Connecting Evolutionary Morphology to Genomics Using Ontologies: A Case Study From Cypriniformes Including Zebrafish

PAULA M. MABEE[1*], GLORIA ARRATIA[2], MILES COBURN[3], MELISSA HAENDEL[4], ERIC J. HILTON[5], JOHN G. LUNDBERG[6], RICHARD L. MAYDEN[7], NELSON RIOS[8], AND MONTE WESTERFIELD[9]

[1]*Department of Biology, University of South Dakota, Vermillion, South Dakota*
[2]*Biodiversity Research Center, The University of Kansas, Lawrence, Kansas*
[3]*Biology Department, John Carroll University, University Hts., Ohio*
[4]*Zebrafish Information Network, University of Oregon, Eugene, Oregon*
[5]*Geology Department, Field Museum of Natural History, Chicago, Illinois*
[6]*Department of Ichthyology, Academy of Natural Sciences, Philadelphia, Pennsylvania*
[7]*Department of Biology, Saint Louis University, St. Louis, Missouri*
[8]*Tulane Museum of Natural History, Belle Chasse, Los Angeles*
[9]*Zebrafish Information Network (ZFIN) and Institute of Neuroscience, University of Oregon, Eugene, Oregon*

ABSTRACT One focus of developmental biology is to understand how genes regulate development, and therefore examining the phenotypic effects of gene mutation is a major emphasis in studies of zebrafish and other model organisms. Genetic change underlies alterations in evolutionary characters, or phenotype, and morphological phylogenies inferred by comparison of these characters. We will utilize both existing and new ontologies to connect the evolutionary anatomy and image database that is being developed in the Cypriniformes Tree of Life project to the Zebrafish Information Network (HYPERLINK "file://localhost/Library/Local%20Settings/Temp/zfin.org" zfin.org) database. Ontologies are controlled vocabularies that formally represent hierarchical relationships among defined biological concepts. If used to recode the free-form text descriptors of anatomical characters, evolutionary character data can become more easily computed, explored, and mined. A shared ontology for homologous modules of the phenotype must be referenced to connect the growing databases in each area in a way that evolutionary questions can be addressed. We present examples that demonstrate the broad utility of this approach. *J. Exp. Zool. (Mol. Dev. Evol.) 308B:655–668, 2007.* © 2007 Wiley-Liss, Inc.

One of the first steps in investigating the developmental basis for a particular evolutionary change in form involves the construction of a hypothesis of a relationship between anatomy (phenotype) and a set of candidate genes (genotype) within a phylogenetic context. The literature in phylogenetic systematics is replete with examples of evolutionary changes in structures at multiple levels of complexity, from the gain

or loss of a tooth or seta to the evolution of limbs or pentaradial symmetry. However, the genetic underpinnings of very few of these evolutionary changes are known. Systematic studies traditionally identify hypothesized morphological homologs and use these data in parsimony analyses to reconstruct relationships of taxa. However, evolutionary statements that arise from these involve only the synapomorphies and character-state transitions on trees. Only rarely does one know what genes are involved in the morphological anagenesis, i.e. succession of character change over evolutionary time; most of what we know about the association between development and genetic change comes from a few model organism species under current study. The scientific communities working with model organism species have undertaken bioinformatic efforts to connect genes to particular phenotypes using ontologies to understand genetic change associated with development (e.g., Gkoutos et al., 2004). An ontology describes a set of concepts or objects and the relationships among those objects, and it is used to reason about the objects. By applying a similar ontology-based approach to evolutionary morphology, one could simultaneously query (question) evolutionary, anatomical, and developmental genetic databases to address evolutionary and developmental (evo-devo) questions in a powerful and novel way (Mabee et al., 2007). The goals of this paper are to describe this approach and to illustrate use cases demonstrating the utility and desirability of such a system that would integrate morphological systematic investigations with genetics, molecular evolution, and development. We advocate this informatics approach for the rapid and efficient integration of comparative evolutionary morphological data with genetic and developmental data.

BACKGROUND

Zebrafish as a model organism

The zebrafish, *Danio rerio*, is a well-established model system for studies of vertebrate genetics and development, with a research community of several thousand investigators internationally. Its rapid development, transparent embryos, short generation time, and ease of genetic manipulation have made it ideal for laboratory study (Nüsslein-Volhard and Dahm, 2002). More significantly, mutagenesis approaches have yielded thousands of mutant lines with a variety of phenotypes (Grunwald and Eisen, 2002). Because it is one of the only representative actinopterygians (the group that includes over half of all living vertebrate species), among model organisms, it is also a key species for evo-devo studies (Parichy and Johnson, 2001; Jackman et al., 2004; Webb and Schilling, 2006).

The Zebrafish Information Network (ZFIN; Table 1) was established by Westerfield (1992) to integrate genetic, gene expression and phenotypic data for zebrafish and to link them to corresponding data in other model organisms and humans. As a model organism database, ZFIN curates data from the zebrafish research community; in other words, it collects, indexes, and enters the data into a publicly accessible database. Other model organisms have model organism databases as well, such as Mouse Genome Informatics for mouse and FlyBase for *Drosophila* (Table 1). Curation of wild type and mutant gene expression patterns and phenotypes are annotated by the model organism databases using a number of standardized ontologies. This provides maximal flexibility not only for curation, but also for complex queries within and between model organism databases.

TABLE 1. Commonly used acronyms, abbreviations, and URLs

CToL—Cypriniformes Tree of Life project, funded through NSF AToL (Assembling the Tree of Life) initiative	http://bio.slu.edu/mayden/ cypriniformes/portal.html
EQ—Entity-Quality syntax where a quality (from PATO) consists of an attribute and a value	http://www.bioontology.org/wiki/ index.php/PATO:About
FlyBase–Model organism database for *Drosophila*	http://flybase.bio.indiana.edu/
GO—Gene Ontology, developed by a consortium of model organism databases for annotating gene products	http://www.geneontology.org/
MGI—Model organism database for mouse (Mouse Genome Informatics)	http://www.informatics.jax.org/
OBO—Open Biomedical Ontologies	http://obo.sourceforge.net/ browse.html
OBO Relations Ontology	http://www.obofoundry.org/ro/
PATO—Phenotype and Trait Ontology, an ontology of qualities that are applicable to any organism	http://www.bioontology.org/wiki/ index.php/PATO: Main_Page
ZFIN—Zebrafish Information Network, the model organism database for zebrafish	http://www.zfin.org/

The use of ontologies in bioinformatics

Ontologies describe some domain of reality and facilitate understanding and interoperability among people and machines (Masolo et al., 2003), where interoperability implies the ability of two or more systems or components to exchange information and to use the information that has been exchanged. A particular ontology represents a standardized vocabulary of the types of entities (features) that exist for a particular system, and the relationships among them (Gruber, '93). The entities in an anatomical ontology for vertebrates, for example, might be systems such as digestive and skeletal and parts of these such as liver, stomach, femur, and skull. One can utilize mathematical logical reasoning to imply additional information because of the use of defined relationships. For example, if we identify a particular bone to be an instance (specific example) of a *basihyal* (the "tongue" of fishes), then we can formally conclude that it is also a part of the branchial arches and an instance of an *endochondral bone*. We can further reason that it *has_part basihyal anterior, basihyal posterior, basihyal teeth*, etc. Ontology terms must be defined and can have synonyms to support different communities' usages, which also aids in querying. The terms can have different relationships to one another, making the ontology more than a simple hierarchical list.

