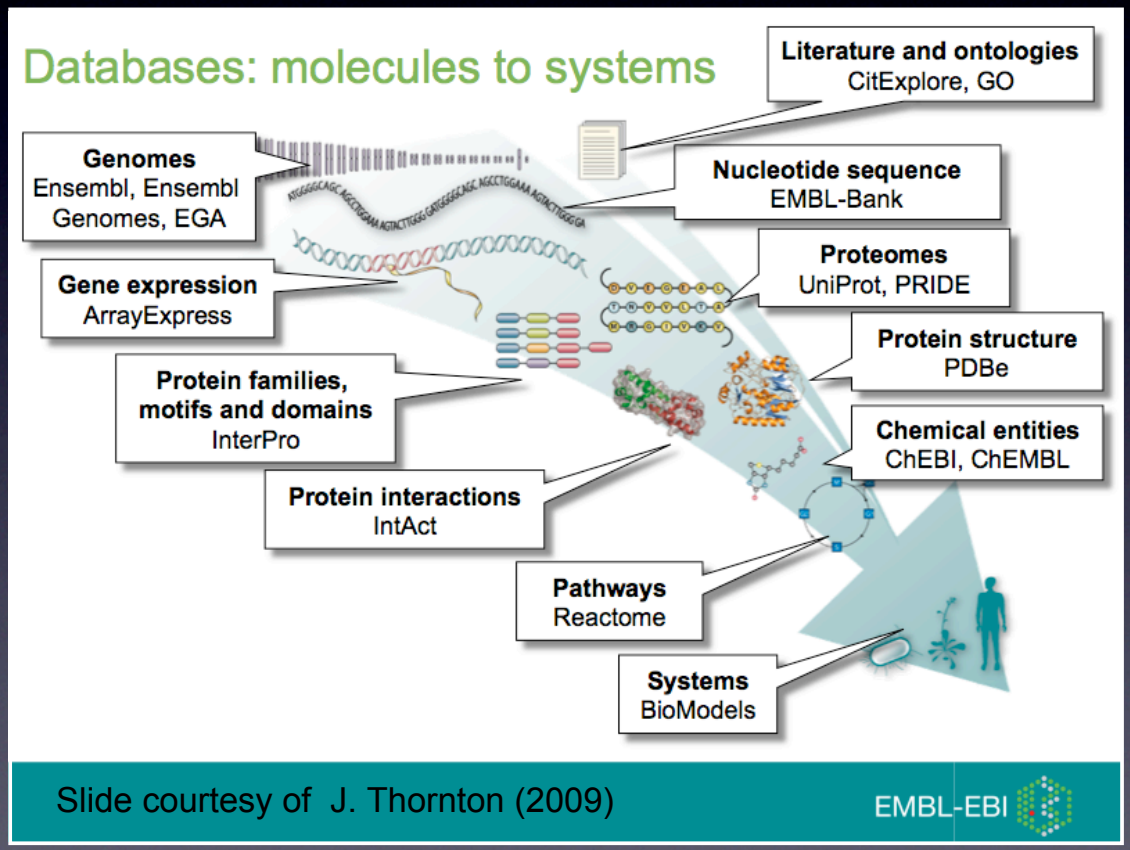
Crotwell & Mabee, 2007

Morphological differences



have their basis in changes in genetic control over development

>1,170 molecular databases



(Galperin & Cochrane 2008)

Zebrafish database

- Est. 1992 by Monte Westerfield and Oregon group
- Community resource
- Curate phenotypes of mutants, transgenics, genes, gene expression

25 July 2009

ntla

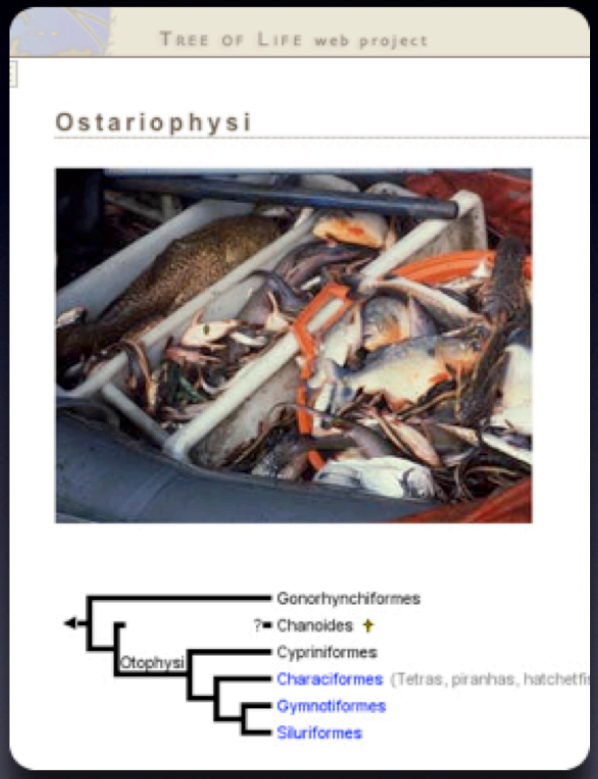
Background:

- Mabee (CToL) and Westerfield (ZFIN), 2005-2006, NESCent working group
- Established communication across traditionally separated scientific communities
- Ontology boot camp...
- Interoperability of evolutionary morphology and mutant phenotype data

Phenoscape

2007

- Curated, ontology-based evolutionary morphology database that maps to genetic databases
- Foster devo-evo synthesis
- Data-mining and discovery



25 July 2009; ToL web

Requirements:

1. Ontologies
2. Curation
3. Database

1. Ontologies

New:

Zebr
Onto
(2196 t

Teleost Anatomy Ontology
(2371 terms; 618 skeletal)

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

Taxonomic Rank Ontology
(8->31 terms)

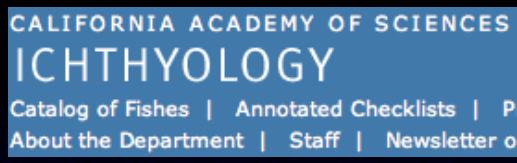
Existing:

Phenotype and Trait Ontology (PATO)
(1,075 terms)

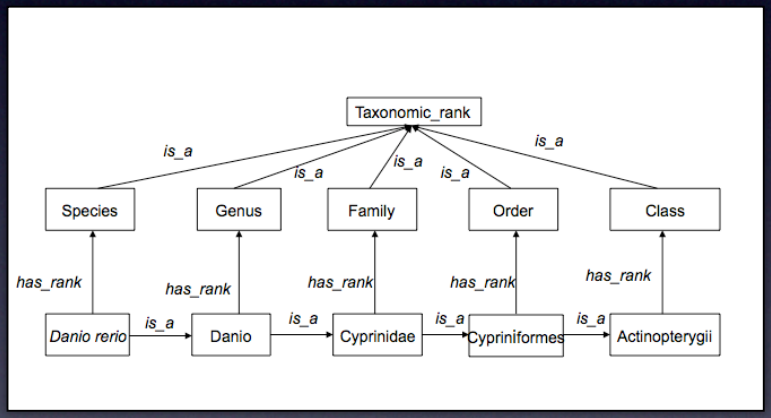
Spatial Ontology
(106 terms)

Evidence Code Ontology

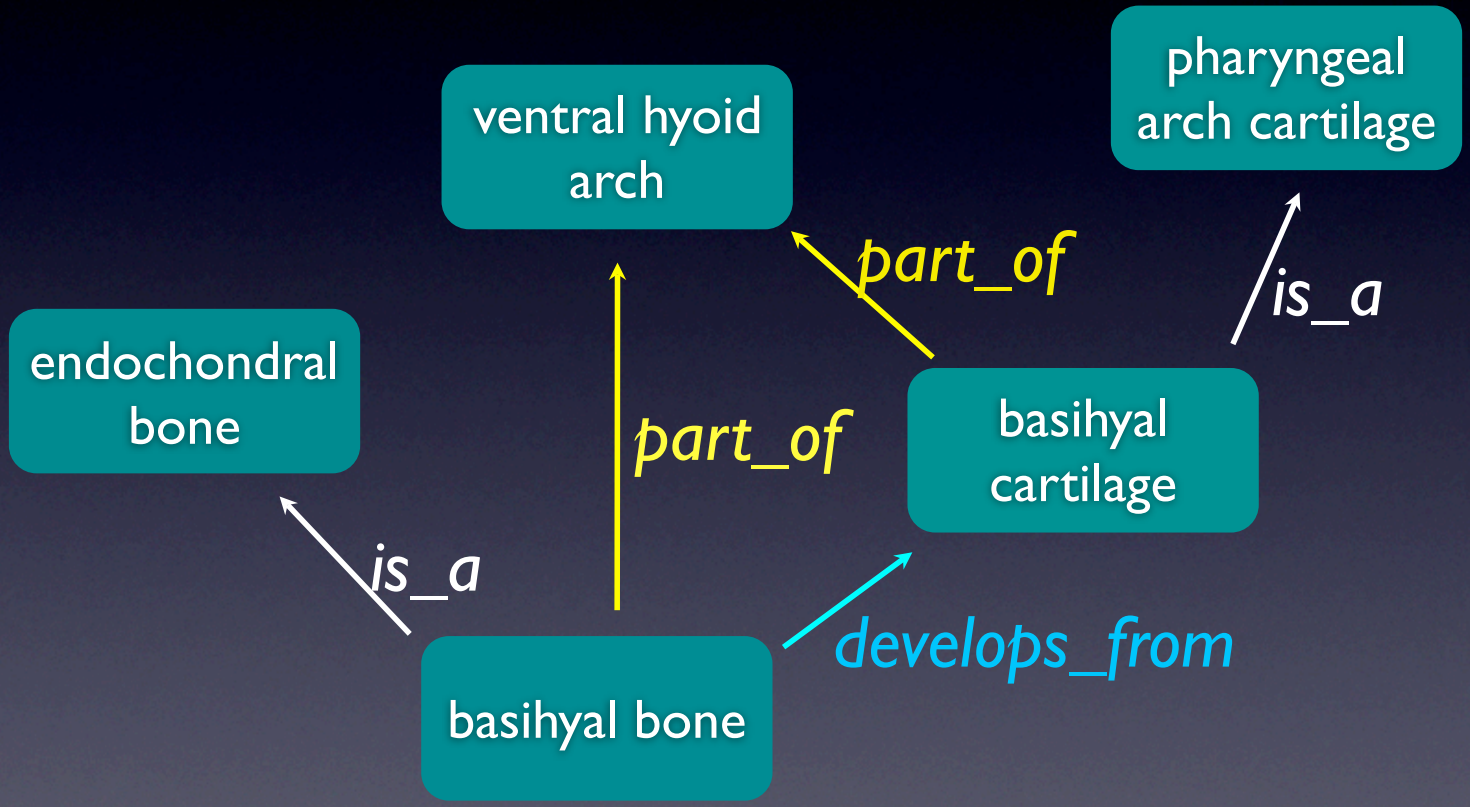
Teleost Taxonomy Ontology (TTO)



