The Importance of Even Highly Incomplete Fossil Taxa in Reconstructing the Phylogenetic Relationships of the Tetraodontiformes (Acanthomorpha: Pisces)¹

FRANCESCO SANTINI* & JAMES C. TYLER†
*Department of Zoology, University of Toronto, 25 Harbord Street, Toronto, MSS 3G5, Canada
†National Museum of Natural History, Smithsonian Institution, Washington, D.C. 20560-0106

SYNOPSIS. The use of fossils in the phylogenetics of extant clades traditionally has been a contentious issue. Fossils usually are relatively incomplete, and their use commonly leads to an increase in the number of equally most parsimonious trees and a decrease in the resolution of phylogenies. Fossils alone, however, provide certain kinds of information about the biological history of a clade, and computer simulations have shown that even highly incomplete material can, under certain circumstances, increase the accuracy of a phylogeny, rather than decrease it.

Because empirical data are still scarce on the effects of the inclusion of fossils on phylogenetic reconstructions, we attempted to investigate this problem by using a relatively well-known group of acanthomorph fishes, the Tetraodontiformes (triggerfishes, pufferfishes, and ocean sunfishes), for which robust phylogenies using extant taxa already exist and that has a well-studied fossil record. Adding incomplete fossil taxa of tetraodontiforms usually increases the number of equally most parsimonious trees and often decreases the resolution of consensus trees. However, adding fossil taxa may help to correctly establish relationships among lineages that have experienced high degrees of morphological diversification by allowing for a reinterpretation of homologous and homoplastic features, increasing the resolution rather than decreasing it. Furthermore, taxa that were scored for 25% or more of their characters did not cause a significant loss of resolution, while providing unique biological information.

INTRODUCTION

The use of fossils in the phylogenetics of extant clades traditionally has been a contentious issue ever since Hennig first proposed the theory of phylogenetic systematics (1966; see also Lovtrup, 1977; Patterson, 1981; Gauthier et al., 1989; Donoghue et al., 1989; Huelsenbeck, 1991; Novacek, 1992; Smith, 1994; Wiens and Reeder, 1995; Wilkinson, 1995; Wiens, 1998, 2003a, b). Kearney, 2002).

Most of the criticism centered on the inclusion of fossil taxa in phylogenetic inference is because fossils are to varying degrees incomplete, with potentially many data missing. For this reason fossils (as well as other incomplete taxa in general) have been shown in certain analyses to increase the number of equally most parsimonious trees (EMPTs) and to decrease the resolution of the resulting consensus trees, thus reducing the accuracy of hypothesized trees (Huelsenbeck, 1991; Wiens and Reeder, 1995; Wiens, 2003b). It has also been claimed that the inclusion of fossil taxa has never overturned hypotheses of phylogenetic relationships based on extant lineages alone (Patterson, 1981).

For these reasons, many neontologists still prefer to produce phylogenetic hypotheses by analyzing extant taxa only, ignoring the information potentially available from the study of the fossil record. In our opinion, however, there are two main compelling reasons that both justify and encourage the use of fossil taxa in our phylogenetic analysis.

First, subsequent to Patterson’s seminal paper, various examples have been presented in which the inclusion of fossils provides additional valuable phylogenetic information that is not recoverable from the study of Recent lineages alone (Gauthier et al., 1988; Donoghue et al., 1989; Novacek, 1992; Smith, 1994). Second, in some of the debate surrounding the incompleteness of fossil material and its negative effect on the accuracy of phylogenetic inferences, we are aware of very few studies that have ever attempted to test the effect of relatively incomplete taxa on phylogenetic reconstructions by comparing phylogenies produced using only fossil or a combination of fossil and extant taxa versus those produced with the use of only extant taxa (e.g., Novacek, 1992; Grande and Bemis, 1998).

