Molecular Phylogenetics and Evolution 48 (2008) 258–269

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

New insights into the organization and evolution of vertebrate IRBP genes and utility of IRBP gene sequences for the phylogenetic study of the Acanthomorpha (Actinopterygii: Teleostei)

Agnès Dettaï *, Guillaume Lecointre

UMR 7138 "Systématique, Adaptation, Evolution", Département Systématique et Evolution, CP26, Muséum National d'Histoire Naturelle, 57 rue Cuvier, Case Postale 26, 75231 Paris Cedex 05, France

ARTICLE INFO

Article history:
Received 20 November 2007
Revised 1 April 2008
Accepted 1 April 2008
Available online 11 April 2008

Keywords:
Molecular phylogeny
Teleostei
Acanthomorpha
IRBP
Interphotoreceptor retinoid-binding protein
Petromyzon marinus

ABSTRACT

The interphotoreceptor retinoid-binding protein (IRBP) coding gene has been used with success for the large-scale phylogeny of mammals. However, its phylogenetic worth had not been explored in Actinopterygians. We explored the evolution of the structure of the gene and compared the structure predicted from known sequences with that of a basal vertebrate lineage, the sea lamprey *Petromyzon marinus*. This sequence is described here for the first time. The structure made up of four tandem repeats (or modules) arranged in a single gene, as present in Chondrichthyes (sharks and rays) and tetrapods, is also present in sea lamprey. In teleosts, one to two paralogous copies of IRBP gene have been identified depending on the genomes. When the sequences from all modules for a wide sampling of vertebrates are compared and analyzed, all sequences previously assigned to a particular module appear to be clustered together, suggesting that the divergence among modules is older than the split between lampreys and other vertebrates. Finally, 92 acanthomorph teleosts were sequenced for the partial module 1 of the gene 2 (713 bp) to assess for the first time the use of this marker for the systematic studies of the Teleostei. The partial sequence is slightly more variable than other markers currently used for this group, and the resulting trees from our sequences recover most of the clades described in the recent molecular multi-marker studies of the Acanthomorpha. We recommend the use of partial sequences from the IRBP gene 2 as a marker for phylogenetic inference in teleosts.

1. Introduction

Teleosts represent approximately one-half of extant vertebrate species. Sixty percent of them, representing 16,908 species, (Froese and Pauly, 2006) are acanthomorphs. This clade is supported by both morphological (Rosen, 1973; Johnson and Patterson, 1993; Stiassny, 1986) and molecular studies (Miya et al., 2003), and includes all fishes with true spines in the dorsal and anal fins. To date, many parts of the interrelationships of the 309 families (Nelson, 2006) remain poorly known. Recent molecular results, while presenting support for some groups defined with morphological characters (for example Tetraodontiformes and Atherinomorpha), also severely challenged several of the accepted hypotheses of acanthomorph relationships (Johnson and Patterson, 1993). Some larger groups where monophyly was poorly supported by morphological data and generally controversial (Stiassny et al., 2004) were not recovered. For instance, the Paracanthopterygii (Greenwood et al., 1966, see for instance the critics by Rosen, 1985), the extanded Zeiformes (Rosen, 1973, 1984; Johnson and Patterson, 1993; Tyler and Santini, 2005), the Scorpaeniformes (see Imamura and Shinohara, 1998 among others), and the Smegmamorpha (Johnson and Patterson, 1993) were absent in the trees inferred in the various molecular studies. In these studies, independent teams working on different markers and with different taxonomic samplings repeatedly recovered several new clades (Chen et al., 2003, 2007; Wiley et al., 2000; Miya et al., 2001, 2003, 2005; Dettaï and Lecointre, 2004, 2005; Holcroft, 2004; Smith and Wheeler, 2004, 2006; Smith and Craig, 2007). Nonetheless, additional work is needed for many parts of the acanthomorph tree, where currently available markers have not permitted the resolution of several nodes.

The power of multi-locus approaches for the resolution of troublesome phylogenetic inferences has been demonstrated on several large-scale groups in the last few years, including mammals (for instance Springer et al., 2001; Madsen et al., 2001), squamates (Vidal and Hedges, 2005) and ray-finned fishes (Chen et al., 2004; Kikugawa et al., 2004; Hurley et al., 2006). While mitochondrial genomes provide valuable information when analyzed as a whole

* Corresponding author. Fax: +33 01 40 79 38 44.
E-mail address: adettai@mnhn.fr (A. Dettaï).

1055-7903/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2008.04.003

(see for instance Inoue et al., 2003; Lavoue et al., 2005; Miya et al., 2003 or Jondeung et al., 2007), mitochondrial genes taken separately might perform poorly for inferring ancient divergences because of a marked saturation at least at the third codon position of protein-coding genes (Orti and Meyer, 1997; Springer et al., 2001). Moreover, mitochondrial genes cannot be considered to evolve completely independently from one another, and additional nuclear datasets are much needed to provide an independent data source.

The partial sequence of the nuclear gene encoding the interphotoreceptor retinoid-binding protein (IRBP) has been used with success for the phylogeny of mammals, alongside numerous other markers, and within families as well as among orders. Stanhope et al. (1992) were the first to use these sequences for phylogenetic inference.

In tetrapods, the gene contains four tandem repeats (called modules hereafter), each approximately 900 bp long. The first three modules and part of the fourth form one large exon, while the last part of the fourth module contains three introns (Fig. 1). The coding sequence for teleost IRBP was first described in *Carassius auratus* (Wagenhorst et al., 1995), followed closely by the expression assays and description in another cyprinid, the zebrafish *Danio rerio* (Rajendran et al., 1996). This hypothesized structure, based on genomic sequences and mRNAs, was composed of only two modules, as opposed to the four modules present in mammals (*Danio rerio* gene 2 in Fig. 1). The second of these modules contains three introns (Rajendran et al., 1996). But very recently, sequence data mining of complete genome databases for a number of model fishes (cyprinid *Danio rerio* and acanthomorphs *Tetraodon nigroviridis, Takifugu rubripes, Gasterosteus aculeatus* and *Oryzias latipes*) led to question this description (Nickerson et al., 2006). A second IRBP locus was detected at a location preceding the first identified locus in some of the species for which complete genomes were available. In *G. aculeatus* and *O. latipes*, no copy of the new locus was found, and in *T. nigroviridis*, only a very partial sequence was present. The new locus was named IRBP gene 1, and the previously identified locus was named gene 2. These names are the ones used in the present study. But the study of Nickerson et al. (2006) was based on a small number of teleost taxa that were available in GenBank and/or complete genomic databases. Here, we explored more thoroughly the presence of this additional copy of the IRBP gene in other species of acanthomorphs, especially focusing on groups that are closely related to the species where no copy or only a partial sequence of gene 1 could be found in the available complete genomes. We also performed a search for the IRBP coding gene in taxa where it had not been described previously, and could identify IRBP-like sequences in the lamprey *Petromyzon marinus*. The new sequences allowed us to reconstruct a tree comparing the modules with one another in a wider range of taxa to study the evolution of the structure of the gene.

In parallel, we examined the phylogenetic usefulness of the IRBP gene 2 partial first module sequences for a wide sampling of acanthomorph species (Table 1). As will be discussed, the inferred clades are largely congruent with previously published molecular but also morphological results for the group. This module can bring valuable information for the reconstruction of the phylogeny of acanthomorph fishes, even with sequences as short as 713 bp.

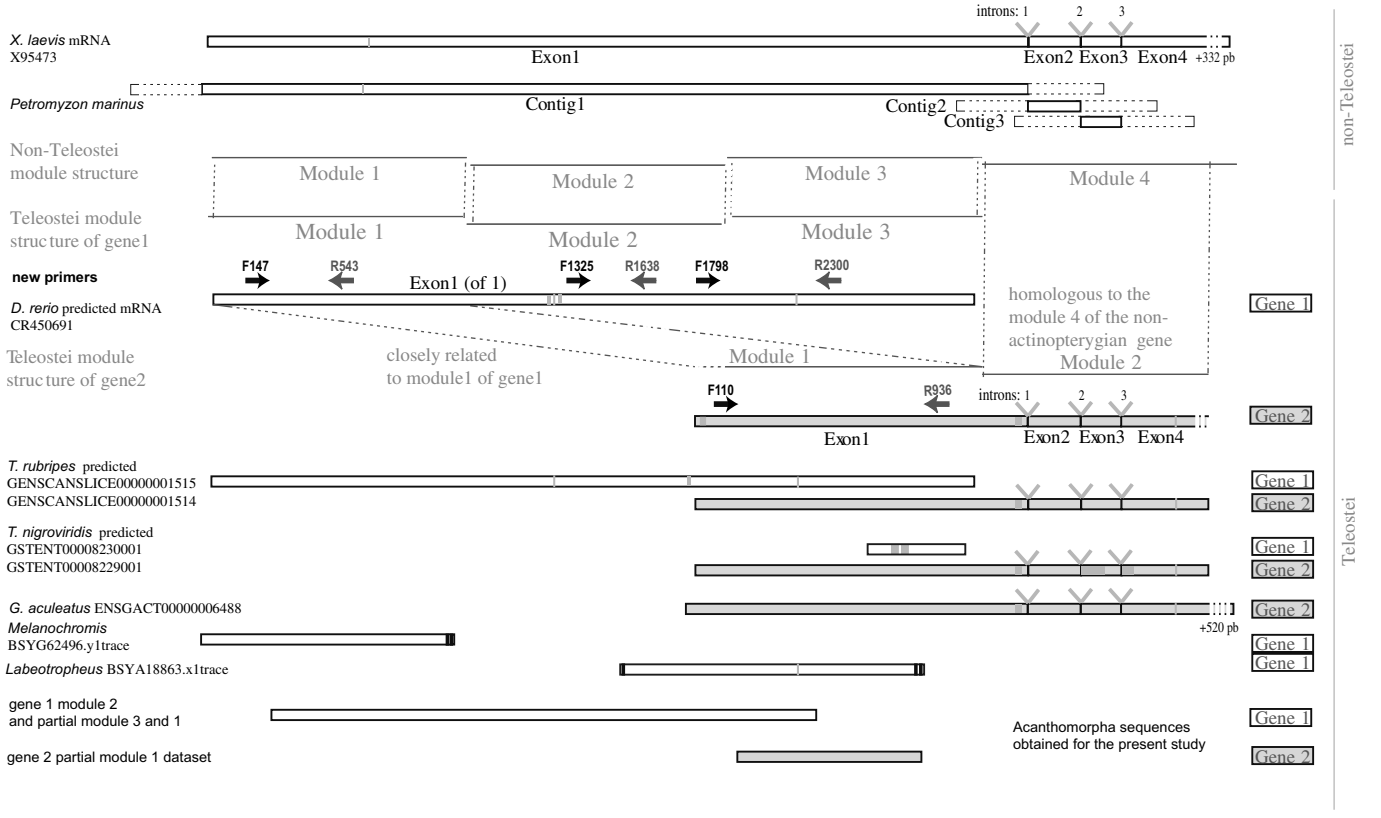

Fig. 1. Representation of the alignment of the full-length tetrapod coding sequence with the coding sequences found in Teleostei and *Petromyzon marinus*. Deletions are represented by grey areas in the affected sequence. The oligonucleotides primers used in the present study are represented by arrows. Sequences of the sarcopterygian single gene, and the teleost gene one are represented by white rectangles, sequences of teleost gene 2 by light-grey rectangles. The position of the sequences obtained for this study is indicated at the bottom of the figure. The module 1 of the actinopterygian gene 2 is the one tested here for acanthomorph phylogeny.

