Molecular Phylogenetics and Evolution 50 (2009) 163–178

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

# Molecular resolution of the systematics of a problematic group of fishes (Teleostei: Osmeridae) and evidence for morphological homoplasy

Katriina L. Ilves *, Eric B. Taylor

*Department of Zoology, Biodiversity Research Centre, Native Fishes Research Group, University of British Columbia, 6270 University Boulevard, Vancouver, BC, Canada V6T 1Z4*

ARTICLE INFO

*Article history:*
Received 16 July 2008
Revised 22 October 2008
Accepted 25 October 2008
Available online 5 November 2008

*Keywords:*
Osmeridae
Smelt
Plecoglossidae
Salangidae
Homoplasy
Mitochondrial and nuclear DNA
Systematics

ABSTRACT

Relationships among the species of Northern Hemisphere smelts (family Osmeridae) have long been debated in the fish systematics literature. Eight independent studies based on morphological characters failed to reach any consensus on osmerid interrelationships. We reconstruct the osmerid phylogeny based on DNA sequence data from three mitochondrial (cyt*b*, 16S, 12S) and three nuclear (ITS2, S71, RAG1) gene regions from multiple individuals of the 14 species in 6 genera, using the Japanese ayu (*Plecoglossus altivelis*) as the outgroup. Analyses with different combinations of nuclear and mitochondrial datasets yielded a generally well-resolved phylogeny of the genera that conflicts with previous hypotheses of osmerid interrelationships, and Shimodaira–Hasegawa tests suggest our topology with the current molecular dataset is significantly better than earlier reconstructions. In addition, mapping 114 morphological characters used in previous studies onto our phylogeny shows widespread homoplasy, which is likely the source of the systematic disagreement produced in earlier works.



## 1. Introduction

The Northern Hemisphere smelts (Osmeridae) are small (<30 cm), elongate, silvery fishes that display diverse life-history characteristics with marine, anadromous and freshwater forms, often within the same species. Osmerids are planktivores found in near-shore marine and adjacent freshwater environments and are important forage fishes for a number of fishes and marine mammals. The family has a Holarctic distribution, meaning they are found in cool-temperate and Arctic waters throughout the Northern Hemisphere. Although the family has a widespread distribution, all species have more limited distributions, generally along single coastlines of the North Pacific and North Atlantic oceans, with the exception of the Holarctic capelin (*Mallotus villosus*) and the Arctic/North Pacific Arctic rainbow smelt (*Osmerus dentex*). The species composition and phylogenetic relationships among the taxa variously attributed to the Osmeridae have been much debated in fish systematics (e.g. McAllister, 1963, 1966; Begle, 1991; Wilson and Williams, 1991; Johnson and Patterson, 1996). In this study we conducted a molecular phylogenetic analysis of all osmerid species and evaluated the phylogenetic content of the morphological characters used in previous systematic analyses of the family. The osmerids are an important group to examine given their widespread geographic distribution and because systematic uncertainty occurs at several taxonomic levels.

Linnaeus (1758) originally classified the smelts within the salmonid genus *Salmo*, and although Cuvier (1817) eventually introduced a separate osmerid genus, *Osmerus*, the grouping of the smelts was still retained within the family Salmonidae. The osmerids continued to receive systematic and taxonomic attention throughout the early 20th century (Hubbs, 1925; Kendall, 1927; Chapman, 1941); the most recent comprehensive revision of the Osmeridae to date is that of McAllister (1963). McAllister's (1963) systematic revision of the entire family was based on meristic and morphometric characteristics and changed the number of recognized species from 14–16 to 10, and included the naming of a new species and subspecies.

However monumental, McAllister's (1963) study did not quell the debate about osmerid interrelationships, as the uncertainty surrounding the phylogenetic placement and biogeography of this group continues to the present day. Although the systematic relationships among the teleost fishes have been the focus of more research effort than any other vertebrate group (Parenti, 1986; Begle, 1991), as noted by Waters et al. (2002), the difficulties encountered by evolutionary biologists studying smelts led Johnson and Patterson (1996) to lament that the "osmerids are unique in the disparity of opinion on their interrelationships".

Under the current Linnean classification system, osmerids fall into the order Osmeriformes and superfamily Osmeroidea (Nelson, 2006) with the Southern Hemisphere Retropinnidae and Galaxii-

* Corresponding author. Present address: Biology Department, Queens College, City University of New York, 65-30 Kissena Blvd., Flushing, NY 11367, USA. Fax: +1 718 997 3445.
*E-mail address:* katriina.ilves@qc.cuny.edu (K.L. Ilves).

1055-7903/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2008.10.021

dae. Below the level of superfamily there is considerable disagreement about how the various putative families and genera are related. The eight published phylogenies of the osmeroids are based on morphological characteristics (Fig. 1). The phylogenies disagree about the relationships among the species and genera, and how the monotypic Plecoglossidae, the Japanese ayu (*Plecoglossus altivelis*), and Salangidae, a temperate and subtropical family of icefishes from the northwestern Pacific, are related to the Osmeridae (Johnson and Patterson, 1996). Uncertainty also extends to the relationships among the Osmeridae and the Southern

