Marine Biology (2005)
DOI 10.1007/s00227-005-0089-z

RESEARCH ARTICLE

Anna L. Bass · Heidi Dewar · Tierney Thys
J. Todd. Streelman · Stephen A. Karl

Evolutionary divergence among lineages of the ocean sunfish family, Molidae (Tetraodontiformes)

Received: 6 May 2005 / Accepted: 8 July 2005

Abstract Ocean sunfish, family Molidae, are enigmatic members of the epipelagic fauna of all tropical and temperate oceans. A study, begun in 1998, initially focused on the population genetics of *Mola mola* Linnaeus 1758 immediately indicated high levels of genetic divergence in the *d-loop* and *cytochrome* b mitochondrial genes. This preliminary effort was expanded to include *Masturus lanceolatus* Liénard 1840, *Ranzania laevis* Pennant 1776, and representative sequences of other Tetraodontiformes. Analysis of the sequence data confirms that there are two species in the genus *Mola*, *Mola mola* and *M. ramsayi* Giglioli 1883, with the latter presumed to be limited to the southern hemisphere. There is an indication of inter-ocean subdivision within both species originating 0.05–0.32 and 1.55–4.10 million years ago, respectively. Given limited sample sizes, however, the divergence estimates are minimums and the isolating mechanisms remain speculative. The systematic analysis provided strong support for the sister taxa relationship between genera *Masturus* and *Mola* and the basal position of the genus *Ranzania* within the family, as well as the sister group relationship of the Tetraodontiform families Tetraodontidae + Diodontidae to the Molidae.

Communicated by J.P.Grassle, New Brunswick

A. L. Bass · S. A. Karl (✉)
Department of Biology, SCA 110,
University of South Florida,
Tampa, FL 33620, USA
E-mail: karl@mail.cas.usf.edu
Tel.: +1-808-2367401
Fax: +1-808-2367443

H. Dewar
Tagging of Pacific Pelagics, Stanford University
c/o Inter-American Tropical Tuna Commission, La Jolla,
92037 CA, USA

T. Thys
Sea Studios Foundation, Monterey, CA 93940, USA

J. Todd. Streelman
School of Biology, The Georgia Institute of Technology,
310 Ferst Drive, Atlanta, GA 30332-0230, USA

S. A. Karl
The Hawai'i Institute of Marine Biology,
University of Hawai'i, Manoa, PO Box 1346,
Kane'ohe, HI 96744, USA

Introduction

Species in the family Molidae are large, primarily pelagic members of the Tetraodontiformes. Commonly referred to as ocean sunfish, the Molidae have a distinctive laterally compressed shape and "chopped off" appearance (Fraser-Brunner 1951; Smith and Heemstra 1986). Relying on median fins for swimming, they lack caudal bones, ribs, pelvic fins, spines or girdles and have fewer vertebrae than any other fish (Tyler 1980). Metamorphosis from larva to adult is remarkable in that, unlike most fish, they pass through two distinct larva phases—a typical *Tetraodon*-like larval and another highly transformative stage resulting in the complete absorption of the tail (Fraser-Brunner 1951).

Knowledge of the natural history of the Molidae is sparse, but they are recognized as the most fecund extant vertebrate with a single female capable of producing as many as 3×10^8 eggs at one time (Schmidt 1921; Parenti 2003). Members of the family are globally distributed in both temperate and tropical waters although there is some indication of restricted distributions for *M. mola* (Fraser-Brunner 1951; Smith and Heemstra 1986). Parin (1968) presented a distribution map compiled from several sources suggesting that spawning areas for *Ma. lanceolatus* and *R. laevis* overlap in the Sargasso Sea but with no clear indication of the spawning regions for *Mola* spp. Five spawning areas have been suggested for the Molidae in the central gyres in the North and South Atlantic, North and South Pacific, and Indian Oceans. Molids are reported to feed primarily on jellyfish and planktonic organisms although a variety of prey such as brittlestars, flounder, and leptocephalus larvae have

been found in their guts (Grassi 1897; Norman and Fraser 1949) indicating that they must sometimes feed on the bottom and among patches of floating seaweeds. Due to their primarily pelagic existence, they are difficult to observe and information on their natural history and behavior is generally lacking. Most frequently, they are sighted basking at the ocean surface (Norman and Fraser 1938). Recent research indicates that basking may be correlated with diving behavior as an attempt to replenish body heat after deep dives into cold water (~2–10°C; personal observation; Cartamil and Lowe 2004). It also appears from our tagging studies, that individual *M. mola* and *Ma. lanceolatus* show no indication of ocean basin scale movements and do not undertake large-scale migrations (unpublished data).

Taxonomically, the family Molidae has a long history in the scientific literature with two of the earliest descriptions in 1758 by Linnaeus and 1766 by Koelreuter (see Parenti 2003). Since that time, numerous genera (19) and species (54) of sunfish have been proposed (Parenti 2003) and the taxonomy of the family has been relatively volatile. The most complete taxonomic revision of the family (Fraser-Brunner 1951) distinguished five species in three genera; *R. laevis*, *Ma. lanceolatus*, *Ma. oxyuropterus*, *M. mola* and *M. ramsayi*. Currently, three species are commonly recognized; *R. laevis*, *Ma. lanceolatus*, *Mola mola*, with a fourth, *Mola ramsayi*, infrequently mentioned in taxonomic treatments (Smith and Heemstra 1986; Nelson 1994; Parenti 2003).

The relationship among currently recognized genera was the focus of two recent studies, one morphological and the other molecular. Santini and Tyler (2002) analyzed 48 morphological characters and found strong support for the traditional hypothesis of a sister taxon relationship between the genera *Masturus* and *Mola*, with *Ranzania* holding the basal position within the family. Using complete mitochondrial genome sequences, Yamanoue et al. (2004) also supported the traditional hypothesis and the morphological analysis. Identification of the sister-group of the Molidae, however, has been controversial and was not resolved even with the addition of complete mitochondrial genomes (Yamanoue et al. 2004). Multiple hypotheses have been presented regarding what family or families are closest to the Molidae: Diodontidae (Breder and Clark 1947), Tetraodontidae + Diodontidae (Winterbottom 1974; Tyler 1980; Santini and Tyler 2003), and Diodontidae + Ostraciidae (Leis 1984).