Table 1
Taxonomic sampling, GenBank CoreSequence and trace archive accession numbers, and Ensembl scaffold numbers of the sequences

Outgroups	Order/sub-order	Family	Genus and species	Gene 2 module 1	Gene 1	Single IRBP-coding gene	
Cypriniformes		Cyprinidae	*Danio rerio*	X85957	CR450691	Accession numbers of the trace files used to reconstruct the IRBP coding gene of *Petromyzon marinus*	
			Carassius auratus	X80802		*Petromyzon marinus*	
Chlorophthalmoidei		Ipnopidae	*Bathypterois dubius*	DQ168042		Contig-containing exon1	
Lampridiformes		Lampridae	*Lampris immaculatus*	DQ168077		1229775045	PMAC-aqw52f02.b1
		Regalecidae	*Regalecus glesne*	DQ168109		1431135039	PMAC-ayu28b06.b1
Polymixiiformes		Polymixiidae	*Polymixia nobilis*	DQ168104		1293713991	PMAC-atn37c06.b1
Paracanthopterygii						1180179031	PMAC-adz75g08.b1
	Batrachoidiformes	Batrachoidae	*Halobatrachus didactylus*	DQ168069		1188732452	PMAC-ahu86e05.b1
	Gadiformes	Gadidae	*Gadus morhua*	DQ168066		1212949869	PMAC-amk23b03.b1
			Gaidropsarus sp.	DQ168067		1484755358	PMAC-bri37g11.b1
		Macrouridae	*Trachyrincus murrayi*	DQ168124		1277028634	PMAC-arg55c05.b1
		Moridae	*Mora moro*	DQ168089		1201242823	PMAC-alm40b05.b1
	Percopsiformes	Aphredoderidae	*Aphredoderus sayanus*	DQ168038		1212919321	PMAC-amg34b01.b1
	Lophiiformes	Ceratiidae	*Ceratias holboelli*	DQ168049		1267811279	PMAC-arn35e12.b1
		Antennariidae	*Antennarius striatus*	DQ168037		1468863562	PMAC-bmg41h07.b1
Zeiformes	Zeioidei	Zeidae	*Zeus faber*	DQ168128		1377120471	PMAC-aut89b08.g1
			Zenopsis conchifer	DQ168127		1255810366	PMAC-air87c05.b1
Beryciformes	Trachichthyoidei	Anomalopidae	*Photoblepharon palpebratus*	DQ168101		1434632667	PMAC-azz26c03.b1
		Diretmidae	*Diretmoides* sp.	DQ168060		1442721743	PMAC-bek61c09.g1
	Berycoidei	Berycidae	*Beryx splendens*	DQ168045	EU625596	1218957550	PMAC-aoh60g03.b1
	Holocentroidei	Holocentridae	*Myripristis botche*	DQ168091		1188594531	PMAC-aeg12c03.g1
Stephanoberyciformes		Barbourisiidae	*Barbourisia rufa*	DQ168041		1197355648	PMAC-aih29h11.b1
		Rondeletiidae	*Rondeletia* sp.	DQ168110		1179789762	PMAC-ady93d05.g1
Mugiloidei		Mugilidae	*Liza* sp.	DQ168082		1423496401	PMAC-axv81d05.b1
Atherinomorpha	Bedotioidei	Bedotiidae	*Bedotia geayi*	DQ168043		1263847863	PMAC-arl14g07.g1
	Belonoidei	Belonidae	*Belone belone*	DQ168044		1205833228	PMAC-alm14c11.b1
	Adrianichthyidae	Adrianichthyidae	*Oryzias latipes*	DQ168094		1485092142	PMAC-brd77e12.b1
	Cyprinodontoidei	Poeciliidae	*Poecilia reticulata*	DQ168102		1180171186	PMAC-ado79e04.b1
Gasterosteriformes	Syngnathoidei	Aulostomidae	*Aulostomus chinensis*	DQ168040		1377461023	PMAC-ava13e06.g1
		Macroramphosidae	*Macroramphosus scolopax*	DQ168083		1229775045	PMAC-aqw52f02.b1
		Syngnathidae	*Syngnathus typhle*	DQ168120		1424824779	PMAC-ayt22f05.g1
			Nerophis ophiodon	DQ168071		1179693405	PMAC-acq51g01.b1
Synbranchiformes	Synbranchoidei	Synbranchidae	*Monopterus albus*	DQ168088		Contig containing exon 2	
	Mastacembeloidei	Mastacembelidae	*Mastacembelus erythrotaenia*	DQ168084		1255405503	PMAC-air87c05.g1
Dactylopteriformes		Dactylopteridae	*Dactylopterus volitans*	DQ168059		1181182584	PMAC-aff28g06.g1
Scorpaeniformes	Scorpaenoidei	Scorpaenidae	*Scorpaena onaria*	DQ168114		1201200882	PMAC-alm40b05.g1
		Triglidae	*Chelidonichthys lucerna*	DQ168053	EU625594	1211948050	PMAC-amg34b01.g1
	Cottoidei	Cottidae	*Taurulus bubalis*	DQ168121		1442361700	PMAC-bek61c09.b1
		Liparidae	*Liparis fabricii*	DQ168081		1175509040	PMAC-aem29g08.b1
Tetraodontiformes	Tetraodontoidei	Tetraodontidae	*Lagocephalus laevigatus*	DQ168076		1464380256	PMAC-bkr18h10.b1
			Tetraodon mbu		EU625595	Contig containing exon 3	
			Tetraodon nigroviridis	SCAF10211	Very partial	1470057826	PMAC-bms02f01.b1
			Takifugu rubripes	CAAB01007779	yes	1489683570	PMAC-brk26a07.b1
		Ostraciidae	*Ostracion* sp.	DQ168095		1179941313	PMAC-ado79e04.g1
		Molidae	*Mola mola*	DQ168087		1200373563	PMAC-alb69h12.b1
		Triacanthodidae	*Triacanthodes* sp.	DQ168125		1422318367	PMAC-aye45a04.b1
Pleuronectiformes	Psettodoidei	Psettodidae	*Psettodes belcheri*	DQ168108	EU625593	1213523859	PMAC-aoh60g03.g1
	Pleuronectoidei	Citharidae	*Citharus linguatula*	DQ168055		1184647640	PMAC-aeg12c03.b1
		Soleidae	*Microchirus variegatus*	DQ168086			
			Solea vulgaris	DQ168117		Lissamphibia	
		Pleuronectidae	*Syacium micrurum*	DQ168119		*Xenopus laevis*	X95473
Elassomatoidei		Elassomatidae	*Elassoma zonatus*	DQ168063		Aves	
Perciformes	Caproidei	Caproidae	*Capros aper*	DQ168048		*Gallus gallus*	AY994153
	Percoidei	Serranidae	*Serranus accraensis*	DQ168115		Mammalia	
			Holanthias chrysostictus	DQ168073		*Mus musculus*	NM_015745
			Epinephelus aeneus	DQ168064		*Homo sapiens*	BC039844
			Pogonoperca punctata	DQ168103		*Canis familiaris*	XM_546201
			Rypticus saponaceus	DQ168111		*Macaca mulatta*	ENSMMUT00000016682
		Centropomidae	*Lates calcarifer*	DQ168075			
		Moronidae	*Lateolabrax japonicus*	DQ168078			
		Percidae	*Perca fluviatilis*	DQ168099			
			Gymnocephalus cernuus	DQ168068			
		Chaetodontidae	*Chaetodon semilarvatus*	DQ168050			
		Drepanidae	*Drepane africana*	DQ168061			
		Pomacanthidae	*Holacanthus ciliaris*	DQ168072			
		Haemulidae	*Pomadasys perotaei*	DQ168105			
		Mullidae	*Mullus surmuletus*	DQ168090			
		Menidae	*Mene maculata*	DQ168085			
		Polynemidae	*Pentanemus quinquarius*	DQ168098			
	Carangoidei	Carangidae	*Chloroscombrus chrysurus*	DQ168054*			
			Trachinotus ovatus	DQ168120*			
		Coryphaenidae	*Coryphaena hippurus*	DQ168056			
		Echeneidae	*Echeneis naucrates*	DQ168062			

Table 1 (*continued*)