- Based on Eschmeyer's Catalog of Fishes
- Contains 36,060 terms
  - 36,508 are species
  - 5,045 are genera
  - 542 are families
  - 47 are extinct
- Contains 43,122 taxonomic synonyms

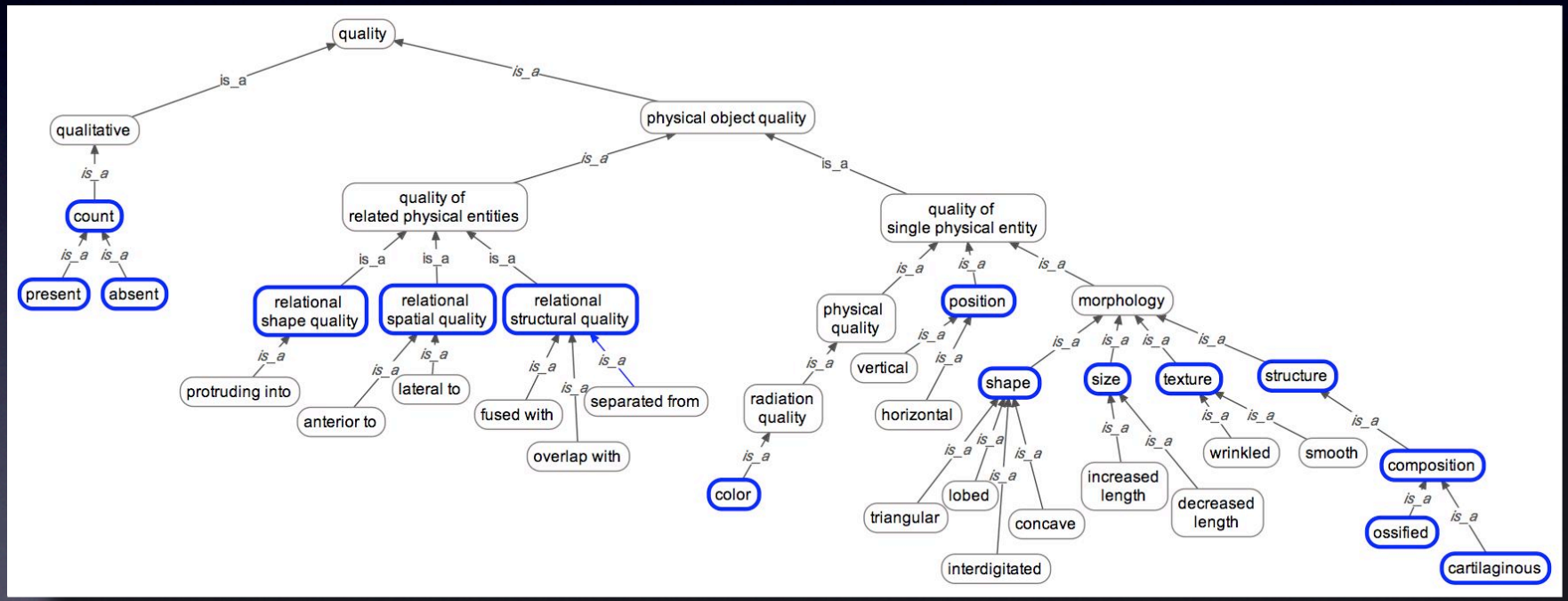


Teleost Anatomy Ontology (TAO)



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*.

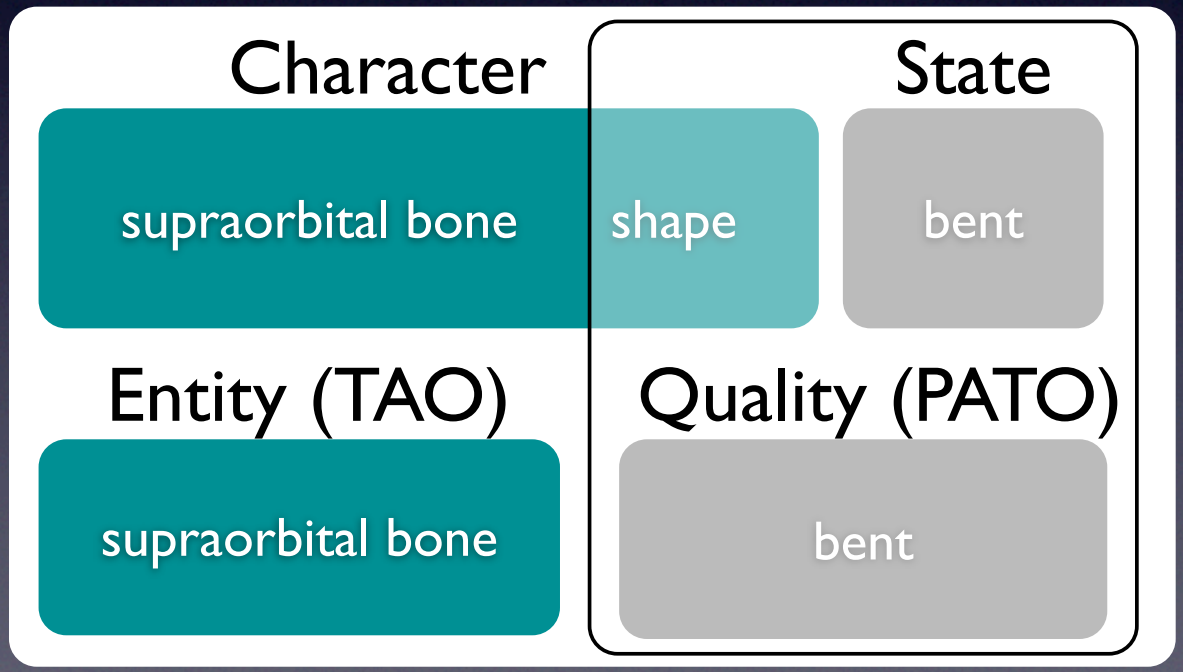
Phenotype and Trait Ontology (PATO)



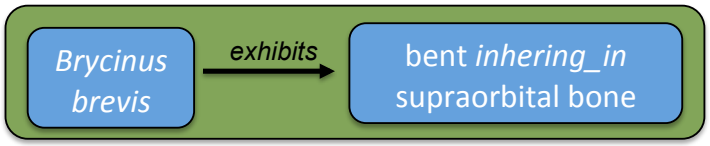
Quality ontology: size, shape, presence-absence, color, etc.

Phenotype = Entity+Quality

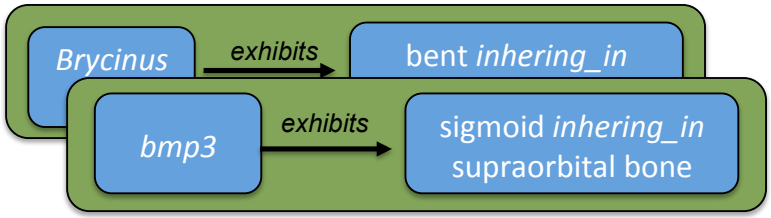
Free text character:
'Supraorbital bone shape: bent (0) or straight (1)'



