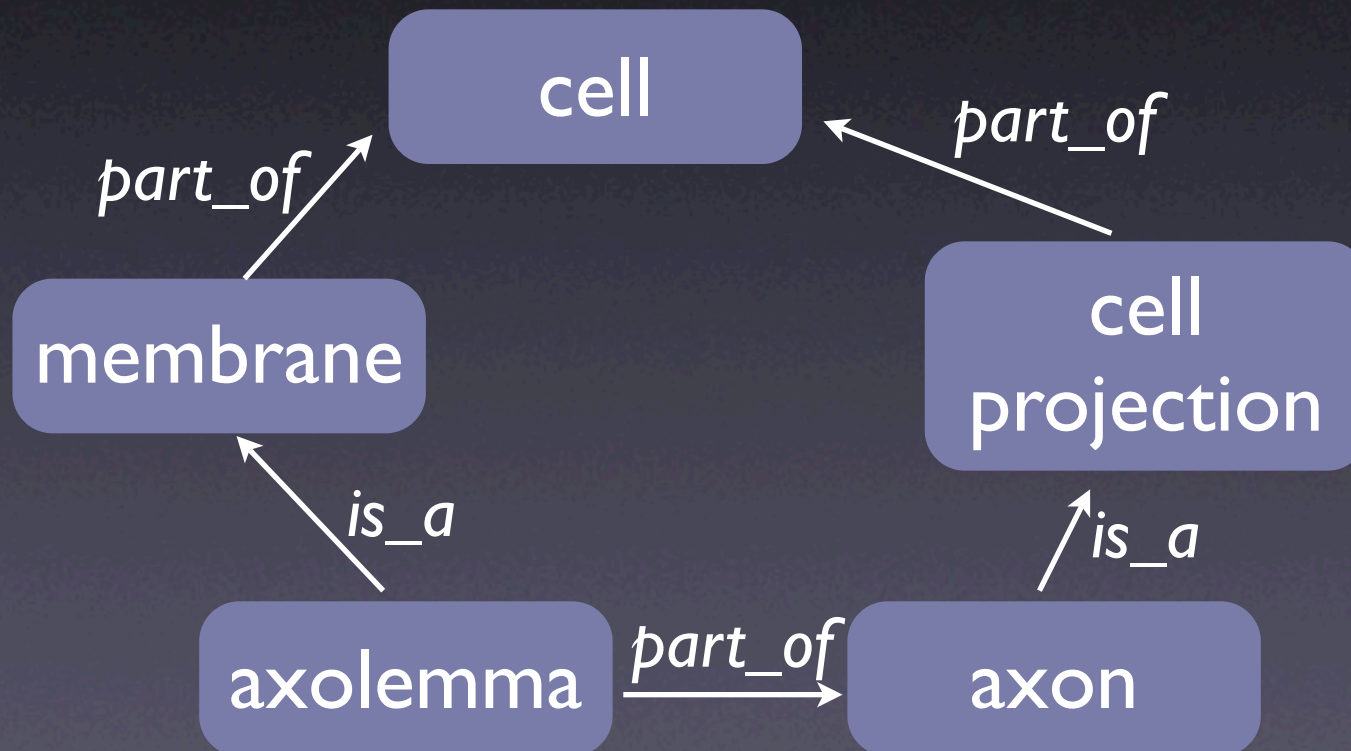


# Annotating Evolutionary Data Matrices Using Ontologies

Jim Balhoff, NESCent

# Ontology

Ontology == controlled vocabulary, defining terms as well as their relationships



# EQ phenotypes

- Ontologies are being used for annotation within genomics communities for mutant descriptions (FlyBase, ZFIN)
- EQ: Entity + Quality (“dorsal fin” “round”)
- Entities are taken from anatomy or gene ontology
- Qualities in PATO (Phenotype and Trait Ontology)