The most widely used ontology in biology is the Gene Ontology (GO; Table 1), developed and maintained by a consortium of model organism databases for annotating gene products. The GO provides a controlled vocabulary to describe the attributes of genes and gene products, for example, cytochrome c is not in the ontology, but attributes of cytochrome c, such as oxidoreductase activity, are. Use of the GO means, for example, that when FlyBase and the Saccharomyces Genome Database describe gene products as having the function "protein tyrosine phosphatase activity; GO:0004725", they mean exactly the same thing. Moreover, a search of these databases for gene products with the more general term "protein phosphatase activity" returns, among other things, all gene products with "protein tyrosine phosphatase activity". The GO is now used by all of the major model organism databases, including ZFIN, and by many other large databases in the genomics domain, including UniProt (Apweiler et al., 2004) and the Protein Data Bank protein structure database (Wolstencroft et al., 2005).

Anatomical ontologies (Bard, 2005) have also been developed by a number of model organism communities (mouse, zebrafish, and *Drosophila*). These include standardized vocabularies of entities (organs, tissues, cell types, and developmental stages) that are related hierarchically, and as such are natural and powerful extensions of purely definitional anatomical vocabularies, for example, Baumel et al. ('93). Relationships that may be specified among entities include *is_a, part_of*, or *develops_from* (Smith et al., 2005). For example, the femur *is_a* bone that *develops_from* mesoderm. These relationships themselves are formally defined in the Open Biomedical Ontologies Relations Ontology (Table 1). Anatomical ontologies, consisting of entities and their relationships, are used to annotate gene expression data and/or phenotypic data within the context of databases. There are currently about 15 anatomical ontologies, many of which are linked to model organism databases (Bard, 2005), including FlyBase, the Edinburgh Mouse Atlas Project, ZFIN, and the Plant Ontology Consortium; they may be accessed at Open Biomedical Ontologies (Table 1). However, anatomical ontologies have been developed for the most part without significant input from comparative evolutionary morphologists or systematists.

Zebrafish as Cypriniformes

Comparative evolutionary questions involving zebrafish require a hypothesis of broader phylogenetic relationships of the species and its close relatives. The Order Cypriniformes is the most diverse order of freshwater fishes, with over 3,000
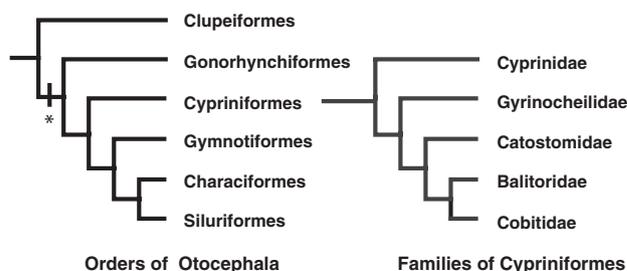

Fig. 1. Phylogenetic relationships among the orders of the Otocephala (left) and Cypriniformes (right) based on nuclear gene sequences of Chen and Orti (unpublished) and whole mitogenomes of Miya et al. (2003), Saitoh et al. (2003), Ishiguro (2005), Lavoué (2005) and Saitoh et al. (2006). The Ostariophysi, a clade within the Otocephala (left panel), is marked with an asterisk.

described species occurring on all landmasses except Central and South America, Antarctica, Australia, New Zealand, and Tasmania. Aside from the importance of the zebrafish for evo-devo studies, many of the fishes in this order represent critically important species in ecosystems and in aquaculture as a protein source for humans. With considerably less than 1% of the Earth's water being continental surface water, the combined diversity and importance of these fishes make them particularly important in both science and human culture.

Cypriniformes are members of the Otophysi (Fig. 1), which also includes the Gymnotiformes (American knifefishes), Characiformes (e.g., tetras), and Siluriformes (catfishes) (Fig. 1). These taxa include approximately 75% of all freshwater fishes (and 27% of *all* fishes) (Berra, 2001). Otophysans are characterized by a novel complex of vertebral characters called the Weberian apparatus (Weber, 1820; Bird and Mabee, 2003). Rudiments of the Weberian apparatus are found in their sister taxon, the Gonorynchiformes, and together they form the more inclusive clade, the Ostariophysi (Fig. 1).

Unfortunately, the species diversity within the Cypriniformes, as well as the evolutionary relationships of species and monophyly of the supraspecific groupings or taxa within the order are poorly known. Currently, five families are recognized (Cyprinidae, Catostomidae, Balitoridae, Cobitidae, Gyrinocheilidae) (Fig. 1). Most of the taxonomic diversity is found in the family Cyprinidae, and the subfamilial taxonomy of all families is in a state of flux (Cavender and Coburn, '92).

The zebrafish, *Danio rerio*, is a small freshwater cyprinid species native to the Himalayan region of Central Asia and is one of several species in the genus, a group that has been historically confused with species in the genera *Devario* and *Microrasbora* (Mayden et al., in press). In fact, the state of our knowledge of diversity and relationships surrounding this species is incredibly poor for such an important animal employed in so many different types of studies (Mayden et al., in press). When the zebrafish was identified as a model organism, the few published phylogenetic studies were unable to identify possible close relatives of either the species or the genus because of very limited taxon sampling and character data (Meyer et al., '93, '95; Sanger and McCune, 2002). Thus, although this species is a critical taxon for model-based studies, it is not possible at this time to conclude whether *D. rerio* represents a single
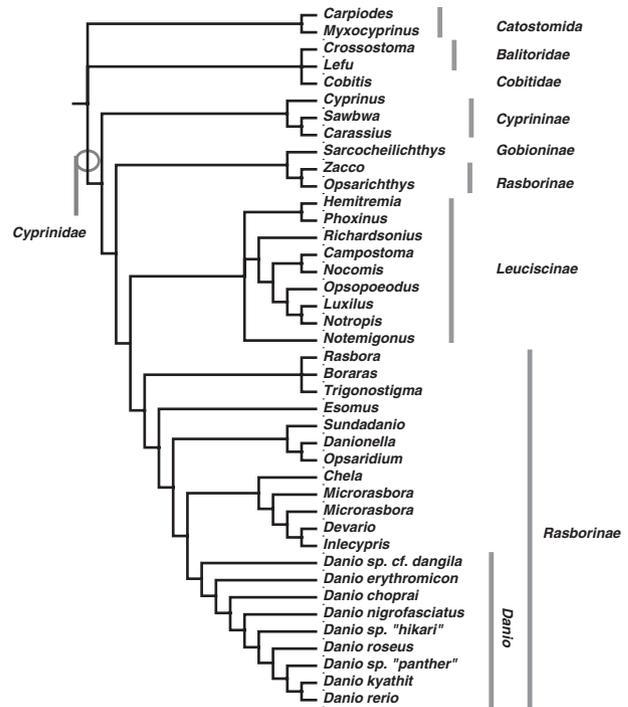

Fig. 2. Phylogenetic relationships among various families, subfamilies, genera, and species of Cyprinidae (Order Cypriniformes) based on nuclear and mitochondrial gene sequences from Mayden et al. (in press).

or multiple species in the wild, nor is it yet possible to determine its closest relative. Mayden et al. (in press) review the diversity and taxonomic history of the species and provide a larger-scale analysis of its phylogenetic relationships (Fig. 2).