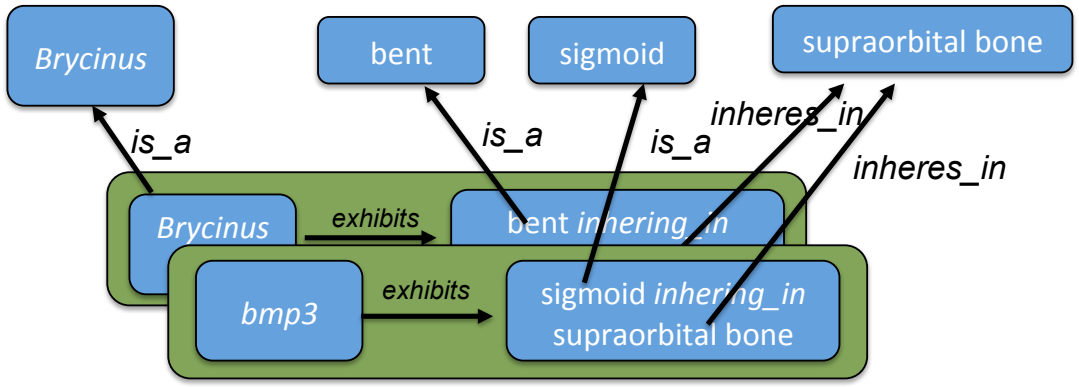
Taxon + Phenotype = Phenotype assertion



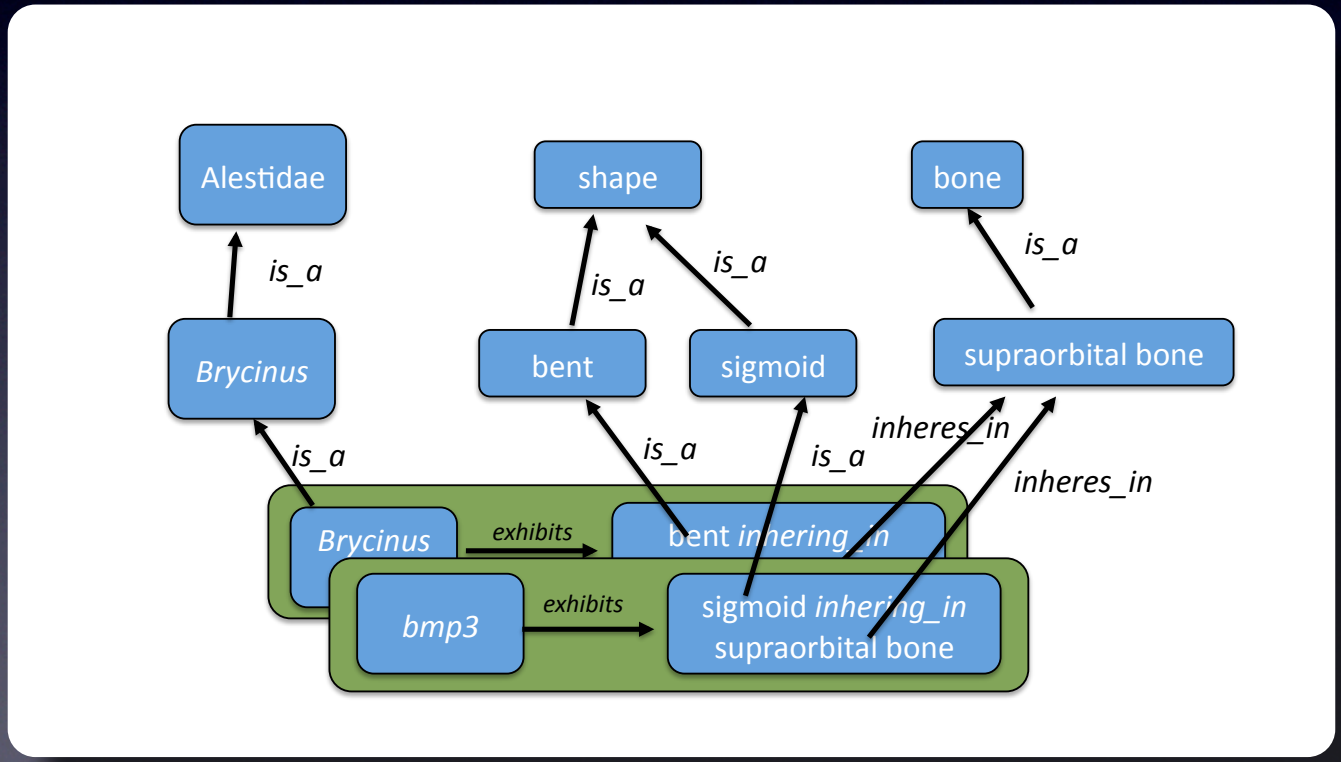
+ Zebrafish phenotype assertions



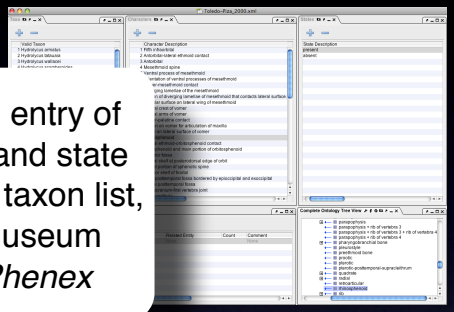
Ontologies


Brycinus
bent
sigmoid
supraorbital bone
is_a
is_a
is_a
inheres_in
inheres_in
Brycinus
exhibits
bent inhering_in
bmp3
exhibits
sigmoid inhering_in
supraorbital bone

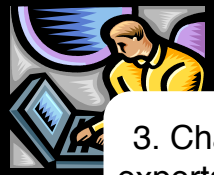
Data + Ontologies in database



2. Curation



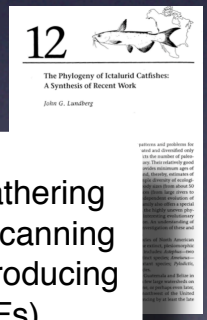
2. Student: Manual entry of free text character and state descriptions, matrix, taxon list, specimens and museum numbers using *Phenex*



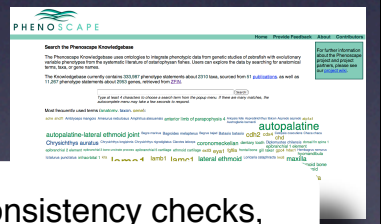
3. Character annotation by experts: Entry of phenotypes and homology assertions using *Phenex*

~ 5 person years

Curators:
Wasila Dahdul
Miles Coburn
Jeff Engemen
Terry Grande
Eric Hilton
John Lundberg
Paula Mabee
Richard Mayden
Mark Sabaj Pérez



