Review

500,000 fish phenotypes: The new informatics landscape for evolutionary and developmental biology of the vertebrate skeleton

By P. Mabee[1,*], J. P. Balhoff[2], W. M. Dahdul[1,2], H. Lapp[2], P. E. Midford[2], T. J. Vision[2,3] and M. Westerfield[4]

[1]*Department of Biology, University of South Dakota, Vermillion, South Dakota, USA;* [2]*National Evolutionary Synthesis Center, Durham, North Carolina, USA;* [3]*Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA;* [4]*Institute of Neuroscience, University of Oregon, Eugene, Oregon, USA*

Summary

The rich phenotypic diversity that characterizes the vertebrate skeleton results from evolutionary changes in regulation of genes that drive development. Although relatively little is known about the genes that underlie the skeletal variation among fish species, significant knowledge of genetics and development is available for zebrafish. Because developmental processes are highly conserved, this knowledge can be leveraged for understanding the evolution of skeletal diversity. We developed the Phenoscape Knowledgebase (KB; http://kb.phenoscape.org) to yield testable hypotheses of candidate genes involved in skeletal evolution. We developed a community anatomy ontology for fishes and ontology-based methods to represent complex free-text character descriptions of species in a computable format. With these tools, we populated the KB with comparative morphological data from the literature on over 2500 teleost fishes (mainly Ostariophysi) resulting in over 500,000 taxon phenotype annotations. The KB integrates these data with similarly structured phenotype data from zebrafish genes (http://zfin.org). Using ontology-based reasoning, candidate genes can be inferred for the phenotypes that vary across taxa, thereby uniting genetic and phenotypic data to formulate evo-devo hypotheses. The morphological data in the KB can be browsed, sorted, and aggregated in ways that provide unprecedented possibilities for data mining and discovery.

Introduction

In response to the enormous challenge presented by the deluge of new data, biologists have embarked on a new voyage of exploration and discovery using bioinformatics. Although the emphasis has been on genomic data (Pennisi, 2011), there is growing recognition that a corresponding sea of phenomic data must also be organized and made computable in relation to genomic data. Phenotypes are the observable features of an organism such as anatomy, behavior, and the development of these traits. Mapping the genome to the phenome of an organism, and integrating such data with evolutionary change-holds a high potential for scientific discovery if the challenges of data organization and access can be overcome. Significant efforts have been mounted to integrate genetic and phenotypic information in evolutionary biology (Mabee et al., 2007a; Dahdul et al., 2010b), biodiversity (Deans et al., 2012), biomedicine (Washington et al., 2009), and agriculture (Faccioli et al., 2009).

Fish skeletal biology provides an ideal testing ground for data integration, given its rich history in comparative anatomy (Cuvier and Valenciènnes, 1846), phylogenetic systematics (Williams and Ebach, 2008), and developmental biology (Grunwald and Eisen, 2002). The purpose of this paper is to describe the informatics approach initiated by the Phenoscape project (Mabee et al., 2007a; Dahdul et al., 2010a,b) that has the potential to transform the way development and evolution of the fish skeleton are studied and understood.

Two problems in evo-devo

Biologists working at the intersection of the fields of evolution and development face two central problems when attempting to integrate data from both fields: (i) the difficulty for an expert in the comparative anatomy and evolution of a taxon to discover candidate genes for evolutionary phenotypes, and (ii) the difficulty for an expert in the molecular development and genetics of a model organism to recover the taxonomic distribution of a particular feature across the tremendous breadth of species.

An example of first problem, the difficulty of candidate gene discovery, is illustrated by numerous publications describing the comparative skeletal anatomy of fishes. For example, Wiley and Johnson (2010) collected and reviewed morphological synapomorphies, most of them skeletal features, which are the basis for recognizing 118 major groups of teleost fishes. The first six characters described are some of the synapomorphies for teleosts and involve aspects of the quadrate, maxilla, coronoid bones, articular, neural spine of preural centrum 1, and the pectoral propterygium. Although quite laborious, a thorough literature search would yield candidate genes that have phenotypic effects on the quadrate and maxilla in zebrafish, as would a search for these bones in the ZFIN database (http://zfin.org). No genes, however, that have a phenotypic effect on the coronoid bones, articular, neural spine of preural centrum 1, or the pectoral propterygium, can be retrieved currently from these sources.

The second problem, i.e., the difficulty of recovering the distribution of a particular feature for a set of taxa, can be generally illustrated by most aspects of the fish skeleton. Of interest to a developmental geneticist might be to identify the variation across fishes of a skeletal phenotype for a particular gene, such as the *brpf1* gene that results in the loss of a basihyal cartilage when mutant in zebrafish (Laue et al., 2008). Phenotypic

data on the skeletal variation of fishes, however, are recorded in an enormous bulk of free-text based literature such as books, monographs, unpublished theses, dissertations, phylogenetic studies, species descriptions, and taxonomic treatments. Although this literature is being digitized, it is still not organized or accessible for finding, aggregating, or comparing phenotypic data across studies. Even fish skeletal data that are available in databases, such as FishBase (Froese and Pauly, 2011), are not easily accessible for effective browsing, comparison, or analysis. From an Internet search of 'basihyal lost absent fish', one might patch together that the basihyal was lost possibly twice in batoid fishes (Miyake and McEachran, 1991), is absent in Pterygotriglini (Richards and Jones, 2002), etc. But it is not immediately obvious from the 5000+ search results that all Siluriformes (catfishes) share the derived loss of the basihyal element (Arratia and Schultze, 1990; de Pinna, 1993). Moreover, it is not possible to collect a hierarchically ordered list of fishes with associated presence or absence of basihyal element, and it is not possible to visualize the distribution of this feature on a phylogenetic tree for fishes. For even the most expert fish anatomist, recovering data on the variation in a particular feature across fishes is extremely difficult, and for students, or for researchers from other disciplines, it is completely unreachable.

