Novel Relationships Among Ten Fish Model Species Revealed Based on a Phylogenomic Analysis Using ESTs

Dirk Steinke, Walter Salzburger, Axel Meyer

Lehrstuhl für Zoologie und Evolutionsbiologie, Department of Biology, University of Konstanz, 78457 Konstanz, Germany

Abstract. The power of comparative phylogenomic analyses also depends on the amount of data that are included in such studies. We used expressed sequence tags (ESTs) from fish model species as a proof of principle approach in order to test the reliability of using ESTs for phylogenetic inference. As expected, the robustness increases with the amount of sequences. Although some progress has been made in the elucidation of the phylogeny of teleosts, relationships among the main lineages of the derived fish (Euteleostei) remain poorly defined and are still debated. We performed a phylogenomic analysis of a set of 42 of orthologous genes from 10 available fish model systems from seven different orders (Salmoniformes, Siluriformes, Cypriniformes, Tetraodontiformes, Cyprinodontiformes, Beloniformes, and Perciformes) of euteleostean fish to estimate divergence times and evolutionary relationships among those lineages. All 10 fish species serve as models for developmental, aquaculture, genomic, and comparative genetic studies. The phylogenetic signal and the strength of the contribution of each of the 42 orthologous genes were estimated with randomly chosen data subsets. Our study revealed a molecular phylogeny of higher-level relationships of derived teleosts, which indicates that the use of multiple genes produces robust phylogenies, a finding that is expected to apply to other phylogenetic issues among distantly related taxa. Our phylogenomic analyses confirm that the euteleostean superorders Ostariophysi and Acanthopterygii are monophyletic and the Protacanthopterygii and Ostariophysi are sister clades. In addition, and contrary to the traditional phylogenetic hypothesis, our analyses determine that killifish (Cyprinodontiformes), medaka (Beloniformes), and cichlids (Perciformes) appear to be more closely related to each other than either of them is to pufferfish (Tetraodontiformes). All 10 lineages split before or during the fragmentation of the supercontinent Pangea in the Jurassic.

Key words: Teleost phylogeny — Phylogenomics — Euteleostei — Molecular clock — EST — Pangea — Gondwana

Introduction

The relative importance of increasing the number of analyzed taxa and the number of characters for accuracy of phylogenetic inferences remains an issue of debate (Hillis 1998; Hillis et al. 2003; Rosenberg and Kumar 2003; Gadagkar et al. 2005; Cummings and Meyer 2005; Rokas et al. 2005). Large-scale phylogenetic analyses inevitably involve a trade-off between taxon sampling and gene sampling. However, recent simulation and empirical studies suggest that increased gene sampling, in general, might have a greater beneficial effect on the rigor of the estimation of phylogenetic topologies than more extensive taxon sampling (Mitchell et al. 2000; Rosenberg and Kumar 2001; Rokas and Carroll 2005). The benefits of sam-

Correspondence to: Axel Meyer, Department of Biology, University of Konstanz, Postbox 5560 M618, D-78457 Konstanz, Germany; *email:* Axel.Meyer@uni-konstanz.de

pling several independent gene genealogies to infer an organismal phylogeny with confidence are widely recognized (Cummings et al. 1995; Takezaki et al. 2003; Chen et al. 2004) because a more complete representation of the whole genome is highly desirable and stochastic errors occurring in data with small sample size will decrease with increasing sample size.

Comparative phylogenomic analyses using expressed sequence tags (ESTs) from taxa across the spectrum of animal diversity promise to yield reliable and robust results. ESTs also provide an economical approach to identify large numbers of genes that can be used in gene expression and phylogenomic studies (Gerhold and Caskey 1996; Renn et al. 2004; Hughes et al. 2006). For this reason, and because of the rapid automated way of data collection and the relatively low costs associated with this technology, many individual scientists as well as large genome sequencing centers have generated large numbers of ESTs that are publicly available and their numbers continue to increase rapidly.

However, the use of ESTs for phylogenetic analyses is limited to the rather small number of species for which EST and genome projects have been conducted. In order to test to test the power of multilocus approaches to reveal phylogenies, it was necessary to choose a group of species for which extensive EST datasets are available. Here we chose teleost fish to conduct EST-based analyses with different phylogenetic approaches such as Bayesian inference and maximum likelihood in an effort to overcome possible pitfalls of one particular method. A recent theoretical study (Mossel and Vigoda 2005) revealed that Bayesian MCMC methods for phylogeny reconstruction could be misleading when the data are generated from a mixture of datasets. Thus, in cases of datasets that contain potentially conflicting phylogenetic signals, phylogenetic reconstruction should be performed separately on each subset according to Mossel and Vigoda (2005).

There are more than 25,000 species of teleost fish, amounting to nearly half of the extant vertebrate species, and about 96% of all extant fish are classified as teleosts (Nelson 1994). Since the pioneering work on the systematics of fish by Greenwood et al. (1966), many studies have proposed novel hypotheses about the relationships among basal teleosts, but the relationships among the derived teleosts are still debated. One particular species-rich monophyletic group of derived teleosts is the Euteleostei, currently ranked as one of the four subdivisions of the Teleostei, along with the more basal groups, Osteoglossomorpha, Elopomorpha, and Clupeomorpha (Nelson 1994; de Pinna 1996; Arratia 1999; Miya et al. 2003, Inoue et al. 2004).

The Euteleostei are the most derived and species-rich group of teleost fish, comprising approximately

Table 1. Diversity and classification of ray-finned fish and species used in this study: The classification follows Nelson (1994)

Class Actinopterygii (23,681 species, 42 orders)
  Division Teleostei (23,637 species, 38 orders)
    Subdivision Euteleostei (22,262 species, 32 orders)
     Superorder Acanthopterygii (13,414 species, 13 orders)
      Order Beloniformes (191 species, 5 families)
       Oryzias latipes
      Order Cyprinodontiformes (807 species, 8 families)
       Fundulus heteroclitus
      Order Perciformes (9,293 species, 148 families)
       Haplochromis sp.
      Order Tetraodontiformes (339 species, 9 families)
       Takifugu rubripes
       Tetraodon nigroviridis
     Superorder Ostariophysi (6,507 species, 5 orders)
      Order Cypriniformes (2,662 species, 5 families)
       Cyprinus carpio
       Danio rerio
      Order Siluriformes (2,405 species, 34 families)
       Ictalurus punctatus
     Superorder Protacanthopterygii (312 species, 3 orders)
      Order Salmoniformes (66 species, 1 family)
       Oncorhynchus mykiss
       Salmo salar

16,000 species. These are placed into 32 orders and nine superorders. Currently used fish model species for developmental, genomic, and comparative genetic studies are assigned to three superorders of the Euteleostei: the Ostariophysi, the Protacanthopterygii, and the Acanthopterygii (Table 1). Ostariophysi are basal euteleosts characterized by the presence of the Weberian apparatus. Protacanthopterygii is a superorder that was established by Greenwood et al. (1966) and originally included a wide array of basal euteleosts. Since then, Rosen and Patterson (1969), Rosen and Greenwood (1970), and Rosen (1973, 1974) have repeatedly removed several of the orders that were originally included in the Protacanthopterygii by them. The resulting Protacanthopterygii (*sensu* Rosen 1974) is the basis of subsequent discussions on monophyly, interrelationships, and intrarelationships (e.g., Fink and Fink 1996; Ishiguro et al. 2003). Among the euteleosts the, by far, most diverse lineage are the Acanthopterygii (spiny rayed fish), comprising approximately 14,800 species, in which both the dorsal and the pelvic fins have true fin spines as well as rays. The majority of Protacanthopterygii has ctenoid scales, and the pelvic fins are thoracic, and the jaws protrusible. To the Acanthopterygii, Johnson and Patterson (1993) assigned five orders (Perciformes, Dactylopteriformes, Scorpaeniformes, Pleuronectiformes, and Tetraodontiformes) as a single clade and putative sister group to the Smegmamorpha (which contains the lineages Synbranchiformes, Mugiloidei, Elassomatidae, Gasterosteiformes, and the Atherinomorpha).