As part of ongoing research into the general ecology, movements, and distribution of *M. mola*, we have been using molecular data to examine population structure and biogeography of this species. We chose to analyze the mitochondrial control region (*d-loop*) because its general fast rate of evolution allows the differentiation of closely related groups and it is commonly used in intraspecific studies (Kocher and Stepien 1997). Our initial results, however, indicated genetic divergence levels within this species that were as great as levels commonly observed between species. We therefore expanded our analysis to include the mitochondrial *cytochrome* b gene sequences representing the identified divergent groups of *M. mola* and selected Tetraodontiform *cyt* b sequences from GenBank. Here, we present the results of these analyses in the context of evolution of globally-distributed, highly-fecund, pelagic marine fishes.

Materials and methods

From 1998 to 2004, tissue samples were taken from 13 putative *M. mola* Giglioli 1883, four *Ma. lanceolatus* Liénard 1840, and one *R. laevis* Pennant 1776. Samples were from multiple locations in the Atlantic and Pacific Oceans and were intended to represent the systematic diversity of the group (Table 1). Samples consisted of fin clips or muscle biopsies and were stored in 95% EtOH.

Initially, genomic DNA was isolated using a standard phenol/chloroform method, precipitated with sodium acetate and 100% EtOH, and resuspended in 1X TE (10 mM Tris HCl, pH 7.5, 1 mM EDTA acid). Later samples were isolated using a non-boiling Chelex method (Walsh et al. 1991) and stored at −20°C. Two regions of the mitochondrial DNA genome were amplified: control region (*d-loop*) and *cytochrome* b (*cyt* b). The primers A (5′-TTCCACCTCTAACTCCCAAA GCTAG-3′), E (5′-CCTGAAGTAGGAACCAGATG-3′) and M (5′-TATGCTTTAGTTAAGGCTACG-3′) of Lee et al. (1995) were used to amplify the *d-loop*. Initially, primers A and M were paired to amplify the entire *d-loop* [~800 base pairs (bp)] from six individuals, however, primers A and E (also from Lee et al. 1995) were used to amplify ~400 bp of the 5′ end of the remainder of the samples. For the *cyt* b gene, the forward primer (5′-GTGACTTGAAAAACCACCGTTG-3′) of Song et al. (1998) and reverse primer (5′-AATAG GAAGTATCATTCGGGTTTGATG-3′) of Taberlet et al. (1992) were used to amplify an approximately 750-bp fragment. Each amplification reaction (25 or 50 μl) of the *d-loop* region consisted of 1X Promega buffer, 1.25 U of Promega *Taq*, 0.5 mM dNTPs, 3 mM $MgCl_2$, 0.5 μM of each primer, 0.12 μg μl^{-1} of bovine serum albumen, and 2 μl of template. Each amplification reaction (25 or 50 μl) of the *cyt* b gene region consisted of 1X Promega buffer, 1.25 U of Enzypol Plus 2000 polymerase, 0.8 mM dNTPs, 2 mM $MgCl_2$, 0.5 μM of each primer, 1.0 M betaine, 0.12 μg μl^{-1} of bovine serum albumen, and 2–4 μl of template. The cycling conditions for all primer pairs consisted of 95° (1 min), 35–45 cycles (95, 30 s; 50, 45 s; 72, 1 min), with a final extension at 72(3 min). A template-free reaction was always included as a negative control. Amplicons were purified using sterile nanopure water and 30,000 MW Millipore filters (Millipore Inc., Bedford, MA, USA). The mass of the amplicons was determined by comparing EtBR staining intensity of 2.0–5.0 μl of each purified reaction relative to a standard mass DNA ladder (Invitrogen Life Technologies, Carlsbad, CA, USA). Cycle sequencing (Amersham ET-Ter-

Table 1 Collection location and GenBank accession numbers for samples used in this study

	Location	Accession Number [a]	
		d-loop	*cyt* b
Mola spp.			
Atlantic	Great Britain; 49°30′30″N, 2°29′ W	AY940822	–[b]
	Italy; 44°20′N, 9°06′E	AY940821	AY940835
	Scotland; Firth of Forth	AY940823	–[b]
	South Africa; Capetown	AY940816	AY940834
	South Africa; Bantry Bay[c]	AY940826	AY940838
	South Africa; Algoa Bay[c]	AY940827	–[b]
	USA; Florida	AY940819	–[b]
Pacific	Australia; Jervis Bay[c]	AY940824	–[b]
	USA; California	AY940815	AY940832
		AY940817	–[b]
		AY940818	–[b]
		AY940814	–[b]
	Japan; Tomiyama	AY940820	AY940836
Masturus lanceolatus	USA; Florida Keys	AY940830	AY940841
	Indonesia; North Sulawesi, Manado	AY940831	–[b]
	Taiwan; 121°37.3′E, 24°2′N	AY940828	AY940839
		AY940829	AY940840
Ranzania laevis	USA; Hawaii	AY940825	AY940842
GenBank Sequences			
Mola mola[c]	–[d]	NC005836	NC005836
Masturus lanceolatus	–[d]	NC005837	NC005837
Sufflamen fraenatus	–[d]	–[b]	AP004456
Takifugu rubripes	–[d]	–[b]	NC004299
Stephanolepis cirrhifer	–[d]	–[b]	AP002952
Diodon holocanthus	–[d]	–[b]	AY267366
Ostracion cubicus	–[d]	–[b]	AY267357

[a] From National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov)
[b] Sequence data not included
[c] Presumed *M. ramsayi*, see text for details
[d] Collection site not reported

minator Kit, Amersham Biosciences Corp., Piscataway, USA) was conducted with approximately 100 ng of purified PCR product according to manufacturers specifications and fluorescently labeled products were visualized using an ABI 310 Genetic Analyzer (Applied Biosystems, Inc., Foster City, CA, USA). Both strands of the amplicon were sequenced, compared, and edited as needed using Sequencher (v4.1; Gene Codes Corp., Ann Arbor, MI, USA).

For the mtDNA *d-loop* region, GenBank sequences for one *M. mola* (accession number NC005836) and one *Ma. lanceolatus* (NC005837) were included in the analysis. GenBank sequences for the *cyt* b gene from one of each of these species and other Tetraodontiformes were included in the analyses (accession numbers are indicated in Table 1). DNA sequences were aligned using Clustal X (Higgins et al. 1996) and the open reading frame for the *cyt* b gene region was identified to aid alignment.