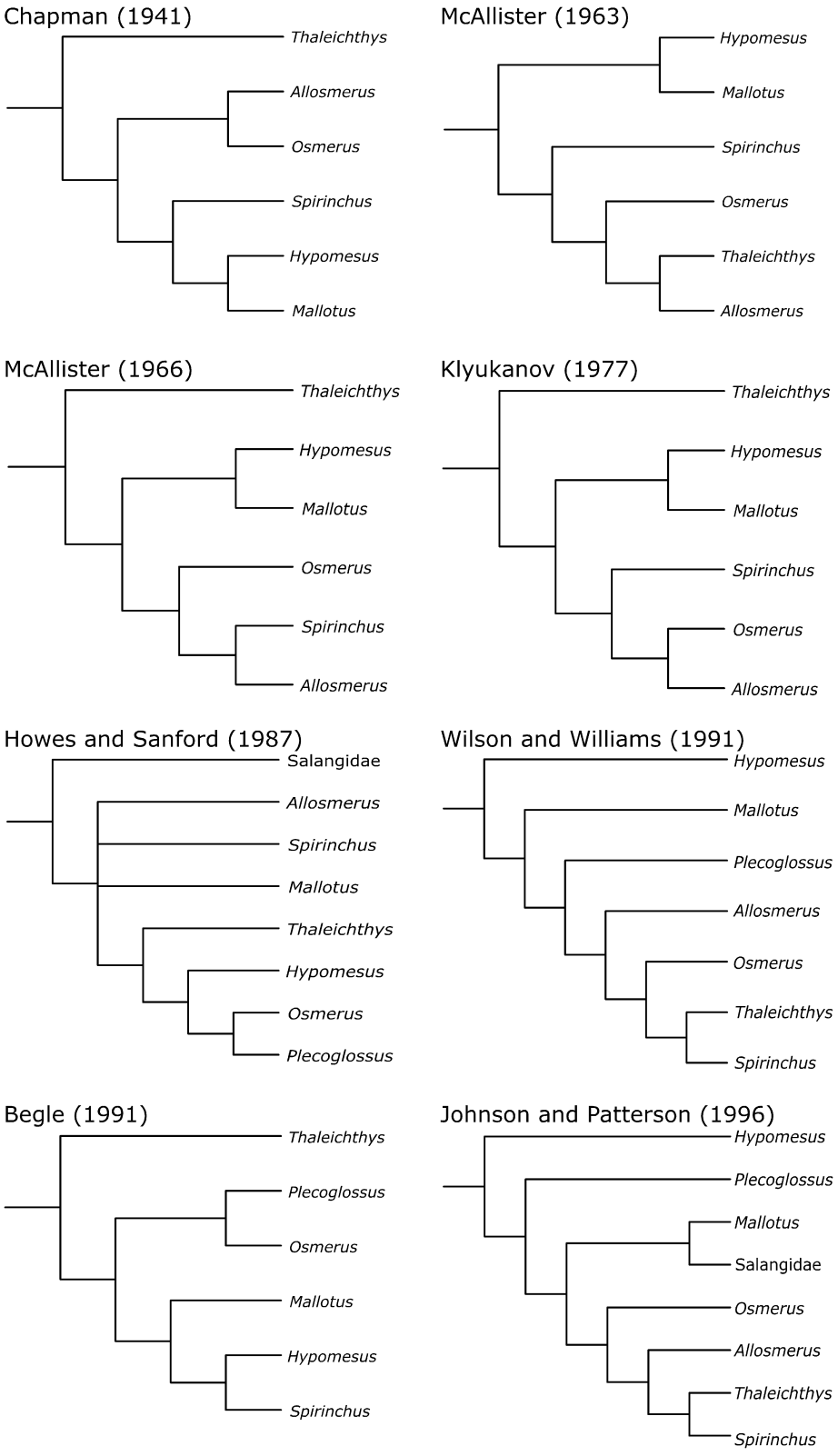


**Fig. 1.** Eight morphology-based hypotheses of systematic relationships among osmerid genera.

Hemisphere smelts and galaxiids (families Retropinnidae and Galaxiidae, respectively). Eschmeyer (2006) lists the Plecoglossidae and Salangidae as separate families, a classification that also has historical support (Chapman, 1941; McAllister, 1963; Klyukanov, 1975); however, other morphological analyses have either nested both families within the Osmeridae (Howes and Sanford, 1987; Johnson and Patterson, 1996) or found the closest affinities of the Salangidae to be with the Southern Hemisphere group, with *Plecoglossus* being more closely related to the Osmeridae (Greenwood et al., 1966; Roberts, 1984; Begle, 1991). Nelson (2006) considers *Plecoglossus* and the salangids as subfamilies within the Osmeridae.

Recent molecular analyses have increased our understanding of the interrelationships among these families, although full agreement has not yet been reached (Fig. 2). Waters et al. (2002) found that the Osmeridae (Northern Hemisphere) and Retropinnidae (Southern Hemisphere) together are sister to the Galaxiidae (Southern Hemisphere). Subsequent phylogenetic analyses (López et al., 2004; Fu et al., 2005) have confirmed that the two Southern Hemisphere families are not sister taxa and also found a more complex relationship among the other families (Fig. 2). Using the same gene regions as Waters et al. (2002) but with more thorough taxonomic sampling of the Salangidae and Osmeridae, Fu et al. (2005) found a poorly-supported sister relationship between these families with parsimony analysis (Fig. 2). In a broader lower euteleostean context, further analyses of mitochondrial (12S, 16S) and nuclear sequences (RAG1) showed strongly supported sister relationships between the Salangidae and Plecoglossidae, and these two families with the Osmeridae (López et al., 2004), which Fu et al. (2005) also found using Bayesian analysis of two mtDNA gene regions (Fig. 2). López et al. (2004) also found surprising evidence that the stomiiform fishes, not the Galaxiidae, are sister to the clade containing the Retropinnidae, Plecoglossidae, Salangidae, and Osmeridae.

Of the six osmerid genera, *Spirinchus*, *Osmerus*, and *Hypomesus* each contain more than one species. While *Spirinchus* has received relatively little systematic attention, the relationships within *Osmerus* and *Hypomesus* have been extensively debated. *Osmerus* in particular has been the focus of considerable taxonomic effort, both morphological and molecular. This genus has a Holarctic distribution, found in near-shore marine habitats and coastal freshwaters of the Pacific, Atlantic, and Arctic Ocean drainages. Studies of the relationships among *Osmerus* species have been complicated by unresolved taxonomic issues. Depending on the evolutionary relationships supported by various studies, the geographic variants have been designated as subspecies: *O. dentex*, *O. eperlanus eperlanus*, and *O. e. mordax* (McAllister, 1963; Luey et al., 1982), or *O. eperlanus*, *O. mordax mordax*, and *O. m. dentex* (Klyukanov, 1969, cited in Scott and Crossman, 1973; McAllister et al., 1980). For this study we considered the three forms as separate species (Nellbring, 1989; Taylor and Dodson, 1994).

A general consensus on the relationships among the species of *Hypomesus* has also been elusive (e.g. McAllister, 1963; Klyukanov, 1970; Saruwatari et al., 1997; Ilves and Taylor, 2007, 2008). Although there are currently six recognized species within the genus, recent morphological (Sidorov and Pichugin, 2004) and molecular analyses (Ilves and Taylor, 2007) found that a newly identified species from the Kuril Islands of Japan, *H. chishimaensis* (Saruwatari et al., 1997), is indistinguishable from *H. nipponensis*. Therefore, in this study we considered the two species as conspecific and contained within *H. nipponensis*.