Table 2. GenBank, JGI, and ENSEMBL accession numbers of the sequences and numbers of amino acids used for the phylogenetic analyses

EST No.	Annotation	Amino acid used in analyses	Cichlid ESTs	*Cyprinus carpio*	*Danio rerio*	*Fundulus heteroclitus*
1	Adenylate kinase 2	189	BJ678728	CF662232	ENSDARP00000010686	CN983450
2	60S ribosomal protein L23a	130	BJ679384	AU240353	ENSDARP00000006624	CN979291
3	60S ribosomal protein L27a	131	BJ682982	CA965998	ENSDARP00000024203	CN983839
4	Triosephosphate isomerase 1	176	BJ683057	CF662476	ENSDARP00000008240	CV816996
5	snRNP core protein	134	BJ683377	CA970312	ENSDARP00000011550	CV819764
6	Acetoacetyl coenzyme A	187	BJ685403	CA968898	ENSDARP00000018064	CN970611
7	Eukaryotic translation initiation factor 4E	180	BJ686392	CA965942	ENSDARP00000031553	CN980908
8	Myosin regulatory light chain MRCL2	179	BJ686415	CA965179	ENSDARP00000020536	CN983930
9	Aldose reductase-like 1	217	BJ688419	CF661491	ENSDARP00000019242	CN985037
10	Eukaryotic translation initiation factor 3G	194	BJ690590	CA965572	ENSDARP00000044633	CN970731
11	F-actin capping protein α-1 subunit	209	BJ692160	CF662014	ENSDARP00000041236	CN983627
12	Eukaryotic translation initiation factor 5A	158	BJ692285	CF662459	ENSDARP00000027654	CN971618
13	Chromosome 5 open reading frame 18	169	BJ697855	CA964372	ENSDARP00000028937	CN983778
14	Myosin light chain 2	171	BJ699515	CA964394	ENSDARP00000023063	CV824448
15	Tyrosine-protein kinase SRC	158	BJ699827	AU081450	ENSDARP00000023656	CN969148
16	60S ribosomal protein L11	178	BJ702068	AU301062	ENSDARP00000003203	CN978772
17	40S ribosomal protein S3a	214	BJ702199	AU183418	ENSDARP00000017987	CN985474
18	Signal sequence receptor, delta	176	BJ702489	CA966252	ENSDARP00000047174	CN961958
19	ras-related nuclear protein	144	A30058AR_P17_15	CF663033	ENSDARP00000017573	CN983151
20	Trypsinogen 2	244	KN-353_16D_C10_T7	CA965313	ENSDARP00000022239	CN990611
21	Tumor suppressor candidate 3	157	KN-353-11A_E12_T7	CA968933	ENSDARP00000014499	CN953901
22	L-3-Hydroxyacyl-CoA dehydrogenase	235	KN-353-11B_F05_T7	CF662461	ENSDARP00000050602	CN985279
23	Cyclophilin B	219	KN-353-12A_F07_T7	CA966741	ENSDARP00000027189	CN984535
24	Elastase 1	234	KN-353-12A_G03_T7	CA967194	ENSDARP00000015328	CN974780
25	Glutathione *S*-transferase M3	100	KN-353-12A_G08_T7	CA967612	ENSDARP00000046181	CN983366
26	Arp2/3 protein complex	173	KN-353-12B_B10_T7	CF661114	ENSDARP00000025212	CN982288
27	Ribosomal protein S15a	135	KN-353-17B_A02_T7	CF661791	ENSDARP00000007879	CN984559
28	Proteasome α 6 subunit	185	KN-353-17C_A05_T7	CA970206	ENSDARP00000006320	CN983790
29	Tubulin α 6	238	KN-353-17D_B12_T7	CA969352	ENSDARP00000042337	CN974950
30	High-mobility group box 1	175	KN-353-17D_F02_T7	CA968420	ENSDARP00000048925	CV822354
31	Superoxide dismutase 1	141	KN-353-18_A02_T7	CA964628	ENSDARP00000049253	CO436130
32	TGF ß-inducible nuclear protein 1	226	KN-353-18A_H01_T7	CA966357	ENSDARP00000003036	CV817035
33	Carboxypeptidase A	197	KN-353-19A_G07_T7	CF660940	ENSDARP00000024981	CV819967
34	ATP synthase, mitochondrial F1, δ	145	KN-353-19C_A03_T7	CA966367	ENSDARP00000022528	CN985063
35	Chymotrypsinogen B1	258	KN-353-19D_D06_T7	CF662498	ENSDARP00000004441	CN981598
36	Malate dehydrogenase	216	KN-353-20_D05_T7	CF660983	ENSDARP00000048494	CV819596
37	Retinol dehydrogenase 4	220	KN-353-20C_B10_T7	CF662451	ENSDARP00000027139	CN981573
38	Cathepsin L preproprotein	251	KN-353-20C_G07_T7	AB128161	ENSDARP00000042856	CN983318
39	Heme oxygenase	245	KN-353-20D_D03_T7	CA964185	ENSDARP00000038993	CN964309
40	Cytochrome *c* oxidase subunit Va	130	KN-353-22_B02_T7	CA965958	ENSDARP00000025342	CV821701
41	LYST-interacting protein LIP2	166	KN-353-25_F09_T7	AU301657	ENSDARP00000020761	CN984213
42	Amylase, α 1A	143	KN-353-25_G04_T7	CA965259	ENSDARP00000021453	CN976675