The Phenoscape project, ontologies, and phenotypes

The Phenoscape project was launched in 2007 to enable large-scale knowledge discovery in the field of evolutionary developmental biology. The aim was to solve the two fundamental problems (above) by semantically integrating data on phenotypic variation among species with the phenotypic effects of genetic variation in model organisms using shared ontologies. To date, the project has produced a proof-of-concept knowledgebase with more than 5 00 000 phenotype assertions from the ichthyological literature.

Information retrieval from free-text is difficult (Washington et al., 2009). Simple text matching will not recognize that the following four phenotypes / character states refer to the same bone: 'lacrymal bone, flat' (Mayden, 1989), 'lacrimal, small, flat' (Grande and Poyato-Ariza, 1999), 'first infraorbital (lachrymal) shape, flattened' (Kailola, 2004), 'suborbital bone is very broad' (Cuvier, 1840). However, if these text strings are annotated with the ontology identifier TAO:0000223, which corresponds to infraorbital 1 and its synonyms (lacrymal, lacrimal, first infraorbital, suborbital), and uniquely references this concept as the first or anterior-most dermal bone that is located adjacent to the orbit in fishes, these differently described phenotypes can be aggregated. Moreover, if both infraorbital 1 and infraorbital 4 are related in an ontology as types of infraorbitals, parts of the infraorbital series, and types of dermal bones, they can be returned in queries for the term 'infraorbital', 'infraorbital series', or 'dermal bone'. Thus, free text terms that are synonymized and related in an ontology can be computationally aggregated and computed in ways that are not possible with free text alone.

An ontology can thus function to relate concepts (terms) in user-defined ways. It is a hierarchical set of well-defined terms and the logical relationships that hold between them. It represents the knowledge of a discipline, in a form that can be understood by both humans and machines. Ontologies are used for standardizing terminology within disciplines and for clarifying and improving communication across domains. Model organism communities have led successful efforts to standardize gene function descriptions (Gene Ontology, Blake and Harris, 2008) and to standardize the names of anatomical structures in model organism specific ontologies, e.g., the zebrafish anatomy ontology (Sprague et al., 2001). More recently, multispecies anatomy ontologies have been developed by the evolutionary community, including one for fishes, the Teleost Anatomy Ontology (Dahdul et al., 2010b) and one for hymenopterans (Yoder et al., 2010). As recently demonstrated by the mouse, fly, and zebrafish databases (Washington et al., 2009), ontologies support interoperability of descriptive data across databases, because, in contrast to natural language, ontologies allow computer processing of the semantic information buried in textual descriptions at a large scale.

The model organism community pioneered an ontology-based approach to represent anatomical phenotypes for the purpose of integrating mutant phenotypes across model organisms (Washington et al., 2009; Mungall et al., 2007). This Entity–Quality (EQ) syntax decomposes phenotype statements into three basic components: a phenotypic quality (Q), such as a 'flattened' shape; the entity that is its bearer (E), such as 'infraorbital bone'; and the organismal entity that exhibits the phenotype, in the case of model organisms, the genotype. Phenotypes in EQ format consist of terms from ontologies for each component, and well-defined relationships (is a, part of, and develops from) that render them formal logic expressions (Burger et al., 2008).

The Phenoscape project adopted this ontology-based approach to represent evolutionary phenotypes described as characters and character states in the systematics literature (Mabee et al., 2007b; Balhoff et al., 2010; Dahdul et al., 2010a,b). Using anatomical and taxonomic terms from teleost-specific ontologies (Teleost Anatomy Ontology and Teleost Taxonomy Ontology) in combination with terms from a taxon-neutral quality ontology (Phenotype and Trait Ontology), we used the EQ formalism to curate all characters and states described in over 50 phylogenetic studies (published between 1981 and 2008) of teleost fishes, primarily ostariophysans but also including clupeomorphs and some euteleosts (percomorphs and salmoniforms). These studies included peer-reviewed publications, book chapters, dissertations, and a M.S. thesis (Dahdul et al., 2010a). We developed the Phenex annotation software (Balhoff et al., 2010) to support a workflow for ontology-based annotation of data from these publications. Specifically, Phenex imports character matrices, loads user-selected anatomy and quality ontologies for representation of phenotypes, and facilitates selection of taxon identifiers from a taxonomic ontology, as well as museum collection IDs for recording specimens (Balhoff et al., 2010). As in other areas, the biocuration process (Howe et al., 2008) of annotating text, here phenotype descriptions, with ontology terms is primarily manual, although efforts to partially automate the work flow are underway (e.g., Dowell et al., 2009). A total of 4 820 phenotypic characters for 2 506 fish taxa, primarily species, have been curated into EQ formalism to date, resulting in 560 485 skeletal and other anatomical phenotype annotations. In the Phenoscape KB these taxon phenotype annotations are combined with 26 934 phenotype annotations for 4 307 zebrafish genes[1]. The Phenoscape KB supports browsing, searching, and analyzing gene and phenotype annotations together. It also allows users to take

---

[1]Numbers of phenotypes, taxa, and genes annotated change as new data are curated into the ZFIN and Phenoscape KB.

advantage of the relations between ontology terms, such as subtype and parthood relations. For example, a search for 'paired fin' will also retrieve data that are tagged with the logical subtypes 'pectoral fin' or 'pelvic fin'. Other logical relations, such as *develops from*, allow more sophisticated searching. A paired fin *develops from* a fin bud, and thus there is the potential to recover genetic and phenotypic data for fin buds from a search on paired fins. From the example above (Wiley and Johnson, 2010), candidate genes associated with the parent terms of coronoid bones, articular, neural spine of preural centrum 1, and pectoral propterygium can be proposed. The pectoral propterygium, for example, is *part of* the pectoral fin and as such is associated with all 153 genes with mutant pectoral fin (including parts) phenotypes in zebrafish.