The evolutionary relationships among the 3,000 plus Cypriniformes species, including zebrafish, are currently the focus of the Cypriniformes Tree of Life (CToL) project (Table 1). CToL is a large-scale phylogenetic initiative funded by NSF since 2004 as part of the NSF Assembling the CToL initiative. With seven US principle investigators, additional core participants, and 40 plus collaborators globally, extensive morphological (external and internal anatomy) and molecular (minimally three mitochondrial and four nuclear genes) data are being collected for 1,000 species. Owing to its joint molecular and morphological approach, CToL is uniquely positioned to integrate its comparative data with that from the zebrafish research community.

Comparing zebrafish mutants and evolutionary phenotypes

A mutant phenotype is the result of an organism's genetic background in a particular environ-

Fig. 3. *No tail* mutant zebrafish. Image from Halpern et al. ('93) (permission granted from Elsevier).

ment and may represent a constellation of differences in morphology, behavior, physiology, and other features relative to the wild type. The well-known zebrafish *no tail* mutant (Fig. 3), for example, lacks not only a tail, but also has U-shaped somites and an undifferentiated notochord. To date, more than 4,000 zebrafish mutant lines have been described with phenotypes in virtually every anatomical system, including the skeletal, cardiovascular, digestive, endocrine, and muscular systems. Most mutants have been identified at an early stage in development, typically up through day 5 (Haffter et al., '96; Stainier et al., '96), although with recent interest in zebrafish as a model for studies as diverse as learning, drug interactions, and visual function, increasing numbers of mutations with late effects are being described (see, e.g., Fisher et al., 2003; DeBruyne et al., 2004; Guo, 2004).

Evolution also produces new phenotypes among individuals in a species. Mutation may initially produce new phenotypes, but the processes of gene flow, selection, and drift maintain this variation, which may result in speciation. Aspects of the phenotype are compared across species, and variant features that are judged homologous are termed ''characters'' and ''character states'' by phylogenetic systematists. Systematic characters may be discovered in any aspect of the organism's phenotype (e.g., molecular, morphological, and behavioral). To discover which characters contain historical information, i.e. are

synapomorphic, at a particular phylogenetic level the condition of each character is assessed in outgroups. To estimate phylogeny, characters are coded in a taxon-by-character matrix format and analyzed with an increasing variety of methods (Page and Holmes, '98; Felsenstein, 2004; Delsuc et al., 2005).

Morphological characters are the raw data of many phylogenetic studies on cypriniform fishes (e.g., Cavender and Coburn, '92) and fishes in general (e.g., Fink and Fink, '81). Individual characters are described using free-form text and illustrations, and they are published in journals and museum monographs. Some of these text and image data are available in web-accessible databanks such as MorphoBank (http://morphobank. informatics.sunysb.edu/) and MorphBank (http://www.morphbank.com/) (Fontal-Cazalla et al., 2002; Hill, 2005), but these data repositories are rare and not yet broadly used by the community. Moreover the images are not annotated with terms from ontologies, nor has a vocabulary for describing anatomy been standardized in this context.

The mutant phenotypes of model organisms are analogous to the phenotypes of naturally occurring species in systematics. That is, individual phenotypic features that differ from wild type in zebrafish mutants are the analogues of individual characters that differ among species or higher level taxa. In contrast to the features contributing to mutant phenotypes, however, the developmental genetic basis of characters that have evolved in species is unknown.

Mutations that affect the skeletal system of zebrafish are of particular interest to ichthyologists because morphologically based evolutionary studies, which include fossil taxa, are frequently focused on similarities and differences in bones and cartilages across species. A variety of mutant phenotypes in the cranium, fins, and axial skeleton have been described in zebrafish (Table 2), and similarly, the skeleton varies in size, shape, and number of elements across cypriniform species (Table 3).

The overlap (intersection in Fig. 4) between similar evolutionary and zebrafish mutant phenotypes can provide candidate genes for evolutionary changes within Cypriniformes and, potentially, more broadly across fishes. Gene conservation across the evolution of animals has been repeatedly demonstrated by comparative studies on, for example, eye, limb, head, and segment formation (Carroll et al., 2005). A recent example of conservation of gene sequence and function

TABLE 2. Selected zebrafish skeletal phenotypes

Affected anatomical part	Mutant/gene
Maxilla: size reduction	*sox9a* [hi1134 a]
Dentary: size reduction	*sox9a* [hi1134 a]
Retroarticular: loss	*edn1, bapx1* [i]
Opercle: size decrease; loss	*sox9a* [hi1134 a], *edn1* [b], *lockjaw* [de],
Opercle: size increase	*edn1* [b]
Ceratohyal: size decrease	*sucker* [j]
Branchiostegals: number decrease	*edn1* [b]
Branchiostegals: shape change	*she, stu, edn1*-MO [b]
Hypobranchials: loss	*val* [f]
Ceratobranchial 5: size reduction	*sox9a* [hi1134 a]
Arches 2—5: reduced or absence	*lockjaw* [b]
Arches 4—6: loss	*duckbill, flathead* [b]
Ethmoid: reduction - loss	*detour, chamelion, youtoo, iguana* [f]
Trabeculae: fused	*detour, chameleon, youtoo, iguana* [f]
Pectoral fin: loss	*Fgf24*-MO [h]
Median fins: loss lepidotrichia	*Finless* [g]
Scapulocoracoid: loss	*sox9a* [a]
Neural & hemal spines: alignment	*chordin* [c]

Deviations in mutant skeletons and associated gene expression patterns are interpreted relative to wild type zebrafish.
[a]Yan et al. (2005).
[b]Kimmel et al. (2003).
[c]Fisher and Halpern ('99).
[d]Knight et al. (2004).
[e]T. Schilling (personal communication).
[f]Kimmel et al. (2001).
[g]van Eeden et al. ('96).
[h]Draper et al. (2003).
[i]Miller et al. (2003).
[j]Piotrowski et al. ('96).

TABLE 3. Selected cypriniform skeletal phenotypes (Siebert, '87)

| Affected skeletal element: aspect |
|---|
| Branchiostegal rays: number |
| Basibranchial 2: T-shaped |
| Basibranchial 4: development |
| Basibranchial 4: shape keeled |
| Basihyal: shape |
| Hypobranchial 3: development |
| Ceratobranchial 5: size |
| Ceratobranchial 5: teeth |
| Infrapharyngeals: number |
| Infrapharyngeal 1: presence |
| Epibranchial 1: uncinate process |
| Pharyngeal teeth: rows |
| Interhyal: size |
| Ceratohyal: shape |
| Anterior ceratohyal: shape |
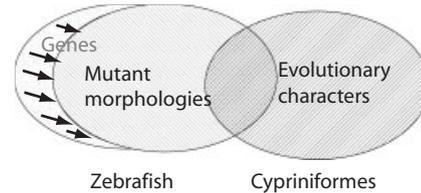
Fig. 4. The intersection of evolutionary and mutant phenotypes provides candidate genes for evolutionary studies.

between humans and fishes involves skin color (Lamason et al., 2005). Little was known about the specific genes that contribute to the variations in human skin color, but by studying the zebrafish mutant "golden", whose pigment stripes are paler than those in wild type, Lamason et al. (2005) found that the altered pigmentation was caused by a mutation in the *slc24A5* gene. The gene is highly conserved in vertebrates, and expression of the human gene in *golden* mutant zebrafish restored wild-type pigmentation (Lamason et al., 2005). European human populations carry a slightly different version of the *slc24A5* gene than do African and East Asian populations, and a genetic polymorphism that changes one amino acid in the coding region of the gene correlates with human skin pigmentation levels (Lamason et al., 2005).