Ictalurus punctatus	Oncorhynchus mykiss	Oryzias latipes	Salmo salar	Takifugu rubripes	Tetraodon nigroviridis	Homo sapiens
CK404937	BX084154	AU167296	CB516091	FRUP00000147185	CAG02308	NP_037543.1
CK421413	CA387238	BJ011898	CA052855	FRUP00000128020	CAG00513	NP_000975.2
CK405718	CA373609	BJ512943	CB509769	FRUP00000158498	CAG05610	NP_000981.1
CB940919	CA375258	BJ713765	CK892449	FRUP00000156180	CAF90849	NP_000356.1
CB940512	BX081447	BJ526519	CB510840	FRUP00000141225	CAF93753	NP_004166.1
CK407109	BX861495	BJ518743	BG935894	FRUP00000152839	CAG03628	NP_005882.1
CK411802	CA343234	BJ711331	BG934363	FRUP00000149241	CAF94272	NP_001959.1
CK425292	BX077116	AU178865	CB503229	FRUP00000164111	CAG10218	NP_291024.1
CK418638	BX077801	BJ002188	BM413709	FRUP00000162086	CAG12115	NP_064695.2
CK412877	BX074310	BJ729066	CK874251	FRUP00000136092	CAG12908	NP_003747.1
CK410486	628259	BJ735067	BG935543	FRUP00000134542	AAR16326	NP_006126.1
CK406655	BX081045	BJ512185	CK893994	FRUP00000158888	CAF89591	NP_001961.1
CF972147	649856	BJ494393	AJ425642	FRUP00000149017	CAG10310	NP_005660.3
CK410726	BX306316	BJ705801	CK881065	FRUP00000162164	AAS90116	NP_037424.2
CF972278	6273112	BJ717913	AF321110	FRUP00000134963	CAG11788	NP_005408.1
CK426309	655997	BJ713881	CD511092	FRUP00000164953	CAF89662	NP_000966.2
CK421043	BX077891	BJ014568	CB503624	FRUP00000129833	CAF90706	NP_000997.1
CB939743	627862	BJ497957	CB514131	FRUP00000161878	CAG07447	NP_006271.1
CK418786	640152	BJ002859	CK888966	FRUP00000147526	CAG04789	NP_006316.1
CK421385	BX074190	BJ709131	CA044506	FRUP00000148709	CAG00063	NP_002761.1
CK409871	BX297096	BJ729148	CA060835	FRUP00000136175	CAG11530	NP_839952.1
CK425291	629060	BJ710935	CB513348	FRUP00000163423	CAG11476	NP_005318.1
BM027882	625265	BJ714574	CB513750	FRUP00000138892	CAF98384	NP_000933.1
CK407314	BX076804	BJ714998	CB504468	FRUP00000138033	CAG06304	NP_001962.2
CK418636	633117	AU170509	CK891233	FRUP00000148010	CAG07510	NP_000840.2
CB938384	BX312030	BJ530477	CK873409	FRUP00000157357	CAG06784	NP_005709.1
CK424864	638935	BJ492753	CD510685	FRUP00000151157	CAG03318	NP_001010.2
CK409903	BX075861	BJ721444	CK875988	FRUP00000165441	CAG00121	NP_002782.1
CB938586	S15341287	AU167720	CB503051	FRUP00000140263	CAG03831	NP_116093.1
CK417131	623331	BJ728213	CK890173	FRUP00000152052	CAG09003	NP_002119.1
BE469461	653657	BJ735553	CB513051	FRUP00000140559	CAG00454	NP_000445.1
CK402128	653564	BJ727573	CB516512	FRUP00000155317	CAG05206	NP_055701.1
CK423133	BX076871	BJ717301	CB510467	FRUP00000132419	AAR16320	NP_001859.1
CK419832	BX076940	BJ729540	CA052975	FRUP00000163451	CAF92415	NP_001678.1
CK402317	BX074786	BJ714978	CB503195	FRUP00000153300	CAG00821	NP_001897.1
CK411748	629835	AJ457305	CK883250	FRUP00000136061	CAG12894	NP_005909.2
CK402358	627431	BJ707018	CB514033	FRUP00000159324	CAF92451	NP_003699.2
BM438275	S15340856	BJ714814	CB516435	FRUP00000136907	CAF88807	NP_001903.1
CK408643	BX317345	AB163431	BG936101	FRUP00000152368	CAF95107	NP_002124.1
CB940074	002027	BJ750384	CA041562	FRUP00000132409	CAG08740	NP_004246.1
CF262675	BX296907	BJ020755	CB504725	FRUP00000148712	CAF98040	NP_071344.1
CK423338	BX075205	BJ516256	CA043016	FRUP00000161922	CAD20312	NP_004029.2

Much controversy persists over the interrelationships among teleosts. The euteleost origin dates back to about 290 million years ago (Kumazawa et al. 1999; Inoue et al. 2005), and due to the extensive variation not only in morphology but also in behavior, ecology, and physiology (see Helfman et al. 1997), it is not surprising that comparative anatomical approaches were faced with a number of difficulties (e.g., lack or paucity of applicable characters for phylogenetic analyses and difficulties in the homology assessment among characters). The same is true for earlier molecular studies (Stepien and Kocher 1997; Miya and Nishida 2000) that used shorter (mostly mitochondrial) DNA sequences (mostly < 1000 positions) based on limited taxonomic representation. However, it is highly desirable to establish the relationships among the fish model systems in order to be able to interpret comparative genomic and developmental processes within the correct phylogenetic framework. It appears that adequate resolution of higher-level relationships among distantly related lineages will require longer stretches of DNA (e.g., Miya et al. 2003), amino acid sequences (e.g., Hoegg et al. 2004), or DNA datasets based on multiple loci (e.g., Chen et al. 2004; Simmons and Miya 2004; Takezaki et al. 2004). Recent molecular studies based on complete mitochondrial genomes have demonstrated the power of this approach since they resulted in highly resolved phylogenies that demonstrated that the Ostariophysi and Protacanthopterygii are sister groups (Ishiguro et al. 2003; Saitoh et al. 2003).

In order to increase the size of the gene sample available for phylogenetic analysis, we took advantage of two complete actinopterygian fish genomes and collections of ESTs available from public databases. In the present study 42 concatenated amino acid sequences retrieved by similarity searches against public DNA and protein sequence databases were used to address the question of the relationships of derived teleosts and to test the power of multilocus approaches in establishing well-supported phylogenies. The phylogenetic signal and the strength of contribution of each of the 42 genes were estimated with randomly chosen data subsets. Our study resulted in a molecular phylogeny among derived teleosts, which indicates that the use of multiple genes produces robust phylogenies. The phylogeny was used to estimate divergence times and to examine the evolutionary history of the component lineages within the teleostean fish.

Materials and Methods

Data Collection

Cichlid EST sequences generated by us (Salzburger et al., in preparation) and in a previous study (Watanabe et al. 2004) were screened against GenBank EST data for *Cyprinus carpio* (Cypriniformes), *Fundulus heteroclitus* (Cyprinodontiformes), *Ictalurus punctatus* (Siluriformes), *Oncorhynchus mykiss* (Salmoniformes), *Oryzias latipes* (Beloniformes), *Salmo salar* (Salmoniformes), and *Tetraodon nigroviridis* (Tetraodontiformes). All of these species are important fish model species. We also used protein data for *Danio rerio* (Zebrafish Sequencing Group at the Sanger Institute) and *Takifugu rubripes* (JGI Fugu v3.0) and genome data for *Homo sapiens* (GenBank). *Homo sapiens* was used as closest related outgroup with available data. We used EverEST (Steinke et al. 2004), a software program, for processing simultaneous database searches based on the BLAST algorithm against all above-mentioned databases to identify the best hits for any given cichlid EST sequence. EverEST was also used to assign query sequences to matched BLAST results. Only those sequences were assigned to the query gene from cichlids that were recovered as "best hits" in a translated BLAST routine using the standard vertebrate code and an e-value $\leq 10^{-50}$. The sequences were aligned using the T-Coffee algorithm (Notredame et al. 2000). Forty-two genes were found to be present in all 11 databases for all taxa and were conserved enough so that an unambiguous alignment was possible. The accession numbers of the analyzed sequences and the number of amino acids used for the phylogenetic analyses are listed in Table 2. Gene sequences were concatenated to form a supergene alignment with a total length of 7726 amino acid positions.

Phylogenetic Analyses

Neighbor-joining (NJ) and maximum parsimony (MP) analyses of the combined amino acid alignment were performed with PAUP* v. 4.10b (Swofford 2002). Maximum likelihood (ML) analyses were performed using PHYML (Guindon and Gascuel 2003). The best-fitting models of sequence evolution for ML were obtained by ProtTest 1.2 (Abascal et al. 2005).