5 00 000 hypotheses: candidate genes, candidate taxa

The wide variety of skeletal phenotypes among fishes are integrated across studies and with genes in the Phenoscape KB, enabling the search for candidate genes underlying evolutionary phenotypes and the query for distributions of phenotypes across taxa in gene expression and function. The shared ontologies integrate these disparate data and yield a rich set of testable hypotheses. Each of the 5 60 485 fish taxon phenotype annotations is associated with one or more genes in the KB, and thus thousands of evolutionary transitions in phenotype and gene associations may be hypothesized. The nine examples below illustrate common evolutionary changes in fish skeletal phenotypes for which candidate genes may be returned from queries (i.e., questions addressed via a software interface) to the Phenoscape KB. The numbers of phenotypes, taxa, and genes that are reported below will change as new data from the literature are curated into the ZFIN and Phenoscape KB (data below from 23 November 2011).

Gill rakers, absent. A query to the Phenoscape KB for taxa that lack gill rakers on one or more gill arches returns 101 taxa, including Anguilliforms, and some Characiformes, Siluriformes, and Tetraodontiformes. Gill rakers are absent from all gill arches of Anguilliforms (Nelson, 2006). In zebrafish, mutationsin two genes, *eda* and *edar*, result in the absence of gill rakers (Harris et al., 2008). One could then ask, for example, 'Are gill rakers absent in anguilliform eels because of changes in *eda, edar*, or regulation of the *eda* signaling pathway?'

Lateral line, variation. A Phenoscape KB query for taxa that vary in some quality of their lateral line, i.e., a change in position, shape, completeness, etc., yields 815 taxa in 18 teleost orders, including, e.g., *Minytrema melanops* (the spotted sucker). Alteration in the function of seven genes in zebrafish, *erbb3b, eya1* (Whitfield et al., 1996), *lef1* (McGraw et al., 2011), *pcsk5a* (Chitramuthu et al., 2010), *rog, sox10,* and *unm_m583* (Driever et al., 1996) results in abnormal lateral line phenotypes. The morpholino-based translation inhibition in *pcsk5a*, for example, disrupts formation of the lateral line, resulting in reduced or complete absence of posterior lateral line neuromasts (Chitramuthu et al., 2010). This query motivates the hypothesis that reduced length of the posterior lateral line in *Minytrema* (Smith, 1992) is due to an alteration in function of *pcsk5a* (or any of the other genes above).

Caudal fin, absent. A caudal fin is primitively present in teleost fishes; lost only a few times during evolution. A query in

Phenoscape KB for absence of a caudal fin returns 28 taxa, including 25 species of gymnotiform knifefishes (Albert, 2001) and three tetraodontiform species (Santini and Tyler, 2002). The loss of the caudal fin in the familiar *Mola mola* has been confirmed through a detailed developmental morphological study (Johnson and Britz, 2005). Five genes, *edar* (van Eeden et al., 1996), *lef1* (McGraw et al., 2011) *smc3, tll 1* (Lele et al., 2001) and *yap1* (Jiang et al., 2009), are associated with caudal fin loss in zebrafish. One might investigate, then, whether the loss of the caudal fin in *Mola mola* is related to changes in regulation of *yap1* (or any of the other genes above).

Ceratobranchial five teeth, absent. A query for taxa that vary in the presence of teeth on their fifth ceratobranchial element results in 43 taxa, including some Characiformes, Gymnotiformes, Tetraodontiformes, and *Gyrinocheilus*. Three genes, *acvr2a* (Albertson et al., 2005), *eda*, and *edar* (Harris et al., 2008), when disrupted, result in phenotypes that include absence of teeth on the ceratobranchial five element in zebrafish. One might ask, for example, whether the loss of ceratobranchial five teeth in the common aquarium fish, the algae-eater *Gyrinocheilus,* is related to changes in regulation of one of these genes.

Dorsal fin, absent. A query for taxa that lack a dorsal fin yields 58 taxa, including all Gymnotiformes (knifefishes) and some Siluridae (catfishes). Two genes, *hoxa13a* (Crow et al., 2009) and *tfap2a* (Li and Cornell, 2007), are linked to zebrafish phenotypes of median fin fold absence. Because the formal *develops from* relation relates dorsal fin to median fin fold in the Teleost Anatomy Ontology, a search for genes associated with dorsal fin phenotypes can return genes associated with the median fin fold phenotypes as well. Thus genes influencing the development of the precursors of structures may be considered as candidates in a search for the basis of the evolutionary novelty of dorsal fin loss in, for example, a gymnotiform knifefish.

