Zootaxa 2946: 79–117 (2011)
www.mapress.com/zootaxa/

Editorial

ISSN 1175-5326 (print edition)
ZOOTAXA
ISSN 1175-5334 (online edition)

The Cause and Effect of Polarization: Thoughts on the "Morphological vs. Molecular Debate" in Systematics, with Examples from the Study of Sturgeons (Actinopterygii: Acipenseridae)

CASEY B. DILLMAN[1] & ERIC J. HILTON[1]

[1] *Department of Fisheries Science, Virginia Institute of Marine Science, College of William & Mary, Gloucester Point, VA 23062, U.S.A. Email: ehilton@vims.edu; dillmanc@vims.edu*

**In*: Carvalho, M.R. de & Craig, M.T. (Eds) (2011) Morphological and Molecular Approaches to the Phylogeny of Fishes: Integration or Conflict?. *Zootaxa*, 2946, 1–142.

The varied possibilities to interpret data about taxa, their interrelationships and their geographic distributions, may be seen as different methods of analysis of data of these kinds… Different methods, even applied to the same data, tend towards different results. In a historical sense, different and conflicting results—different histories—cannot all be true. At least some must be artifactual and, therefore, method-generated (Nelson & Ladiges, 2001: 389)

Introduction

As eloquently stated by Nelson and Ladiges (2001) in the passage above, the way that systematic data is analyzed almost certainly will affect how it becomes interpreted. Often the way in which data are analyzed will depend on the type of data being considered. Regardless of the methods employed or data under consideration, there are several levels to analysis and systematic studies - whether based on morphology or molecules. For Grande and Bemis (1998:647), this followed an empirical exploratory-analytical-explanatory progression, so that "no conscious assumptions about processes of speciation, completeness of the fossil record, geographical dispersal, specific ancestors, specific areas of endemism, or 'true' phylogenies" will bias the results of the analysis. We agree with this as a starting point, but rephrase and add to it here. In our view, systematic studies encompass four phases: 1, Data generation, which is equivalent to Grande and Bemis' "empirical exploratory" phase, and for morphologically based studies this involves specimen examination and comparisons, whereas in molecular studies, this involves amplification and sequencing of specific genes; 2, Data analysis, which is similar to Grande and Bemis' analytical phase, and involves "sorting" of data, already with *hypotheses* of homology in place (similarly scored characters, sequence alignment, etc.), by some method (parsimony, likelihood, or Bayesian), to search for patterns in the data. We see Grande and Bemis' "explanatory" phase as encompassing two distinct yet interrelated activities: 3, data examination—an iterative process of checking and re-checking character definitions or character coding, often spurred on by primary homology statements being interpreted as homoplasies in light of other characters; and 4, data explanation, or interpretation of evolutionary scenarios (e.g., life history evolution, biogeography, etc.) based on the pattern of phylogeny resulting from the analysis of character data.

Although these four steps can be delimited (among others, we suppose), and give the impression that systematic biology is a straightforward process, in reality no phase is unto itself. For example, the exploratory phase is actually a combination of naïve observation (data generation) and data examination, in that hypotheses of homology are already implicit in comparisons between taxa (comparing features of any particular structure presupposes that the two structures are the "same" thing; this presupposition comes from various lines of evidence, including topographic relationships and ontogeny (Rieppel and Kearney 2002). coupled to experience and/or background knowledge. It is in this regard that systematics is an iterative discipline, with each step coming closer to the "truth,"

Accepted by M.R. de Carvalho: 2 May 2011; published: 8 Jul. 2011

which, of course, is unknowable. With truth as an unattainable goal, we do the best we can with the data that are available—both original and from background knowledge – to better refine our hypotheses of homology and of relationships.