Phylogenetic relationships were inferred using Bayesian analysis, as implemented in MrBayes (v3.0b4; Ronquist and Huelsenbeck 2003), and maximum likelihood (ML) and maximum parsimony, as implemented in PAUP* (v4.0b10; Swofford 1998). For the Bayesian analysis, the data were partitioned into four sets corresponding to the *d-loop* gene and putative 1st, 2nd, and 3rd codon positions of *cyt* b. Parameters of the models of evolution such as proportion of invariant sites, gamma distribution, and nucleotide substitution rates were estimated from the data using MrBayes. Each gene was analyzed separately and in a combined analysis. Five million generations were run and sampled every 1,000 generations. Three runs starting from random trees were used to determine if independent chains had converged after an average burn-in of 3,000 generations. For the ML analysis, the model of evolution was determined for the *d-loop* and *cyt* b gene region separately and as a combined dataset using the Akaike Information Criteria within Modeltest (v3.06; Posada and Crandall 1998). Maximum parsimony was conducted on the *cyt* b dataset only. Heuristic tree searches were conducted on ten trees generated from a stepwise addition of randomly chosen taxa using the tree-bisection-reconnection (TBR) branch swapping method for both the ML and MP analyses. In all analyses, gaps were treated as missing data. Node support for the maximum likelihood analyses was estimated from 100 bootstrap replicates under a fast stepwise heuristic method option in PAUP*. Node support for the maximum parsimony analysis was estimated from 1,000 bootstrap replicates under a full heuristic search option in PAUP*. Estimates of genetic divergence distances were generated using an ML estimator with PAUP*.

Results

Control region (*d-loop*)

The data consisted of an aligned region of 384 nucleotides from 14 *Mola* sp., five *Ma. lanceolatus*, and one *R. laevis* (Table 1). To maximize sequence similarity

among all Molidae sequences, the alignment required 21 indels ranging in size from 1 to 36 bp. Nucleotide frequencies as estimated by Modeltest were strongly A-T biased: $\pi_A = 0.3840$, $\pi_T = 0.3182$, $\pi_G = 0.1309$ and $\pi_C = 0.1669$. Under the Akaike Information Criteria (AIC), the best-fit model of evolution and the one used in ML analyses was the general time reversible plus gamma model (GTR + G) with a shape parameter of 0.6846. Reflecting the large number of polymorphisms within the *d-loop*, the proportion of invariable sites indicated by Modeltest was zero.

The average ML divergence ranged from $d = 0.023 \pm 0.012$ (within the genus *Masturus*) to $d = 1.287 \pm 0.108$ (for the genera *Masturus* versus *Ranzania*; Table 2). The average ML divergences within specimens initially identified as *M. mola* appeared to fall into two distinct groupings each with small, within-group, pairwise divergences but larger between group values (Fig. 1). We tentatively designated the divergent samples taken from the southern hemisphere ($n = 5$) as *M. ramsayi* (sensu Giglioli 1883). Average within-species divergence was 0.063 ± 0.033, 0.136 ± 0.090, and 0.023 ± 0.012 for *M. mola*, *M. ramsayi*, and *Masturus lanceolatus*, respectively (Table 2). For both *M. mola* and *M. ramsayi* there were significantly lower levels of within versus between ocean divergence ($t = -14.97$, $P = 0.00$ and $t = -17.218$, $P = 0.00$, respectively) with an average between ocean ML distance for each of the species of 0.090 ± 0.013 and 0.194 ± 0.012, respectively (Table 2). There is clear evolutionary divergence both within and between ocean basins as well as between species. ML distances between genera were considerably larger than those seen within species and ranged from 0.579 ± 0.046 (*Ma. lanceolatus* versus *M. ramsayi*) to 1.287 ± 0.108 (*R. laevis* versus *Ma. lanceolatus*; Table 2). In addition to a large number of nucleotide substitutional differences between the *Mola* spp., *M. ramsayi* differed from the other Molidae species by the presence of one or sometimes two copies of the 25-bp insertion ($[AT]_7$GTATTATCACC) approximately 8 bp from the ^{PRO}tRNA end of the control region. One individual from Australia, also designated as *M. ramsayi*, did not contain the 25-bp insertion and was similar to the GenBank *M. mola* sequence.

The Bayesian phylogram of the *d-loop* sequences generally was unresolved at the deeper nodes (Fig. 2), but clearly recovered four distinct clades corresponding to the southern hemisphere *Mola* sp. samples (putative *M. ramsayi*), *M. mola*, *Ma. lanceolatus*, and *R. laevis*. The three clades formed by the *Masturus* and *Mola* species were associated with large posterior probability values indicating strong support for these groups. Within the *M. mola* clade, there were two groups of samples with high probability values corresponding to individuals collected in the Pacific and Atlantic Ocean basins (Fig. 2). The ML analysis recovered an identical topology with large bootstrap values supporting the major clades (Fig. 2).

Cytochrome *b* region

To better assess the deeper level systematic relationships, a subset of the individuals was assayed for *cyt* b DNA sequence variation. A total of 741 nucleotides were examined from seven putative *M. mola*, two putative *M. ramsayi*, four *Ma. lanceolatus*, and one *R. laevis* (Table 1). Each was chosen to represent the major clades identified by the *d-loop* analysis (Fig. 2). Nucleotide frequencies estimated by Modeltest were $\pi_A = 0.2660$, $\pi_T = 0.2582$, $\pi_G = 0.1247$ and $\pi_C = 0.3510$. Under the AIC, the best-fit model was the TVM + I + G with a gamma distribution and shape parameter of 1.661. The assumed proportion of invariable sites was equal to 0.5054 and six substitution types and rates were estimated.

Within the Molidae, ML divergence estimates for *cyt* b were highly and significantly correlated with *d-loop* estimates ($P < 0.01$, $R^2 = 0.961$; Fig. 1). The per group *cyt* b ML divergence ranged from $d = 0.001 \pm 0.001$ (within *M. mola*) to 0.396 ± 0.004 (*Ma. lanceolatus* versus *R. laevis*; Table 2) and were, on average, four times smaller than *d-loop* values (excluding intraspecific values for *M. mola* where divergences were essentially zero; Table 2). Both Bayesian and ML analyses of the *cyt* b data recovered a topology identical to the *d-loop* analysis except, as expected for a more slowly evolving gene, there was less resolution for nodes leading to terminal

Table 2 Average divergence ($\bar{d}$), standard deviation (SD), and range of maximum likelihood divergence estimates for *d-loop* and *cyt* b regions. Best-fit models were chosen based on AIC in Modeltest (v3.06; Posada and Crandall 1998) and divergence estimates were generated with PAUP* (v4.0b10; Swofford 1998)