The Lamason et al. (2005) study shows that phenotypic and genetic variation in mutant zebrafish is meaningful to variation on an evolutionary scale. Thus, querying a zebrafish database of phenotypic variants holds significant promise for providing insights into the developmental and genetic basis of evolutionary variants. The joining of multi-species evolution databases to single-species model organism databases will result in the ability to address a variety of important questions. We describe four use cases below that illustrate the utility of connecting these databases.

Use Case 1: developmental candidate genes

Ceratobranchial 5

Extensive variation in the size and shape of bones characterizes vertebrates, and aspects of the skeletal system are commonly coded as character data in systematic analyses of fishes. The fifth ceratobranchial, the lower pharyngeal arch bone bearing the pharyngeal teeth, varies in size and shape among Cypriniformes and relatives. In other ostariophysan fishes, the Siluriformes, Characiformes, and Gonorynchiformes, the fifth

ceratobranchial is of "normal" size, but it appears to be independently enlarged in two groups of Cypriniformes, the catostomids and cyprinids (Siebert, '87).

Investigating the underlying developmental genetic basis of relative size increase in the fifth ceratobranchial in these two evolutionary lineages (catostomids vs. cyprinids) requires an initial hypothesis regarding the particular genes that might be involved. Significant genetic information about skeletal development in zebrafish is available from ZFIN. One can search for gene expression and mutant phenotypes in the skeleton using the zebrafish anatomical ontology. An investigator can ask, "Which genes are associated with size change of the fifth ceratobranchial (in mutant lines of zebrafish)?" The reply is that the zebrafish phenotype "ceratobranchial 5 reduced" is observed in the mutant line, $sox9a^{hi1134}$, and these data are attributed to a particular publication (Yan et al., 2005). The information that $sox9a$ has a role in size reduction in the fifth ceratobranchial provides a starting point for comparative evo-devo studies on size differences in this bone across taxa. The comparative gene expression data from this publication are retrieved from this search of ZFIN as well, and thus it is available for comparison with gene expression patterns from other species. Comparative gene expression data are critical for formulating hypotheses of evolutionary change. One might then hypothesize that the expression or sequence of $sox9a$ has been altered to result in the enlargement of this bone in these two cypriniform lineages, and developmental genetic work on species in these lineages could test this hypothesis.

Branchiostegal rays

Branchiostegal rays are long, flat dermal bones that develop in and support the ventral portion of the branchiostegal (gill) membrane. They vary in size, shape, and number across fishes (McAllister, '68). Zebrafish and all other cypriniforms have three branchiostegals. Gonorynchiforms have three or four, Characiformes have three, four, or five, and Siluriformes species may possess between 3 and 20 branchiostegals (McAllister, '68). Understanding the developmental and genetic basis of these evolutionary differences requires a starting point. We might query ZFIN: "What gene mutations modify the number of branchiostegal rays in mutant zebrafish"? Our query would result in the following information: there is a mutant line known as "sucker", described by Kimmel et al.

(2003), which has a reduced number of branchiostegal rays from 3 to 1 and concomitant changes in gene expression. The sucker mutation is a change in the *endothelin-1* gene. An evolutionary biologist might then hypothesize that changes in *endothelin-1* or its expression is the basis for changes in number of branchiostegals and their homologues (in the opercular series) across fish species (Kimmel et al., in press).

Support of homology statements

In some instances, the identification of particular anatomical features across taxa is obscured by the wide phylogenetic distance between taxa. For example, the large bony element of the gill cover in sturgeons and paddlefishes (Acipenseriformes) has variously been called the "subopercle" and the "opercle" of other actinopterygian fishes. Identification of this bone as the opercle has followed from its position and size (i.e., the opercle is typically the largest bone of the gill cover series). Conversely, its interpretation in acipenseriforms as representing the subopercle of other taxa follows from its position relative to a small bone interpreted to be the opercle in fossil members of the group (Bemis et al., '97; Grande et al., 2002). By querying and comparing the genetic underpinnings of opercle bone development in mutant phenotypes, it may be possible to identify genes involved in the development of (or absence of) these bones in zebrafish, such as *lockjaw* (Table 2), and ultimately to compare expression patterns of these same genes in acipenseriform fishes, thereby adding evidence in support of homology statements.

Use case 2: candidate taxa

A developmental biologist who observes branchiostegal ray number reduction in a zebrafish mutant might want to know which species show a reduction in branchiostegal ray number and what evolutionary pattern could account for this change. Through such an investigation they might gain further insight into the functions of the genes involved. By querying the evolutionary database of skeletal characters by species, they would find that reduction in number has occurred multiple times (Fig. 5), but that three lineages, the solenostomids and syngnathids (ghost pipefishes and pipefishes), giganturids (telescopefishes), and saccopharyngoid (gulper and swallower) eels have the fewest branchiostegal rays (McAllister, '68). Members of these three unrelated lineages exhibit a number of

interesting parallels including the loss of the swim bladder, pelvic fins, and scales, as well as elongation of the mandibular or hyoid arches and reduction or loss of the opercle in syngnathids and saccopharyngoids. In addition, a variety of other bones and soft tissues are lost or greatly modified in these fishes, suggestive of alterations in the expression of broadly acting genes. If such comparative morphological data were available and searchable in phylogenetic and tabular formats, then an investigator could make the best choice of taxa in which to investigate development further. If through such an investigation of these fishes, a particular gene is implicated in an evolutionary change, this information becomes useful to the developmental biologist who seeks to understand the function and evolution of that gene better.

Use case 3: correlated characters

Determining whether morphological characters are correlated because of functional, genetic, or developmental constraints, or whether they are independent from one another has been a long-standing issue in phylogenetic systematics. Independence of characters is a fundamental necessity—and an assumption of—phylogenetic analysis (Sneath and Sokal, '73; Wiley, '81; Farris, '83); see also more recent discussions of character

concepts and identification (e.g., Wagner, 2001; Rieppel and Kearney, 2002). Often the correlation between two or more characters is difficult to ascertain, particularly when the characters are linked together at the molecular level. However, computerized access to genetic and developmental data such as in ZFIN, supports queries that can allow assessment of such genetically and developmentally correlated characters. For example, an investigator might observe that four characters, the reduction in size of the dentary, maxilla, ceratohyal, and opercle, supports the monophyly of a particular clade. The question is whether the size reduction in each of these bones represents independent support for this phylogenetic hypothesis, or whether they are correlated because of a common genetic or developmental basis. One might use these features and their qualities (e.g., size, shape) as queries of ZFIN to ask whether there is information from zebrafish mutant phenotypes that addresses this question. For instance, one would find that in $sox9a^{hi1134}$ mutants the dentary, maxilla, and opercle bones are reduced in size relative to wild-type zebrafish, whereas other bones are relatively unaffected (Yan et al., 2005). This suggests that size of the dentary, maxilla, and opercle, may be co-regulated in part by $sox9a$ (Yan et al., 2005). An investigator might then conclude that support for the monophyly of this clade is not as strong as it appears and that perhaps only two characters (the ceratohyal and the correlated character, the dentary + maxilla + opercle) support the hypothesis. These data could be reanalyzed using this information (recoding the data for two vs. four characters) and other aspects of the possible correlation of these characters (e.g., ecological, functional, etc.) could be considered by the investigator.