1. Student: gathering publications (scanning hard copies, producing OCR PDFs)



4. Consistency checks, upload of data to public view of Phenoscape KB

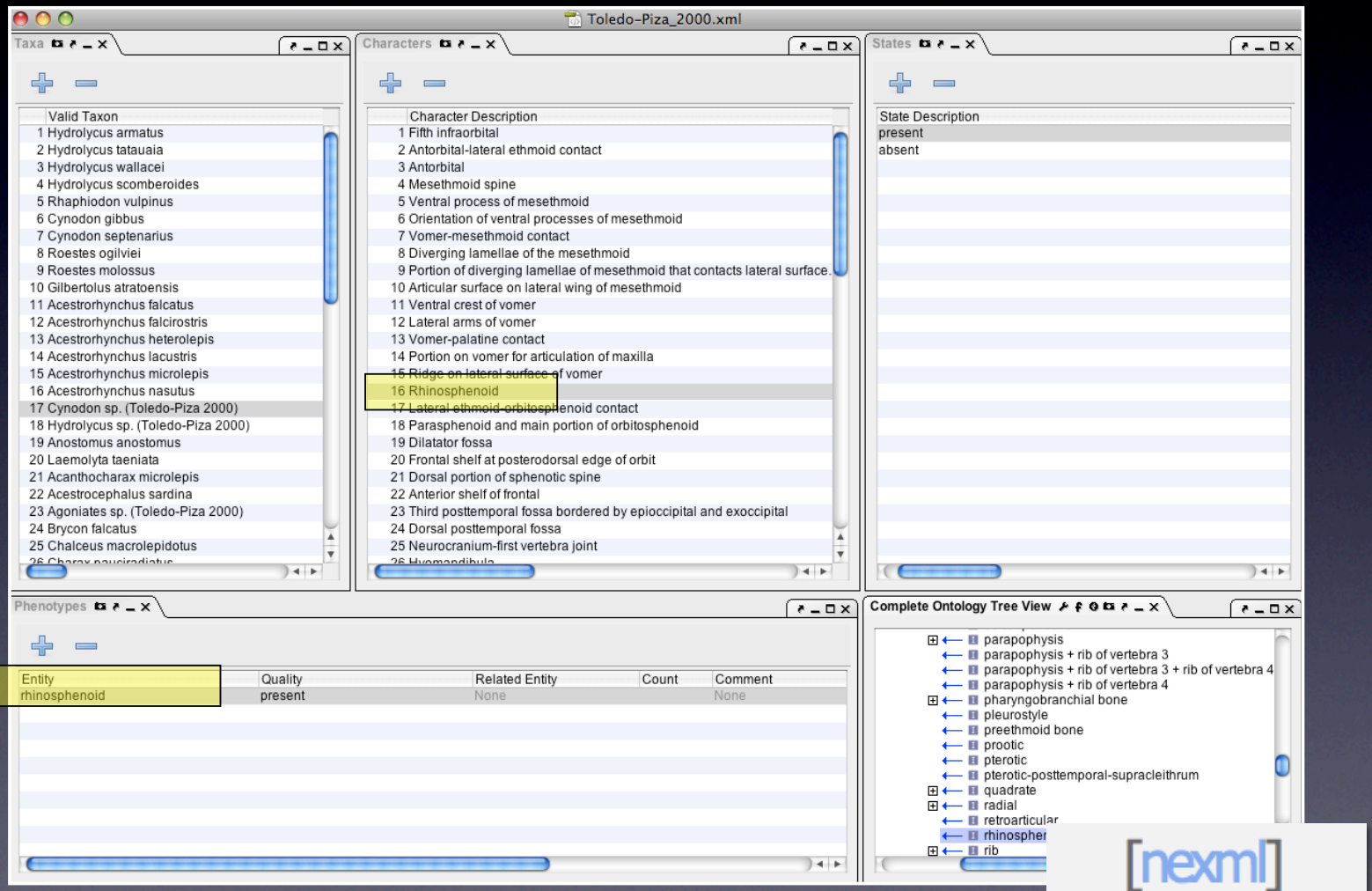
Free-text characters to phenotypes

APPENDIX 1. CHARACTER SUMMARY

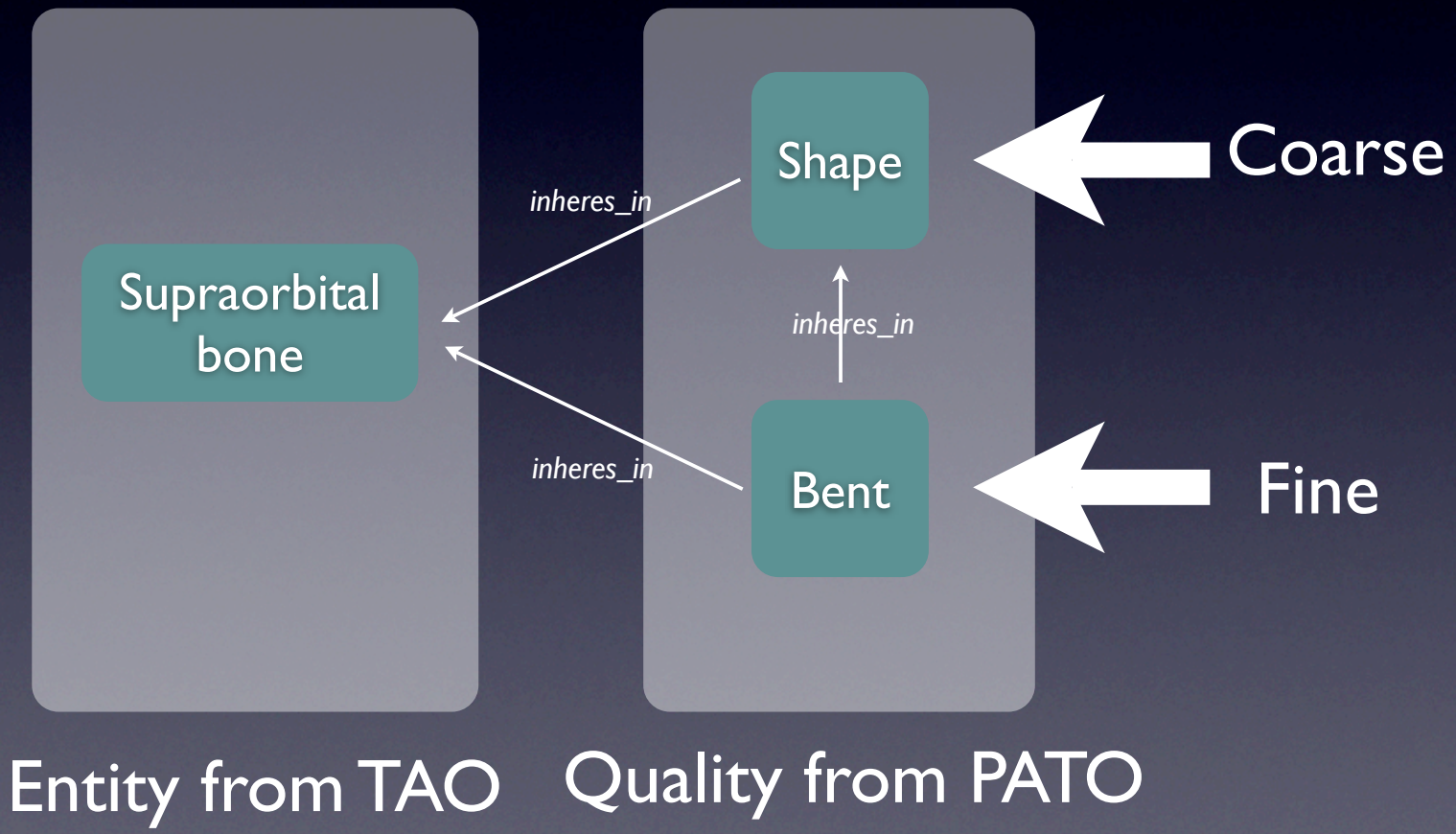
1. Fifth infraorbital. 0, well developed, without contact between fourth and sixth infraorbitals; 1, greatly reduced, with posteroventral margin of sixth infraorbital in contact with posterodorsal margin of fourth infraorbital.
2. Antorbital-lateral ethmoid contact. 0, no contact; 1, antorbital contacting 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 diftilaginous 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)

Phenex software for curation



Curated to a coarse level



Curate homology assertions

- Type of annotation supported by evidence (e.g., topological, morphological, developmental similarity)
- Enables user to view homology assertions, examine evidence