Confidence in estimated relationships of NJ, MP, and ML tree topologies was evaluated by a bootstrap analysis with 2,000 replicates (Felsenstein 1985) and Bayesian methods of phylogeny inference (Larget and Simon 1999). Bayesian analyses were initiated with random seed trees and were run for 200,000 generations. The Markov chains were sampled at intervals of 100 generations with a burn-in of 1000. Bayesian phylogenetic analyses were conducted with MrBayes 3.0b4 (Huelsenbeck and Ronquist 2001) using the Whelan and Goldman (2001) model + I + Γ. Alternative topologies were compared applying the approximately unbiased test (Shimodaira 2002) as implemented in the CONSEL package (Shimodaira and Hasegawa 2001), using the sidewise likelihood values estimated by PAML (Yang 1997).

In order to test the phylogenetic signal and the contribution of each of the 42 genes and combinations of those to the general topology, we randomly selected 100 subsets, each containing six EST loci and constructed ML trees for every subset and every single-gene using PHYML with the model settings estimated as described above. The subset size of six represents a trade-off between the computational power and the likelihood to retrieve every possible pair of loci. The number of subset gene trees supporting the basal dichotomy was used to evaluate those contributing loci by counting the number of locus pairs represented in correctly inferred subset topologies. This amount was used for a graphical matrix representation of the contribution of all possible loci pairs to infer the phylogenetic signal of loci combinations. We also calculated the number of single-gene trees supporting a given partition of the general topology (see Gadagkar et al. 2005). Using Poisson corrected average pairwise distances we also generated six subsets, each containing seven loci of genes with different evolutionary rates. The loci were
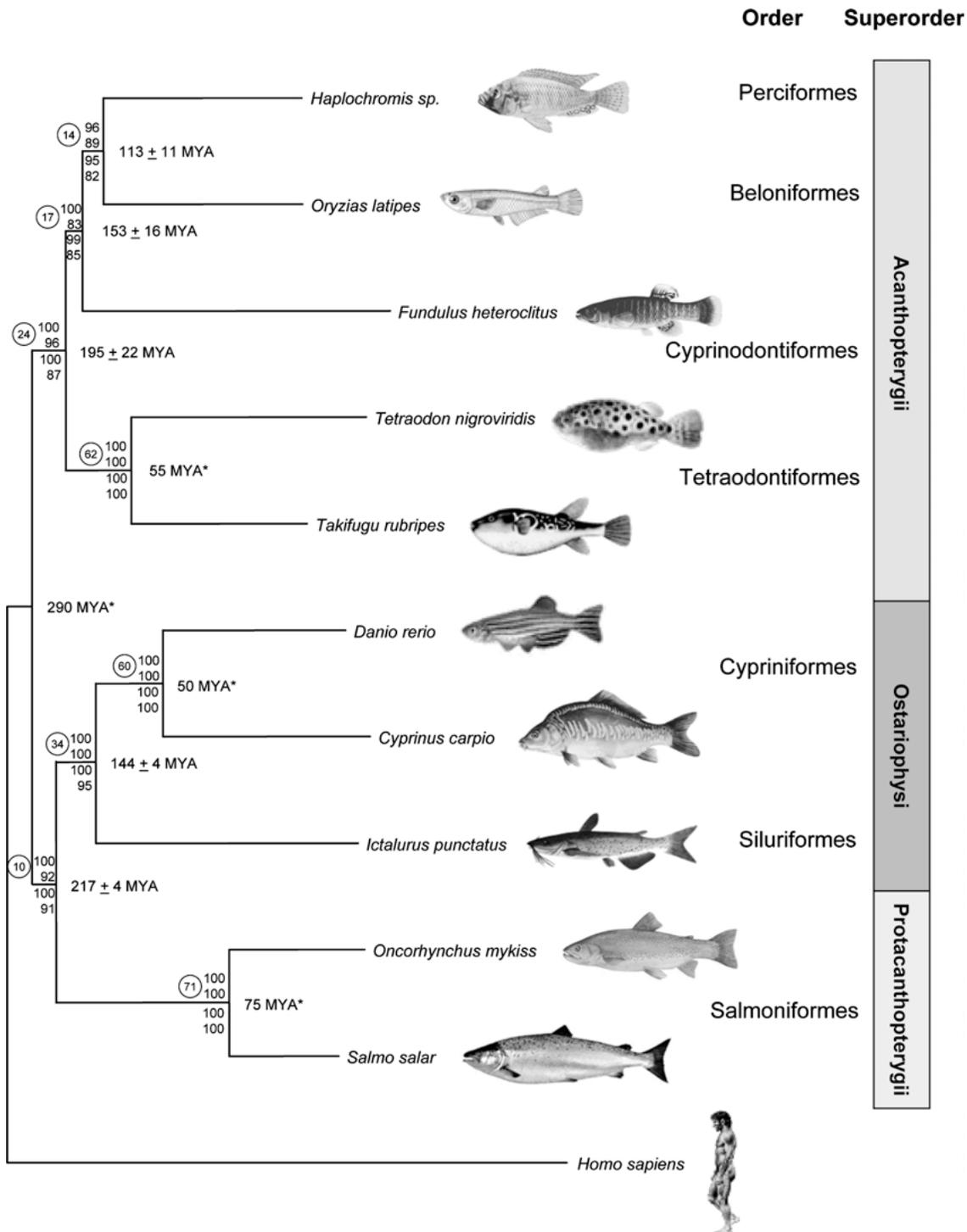
Order   Superorder

Fig. 1. Phylogeny based on a combined dataset of 42 loci with a total of 7726 amino acid positions. Values above branch indicate posterior probabilities (MrBayes; upper value of quartet) and bootstrap values from maximum likelihood (PHYML; second value of quartet). Numbers below branch represent bootstraps from neighbor joining (third value of quartet) and maximum parsimony (both PAUP*; lowest value of quartet). Value to the right of the node represents the estimated age (million years ago; MYA) calculated using the local molecular clock method of age estimation with an optimization via the truncated Newton method with r8s (Sanderson 2003). Confidence intervals were assessed by means of a bootstrap approach with 25 replicates. Calibration points are indicated by an asterisk. Number in a circle to the left of the node represents the percentage of single-gene trees supporting that node (see Gadagkar et al. 2005).

grouped according to their distances, which were estimated using MEGA 3.0 (Kumar et al. 2004). This analysis was performed to test relationships between substitution rates and topology by constructing ML trees as described above.

Molecular Clock

To estimate a local molecular clock a method of estimation with an optimization via the truncated Newton method was employed, as implemented in r8s (Sanderson 2003). The truncated Newton (TN) algorithm tolerates age constraints. Divergence time algorithms require at least one internal node to be fixed or constrained. We used three dates: 55 MYA marking the earliest known fossil evidence for the Tetraodontidae (Berg 1958), 75 MYA for the earliest known fossil evidence of the Salmonidae (Resetnikov 1988), and 50 MYA as the age of the last common ancestor of *Danio* and *Cyprinus* (Cavender 1991; Kruiswijk et al. 2002). The first two calibration nodes were only constrained by the max_age function in r8s; the latter one was fixed because the fossil represents the last common ancestor of both lineages. Based on these fossil calibrations, trees were constrained at the basal node at 290 MYA, the date at which pufferfish and zebrafish, a representative of the most basal lineage in this study, shared a last common ancestor. This estimation is based on a previous calibration from molecular data (Kumazawa et al. 1999; Inoue et al. 2005) and the data should therefore be treated with caution. Confidence intervals were assessed by means of a bootstrap approach. We simulated 25 bootstrap matrices with Seqboot (PHYLIP 3.63 package; Felsenstein 1989) and, for each matrix, constructed a ML tree. The resulting trees were then analyzed with r8s as described above. The minimum and maximum values are represented by the minimum and maximum age estimates of the simulation matrices. Consistency between fossil and molecular age estimates for the three fossil calibration points was examined using the fossil cross-validation method (Near and Sanderson 2004; Near et al. 2005). The calibration points are approximately equally accurate because the magnitude of the squared deviation is only decreasing by a small fraction as fossils are removed (Near et al. 2005).