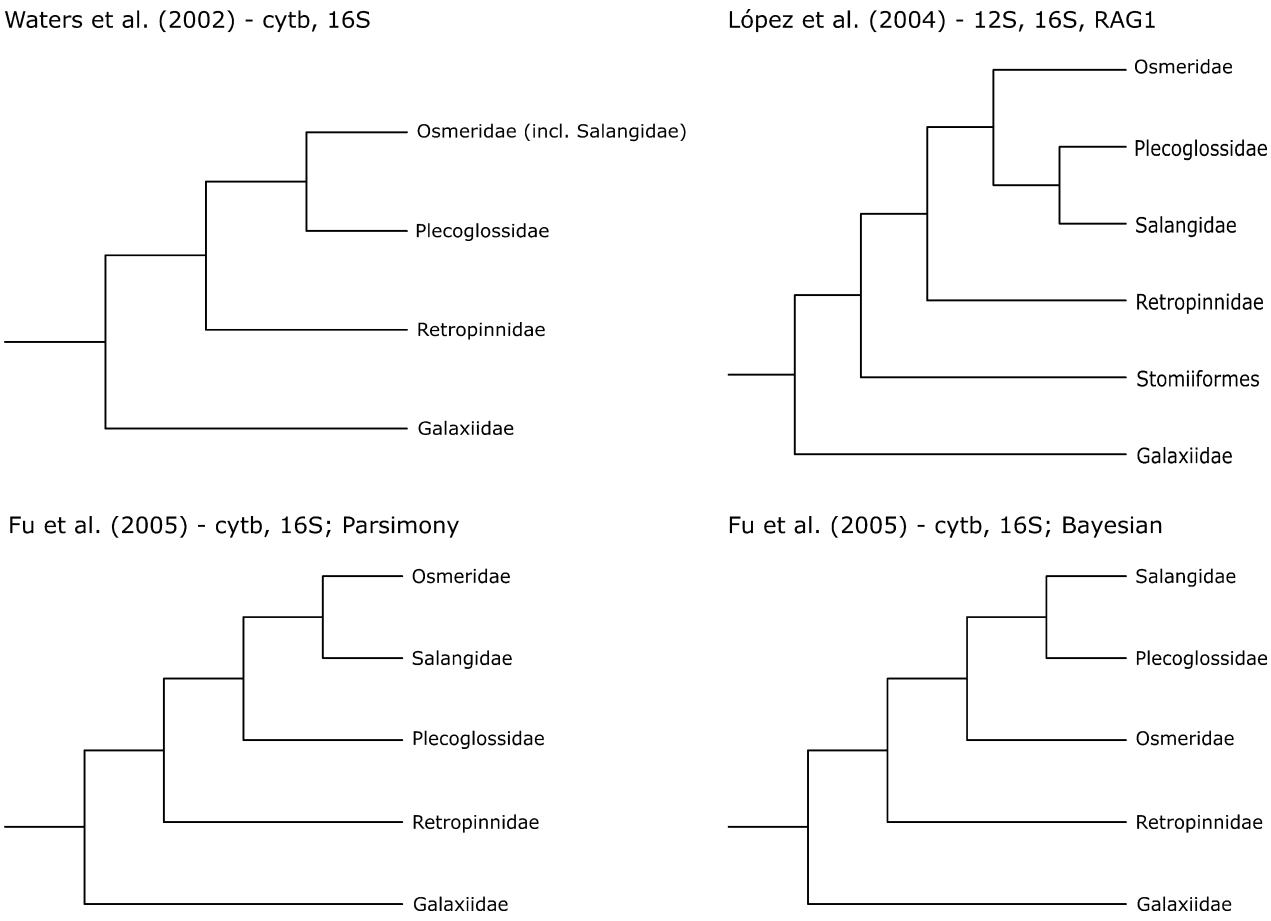


**Fig. 2.** Molecular phylogenies that have included osmerid, salangid and galaxiid taxa.

The aims of our study were to (1) produce a molecular phylogeny of the Osmeridae based on three mitochondrial and three nuclear gene regions, (2) compare the resulting phylogeny to previous hypotheses of osmerid interrelationships, and (3) with the assumption that the molecular phylogeny accurately reflects the osmerid phylogeny, examine the evolution of characters used in previous studies and assess their utility for determining the systematic relationships within the Osmeridae.

## 2. Materials and methods

### 2.1. Taxon sampling

To take into account potential problems with lineage sorting (Maddison, 1997), 4–25 individuals per species from several localities were sequenced for at least a subset of the six gene regions ($N = 440$ sequences; Table 1). The samples were obtained from at least two geographic locations per species, with the following exceptions: (1) *Hypomesus transpacificus*, which is endemic to the Sacramento—San Joaquin River basin, CA, (2) *H. japonicus*, for which samples were only available from a single location in Hokkaido, Japan, (3) *Spirinchus lanceolatus*, found only along eastern Hokkaido, and (4) the outgroup *P. altivelis*, for which samples were only available from Lake Biwa, Japan.

### 2.2. DNA sequence data

DNA extraction, PCR, and sequencing protocols followed Ilves and Taylor (2007) with the following additions and modifications. The mitochondrial 12S gene was amplified with the primers 12SF and 12SR (Table 2) in 50 μl reactions containing 50–300 ng of genomic DNA and final concentrations of 800 μM of dNTPs, 800 nM of each primer, 1.25 U of New England Biolabs (NEB) *Taq* DNA polymerase, 20 mM Tris–HCl (pH 8.4), 50 mM KCl, and 2.5 mM $MgCl_2$ under the following conditions: 95 °C for 5 min, followed by 30 cycles of 1 min each of 95 °C, 50 °C and 72 °C, and a final extension at 72 °C for 5 min.

Three species (*Osmerus dentex*, *O. eperlanus*, and *Plecoglossus altivelis*) were amplified with primers from López et al. (2004). *O. dentex* was amplified with primers RAG1F1 and RAG1R1 (Table 2) in 25 μl reactions with final concentrations of 800 μM of dNTPs, 800 nm of each primer, 1 U of NEB *Taq* DNA polymerase, 20 mM Tris–HCl (pH 8.4), 50 mM KCl, and 2.5 mM $MgCl_2$ under the following conditions: 94 °C for 3 min, followed by 35 cycles of 45 s at 95 °C, 45 s at 52 °C and 1 min 15 s at 72 °C, and a final extension at 72 °C for 5 min. *O. eperlanus* and *P. altivelis* were amplified with the primers RAG1F1 and RAG1R2 (Table 2) with the same reaction conditions and PCR profile described for *O. dentex*, except a 53 °C annealing temperature was used for *P. altivelis*. The mitochondrial 16S and nuclear ITS2 regions were amplified with the primers and conditions described in Ilves and Taylor (2007), except that a 55 °C annealing temperature was used for most samples.

### 2.3. Phylogenetic analyses

#### 2.3.1. Sequence alignment and general phylogenetic methods

Sequences were aligned using ClustalX (Thompson et al., 1997) or manually with MacClade v4.06 (Maddison and Maddison, 2003), and edited with Se-Al v2.0a11 (Rambaut, 1996) or MacClade v4.06 (Maddison and Maddison, 2003). Alignments of protein-coding genes cyt*b* and RAG1 and the 12S gene region were unambiguous. By contrast, positional homology, due to the presence of indels and repeats in the 16S rDNA, ITS2 and S71 regions, was difficult to determine. As a result, ambiguously aligned characters were excluded from analyses.

Data were analyzed by gene region and by different data partitions: cyt*b* + 16S, all mitochondrial (mtDNA) data (cyt*b* + 16S + 12S), all nuclear (nDNA) data (ITS2 + S71 + RAG1) and all data combined (allDNA). There were two combined mtDNA partitions because 12S sequences were only available for a subset of samples. Modeltest v3.6 (Posada and Crandall, 1998) was used to select a model of sequence evolution for each data partition (individual genes and combined data) for neighbor-joining, maximum likelihood (ML) and Bayesian analyses. For each analysis we used the best-fit model chosen by the Akaike Information Criterion (AIC) method (Posada and Buckley, 2004). Neighbor-joining, parsimony, and ML phylogenetic analyses were all implemented in PAUP* v.4.0b10 (Swofford, 2002). Bayesian analysis was conducted using MrBayes v.3.1.2 (Huelsenbeck and Ronquist, 2001). The models corresponding to each partition were as follows: TrN + I + G (cyt*b* and cyt*b* + 16S), TrNef + I (16S), TVMef + I + G (12S), TVM + I + G (ITS2 and all mtDNA), TrN + G (S71 and RAG1), GTR + I + G (all nDNA and allDNA).

Due to the large amount of sequence data (Table 1), NJ and Bayesian methods were used to identify individuals of each species that represented a range of the intraspecific variation. From these individuals we constructed a reduced dataset on which we conducted more rigorous ML and parsimony analyses. The first step of analysis included 1000 NJ bootstrap pseudo-replicates and Bayesian analyses ($2 \times 10^6$ generations, burn-in 2000, 18,000 sampled trees) of all sequences for each gene separately (Supplementary Figs. 1–6). We then created a reduced dataset, consisting of two to three individuals per species that included a range of haplotypic variation based on branching patterns within each species for each gene. In most cases the same individuals were used for all genes, although in a few cases a sequence from a conspecific was substituted in order to capture intraspecific variation and sequence quality (Supplementary Table 1). Due to computational limits of ML analysis of the combined mtDNA, nDNA, and allDNA datasets, a single individual was chosen as an exemplar for each species. NJ, Bayesian, and parsimony analyses were also performed for these further reduced partitions to compare topologies resulting from taxon removal and different reconstruction methods.