Example: Weberian apparatus

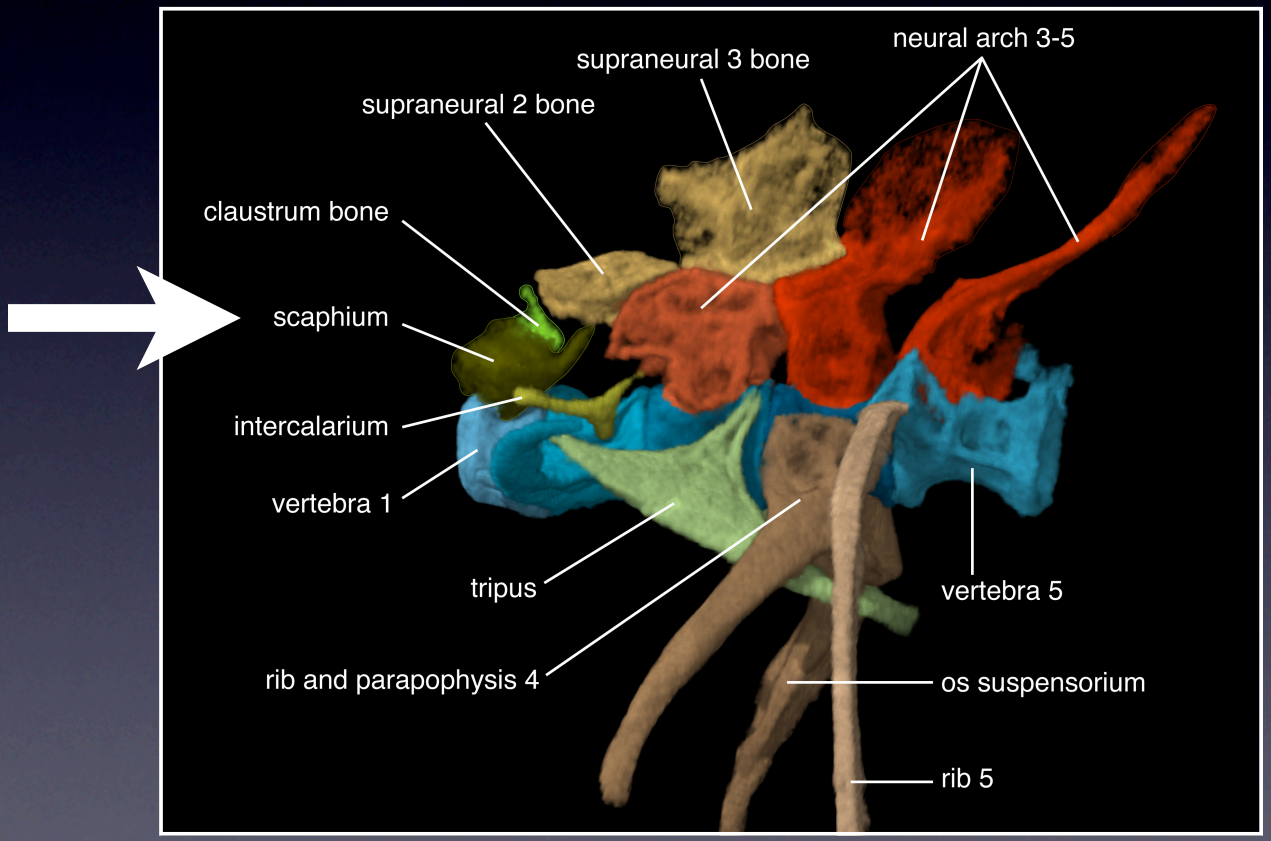


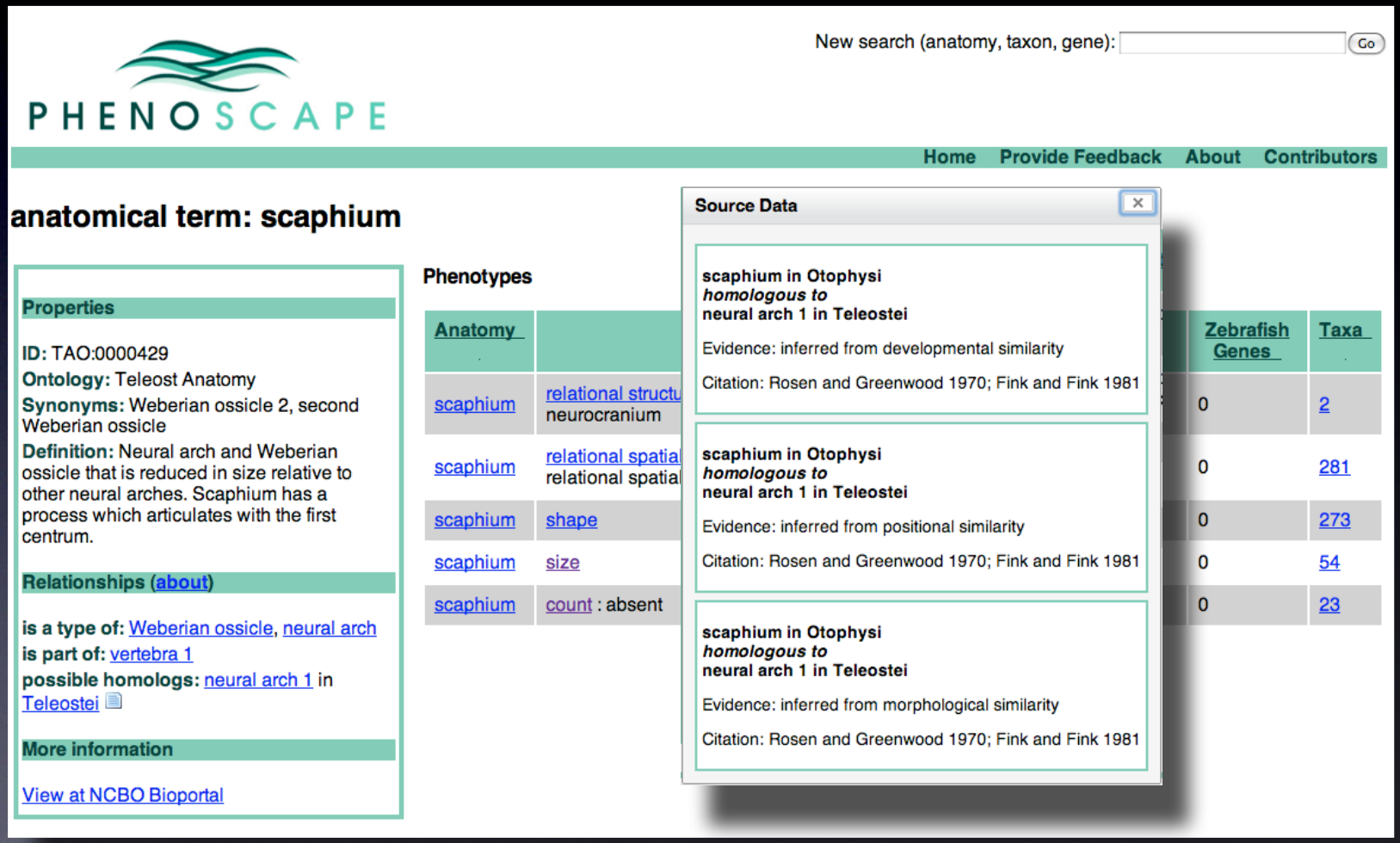
image by Kyle Luckenbill, ANSP

Curate homology assertions and evidence

Entity 1	Taxon 1	Relationship	Entity 2	Taxon 2	Evidence	Reference(s)
scaphium	Otophysi	homologous_to	neural arch 1	Teleostei	IDS, IMS, IPS	(Fink and Fink, 1981; Rosen and Greenwood, 1970)
intercalarium	Otophysi	homologous_to	neural arch 2 (ventral portion)	Teleostei	IDS, IMS, IPS	(Rosen and Greenwood, 1970)
intercalarium	Otophysi	homologous_to	neural arch 2	Teleostei	NAS	(Fink and Fink, 1981)
intercalarium	Otophysi	homologous_to	neural arch 2	Teleostei	IMS	(Hora, 1922)
intercalarium	Otophysi	homologous_to	rib of vertebra 2	Teleostei	TAS	(Hora 1922)
tripus	Otophysi	homologous_to	parapophysis + rib of vertebra 3	Teleostei	IDS, IMS, IPS	(Fink and Fink, 1981; Rosen and Greenwood, 1970)

Dahdul et al., in press. 'The Teleost Anatomy Ontology: Anatomical representation for the genomics age.' *Systematic Biology*

Homology online...



Curated 4,208 characters in 2,310 species from 51 papers

Taxon	# Papers	# Species	# Characters
Cypriniforms	8	676	794
Siluriforms	20	1724	2,110
Characiforms	10	754	1,156
Gymnotiforms	1	116	231
Gonorynchiforms	3	41	467
Clupeiforms	5	200	439
Euteleosts	3	145	582
Total	51	3,656	5,779
TOTAL ANNOTATED	**51**	**2310**	**4208**

Phenoscape KB

333,987 evolutionary phenotype assertions

11,267 phenotype statements about 2,953 genes

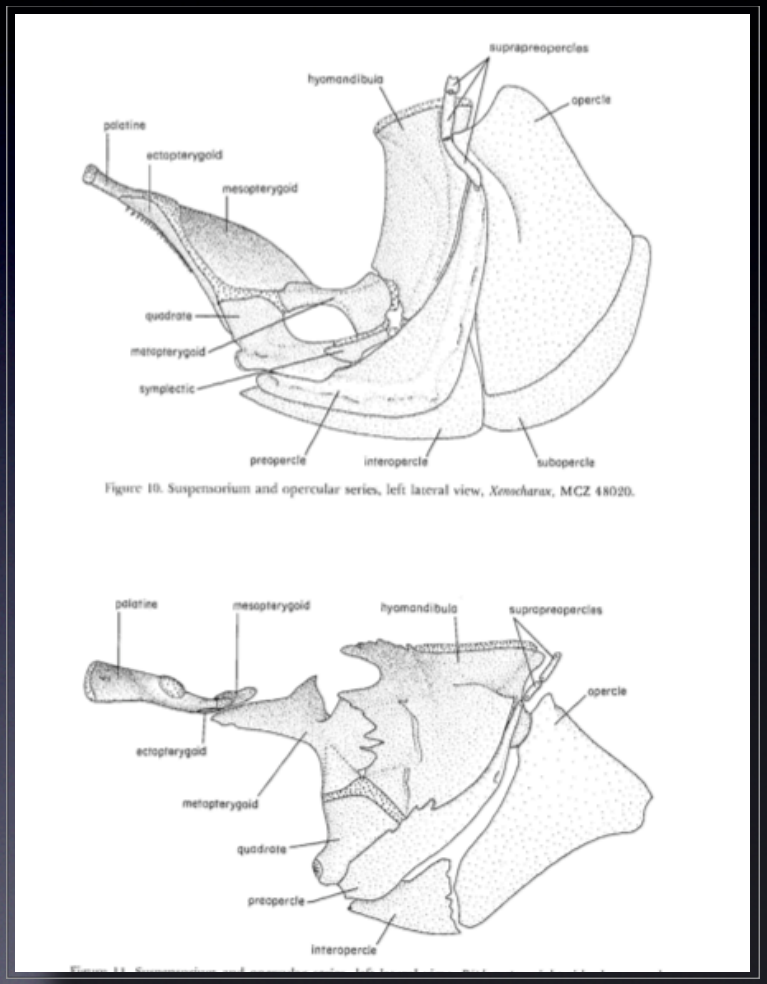
25 July 2008

Basic overview of data

1 Character state: 1 Phenotype

Character 36 (Fink & Fink, 1981). "In siluriforms the **opercle** is approximately **triangular** in shape (0) rather than approximately rectangular (1) as in other ostariophysans and primitive teleosts."

Phenotype:
E: **opercle**; Q: **triangular**



(Fink and Fink 1981)

1 Character state: 2 Phenotypes

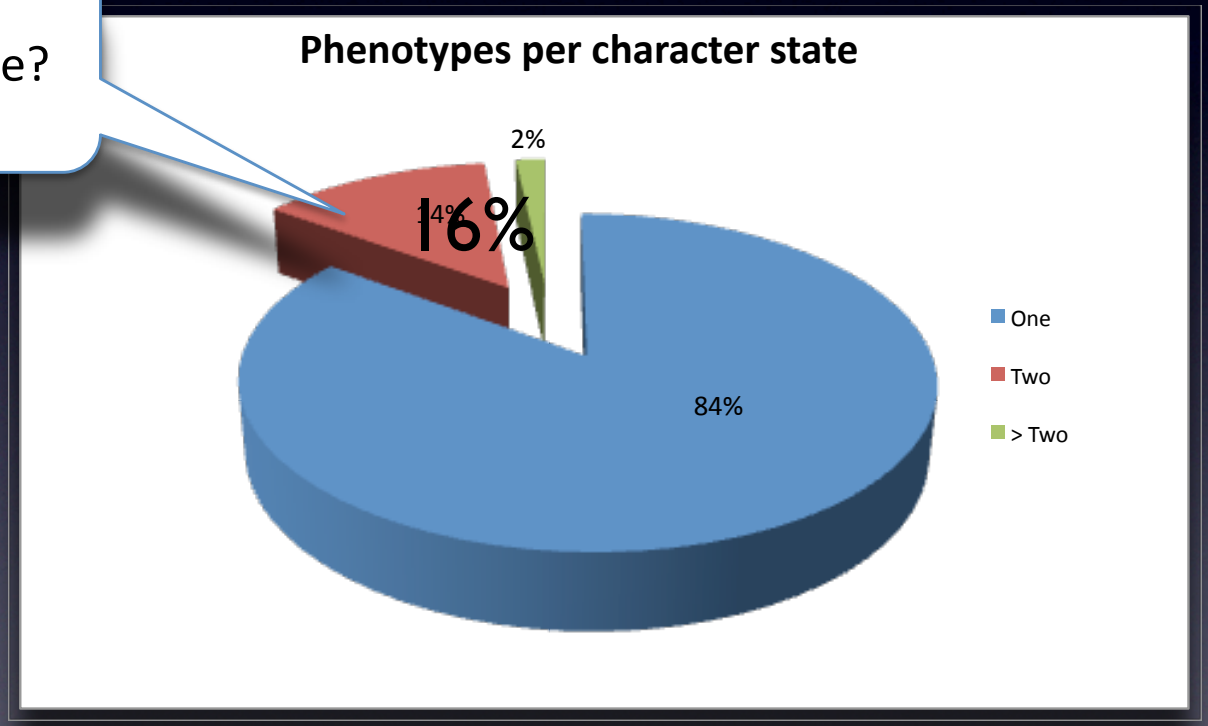
Character 80 (Zanata & Vari, 2005). "Form and area of attachment of **primordial ligament**: (0) ligament relatively **narrow** and **attaching to** posteromedial portion of **ascending process of maxilla**.....

2 Phenotypes:
E: **primordial ligament**; Q: **size, narrow**

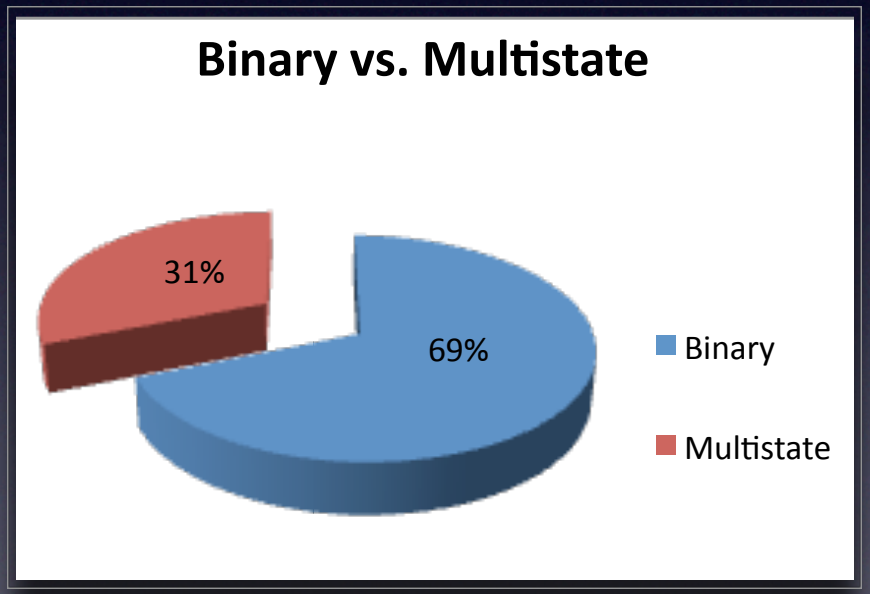
E1: **primordial ligament**; Q: **attached to;** E2**: maxilla ascending process**