Results

The alignment of the dataset consisted of 42 orthologous groups of eukaryotic protein fragments of 10 teleost species and 1 outgroup species. The total length of the combined dataset was 7726 amino acid positions; 4778 positions were invariant. Of the remaining 2948 variant positions, 1557 were phylogenetically informative (shared by at least two taxa).

MP, ML, NJ, and Bayesian inference analyses produced identical tree topologies. The phylogenetic analyses of the complete dataset (Fig. 1) strongly supported the monophyly of the teleost fish used in this study. Our analyses recovered two major clades in the teleosts. The first clade includes members of the Salmoniformes, Siluriformes, and Cypriniformes and is supported by high bootstrap and posterior probability values. Within this clade, the representatives of the Salmoniformes (*Oncorhynchus mykiss* and *Salmo salar*) appeared as sister group to a clade comprised by the Siluriformes (*Ictalurus punctatus*) and the Cypriniformes (*Cyprinus carpio* and *Danio rerio*). In the second clade, the representatives of the Tetraodontiformes (*Tetraodon nigroviridis* and *Takifugu rubripes*) were placed as sister group to a clade formed by the Cyprinodontiformes, Beloniformes,

Table 3. Comparison of the likelihood values of different topologies among the different superorders within the euteleosts, applying the approximately unbiased test

Topology	Loglk	Δloglk	P
(Acan (Prot + Osta))	−58,999.377	0.000	
(Prot (Osta + Acan))	−59,234.660	−235.283	0.002
(Osta (Prot + Acan))	−59,456.775	−457.398	0.001

Note. The first topology is the maximum likelihood tree. Acanthopterygii, Acan; Ostariophysi, Osta; Protacanthopterygii, Prot. loglk, likelihood; Δloglk, difference of likelihood; p, p value, approximately unbiased test.

and Perciformes. All nodes in this clade were strongly supported as well. The members of the Perciformes (*Haplochromis sp.*) and Beloniformes (*Oryzias latipes*) formed a monophyletic group, and the representative of the Cyprinodontiformes (*Fundulus heteroclitus*) branched basal to this clade.

Comparing different topologies within the euteleost fish with the approximately unbiased test significantly ruled out possible alternative Superorder relationships ((Protacanthopterygii (Ostariophysi + Acanthopterygii)) or (Ostariophysi (Protacanthopterygii + Acanthopterygii))). Thus a sister group relationship between the Ostariophysi and the Acanthopterygii or between the Protacanthopterygii and the Acanthopterygii were rejected (Table 3).

The relative ages of the main clades within the teleostean fish as revealed from our molecular clock analyses were also estimated (Fig. 1). The split between the Ostariophysi/Protacanthopterygii clade and the Acanthopterygii was dated to the early Triassic (approximately 217 ± 4 MYA), whereas all other splits were estimated to have occurred in the Jurassic (135–190 MYA). Based on our calibrations, the split between the Cypriniformes and the Siluriformes was estimated to have occurred at 141 ± 4 MYA. The time estimate for the split between the Tetraodontiformes and all other Acanthopterygian species was 195 MYA, whereas the Cyprinodontiformes diverged from the latter group 153 ± 16 MYA. The estimated divergence time between the cichlids and the Beloniformes was dated to 113 ± 11 MYA. Based on our calibration points, these age estimates are relatively robust; the mean age estimated from 25 bootstrap trees for which we repeated the age estimation procedure outlined above reveals a maximum of 16 MYA standard deviation and a fossil cross-validation (Near and Sanderson 2004; Near 2005) resulted in inconsistent molecular age estimates. Despite the fact that our results correspond well with recent studies, all molecular clock estimations should be treated with caution, because we used a molecular calibration (290 MYA for the last common ancestor of zebrafish and pufferfish) to constrain the basal node.

Fig. 2. Matrix plot of EST pairs of 100 simulated subsets containing six loci. The gray scale corresponds to the frequency of particular gene pairs occurring in maximum likelihood subset topologies congruent with the basal dichotomy depicted in Fig 1. The histogram below depicts the absolute number of appearances of single genes in congruent topologies. The trees on the right side exemplify two maximum likelihood trees of combinations with a high phylogenetic signal (EST 11 + 42) and a low phylogenetic signal (EST 7 + 12). Numbers correspond to the first column in Table 2.

The strength of the phylogenetic signal and the contribution of each subset of genes to the general topology in 100 random subsets each containing six EST loci are rather weak as depicted in Fig. 2 by a matrix representation of the frequency of occurrence of loci in "correct" topologies. Only combinations of a few loci (e.g., EST 11 and EST 42) showed enough resolution to reproduce the basal dichotomy or at least one of three subgroups (superorders), however, the complete estimated topology as depicted in Fig. 1 was not found with any of the 100 subsets above. The percentage of single-gene trees supporting a given partition of the general topology ranges from 10% to 71% (Fig. 1), with terminal nodes being more often correctly inferred than basal nodes. Substitution rates of loci with high (e.g., EST 11 or EST 42) and low (EST 7 or EST 12) phylogenetic signal were similar and ranged from 0.13 to 0.17. The Poisson corrected amino acid substitution rate among all loci ranged from 0.01 to 0.31 (Fig. 3). Phylogenetic analyses of six subsets containing seven loci each according to the Poisson corrected amino acid substitution rate supported the two major clades also with low substitution rates. However, relationships within the two

clades varied with the substitution rate (Fig. 3). The overall topology as depicted in Fig. 1 was not recovered. However, the topologies were similar to that supported by the analysis of the combined dataset, whenever loci with lower amino acid substitution rates were used. Figure 4 shows that the length of the EST groups used does not correlate ($R^2 = 0.0081$) with the amino acid substitution rates and therefore we conclude that the analyses of subsets are not biased due to length differences.