Support for monophyly was assessed with 1000 bootstrap pseudo-replicates (Felsenstein, 1985) for NJ, parsimony, and ML analyses, and with posterior probabilities for Bayesian analyses. We considered nodes with bootstrap values $\geqslant 70\%$ (Hillis and Bull, 1993) and Bayesian posterior probabilities $\geqslant 95\%$ to be well-supported. Heuristic searches for parsimony and ML analyses were conducted with 10 random replicates of stepwise taxon addition; five replicates were used for ML bootstrap analyses. NJ and Bayesian analyses were conducted on all datasets, and parsimony bootstrap analysis was used for all reduced datasets. ML heuristic searches were conducted for all reduced data partitions, and ML bootstrapping was performed for the combined mtDNA, nDNA, and allDNA (mtDNA + nDNA) partitions using a single individual per species. Bayesian analysis for all datasets was run in two parallel analyses ($2 \times 10^6$ generations, trees sampled every 100 generations, burn-in of 2000). Stabilization of likelihoods apparently occurred by 2000 generations, meaning that all trees sampled are from well after this stationarity.

#### 2.3.2. Outgroup choice

Our preliminary analyses supported choosing *Plecoglossus altivelis* as a suitable outgroup as this species falls outside the Osmeridae and is one of the most closely related taxa to this group. NJ, parsimony, and Bayesian analyses of *cytb*, 16S, 12S, and RAG1 sequences from *Galaxias fasciatus*, *Retropinna retropinna*, *Salangichthys microdon*, *P. altivelis*, and one individual from each of the Osmeridae 'proper' were conducted. We used GenBank sequences for *G. fasciatus*, *S. microdon* and 12S of *R. retropinna* (Table 1). Addi-

**Table 1**
List of samples used in this study, including species, sample ID, museum voucher information, sample locality and GenBank Accession Nos.

| Species (sample ID) | Museum Voucher | Locality | GenBank Accession Nos | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | cyt*b* | 16S | 12S | ITS2 | S71 | RAG1 |
| *Allosmerus elongatus* (AE4L) | BC Uncat. | BC coast, Canada | EU621530 | EU621407 | EU621477 | EU621441 | EU621503 | EU621576 |
| *Allosmerus elongatus* (AEBMS1) | BC Uncat. | BC, Imperial Eagle Channel, Canada | EU621531 | EU621408 | EU621478 | EU621442 | EU621504 | EU621577 |
| *Allosmerus elongatus* (AEOR1) | BC Uncat. | OR coast, USA | EU621532 | – | – | EU621443 | – | – |
| *Allosmerus elongatus* (AEOR2) | BC Uncat. | OR coast, USA | EU621533 | – | – | EU621444 | – | – |
| *Hypomesus japonicus* (HJ5) | BC Uncat. | Hokkaido, off Ohtanoshike, near Kushiro, Japan | DQ010183* | DQ010215* | EU621479 | DQ010254* | DQ010284* | EU621578 |
| *Hypomesus japonicus* (HJ6) | BC Uncat. | Hokkaido, off Ohtanoshike, near Kushiro, Japan | DQ010184* | DQ010216* | – | DQ010255* | DQ010285* | DQ010305* |
| *Hypomesus japonicus* (HJ7) | BC Uncat. | Hokkaido, off Ohtanoshike, near Kushiro, Japan | DQ010185* | DQ010217* | EU621480 | DQ010256* | DQ010286* | DQ010306* |
| *Hypomesus japonicus* (HJ8) | BC Uncat. | Hokkaido, off Ohtanoshike, near Kushiro, Japan | DQ010186* | DQ010218* | – | DQ010257* | DQ010287* | DQ010307* |
| *Hypomesus japonicus* (HJ9) | BC Uncat. | Hokkaido, off Ohtanoshike, near Kushiro, Japan | DQ010187* | DQ010219* | – | DQ010258* | DQ010288* | DQ0103081* |
| *Hypomesus japonicus* (HJ10) | BC Uncat. | Hokkaido, off Ohtanoshike, near Kushiro, Japan | DQ010188* | DQ010220* | – | DQ010259* | DQ010289* | DQ010309* |
| *Hypomesus nipponensis* [*Hypomesus chishimaensis*] (HC1) | UW 043710 | Iturup Island, Kuybyshevskoe Lake, Japan | DQ010175* | DQ010202* | – | DQ010234* | DQ010262* | DQ010293* |
| *Hypomesus nipponensis* [*Hypomesus chishimaensis*] (HC2) | UW 043710 | Iturup Island, Kuybyshevskoe Lake, Japan | DQ010176* | DQ010203* | – | DQ010235* | DQ010263* | DQ010294* |
| *Hypomesus nipponensis* [*Hypomesus chishimaensis*] (HC3) | UW 043710 | Iturup Island, Kuybyshevskoe Lake, Japan | DQ010177* | DQ010204* | – | DQ010236* | DQ010264* | DQ010295* |
| *Hypomesus nipponensis* [*Hypomesus chishimaensis*] (HC4) | UW 043710 | Iturup Island, Kuybyshevskoe Lake, Japan | DQ010178* | DQ010205* | – | DQ010237* | DQ010265* | DQ010296* |
| *Hypomesus nipponensis* [*Hypomesus chishimaensis*] (HC5) | UW 041862 | Kunashir Island, Lake Serebryanoye, Japan | DQ010179* | DQ010206* | – | DQ010238* | DQ010266* | – |
| *Hypomesus nipponensis* [*Hypomesus chishimaensis*] (HC6) | UW 041862 | Kunashir Island, Lake Serebryanoye, Japan | DQ010180* | DQ010207* | – | DQ010239* | DQ010267* | DQ010297* |
| *Hypomesus nipponensis* [*Hypomesus chishimaensis*] (HC7) | UW 041862 | Kunashir Island, Lake Serebryanoye, Japan | – | DQ010208* | – | DQ010240* | DQ010268* | – |
| *Hypomesus nipponensis* [*Hypomesus chishimaensis*] (HC10) | UW 041869 | Zelionyi Island, stream of Lake Srednoye, Japan | – | DQ010209* | – | DQ010241* | DQ010269* | – |
| *Hypomesus nipponensis* [*Hypomesus chishimaensis*] (HC11) | UW 041869 | Zelionyi Island, stream of Lake Srednoye, Japan | – | DQ010210* | – | DQ010242* | DQ010270* | – |
| *Hypomesus nipponensis* [*Hypomesus chishimaensis*] HC12) | UW 041869 | Zelionyi Island, stream of Lake Srednoye, Japan | – | DQ010211* | – | DQ010243* | – | – |
| *Hypomesus nipponensis* [*Hypomesus chishimaensis*] (HC18) | UW 046336 | Sakhalin Island, Shlyuzovka River, Russia | DQ010181* | DQ010212* | – | DQ010244* | DQ010271* | DQ010298* |
| *Hypomesus nipponensis* [*Hypomesus chishimaensis*] (HC19) | UW 046336 | Sakhalin Island, Shlyuzovka River, Russia | DQ010182* | DQ010213* | – | DQ010245* | DQ010272* | – |
| *Hypomesus nipponensis* [*Hypomesus chishimaensis*] (HC20) | UW 046336 | Sakhalin Island, Shlyuzovka River, Russia | – | DQ010214* | – | DQ010246* | DQ010273* | – |
| *Hypomesus nipponensis* (HN1) | BC Uncat. | Hokkaido, Harutori River, Japan | DQ010189* | DQ010221* | – | DQ847513* | DQ010274* | DQ847515* |
| *Hypomesus nipponensis* (HN2) | BC Uncat. | Hokkaido, Harutori River, Japan | DQ836403* | DQ836424* | – | DQ836442* | DQ836460* | DQ836479* |
| *Hypomesus nipponensis* (HN3) | BC Uncat. | Hokkaido, Harutori River, Japan | DQ836404* | DQ836425* | – | DQ836443* | DQ836461* | DQ836480* |
| *Hypomesus nipponensis* (HN5) | BC Uncat. | Hokkaido, Harutori River, Japan | DQ010190* | DQ010222* | EU621481 | DQ010247* | DQ010275* | DQ010299* |
| *Hypomesus nipponensis* (HN6) | BC Uncat. | Hokkaido, Harutori River, Japan | DQ010191* | DQ010223* | – | DQ010248* | DQ010276* | DQ010300* |
| *Hypomesus nipponensis* (HN7) | BC Uncat. | Hokkaido, Harutori River, Japan | DQ010192* | DQ010224* | EU621482 | DQ010249* | DQ010277* | DQ010301* |
| *Hypomesus nipponensis* (HN8) | BC Uncat. | Hokkaido, Harutori River, Japan | DQ010193* | DQ010225* | – | DQ010250* | DQ010278* | – |
| *Hypomesus nipponensis* (HN9) | BC Uncat. | Hokkaido, Harutori River, Japan | DQ010194* | DQ010226* | – | DQ847514* | DQ010279* | – |
| *Hypomesus nipponensis* (HN11) | BC Uncat. | Hokkaido, off Ohtanoshike, near Kushiro, Japan | DQ836399* | DQ836420* | – | DQ836438* | DQ836456* | DQ836477* |
| *Hypomesus nipponensis* (HN12) | BC Uncat. | Hokkaido, off Ohtanoshike, near Kushiro, Japan | DQ836400* | DQ836421* | – | DQ836439* | DQ836457* | DQ836478* |