Outgroups	Order/sub-order	Family	Genus and species	Gene 2 module 1	Gene 1	Single IRBP-coding gene
	Acanthuroidei	Acanthuridae	*Ctenochaetus striatus*	DQ168057		
		Siganidae	*Siganus vulpinus*	DQ168116		
	Labroidei	Labridae	*Labrus bergylta*	DQ168075		
		Scaridae	*Scarus hoefleri*	DQ168112		
		Cichlidae	*Haplochromis nubilus*	DQ168070	EU625591	
			Labeotropheus fuellebornii		gnl|ti|1368495966 BSYA18863.x1*	
			Melanochromis auratus		gnl|ti|1370326225 BSYG62496.y1*	
	Zoarcoidei	Pholidae	*Pholis gunnellus*	DQ168100		
	Notothenioidei	Bovichtidae	*Bovichtus variegatus*	DQ168046		
		Nototheniidae	*Notothenia coriiceps*	DQ168093	EU625597*	
			Trematomus bernacchii			
		Channichthyidae	*Neopagetopsis ionah*	DQ168092		
	Trachinoidei	Trachinidae	*Trachinus draco*	DQ168123		
		Uranoscopidae	*Uranoscopus albesca*	DQ168126	EU625592	
		Cheimarrichthyidae	*Cheimarrichthys fosteri*	DQ168052		
		Chiasmodontidae	*Kali macrura*	DQ168074		
	Blennioidei	Blenniidae	*Parablennius gattorugine*	DQ168097		
		Tripterygiidae	*Forsterygion lapillum*	DQ168065		
	Gobiesocoidei	Gobiesocidae	*Lepadogaster lepadogaster*	DQ168080		
			Apletodon dentatus	DQ168039		
	Callionymoidei	Callionymidae	*Callionymus lyra*	DQ168047		
	Gobioidei	Gobiidae	*Pomatoschistus* sp.	DQ168106		
	Scombroidei	Sphyraenidae	*Sphyraena sphyraena*	DQ168118		
		Scombridae	*Scomber japonicus*	DQ168113		
	Stromateoidei	Stromateidae	*Pampus argenteus*	DQ168096		
		Centrolophidae	*Psenopsis anomala*	DQ168107		
	Channoidei	Channidae	*Channa striata*	DQ168051		
	Anabantoidei	Anabantidae	*Ctenopoma* sp.	DQ168058		

New sequences obtained in this study are indicated in bold. Classification follows Nelson (1994) except for the Caproidei (Johnson and Patterson, 1993). A star (*) indicates sequences that are not full-length.

2. Material and methods

2.1. PCR and sequencing

2.1.1. Amplification and sequencing of the partial teleost IRBP gene 1

The RNA sequences for the IRBP gene 2 described in Wagenhorst et al. (1995) and Rajendran et al. (1996) were retrieved from GenBank using the NCBI portal (http://www.ncbi.nlm.nih.gov/) and used to BLAST-search (Altschul et al., 1997) the complete and nearly complete genomes of *D. rerio*, *T. nigroviridis, T. rubripes, G. aculeatus* and *O. latipes* through the ENSEMBL (Hubbard et al., 2005) MultiBlastview Tool (http://www.ensembl.org/Multi/blastview, releases 40–42). We then used the sequence of gene 1 in the *D. rerio* complete genome to query all the other available teleost complete genomes using the ENSEMBL portal, and also the trace archives for all the available actinopterygians, as well as for available non-sarcopterygian and non-actinopterygian sequences using the Discontiguous Megablast tool of the NCBI portal. Sequences producing significant hits were retained, and used in turn to query the sequence databases. We assembled the trace sequences for *P. marinus* using SEQUENCHER (Gene Codes Corporation). All these sequences were aligned by hand using BioEdit (Hall, 1999), and used to design primers for the partial gene 1 for teleosts (see Table 2), using Primer3 (Rozen and Skaletsky, 2000; http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). Samples listed in Table 1 were conserved in 70% ethanol and extracted following the protocol of Winnpennminck et al. (1993). Sequence-specific amplifications were performed by PCR in a final 50-μL volume containing 5% DMSO, 300 μM of each dNTP, 0.3 μM of *Taq* DNA polymerase (Qiagen) and 5 μL of 10× CoralLoad buffer (Qiagen) and 1.7 pM of each of the two primers (see Table 2 for the primers), with 0.1–0.3 μg of DNA added depending on species. After denaturation for 2 min at 94 °C, the PCR ran for 45 cycles of (20 s, 94 °C; 30 s, 58 °C; 5 min, 72 °C), with a terminal elongation of 10 min. The result was visualized on ethidium bromide-stained agarose gels. The PCR products were sequenced through commercial contract (Genoscreen, Lille, France). Each sequence was obtained at least twice and checked against its chromatograms in BioEdit (Hall, 1999). We controled for potential contaminations and mix-ups by pairwise sequence comparison. All new sequences were deposited in GenBank (accession numbers listed in Table 1).

2.1.2. Amplification and sequencing of the partial teleost IRBP gene 2

The sequences of IRBP gene 2 recovered in the genomic databases of *D. rerio*, *T. nigroviridis*, *T. rubripes*, *G. aculeatus* and *O. latipes* were used as references to design oligonucleotide primers (Table 2) using the same computer programs as for gene 1 for a fragment of approximately 800 bp of the first module of the gene 2. We amplified the partial gene 2 from genomic DNAs extracted with the same protocol as for gene 1, but for a much wider and more representative sampling within Acanthomorpha (see Table 1). Sequence-specific amplifications were performed by PCR in a final 25-μL volume containing the same products as before except for the polymerase and buffer: 0.3 μM of Qbiotaq (QBioGen), 2.5 μL of 10× buffer (QBioGen); 0.01–0.1 μg of DNA were added depending on species. After denaturation for 2 min, the PCR ran for 40 cycles of (30 s, 94 °C ; 30 s, 50 or 54 °C; 1 min, 72 °C), with a terminal elongation of 3 min. The result was visualized on ethidium bromide-stained agarose gels, and purified with the MinElute PCR Purification kit or the MinElute Gel purification kit (Qiagen). Sequencing was performed on a CEQ2000 sequencer (Beckman Coulter), version 4.3.9, with the products and according to the instructions of the manufacturers' kit. Sequence assembly and checking was as before (accession numbers in Table 1).

2.1.3. Alignment and analyses

Alignment was performed by hand under BioEdit (Hall, 1999). The partial gene 2 IRBP contains almost no gaps (a single 3-bp deletion for *Pentanemus quinquarius)* and therefore presented no alignment ambiguities. The alignment of the modules of gene 1 and 2 with one another, as well as with the modules of other vertebrates, was done using the sequences translated in amino-acids and aligned as guideline. We analyzed several datasets: partial nucleic acids sequence of the teleost IRBP gene 2 (this dataset will be referred to hereafter as partial gene 2), and the almost complete sequence of the modules for a large sampling of vertebrates, under both nucleic acids and amino acids form. Because there was an ambiguously aligned region for the two vertebrate module datasets, analyses were performed with and without this region for the amino acids dataset.

We evaluated saturation (Philippe and Douzery, 1994; Hassanin et al., 1998) for the partial gene 2 separately for transitions and transversions, and also separately for each codon position (Fig. 2). Mean pairwise differences among three pairs of taxa (*D. rerio* and *T. rubripes, T. rubripes* and *T. nigroviridis*, and *T. rubripes* and a cichlid) were calculated using PAUP*4.0b10 (Swofford, 2002) for all modules of the IRBP-coding genes, MLL4 (Dettaï and Lecointre, 2005), rhodopsin retrogene (Chen et al., 2003), RAG1 and eight markers described by Li et al. (2007).

Gene 2 was analyzed by maximum likelihood (ML) and Bayesian phylogenetic inference method (BA). Both AIC and BIC approaches as implemented in Modeltest 3.7 (Posada and Crandall, 1998; Posada and Buckley, 2004) were used to identify the level of complexity of the model of nucleotide substitution that best fit the gene 2 dataset. We tested the dataset first as a whole, and then tested 1st, 2nd and 3rd positions separately (Table 3). BA was used as implemented in MrBayes 3.1 (Huelsenbeck and Ronquist, 2001). GTR + I + Γ were set as model, with the defaults settings for the priors for the proportion of invariable sites and for the gamma shape parameter. The analyses were run, using the 1st, 2nd and 3rd codon positions as different partitions. Eight analyses were run with the following parameters: four chains, 8 million generations, sampling of every 80th tree and discarding of the first 40,000 trees after checking the burn-in zone was included in this interval. After checking convergence had been reached, the trees and parameters resulting from the eight analyses were pooled and combined in a consensus. The ML analysis was run with PhyML (Guindon and Gascuel, 2003) on the PhyML online web server (Guindon et al., 2005) with a GTR + I + Γ model and estimation of all the parameters.

Table 2
Primers used for the amplification and sequencing of the partial first module of the IRBP-coding gene 2, and partial gene 1 in various teleosts

	Primer name		5′–3′ sequences
Oligonucleotide primers for the 1st module of the IRBP coding gene 2			
Forward primers	IRBP	U104	ATA GTY NTG GAC AAN TAC TGC TC
	IRBP	U110	TGG ACA AYT ACT GCT CRC CAG A
Reverse primers	IRBP	L953	CNG GAA YYT GAR CAC GGA GG
	IRBP	L936	CAC GGA GGY TGA YNA TCT TGA T
	IRBP	L922	TGA TNN CRG TKG CRA GGG CAT Y
	IRBP	L916	GTG KCA AGG GCA TCT TCT GC
Oligonucleotide primers for the partial 1st, 2nd and 3rd module of gene 1			
Forward primers	IRBP	F147	AAA CTA CTG CWT CCC KGA GAA
	IRBP	F1325	ATA YCT GCG YTT KGA CAG GTT
	IRBP	F1798	CAA CAA CGG DGA RTD CTG GCT
Reverse primers	IRBP	R543	AVA GYY CTC CMG CTG TGC TG
	IRBP	R1638	GAC TGB ATY ARG TAA GCA AAC TC
	IRBP	R2300	TGT TGT ACC TCA TRT CRA TTA TGA G

All primers were designed for this study. See the position of the primers on the gene on Fig. 1.

We analyzed the module datasets using maximum parsimony only. Heuristic searches (TBR search, 1000 replicates, gaps considered as missing) were conducted with PAUP*4.0b10 (Swofford, 2002), and a 1000 bootstrap replicates were performed on the dataset.