Description of mutant zebrafish phenotypes using Entity + Quality syntax

Because gene function is indexed by the model organism databases using the GO (Table 1), it is now possible to find mutant phenotypes that affect the same biological process, function, or subcellular location in different model organisms. However, the GO is used to describe gene function; it is not designed to capture the richness of phenotype description. To address this need, the phenotype and trait ontology (PATO; Table 1) has been developed for cross-species use in combination with other ontologies.
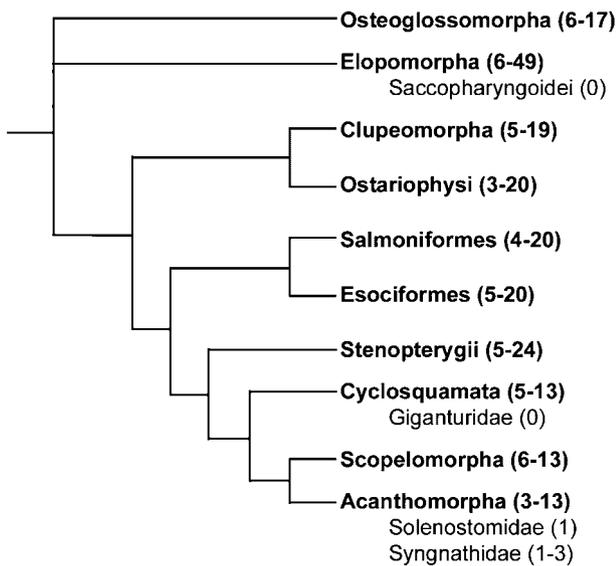

Fig. 5. Phylogeny of teleost fishes, showing the pattern of evolution in number of branchiostegal rays, modified from McAllister ('68), Nelson (2006) and the Teleostei Tree of Life page (http://tolweb.org/Teleostei/15054).

A zebrafish phenotype can be described using an Entity + Quality (EQ) syntax. Entity and Quality terms reference two different kinds of ontologies. Entities may be obtained from the zebrafish anatomical ontology or cross-species ontologies such as GO or the cell type ontology. Qualities come from PATO. An example of a phenotype annotated with EQs is shown in Fig. 6.

Description of evolutionary characters using EQ syntax

The combinatorial logic and structure of the EQ syntax can be expanded to encompass evolutionary morphological characters. Traditional morphological characters in systematics are framed using free-form text. For example, a character might be phrased "Number of basibranchial elements present" with states "two" "three", and "four" (Sawada, '81). This means that some species have two basibranchials, some have three, and others have four basibranchials. This simple character can easily be rephrased in the EQ syntax. In this case, the entity "basibranchial" would be drawn from an anatomical ontology and the quality "number" with values 2, 3, or 4, drawn from the PATO ontology. The rephrased character in EQ syntax would thus be "Basibranchial, number" and a particular species character state would be the values 2, 3, or 4.

Other morphological characters, particularly those dealing with qualitative aspects, are more problematic to describe with the EQ syntax. Variation in size of particular bones is common among fishes. For example, the opercular bone
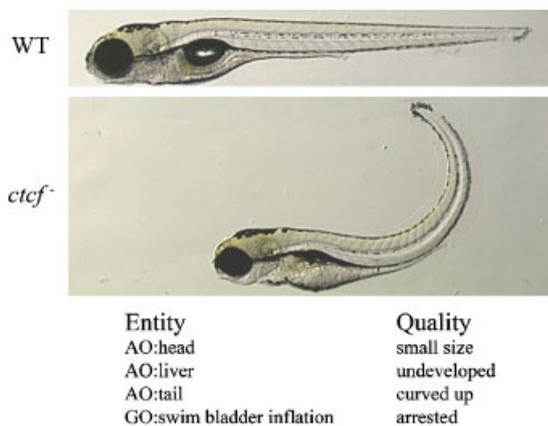

Fig. 7. Normal neural arch 2 in (**A**) *Chanos chanos* (11.4 mm SL, Standard Length) (arrow) and (**B**) modified neural arch 2, termed the intercalarium in *Catostomus commersoni* (15.8 mm SL) (arrow).

Fig. 6. Entity-Quality coding of a mutant phenotype in zebrafish. Wild type (WT) is shown in upper panel; *ctcf⁻* mutant phenotype in lower panel. AO, zebrafish anatomical ontology; GO, Gene Ontology (Table 1).
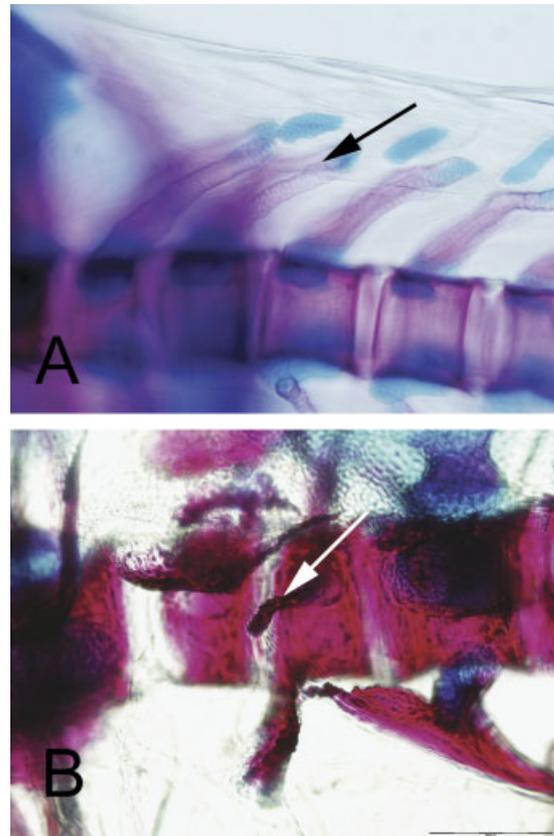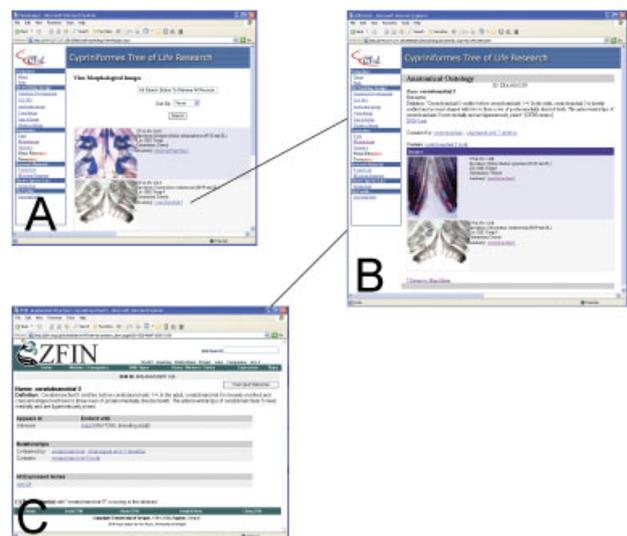

Fig. 8. Implementation of the zebrafish anatomical ontology in the Cypriniformes Tree of Life -Zebrafish Information Network working group database.

may be small in some species but large in others. The character would then be "opercular" with states "small" or "large". In EQ syntax, for one species, the entity would be opercular bone and the quality would be small [size]. For the other species, the entity would also be opercular bone, but the quality would be large. This example highlights the problem of a reference in relation to "small" or "large". For mutant phenotypes, the reference is the wild type, but when comparing among species, the reference is dependent on the species sampled in the group of interest, in our case, Cypriniformes. If taxa outside Cypriniformes are then sampled, and a smaller opercular bone is observed in some species, then a new value "very small" must be added or "small" within Cypriniformes must be redefined as "medium". Ideally the reference or standard is the most recent common ancestor of the taxa being compared, and that will change as the group under consideration changes.