Preopercle, shape. The preopercle varies in shape in 546 taxa, including Amiiformes, Aspidorhynchiformes, Characiformes, Clupeiformes, Cypriniformes, Ellimmichthyiformes, Elopiformes, Esociformes, Gonorynchiformes, Gymnotiformes, Hiodontiformes, Lepisosteiformes, Osteoglossiformes, Salmoniformes, Semionotiformes, and Siluriformes. No phenotypes are recorded for preopercle shape in zebrafish. However, if the search is broadened to include shape phenotypes from other parts of the opercular series (branchiostegal rays, interopercle, opercle, preopercle, and subopercle) then nine genes with mutant phenotypes are revealed: *acvr2a, acvr2b* (Albertson et al., 2005), *edn1* (Walker et al., 2007), *furina* (Walker et al., 2006), *jag1b, notch2* (Zuniga et al., 2010), *mef2ca* (Kimmel et al., 2003), *plcb3* (Walker et al., 2007), and *unm_t3153*. On the other hand, if the search is broadened from preopercle to include shape changes in other components of the dorsal hyoid arch, ten genes, some of them different from above, are found. Any of these genes provide a starting point for investigating the possible bases for shape changes in this dermal bone.

In fact, much of the skeletal variation that exists among fish species involves changes in shape, and approximately a fifth of the species phenotypes in the KB reflect this. Mutations in 920 of the 4307 zebrafish genes in the KB produce a change in shape of some aspect of anatomy in 1435 (of 26 934 total) gene

phenotypes. Changes in shape, especially in integrated skeletal structures may be very indirect, reflected in the increase in number of potential links. This lack of direct causality makes hypothesis testing difficult for complex structures.

Eye, decreased in size. A query of the Phenoscape KB informs the user that in all Siluriformes and Gymnotiformes, generalized from 57 studied species, the eye is reduced in size relative to the surrounding infraorbital bones (Fink and Fink, 1981). Mutations in 574 zebrafish genes produce reduced eye size, including, for example, *pbx4* (French et al., 2007).

Scales, absent. Scales are frequently lost on the head and body of fishes, and Phenoscape KB lists 336 taxa from 27 teleostean orders, including e.g., Siluriformes (catfishes), in which all scales are absent except for bony tubes of the lateral line (Fink and Fink, 1981). Three zebrafish genes, *eda*, *edar* (Harris et al., 2008), and *unm_t31273*, are candidates for this phenotype.

Basihyal, absent. All Siluriformes (catfishes) and four other teleostean species with data in the Phenoscape KB have lost the basihyal element, i.e., the anterior-most median element of the gill arches, or 'tongue' of the fish. The disruption of eleven zebrafish genes, including *brpf1* (Laue et al., 2008), *disc1* (Wood et al., 2009), *disp1* (Schwend and Ahlgren, 2009), *fac* (Schilling et al., 1996), *foxd3* (Neuhauss et al., 1996), *hand2* (Miller et al., 2003), *kat6a* (Laue et al., 2008), *sox9a* (Yan et al., 2002), *unm_th9*, *unm_tn20c*, and *unm_ty5* result in an aplastic or absent basihyal phenotype.

Limitations and biases in candidate genes –taxon approach

These examples demonstrate the ease with which genetic phenotypes may be aligned with diverse taxon phenotypes using ontologies to yield testable hypotheses. They also expose possible limitations, such as a bias toward well-studied pleiotropic genes such as *eda* or *edar*, which are proposed as candidates in several of the above cases. Missing phenotype data for taxa, whether because the taxa have never been surveyed for particular features or because these data have not been entered into the Phenoscape KB, also limit this approach, as does the coarseness of phenotype annotations for both genes and taxa. Other sources of bias include the different focus of phenotypic study in zebrafish developmental biology (neural system) vs. fish comparative morphology (skeletal system) and the difference in developmental stage under study, i.e., embryos and larvae in zebrafish vs. adults in comparative studies.

Integration of taxon phenotypes across studies

Ontology-annotations of phenotypic data surmount the difficulty of recovering the distribution of any skeletal feature across a set of taxa. The data can easily be viewed, summarized, and synthesized across studies at a scale not previously possible. As an example, the distribution of skeletal data, stratified by skeletal region across the five orders of ostariophyan fishes and their sister taxon order, the Clupeiformes (Fig. 1) reveals the disproportionate level of data on paired fins in the Siluriformes, reflecting the rich variation in pectoral fin spine ornamentation in this clade. The distribution of species phenotypes across particular skeletal elements, body
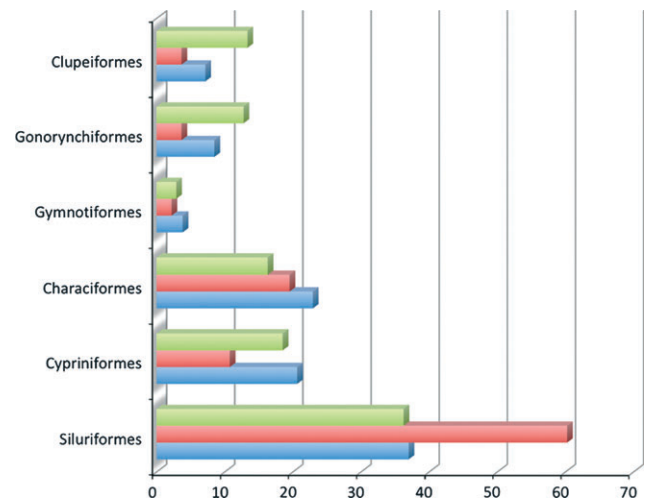

Fig. 1. Distribution of skeletal phenotypes in the Phenoscape Knowledgebase across the cranium (blue), paired fins (red), and postcranial axial skeleton (green) for Clupeiformes and the five ostariophysan orders

regions, etc. can be ascertained easily for any sample of publications in the KB.