# Character Matrix

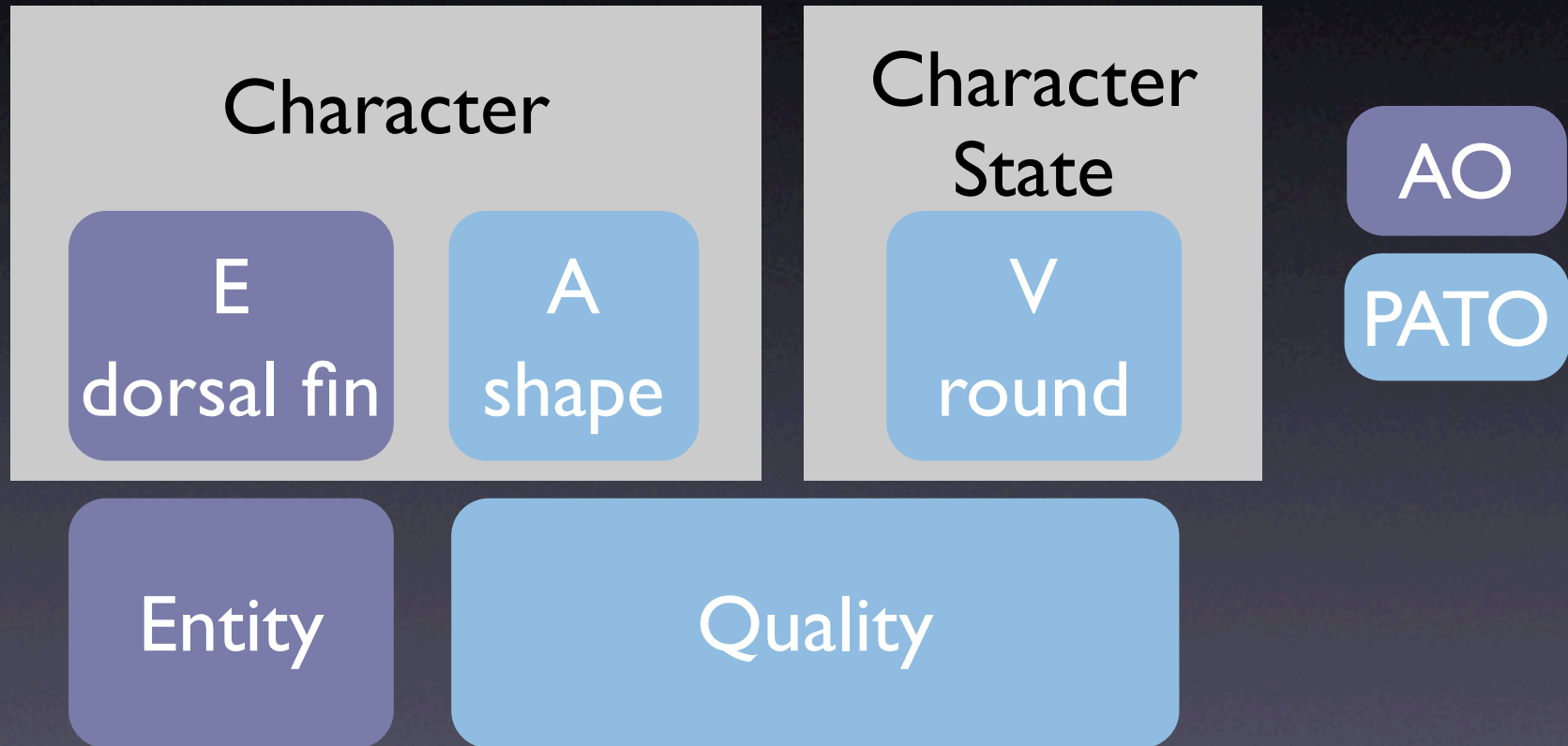
- Common phenotypic annotation method in evolutionary biology
- Characters + character states, similar to EQ

	Dorsal fin shape
<i>Species one</i>	round
<i>Species two</i>	pointed
<i>Species three</i>	undulate

# EQ for evolutionary phenotypes

- Fish Evolution working group at NESCent - ZFIN + Cypriniformes Tree of Life
- Obvious benefits of data integration - candidate gene identification, character independence, etc.