We discovered that the EQ syntax must be expanded to accommodate the more complex characters of systematics. Frequently an additional entity (E_2) and quality are required. A complex character might thus read: $E_1Q_1E_2Q_2$. An example comes from the vertebral column of fishes, where the vertebral number at which the first hemal spine is located varies. Such a character might read "Vertebral position of first hemal spine" with states "vertebra 16", "vertebra 17", or "vertebra 18". The EQ description of this character, because it relates two features, would involve two anatomical entities. For species in which the first hemal spine is located on the 16th vertebra, the syntax would be [E_1: first hemal spine] [Q: located at] [E_2: vertebra 16]. The quality "located at" is relational and thus requires a second entity.

Using a consistent language or vocabulary and consistent format (syntax) to describe systematic characters has benefits beyond interoperability with model organism databases. The difficulty of comparing characters across studies owing to the lack of standardization (in addition to other factors) is well recognized by systematists. There are currently no agreed upon terminologies in most evolutionary disciplines, and ontologies must be developed by these communities. The codification of systematic characters using taxon-specific anatomy ontologies and taxon-independent qualifiers promises great flexibility because different combinations can be used to describe different characters.

Issues concerning homology

Homology (i.e., similarity due to common ancestry) is a central concept in comparative evolutionary biology. Homology assignments, or "characters", are hypotheses that are the data for phylogenetic inference. Standard lines of evidence are used to assess homology a priori (Mayden and Wiley, '92) and include similarity in shape and size, topographic position, complexity, and development (Remane, '52; Roth, '84; Patterson, '88). A posteriori, homology is tested by the distribution of characters on phylogenies resulting from character analysis, and it is reassessed when required (Mayden and Wiley, '92).

Within major clades of organisms, the definition of homology for most structures will not likely represent a problem; however, when comparing across multiple clades, homologous features may be referred to by different names. One possible solution to the complexity of homology in multispecies databases is to treat terms as synonymies.

There are several lines of evidence to suggest that the so-called frontal bone in actinopterygian fishes (including zebrafish) is the homologue of the parietal bone, and not the "frontal" bone, in advanced "piscine" sarcopterygians as well as in "tetrapod" sarcopterygians such as frog, bird, mouse, or human. The so-called parietal bone of actinopterygians is the homologue of the postparietal bone of piscine sarcopterygians and tetrapods (Jollie, '62; Schultze and Arsenault, '85). Actinopterygians and basal sarcopterygians do not have the bone that is defined as the frontal in tetrapods.

Although the term "parietal" could be added as the synonym of "frontal" and vice versa, the frontal and parietal bones within the same species are not synonyms. Moreover, if, for example, a user wanted to search for all of the genes expressed in all homologues of the frontal bone in vertebrates, they would not want to find the genes expressed in both frontal and parietal bones of sarcopterygians. Thus, some mechanism must be put in place to inform the computer of this special type of relationship, i.e. homology.

We propose to locate homology relationships in a relational table rather than in the anatomical ontology itself. The ontology will contain all possible anatomical terms, but it will not imply homology between terms—it will be homology-neutral. Thus the frontal and parietal bones will both be listed in the anatomical ontology, but their homology relationship will be encoded separately. Evidence codes for homology (e.g., position, devel-

opment, composition, etc.), analogous to the evidence codes used for gene orthology by the model organism databases, will be required, and they will be referenced to the literature.

Building a taxonomic ontology

Additionally, we propose to create a taxonomic ontology of zebrafish and close relatives from the classification in the Catalog of Fishes database for ichthyology (Eschmeyer, '98) and subsequent web-based updates at http://www.calacademy.org/RESEARCH/ichthyology/catalog/fishcatsearch.html. This taxonomic ontology will be in the form of *is_a* relationships (i.e., *Cyprinidae is_a Cypriniformes*). We will associate particular entities from the multi-species anatomical ontology with particular taxonomic groups in which they are present. Thus, terms will be available for only those taxa in which the corresponding anatomical structures are present. For example, fishes outside of the Otophysi have normal neural arches on the second vertebra. In contrast, zebrafish and other members of the Ostariophysi (Fig. 1) have the second vertebral neural arch modified as an "intercalarium" (Fig. 7), part of a complex of features (the Weberian apparatus) that are modified for sound transmission and hearing in these fishes. The intercalarium would be specified as the homologue of the second neural arch in the relational homology table. It would be annotated to the taxonomic group "Otophysi" in the taxonomic ontology, and second neural arch would be annotated to the level of "Vertebrata". Therefore, if a systematist wanted to describe a character involving the second neural arch from an otophysan species such as *D. rerio*, only the intercalarium would be available as a choice from the multi-species anatomical ontology. In queries relating to the second neural arch, however, the intercalarium would be searched.

Preliminary implementation of a multi-species ontology

Multi-species anatomical ontologies must be developed to formulate morphological characters using the EQ syntax if evolutionary questions are to be addressed (Haendel et al., 2007). We are prototyping a multi-species anatomical ontology for ostariophysans fishes, maintaining the zebrafish anatomical ontology as the "core" ontology, because it contains many basal features for fishes (Mabee et al., 2007). The internal CToL user interface currently supports the association of up to three anatomical terms from this ontology with images that are uploaded by investigators. These terms may later be used to describe characters states using the EQ syntax. The interface facilitates browsing the morphological data either by terms in the anatomical ontology or by the images submitted (Fig. 8A). Each term within the anatomical ontology (Fig. 8B) is linked back to ZFIN, allowing users direct access to ZFIN gene expression, mutant phenotype, genetics, genomics, and literature citation data (Fig. 8C).

CONCLUSIONS

Collaborative systematics requires the development of databases for both images and text-based character data. Multiple small-scale databases have been developed by many collaborative groups, including CToL, but little attention has been paid to joining these databases to genetic databases. By referencing the morphological characters of systematics, currently embedded as free-form text in character matrices and literature, to ontologies, a variety of queries (i.e., questions using computers) that integrate genetics with evolution can be made (Mabee et al., 2007). Such infrastructure will also support computerized data mining. Questions regarding the relationships between genes and morphology form the basis of the field of evo-devo (Carroll et al., 2001). Queries such as "What are the developmental genetic mechanisms responsible for evolutionary changes in morphology in a particular clade"? can be answered only if data from molecular developmental studies (genetic, gene expression, and phenotype) are integrated and queried simultaneously with morphological evolutionary data. Microarrays and quantitative trait loci analyses are also used to identify factors that underlie species-specific differences in morphology. Bioinformatics will complement results of such analyses by facilitating or verifying the identification of potential gene candidates. For example, are there known mutations in genes that are revealed to be up or down regulated in a microarray experiment? Are there mutant phenotypes for genes associated with an evolutionarily salient quantitative trait loci? Specifically for our study of cypriniform evolution, the use of ontologies will allow interoperability with ZFIN and enable simultaneous queries of the ZFIN and CToL databases. We propose to create a connection between the model organism (zebrafish) and ichthyology/phylogeny communities that will expand to include

researchers in other model fish communities (*Fugu*, medaka) and ichthyologists who specialize in studies of other clades. The tools that we develop will be generalizable to other taxa and evo-devo questions.