Comparison	*d-loop* (GTR + G)			*cyt* b (TVM + I + G)		
	$\bar{d}$	SD	Range	$\bar{d}$	SD	Range
Within *Mola mola*	0.063	0.033	0.012–0.114	0.001	0.001	0.000–0.002
Pacific vs. Atlantic	0.090	0.013	0.066–0.114	0.001	0.000	0.001–0.002
Within *Mola ramsayi*	0.136	0.090	0.013–0.206	0.031	–[a]	–[a]
Pacific vs. Atlantic	0.194	0.012	0.181–0.206	0.031	–[a]	–[a]
M. mola vs. *M. ramsayi*	0.298	0.002	0.054–0.058	0.056	0.002	0.054–0.058
Within *Masturus*	0.023	0.012	0.007–0.043	0.003	0.002	0.000–0.006
Masturus vs. *Mola mola*	0.598	0.041	0.501–0.685	0.144	0.005	0.139–0.150
Masturus vs. *Mola ramsayi*	0.579	0.046	0.484–0.642	0.140	0.006	0.128–0.145
Ranzania vs. *Mola* sp.	0.772	0.059	0.685–0.861	0.352	0.030	0.305–0.391
Ranzania vs. *Masturus*	1.287	0.108	1.176–1.401	0.396	0.004	0.391–0.401

[a] Only a single, pairwise comparison

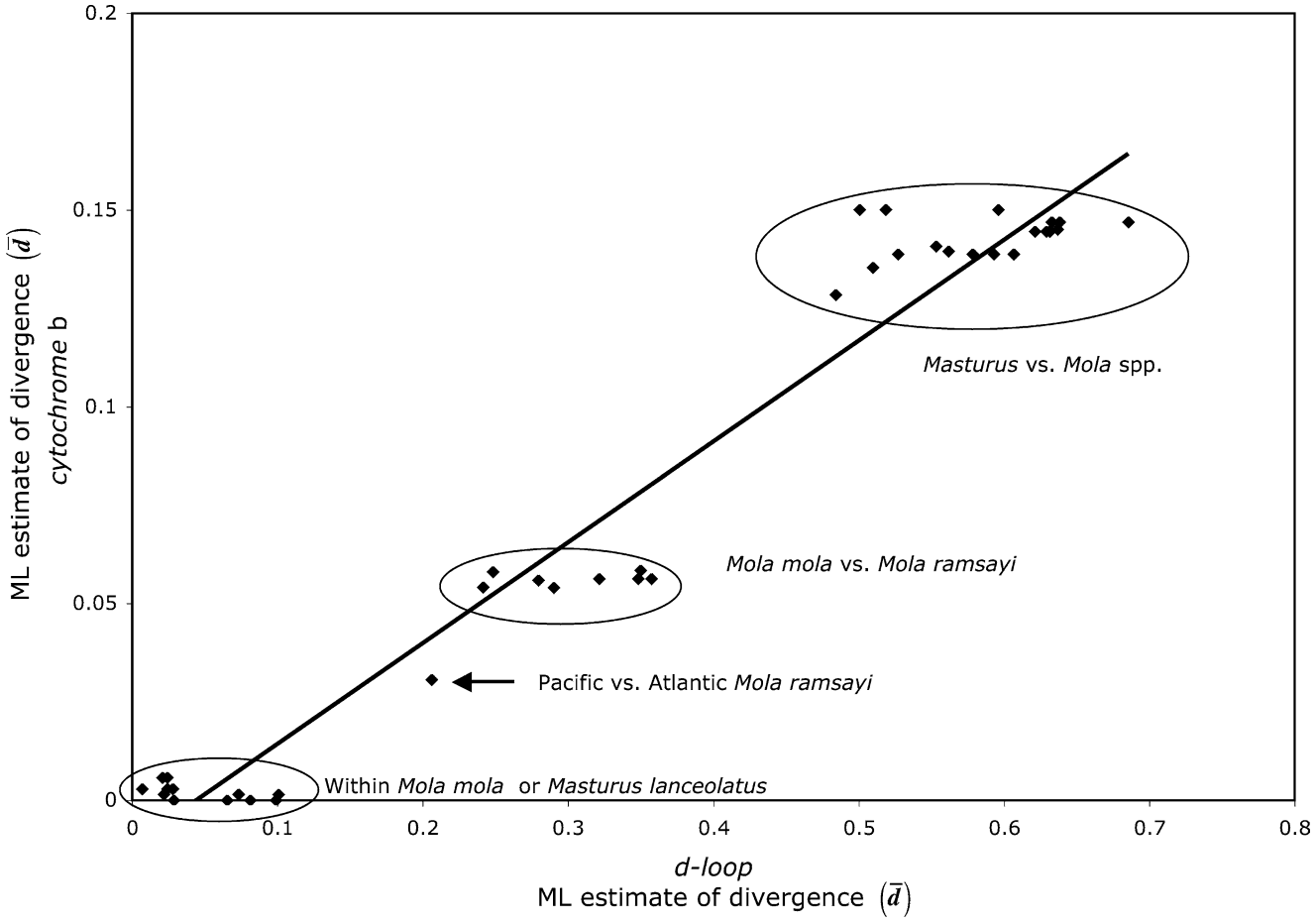

Fig. 1 Maximum likelihood estimates of divergence for the *d-loop* versus the *cyt* b gene region ($y = 0.2561\times-0.0113$; $R^2=0.961$). Clusters of data points correspond to taxonomic hierarchies

branches and more resolution at deeper nodes (tree not shown). While not strongly resolving the within ocean basin relationships, the *cyt* b data alone clearly supported clades consisting of *M. mola* (99 and 93 for posterior probability and bootstrap support, respectively; tree not shown), *M. ramsayi* (97 and 98), *Ma. lanceolatus* (100 and 98), and *R. laevis* as the basal member of the family (97 and 60). Some deeper level taxonomic aspects also were resolved in the *cyt* b analysis. There was strong support for Tetraodontidae (*Takifugu rubripes*) + Diodontidae (*Diodon holocanthus*) as the sister-group to the Molidae. Maximum parsimony analysis yielded an identical topology within the family Molidae with 100% support for terminal relationships (*R. laevis* + *Ma. lanceolatus* + (*M. mola* + *M. ramsayi*))), but failed to resolve deeper nodes.