These studies mostly showed that the influence of incomplete fossils cannot easily be predicted before the computer-based cladistic analysis, but one common indication seems to be that even fairly incomplete materials can be included before many nodes in the cladograms start to collapse (i.e., Grande and Bemis, 1998, were able to successfully include in their study taxa scored for as little as 64% of their characters, even though they commented that the cut-off point would probably differ from study to study). Moreover, there have been theoretical studies (Wiens, 1998, 2003a, b) that have shown how even very incomplete taxa can provide highly useful phylogenetic information, and sometimes even increase the accuracy of phylogenetic reconstructions. In particular, Wiens (2003a, b) showed that the most important factor that impacts on the ability to recover a phylogeny is not the percentage of
unknown characters, but the overall number of characters that can be scored. In these simulations even taxa that were scored for only 5% of their characters could still be included and correctly placed in the topology when the number of characters that could be scored was relatively high (italics ours). We use the term relatively because in Wiens simulations the data sets contained 16 taxa and 100, 500, 1,000 or 2,000 characters, and having taxa scored for just 5% of 1,000 identified characters still means having 50 potentially informative features at ones disposal. Few paleontological data sets will approach the size of the largest of Wiens data sets (at least for the number of characters) but an intriguing aspect is that even with the smaller (and more realistic) data sets (i.e., 100 and 500 characters), very incomplete materials can still produce accurate reconstruction of the phylogeny. Wiens (2003b) also commented that in some cases perhaps only a single synapomorphy may be needed to accurately place a taxon on a tree. Thus, the larger number of characters increases the probability that such key characters will be found. This theoretical result certainly deserves additional empirical testing. Furthermore, the principle of “total evidence” (Kluge, 1989) often is invoked to justify the inclusion of various data sets into a common matrix, and it calls for the utilization in phylogenetic analyses of all available sources of data (and this includes both taxa and characters).

In this paper we use a recently assembled data set of 210 morphological features (mostly osteological) for a group of 22 extant and 36 fossil tetraodontiform fish taxa to test the effect of the inclusion of variously highly incomplete to relatively complete fossil materials on phylogenetic inference and to attempt to confront our empirical results with Wiens theoretical findings. Obviously, in simulation studies the topology inferred from the analysis of the incomplete data set can be compared with the known “true” topology in order to assess the ability to recover the correct phylogeny. No such comparison is possible in actual case studies, when the “true” phylogeny is not known. In our case, we will use congruence among the topologies produced by the analyses of various data sets and the previous classifications and phylogenetic hypotheses that have been performed with various methods (both evolutionary or pre-cladistic and more modern) and sources of data (osteological, myological, larval) (Fraser-Brunner, 1950; Tyler, 1968, 1980; Winterbottom, 1974; Winterbottom and Tyler, 1983; Leis, 1984; Klassen,1995; Tyler and Santini, 2002; Santini and Tyler, 2003) as a way of inferring the correctness of analyses that include both fossil and extant taxa. Specifically, for each fossil taxon, we assume that it is correctly placed when these authors and our analyses place it in the same non-controversial clade.

Our phylogenetic work with fossil tetraodontiforms is relatively unusual within the Acanthomorpha (or spiny-rayed fishes). This large clade of fishes, with over 16,000 species described (Froese and Pauly, 2002), appears in the fossil record towards the end of the Cretaceous (Patterson, 1993), and seems to have undergone an explosive diversification during the beginning of the Tertiary (Bellwood, 1996). To date, very little of the fossil diversity of the Acanthomorpha has been described, and even less has been incorporated in phylogenetic analyses that also include extant taxa, while fossil materials have been more commonly included in investigations of the phylogenetic relationships of basal Actinopterygii and lower Teleostei (e.g., many papers in Arratia and Schultze, 1999; also Arratia, 1995, 1997; Stiassny et al., 1996; Grande and Bemis, 1998).

The study group

Extant Tetraodontiformes is a group of approximately 400 species (Froese and Pauly, 2002) that are divided among ten families of great diversity in body form and osteological structure (Tyler, 1980; Santini and Tyler, 2003): the Triacanthodidae (spikefishes); Triacanthidae (triplespines); Balistidae (triggerfishes); Monacanthidae (filefishes); Aracanidae (boxfishes); Ostraciidae (trunkfishes); Tetraodontidae (pufferfishes); Diodontidae (porcupinefishes); Triodontidae (three-tooth puffers); and Molidae (ocean sunfishes).