Different phases of the systematic process are emphasized by different research programs, and more specifically by different researchers. Researchers develop expertise in collecting particular forms of data, and therefore labs become specialized. This specialization may be reflected in the data themselves. For instance, the types of data collected in a morphological study are in some ways different than those collected in a molecular study. Often, molecular data are more amenable to statistical analyses and modeling than morphological data, and to treat them identically would be to disregard these differences. The questions that can be addressed with types of data are also different. Is this to say that one or the other of the two approaches is any less valuable? We use this forum to at least tangentially explore this question and to comment on the "morphology vs. molecules" debate in systematic ichthyology. Although we come to this issue from the perspectives of a comparative morphologist (EJH) and a molecular systematist (CBD), we together firmly contend that there is a wealth of systematic information to be found in both data sources, and that this divide is to some degree arbitrary and unhelpful for the progress of systematic biology generally. Indeed, each data source provides information that the other cannot, and to view this debate as a simple dichotomy of data type is overly simplistic. Additionally, we see the polarization of different "camps" of systematists as not helpful to the task at hand—to discover patterns in nature that may be used to explain evolutionary relationships among organisms and as a basis for other evolutionary studies. As evidence in support of our perspective, we draw on examples from our past studies on sturgeons and their relatives, and use these data to reciprocally illuminate character distributions on reconstructed hypotheses of relationships.

The family Acipenseridae, the sturgeons, comprises 25 extant species classified in four genera (*Acipenser*, *Huso*, *Scaphirhynchus*, and *Pseudoscaphirhynchus*). Only two monotypic fossil genera (the very poorly preserved taxon †*Protoscaphirhynchus* and the well-preserved †*Priscosturion*) are known despite having a nearly continuous fossil record from the Late Cretaceous to the Recent (see Hilton and Grande 2006; Grande and Hilton 2006, 2009). The family Acipenseridae is the sister group of the family Polyodontidae (paddlefishes), and together with two entirely extinct families, form the order Acipenseriformes. Collectively, Acipenseriformes are the most species rich group of basal (i.e., non-teleostean) actinopterygian fishes to have extant members, with the sturgeons forming the bulk of this group (25 of 27 extant species). Most species of sturgeons are imperiled, with some members critically endangered, and at least one member (*P. fedtschenkoi*) thought to have recently gone extinct. Because they are culturally and economically important (e.g., as the source of caviar) there is general interest in sturgeon biology in both scientific and public circles (e.g., Saffron 2002, Carey 2005). These long-lived, highly migratory fishes have been heavily impacted by human activities, and most are endangered or threatened. While much research has been conducted on these fishes from ecological, behavioral, and conservation perspectives (Birstein 1993; Billard and Lecointre 2001), there are significant gaps in our understanding of sturgeon evolution.

Taxonomy and classification are essential tools for communication of biological data, and these need to convey the phylogenetic history of a group so that generalization of known information may be extended to taxa for which data is unknown. The modern taxonomic conceptualization of the family Acipenseridae has been used for over 100 years (Berg 1904), which in and of itself is not a problem, and suggests a stable classification. However, what is troubling is that the four-genera system for extant taxa does not reflect what is known about the phylogenetic relationships, as they are currently understood from all data sources. Most significantly, the fact that *Acipenser* – which with 17 species is the most species-rich genus of sturgeons—is paraphyletic is well established (Bemis *et al.* 1997; Findeis 1997; Birstein *et al.* 2002; Dillman *et al.* 2007; Hilton and Forey 2009; Hilton *et al.* unpublished). When Birstein *et al.* (2002) published their analysis, which concluded that the Scaphirhynchini (= *Scaphirhynchus* + *Pseudoscaphrirhynchus*) was paraphyletic, with *Psuedoscaphirhynchus* being most closely related to *A. stellatus*, one of us (EJH), who had been working on the anatomy of sturgeons for a few years and was just embarking on a morphological systematic study of the family, recalls being unconvinced despite the 100% bootstrap value for this node: this could not be "true." However, as described in Hilton (2005), upon examination of specimens of *A. stellatus* and *P. kaufmanni*, all characters that were at the time considered to be synapomorphies of the genus *Pseudoscaphirhynchus* (Findeis, 1993) were also found in specimens of *A. stellatus*. This has been supported in subsequent parsimony analyses of an expanded data matrix (Hilton and Forey, 2009; Hilton *et al.*, 2011). This example was used by Hilton (2005: 143) to justify the statement that "… novel sister-group hypotheses based on molecular data can serve as a useful impetus to reexamine anatomical features to see if there is morphological sup-

port for such hypotheses. It also shows that there is much still to be learned from morphology." We further use this as the focal point for discussion of what precisely can be learned from closer examination of these competing data sets.