Combined data analysis

We combined the *d-loop* and *cyt* b data and performed both a Bayesian and a ML analysis with the addition of other members of the suborder Tetraodontoidei sensu Santini and Tyler (2003) as outgroup taxa. Control region sequences were coded as missing data for non-Molidae taxa because the high level of polymorphism resulted in substitutional saturation and thus the sequences did not provide useful information. For the Bayesian analysis, the data were partitioned as previously described. For the ML analysis, the best fit model under AIC indicated by Modeltest was a TrN + I + G model with six substitution types. Nucleotide frequencies were estimated as $\pi_A=0.3084$, $\pi_T=0.2709$, $\pi_G=0.1184$, and $\pi_C=0.3022$. The proportion of invariable sites was 0.4118 and rate variation followed a gamma distribution with a shape equal to 0.8051. Both the ML and the Bayesian analyses resulted in identical topologies (Fig. 3). As with the separate analyses, two *M. mola*, two *M. ramsayi*, one *Ma. lanceolatus*, and the *R. laevis* clades were recovered. All non-terminal nodes except the deepest ones were strongly supported with high posterior probabilities and bootstrap values.

Discussion

Systematics and evolution of the suborder Tetraodontoidei

Our initial research focused on global population subdivision within *M. mola*, but subsequently shifted to

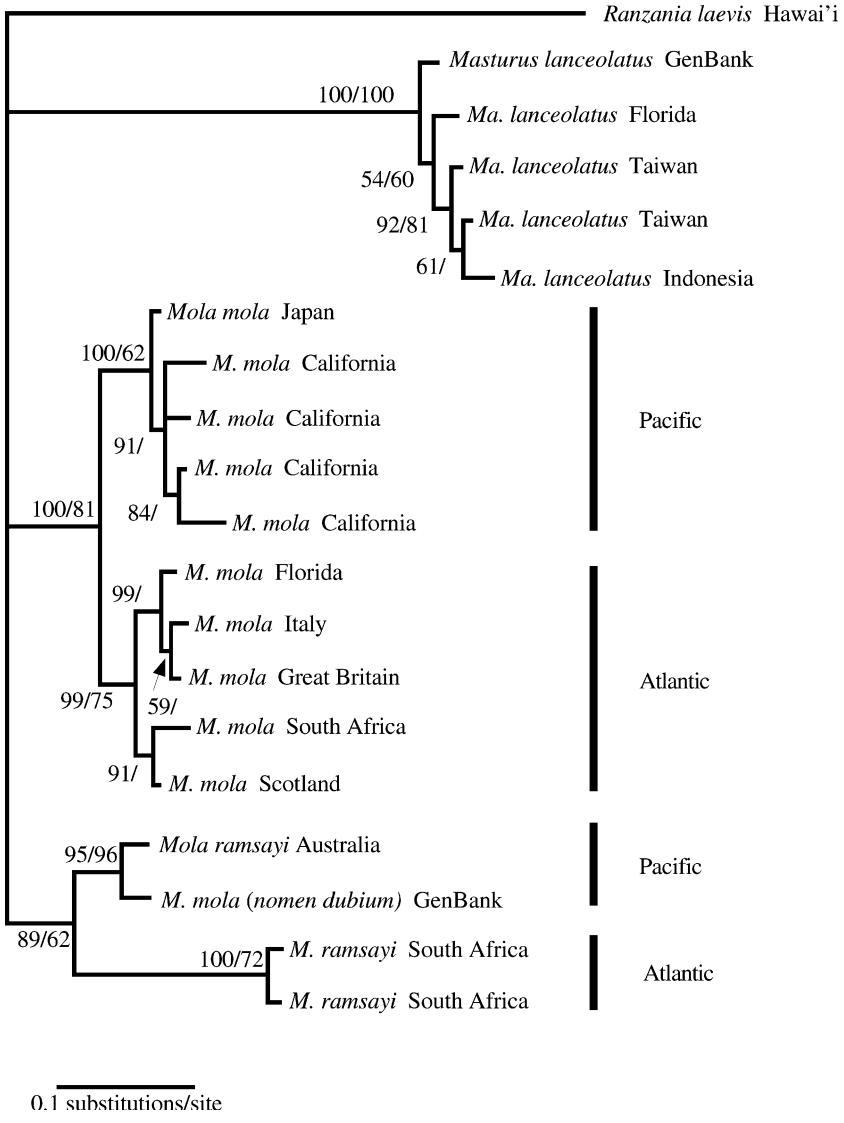

Fig. 2 Bayesian phylogram for *d-loop* data. Posterior probability values (*before slash*) and ML bootstrap values > 50 (*after slash*) are shown at nodes. When only one number is shown, the other value was ≤ 50

deeper taxonomic questions involving the unique and interesting family Molidae. The results of our broader phylogenetic assay, including members of all but one (Triodontidae) of the extant families in the suborder Tetraodontoidei, are consistent with accepted tetradontiform systematic relationships. The genera *Ranzania*, *Masturus*, and *Mola*, along with single representatives of the genera *Takifugu* and *Diodon*, formed a well-supported clade consistent with previous suggestions that Tetraodontidae + Diodontidae is the sister-group to the Molidae (Winterbottom 1974; Tyler 1980; Santini and Tyler 2003). The generic level distinctions within the Molidae are also well supported and the family is monophyletic. Also consistent with previous assessments, *Mola* and *Masturus* are sister taxa with *Ranzania* placed as the basal member of the family (Fraser-Brunner 1951; Santini and Tyler 2002; Yamanoue et al. 2004). Applying a rate of sequence divergence of 0.76% to 2.00% million years^{-1} (Zardoya and Ignacio 1999; Bowen et al. 2001; Dowling et al. 2002) to the *cyt* b data suggests that *Ranzania* diverged from *Masturus* approximately 19.8–52.1 million years ago (mya) and from *Mola* 17.6–46.3 mya. The age of *Ranzania* could place the origin of the Molidae near the time of origin of the Tetraodontiformes (~65 mya; Santini and Tyler 2003). This lends further support to the assertion by Santini and Tyler (2003) that most of the diversification into extant tetraodontids occurred prior to 35 mya. Unfortunately, the Molidae has one of the least known fossil records of any tetraodontiform family. Nonetheless, the *cyt* b dating is consistent with known fossils of *Ranzania* sp. (35 mya) and the putative, extinct sister-group *Eomola* (50 mya; Tyler and Bannikov 1992; Santini and Tyler 2003). Most fossil molids from the Miocene and Pliocene (5–24 mya) are morphologically nearly identical to recent genera (Santini and Tyler 2002) and these dates correspond to our estimates of divergence. The *Masturus* and *Mola* split is more recent and appears to have occurred 7.2–18.95 mya. This corresponds fairly well with the timing of the Terminal