(*continued on next page*)

**Table 1** (*continued*)

| Species (sample ID) | Museum Voucher | Locality | GenBank Accession Nos | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | cyt*b* | 16S | 12S | ITS2 | S71 | RAG1 |
| *Hypomesus nipponensis* (HN13) | BC Uncat. | Hokkaido, off Ohtanoshike, near Kushiro, Japan | DQ836401[+] | DQ836422[+] | – | DQ836440[+] | DQ836458[+] | – |
| *Hypomesus nipponensis* (HN14) | BC Uncat. | Hokkaido, off Ohtanoshike, near Kushiro, Japan | DQ836402[+] | DQ836423[+] | – | DQ836441[+] | DQ836459[+] | – |
| *Hypomesus olidus* (HO1) | UW 043724 | Kamchatka, Yavinskoye Lake, Russia | DQ010195[*] | DQ010227[*] | – | – | DQ010280[*] | DQ010302[*] |
| *Hypomesus olidus* (HO2) | UW 043724 | Kamchatka, Yavinskoye Lake, Russia | DQ010196[*] | DQ010228[*] | EU621483 | DQ010251[*]5 | DQ010281[*] | – |
| *Hypomesus olidus* (HO3) | UW 043724 | Kamchatka, Yavinskoye Lake, Russia | DQ010197[*] | DQ010229[*] | – | DQ010252[*] | DQ010282[*] | DQ010303[*] |
| *Hypomesus olidus* (HO4) | UW 043724 | Kamchatka, Yavinskoye Lake, Russia | DQ010198[*] | DQ010230[*] | EU621484 | DQ010253[*] | DQ010283[*] | DQ010304[*] |
| *Hypomesus olidus* (HOAK1) | BC Uncat. | AK, Chignik Lake, USA | DQ836405[+] | DQ836426[+] | – | DQ836444[+] | DQ836462[+] | – |
| *Hypomesus olidus* (HOAK2) | BC Uncat. | AK, Chignik Lake, USA | DQ836406[+] | – | – | – | DQ836463[+] | – |
| *Hypomesus olidus* (HOAK3) | BC Uncat. | AK, Chignik Lake, USA | DQ836407[+] | – | – | – | DQ836464[+] | – |
| *Hypomesus olidus* (HOAK4) | BC Uncat. | AK, Chignik Lake, USA | DQ836408[+] | – | – | – | DQ836465[+] | – |
| *Hypomesus pretiosus* (HP1) | BC Uncat. | AK Sumner Strait, USA | DQ836409[+] | DQ836427[+] | EU621486 | DQ836445[+] | DQ836466[+] | DQ8364812[+] |
| *Hypomesus pretiosus* (HP3b) | BC Uncat. | BC, Wreck Beach, Vancouver, Canada | DQ010199[+] | DQ010232[+] | EU621485 | DQ010260[+] | DQ010290[+] | DQ010310[+] |
| *Hypomesus pretiosus* (HP4) | BC Uncat. | BC Coast, Canada | – | EU621409 | – | – | – | – |
| *Hypomesus pretiosus* (HP5) | BC Uncat. | BC Coast, Canada | – | EU621410 | – | – | – | – |
| *Hypomesus pretiosus* (HP7) | BC Uncat. | BC Coast, Canada | – | EU621411 | – | – | – | – |
| *Hypomesus pretiosus* (HPBMS2) | BC Uncat. | BC, Imperial Eagle Channel, Canada | DQ836412[+] | DQ836430[+] | – | DQ836448[+] | DQ836469[+] | – |
| *Hypomesus pretiosus* (HPJB1) | BC Uncat. | Puget Sound, San Juan Islands, WA, USA | DQ836410[+] | DQ836428[+] | – | DQ836446[+] | DQ836467[+] | – |
| *Hypomesus pretiosus* (HPJB2) | BC Uncat. | Puget Sound, San Juan Islands, WA, USA | DQ836411[+] | DQ836429[+] | – | DQ836447[+] | DQ836468[+] | – |
| *Hypomesus transpacificus* (HT1) | BC Uncat. | Sacramento River, CA, USA | DQ836413[+] | DQ836431[+] | EU621488 | DQ836449[+] | DQ836470[+] | – |
| *Hypomesus transpacificus* (HT2) | BC Uncat. | Sacramento River, CA, USA | DQ010200[+] | DQ010231[+] | – | DQ010261[+] | DQ010291[+] | DQ0103113[+] |
| *Hypomesus transpacificus* (HT3) | BC Uncat. | Sacramento River, CA, USA | DQ836414[+] | DQ836432[+] | EU621487 | DQ836450[+] | DQ836471[+] | – |
| *Hypomesus transpacificus* (HT4) | BC Uncat. | Sacramento River, CA, USA | DQ836415[+] | DQ836433[+] | – | DQ836451[+] | DQ836472[+] | DQ836482[+] |
| *Hypomesus transpacificus* (HT5) | BC Uncat. | Sacramento River, CA, USA | DQ836416[+] | DQ836434[+] | – | DQ836452[+] | DQ836473[+] | DQ836483[+] |
| *Hypomesus transpacificus* (HT6) | BC Uncat. | Sacramento River, CA, USA | EU621534 | – | – | – | – | – |
| *Hypomesus transpacificus* (HT7) | BC Uncat. | Sacramento River, CA, USA | DQ836417[+] | DQ836435[+] | – | DQ836453[+] | DQ836474[+] | DQ8364844[+] |
| *Hypomesus transpacificus* (HT8) | BC Uncat. | Sacramento River, CA, USA | DQ836418[+] | DQ836436[+] | – | DQ836454[+] | DQ836475[+] | DQ836485[+] |
| *Mallotus villosus* (MV1) | BC Uncat. | BC, Central Coast, Canada | EU621537 | – | – | EU621446 | EU621505 | |
| *Mallotus villosus* (MV2) | BC Uncat. | BC, Central Coast, Canada | DQ836419[+] | DQ836437[+] | EU621490 | DQ836455[+] | DQ836476[+] | DQ836486[+] |
| *Mallotus villosus* (MV3) | BC Uncat. | BC, Central Coast, Canada | EU621535 | EU621412 | – | – | – | |
| *Mallotus villosus* (MVTC2) | BC Uncat. | Atlantic Ocean, Canada | DQ397093[*] | DQ397094[*] | EU621489 | DQ397095[*] | DQ397096[*] | DQ397097[*] |
| *Mallotus villosus* (MVAT1) | BC Uncat. | Atlantic Ocean, Canada | EU621536 | EU621413 | – | EU621445 | – | – |
| *Osmerus dentex* (ODAK27) | BC Uncat. | AK, USA | EU621539 | EU621416 | – | EU621455 | – | – |
| *Osmerus dentex* (ODAK28) | BC Uncat. | AK, USA | EU621540 | – | – | – | – | – |
| *Osmerus dentex* (ODAK31) | BC Uncat. | AK, USA | EU621538 | EU621414 | EU621491 | EU621452 | – | – |
| *Osmerus dentex* (ODAK35) | BC Uncat. | AK, USA | EU621541 | EU621415 | – | EU621453 | – | – |
| *Osmerus dentex* (ODAK43) | BC Uncat. | AK, USA | EU621542 | – | – | – | – | – |
| *Osmerus dentex* (ODBER1) | UW 112173 | AK, Bering Sea, USA | EU621543 | – | – | EU621456 | – | – |