Translational biodiversity

One of the most significant discoveries from the past decade is that genes, intron–exon structure, synteny, gene expression patterns, networks, and developmental processes, are highly conserved, well beyond expectation, across very distantly related organisms (Degnan, 2010). Examples of deep conservation or 'deep homology' (Shubin et al., 2009) have become common and familiar across greatly divergent species, e.g., the function of *Irx* and *hnf1* genes in lampreys are conserved across vertebrates for positioning the r4/r5 boundary (Jimenez-Guri and Pujades, 2011); the pigment gene *slc24a5* is functionally conserved between zebrafish and humans (Lamason et al., 2005); common genes control eye development (Gehring, 2004) and appendage development (Pueyo and Couso, 2005) from insects to humans (Carroll et al., 2005).

Biomedical researchers have leveraged this conservation to translate studies from model organisms to human medicine, so-called 'translational medicine' (Washington et al., 2009). For example, because important pathways such as insulin signaling have remained relatively unchanged during evolution, many human diseases, including cancer, are studied effectively in the nematode *C. elegans* (Markaki and Tavernarakis, 2010). Human craniofacial defects have been elucidated significantly by studies of zebrafish (Ghassibe-Sabbagh et al., 2011; Laue et al., 2011; Petrey et al., 2011), mouse (Tuveson and Hanahan, 2011), and *Xenopus* (Warkman and Krieg, 2007).

Further leveraging the conservation of developmental genes and networks to translate from model organisms to the breadth of species beyond humans, an approach that could be termed 'translational biodiversity,' is supported by the Phenoscape KB. Here, and demonstrated by the examples above, genetic and developmental data from zebrafish are leveraged to propose candidate genes for evolutionary changes in skeletal phenotypes across fish species, and the descriptive data from the field of comparative and evolutionary morphology are made accessible for searching and aggregating data across the breadth of species. These two disparate data-stores, one from molecular genetics and development and the other from

comparative morphology and evolution, can thus be rendered explorable and usable to researchers in other domains.

Although some differences in developmental mechanisms among similar phenotypes will certainly have arisen in the course of evolution (e.g., Tanaka et al., 2002), this computational approach to generating candidate genes is attractive, because studying the genetic and developmental bases of evolutionary phenotypes using the laboratory approaches from model organisms is simply impractical for the millions of extant species on earth. A fully developed 'translational biodiversity' will require databases of computable phenotypes for both model organisms and taxonomic groups that computers can understand and reason across. Such an ontology-based approach promises powerful data synthesis and discovery at a scale not otherwise possible. It also makes data accessible for broad groups of researchers and creates opportunities for new and synthetic research.

Acknowledgements

References

Albert, J. S., 2001: Species diversity and pylogenetic systematics of American knifefishes (gymnotiformes, teleostei). Miscellaneous Pulbications, Museum of Zoology, Universiyt of Michigan *no. 190*, 140 pages.

Albertson, R. C.; Payne-Ferreira, T. L.; Postlethwait, J.; Yelick, P. C., 2005: Zebrafish *acvr2a* and *acvr2b* exhibit distinct roles in craniofacial development. Dev. Dyn. **233**, 1405–1418.

Arratia, G.; Schultze, H. P., 1990: The urohyal development and homology within osteichthyans. J. Morphol. **203**, 247–282.

Balhoff, J. P.; Dahdul, W. M.; Kothari, C. R.; Lapp, H.; Lundberg, J. G.; Mabee, P.; Midford, P. E.; Westerfield, M.; Vision, T. J., 2010: Phenex: ontological annotation of phenotypic diversity. PLoS ONE **5**, e10500.

Blake, J. A.; Harris, M. A., 2008: The Gene Ontology (GO) project: structured vocabularies for molecular biology and their application to genome and expression analysis. Curr. Protoc. Bioinform. **23**, 7.2.1–7.2.9.

Burger, A.; Davidson, D.; Baldock, R., (Ed.). 2008: Anatomy ontologies for bioinformatics: Principles and practice, Springer.

Carroll, S. B.; Grenier, J. K.; Weatherbee, S. D., 2005: From DNA to diversity: molecular genetics and the evolution of animal design. Blackwell Publishing, Malden, MA, USA.

Chitramuthu, B. P.; Baranowski, D. C.; Cadieux, B.; Rousselet, E.; Seidah, N. G.; Bennett, H. P., 2010: Molecular cloning and embryonic expression of zebrafish PCSK5 co-orthologues: functional assessment during lateral line development. Dev. Dyn. **239**, 2933–2946.

Crow, K.; Amemiya, C. T.; Roth, J.; Wagner, G. P., 2009: Hypermutability of *HoxA13A* and functional divergence from its paralog are associated with the origin of a novel developmental feature in zebrafish and related taxa (Cypriniformes). Evolution **63**, 1574–1592.

Cuvier, G., 1840: Animal Kingdom: arranged according to its organization, forming the basis for a natural history of animals, and introduction to comparative anatomy, Amen Corner, 1840; digitized from copy at Library of Catalonia, 18 Jul 2008 by google ebooks.

Cuvier, L. B.; Valenciènnes, A., 1846: Histoire naturelle des poissons. Chez F. G. Levrault, Paris, France.

Dahdul, W. M.; Balhoff, J. P.; Engeman, J.; Grande, T.; Hilton, E. J.; Kothari, C.; Lapp, H.; Lundberg, J. G.; Midford, P. E.; Vision, T. J.; Westerfield, M.; Mabee, P. M., 2010a: Evolutionary characters, phenotypes and ontologies: curating data from the systematic biology literature. PLoS ONE **5**, e10708.

Dahdul, W. M.; Lundberg, J. G.; Midford, P. E.; Balhoff, J. P.; Lapp, H.; Vision, T. J.; Haendel, M. A.; Westerfield, M.; Mabee, P. M., 2010b: The Teleost Anatomy Ontology: Anatomical representation for the genomics age. Syst. Biol. **59**, 369–383.

