Phenoscape Data Jamboree - Final Report

April 18-21, 2008
NESCent, Durham, NC

A hosted workshop supported by NSF BDI-0641025

1. Participants

Project leaders
- Paula Mabee, pmabee@usd.edu, University of South Dakota
- Monte Westerfield, monte@uoneuro.uoregon.edu, University of Oregon
- Todd Vision, tjv@bio.unc.edu, NESCent and UNC Chapel Hill

Phenoscape personnel
- Jim Balhoff, balhoff@nescent.org, NESCent
- Wasila Dahdul, dahdul@acnatsci.org, Academy of Natural Sciences
- Hilmar Lapp, hlapp@nescent.org, NESCent
- John Lundberg, lundberg@acnatsci.org, Academy of Natural Sciences
- Peter Midford, peteremidford@yahoo.com, University of Kansas

Guest data curators
- Miles Coburn, coburn@jcu.edu, Jonh Carroll University
- Kevin Conway, conwaykw@gmail.com, St. Louis University
- Maŕio de Pinna, pinna@ib.usp.br, Universidade de Saõ Paulo
- Brian Sidlauskas, bls16@duke.edu, NESCent

Advisors
- Martin Ringwald, ringwald@informatics.jax.org, Jackson Laboratories
- Nicole Washington, nlwashington@lbl.gov, University of California - Berkeley

2. Workshop goals

Phenoscape (http://phenoscape.org) is a project (funded by the NSF Biological Databases & Informatics program) that arose from the NESCent Working Group "Towards an Integrated Database for Fish Evolution", led by Paula Mabee (a PI on the NSF Cypriniformes Tree of Life grant) and Monte Westerfield (head of the Zebrafish Information Network, zfin.org). The aim of Phenoscape is to develop tools for machine-reasoning on phenotype data from evolutionary morphology and model organism developmental genetics (Mabee et al. 2007a, Mabee et al. 2007b), using ostariophysan fishes as a proof of principle. In its first year, the project has developed customized data curation software (using Phenote as a framework, http://phenote.org), developed ontology resources (most importantly a new multi-species Teleost Anatomy Ontology and a Teleost Taxonomy Ontology), and established a curatorial workflow for annotating systematic character data using the same entity-quality syntax being used by genetic model organism databases, in particular the zebrafish database, ZFIN.

The aim of the jamboree was to test the data curation workflow and tools with the help of several ichthyological morphologists serving in the role of guest data curators. Project personnel, together with two external advisors, were on hand to help acquaint the guest data curators with the software, concepts, workflow and tools, and to record and discuss issues arising as the guest curators wrestled with real character data. An experiment that tested consistency of character representation among curators was conducted. This was followed by discussion and wrap-up, and later a general project meeting.

3. Summary of activities and discussion

User documentation was collected on the wiki in preparation for the jamboree (http://phenoscape.org/wiki/Data_Jamboree_1, section "Resources"). The workshop started with introductions by participants and a tutorial in the use of Phenote and the associated ontologies to annotate phenotypes using EQ syntax. Once guest curators were familiar with the workflow and tools, they were paired with project personnel and began entering data from pre-selected publications. The selected publications were either authored by the guest data curators themselves or within their area of specialty:

- Miles Coburn (Cavender and Coburn 1992, Coburn and Cavender 1992, Hoffman and Britz 2006, Sawada 1982, Weitzman 1962)
- Kevin Conway (Conway 2005, Conway and Mayden 2007, Conway et al. 2008)
- Mario de Pinna (de Pinna 1992, de Pinna 1996a, de Pinna 1996a, de Pinna and Grande 2003, de Pinna et al 2007)
- Brian Sidlauskas (Vari et al 1995, Sidlauskas and Vari in press).

Problems and issues were discussed as they arose, and collected on the project wiki. Topics of discussion included issues with the structure of the ontologies, such as the sometimes ambiguous distinction between shape and size annotations, which are currently considered distinct by PATO (the Phenotype and Trait Ontology, from which phenotype qualities are derived). A major topic of discussion included the question of what is implied, or not implied, about homology when the same anatomical term is used for annotations in different species. Other issues involved challenges in translating complex phenotypic descriptions from the context of a comparative systematic publication to the context of the Phenoscape database. For instance, when a published character state requires comparison with alternative states in different taxa, how can those be defined by reference only to the anatomy of the taxon under consideration? A further set of issues concerned the Phenote user interface, such as how to facilitate annotation of phenotypes for large collections of taxa.

Following this activity, we conducted an experiment to determine how often, and for what reasons, curators choose divergent EQ conceptualizations for the same character and character states. Four curators (Coburn, Conway, dePinna and Lundberg) used Phenote to encode EQ annotations for the same 10 character/state descriptions (plus one extra credit description), and the results were compiled and reviewed immediately afterward with the group. Only one of the 10 characters was annotated identically among all four curators. The reasons why the other annotations differed among curators were revealing, and ranged from Phenote interface bugs, to difficult aspects of the ontologies (e.g. lack of quality terms), to a lack of standardized guidelines for certain special cases, to differing interpretation of the text descriptions. These results were discussed with the advisors in order to prioritize effort on Phenote development, ontology refinement, and improvement of the curation process more generally.