To date, the molecular developmental biologists who have focused on model organisms have driven many of the advances in bioinformatics methods that support the large quantities of accumulated molecular and genetic data; the informatics effort by evolutionary biologists lags far behind. Only by scaling up informatics support for evolutionary morphology and by developing evolutionary databases annotated with ontologies, we can use data, databases, and the mining potential of bioinformatics to get at the molecular developmental underpinnings of evolutionary change. Ultimately these informatics tools and ontologies will be useful to other communities of scientists that need to do the same thing: communicate across disciplines that have been separated previously, to address fundamental questions in evolution and development.

ACKNOWLEDGMENTS

LITERATURE CITED

Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Natale DA, O'Donovan C, Redaschi N, Yeh LS. 2004. UniProt: the universal protein knowledgebase. Nucleic Acids Res 32(Database issue):D115–D119.

Bard JBL. 2005. Anatomics: the intersection of anatomy and bioinformatics. J Anat 206:1–16.

Baumel JJ, King AS, Breazile JE, Evans HE, Vanden Berge JC, editors. 1993. Handbook of avian anatomy: nomina anatomica avium, 2nd edition. Cambridge, MA: Harvard Univ; Publications of the Nuttall Ornithological Club. 779p.

Bemis WE, Findeis EK, Grande L. 1997. An overview of acipenseriformes. Environ Biol Fishes 48:25–71.

Berra TM. 2001. Freshwater fish distribution. San Diego: Academic Press. 604p.

Bird NC, Mabee PM. 2003. The developmental morphology of the axial skeleton of the zebrafish, *Danio rerio* (Ostariophysi: Cyprinidae). Dev Dyn 228:337–357.

Carroll SB, Grenier JK, Weatherbee SD. 2001. From DNA to diversity: molecular genetics and the evolution of animal design. London: Blackwell Science. 214p.

Carroll SB, Grenier JK, Weatherbee SD. 2005. From DNA to diversity: molecular genetics and the evolution of animal design. Oxford: Blackwell Publishing. 258p.

Cavender TM, Coburn MM. 1992. Phylogenetic relationships of North American Cyprinidae. In: Mayden RL, editor. Systematics, historical ecology, and North American fresh-water fishes. Stanford: Stanford University Press. p 293–327.

DeBruyne J, Hurd MW, Gutierrez L, Kaneko M, Tan Y, Wells DE, Cahill GM. 2004. Isolation and phenogenetics of a novel circadian rhythm mutant in zebrafish. J Neurogenet 18: 403–428.

Delsuc F, Brinkmann H, Philippe H. 2005. Phylogenomics and the reconstruction of the tree of life. Nat Rev Genet 6: 361–375.

Draper BW, Stock DW, Kimmel CB. 2003. Zebrafish fgf24 functions with fgf8 to promote posterior mesodermal development. Development 130:4639–4654.

Eschmeyer WN. 1998. Catalog of fishes. San Francisco: Special Publication, California Academy of Sciences. 2905p.

Farris JS. 1983. The logical basis of phylogenetic analysis. In: Platnick NI, Funk VA, editors. Advances in Cladistics. New York: Columbia University Press. p 7–36.

Felsenstein J. 2004. Inferring phylogenies. Sunderland, MA: Sinauer Associates. 664p.

Fink SV, Fink WL. 1981. Interrelationships of the ostario-physan fishes (Teleostei). Zool J Linnean Soc 72:297–353.

Fisher S, Halpern ME. 1999. Patterning the zebrafish axial skeleton requires early *chordin* function. Nat Genet 23:442–446.

Fisher S, Jagadeeswaran P, Halpern ME. 2003. Radiographic analysis of zebrafish skeletal defects. Dev Biol 264: 64–76.

Fontal-Cazalla FM, Buffington ML, Nordlander G, Liljeblad J, Ros-Farré P, Nieves-Aldrey JL, Pujade-Villar J, Ronquist F. 2002. Phylogeny of the Eucoilinae (Hymenoptera: Cynipoi-dea: Figitidae). Cladistics 18:154–199.

Gkoutos GV, Green EC, Mallon AM, Hancock JM, Davidson D. 2004. Using ontologies to describe mouse phenotypes. Genome Biol 6:R8.

Grande L, Jin F, Yabumoto Y, Bemis William E. 2002. *Protosephurus liui*, a well-preserved primitive paddlefish (Acipenseriformes: Polyodontidae) from the lower Cretaceous of China. J Vert Paleontol 22:209–237.

Gruber TR. 1993. A translation approach to portable ontologies. Knowl Acquisition 5:199–220.

Grunwald DJ, Eisen JS. 2002. Timeline: headwaters of the zebrafish—emergence of a new model vertebrate. Nat Rev Genet 3:717–724.

Guo S. 2004. Linking genes to brain, behavior and neurological diseases: what can we learn from zebrafish? Genes, Brain Behav 3:63–74.

Haendel MA, Neuhaus F, Osumi-Sutherland DS, Mabee PM, Mejino Jr JLV, Mungall CJ, Smith B. 2007. CARO—The Common Anatomy Reference Ontology. In: Burger A, Davidson D, Baldock R, editors. Anatomy Ontologies for Bioinformatics: principles and practice. In press.

Haffter P, Granato M, Brand M, Mullins MC, Hammersch-midt M, Kane DA, Odenthal J, van Eeden FJ, Jiang YJ, Heisenberg CP, Kelsh RN, Furutani-Seiki M, Vogelsang E, Beuchle D, Schach U, Fabian C, Nusslein-Volhard C. 1996.

The identification of genes with unique and essential functions in the development of the zebrafish, *Danio rerio*. Development 123:1–36.

Halpern ME, Ho RK, Walker C, Kimmel CB. 1993. Induction of muscle pioneers and floor plate is distinguished by the zebrafish no tail mutation. Cell 75:99–111.

Hill RV. 2005. Integration of morphological data sets for phylogenetic analysis of amniota: the importance of integumentary characters and increased taxonomic sampling. Syst Biol 54:530–547.

Ishiguro NB, Miya M, Inoue JG, Nishida M. 2005. *Sundasalanx* (Sundasalangidae) is a progenetic clupeiform, not a closely related group of salangids (Osmeriformes): Mitogenomic evidence. J Fish Biol 67:561–569.

Jackman WR, Draper BW, Stock DW. 2004. Fgf signaling is required for zebrafish tooth development. Dev Biol 274:139–157.

Jollie M. 1962. Chordate morphology. New York: Reinhold Books.

Kimmel CB, Ullmann B, Walker M, Miller CT, Crump JG. 2003. Endothelin 1-mediated regulation of pharyngeal bone development in zebrafish. Development 130:1339–1351.

Kimmel CB, Walker MB, Miller CT. In press. Morphing the hyomandibular skeleton in development and evolution. J Exp Biol Mol Dev Evol (this volume).

Knight RD, Javidan Y, Nelson S, Zhang T, Schilling T. 2004. Skeletal and pigment cell defects in the lockjaw mutant reveal multiple roles for zebrafish tfap2a in neural crest development. Dev Dyn 229:87–98.

Lamason RL, Mohideen M-APK, Mest JR, Wong AC, Norton HL, Aros MC, Jurynec MJ, Mao X, Humphreville VR, Humbert JE, Sinha S, Moore JL, Jagadeeswaran P, Zhao W, Ning G, Makalowska I, McKeigue PM, O'Donnell D, Kittles R, Parra EJ, Mangini NJ, Grunwald DJ, Shriver MD, Canfield VA, Cheng KC. 2005. SLC24A5, a Putative Cation Exchanger, affects pigmentation in zebrafish and humans. Science 310:1782–1786.