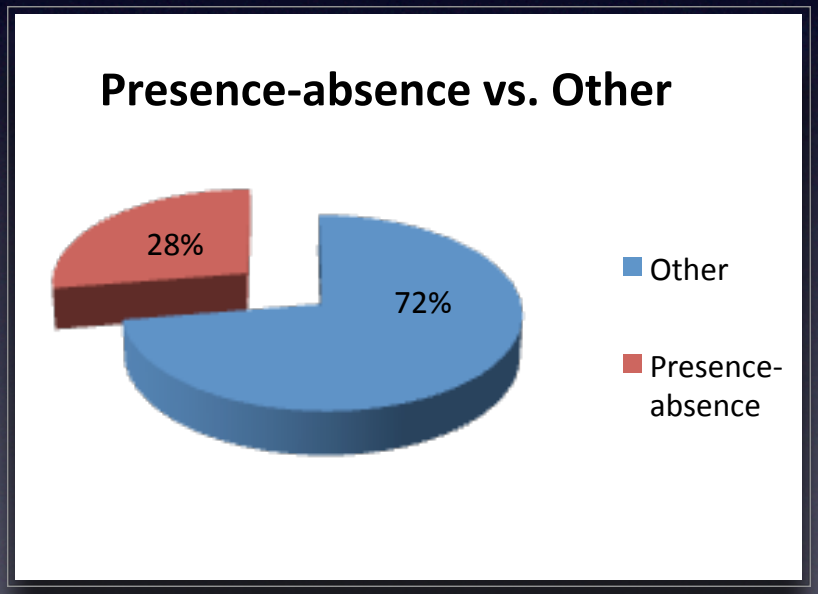
1 character state: >1 phenotype 16%

Significance?



Binary characters 69%; Presence-absence characters: 28%





Relational vs. Non-relational phenotypes

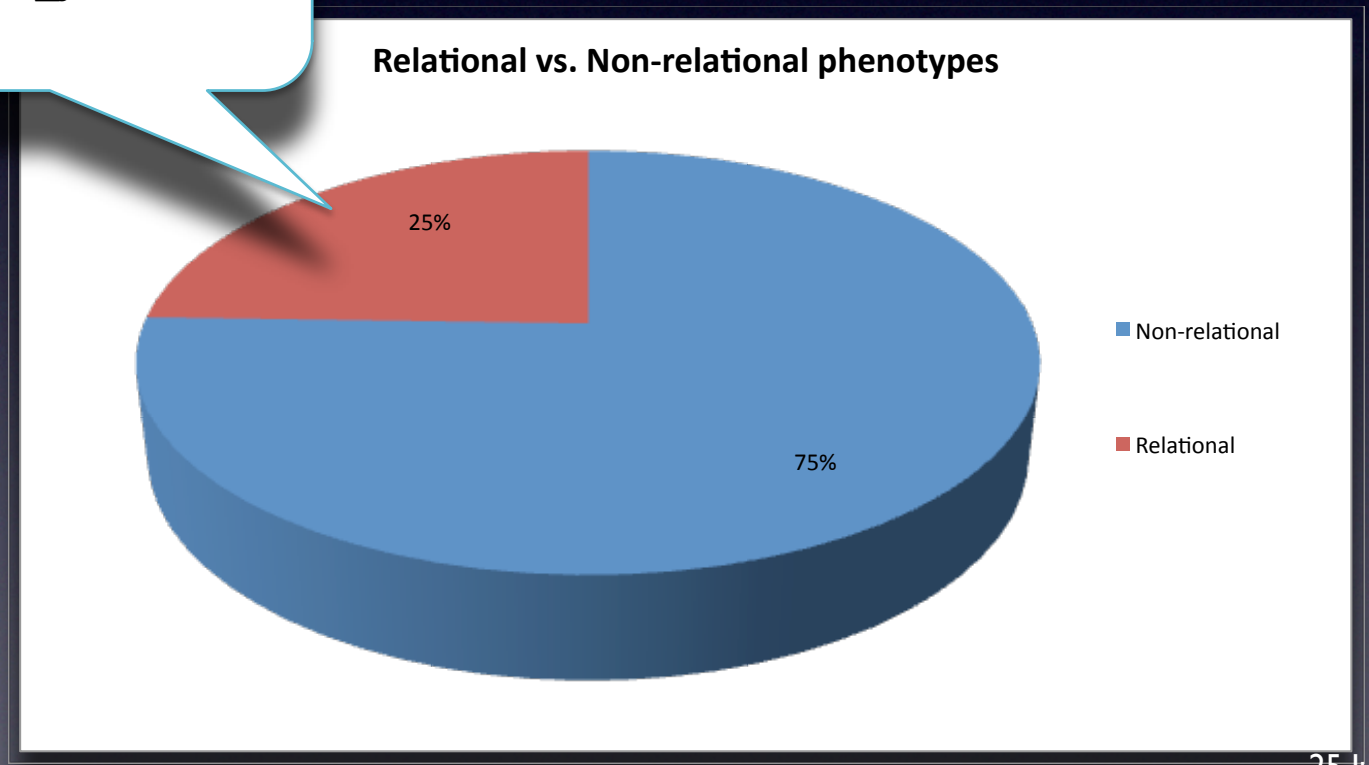
Character 80 (Zanata & Vari, 2005). "Form and area of attachment of **primordial ligament**: (0) ligament relatively **narrow** and **attaching to** posteromedial portion of **ascending process of maxilla**.....

2 Phenotypes:
E: **primordial ligament**; Q: **size, narrow**

E1: **primordial ligament**; Q: **attached to;** E2**: maxilla ascending process**