| *Osmerus dentex* (ODSAK1) | UW 044763 | Sakhalin Island, environs of Lake Spenskoye, Russia | EU621544 | EU621417 | EU621492 | EU621454 | EU621506 | EU621579 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *Osmerus eperlanus* (OE1) | BC Uncat. | Lake Ijsselmeer, Netherlands | EU621545 | EU621418 | – | EU621457 | EU621507 | – |
| *Osmerus eperlanus* (OE14) | BC Uncat. | Lake Ijsselmeer, Netherlands | EU621546 | EU621419 | EU621494 | EU621458 | EU621508 | EU621581 |
| *Osmerus eperlanus* (OERT) | BC Uncat. | Kryonjoki River, Finland | U05667 | – | – | – | – | – |
| *Osmerus eperlanus* (OELK1) | BC Uncat. | Lake Kolovesi, Finland | EU621547 | – | – | EU621461 | EU621510 | – |
| *Osmerus eperlanus* (OELP1) | BC Uncat. | Lake Paasivesi, Finland | EU621549 | EU621420 | EU621493 | EU621459 | EU621509 | EU621580 |
| *Osmerus eperlanus* (OELP2) | BC Uncat. | Lake Paasivesi, Finland | EU621550 | – | – | – | – | – |
| *Osmerus eperlanus* (OEPJ1) | BC Uncat. | Lake Peipsi, Estonia | EU621548 | – | – | EU621460 | – | – |
| *Osmerus mordax* (OM13) | BC Uncat. | NB, Canada | EU621551 | EU621421 | – | EU621449 | EU621511 | EU621582 |
| *Osmerus mordax* (OM19) | BC Uncat. | NB, Canada | EU621552 | EU621422 | – | EU621447 | – | – |
| *Osmerus mordax* (OMLH3) | BC Uncat. | QC, Lac Heney, Canada | EU621553 | EU621423 | EU621496 | EU621448 | EU621512 | EU621583 |
| *Osmerus mordax* (OMLH18) | BC Uncat. | QC, Lac Heney, Canada | EU621554 | – | – | EU621451 | – | – |
| *Osmerus mordax* (OMML8) | BC Uncat. | ON, Meech Lake, Canada | EU621555 | – | EU621495 | EU621450 | EU621513 | – |
| *Osmerus mordax* (OMRT) | BC Uncat. | ON, Canada | U05666 | – | – | – | – | – |
| *Spirinchus lanceolatus* (SL1) | BC Uncat. | Pacific coast, Japan | EU621556 | EU621424 | – | EU621462 | EU621514 | EU621587 |
| *Spirinchus lanceolatus* (SL2) | BC Uncat. | Pacific coast, Japan | EU621557 | EU621425 | – | EU621463 | EU621515 | EU621588 |
| *Spirinchus lanceolatus* (SL3) | BC Uncat. | Pacific coast, Japan | EU621558 | – | – | EU621464 | – | – |
| *Spirinchus lanceolatus* (SL4) | BC Uncat. | Pacific coast, Japan | EU621559 | EU621426 | – | – | EU621516 | – |
| *Spirinchus lanceolatus* (SL5) | BC Uncat. | Pacific coast, Japan | EU621560 | EU621427 | EU621497 | EU621465 | EU621517 | EU621589 |
| *Spirinchus starksi* (SSVI1) | BC Uncat. | BC, Vancouver Island, Canada | – | – | – | EU621467 | EU621518 | – |
| *Spirinchus starksi* (SSVI2) | BC Uncat. | BC, Vancouver Island, Canada | EU621561 | EU621428 | – | – | – | – |
| *Spirinchus starksi* (SSCA1) | BC Uncat. | CA coast, USA | EU621562 | EU621429 | EU621498 | EU621466 | EU621520 | EU621590 |
| *Spirinchus starksi* (SSCA2) | BC Uncat. | CA coast, USA | EU621563 | EU621430 | – | EU621468 | EU621519 | – |
| *Spirinchus starksi* (SSPUG1) | UW 048781 | WA, Puget Sound, USA | EU621564 | EU621431 | – | – | EU621521 | – |
| *Spirinchus thaleichthys* (STAN1) | BC Uncat. | BC, Canada | EU621565 | EU621432 | EU621499 | EU621469 | EU621522 | EU621585 |
| *Spirinchus thaleichthys* (STAN2) | BC Uncat. | BC, Canada | EU621566 | EU621435 | – | – | EU621523 | – |
| *Spirinchus thaleichthys* (STFR1) | BC Uncat. | BC, Fraser River, Canada | EU621569 | – | – | EU621471 | – | – |
| *Spirinchus thaleichthys* (STHL1) | BC Uncat. | BC, Harrison Lake, Canada | EU621568 | EU621434 | – | – | – | – |
| *Spirinchus thaleichthys* (STLW1) | BC Uncat. | WA, Lake Washington, USA | EU621567 | EU621433 | EU621500 | EU621470 | EU621524 | EU621586 |
| *Thaleichthys pacificus* (TPBR2) | BC Uncat. | AK, Bering Sea, USA | EU621571 | EU621437 | – | EU621473 | EU621526 | |
| *Thaleichthys pacificus* (TPBMS2) | BC Uncat. | BC, Bamfield, Canada | EU621570 | EU621438 | – | EU621472 | EU621525 | EU621591 |
| *Thaleichthys pacificus* (TPCOWA) | BC Uncat. | WA, Cowlitz River, USA | EU621574 | EU621436 | EU621501 | EU621474 | EU621527 | EU621592 |
| *Thaleichthys pacificus* (TPPUG1) | UW 048013 | WA, Puget Sound, USA | EU621572 | EU621439 | – | EU621475 | – | – |
| *Thaleichthys pacificus* (TPPUG2) | UW 048014 | WA, Puget Sound, USA | EU621573 | – | – | – | EU621528 | – |
| *Plecoglossus altivelis* (PLEC3) | UW 012704 | Honshu, Lake Biwa, Japan | DQ010201 | DQ010233 | – | – | DQ010292 | – |
| *Plecoglossus altivelis* (PLEC4) | UW 012704 | Honshu, Lake Biwa, Japan | EU621575 | EU621440 | EU621502 | EU621476 | EU621529 | EU621584 |
| *Retropinnia retropinna* (RR1) | BC Uncat. | New Zealand | FJ392549 | FJ392551 | – | FJ392553 | FJ392555 | FJ392557 |
| *Retropinnia retropinna* (RR2) | BC Uncat. | New Zealand | FJ392550 | FJ392552 | – | FJ392554 | FJ392556 | FJ392558 |
| *Retropinnia retropinna* | – | See GenBank Accession | – | – | NC004598‡ | – | – | – |
| *Galaxias fasciatus* | – | See GenBank Accessions | AF267350‡ | AF112333‡ | AY430265‡ | – | – | AY430218‡ |
| *Salangichthys microdon* | – | See GenBank Accessions | AF454838‡ | AY443566‡ | AY430267‡ | – | – | AY380539‡ |