Deans, A. R.; Yoder, M. J.; Balhoff, J. P., 2012: Time to change how we describe biodiversity. Trends Ecol. Evol. **27**, 78–84.

Degnan, B. M., 2010: The rise of genomics sheds light on the dawn of animals. Evol Dev. **12**, 425–427.

Dowell, K. G.; McAndrews-Hill, M.S.; Hill, D.P.; Drabkin, H. J.; Blake, J. A., 2009: Integrating text mining into the MGI biocuration workflow. Database **2009**, bap019.

Driever, W.; Solnica-Krezel, L.; Schier, A.F.; Neuhauss, S. C.; Malicki, J.; Stemple, D.L.; Stainier, D. Y.; Zwartkruis, F.; Abdelilah, S.; Rangini, Z.; Belak, J.; Boggs, C., 1996: A genetic screen for mutations affecting embryogenesis in zebrafish. Development **123**, 37–46.

van Eeden, F. J.; Granato, M.; Schach, U.; Brand, M.; Furutani-Seiki, M.; Haffter, P.; Hammerschmidt, M.; Heisenberg, C. P.; Jiang, Y. J.; Kane, D. A.; Kelsh, R. N.; Mullins, M. C.; Odenthal, J.; Warga, R. M.; Nusslein-Volhard, C., 1996: Genetic analysis of fin formation in the zebrafish, *Danio rerio*. Development **123**, 255–262.

Faccioli, P.; Stanca, A. M.; Morcia, C.; Terzi, V., 2009: From DNA sequence to plant phenotype: bioinformatics meets crop science. Curr. Bioinform. **4**, 173–176.

Fink, S. V.; Fink, W. L., 1981: Interrelationships of the Ostariophysan fishes (Teleostei). Zool. J. Linn. Soc. **72**, 297–353.

French, C. R.; Erickson, T.; Callander, D.; Berry, K. M.; Koss, R.; Hagey, D. W.; Stout, J.; Wuennenberg-Stapleton, K.; Ngai, J.; Moens, C. B.; Waskiewicz, A. J., 2007: *Pbx* homeodomain proteins pattern both the zebrafish retina and tectum. BMC Dev. Biol.. **7**, 85.

Froese, R.; Pauly, D., 2011, 1 / 20 / 10. "FishBase." http://www.fishbase.org/. 2011.

Gehring, W. J., 2004: Historical perspective on the development and evolution of eyes and photoreceptors. Int. J. Dev. Biol.. **48**, 707–717.

Ghassibe-Sabbagh, M.; Desmyter, L.; Langenberg, T.; Claes, F.; Boute, O.; Bayet, B.; Pellerin, P.; Hermans, K.; Backx, L.; Mansilla, M. A.; Imoehl, S.; Nowak, S.; Ludwig, K. U.; Baluardo, C.; Ferrian, M.; Mossey, P. A.; Noethen, M.; Dewerchin, M.; Francois, G.; Revencu, N.; Vanwijck, R.; Hecht, J.; Mangold, E.; Murray, J.; Rubini, M.; Vermeesch, J. R.; Poirel, H. A.; Carmeliet, P.; Vikkula, M., 2011: FAF1, a gene that is disrupted in cleft palate and has conserved function in zebrafish. Am. J. Hum. Genet.. **88**, 150–161.

Grande, T.; Poyato-Ariza, F. J., 1999: Phylogenetic relationships of fossil and Recent gonorynchiform fishes (Teleostei: Ostariophysi). Zool. J. Linn. Soc. **125**, 197–238.

Grunwald, D. J.; Eisen, J. S., 2002: Headwaters of the zebrafish – emergence of a new model vertebrate. Nat. Rev. Genet.. **3**, 717–724.

Harris, M. P.; Rohner, N.; Schwarz, H.; Perathoner, S.; Konstantinidis, P.; Nüsslein-Volhard, C., 2008: Zebrafish *eda* and *edar* mutants reveal conserved and ancestral roles of ectodysplasin signaling in vertebrates. PLoS Genet.. **4**, e1000206.

Howe, D.; Costanzo, M.; Fey, P.; Gojobori, T.; Hannick, L.; Hide, W.; Hill, D. P.; Kania, R.; Schaeffer, M.; St Pierre, S.; Twigger, S.; White, O.; Rhee, S. Y., 2008: Big data: The future of biocuration. Nature **455**, 47–50.

Jiang, Q.; Liu, D.; Gong, Y.; Wang, Y.; Sun, S.; Gui, Y.; Song, H., 2009: Yap is required for the development of brain, eyes, and neural crest in zebrafish. Biochem. Biophys. Res. Commun. **384**, 114–119.

Jimenez-Guri, E.; Pujades, C., 2011: An ancient mechanism of hindbrain patterning has been conserved in vertebrate evolution. Evol. Dev. **13**, 38–46.

Johnson, G. D.; Britz, R., 2005: Leis' conundrum: homology of the clavus of the ocean sunfishes. 2 Ontogeny of the median fins and axial skeleton of Ranzania laevis (Teleostei, Tetraodontiformes, Molidae). J. Morphol. **266**, 11–21.

Kailola, P. J., 2004: A phylogenetic exploration of the catfish family Ariidae (Otophysi; Siluriformes). The Beagle, Records of the Museums and Art Galleries of the Northern Territory **20**, 87–166.

Kimmel, C. B.; Ullmann, B.; Walker, M.; Miller, C. T.; Crump, J. G., 2003: Endothelin 1-mediated regulation of pharyngeal bone development in zebrafish. Development **130**, 1339–1351.