Relational phenotypes: 25%

A *attached_to* B
A *separated_from* B

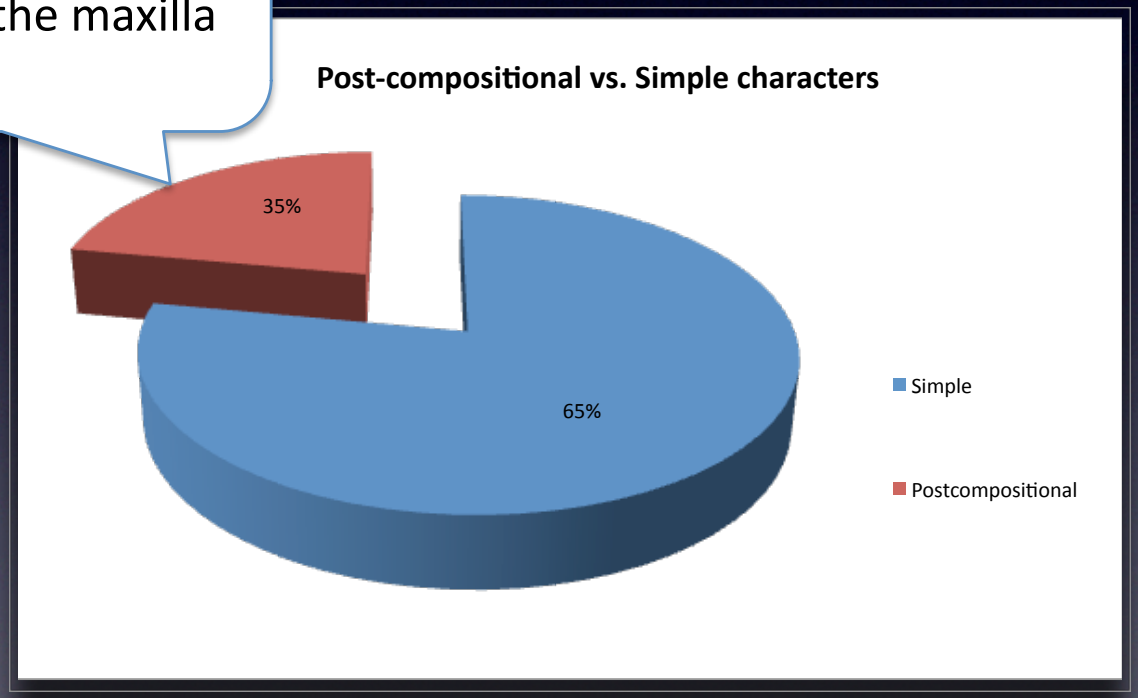


Post-compositional characters: e.g., process *part_of* maxilla

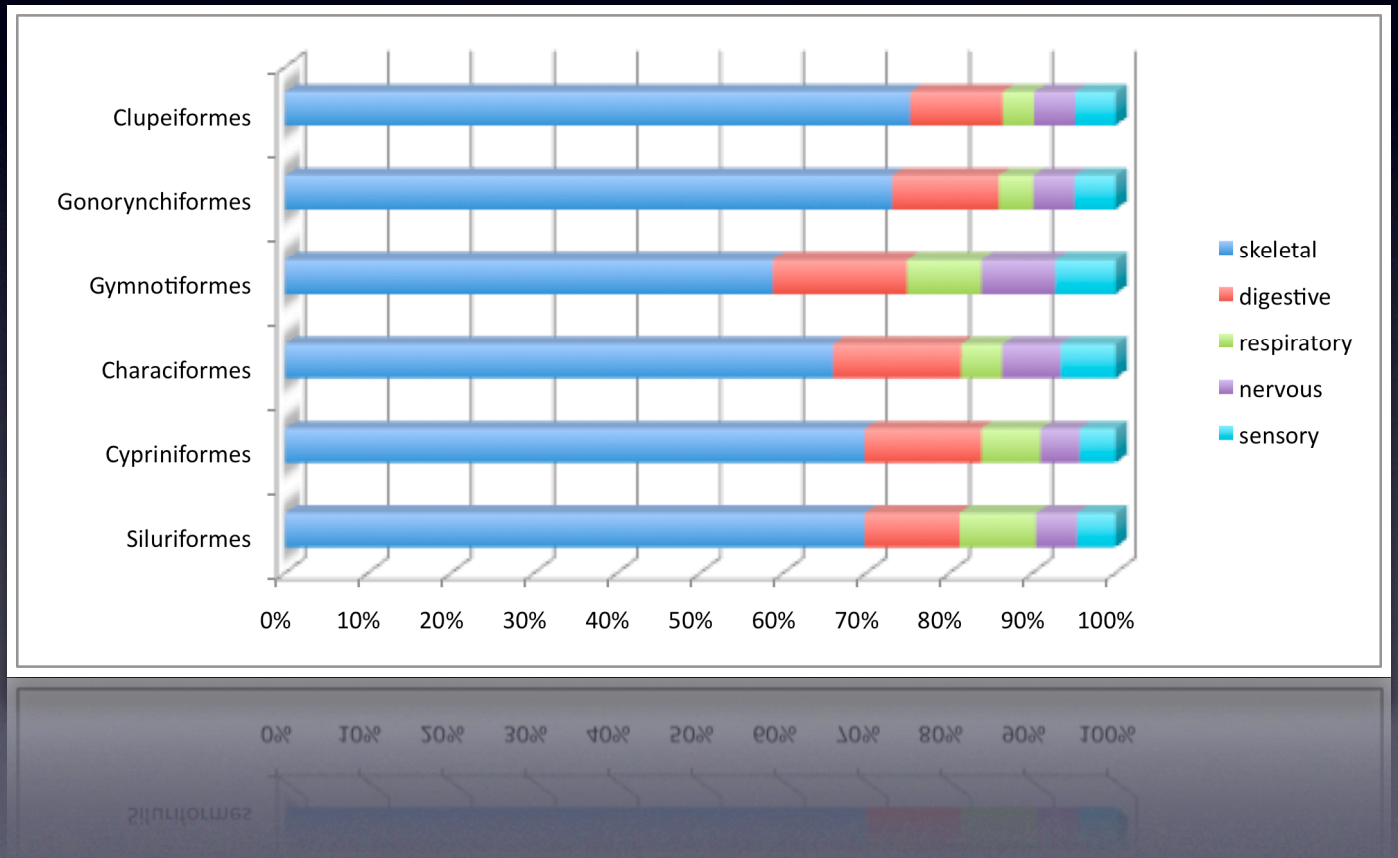
25 July 2008

Doo-dads: 35%

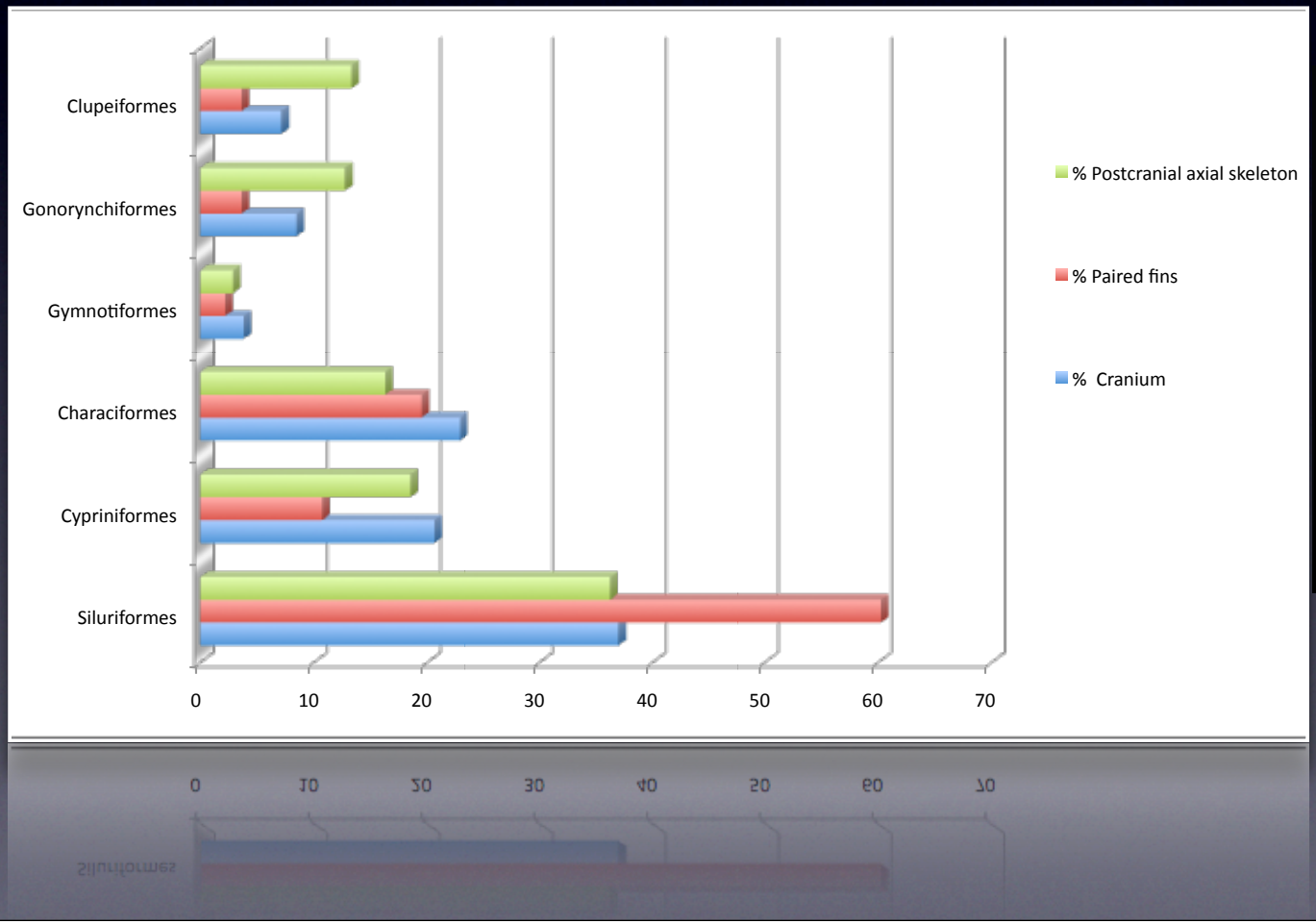
e.g., anterior process of the lateral edge of the maxilla



Distribution of all characters across anatomical systems in taxa



Distribution of skeletal characters in broad regions across taxa



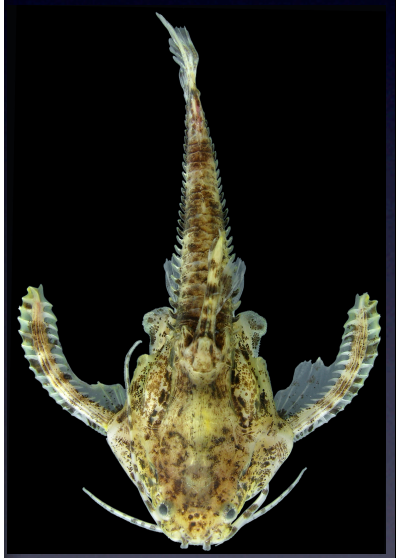


Image from Sabaj-Perez

25 July 2008

Devo-Evo applications

Search for zebrafish genes involved in fish scale development

Phenotypes

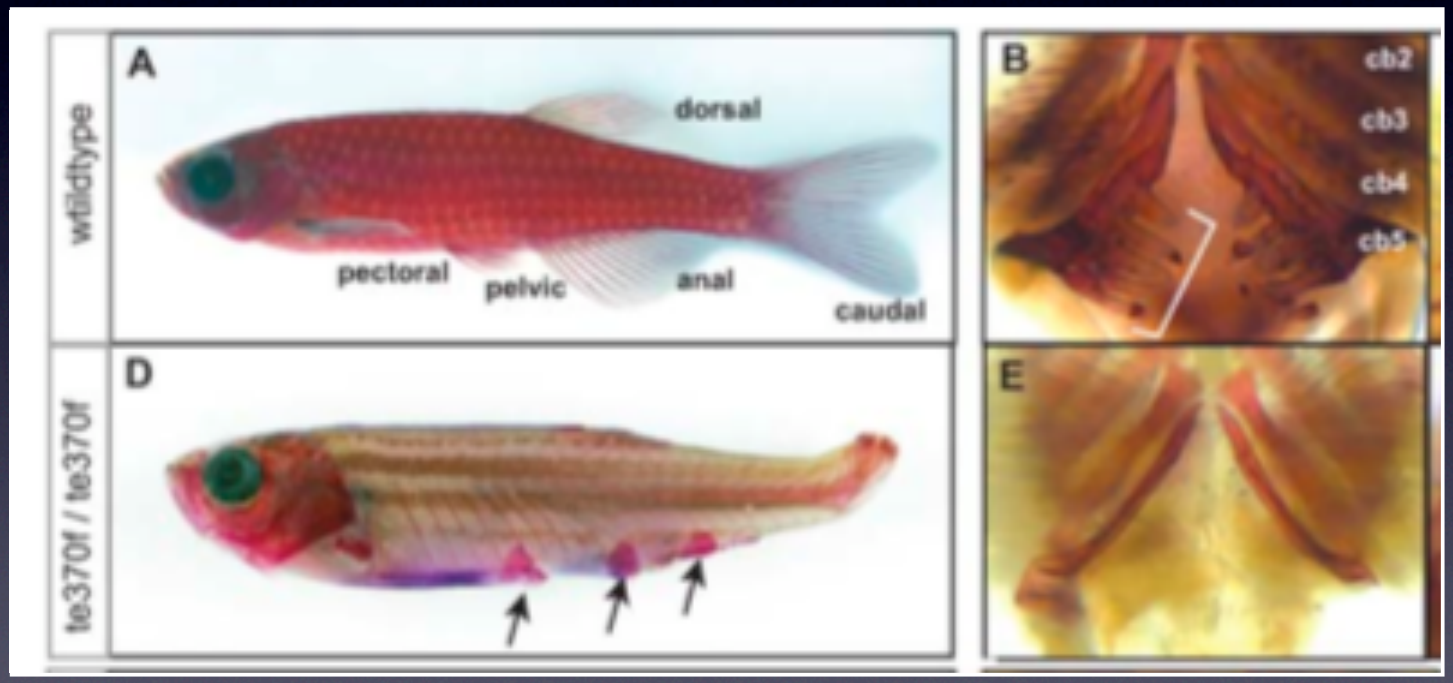
Anatomy	Quality	Zebrafish Genes	Taxa
scale	relational spatial quality : angular placement, relational spatial quality towards process of parietal bone, relational spatial quality towards supraoccipital crest	3	19
scale	texture : texture of	0	91
scale	structure	0	95
scale	position : inverted, spatial pattern	10	79
scale	shape : round	17	372
scale	quality : discontinuous, malformed, occurrence quality	6	0
scale	count : absent, count of 0, count of 0-3, count of 4-5, count of 6 and 3 more	5	74
scale	size : decreased size	5	215