Discussion

Implications for Multilocus Phylogenies

The approach used in this study led to a well-supported but novel hypothesis of evolutionary relationships among the euteleostean fish (Fig. 1). The sampling of multiple genes with a comparatively large number of sequence positions is likely to improve phylogenetic robustness (Lake and Moore 1998). The large amounts of ESTs being produced through automated sequencing technologies is therefore likely
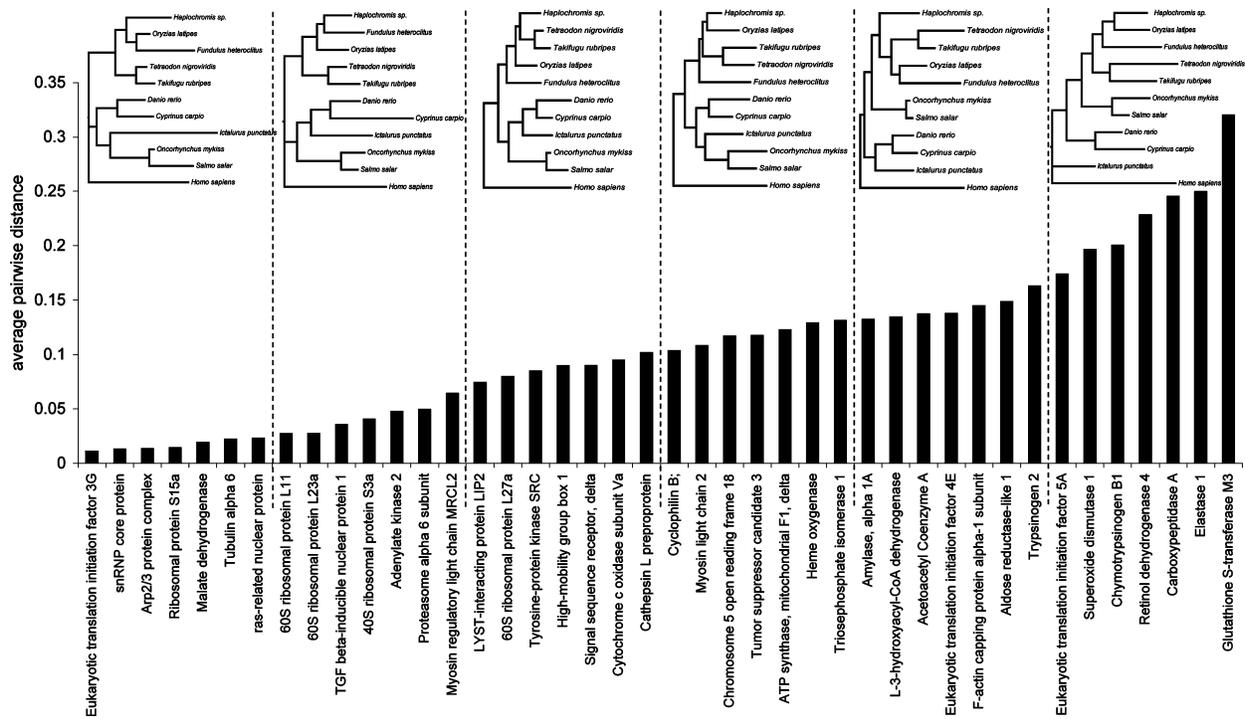
Fig. 3. Average Poisson corrected pairwise distances for the 42 ORF groups in ascending order. The trees show estimated maximum likelihood topologies for six subsets, each of which is based on the phylogenetic analysis of seven loci.
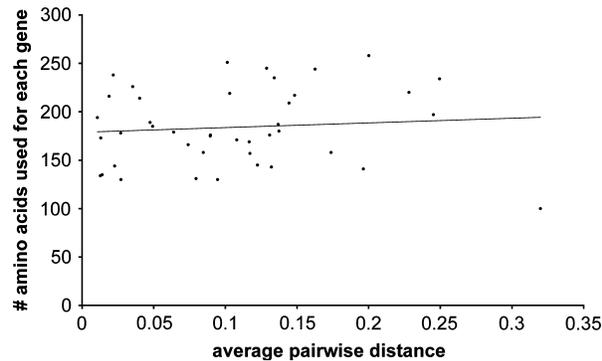

Fig. 4. Plot of Poisson corrected amino acid substitution rates versus the number of amino acids used for each gene.

to provide scientists with sufficient data to calculate reliable multi locus phylogenies. The power of comparative phylogenomic analyses using ESTs from different taxa is a function of the number of data available. Here, we were able to show that robustness increases with the amount of available sequences independent of their length and rate of amino acid substitution (Fig. 4). Remarkably, all analyses produced congruent tree topologies with confidence values not lower than 82 (Fig. 1). Given the fact that subsets of EST groups containing six sequences simulated in this study were not able to recover the phylogeny of the concatenated dataset, we conclude that comparably high numbers of loci are needed to infer robust phylogenies from EST based studies

among distantly related taxa. This inference should be tested in the future with even more inclusive phylogenetic issues, such as the relationships among animal phyla and the questions surrounding the "Cambrian explosion" (Rokas et al. 2005). The number of single-gene trees supporting partitions of the general topology corroborates this observation (Fig. 1). Only terminal nodes could be resolved with single-gene trees. The discrepancy between high confidence values within the combined analysis and the low number of single-gene trees supporting the general topology shows that multilocus analyses perform better in resolving higher-level relationships among distantly related lineages than single-locus analyses as was suggested before repeatedly. However, the single-gene trees are based on relatively short amino acid sequences due to the fact that we used EST data, which are usually not longer than ~600 bp. Therefore single-locus datasets might not contain sufficient information to produce robust phylogenies, also indicated by the relative small total number of informative sites (1557) compared to that in a study at a comprehensive phylogenetic scale (Miya et al. 2003). Although each of the loci included in the 42-gene set was carefully screened and orthology among the sequences derived from the different species seemed rigorously established, the possibility of unrecognized paralogy at a few of the loci still cannot be fully excluded. The contribution to phylogenetic uncertainty of such paralogous sequences

could, however, not have been large enough to influence the phylogenetic reconstruction of the concatenated dataset. Yet it might have been large enough to influence the reconstruction of some single-gene trees.

Part of the increase in robustness afforded by concatenating multiple genes is due to the fact that many branches in individual gene trees may have experienced only a few substitutions. Adding genes to a dataset by concatenation increases the absolute number of evolutionary changes on such branches and makes it possible to infer them with greater accuracy. Furthermore, an overall increase in sequence length leads to an overall smaller variance in evolutionary rates and other parameters in model based methods. Therefore, it may be better not to discard genes producing incongruent phylogenies, as they may provide additional information for resolving some short branches (Shevchuk and Allard 2001; Rokas et al. 2003). On the other hand, if individual gene trees contain systematic errors that may result in similar (but erroneous) phylogenies, then the use of congruent phylogenies may actually result in an attenuation of this error. Despite the fact that we did not make an effort to account for large variation in evolutionary rates, sequence length, transition–transversion ratio, and base composition (G + C content) among the single sequences, the concatenated dataset still performed well. This indicates that the increase in phylogenetic signal or signal/noise ratio due to the concatenation is much higher than any bias introduced by using a single substitution pattern applied to the entire concatenated sequence. It is possible that the use of gene-specific evolutionary models in a partitioned approach may improve the accuracy of concatenated sequence analysis, but to date this is not possible with the available methods and software given the number of loci used in this study.

Implications for the Teleost Phylogeny

The results of phylogenomic analyses based on 42 orthologous groups of nuclear protein-coding genes confirmed the basal placement of Ostariophysi and Protacanthopterygii but revealed some unexpected relationships among acanthoptergyian species. Recent molecular studies have demonstrated that Ostariophysi and Protacanthopterygii are sister groups (Ishiguro et al. 2003; Saitoh et al. 2003), a finding that was confirmed in this study. According to the molecular clock analyses the basal divergence of the Ostariophysi and Protacanthopterygii took place no later than the middle Triassic (213–221 MYA). Pangean separation in the middle Jurassic may have been responsible for the present geographic patterns, in which cypriniform fish show a largely Laurasian distribution, whereas siluriform fish are likely to have originated in Gondwanaland, leading to their present South American distribution, on one hand, and African lineages that subsequently dispersed into the Eurasian continent following land connections or accretion, on the other (Saitoh et al. 2003). All members of the Ostariophysi share four or five modified vertebrae, aiding in hearing, which connect the swim bladder to the inner ear and convey pressure changes and sound (Weberian apparatus). Basal lineages maintained an adipose fin posterior to the dorsal fin, which is considered to be the ancestral character state for euteleosts. This enigmatic fin is not found in all basal euteleosts, however, since, e.g., esociforms and alepocephaloids lack it (Johnson and Patterson 1996), it is likely that it has been lost secondarily. In all other lineages of the Ostariophysi, especially in basal orders such as the Characiformes, an adipose fin is usually present.