# Character Matrix EQ



# EQ software

- EQ annotations typically done via custom web interfaces (for genome projects) or Phenote
- Phenote - Java-based standalone application, loads ontologies and constructs list of EQ statements
- Has no matrix perspective

# Phenote

Phenote 0.9.2 dev

File

Pub

Genotype

Genetic Context

Entity **Fly**

Quality

Search:

- Terms
- Synonyms
- Definitions
- Obsolete

Pub	Genotype	Genetic Context	Entity	Quality
			leg	long length
			eye disc	irregular
			anterior notopleural bri...	green

**Term Info**

**TERM:** green  
**ID:** PATO:0000320

**PARENTS:**  
**Superclass(ISA):** [color](#)

**CHILDREN:**  
**Subclass(ISA):** [dark green](#)  
**Subclass(ISA):** [fluorescent green](#)  
**Subclass(ISA):** [bright green](#)  
**Subclass(ISA):** [light green](#)  
**Subclass(ISA):** [dull green](#)

**Definition:** A hue of that portion of the visible spectrum lying between yellow and blue, evoked in the human observer by radiant energy with wavelengths of approximately 490 to 570 nanometers;



# Character matrix software

- MacClade, NDE, Mesquite
- None have any facility for ontology use
- Mesquite - Java-based, open-source, plug-in driven

# Mesquite

Character Matrix "Matrix in file "TahitiCyrtandraDRAFT1.nex""

Graphics Text Parameters Modules Citations

Taxon \ Character	1	2	3	4	5	6
	Leaf blade base	Leaf margins	Hairiness of calyx inner surface	Pubescence of foliage buds	Corolla length	*Shape of leaf blade
1 *C. apiculata	free	serrate	pubescent inner surface	present	2-5cm	lanceolate-elliptic
2 *C. bidwillii	free	serrate	pubescent inner surface	present	2-5cm	lanceolate-elliptic
3 *C. biflora	free	entire to serrate	pubescent outer surface	absent	>5cm	lanceolate-elliptic
4 *C. connata	connate	entire	pubescent outer surface	absent	?	lanceolate-oblong
5 *C. glabra	free	entire	pubescent outer surface	present	2-5cm	elliptic
6 *C. geminiflora	free	entire	?	present	2-5cm	elliptic
7 *C. induta	free	serrate	pubescent inner surface	present	>5cm	elliptic-ovate
8 *C. mooreaensis	free	entire to serrate	pubescent outer surface	present	<2cm	lanceolate-elliptic
9 *C. mucronata	free	entire to serrate	pubescent inner surface/pubescent outer surface	present	2-5cm	elliptic-ovate
10 *C. nadeaudii	free	entire to serrate	pubescent inner surface	present	2-5cm	lanceolate-elliptic
11 *C. taitensis	free	entire to serrate	pubescent outer surface	present	2-5cm	elliptic-ovate
12 *C. tohivaensis	free	entire to serrate	pubescent inner surface	present	>5cm	elliptic-ovate
13 *C. vairiae	free	entire	pubescent inner surface	absent	2-5cm	lanceolate
14 *C. vescoi	free	entire to serrate	pubescent inner surface	present	2-5cm	elliptic

[t.14 c.5 s.1] Corolla length: (0) <2cm; (1)\* 2-5cm; (2) >5cm; [in taxon "C. vescoi"]

# Ontology-based character matrix annotation

- Increase semantic structure of character matrix data
- Evolutionary biologist should be able to use matrix perspective while choosing terms from ontologies
- Use Phenote to add ontology capabilities to Mesquite?

# Acknowledgments

- NESCent working group: “Towards an Integrated Database for Fish Evolution” - Paula Mabee, Monte Westerfield
- Todd Vision, Hilmar Lapp (NESCent)
- Chris Mungall, Mark Gibson (NCBO)