Find 46 genes of interest

25 July 2008

Zebrafish *finless* mutants: scale loss

eda gene



Harris et al., 2007

Search for fishes that lack scales:

All phenotypes for scale count

- Teleostei (645).....circuli absent, circuli count, circuli present, circuli of posterior surface of scale absent, circuli of posterior surface of scale present, lateral line scale absent, lateral line scale count, lateral line scale count, 29-32, lateral line scale count, 31-58, lateral line scale count, 35-41, lateral line scale count, 42-53, lateral line scale count, 55-65, lateral line scale count, mean: 125 (range: 87–139), lateral line scale count, mean: 142 (range: 130–168), lateral line scale count, mean: 82 (range: 78–115), lateral line scale count, ≤64, lateral line scale count, ≥33, lateral line scale count, ≥76, lateral line scale present, process of scale count, process of scale of dorsal side of caudal peduncle absent, process of scale of dorsal side of caudal peduncle present, process of scale of ventral side of caudal peduncle absent, process of scale of ventral side of caudal peduncle present, scale absent, scale count, 0, scale count, 0-3, scale count, 0-4, scale count, 3-4, scale count, 4-5, scale count, 6, scale of caudal fin absent, scale of caudal fin present, scale of caudal peduncle count, scale of caudal peduncle count, >17, scale of dorsal fin count, 30-39, scale of dorsal fin count, 40-49, scale of dorsal fin count, 50-56, scale on antero-dorsal region of body absent, scale on body absent, scale on body present, scale on dorsal region of body absent, scale on dorsal region of trunk absent, scale on head absent, scale on head present, scale on post-vent region count, 1-2, scale on post-vent region count, 4-5, scale on posterior region of head absent, scale on posterior region of head present, scale on postero-lateral region of body absent, scale on postero-lateral region of body present
  - Euteleostei (16).....circuli absent, circuli of posterior surface of scale absent, lateral line scale count, 31-58, lateral line scale count, mean: 125 (range: 87–139), lateral line scale count, mean: 142 (range: 130–168), lateral line scale count, mean: 82 (range: 78–115), lateral line scale present, scale absent, scale count, 0, scale count, 0-3, scale count, 0-4, scale count, 3-4, scale count, 4-5, scale count, 6
    - Protacanthopterygii (8).....circuli absent, lateral line scale count, 31-58, lateral line scale count, mean: 125 (range: 87–139), lateral line scale count, mean: 142 (range: 130–168), lateral line scale count, mean: 82 (range: 78–115), lateral line scale present
    - Neoteleostei (8).....circuli of posterior surface of scale absent, scale absent, scale count, 0, scale count, 0-3, scale count, 0-4, scale count, 3-4, scale count, 4-5, scale count, 6
      - Percomorpha (8).....circuli of posterior surface of scale absent, scale absent, scale count, 0, scale count, 0-3, scale count, 0-4, scale count, 3-4, scale count, 4-5, scale count, 6
        - Perciformes (1).....circuli of posterior surface of scale absent
        - Smegmamorpha (7).....scale absent, scale count, 0, scale count, 0-3, scale count, 0-4, scale count, 3-4, scale count, 4-5, scale count, 6
          - Gasterosteiformes (7).....scale absent, scale count, 0, scale count, 0-3, scale count, 0-4, scale count, 3-4, scale count, 4-5, scale count, 6
            - Hypoptychidae (1).....scale count, 6
            - Aulorhynchidae (1).....scale count, 6
            - Gasterosteidae (5).....scale absent, scale count, 0, scale count, 0-3, scale count, 0-4, scale count, 3-4, scale count, 4-5
              - *Pungitius* (1).....scale count, 0-3
              - *Culaea* (1).....scale count, 3-4
              - *Apeltes* (1).....scale absent, scale count, 0
              - *Spinachia* (1).....scale count, 4-5
              - *Gasterosteus* (1).....scale count, 0-4

Find sticklebacks, catfishes, etc.

25 July 2008

Is *eda* involved in scale loss in catfishes or sticklebacks?

Ictalurus punctatus

?

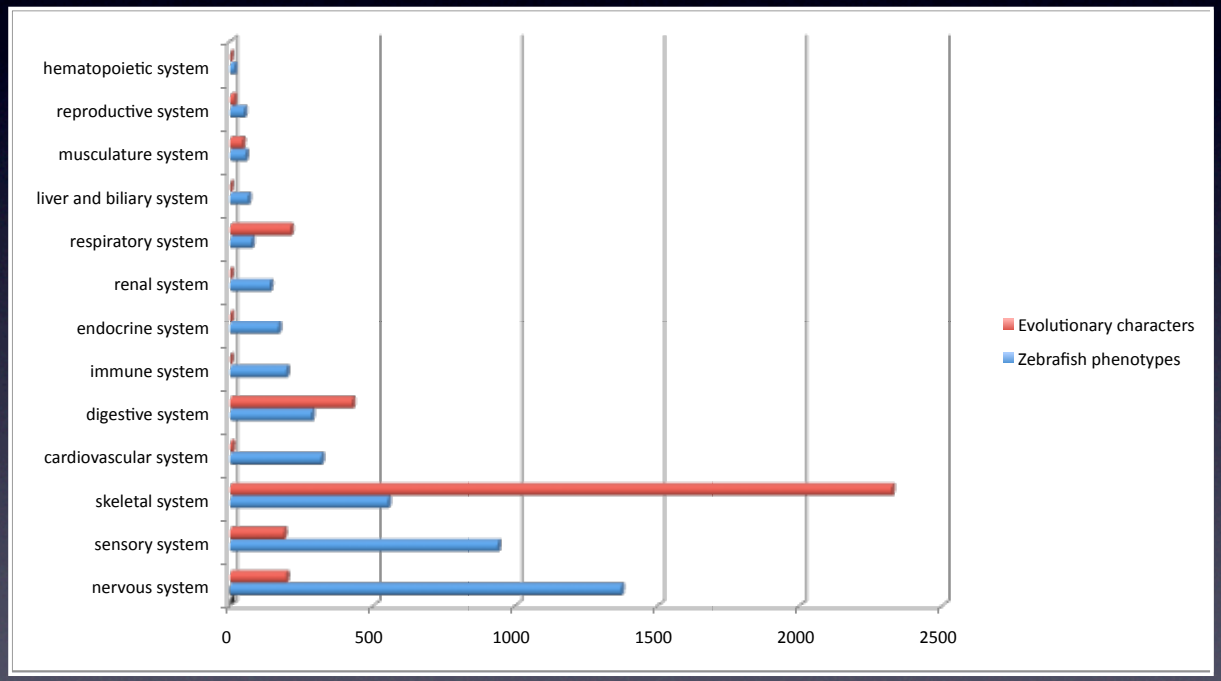
Research needed

Is *eda* involved in scale loss in sticklebacks?

Gasterosteus aculeatus

Yes -- *eda* is associated with scale loss in sticklebacks (Colosimo et al., 2005)

Potential for identifying genetic basis of evolutionary characters



4,217 zebrafish phenotypes; 3,405 evolutionary characters

25 July 2008

Summary & Conclusions

- Developed taxonomy and anatomy ontologies
- Curated character data using ontologies
- Developed database and interface to integrate searching genes and anatomy
- Makes data accessible for broad group of researchers and creates opportunities for new and synthetic research

Acknowledgements

National Science Foundation (BDI-0641025)
National Evolutionary Synthesis Center

Contributors to Teleost Ontologies

Gloria Arratia
Stan Blum
Miles Coburn
Kevin Conway
Wasila Dahdul
Mário de Pinna
Jeff Engemen
Bill Eschmeyer
Terry Grande
Melissa Haendel
Brian Hall
Eric Hilton
John Lundberg
Richard Mayden
Mark Sabaj Pérez
Brian Sidlauskas
Richard Vari
Jacqueline Webb
Edward Wiley

Curators:

Miles Coburn
Jeff Engemen
Terry Grande
Eric Hilton
John Lundberg
Paula Mabee
Richard Mayden
Mark Sabaj

Phenoscape Workshop Participants & Contributors

Arhat Abzhanov
Michael Ashburner
Judith Blake
Stan Blum
Quentin Cronk
Mário de Pinna
Andy Deans
George Gkoutos
Melissa Haendel
Hopi Hoekstra
Hans Hofmann
Elizabeth Jockusch
Elizabeth Kellogg
Chuck Kimmel
Suzanna Lewis
Anne Maglia
Austin Mast
Chris Mungall
Martin Ramirez
Sue Rhee
Martin Ringwald
Nelson Rios
Mark Sabaj Pérez
Eric Segerdell
Brian Sidlauskas
Barry Smith
David Stern
Peter Vize
Gunter Wagner
Nicole Washington
Edward Wiley

More....

- Phenoscape poster Sunday evening
- Phenoscape hands-on demonstrations, this afternoon, Exec Towers, Salon II
- phenoscape.org
- kb.phenoscape.org