* Identify sequences from Ilves and Taylor (2007, 2008).
+ Identify sequences from Ilves and Taylor (2007, 2008).
‡ Indicates sequences obtained from GenBank.

**Table 2**
Primers used to amplify the six gene regions sequenced in this study.

| Gene | Primer | Sequence (5′–3′) | Source |
|---|---|---|---|
| cyt*b* | cyt*b*2 | CCC TCA GAA TGA TAT TTG TCC TCA | Kocher et al. (1989) |
| | GluDG | TGA CTT GAA RAA CCA YCG TTG | |
| 16S | 16Sar | CGC CTG TTT ATC AAA AAC AT | Waters et al. (2002) |
| | 16Sbr | CCG GTC TGA ACT CAG ATC ACG T | |
| 12S | 12SF | AAA AAG CTT CAA ACT GGG ATT AGA TAC CCC ACT | Kocher et al. (1989) |
| | 12SR | TGA CTG CAG AGG GTG ACG GGC GGT GTG T | |
| ITS2 | 5.8sr | CTA CGC CTG TCT GAG TGT C | Presa et al. (2002) |
| | 28s | ATA TGC TTA AAT TCA GCG GG | |
| S71 | S7RPEX1F | TGG CCT CTT CCT TGG CCG TC | Chow and Hazama (1998) |
| | S7RPEX2R | AAC TCG TCT GGC TTT TCG CC | |
| RAG1 | RAG1F | AGC TGT AGT CAG TAY CAC AAR ATG | Quenouille et al. (2004) |
| | RAG9R | GTG TAG AGC CAG TGR TGY TT | |
| | RAG1F1 | CTG AGC TGC AGT CAG TAC CAT AAG ATG T | López et al. (2004) |
| | RAG1R1 | CTG AGT CCT TGT GAG CTT CCA TRA AYT T | |
| | RAG1R2 | TGA GCC TCC ATG AAC TTC TGA AGR TAY TT | |
| | RAG1R3 | GTC TTG TGS AGG TAG TTG GT | |

tional sequences for *R. retropinna*, *P. altivelis*, and all species of the Osmeridae were from Ilves and Taylor (2007, 2008) or were generated for the current study (Table 1).

The mitochondrial data suggested a sister relationship between *Plecoglossus altivelis* and the salangids and nested this clade within the Osmeridae with very low posterior probability (0.51; Supplementary Fig. 7A). Analysis of RAG1, on the other hand, showed that *P. altivelis* was a member of the Osmeridae sister group with a posterior probability of 0.98, likely as sister to the Salangidae (Supplementary Fig. 7B). Similarly, using *Retropinna retropinna* as the outgroup, analysis of all individual and combined mtDNA (cyt*b*, 16S, 12S) and nDNA (ITS2, S71, RAG1) gene regions sequenced in this study confirmed that *P. altivelis* was not nested within the Osmeridae (data not shown). Furthermore, Bayesian analysis of the nDNA and allDNA datasets with *R. retropinna* as the outgroup yielded virtually identical topologies to the analysis with *P. altivelis* as the outgroup, the only difference being that the placement of *Mallotus villosus* was unresolved from analysis of the allDNA data-