Lavoué S, Miya M, Inoue JG, Saitoh K, Ishiguro NB, Nishida M. 2005. Molecular systematics of the gonorynchiform fishes (Teleostei) based on whole mitogenome sequences: implications for higher-level relationships within the Otocephala. Mol Phylogene Evol 37:165–177.

Mabee PM, Ashburner M, Cronk Q, Gkoutos GV, Haendel M, Segerdell E, Mungall C, Westerfield M. 2007. Phenotype ontologies: the bridge between genomics and evolution. Trends Ecol Evol. In press (July).

Masolo C, Borgo S, Gangemi A, Guarino N, Oltramari A. 2003. Ontology Library (final). WonderWeb Deliverable D18 http://wonderwebsemanticweborg/deliverables/documents/D18pdf

Mayden RL, Wiley EO. 1992. The fundamentals of phylogenetic systematics. In: Mayden RL, editor. Systematics, historical ecology, and North American freshwater fishes. Stanford, CA: Stanford University Press. p 114–185.

Mayden RL, Tang K, Conway KW, Freyhof J, Chamberlain S, Haskins M, Schneider L, Sudkamp M, Wood RM, Agnew M. et al. In press. Phylogenetic relationships of *Danio* within the Order Cypriniformes: A framework for comparative and evolutionary studies of a model species. J Exp Biol: Mol Dev Evol (this volume).

McAllister DE. 1968. Evolution of branchiostegals and classification of teleostome fishes. Bull Nat Museum Canada, Biol Series 221:1–239.

Meyer A, Biermann C, Orti G. 1993. The phylogenetic position of zebrafish (*Danio rerio*), a model system in developmental biology: an invitation to the comparative method. Proc R Soc Lond B 252:231–236.

Meyer A, Ritchie PA, Witte K-E. 1995. Predicting developmental processes from evolutionary patterns: a molecular phylogeny of the zebrafish (*Danio rerio*) and its relatives. Philos Trans R Soc LondB Biol Sci 349:103–111.

Miller CT, Yelon D, Stainier DY, Kimmel CB. 2003. Two endothelin 1 effectors, hand2 and bapx1, pattern ventral pharyngeal cartilage and the jaw joint. Development 130:1353–1365.

Miya M, Takeshima H, Endo H, Ishiguro NB, Inoue JG, Mukai T, Satoh TP, Yamaguchi M, Kawaguchi A, Mabuchi K, Shirai SM, Nishida M. 2003. Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences. Mol Phylogenet Evol 26:121–138.

Nelson JS. 2006. Fishes of the World. Hoboken, NJ: Wiley.

Nüsslein-Volhard C, Dahm R. 2002. Zebrafish. Oxford: Oxford University Press. 303p.

Page RDM, Holmes EC. 1998. Molecular evolution: a phylogenetic approach. Oxford, UK: Blackwell Science. 346p.

Parichy D, Johnson S. 2001. Zebrafish hybrids suggest genetic mechanisms for pigment pattern diversification in *Danio*. Dev Genes Evol 211:319–328.

Patterson C. 1988. Homology in classical and molecular biology. Mol Biol Evol 5:603–625.

Piotrowski T, Schilling TF, Brand M, Jiang YJ, Heisenberg CP, Beuchle D, Grandel H, van Eeden FJ, Furutani-Seiki M, Granato M, Haffter P, Hammerschmidt M, Kane DA, Kelsh RN, Mullins MC, Odenthal J, Warga RM, Nusslein-Volhard C. 1996. Jaw and branchial arch mutants in zebrafish II: Anterior arches and cartilage differentiation. Development 123:345–356.

Remane A. 1952. Die Grundlagen des naturlichen Systems der vergleichenden Anatomie und der Phylogenetick. Leipzig: Geest und Portig.

Rieppel O, Kearney M. 2002. Similarity. Biol J Linnean Soc 75:59–82.

Roth VL. 1984. On homology. Biol J Linnean Soc 22:13–29.

Saitoh K, Miya M, Inoue JG, Ishiguro NB, Nishida M. 2003. Mitochondrial genomics of ostariophysan fishes: perspectives on phylogeny and biogeography. J Mol Evol 56:464–472.

Saitoh K, Sado T, Mayden RL, Hanzawa N, Nakamura K, Nishida M. 2006. Mitogenomic evolution and interrelationships of the Cypriniformes (Actinopterygii: Ostariophysi): The first evidence toward resolution of higher-level relationships of the world's largest freshwater fish clade based on 59 whole mitogenome sequences. J Mol Evol 63:826–841.

Sanger TJ, McCune AR. 2002. Comparative osteology of the *Danio* (Cyprinidae: Ostariophysi) axial skeleton with comments on *Danio* relationships based on molecules and morphology. Zool J Linnean Soc 135:529–546.

Sawada Y. 1981. Phylogeny and zoogeography of superfamily Cobitoidea (Cyprinoidei, Cypriniformes). Mem Fac Fish Hokkaido Univ 28:65–223.

Schultze HP, Arsenault M. 1985. The panderichthyid fish *Elpistostege*: a close relative of tetrapods? Paleontology 28:293–309.

Siebert DH. 1987. Interrelationships among families of the order Cypriniformes (Teleostei) [Ph.D.]: City University of New York, New York, New York.

Smith B, Ceusters W, Klagges B, Kohler J, Kumar A, Lomax J, Mungall C, Neuhaus F, Rector AL, Rosse C. 2005. Relations in biomedical ontologies. Genome Biol 6:R46.

Sneath PHA, Sokal RR. 1973. Numerical taxonomy. San Francisco: Freeman.

Stainier DY, Fouquet B, Chen JN, Warren KS, Weinstein BM, Meiler SE, Mohideen MA, Neuhauss SC, Solnica-Krezel L, Schier AF, Zwartkruis F, Stemple DL, Malicki J, Driever W, Fishman MC . 1996. Mutations affecting the formation and function of the cardiovascular system in the zebrafish embryo. Development 123:285–292.

van Eeden FJ, Granato M, Schach U, Brand M, Furutani-Seiki M, Haffter P, Hammerschmidt M, Heisenberg CP, Jiang YJ, Kane DA, Kelsh RN, Mullins MC, Odenthal J, Warga RM, Nusslein-Volhard C. 1996. Genetic analysis of fin formation in the zebrafish, *Danio rerio*. Development 123:255–262.

Wagner GP. 2001. The character concept in evolutionary biology. San Diego, CA: Academic Press. 622p.

Webb JF, Schilling TF. 2006. Zebrafish in comparative context: a symposium. Integr Comp Biol 46:569–576.

Weber EH. 1820. De aure et auditu hominis et animalium, Pars I. De Aure Animalium Aquatilium. Leipzig.

Wiley EO. 1981. Phylogenetics: the theory and practice of phylogenetic systematics. New York: Wiley.

Wolstencroft K, McEntire R, Stevens R, Tabernero L, Brass A. 2005. Constructing ontology-driven protein family databases. Bioinformatics 21:1685–1692.

Yan YL, Willoughby J, Liu D, Crump JG, Wilson C, Miller CT, Singer A, Kimmel C, Westerfield M, Postlethwait JH. 2005. A pair of Sox: distinct and overlapping functions of zebrafish sox9 co-orthologs in craniofacial and pectoral fin development. Development 132:1069–1083.