Lamason, R. L.; Mohideen, M.-A. P. K.; Mest, J. R.; Wong, A. C.; Norton, H. L.; Aros, M. C.; Jurynec, M. J.; Mao, X.; Humphreville, V. R.; Humbert, J. E.; Sinha, S.; Moore, J. L.; Jagadeeswaran, P.; Zhao, W.; Ning, G.; Makalowska, I.; McKeigue, P. M.; O'Donnell, D.; Kittles, R.; Parra, E. J.; Mangini, N. J.; Grunwald, D. J.; Shriver, M. D.; Canfield, V. A.; Cheng, K. C., 2005: SLC24A5, a Putative Cation Exchanger, Affects Pigmentation in Zebrafish and Humans. Science **310**, 1782–1786.

Laue, K.; Daujat, S.; Crump, J. G.; Plaster, N.; Roehl, H. H.; Kimmel, C. B.; Schneider, R.; Hammerschmidt, M., 2008: The multidomain protein *Brpf1* binds histones and is required for Hox gene expression and segmental identity. Development **135**, 1935–1946.

Laue, K.; Pogoda, H. M.; Daniel, P. B.; van Haeringen, A.; Alanay, Y.; von Ameln, S.; Rachwalski, M.; Morgan, T.; Gray, M. J.; Breuning, M. H.; Sawyer, G. M.; Sutherland-Smith, A. J.; Nikkels, P. G.; Kubisch, C.; Bloch, W.; Wollnik, B.; Hammerschmidt, M.; Robertson, S. P., 2011: Craniosynostosis and multiple skeletal anomalies in humans and zebrafish result from a defect in the localized degradation of retinoic acid. Am. J. Hum. Genet.. **89**, 595–606.

Lele, Z.; Bakkers, J.; Hammerschmidt, M., 2001: Morpholino phenocopies of the swirl, snailhouse, somitabun, minifin, silberblick, and pipetail mutations. Genesis **30**, 190–194.

Li, W.; Cornell, R. A., 2007: Redundant activities of *Tfap2a* and *Tfap2c* are required for neural crest induction and development of other non-neural ectoderm derivatives in zebrafish embryos. Dev. Biol. **304**, 338–354.

Mabee, P.; Ashburner, M.; Cronk, Q.; Gkoutos, G.; Haendel, M.; Segerdell, E.; Mungall, C.; Westerfield, M., 2007a: Phenotype ontologies: the bridge between genomics and evolution. Trends Ecol. Evol. **22**, 345–350.

Mabee, P. M.; Arratia, G.; Coburn, M.; Haendel, M.; Hilton, E. J.; Lundberg, J. G.; Mayden, R. L.; Rios, N.; Westerfield, M., 2007b: Connecting evolutionary morphology to genomics using ontologies: A case study from Cypriniformes including zebrafish. J Exp Zool B Mol Dev Evol. **308B**, 655–668.

Markaki, M.; Tavernarakis, N., 2010: Modeling human diseases in *Caenorhabditis elegans*. Biotechnol. J.. **5**, 1261–1276.

Mayden, R. L., 1989: Phylogenetic studies of North American minnows, with emphasis on the genus *Cyprinella* (Teleostei: Cypriniformes). Misc. Publ. Mus. Nat. Hist., Univ. Kansas **80**, 1–189.

McGraw, H. F.; Drerup, C. M.; Culbertson, M. D.; Linbo, T.; Raible, D. W.; Nechiporuk, A. V., 2011: *Lef1* is required for progenitor cell identity in the zebrafish lateral line primordium. Development **138**, 3921–3930.

Miller, C. T.; Yelon, D.; Stainier, D. Y.; Kimmel, C. B., 2003: Two endothelin 1 effectors, *hand2* and *bapx1*, pattern ventral pharyngeal cartilage and the jaw joint. Development **130**, 1353–1365.

Miyake, T.; McEachran, J. D., 1991: The morphology and evolution of the ventral gill arch skeleton in batoid fishes (Chondrichthyes: Batoidea). Zool. J. Linn. Soc. **102**, 75–100.

Mungall, C.; Gkoutos, G.; Washington, N.; Lewis, S., Year: Representing phenotypes in OWL. Proceedings of the OWLED 2007 Workshop on OWL: Experience and Directions, Innsbruck, Austria.

Nelson, J. S., 2006: Fishes of the World., John Wiley & Sons, Inc., Hoboken, NJ.

Neuhauss, S. C.; Solnica-Krezel, L.; Schier, A. F.; Zwartkruis, F.; Stemple, D. L.; Malicki, J.; Abdelilah, S.; Stainier, D. Y.; Driever, W., 1996: Mutations affecting craniofacial development in zebrafish. Development **123**, 357–367.

Pennisi, E., 2011: Will computers crash genomics? Science **331**, 666–668.

Petrey, A. C.; Flanagan-Steet, H.; Johnson, S.; Fan, X.; De la Rosa, M.; Haskins, M. E.; Nairn, A. V.; Moremen, K. W.; Steet, R., 2012: Excessive activity of cathepsin K is associated with cartilage defects in a zebrafish model of mucolipidosis II. Dis Model Mech. **5**, 177–190.

de Pinna, M. C. C., 1993: Higher-level Phylogeny of Siluriformes (Teleostei: Ostariophysi), with a New Classification of the Order. New York, City University of New York. PhD.

Pueyo, J. I.; Couso, J. P., 2005: Parallels between the proximal-distal development of vertebrate and arthropod appendages: homology without an ancestor? Curr. Opin. Genet. Dev. **15**, 439–446.