Following the departure of the guest data curators and advisors, the project personnel conducted an all-hands meeting to review progress and plan for activities over the coming months. Major items for discussion included community engagement activities, future project meetings, participation of project members in two upcoming workshops (on OBO Relations and Morphbank), planned publications (see below), and revisions to curation workflow.

4. Strategy and plans for follow-up activities

The meeting suggested several major changes to the curation workflow, and these will be implemented by changes to the software and curation guidelines in the coming months. In addition, some of the preparatory work required will be the basis for summer internship projects planned for this summer. The revised workflow will be tested at the next data jamboree, tentatively planned for September 2008 in South Dakota.

5. Anticipated outcomes and products

We plan to document the outcomes of the curation experiment on the project wiki (http://phenoscape.org). Plans were further made to syndicate blog posts and establish a mailing list for wider community engagement. We anticipate considerable changes to Phenote being released in the coming months (including both bug fixes and major revision of the Phenoscape curation interface), with possible work on a Mesquite module, as well. Balhoff and Washington began work on a data model for Phenote which will, over the coming months, form the basis for the Phenoscape database. This database will eventually hold all the evolutionary morphology data entered by our collaborators and will serve as the backend to the user interface that will enable the queries and visualization tools. The revised workflow will enable us, together with curators from the community, to begin entering data into the Phenoscape system, and we plan to start by curating several hundred characters from several dozen papers on the ostariophysans (see below). Midford will propose major changes to taxonomy ontologies in OBO and describe the system in a forthcoming publication. Similarly, project members plan to draft a paper discussing some of the subtleties involved in encoding homology statements internally or externally to a multispecies ontology.

6. References

Cavender, Ted M., and Miles M. Coburn. 1992. Phylogenetic relationships of North American cyprinids. Pp. 293-327. In: Systematics, Historical Ecology, and North American Freshwater Fishes. R.L. Mayden (Ed.). Stanford University Press. Stanford, California. 969 p.

Coburn, Miles M., and Ted M. Cavender. 1992. Interrelationships of North American cyprinid fishes. Pp. 328-373 In: Systematics, Historical Ecology, and North American Freshwater Fishes. R.L. Mayden (Ed.). Stanford University Press. Stanford, California. 969 p.

Conway, K. W., W-J Chen, and R. L. Mayden. 2008. The "Celestial Pearl danio" is a miniature Danio (s.s) (Ostariophysi: Cyprinidae): evidence from morphology and molecules. Zootaxa 1686:1-28

Conway, K. W., and R. L. Mayden. 2007. Gill Arches of Psilorhynchus (Ostariophysi: Psilorhynchidae). Copeia: 2007:267-280

Conway, K. W. 2005. Monophyly of the genus Boraras (Teleostei: Cyprinidae)

de Pinna MCC. 1996. A phylogenetic analysis of the Asian catfish families Sisoridae, Akysidae, and Amblycipitidae, with a hypothesis on the relationships of the neotropical Aspredinidae (Teleostei, Ostariophysi). Fieldiana: Zoology (New Series) 84:1-83.

de Pinna MCC, Ferraris CJJ, Vari RP. 2007. A phylogenetic study of the neotropical catfish family Cetopsidae (Osteichthys, Ostariophysi, Siluriformes), with a new classification. Zoological Journal of the Linnean Society 150:755-813.

de Pinna MCC. 1992. A new subfamily of Trichomycteridae (Teleostei, Siluriformes), lower loricarioid relationships and a discussion on the impact of additional taxa for phylogenetic analysis. Zoological Journal of the Linnean Society 106:175-229.

de Pinna MCC, Grande T. 2003. Ontogeny of the accessory neural arch in pristigasteroid clupeomorphs and its bearing on the homology of the otophysan claustrum (Teleostei). Copeia:838-845.

de Pinna MCC. 1996. Teleostean monophyly Interrelationships of Fishes. New York: Academic Press. 147-162.

Hoffmann, M. and R. Britz. 2006. Ontogeny and homology of the neural complex of otophysan Ostariophysi. Zool. J. Linn. Soc. 147(3):301-330.

Mabee PM, Ashburner M, Cronk Q, Gkoutos GV, Haendel M, Segerdell E, Mungall C, and Westerfield M. 2007a. Phenotype ontologies: the bridge between genomics and evolution. Trends Ecol Evol 22:345-50.

Mabee PM, Arratia G, Coburn M, Haendel M, Hilton EJ, Lundberg JG, Mayden RL, Rios N, and Westerfield M. 2007b).Connecting evolutionary morphology to genomics using

ontologies: a case study from Cypriniformes including zebrafish. J Exp Zoolog B Mol Dev Evol 308B:655–668.
Sawada, Y. 1982. Phylogeny and zoogeography of the superfamily Cobitoidea (Cyprinoidei: Cypriniformes). Mem. Fact. Fish. Hokkaido Univ. 28:65-223.
Sidlauskas B, Vari RP. in press. Phylogenetic relationships within the South American fish family Anostomidae (Teleostei, Ostariophysi, Characiformes). Zoological Journal of the Linnean Society.
Vari RP, Castro RMC, Raredon SJ. 1995. The Neotropical fish family Chilodontidae (Teleostei: Characiformes): a phylogenetic study and a revision of Caenotropus Günther. Smithsonian Contributions to Zoology 577:32.
Weitzman, S.H. 1962. The osteology of Brycon meeki, a generalized characid fish, with an osteological definition of the family. Stanford Ichthy. Bull. 8(1):1-77.