3. Results and discussion

3.1. New IRBP sequences

While the size of the IRBP protein of the skate (Chondrichthyes) has been published (Sun and Ripps, 1992), its sequence is not known yet. Searches for chondrichthyan sequences using the whole gene 1 and gene 2 sequences for *D. rerio*, as well as with the sequences of the IRBP gene of *Macaca mulatta* and the *Xenopus laevis*, yielded no result, neither in the CoreSequence database nor in the trace archives of GenBank. Searches in the two available complete genomes of tunicates (*Ciona savignyi* and *Ciona intestinalis*), and in the currently available trace archives for *Oikopleura dioika* and *Branchiostoma floridae* were also unfruitful, regardless of the Blast-search settings. The absence of IRBP-like sequences in the available *Ciona* genomes had already been described by Nickerson et al. (2006).

3.2. Petromyzon marinus

However, similar searches in the *Petromyzon marinus* whole genome shotgun sequences (WGS) produced positive hits that could be assembled in three contigs. Using the available whole genome shotgun sequences recovered by Blast-search, it was possible to reconstruct a partial sequence of the IRBP gene for *P. marinus*. The first of these contigs is 4371 bp long, and can be aligned unambiguously with the first exon of the tetrapod IRBP gene from its beginning to its end (see Fig. 2). Comparisons of each of the modules of *Petromyzon* with all the modules of *Danio*, *Xenopus* and *Homo,* using Blast2Sequences (Tatusova and Madden, 1999) and the phylogeny of the modules (Fig. 3) corroborated this homology hypothesis. It has no stop codons along this aligned part, but the 9th codon after what corresponds to the end of the exon in tetrapods and teleosts is a stop codon. It also shares with all the tetrapods (except *Gallus*) a conserved six base pairs (ATGCAG) next to where the exon ends in these.

The second contig is 997 bp, and can be aligned unambiguously with the complete exon 2 of the single gene present in sarcopterygians, and the exon 2 of gene 2 of Teleostei (201 bp). There are stop codons 58 codons before the beginning of the alignment and 19 codons after its end.

The third contig is 1602 bp long, and can be aligned unambiguously with the complete exon 3 of the single gene present in Sarcopterygians, and the exon 3 of gene 2 of Teleostei (141 bp). There are stop codons 37 codons before the beginning of the alignment and 51 codons after its end.

Unfortunately, the contigs could not be assembled to one another, even using as query for the search the ends of the contigs. It can not yet be excluded that they are not located close to one another, or even are part of different genes in the *Petromyzon* genome. Blast searches on the newly available assembly of the *Petromyzon* genome brought no answer, as the coding sequence for exon 1 is surrounded by unknown sequences (Ns). The sequence corresponding to the fourth and last exon could not be recovered in *P. marinus*. But if they are indeed all part of a single gene, then the structure of the IRBP gene in *P. marinus* is similar to the one in Sarcopterygii and the one hypothesized in sharks and rays (see Fig. 2). The two gene structure detected in teleosts is derived, as already proposed in Rajendran et al. (1996) and in Nickerson et al. (2006).

The high conservation in length of the predicted exons of the IRBP gene, and of the position of the potential introns, led us to

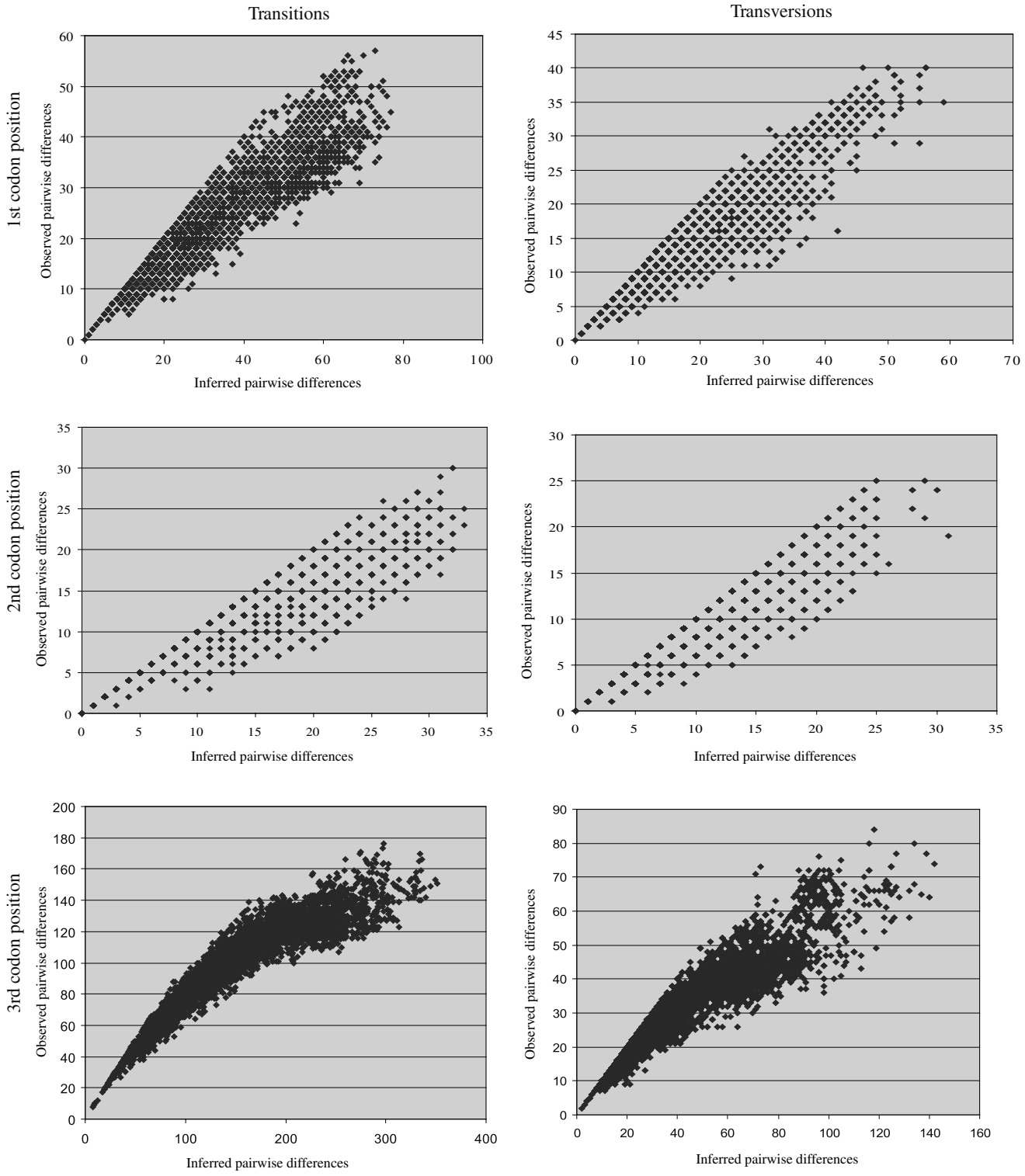

Fig. 2. Saturation plots for the partial module 4 of the IRBP coding gene. The observed pairwise differences were calculated with PAUP*, the inferred pairwise differences were obtained using the maximum parsimony tree.

concatenate the "coding" part of the three contigs in a single sequence for the phylogenetic analysis of the modules.

3.3. New actinopterygian gene 1 sequences

All available actinopterygian sequences, whether in the CoreSequence database or in the trace archives were Blast-searched for gene 1. Very few new sequences were recovered: only two trace files for *Labeotropheus fuellebornii* and *Melanochromis auratus*. We recovered the same results as Nickerson et al. (2006) as to the presence of gene 1 in the available complete genomes of teleosts: no sequences for *G. aculeatus* and *O. latipes*, and a very partial sequence for *T. nigroviridis* (see Fig. 2). Yet, *T. rubripes*, where a full-length gene 1 is present, is much more closely related to *Tetraodon* than *Gasterosteus* or *Oryzias*, raising some questions about the distribution of gene 1 in acanthomorphs. To explore this, a number of

Table 3
Parameters and models as selected by Modeltest 3.7 and as inferred in the BA with MrBayes 3.1

Estimated with Modeltest	AIC Model	Alpha *p*	Value of *I*	BIC Model	Alpha *p*.	Value of *I*	
Confidence level 0.01							
complete seq.	GTR+I+G	0.927	0.201	TrN+I+G	0.326	0.203	
1st codon pos.	TrN+G	0.543		TrN+G	0.543		
2nd codon pos.	GTR+G	0.369		GTR+G	0.369		
3rd codon pos.	GTR+G	2.939		TVM+G	2.937		
	Selected model	Alpha p.	Value of I	Freq. A	Freq. C	Freq. G	Freq T
Means estimated through Mr. Bayes 3.1							
1st codon pos.	GTR + I + G	0.732	0.117	0.374	0.218	0.266	0.266
2nd codon pos.	GTR + I + G	0.583	0.264	0.291	0.309	0.161	0.239
3rd codon pos.	GTR + I + G	2.897	0.007	0.166	0.359	0.257	0.217

Complete results of the analyses are available upon request.

species which are representative of the main clades of acanthomorphs as established in several molecular studies (Chen et al., 2003; Dettaï and Lecointre, 2004, 2005; Miya et al., 2003, 2005; Smith and Wheeler, 2006) were selected, especially focusing on species closely related to the species where no gene 1 could be detected in the available complete and almost complete genomic sequences. Amplification and sequencing of a large part of gene 1 was possible for specimens representing most clades (see Table 1) published in previous molecular studies (Chen et al., 2003; Miya et al., 2003, 2005; Dettaï and Lecointre, 2004, 2005; Smith and Wheeler, 2006; Smith and Craig, 2007; Chen et al., 2007). The hypothesis that the different structure of the IRBP coding genes in teleosts is specific to this group has been put forward by Nickerson et al. (2006). This remains to be tested by the exploration of the gene or genes in basal actinopterygians, where it has not been studied yet.

3.4. Analysis of the modules

Previous publications have identified the three dimensional structure of IRBP modules in mammals (Loew and Gonzalez-Fernandez, 2002). Comparison of structure and sequences helped identify a larger family of proteins with similarities to the modules of the gene. Among them are Tsp, a tail-specific protease degrading selectively proteins with nonpolar C termini (Silber et al., 1992), enoyl-coenzyme A hydratase (Engel et al., 1996), dienoyl coenzyme A isomerase (Modis et al., 1998), 4-chlorobenzoyl coenzyme A dehalogenase (Benning et al., 1996) and C-terminal processing protease (Liao et al., 2000). Unfortunately, the sequences identified as having similarities with IRBP were so distant that alignment was highly unreliable and therefore they were of limited use as outgroups. The trees were thus left unrooted.