The molecular data support a close relationship between the Atherinomorpha (Beloniformes and Cyprinodontiformes) and a representative (*Haplochromis*) of the Percomorpha, a sister group of the Smegmamorpha (Johnson and Patterson 1993). The monophyly of the Smegmamorpha is not supported by the present study or any previous molecular phylogeny (Wiley et al. 2000; Chen et al. 2003; Miya et al. 2003). Ancestral features among the atherinomorphs like a protrusible upper jaw and flexible spines on dorsal and pelvic fins in abdominal or subabdominal position are shared with basal teleosts (Nelson 1994). These features could be the result of a secondary loss that occurred during the evolution of ray finned fish because *Oryzias* and *Fundulus* are nested among perchlike fish like *Takifugu, Tetraodon*, and *Haplochromis*, just as recently hypothesized by Chen et al. (2004).

The splits in the Acanthopterygii group correspond well with the beginning breakup of Laurasia and the enlarging Turgai Sea in the Jurassic except the split of the tetraodontiform lineage (max. 216 MYA). Most of the tetraodontiform families are found in warm and temperate marine waters worldwide, with a few families absent from the Atlantic and eastern Pacific. The earliest known fossil evidence for the Tetraodontidae (Berg 1958; Santini and Tyler 2003; Santini 2004) dates back to the early Tertiary. The relatively long branches among the Tetraodontidae as depicted in Fig. 1 might be the result of independent and unique evolution along this lineage leading to rather compact genomes (Aparicio et al. 2002; Jaillon et al. 2004). Extant species of killifish and cichlids show a Gondwanan distribution (Streelman et al. 1998; Zardoya et al. 1996) that is in concordance with our paleophylogenetic reconstructions. The majority of the beloniform species are found in marine waters worldwide, and the family

Adrianichthyidae and members of the Belonidae are known to be secondary freshwater fish with Gondwanan distribution (Collette 2003).

Conclusion

We showed that multigene EST phylogenies represent a powerful method to increase the robustness of topologies. Our evaluations have demonstrated that inference of phylogeny robustness increases with the number of loci and that these loci should be chosen according to their rate of amino acid substitution. This study identified several more slowly evolving genes that are suitable candidates for future phylogenetic analyses of fish and, possibly, other taxa of similar age. The results of the genome-wide phylogenetic analysis described here indicate that the available data support previous findings in mtDNA based molecular studies for the Ostariophysi/Protacanthopterygii relationship (e.g., Ishiguro et al. 2003) and concatenated nuclear loci among the Acanthopterygii (e.g., Chen et al. 2004). To reach a new level of confidence for phylogenetic purposes, representative samples of genome sequences or EST sequences from additional relevant taxa are required. The rapid progress of genomic resources for an increasing number of species also emphasizes the importance of a reliable phylogenetic framework in which to interpret comparative results correctly.

References

Abascal F, Zardoya R, Posada D (2005) ProtTest: Selection of best-fit models of protein evolution. Bioinformatics 21:2104–2015

Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Rash S, Hoon S, Smit A, Gelpke MD, Roach J, Oh T, Ho IY, Wong M, Detter C, Verhoef F, Predki P, Tay A, Lucas S, Richardson P, Smith SF, Clark MS, Edwards YJ, Doggett N, Zharkikh A, Tavtigian SV, Pruss D, Barnstead M, Evans C, Baden H, Powell J, Glusman G, Rowen L, Hood L, Tan YH, Elgar G, Hawkins T, Venkatesh B, Rokhsar D, Brenner S (2002) Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. Science 297:1283–1285

Arratia G (1999) The monophyly of teleostei and stem-group teleosts. In: Arratia G, Schultze HP (eds) Mesozoic fish 2: Systematics and fossil record. Pfeil, Munich

Benton MJ (1993) The fossil record, vol. 2 Chapman and Hall, London

Berg LS (1958) System der rezenten und fossilen Fischartigen und Fische. Verlag der Wissenschaften, Berlin

Cavender (1991) The fossil record of the Cyprinidae. In: Winfield IJ, Nelson JS (eds) Cyprinid fish: Systematics, biology and exploitation. Fish and Fisheries Series 3. Chapman and Hall, London

Chen WJ, Bonillo C, Lecointre G (2003) Repeatability of clades as a criterion of reliability:a case study for molecular phylogeny of Acanthomorpha (Teleostei) with larger number of taxa. Mol Phylogenet Evol 26:262–288

Chen WJ, Orti G, Meyer A (2004) Novel evolutionary relationship among four fish model systems. Trends Genet 20:424–431

Collette BB (2003) Family Belonidae Bonaparte 1832—needlefish. Calif Acad Sci Annot Checklists Fish 16:1–22

Cummings MP, Meyer A (2005) Magic bullets and golden rules: data sampling in molecular phylogenetics. Zoology 108:329–336

Cummings MP, Otto SP, Wakeley J (1995) Sampling properties of DNA sequence data in phylogenetic analysis. Mol Biol Evol 12:814–822

De Pinna MCC (1996) Teleostean monophyly. In: Stiassny MLJ, Parenti LR, Johnson GD (eds) Interrelationships of fish. Academic Press, San Diego

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Felsenstein J (1989) Phylogenetic inference programs (PHYLIP). University of Washington, Seattle, and University Herbarium, University of California, Berkeley

Fink SV, Fink WL (1996) Interrelationships of ostariophysan fish (Teleostei). In: Stiassny MLJ, Parenti LR, Johnson GD (eds) Interrelationships of fish. Academic Press, San Diego

Gadagkar SR, Rosenberg MS, Kumar S (2005) Inferring species phylogenies from multiple genes:Concatenated sequence tree versus consensus gene tree. Mol Dev Evol 304B:64–74

Gerhold D, Caskey CT (1996) It's the genes! EST access to human genome content. Bioessays 18:973–981

Greenwood PH, Rosen DE, Weitzman SH, Mayers GS (1966) Phyletic studies of teleostean fish, 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 52:696–704

Helfman GS, Collette BB, Facey DE (1997) The diversity of fish Blackwell Science, Cambridge, MA

Hillis DM (1998) Taxonomic sampling, phylogenetic accuracy, and investigator bias. Syst Biol 47:3–8

Hillis DM, Pollock DD, McGuire JA, Zwickl DJ (2003) Is sparse taxon sampling a problem for phylogenetic inference? Syst Biol 52:124–126

Hoegg S, Brinkmann H, Taylor JS, Meyer A (2004) Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. J Mol Evol 59:190–203

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17:754–755

Hughes J, Longhorn SJ, Papadopoulou A, Theodorides K, de Riva A , Mejia-Chang M, Foster PG, Vogler AP (2006) Dense taxonomic EST sampling and its applications for molecular systematics of the Coleoptera (beetles). Mol Biol Evol 23:268–278

Inoue JG, Miya M, Tsukamoto K, Nishida M (2004) Mitogenomic evidence for the monophyly of elopomorph fish (Teleostei) and the evolutionary origin of the leptocephalus larva. Mol Phylogenet Evol 32:274–286

Inoue JG, Miya M, Venkatesh B, Nishida M (2005) The mitochondrial genome of Indonesian coelacanth *Latimeria menadoensis* (Sarcopterygii: Coelacanthiformes) and divergence time estimation between the two coelacanths. Gene 349:227–235

Ishiguro NB, Miya M, Nishida M (2003) Basal euteleostean relationships: a mitogenomic perspective on the phylogenetic reality of the "Protacanthopterygii." Mol Phylogenet Evol 27:476–488