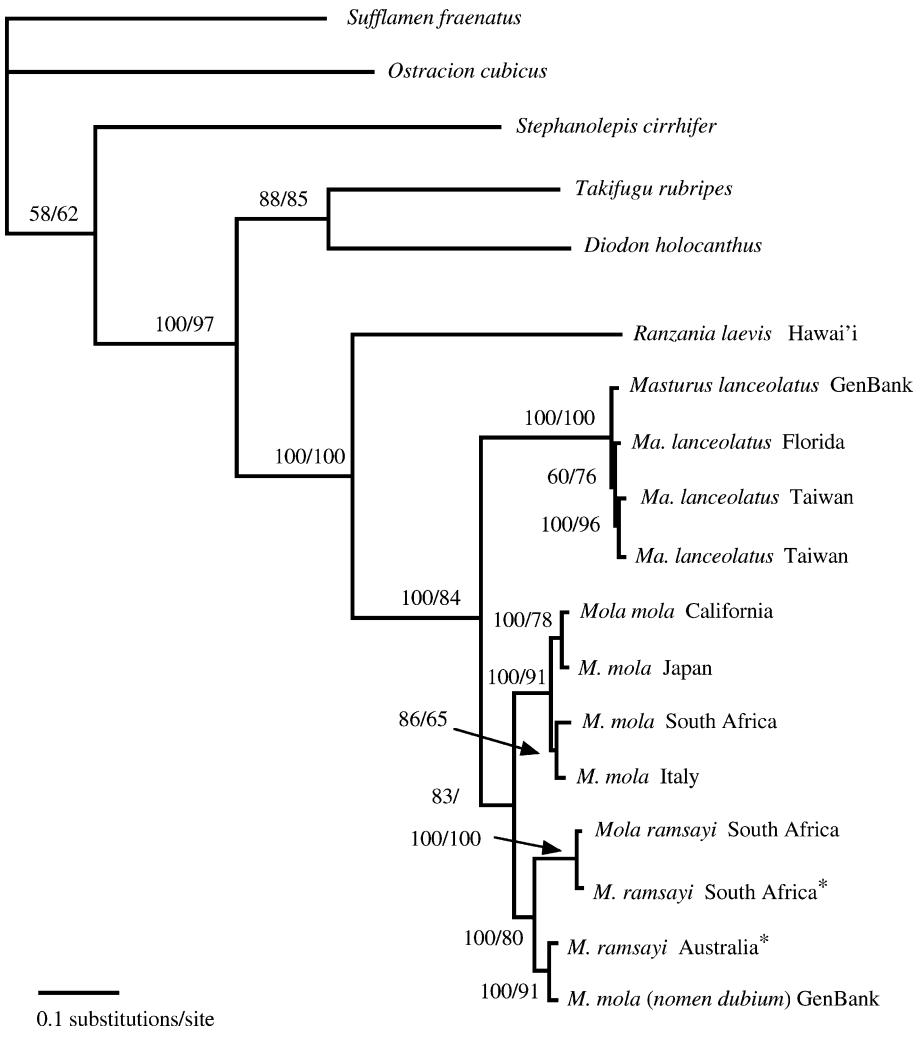

Fig. 3 Bayesian phylogram based on combined *d-loop* and *cyt b* data. Only *d-loop* data were available for taxa marked with asterisks and only *cyt b* sequences were available for non-Molidae species. Posterior probability values (*before slash*) and ML bootstrap values > 50 (*after slash*) are shown at nodes. When only one number is shown, the other value was ≤ 50

Tethyan event (12 –18 mya), but how this may have played a role in the diversity of a globally distributed, pelagic species is unclear.

A southern hemisphere species?

It is often quite challenging to assemble even a basic understanding of the biology of a large pelagic animal such as the ocean sunfish. Here, we began a population genetics study only to discover unexpectedly large degrees of genetic divergence among specimens of a single putative species, *M. mola*. Further examination of the degree and pattern of divergence among all members of the family led us to conclude that some of the individuals collected off the coasts of Australia and South Africa (i.e., southern hemisphere) and the sequence in GenBank (unspecified collection location) are not *M. mola* but probably are a previously described but commonly overlooked sister species, *M. ramsayi* (Giglioli 1883). Several lines of data support this conclusion. First, in comparing the DNA divergence estimates within and between species there clearly are three distinct clusters of points (Fig. 1). The cluster with the least pairwise divergence includes all intra-specific *Ma. lanceolatus* and some of the intra-specific *Mola* sp. comparisons. The group with the most divergent pairwise distances includes all inter-generic comparisons (i.e., *Ma. lanceolatus* versus *Mola* spp.). The remaining cluster, which is intermediate in divergence, includes comparisons between some of the southern hemisphere *Mola* spp. samples (i.e., *M. ramsayi*) with other individuals (i.e., *M. mola*) from throughout the range. Clearly, the individuals that were initially field-identified as *M. mola* are divided into two groups not as distantly separated as genera but more divergent than conspecifics; the magnitude of divergence is consistent with the species level of other teleosts (Johns and Avise 1998). As such, we believe that some of the individuals from Australia and South Africa are *M. ramsayi*. Although considerably more geographic sampling needs to be done before a firm conclusion can be reached, it appears that *M. mola* is globally distributed while *M. ramsayi* may, as previously considered, be a southern hemisphere species. The sequence in GenBank, identified as *M. mola*, should also be considered *M. ramsayi*.