Materials and Methods

To investigate "The Crisis" (Mooi and Gill, 2010) a comparative approach between methodologies was taken. To achieve this, the distribution of characters supporting hypothesized relationships and hypothesized homologies among acipenserids were investigated using molecular (Dillman *et al*. 2007) and morphological (Hilton and Forey 2009) data. First, the data sets were matched in terms of terminal taxa. Second, the molecular data set was reduced to Cytochrome *b* (cyt *b*) sequences only, primarily due to the high number of indels in the control region of Acipenseriformes (which could present a problem with alignment as noted in Mooi and Gill, 2010). Additionally, the data set of Hilton and Forey (2009) was reduced to extant taxa only to allow for direct comparisons to the molecular data set. Third, these data sets were re-analyzed utilizing only parsimony reconstructions (as to avoid the criticism of variation in statistical approaches). The resulting topologies were compared to the original publications for topological consistency. We also conducted statistical tests of topological congruence utilizing both the Templeton test and the K-H test (Templeton 1983; Kishino & Hasegawa 1989). The topologies constructed from each matrix were compared to the topologies from the matrix of the other data set to determine if the topologies were significantly different across the data sets. The final data sets included 12 terminal taxa (Table 1) with *Polyodon* serving as the outgroup in all investigations. Nodes were coded numerically for ease of discussion between topologies; for novel nodes (i.e., those nodes that didn't occur in both data molecular and morphological topologies, or sets of topologies) increasing consecutive numbers were used (Tables 2 & 3).

TABLE 1. Species and specimens of Acipenseridae examined in this study. Note that additional specimens were examined (e.g., over 100 specimens of *A. brevirostrum* were examined in the context of morphological and character descriptions) but are not listed here for brevity. Institutional abbreviations: AMNH, American Museum of Natural History; CAS, California Academy of Sciences; FMNH, Field Museum of Natural History; MCZ, Museum of Comparative Zoology, Harvard University; UAIC, University of Alabama Ichthyology Collection; UMA, University of Massachusetts Amherst; UMMZ, University of Michigan Museum of Zoology; USNM, National Museum of Natural History, Smithsonian Institution; VIMS, Virginia Institute of Marine Science.

Taxon	Specimens examined	Genbank Number	Voucher specimen
Polyodon spathula	AMNH 218237, FMNH 98245, FMNH 98287	NC_004419	N/A
Huso huso	AMNH 1553, CAS 211810, MCZ 54269	AJ245840	N/A
Scaphirhynchus platorhynchus	FMNH 45024, FMNH 98286, VIMS 12098	U56988	N/A
Pseudoscaphirhynchus kaufmanni	AMNH 97566, UAIC 13265.01	EF484339	
Pseudoscaphirhynchus hermanni	UAIC 13252.01, UAIC uncat.	EF484338	
Acipenser brevirostrum	FMNH 112209, MCZ 54167, VIMS 12090,	AJ245828	N/A
Acipenser stellatus	MCZ uncat., UMMZ 184980	AY846701	N/A
Acipenser ruthenus	AMNH 1554, UMA F10369	EU733251	N/A
Acipenser fulvescens	FMNH 85157, FMNH 98256, UMMZ 223764	AJ245829	N/A
Acipenser oxyrinchus	USNM 94726, USNM 110207, VIMS uncat.	AJ245838	N/A
Acipenser transmontanus	FMNH 117777, FMNH 117779, VIMS 12097	AF184107	N/A
Acipenser baerii	FMNH 117786, FMNH 117785, VIMS 12083	AF238644	N/A