Jaillon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A, Nicaud S, Jaffe D, Fisher S, Lutfalla G, Dossat C, Segurens B, Dasilva C, Salanoubat M, Levy M, Boudet N, Castellano S, Anthouard V, Jubin C, Castelli V, Katinka M, Vacherie B, Biemont C, Skalli Z, Cattolico L, Poulain J, De Berardinis V, Cruaud C, Duprat S, Brottier P, Coutanceau JP, Gouzy J, Parra G, Lardier G, Chapple C, McKernan KJ, McEwan P, Bosak S, Kellis M, Volff JN, Guigo R, Zody MC, Mesirov J, Lindblad-Toh K, Birren B, Nusbaum C, Kahn D, Robinson-Rechavi M, Laudet V, Schachter V, Quetier F, Saurin W, Scarpelli C, Wincker P, Lander ES, Weissenbach J, Roest Crollius H (2004) Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. Nature 431:946–957

Johnson GD, Patterson C (1993) Percomorph phylogeny: a survey of acanthomorphs and a new proposal. Bull Mar Sci 52:554–626

Johnson GD, Patterson C (1996) Relationships of lower euteleostean fish. In: Stiassny MLJ, Parenti LR, Johnson GD (eds) Interrelationships of fish. Academic Press, San Diego

Kruiswijk CP, Hermsen TT, Westphal AH, Savelkoul FJ, Steti RJM (2002) A novel functional class I cineage in Zebrafish (Danio rerio), Carp (Cyprinus carpio), and large Barbus (Barbus intermedius) showing an unusual conservation of the peptide binding domains. J Immunol 169:1936–1947

Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief Bioinform 5:150–163

Kumar S, Gadagkar SR (2000) Efficiency of the neighbor-joining method in reconstructing deep and shallow evolutionary relationships in large phylogenies. J Mol Evol 51:544–553

Kumazawa Y, Yamaguchi M, Nishida M (1999). Mitochondrial molecular clocks and the origin of euteleostean biodiversity: familial radiation of Perciforms may have predated the Cretaceous/Tertiary boundary. In: Kato M (ed) The biology of biodiversity. Springer, Hong Kong

Lake LA, Moore JE (1998) Phylogenetic analysis and comparative genomics. Trends guide to Bioinformatics, Trends Supplement 1998:22–23

Larget B, Simon DL (1999) Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. Mol Biol Evol 16:750–759

Mitchell A, Mitter C, Regier JC (2000) More taxa or more characters revisited: combining data from nuclear protein-encoding genes for phylogenetic analyses of Noctuoidea (Insecta: Lepidoptera). Syst Biol 49:202–224

Miya M, Nishida M (2000) Use of mitogenomic information in teleostean molecular pyhlogenetics: a tree-based exploration under the maximum-parsimony optimality criterion. Mol Phylogenet Evol 17:437–455

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

Mossel E, Vigoda E (2005) Phylogenetic MCMC algorithms are misleading on mixtures of trees. Science 309:2207–2209

Near TJ, Sanderson MJ (2004) Assessing the quality of molecular divergence time estimates by fossil calibrations and fossil-based model selection. Phil Trans R Soc Lond B Biol Sci 359:1477–1483

Near TJ, Meylan PA, Shaffer HB (2005) Assessing concordance of fossil calibration points in molecular clock studies: an example using turtles. Am Nat 165:137–146

Nelson J (1994) Fishes of the world. Wiley, New York

Notredame C, Higgins DG, Heringa J (2000) T-Coffee: a novel method for fast and accurate multiple sequence alignment. J Mol Biol 302:205–217

Renn SCP, Aubin-Horth N, Hofmann HA (2004) Biologically meaningful expression profiling across species using heterologous hybridization to a cDNA microarray. BMC Genomics 5:42

Resetnikov JS (1988) Coregonid fish in recent conditions. Finnish Fish Res 9:11–16

Rokas A, Carroll SB (2005) More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. Mol Biol Evol 22:1337–1344

Rokas A, William BL, King N, Carroll SB (2003) Genome scale approaches to resolving incongruence in molecular phylogenies. Nature 425:798–804

Rokas A, Kruger D, Carroll SB (2005) Animal evolution and the molecular signature of radiations compressed in time. Science 310:1933–1938

Rosen DE (1973) Interrelationships of higher teleostean fish. In: Greenwood PH, Miles RS, Patterson C (eds) Interrelationships of fish. Academic Press, London

Rosen DE (1974) Phylogeny and zoogeography of salmoniform fish and relationships of Lepidogalaxias salamandroides. Bull Am Mus Nat Hist 153:265–326

Rosen DE, Greenwood PH (1970) Origin of the Weberian apparatus and the relationships of the ostariophysan and gonorynchiform fish. Am Mus Novit 2428:1–25

Rosen DE, Patterson C (1969) The structure and relationships of the paracanthopterygian fish. Bull Am Mus Nat Hist 141:357–474

Rosenberg MS, Kumar S (2001) Incomplete taxon sampling is not a problem for phylogenetic inference. Proc Natl Acad Sci USA 98:10751–10756

Rosenberg MS, Kumar S (2003) Taxon sampling, bioinformatics, and phylogenomics. Syst Biol 52:119–124

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

Sanderson MJ (2003) r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics 19:301–302

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

Shevchuk NA, Allard MW (2001) Sources of incongruence among mammalian mitochondrial sequences: COII, COIII, and ND6 genes are main contributors. Mol Phylogenet Evol 21:43–54

Shimodaira H (2002) An approximately unbiased test of phylogenetic tree selection. Syst Biol 51:492–508

Shimodaira H, Hasegawa M (2001) CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17:1246–1247

Simmons MP, Miya M (2004) Efficiently resolving the basal clades of a phylogenetic tree using Bayesian and parsimony approaches: a case study using mitogenomic data from 100 higher teleost fish. Mol Phylogenet Evol 31:351–362

Steinke D, Salzburger W, Meyer A (2004) EverEST—A phylogenomic EST database approach. Phyloinformatics 6:1–4

Stepien CA, Kocher TD (1997) Molecules and morphology in studies of fish evolution. In: Kocher TD, Stepien CA (eds) Molecular systematics of fish. Academic Press, San Diego

Swofford DL (2002) PAUP*. Phylogenetic analysis using parsimony (*and other methods), version 4.10b Sinauer Associates, Sunderland, MA

Takezaki N, Figueroa F, Zaleska-Rutczynska Z, Klein J (2003) Molecular phylogeny of early vertebrates: monophyly of the Agnathans as revealed by sequences of 35 genes. Mol Biol Evol 20:287–292

Takezaki N, Figueroa F, Zaleska-Rutczynska Z, Takahata N, Klein J (2004) The phylogenetic relationship of tetrapod, coelacanth, and lungfish revealed by the sequences of forty-four nuclear genes. Mol Biol Evol 21:1512–1524

Taylor JS, Braasch I, Frickey T, Meyer A, van de Peer Y (2003) Genome duplication, a trait shared by 22,000 species of ray-finned fish. Genome Res 13:382–390

Watanabe M, Kobayashi N, Shin-i T, Horiike T, Tateno Y, Kohara Y, Okada N (2004) Extensive analysis of ORF sequences from two different cichlid species in Lake Victoria provides molecular evidence for a recent radiation event of the Victoria species flock identity of EST sequences between *Haplochromis chilotes* and *Haplochromis* sp. ''Redtailsheller.'' Gene 343:263–269

Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol 18:691–699

Wiley EO, Johnson GD, Dimmick WW (2000) The interrelationships of Acantomorph fish: a total evidence approach using molecular and morphological data. Biochem Syst Ecol 28:319–350