The tetraodontiforms have a fairly well-known fossil record, dating back to the Cretaceous, which was recently reviewed by Tyler and Santini (2002), and that is among the reasons that contribute to make this clade the focus of great attention by the community of evolutionary developmental biologists (Santini and Stiassny, 2002). Tyler and Santini (2002) assigned the fossil taxa to various extant groups that had been given status at higher taxonomic levels (subfamilial to sub-btatal) by previous workers (most notably Tyler, 1968, 1980; Winterbottom, 1974; Tyler and Santini, 2002). These taxonomic assignments were made before a comprehensive phylogenetic hypothesis was available for extant and fossil taxa combined. Subsequently, Santini and Tyler (2003) proposed a phylogeny and classification based on a cladistic analysis of a large data set of 210 morphological characters for 20 extant and 36 extinct species of tetraodontiforms, using two extant outgroup taxa (a zeiform and a caproid). In that paper we provided phylogenies for the extant taxa alone and for all of the extant and fossil taxa together. Because of space constraints, we could not include in that paper a detailed discussion of the effect that the inclusion of some of these fossil materials, especially the most highly incomplete ones (scored for only 10 or 17 of the 210 characters), have on the attempt to infer phylogenetic relationships. The same constraints also prevented us from giving a comprehensive comparison of the topologies produced by various combinations of extant and fossil taxa. Thus, what remains to be explored are the phylogenetic results when only the best known fossils are included in the analysis, which is often the case if fossils are included at all. As opposed to results when very incomplete taxa are added to those with a fuller complement of known features. In this paper we will build on that
Fig. 1a. Strict consensus tree of the three EMPT, with L 457, CI 0.69, RI 0.82, produced by the analysis of the extant species only. b. Strict consensus tree of the three EMPT, with L 546, CI 0.59, RI 0.79, produced by the analysis of all taxa scored for 100 or more characters. c. Strict consensus tree of the 15 EMPT, with L 533, CI 0.59, RI 0.79, produced by the analysis of all taxa scored for 90 or more characters. d. Strict consensus tree of the eight EMPT, with L 564, CI 0.57, RI 0.79, produced by the analysis of all taxa scored for 80 or more characters.
earlier work (Santini and Tyler, 2003) and present a comparison of the topologies produced by the analyses of data sets containing various combinations of fossil and extant taxa.

**Materials and Methods**

Santini and Tyler (2003) developed the character data matrix for living and fossil Tetraodontiformes that is the basis for the analysis presented here, and we refer the reader to that paper for explanations concerning the selection of intra- and extraordinary taxa and of rooting method. The phylogenetic analyses were performed with WINCLADA (Nixon, 2002) in combination with NONA (Goloboff, 1999). All characters were given equal weight (1), and all multistate characters were treated as unordered, given that the direction of the evolutionary transformation is not known. Heuristic searches, with random addition of taxa, TBR branch swapping, and 1,000 replications were performed. Tree length (L), consistency index (CI), and retention index (RI) are provided for each analysis (Klug and Farris, 1969; Farris, 1989). A strict consensus tree was calculated when more than one most parsimonious tree was obtained after an analysis, and the decay index (Bremer, 1994; “Bremer support,” Källersjö et al., 1992) is shown above each branch to indicate the degree of support for the individual clades. Cladograms for publication were produced using TREEVIEW (Page, 1996).

**Results and Discussion**

Analysis of the 210 characters for the 22 extant species only (20 tetraodontiforms plus two outgroups) recovers two EMPTs, with L 457, CI 0.69, RI 0.82, and each of the extant families is recovered as monophyletic. The strict consensus tree (Fig. 1a) differs from the previously most widely accepted cladistic hypothesis for the phylogeny of the Tetraodontiformes (Winterbottom, 1974) mostly for the tricotomy that includes Balistoida (trigger and filefish), Ostracioida (box and cowfishes) and the Tetraodontoida (puffer, porcupine fishes, and molas). In Winterbottom’s (1974) myologically based phylogeny, the Ostracioida and Balistoida were each other’s sister groups—as they also appeared in Tyler’s (1980) non-cladistic osteological phylogeny, in Winterbottom and Tyler’s (1983) analysis of a combined myological and osteological data-set, and in Klassen’s (1995) treatment of ostracoid relationships. A second difference between this analysis and Winterbottom’s concerns the placement of the Triacanthidae, which appear here as the sister group to all tetraodontiforms with the exception of the Triacanthodidae, as opposed to being the sister group of the Triacanthodidae; Winterbottom did, however, point out that the Triacanthodidae + Triacanthidae sister group relationships was only very weakly supported, and proposed that the Triacanthidae might be the sister group to all remaining tetraodontiforms, to the exclusion of the triacanthodids. We note that while most familial and superfamilial clades are recovered with a moderate to high decay index, the support for the most basal relationships is very weak.