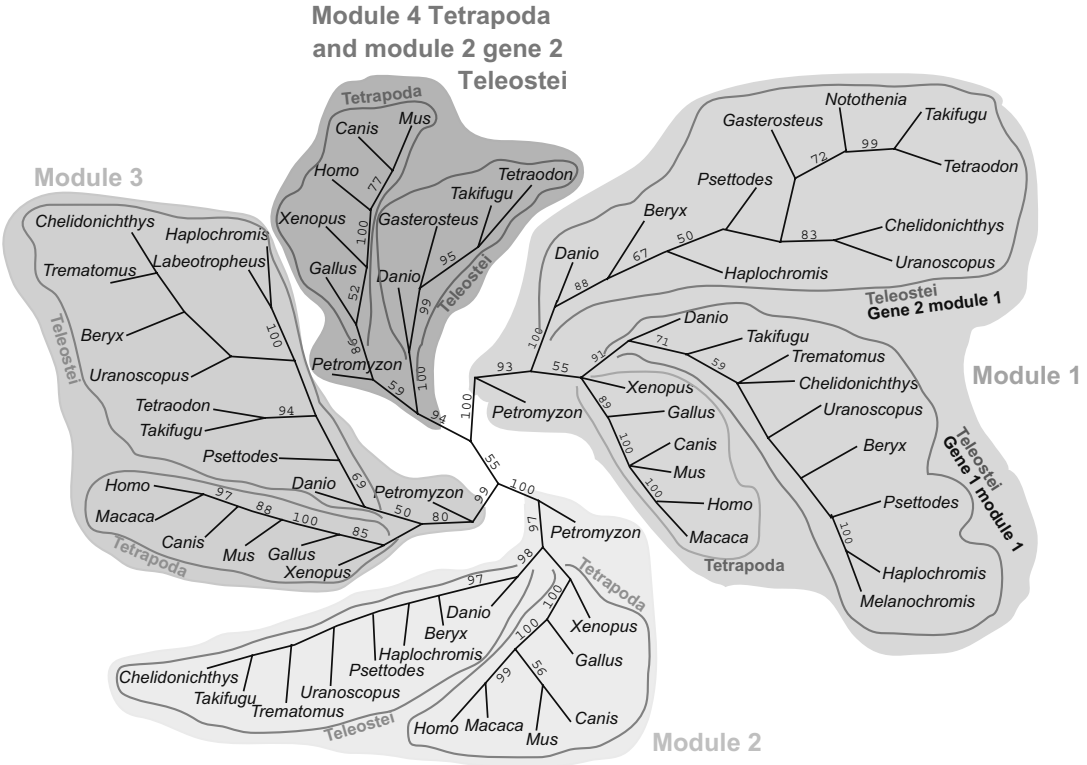

Fig. 3. Strict consensus of eight equiparsimonious trees of 3441 steps obtained through maximum parsimony analysis with PAUP* of the amino acid sequences of the modules of the IRBP coding genes for *Petromyzon marinus*, teleosts and tetrapods. The dataset comprises 68 sequences and is 324-aa-long. Values associated with the branches are bootstrap proportions obtained with 1000 bootstrap replicates. The tree is unrooted. The module names are given indicated according to the position of the sequence in the gene(s).

The unrooted tree depicting module interrelationships of vertebrate IRBP genes is shown in Fig. 3. The trees have the same general structure whether inferred with the amino-acid sequences or the nucleic acid sequences. All sequences from a given module cluster together in both cases (considering a hypothetic root on one of the five central branches), and tetrapods are separated from teleosts within each module except for the module 2 with inference from the nucleic acid sequences where *Gallus* and *Xenopus* are sister groups of the teleost clade. Very deep divergences are generally difficult to recover from a single molecular marker using unweighted parsimony, which is why multiple markers are generally used for this type of studies. In the present case, the success of our analyses in recovering basal gnathostome interrelationships within each module using no model corrections suggests that these sequences have suitable properties for deep phylogeny. For the trees inferred from amino acid sequences, there are high bootstrap supports for the basal branches of each module. *Petromyzon* module sequences are basal in all module clades except for the module 4/module 2 gene 2 clade, where it is a sister group of tetrapods, but with a very low bootstrap proportion. In all the other clades, the basal position is much better supported. This and the high conservation of the limits of the exons, leads to the conclusion that the four module structure of the IRBP gene is older than the lampreys/gnathostomes split. Beside in Osteichthyes, previous studies had only found IRBP in Chondrichthyes, and the gene was not sequenced. The molecular weight of the protein in this group pointed to a single, four-module protein (Sun and Ripps, 1992). The present result gives a considerably older age for the formation of the gene with its four-module structure.

The pattern within teleosts and tetrapods in each module/cluster is more variable, but mostly compatible with known relationships among vertebrates. Bootstrap supports are low except for the groupings of closely related taxa (Tetraodontidae, Primates, and Cichlidae).

3.5. Properties and structure of the IRBP-coding gene

Fig. 1 summarizes the previously published results (Rajendran et al., 1996; Nickerson et al., 2006), the results of the alignment and of the modules analyses. Module 1 of teleost IRBP gene 2 is

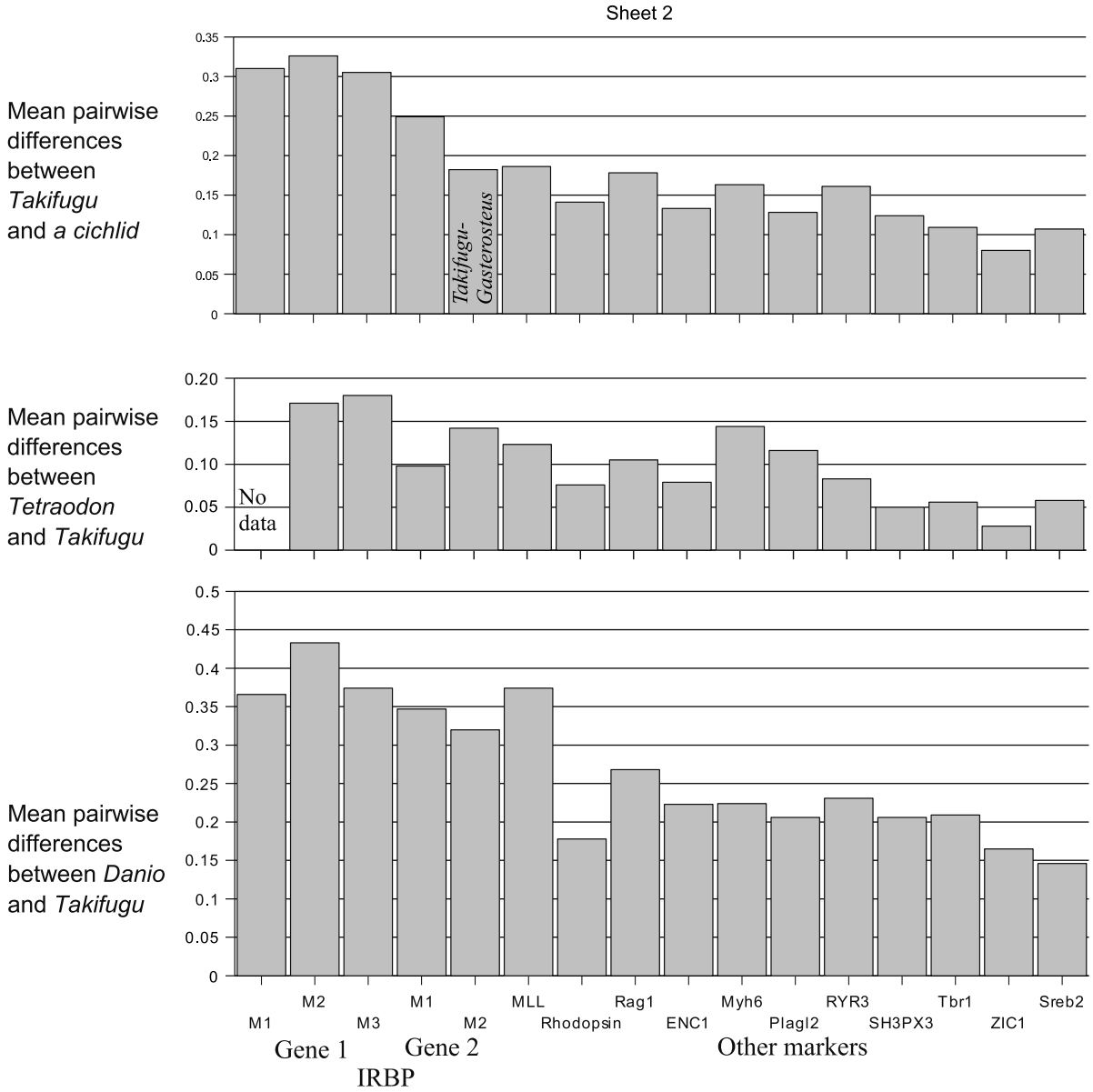

Fig. 4. Comparison of the mean pairwise differences for three pairs of taxa for the five modules of the IRBP coding gene and for 11 other nuclear markers used for (MLL, rhodopsin retrogene, RAG1) or proposed (Li et al., 2007) for use in acanthomorph phylogeny. The values were calculated with PAUP*.

closest to the clade which groups module 1 teleost IRBP gene 1 and module 1 of other vertebrate IRBP gene, while module 2 of the teleost IRBP gene 2 is closest to module 4 of other vertebrate IRBP gene. This corroborates the findings of Nickerson et al. (2006), and supports the hypothesis of a single-ancestral IRBP-coding gene made of four modules.

It is easy to distinguish the sequences belonging to each module in the alignment, no matter how deep the divergence, as there are module-specific patterns and deletions shared by all included taxa. The clear cut difference between module sequences, and the clustering of all the sequences of each module in clades makes their use as phylogenetic markers very interesting. Indeed, it is not possible to confuse the modules with one another, and no gene even remotely similar to IRBP was found in all the genomes searched.