Bayesian and ML analyses indicate strong support for two *Mola* clades, both clearly separate from an equally strongly supported *Masturus* clade (Fig. 2). The Bayesian and ML analyses of the combined *d-loop* and *cyt* b data also strongly support a sister taxon relationship for the two *Mola* sp. clades. One clade is composed exclusively of individuals collected in the southern hemisphere whereas the other includes individuals from both hemispheres. The earliest reference to a southern hemisphere species of ocean sunfish was by Giglioli (1883). He reported that *Orthragoriscus ramsayi* differed morphologically from *M. mola* in the shape, size, and form of the caudal rays and that *O. ramsayi* was covered with small horny scales. Fraser-Brunner (1951) concluded that the genus *Mola* was represented by two species with *M. ramsayi* differing from *M. mola* mainly in having more fin rays in the clavus, larger bony ossicles at the distal ends of these rays, and no band of smaller scales along the base of the clavus. Fraser-Brunner (1951) referred to the *M. mola* as a wide-ranging species that was largely or entirely replaced by *M. ramsayi* in the South Pacific. Parenti (2003) listed *Orthagoriscus eurypterus* Philippi 1892 as a synonym of *M. ramsayi* and Chile as the type location, but notes that no type specimen is known for the fish mentioned by either Giglioli or Philippi. Smith and Heemstra (1986) included the southern sunfish, *M. ramsayi*, and stated that it is known only from New Zealand, Australia, and South Africa. Our molecular data are entirely consistent with these observations with the exception that *M. mola* apparently is not absent from the southern hemisphere (Gomon et al. 1994). The data indicate that both species occur sympatrically in the south Atlantic at least around South Africa. The species distributions in the south Pacific are unclear given that we have only a single sample from this region. Unfortunately, we do not have voucher specimens for any of the putative *M. ramsayi* individuals and are unable to report the morphological differences described by Fraser-Brunner (1951). Nonetheless, we believe that the genetic data clearly indicate a species level distinction within the genus *Mola* and that some of the individuals from the southern hemisphere are most appropriately designated *M. ramsayi*.

It is interesting to note that trained professionals (fisheries scientists, museum curators, and fishermen) who collected many of the tissue samples for this study were asked to target *M. mola*. Samples were field identified as *M. mola* usually with unfettered access to the entire fish. The consistent identification of these animals as *M. mola* suggests that the large degree of genetic divergence we uncovered between *M. mola* and *M. ramsayi* is not reflected strongly in the overall morphology. The external differences noted by Giglioli (1883) were limited to the shape, size and form of the caudal rays and the presence of horny scales on the dermis, and field personnel were unable to distinguish the species.

Phylogeography of the molids

Comparing the timing of divergence to paleogeography can provide some insight into factors contributing to speciation. Using the previously described molecular clock conversions, the split between *Mola mola* and *M. ramsayi* would have occurred 2.8–7.5 mya, close to the estimated time of glacial maxima changes in the Pleistocene (~2 mya; Hallam 1994). Interestingly, the estimated dates for the divergence between the Atlantic and Indo-Pacific individuals within the two *Mola* species are different for *M. mola* (0.05–0.32 mya) than *M. ramsayi* (1.55–4.10 mya) and may indicate different proximate causes. As the divergences are of recent origin, the isolating mechanism for *M. mola* is unclear. The timing of the *M. ramsayi* Atlantic and Indo-Pacific divergence, however, coincides with the formation of the Isthmus of Panama (3.1–3.5 mya; Coates and Obando 1996) although given that *M. ramsayi* is believed to be a southern hemisphere species, this may be an unlikely explanation. The only other connection between the Indo-Pacific and Atlantic basins at this time was around the southern tip of Africa. Approximately 2.5 mya, however, southern ocean circulation and climate changes resulted in the establishment of cold-water upwelling in South Africa and the establishment of opposing current patterns (Shannon 1985). At present, the warm southwestern flowing Indian Ocean Agulhas current meets the cold southeastern Atlantic Benguela system at the Cape of Good Hope (Gordon 1985; Lutjeharms and Gordon 1987) separating the two ocean basins. Given their distribution and the age of the separation of the Atlantic and Indo-Pacific *M. ramsayi*, the soft-barrier isolation of currents and temperature at the tip of Africa may be a better explanation than the hard-barrier of the Isthmus of Panama. This allopatric isolation has also been proposed to explain a similar timing of speciation in a reef fish species (Bowen et al. 2001). It is important to remember, however, that the estimated time of divergence here is based only on two individuals and may be considerably older. In addition, although the tip of Africa is generally considered a barrier for tropical taxa, several examples of a close relationship between eastern Atlantic and western Indo-Pacific species have been documented (see Roberts et al. 2004).

Although our sample sizes are small, they provide insights at an intraspecific or population level for *M. mola*, *M. ramsayi*, and *Ma. lanceolatus*. On a global scale, both *Mola* species appear to be subdivided into Atlantic-Mediterranean and Indo-Pacific Ocean clades (Figs. 1 and 2). The inter-ocean divergence is even more pronounced for *M. ramsayi* than for *M. mola*. Although the global pattern for *Ma. lanceolatus* is less clear, the average intra-specific divergence for *Ma. lanceolatus* is 2.7 and 5.9 times smaller than seen in *M. mola* and *M. ramsayi*, respectively (Table 2). Regardless of whether the GenBank sample of *Ma. lanceolatus* was from the Atlantic or Pacific, the phylogenetic analysis of the

d-loop does not indicate clear separation between the ocean basins for this species. The number of *Ma. lanceolatus* samples analyzed, however, is probably too small to reveal subtle divisions, if they exist. Clearly, *Mola* spp. are subdivided on a global scale and any view of them as passively drifting members of the plankton needs revision. The genetic data also are consistent with our satellite tagging data indicating only regional movement of individuals on the scale of 100s of km over periods of 4–6 months (unpublished data). If molas, in general, are limited to local areas it is possible that several, separate populations may exist even within an ocean basin, making it likely that direct or indirect overexploitation could lead to local depletion or extirpation. An accurate understanding of the biogeographic structure will require larger sample sizes and more complete geographic sampling.

Acknowledgements We thank the major contributors to this study including those who collected tissue samples: D. Adams, C. Bartels, L. Benson, C. Brown, D. Christie, M. Farquhar, G. Farwell, P. Garratt, L. Greene, R. Horn, S. Huang, R. Lord, M. de Maine, A. Mariot, M. McGrouther, J. Seeto, M. S. Shie, I. Smith, G. Swinney, M. Tringali, J. Wissema, H. Y. Young and the staff from Kamogawa Sea World, Two Oceans Aquarium, Taiwan Fisheries Research Institute, and Capricorn Fisheries. Thank you to A. Castro, C. Curtis, K. Hayes, C. Puchulutegui and T. Schwartz who provided both lab and data analysis assistance. P. Motta and J. Tyler provided hard-to-find references. J. Grassle and two anonymous reviewers provided many helpful comments on the text. All experiments complied with current laws of the country from which samples were collected and were approved under USF IACUC permit number 1972. This research was supported by grants from The National Geographic Committee for Research and Exploration, the American Association for the Advancement of Science and the National Science Foundation Women's International Science Collaboration Program, Smithsonian Institution, Arcadia Wildlife Preserve Inc., and National Science Foundation grants DEB 9806905 and DEB 0321924.