When the most complete fossil taxa are added to the analysis (in this case the 11 fossil taxa scored for at least 100 out of 210 characters; total number of taxa 33, two outgroups plus 31 tetraodontiforms), three EMPTs were recovered, with L 546, CI 0.59 and RI 0.79 (Fig. 1b). The Cretaceous Plectocretacicoidea are shown to be the most basal tetraodontiform lineage, followed by: the Triacanthodidae, which is the sister group to all of the other tetraodontiforms, both extant and fossil; the Triacanthidae (including Protacanthodes); Bolcabalistes; the two clades Balistoida + Ostracioida; and the Tetraodontoida. All extant families are recovered as monophyletic. The only unexpected result produced by this analysis is the placement of Bolcabalistes, which we had expected to be a sister group of the Balistoida.

When the three additional fossil taxa scored for 90 or more characters are included (36 total taxa; Fig. 1c), 15 EMPTs are recovered, with L 553, CI 0.59, RI 0.79. The Triacanthidae are again the only extant family that does not appear as monophyletic in the strict consensus tree, and it changes position in the topology of this tree, with the two Eocene Protacanthodes and the clade of Oligocene and extant triacanthids appearing in a polytomy with the Bolcabalistes + (Balistoida + Ostracioida) clade. Bolcabalistes, as intuitively expected, appears as the sister group to Balistoida + Ostracioida, while the newly analyzed Eoplectus is shown to be the sister group of all Tetraodontoidae.

When three more taxa scored for 80 or more characters are included (total number of taxa 39; Fig. 1d), the analysis produces eight EMPTs, with L 564, CI 0.57, RI 0.79. This topology is virtually identical to the preceding ones, except for the Triacanthidae being a monophyletic group again. The first two fossils that appear to belong within the clade of extant Tetraodontoei are included in this data set (Eotetraodon and Archaeotetraodon jamestyi) and they both appear to be basal Tetraodontidae.

When all taxa scored for 70 or more characters—one third or more of the total number of characters—are included in the data set (total number of taxa 42; Fig. 2a), 17 EMPTs are recovered, with L 571, CI 0.57, RI 0.82. In spite of the much higher number of
EMPTs recovered than in the previous analyses, the topology of the consensus tree shows how most of the uncertainty is confined to the relationships among the three species-richest clades (Triacanthidae, Balistoidea + Ostracioidae and Tetraodontoidae), and among the species within the Balistidae and the Triacanthidae, the two families with the greatest number of fossils at this stage in the series of analyses (four and six, respectively). A new taxon, Moclaybalistes, is introduced in the analysis, and it appears in a polytomy with: Eoplectus; the Triacanthidae; Balistoidea + Ostracioidae; and the Tetraodontoidae.

When all taxa scored for 60 or more characters are included (45 total taxa), 862 EMPTs are recovered, with L 579, CI 0.56, and RI 0.80 (consensus tree not shown). In this case a great source of uncertainty is represented by the Tetraodontidae, which now comprise three extant plus four fossil species, all appearing in a polytomy. The only unexpected result in this case is the newly analyzed Zignoichthys, which has always been classified as a tetraodontoid by previous workers (see its history in Tyler and Santini, 2002), now appears as the sister group of the Ostracioida, even though the node that supports such a clade has a minimal decay index of one.

When all taxa scored for 50 or more characters are included (47 total taxa; Fig. 2b), bringing the number of known characters in some of the species in the data set to about 24% of the total, 3,028 EMPTs are recovered, with L 585, CI 0.55, and RI 0.80. The Triacanthidae, Balistoidea, Ostracioida, Diodontidae, and Molidae still appear as monophyletic groups, but they all appear in a polytomy with Bolcabalistes, Moclaybalistes, Eoplectus, the extant Triodon macroperius, Zignoichthys, and the three extant and four fossil putative Tetraodontidae.