It is especially noteworthy that it was possible to amplify and sequence partial IRBP gene 1 for samples selected from most of the main clades of acanthomorphs, as highlighted on Fig. 5 (Chen et al., 2003; Miya et al., 2003, 2005; Dettaï and Lecointre, 2004, 2005; Smith and Wheeler, 2006). This gene 1 might also be useful as a phylogenetic marker in the acanthomorph groups where it is present. But for larger scale studies, it would be necessary to further investigate the exact distribution among teleosts of the complete IRBP gene 1. In the absence of further data, it is not possible to say if unsuccessful amplification in a group is due to technical problems (non-adapted primers) or to the absence of the gene 1. It is therefore better to use, at least provisionally, the first module of gene 2 for phylogenetic purposes, as this module is present in all the groups tested in this study.

Comparison of mean pairwise differences among pairs of sequences within modules (Fig. 4) showed some variability among the modules. The differences were the smallest within module 1 and 2 of gene 2, and greatest within module 2 of gene 1 for most pairs of taxa. This is of interest if the gene is to be used for phylogenetic reconstruction, as some modules might be more interesting for reconstructing deeper divergences (module 1 of gene 2 for instance), while others might present enough variability for more recent events.

3.6. Properties of the gene 2 dataset for teleosts

The Acanthomorpha dataset for the first module of gene 2 is composed of 96 terminals, 713 positions, 471 positions informative for parsimony. Partial results for the model search performed with Modeltest 3.7 and *a posteriori* values of the model parameters estimated during the BA are given in Table 3. The models and parameters recommended by Modeltest are different for each codon position in this part of the gene. The BA also gave different end-values for the parameters and nucleotide frequencies for the three codon positions. This might explain the better recovery of previously defined clades by the BA than by the ML analysis. MrBayes 3.1 allows definition of different codon partitions with different models and parameters, but PhyML does not. In this case, the different properties of the three codon partitions call for a different treatment of each of these partitions, and therefore MrBayes 3.1 is probably better suited to the analysis.

The IRBP dataset only shows a beginning of saturation for the 3rd codon position transitions (see Fig. 2). A χ^2 heterogeneity test performed with PAUP on the nucleotidic composition heterogeneity shows a highly significant heterogeneity for 3rd codon positions, but not for 1st and 2nd. The GC content varies from 47% (for *Apletodon dentatus*) to 89.7% for *Gaidropsarus*, with an average of 65%.

But more significantly, the inferred trees, even from sequences as short as 713 bp, show a very high congruence with groups published on other markers on this group (Chen et al., 2003, 2007; Miya et al., 2003, 2005; Dettaï and Lecointre, 2004, 2005; Smith and Wheeler, 2006; Smith and Craig, 2007). Table 4 compares the clades present in three previous studies (Dettaï and Lecointre, 2005; Miya et al., 2005; Smith and Wheeler, 2006) using combined analyses of multiple genes and the clades present in the trees inferred from the first module of gene 2 of IRBP with BA and ML.

Table 4
Clades within the Acanthomorpha recovered by previous multigene combined analyses (Dettaï and Lecointre, 2005; Miya et al., 2005; Smith and Wheeler, 2006) and with the partial 1st module of the IRBP coding gene 2

	Dettaï and Lecointre (2005) Partial MLL, rhodopsin, 12S, 16S and 28S rDNA		Complete mitochondrion Miya et al. (2005)	Present study	
	Presence	Name		IRBP BA	IRBP ML
Zeioidei, Gadiformes	Yes	A	Yes	Yes	Yes
Mugiloidei, Atherinomorpha	Yes − 1	C	No	?	Yes
Blennioidei, Gobiesocoidei	Yes	D	Yes	Yes + 1	No
Cichlidae, Mugiloidei, Atherinomorpha, Blennioidei,Gobiesocoidei	Yes	Q	Yes*	Yes + 1	No
Syngnathoidei, Callionymoidei, Mullidae	Yes	E'	*	?	No
Aulostomidae, Macrorhamphosidae, Dactylopteridae	Yes	E	*	?	No
Stromateoidei, Scombridae, Chiasmodontidae	Yes	H	*	Yes	Yes − 1
Syngnathoidei, Callionymoidei, Mullidae, Stromateoidei, Scombridae, Chiasmodontidae	Yes	E+E'+H	*	?	Yes − *E'*
Channidae, Anabantoidei	Yes	f1	*	Yes	Yes
Channidae, Anabantoidei, Symbranchiformes	Yes	F	Yes*	Yes*	Yes
Uranoscopidae, Ammodytidae, Cheimarrichthyidae	Yes	G	*	Yes*	No
Cottoidei, Zoarcoidei	?	I	Yes + 1	Yes*	Yes
Cottoidei, Zoarcoidei, Gasterosteiformes, Triglidae	Yes	Isc	Yes + 1	Yes* +1	No
Notothenioidei, Percidae	Yes	K	*	No	No
Pleuronectiformes, Centropomidae, Carangidae, Menidae, Sphyraenidae, Polynemidae, Echeneidae	Yes	L	Yes*	Yes	Yes − 1
Tetraodontiformes, Lophiiformes, Caproidei, Elassomatidae, Acanthuridae, Siganidae, Pomacanthidae, Drepanidae, Chaetodontidae	Yes	N	Yes*	?	No
Labridae, Scaridae	Yes	M	*	Yes	Yes
Polymixiiformes, Zeioidei, Gadiformes	Yes	O	Yes + 2	Yes	Yes
Notothenioidei, Percidae, Triglidae, Trachinidae, Scorpaenidae, Serranidae	?	X	Yes*	Yes* − 2	Yes + 3
Zeioidei, Gadiformes, Percopsiformes, Lampridiformes, Polymixiiformes	Yes − 1		No	Yes	No

Presence of a clade is indicated by yes, irresolution in the tree by ?, incomplete answer due to incomplete taxonomic sampling by a *. Clades present, but with intruder taxon or missing taxon are indicated by Yes + 1 or Yes − 1. Clades M and Q are also recovered with the mitogenome in the more complete sampling of Mabuchi et al. (2007).

The BA analysis yields a tree that is considerably more congruent with other studies than the ML analysis. Out of 20 scored nodes, 10 are perfectly recovered (vs. 6 for ML), four are imperfectly recovered (vs. 4), and only one is contradicted (vs. 9). With BA, some groups that have proven hard to recover using other molecular datasets, like Tetraodontiformes, for instance, are recovered with high posterior probability. But overall, the posterior probabilities of larger recovered clades are low (Fig. 5). The analyses of IRBP recover several clades that had been first proposed based on morphological characters like the Atherinomorpha (Parenti, 1993) and their relationship with mugilids (Stiassny, 1993), or the relationship of Tetraodontiformes with siganids, acanthurids and chaetodontids (Winterbottom, 1974), although other groups also insert in this clade (Caproidei, Lophiiformes). Other, less consensual morphologically defined groups are not recovered here, as they were also not in previous molecular studies (Miya et al., 2003, 2005; Dettaï and Lecointre, 2004, 2005; Smith and Wheeler, 2006; Smith and Craig, 2007; Mabuchi et al., 2007): Scorpaeniformes, Smegmamorpha, Perciformes or

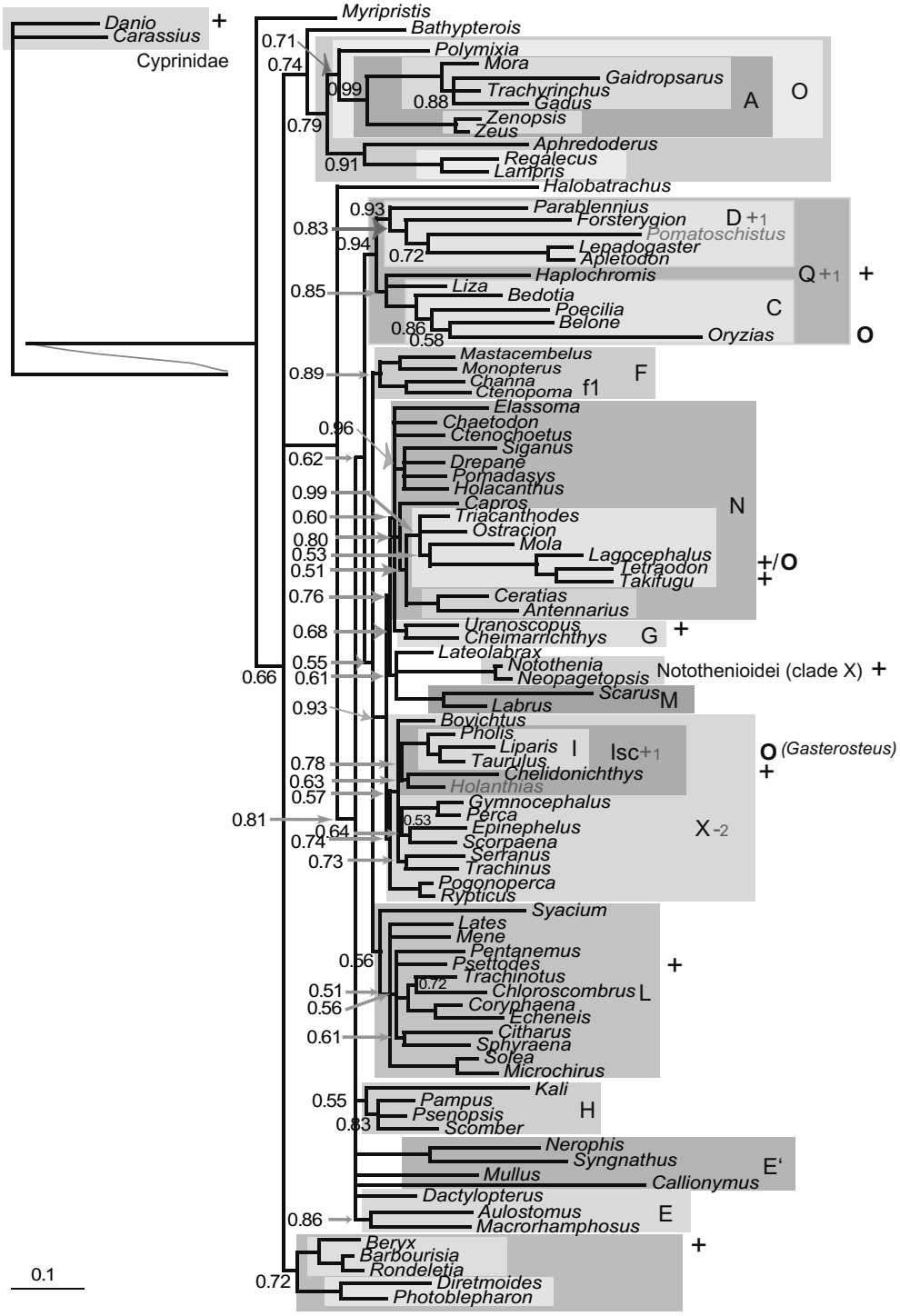

Fig. 5. Bayesian analysis majority rule consensus of the analysis of the partial module 1 of the second teleost IRBP coding gene for a wide sampling of Acanthomorpha. Values indicated next to the nodes are posterior probabilities. Nodes with no indication have a posterior probability of 1. Letters refer to the clades recovered by other studies (see Table 4). + Indicates groups where an almost complete first gene could be identified, O points species where the first gene could not be found in the complete genome available through Ensembl nor in the WGS sequences.