Richards, W. J.; Jones, D. L., 2002: Preliminary classification of the gurnards (Triglidae:Scorpaeniformes). Mar. Freshwater Res. **53**, 275–282.

Santini, F.; Tyler, J., 2002: Phylogeny of the ocean sunfishes (Molidae, Tetraodontiformes), a highly derived group of teleost fishes. Italian Journal of Zoology **69**, 37–43.

Schilling, T. F.; Piotrowski, T.; Grandel, H.; Brand, M.; Heisenberg, C.-P.; Jiang, Y.-J.; Beuchle, D.; Hammerschmidt, M.; Kane, D. A.; Mullins, M. C.; van Eeden, F. J. M.; Kelsh, R. N.; Furutani-Seiki, M.; Granato, M.; Haffter, P.; Odenthal, J.; Warga, R. M.; Trowe, T.; Nüsslein-Volhard, C., 1996: Jaw and branchial arch mutants in zebrafish I: Branchial arches. Development **123**, 329–344.

Schwend, T.; Ahlgren, S. C., 2009: Zebrafish *con*/disp1 reveals multiple spatiotemporal requirements for Hedgehog-signaling in craniofacial development. BMC Dev. Biol.. **9**, 59.

Shubin, N.; Tabin, C.; Carroll, S., 2009: Deep homology and the origins of evolutionary novelty. Nature **457**, 818–823.

Smith, G. R., 1992: Phylogeny and biogeography of the Catostomidae, freshwater fishes of North America and Asia. In: Systematics, HIstorical Ecology, and North American Freshwater Fishes. R. L. Mayden (Ed.). Stanford University Press, Stanford, CA, 778–826.

Sprague, J.; Doerry, E.; Douglas, S.; Westerfield, M., 2001: The Zebrafish Information Network (ZFIN): a resource for genetic, genomic and developmental research. Nucleic Acids Res. **29**, 87–90.

Tanaka, M.; Munsterberg, A.; Anderson, W. G.; Prescott, A. R.; Hazon, N.; Tickle, C., 2002: Fin development in a cartilaginous fish and the origin of vertebrate limbs. Nature **416**, 527–531.

Tuveson, D.; Hanahan, D., 2011: Translational medicine: Cancer lessons from mice to humans. Nature **471**, 316–317.

Walker, M. B.; Miller, C. T.; Coffin Talbot, J.; Stock, D. W.; Kimmel, C. B., 2006: Zebrafish furin mutants reveal intricacies in regulating Endothelin1 signaling in craniofacial patterning. Dev. Biol. **295**, 194–205.

Walker, M. B.; Miller, C. T.; Swartz, M. E.; Eberhart, J. K.; Kimmel, C. B., 2007: phospholipase C, beta 3 is required for Endothelin1 regulation of pharyngeal arch patterning in zebrafish. Dev. Biol. **304**, 194–207.

Warkman, A. S.; Krieg, P. A., 2007: Xenopus as a model system for vertebrate heart development. Semin. Cell Dev. Biol.. **18**, 46–53.

Washington, N. L.; Haendel, M. A.; Mungall, C. J.; Ashburner, M.; Westerfield, M.; Lewis, S. E., 2009: Linking human diseases to animal models using ontology-based phenotype annotation. PLoS Biol. **7**, 1–20.

Whitfield, T. T.; Granato, M.; van Eeden, F. J.; Schach, U.; Brand, M.; Furutani-Seiki, M.; Haffter, P.; Hammerschmidt, M.; Heisenberg, C. P.; Jiang, Y. J.; Kane, D. A.; Kelsh, R. N.; Mullins, M. C.; Odenthal, J.; Nusslein-Volhard, C., 1996: Mutations affecting development of the zebrafish inner ear and lateral line. Development **123**, 241–254.

Wiley, E. O.; Johnson, G. D., 2010: A teleost classification based on monophyletic groups. In: Origin and Phylogenetic Interrelationships of Teleosts. J. S. Nelson, H.-P. Schultze, M. V. H. Wilson (Eds.). Verlag Dr. Friedrich Pfeil, München, 123–182.

Williams, D. M.; Ebach, M. C., 2008: Foundations of Systematics and Biogeography., Springer.

Wood, J. D.; Bonath, F.; Kumar, S.; Ross, C. A.; Cunliffe, V. T., 2009: Disrupted-in-schizophrenia 1 and neuregulin 1 are required for the specification of oligodendrocytes and neurones in the zebrafish brain. Hum. Mol. Genet.. **18**, 391–404.

Yan, Y. L.; Miller, C. T.; Nissen, R. M.; Singer, A.; Liu, D.; Kirn, A.; Draper, B.; Willoughby, J.; Morcos, P. A.; Amsterdam, A.; Chung, B. C.; Westerfield, M.; Haffter, P.; Hopkins, N.; Kimmel, C.; Postlethwait, J. H.; Nissen, R., 2002: A zebrafish *sox9* gene required for cartilage morphogenesis. Development **129**, 5065–5079.

Yoder, M. J.; Miko, I.; Seltmann, K. C.; Bertone, M. A.; Deans, A. R., 2010: A gross anatomy ontology for Hymenoptera. PLoS ONE **5**, e15991.

Zuniga, E.; Stellabotte, F.; Crump, J. G., 2010: Jagged-Notch signaling ensures dorsal skeletal identity in the vertebrate face. Development **137**, 1843–1852.

Author's address: Paula Mabee, Department of Biology, 414 East Clark Street, University of South Dakota, Vermillion, South Dakota, United States of America. Email: pmabee@usd.edu