References

Breder CM Jr, Clark E (1947) A contribution to the visceral anatomy, development, and relationships of the Plectrognathi. Am Mus Nat His 88:287–319

Bowen BW, Bass AL, Rocha LA, Grant WS, Robertson DR (2001) Phylogeography of the trumpetfishes (*Aulostomus*): ring species complex on a global scale. Evolution 55:1029–1039

Cartamil DP, Lowe CG (2004) Diel movement patterns of ocean sunfish *Mola mola* off southern California. Mar Ecol Prog Ser 266:245–253

Coates AG, Obando JA (1996) The geological evolution of the Central American isthmus. In: Jackson JBC, Budd AF, Coates AG (eds) Evolution and the environment in Tropical America. University of Chicago Press, Chicago, pp 21–56

Dowling TE, Tibbets CA, Minckley WL, Smith GR (2002) Evolutionary relationships of the plagopterine (Teleostei: Cyprinidae) from cytochrome *b* sequences. Copeia 2002:665–678

Fraser-Brunner A (1951) The ocean sunfishes (family Molidae). Bull Br Mus 1:89–121

Giglioli HH (1883) Zoology at the fisheries exhibition. Nature 28:313–316

Gomon MF, Glover JCM, Kuiter RH (1994) The fishes of Australia's south coast. The Flora and Fauna of South Australia Handbooks Committee, State Print, Adelaide

Gordon AL (1985) Indian-Atlantic transfer of thermocline water at Agulhas retroflection. Science 227:1030–1033

Grassi GB (1897) The reproduction and metamorphosis of the common eel (*Anguilla vulgaris*). Proc Royal Soc Lond 60:262–271

Hallam A (1994) An outline of Phanerozoic biogeography. Oxford University Press, New York

Higgins DG, Thompson JD, Gibson TJ (1996) Using CLUSTAL for multiple sequence alignments. Methods Enzymol 266:383–402

Johns GC, Avise JC (1998) A comparative summary of genetic distances in the vertebrates from the mitochondrial cytochrome *b* gene. Mol Biol Evol 15:1481–1490

Kocher TD, Stepien CA (1997) Molecular systematics of fishes. Academic Press, New York

Lee WJ, Conroy J, Howell WH, Kocher TD (1995) Structure and evolution of teleost mitochondrial control regions. J Mol Evol 41:54–66

Leis JM (1984) Tetraodontiformes: Relationships. In: Moser HG, Richards WJ, Cohen DM, Fahay MP, Kendall AW Jr, Richardson SL (eds) Ontogeny and systematics of fishes. Special publication 1. American Society of Ichthyologists and Herpetologists, Kansas, pp 459–463

Lutjeharms JRE, Gordon AL (1987) Shedding of an Agulhas ring observed at sea. Nature 325:138–140

Nelson JS (1994) Fishes of the World. 3rd edn. Wiley, New York

Norman JR, Fraser FC (1938) Giant fishes, whales and dolphins. WW Norton, New York

Norman JR, Fraser FC (1949) Field Book of Giant Fishes. GP Putnam and Sons, New York

Parenti P (2003) Family Molidae Bonaparte 1832–molas or ocean sunfishes. Calif Acad Sci Annotated Checklists of Fishes No. 18

Parin NV (1968) Ichthyofauna of the epipelagic zone. Academy of Sciences of the USSR, Institute of Oceanography. Translated from Russian (by M. Ravah. Edited by H. Mills). Jerusalem, Israel Program for Scientific Translations. U.S. Dept. of Commerce, Clearinghouse for Federal Scientific and Technical Information, Springfield, Va. 1970

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817–818

Roberts MA, Schwartz TS, Karl SA (2004) Global population genetic structure and male-mediated gene flow in the green sea turtle (*Chelonia mydas*): analysis of microsatellite loci. Genetics 166:1857–1870

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574

Santini F, Tyler JC (2002) Phylogeny of the ocean sunfishes (Molidae, Tetraodontiformes), a highly derived group of teleost fishes. Ital J Zool 69:37–43

Santini F, Tyler JC (2003) A phylogeny of the families of fossil and extant tetraodontiform fishes (Acanthomorpha, Tetraodontiformes), Upper Cretaceous to recent. Zool J Linn Soc 139:565–617

Schmidt J (1921) New studies of sun-fishes made during the "Dana" Expedition, 1920. Nature 107:76–79

Shannon LV (1985) The Benguela ecosystem. I. Evolution of the Benguela physical features and processes. Oceanogr Mar Biol Annu Rev 23:105–182

Smith MM, Heemstra PC (1986) Smith's Sea Fishes, 6th edn. Springer, Berlin Heidelberg New York

Song CB, Near TJ, Page LM (1998) Phylogenetic relations among Percid fishes as inferred from mitochondrial cytochrome *b* DNA sequence data. Mol Phylogenet Evol 10:343–353

Swofford DL (1998) PAUP* 4.0: phylogenetic analysis using parsimony and other methods. Sinauer Associates, Massachusetts

Taberlet P, Meyer A, Bouvet J (1992) Unusually large mitochondrial variation in populations of the blue tit, *Parus caeruleus*. Mol Ecol 1:27–36

Tyler JC (1980) Osteology, phylogeny, and higher classification of the fishes of the order Plectognathi (Tetraodontiformes). NOAA Tech Report, NMFS Circ 434:1–422

Tyler JC, Bannikov AF (1992) New genus of primitive ocean sunfish with separate premaxillae from the Eocene of southwest Russia (Molidae, Tetraodontiformes). Copeia 4:1014–1023

Walsh HE, Metzger DA, Higuchi R (1991) Chelix medium for simple extraction of DNA for PCR-based typing from forensic material. BioTechniques 10:506–513

Winterbottom AS (1974) The familial phylogeny of the Tetraodontiformes (Acanthopterygii: Pisces) as evidenced by their comparative myology. Smithsonian Contrib Zool 155:1–201

Yamanoue Y, Miya M, Matsuura K, Katoh M, Sakai H, Nishida M (2004) Mitochondrial genomes and phylogeny of the ocean sunfishes (Tetraodontiformes: Molidae). Ichthyol Res 51:269–273

Zardoya R, Ignacio D (1999) Molecular evidence on the evolutionary and biogeographical patterns of European cyprinids. J Mol Evol 49:227–237