Paracanthopterygii, all already criticized in morphological studies (Imamura and Shinohara, 1998; Rosen, 1985). Labroids sensu Stiassny and Jensen (1987), an another doubtful group because of a possible high level of homoplasy in its proposed synapomorphies, is also polyphyletic as already proposed by Dettaï and Lecointre (2005) and Mabuchi et al. (2007). The cichlids group with the Atherinomorpha in clade Q, while the labrids and scarids form their own distinct clade. Finally, the sometimes surprising new clades proposed solely based on molecular studies (Chen et al., 2003, 2007; Miya et al., 2003, 2005; Dettaï and Lecointre, 2004, 2005; Smith and Wheeler, 2006) are recovered here (see Table 4).

When the divergence between chosen pairs of taxa is compared with other nuclear markers used in teleosts, the first module of the IRBP gene 2 (Fig. 4) has one of the highest mean of pairwise differences, well above the Rag1 and rhodopsin sequences, or most markers proposed by Li et al. (2007). It scores slightly below MLL, another marker (Dettaï and Lecointre, 2005) very efficient at recovering clades found by analyses of much larger datasets (Dettaï and Lecointre, 2005; Miya et al., 2003, 2005; Smith and Wheeler, 2004, 2006). In both IRBP and MLL, the differences are spread evenly along the sequences, a desirable property for a phylogenetic marker (Collins et al., 2005). Only one-third of all the codon positions of the IRBP module 1 of gene 2 dataset is conserved across our whole taxonomic sampling.

The exons of the IRBP coding gene are long, so it is not necessary to include an intron to have a sequence long enough for phylogenetic purposes. This limits the risk of amplification problems on nuclear DNA due to unexpected intron size variation in some species. Moreover, the gene does not belong to a multigenic family. The modules are also clearly distinct from one another in sequence, so specific primers for one or the other module can be designed. In the worst case scenario, where amplification of an unwanted module still occur, it can be detected at the alignment stage or by using a phylogenetic tree of the modules, as all sequences of a module cluster together in analyses.

4. Conclusions

By showing the presence of a single four-module IRBP gene in the genome of the lamprey *Petromyzon marinus*, our study has corroborated that the two-gene condition observed in Teleostei is derived compared to the one-gene condition, and increased considerably the known age of the apparition of the gene. The good support of each of the branches of the modules hints at an even older age, and the basal branches of the module clades show a noticeable divergence.

IRBP could not be detected in the genomes of available *Ciona* species nor in other available non-vertebrates. It would be of the highest interest to test for its presence in hagfishes.

The exploration of the absence or presence of the first three modules of IRBP in Teleostei must be extended, and assessed using other methods like high stringency Northern Blot for instance. If this first gene was really lost in some groups, this has happened repeatedly in the teleost tree. Yet both gene 1 and gene 2 are transcriptionally active in zebrafish adult eyes as well as in whole larval bodies, with different timing and pattern of the expression (Nickerson et al., 2006) hinting at possible subfunctionalization. Corroborating the absence of this module in some species might lead to new insights about the possible subfunctionalization of the two genes in teleosts. The timing of the acquisition of the derived structure of the genes in teleosts also needs to be investigated by studying the sequence of the IRBP coding gene or genes in basal Actinopterygians.

The current problems with the phylogeny of acanthomorph fishes will probably require new nuclear markers to supplement those most often used at present: almost complete mitogenomes, RAG1 and RAG2, Tmo4c4, 28S rDNA, rhodopsin retrogene, MLL4, and, very recently, the markers proposed by Li et al. (2007). While no single gene is sufficient for phylogenetic reconstruction, the partial IRPB sequences provide an interesting marker, less saturated than the mitogenome but more variable than most markers used for teleosts, and nonetheless recovering efficiently the clades proposed by other studies. The modules are different enough from one another to avoid confusion. The interrelationship tree of the modules shows that each of the IRBP modules is an interesting candidate for use with a wider sampling in the very diverse Acanthomorpha without risk of confusion between modules and genes. Even short sequences (713 bp) give a satisfactory result, comparable to those derived from multiple datasets combination. We therefore recommend this marker for large-scale phylogenies with wide taxonomic sampling in Acanthomorpha and Teleostei.

Acknowledgments

We thank Bruno Chanet, Samuel Iglésias, Leo W. Smith and Michel Hignette for samples, Sylvie Mazan for help with the *Petromyzon marinus* searches, Martine Dessoutter for specimen identifications, Frédéric Busson for FishBase data extraction, Wei-Jen Chen and Gaël Lancelot for manuscript improvement. We thank the Service de Systématique moléculaire for technical support. A.D. was the recipient for this project of a BQR grant of the Muséum National d'Histoire Naturelle: "Isolation de marqueurs moléculaires pour la phylogénie inter-ordinale et la cytogénétique intra-familiale des acanthomorphes (teleostéens)".

References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Benning, M.M., Taylor, K.L., Liu, R.-Q., Yang, G., Xiang, H., Wesenberg, G., Dunaway-Mariano, D., Holden, H.M., 1996. Structure of 4-chlorobenzoyl coenzyme A dehalogenase determined to 1.8 Å resolution: an enzyme catalyst generated via adaptive mutation. Biochemistry 35, 8103–8109.

Chen, W.J., Bonillo, C., Lecointre, G., 2003. Repeatability as a criterion of reliability of new clades in the acanthomorph (Teleostei) radiation. Mol. Phylogenet. Evol. 26 (2), 262–288.

Chen, W.-J., Ortí, G., Meyer, A., 2004. Novel evolutionary relationships among four fish model systems. Trends Genet. 20 (9), 424–431.

Chen, W.-J., Ruiz-Carus, R., Ortí, G., 2007. Relationships among four genera of mojarras (Teleostei: Perciformes: Gerreidae) from the western Atlantic and their tentative placement among percomorph fishes. J. Fish Biol. 70 (sB), 202–218.

Collins, T.M., Fedrigo, O., Naylor, G.J., 2005. Choosing the best genes for the job: the case for stationary genes in genome-scale phylogenetics. Syst. Biol. 54 (3), 493–500.

Dettaï, A., Lecointre, G., 2004. In search of the Notothenioid (Teleostei) relatives. Antarctic Sci. 16, 71–85.

Dettaï, A., Lecointre, G., 2005. New clades in the acanthomorph bush inferred by the study of taxonomic congruence. C.R. Biol. 328 (7), 674–689.

Engel, C.K., Mathieu, M., Zeelen, J.P., Hiltunen, J.K., Wierenga, R.K., 1996. Crystal structure of enoyl-coenzyme A (CoA) hydratase at 2.5 angstroms resolution: a spiral fold defines the CoA-binding pocket. EMBO J. 15, 5135–5145.

Froese, R., Pauly, D., 2006. FishBase. World Wide Web electronic publication (accessed in 2006).

Greenwood, P.H., Rosen, D.E., Weitzman, S.H., Mayers, G.S., 1966. Phyletic studies of teleostean fishes, with a provisional classification of living forms. Bull. Am. Mus. Nat. Hist. 131, 339–455.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52 (5), 696–704.

Guindon, S., Lethiec, F., Duroux, P., Gascuel, O., 2005. PHYML online—a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res. 33 (Web Server issue), W557–W559.

Hall, T.A., 1999. BIOEDIT: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Res. Symp. 41, 95–98.

Hassanin, A., Lecointre, G., Tillier, S., 1998. The 'evolutionary signal' of homoplasy in protein-coding gene sequences and its consequences for a priori weighting in phylogeny. C.R. Biol. 321 (7), 611–620.