Analysis of all taxa scored for 40 or more characters (51 total taxa) or 30 or more characters (55 total taxa) shows a drastic increase in the number of EMPT (up to 7,950 for the analysis of taxa scored for 40 or more characters) but no significant changes in the overall topology of the consensus trees, hence due to reasons of space these trees are not shown.

Analysis of all taxa scored for 20 or more characters (56 total taxa; Fig. 2c) recovers 19,056 EMPTs, with L 602, CI 0.54, and RI 0.80. The topology of the strict consensus tree is virtually identical to the previous one, with the exception of a new taxon, Eospinus, being added to Moclaybalistes and Bolcabalistes in a polytomy with the Balistoida + Ostracioida clade.

Analysis of all 58 extant and fossil taxa (Santini and Tyler [2003], using PAUP recovered 9,332 EMPTs. Given that the topologies of the strict consensus trees produced by these two analyses do not differ from one another, we reproduce here in Fig. 2d the result from Santini and Tyler [2003],) recovered 13,200 EMPTs, with L 602, CI 0.54, RI 0.80. The topology is very similar to those of the previous two analyses, with monophyletic Diodontidae and Molidae recovered. Of special interest to us is the fact that Eomola, known from only a few jaw bones and that could be scored for only 10 of the 210 characters, nine of which features appear to be plesiomorphic and only one derived (individual fused teeth in parrot-like beak indistinguishable), still appears in the strict consensus tree as the sister group of the two extant Molidae investigated (thus confirming one of the predictions of Wiens, 2003b, that a single synapomorphy may be enough to correctly place a taxon in spite of the phylogenetic “noise” that may be present in a data set).

Various other combinations of extant and fossil taxa have been investigated, even though space limitations prevent us from going into greater detail here. The most interesting cases that we wish to discuss are those that involve analyses of fossil only data sets. Two data sets including all of the fossil taxa plus outgroups were analyzed: one of these also included the two extant taxa of Monacanthidae, which is the only extant family not represented by fossils in this work, whereas the other data set excluded the two extant Monacanthidae. Analysis of the data set with the Monacanthidae recovered 9,146 EMPTs, with L 366, CI of 0.63 and RI of 0.76 (Fig. 3a). In this analysis three main lineages of tetraodontiforms are recovered: the Plectocretacioid is the most basal lineage, followed by the Triacanthodidae and then by the clade formed by all other Tetraodontiformes. Within this monophyletic latter clade, the Triacanthidae, Balistoidea, and Ostracioida can be recovered, all originating in a polytomy with all other taxa. While the congruence of this topology with those recovered by analyzing exclusively or mostly extant species is rather surprising, even more stunning is the congruence of the topology of the strict consensus of the 7,371 EMPTs, with L 328, CI 0.63, and RI 0.74, recovered from analysis of the data set that contains only fossil tetraodontiforms plus the two outgroups (Fig. 3b). In this case, all extant families are recovered as monophyletic by the analysis of their extant taxa only, and the only major departure from “accepted wisdom” is the fact that the highly incomplete (for characters) fossil Ostracioida form a clade nested within the equally highly incomplete Diodontidae. In spite of this, we find it amazing that had the Tetraodontiformes become completely extinct, we would still be able to recover their infrafamilial relationships almost as well as we do by analyzing their living relatives.

The final analysis that we wish to show, and in this case also a cautionary tale, concerns the data set formed by extant taxa only plus the relatively well known Eoplectus. Even though Tyler (1980) placed Eoplectus in the Triacanthodidae in his evolutionary taxonomic analysis, he stressed that it was the ancestor of the Tetraodontoidae, and all cladistic analyses have always placed it as the most basal member of the Tetraodontoidae (Winterbottom, 1974; Santini and Tyler, 2003; this paper). When Eoplectus is the only fossil in a data set of extant forms, the analysis recovers one most parsimonious tree, and in this Eoplectus appears as the sister group to all Tetraodontiformes except the
Triacanthodidae (Fig. 3c). The causes of this are not yet fully evident, but they are probably related to the presence in this very basal tetraodontoid of many features that appear as plesiomorphic when this limited data set is analyzed. When more fossils are added to the data set, several characters can be reinterpreted and Eoplectus is recovered in our topology where we would expect it to be, which is as a stem tetraodontoid. This example may serve to illustrate the importance of adding as many taxa as possible when dealing with...
highly derived groups, and not just one or few of the best known fossils.