Holcroft, N.I., 2004. A molecular test of alternative hypotheses of tetraodontiform (Acanthomorpha: Tetraodontiformes) sister group relationships using data from the RAG1 gene. Mol. Phylogenet. Evol. 32, 749–760.
Hubbard, T., Andrews, D., Caccamo, M., Cameron, G., Chen, Y., Clamp, M., Clarke, L., Coates, G., Cox, T., Cunningham, F., Curwen, V., Cutts, T., Down, T., Durbin, R., Fernandez-Suarez, X.M., Gilbert, J., Hammond, M., Herrero, J., Hotz, H., Howe, K., Iyer, V., Jekosch, K., Kahari, A., Kasprzyk, A., Keefe, D., Keenan, S., Kokocinsci, F., London, D., Longden, I., McVicker, G., Melsopp, C., Meidl, P., Potter, D., Proctor, G., Rae, M., Rios, D., Schuster, M., Searle, S., Severin, J., Slater, G., Smedley, D., Smith, J., Spooner, W., Stabenau, A., Stalker, J., Storey, R., Trevanion, S., Ureta-Vidal, A., Vogel, J., White, S., Woodwark, C., Birney, E., 2005. Ensembl 2005. Nucleic Acids Res. 33 (Database issue), D447–D453.
Huelsenbeck, J.P., Ronquist, F., 2001. MrBayes: Bayesian inference of phylogeny. Bioinformatics 17, 754–755.
Hurley, H.a., Lockridge Mueller, R., Dunn, K.A., Schmidt, E.J., Friedman, M., Ho, R.K., Prince, V.E., Yang, Z., Thomas, M.G., Coates, M.I., 2006. A new time-scale for ray-finned fish evolution. Proc. R. Soc. B. 274, 489–498.
Imamura, H., Shinohara, G., 1998. Scorpaeniform fish phylogeny: an overview. Bull. Nat. Sci. Mus., Tokyo, Ser. A 24, 185–212.
Inoue, J.G., Miya, M., Tsukamoto, K., Nishida, M., 2003. Basal actinopterygian relationships: a mitogenomic perspective on the phylogeny of the "ancient fish". Mol. Phylogenet. Evol. 26 (1), 110–120.
Johnson, G.D., Patterson, C., 1993. Percomorph phylogeny: a survey of acanthomorphs and a new proposal. Bull. Mar. Sci. 52, 554–626.
Jondeung, A., Sangthong, P., Zardoya, R., 2007. The complete mitochondrial DNA sequence of the Mekong giant catfish (*Pangasianodon gigas*), and the phylogenetic relationships among Siluriformes. Gene 387 (1–2), 49–57.
Kikugawa, K., Katoh, K., Kuraku, S., Sakurai, H., Ishida, O., Iwabe, N., Miyata, T., 2004. Basal jawed vertebrate phylogeny inferred from multiple nuclear DNA-coded genes. BMC Biol. 2, 3.
Lavoue, S., Miya, M., Inoue, J.G., Saitoh, K., Ishiguro, N.B., Nishida, M., 2005. Molecular systematics of the gonorynchiform fishes (Teleostei) based on whole mitogenome sequences: implications for higher-level relationships within the Otocephala. Mol. Phylogenet. Evol. 37 (1), 165–177.
Li, C., Orti, G., Zhang, G., Lu, G., 2007. A practical approach to phylogenomics: the phylogeny of ray-finned fish (Actinopterygii) as a case study. BMC Evol. Biol. 20, 7–44.
Liao, D.I., Qian, J., Chisholm, D.A., Jordan, D.B., Diner, B.A., 2000. Crystal structures of the photosystem II D1 C-terminal processing protease. Nat. Struct. Biol. 7, 749–753.
Loew, A., Gonzalez-Fernandez, F., 2002. Crystal structure of the functional unit of interphotoreceptor retinoid binding protein. Structure 10, 43–49.
Mabuchi, K., Miya, M., Azuma, Y., Nishida, M., 2007. Independent evolution of the specialized pharyngeal jaw apparatus in cichlid and labrid fishes. BMC Evol. Biol. 7, 10.
Madsen, O., Scally, M., Douady, C.J., Kao, D.J., DeBry, R.W., Adkins, R., Amrine, H.M., Stanhope, M.J., deJong, W.W., Springer, M.S., 2001. Parallel adaptive radiations in two major clades of placental mammals. Nature 409 (6820), 610–614.
Miya, M., Kawaguchi, A., Nishida, M., 2001. Mitogenomic exploration of higher teleostean phylogenies: a case study for moderate-scale evolutionary genomics with 38 newly determined complete mitochondrial DNA sequences. Mol. Biol. Evol. 18, 1993–2009.
Miya, M., Takeshima, H., Endo, H., Ishiguro, N.B., Inoue, J.G., Mukai, T., Satoh, T.P., Yamaguchi, M., Kawaguchi, A., Mabuchi, K., Shirai, S.M., Nishida, M., 2003. Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences. Mol. Phyl. Evol. 26, 121–138.
Miya, M., Satoh, T., Nishida, M., 2005. The phylogenetic position of toadfishes (order Batrachoidiformes) in the higher ray-finned fish as inferred from partitioned Bayesian analysis of 102 whole mitochondrial genome sequences. Biol. J. Linn. Soc. 85, 289–306.
Modis, Y., Filppula, S.A., Novikov, D.K., Norledge, B., Hiltunen, J.K., Wierenga, R.K., 1998. The crystal structure of dienoyl-CoA isomerase at 1.5 Å resolution reveals the importance of aspartate and glutamate sidechains for catalysis. Structure 6, 957–970.
Nelson, J.S., 1994. Fishes of the World, third ed. John Wiley and Sons, Inc., Hoboken, New Jersey.
Nelson, J.S., 2006. Fishes of the World, fourth ed. John Wiley and Sons, Inc., Hoboken, New Jersey.
Nickerson, J.M., Frey, R.A., Ciavatta, V.T., Stenkamp, D.L., 2006. Interphotoreceptor retinoid-binding protein gene structure in tetrapods and teleost fish. Mol. Vision 12, 1565–1585.
Orti, G., Meyer, A., 1997. The radiation of characiform fishes and the limits of resolution of mitochondrial ribosomal DNA sequences. Syst. Biol. 46, 75–100.
Parenti, L.R., 1993. Relationships of atherinomorph fishes (Teleostei). Bull. Mar. Sci. 52, 170–196.
Philippe, H., Douzery, E., 1994. The pitfalls of molecular phylogeny based on four species, as illustrated by the Cetacea/Artiodactyla relationships. J. Mamm. Evol. 2, 133–152.
Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14 (9), 817–818.
Posada, D., Buckley, T.R., 2004. Model selection and model averaging in phylogenetics: advantages of the AIC and Bayesian approaches over likelihood ratio tests. Syst. Biol. 53, 793–808.
Rajendran, R.R., Van Niel, E.E., Stenkamp, D.L., Cunningham, L.L., Raymond, P.A., Gonzalez-Fernandez, F., 1996. Zebrafish interphotoreceptor retinoid-binding protein: differential circadian expression among cone subtypes. J. Exp. Biol. 199, 2775–2787.
Rosen, D.E., 1973. Interrelationships of higher euteleostean fishes. In: Greenwood, P.H., Miles, S., Patterson, C. (Eds.), Interrelationships of Fishes. J. Linn. Soc. (London) 53 Suppl. 1. Academic Press, New York, pp. 397–513.
Rosen, D.E., 1984. Zeiformes as primitive plectognath fishes. Am. Mus. Novit. 2782, 1–45.
Rosen, D.E., 1985. An essay on euteleostean classification. Am. Mus. Novit. 2827, 1–57.
Rozen, S., Skaletsky, H., 2000. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz, S., Misener, S. (Eds.), Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, pp. 365–386.
Silber, K.R., Keiler, K.C., Sauer, R.T., 1992. Tsp: a tail-specific protease that selectively degrades proteins with nonpolar C termini. Proc. Natl. Acad. Sci. USA 89, 295–299.
Smith, W.L., Craig, M.T., 2007. Casting the percomorph net widely: the importance of broad taxonomic sampling in the search for the placement of serranid and percid fishes. Copeia 1, 35–55.
Smith, W.L., Wheeler, W.C., 2004. Polyphyly of the mail-cheeked fishes (Teleostei: Scorpaeniformes): evidence from mitochondrial and nuclear sequence data. Mol. Phylogenet. Evol. 32, 627–646.
Smith, W.L., Wheeler, W.C., 2006. Venom evolution widespread in fishes: a phylogenetic road map for the bioprospecting of piscine venoms. J. Hered. 97 (3), 206–217.
Springer, M.S., DeBry, R.W., Douady, C., Amrine, H.M., Madsen, O., de Jong, W.W., Stanhope, M.J., 2001. Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny reconstruction. Mol. Biol. Evol. 18 (2), 132–143.
Stanhope, M.J., Czelusniak, J., Si, J.S., Mickerson, J., Goodman, M., 1992. A molecular perspective on mammalian evolution from the gene encoding interphotoreceptor retinoid binding protein, with convincing evidence for bat monophyly. Mol. Phylogenet. Evol. 1, 148–160.
Stiassny, M.L.J., 1986. The limits and relationships of the acanthomorph teleosts. J. Zool. (Lond.) (B) 1, 411–460.
Stiassny, M.L.J., 1993. What are grey mullets? Bull. Mar. Sci. 52, 197–219.
Stiassny, M.L.J., Jensen, J.S., 1987. Labroid interrelationships revisited: morphological complexity, key innovations, and the study of comparative diversity. Bull. Mus. Comp. Zool. 151, 261–319.
Stiassny, M.L.J., Wiley, E.O., Johnson, G.D., de Carvalho, M.R., 2004. Gnathostome fishes. In: Donaghue, M.J., Cracraft, J. (Eds.), Assembling the Tree of Life. Oxford University Press, New York.
Sun, Y., Ripps, H., 1992. Rhodopsin regeneration in the normal and in the detached/replaced retina of the skate. Exp. Eye Res. 55 (5), 679–689.
Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods), version 4.0b10. Sinauer Associates, Sunderland, Massachusetts.
Tatusova, T.A., Madden, T.L., 1999. Blast 2 sequences—a new tool for comparing protein and nucleotide sequences. FEMS Microbiol. Lett. 174, 247–250.
Tyler, J.C., Santini, F., 2005. A phylogeny of the fossil and extant zeiform-like fishes, Upper Cretaceous to Recent, with comments on the putative zeomorph clade (Acanthomorpha). Zool. Scripta 34 (2), 157–175.
Vidal, N., Hedges, S.B., 2005. The phylogeny of squamate reptiles (lizards, snakes, and amphisbaenians) inferred from nine nuclear protein-coding genes. C.R. Biol. 328 (10–11), 1000–1008.
Wagenhorst, B.B., Rajendran, R.R., Van Niel, E.E., Hessler, R.B., Bukelman, A., Gonzalez-Fernandez, F., 1995. Goldfish cones secrete a two-repeat interphotoreceptor retinoid-binding protein. J. Mol. Evol. 41, 646–656.
Wiley, E.O., Johnson, G.D., Dimmick, W.W., 2000. The interrelationships of acanthomorph fishes: a total evidence approach using molecular and morphological data. Biochem. Syst. Ecol. 28, 319–350.
Winnpennminck, B., Backeljau, T., Wachter, R.D., 1993. Extraction of high molecular weight DNA from molluscs. Trends Genet. 9, 407.
Winterbottom, R., 1974. A descriptive synonymy of the striated muscles of the Teleostei. Proc. Acad. Nat. Sci. Philadelphia 125 (12), 225–317.