To summarise briefly our findings, the main areas of disagreements between the results of our analyses and the previous phylogenetic hypotheses concern mostly the placement of the family Triacanthidae, and the relationships between the Balistoidae, the Ostracioidae and the Tetraodontoidae. The Triacanthidae never appear as the sister group of the deep-water Triacanthodidae in our analyses, and they are likely part of a clade containing the Balistoidae and Ostracioidae. The Ostracioidae appear in most analyses of combinations of extant and fossil taxa to be the sister group of the Balistoidae, supporting the earlier hypotheses of Winterbottom (1974), Tyler (1980), Winterbottom and Tyler (1983), and Klassen (1995), while the analysis of only the extant taxa in our data set suggests that the Ostracioidae might be related to the Tetraodontoidae (in agreement with the findings of Leis, 1984). We believe that this difference between the topologies recovered from the analysis of extant taxa only, and the topologies recovered from the analyses of extant and fossil taxa may be due to the fact that both the extant Ostracioidae and Tetraodontoidae are clades that are osteologically very derived when compared to all other tetraodontiforms, with many osteological changes that appear on the branches that lead to these two groups. Many of those changes, especially those connected to the reductive trends in the spiny dorsal fin and in the pelvic complex, appear to be convergences between the two lineages, but in a cladistic analysis may be more parsimoniously reconstructed as synapomorphies, and produce the clustering of Ostracioidae and Tetraodontoidae as sister groups (this is a phenomenon similar to what is known as “long-branch attraction” in molecular data; see Felsenstein, 1978). However, when fossil taxa that often show intermediate states for some structures that are fully developed in the basal Tetraodontiformes, and are very derived in Ostracioidae and Tetraodontoidae, are added to the analysis, these long branches are broken (as already discussed in Gauthier et al., 1988). As a result, greater congruence with the myological data is obtained, probably because the myological features in this case are more conservative than the osteological ones (an argument already briefly advanced by Borden, 1999, in his analysis of the surgeonfish genus *Naso*). If this observation is correct, we would hypothesize that myological features, and other kinds of morphological data that may be even more conservative, such as from nerves and soft tissues, may prove more useful when dealing with groups that have undergone extensive morphological diversification, if no fossil record is available for these taxa.

**Conclusion**

This work shows how inclusion of an increasing number of even poorly known fossils does not automatically result in poor resolution of the phylogeny and widespread collapse of many nodes. On the contrary, including large numbers of fossil taxa that are scored for about 25% or more of their characters does not appear to cause loss of information for the interfamilial relationships of the Tetraodontiformes. Furthermore, even a very low number of synapomorphic character states (down to just one as in the case of *Eomola*) may be enough to “correctly” place a very incomplete fossil taxon on the cladogram. Both these findings appear to corroborate the results of Wiens (2003a,b) simulations, and we hope to soon see more empirical tests of these.

An intriguing result of this work is that when fossils are added to the phylogeny, even if only the best known fossils are included, the topology that we recover is in greater agreement with that produced by a previous myological analysis (Winterbottom, 1974) than with the analysis of only the extant species in our data set. We think that this may be due to the fact that in a group that has likely experienced a rapid diversification of its main clades, characters such as musculature that are more conservative than those of the skeletal system may offer better chances of recovering the correct phylogeny, if fossil taxa have not been included to break long branches.

**Acknowledgments**

This work was made possible by many people at museums and universities around the world that made fossil and extant materials of tetraodontiforms available for examination throughout the years, as given in Tyler and Santini (2002). J. Wiens kindly made page proofs of his most recent papers available for examination, so that we could comment on the result of his most recent simulation studies. This work was supported by the Smithsonian Institution through two short term visitor fellowships to F. Santini for work with J.C. Tyler. F. Santini is currently funded by a Marie Curie fellowship for a project on “Paleontological and molecular approaches to the phylogeny of Acanthomorpha (Pisces).” Comments by J. Wiens and two other anonymous reviewers greatly improved the manuscript.

**References**