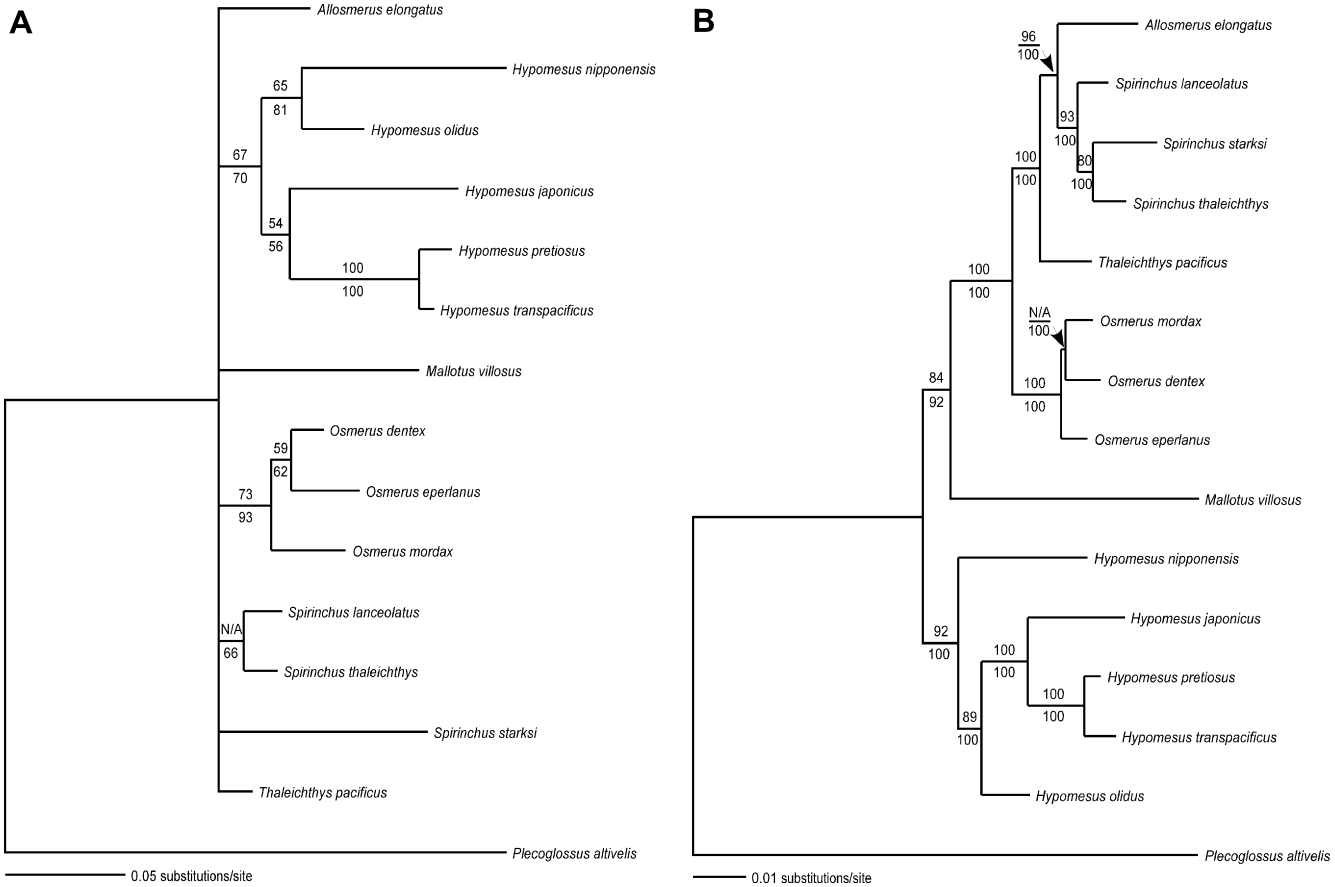


**Fig. 3.** (A) Combined mtDNA (cyt*b*, 16S, 12S) and (B) combined nuclear DNA (ITS2, S71, RAG1) osmerid phylogenies based on a single individual/species resulting from Bayesian and maximum likelihood (ML) reconstruction. Numbers above nodes represent support values from 1000 bootstrap pseudo-replicates (ML). Numbers below nodes represent posterior probabilities (%) from a consensus of 18,000 trees (Bayesian analysis).

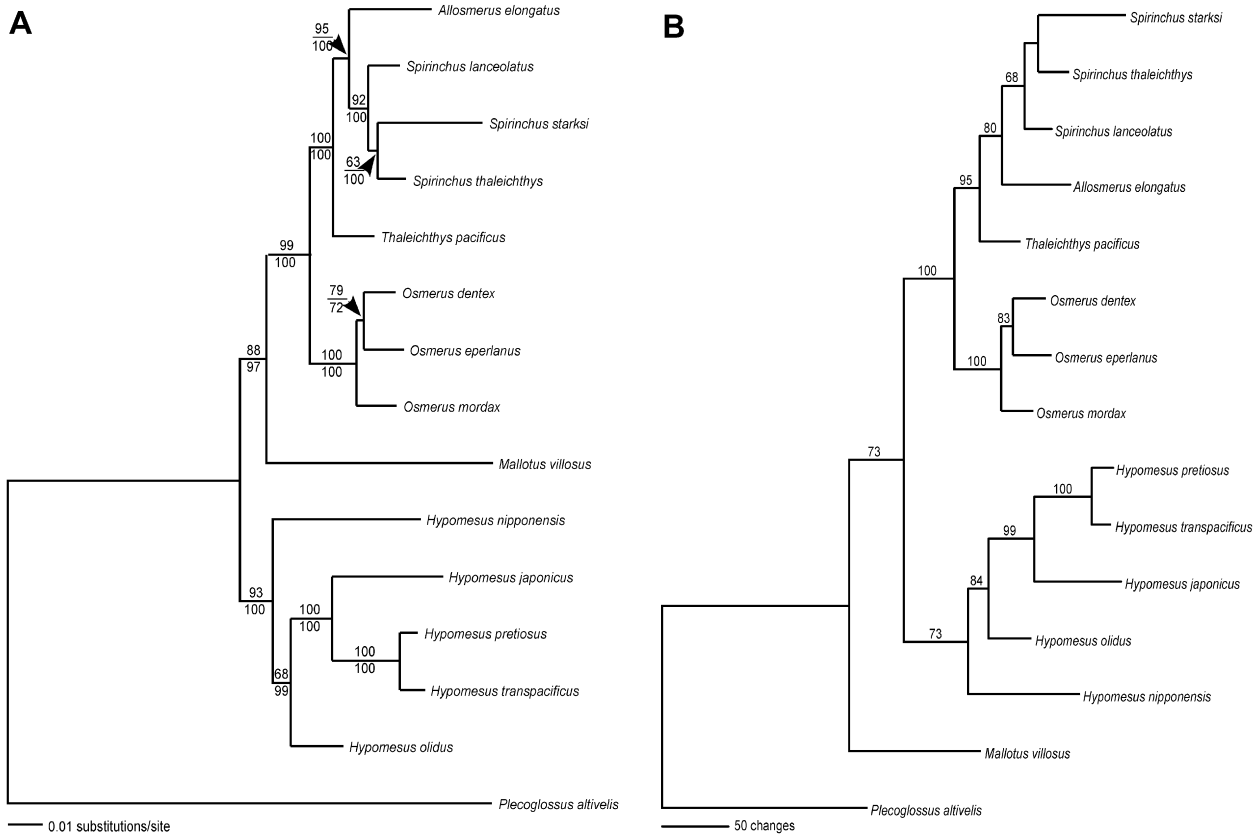


**Fig. 4.** Combined mtDNA (cyt*b*, 16S, 12S) and nDNA (ITS2, S71, RAG1) Osmeridae phylogenies resulting (A) Bayesian and maximum likelihood (ML) reconstruction and (B) parsimony reconstruction. Numbers above nodes represent support values from 1000 bootstrap pseudo-replicates (ML, parsimony) or posterior probabilities (%) from a consensus of 18,000 trees (Bayesian analysis).

**Table 3**
Results of Shimodaira–Hasegawa tests (Shimodaira and Hasegawa, 1999) comparing the molecular topology (Fig. 4A) with earlier morphology-based hypotheses (Fig. 1).

| Comparison | −ln likelihood | *P*-value |
|---|---|---|
| *Without Plecoglossus* | | |
| allDNA phylogeny | 10482.1[+] | |
| Chapman (1941) | 10518.5 | 0.005[*] |
| McAllister (1963) | 10519.1 | 0.004[*] |
| McAllister (1966) | 10502.5 | 0.024[*] |
| Klyukanov (1977) | 10518.5 | 0.005[*] |
| *With Plecoglossus* | | |
| allDNA phylogeny | 11825.8[+] | |
| Wilson and Williams (1991) | 11859.9 | 0.027[*] |
| Begle (1991) | 11896.0 | 0.0004[*] |
| *With Salangidae* | | |
| cyt*b*, 16S, 12S, RAG1 phylogeny | 8681.5[+] | |
| Howes and Sanford (1987) | 8709.8 | 0.015[*] |
| Johnson and Patterson (1996) | 8699.1 | 0.045[*] |

[+] Indicates highest likelihood score.
[*] Indicates significant difference ($P < 0.05$).

set (data not shown). Although the use of multiple outgroups would be ideal, particularly for examining osmerid–salangid–retropinnid–galaxiid relationships, sequences for two of the nDNA genes (ITS2, S71) used in this study were not available for galaxiid or salangid species.

### 2.3.3. Partition homogeneity and topology tests

Partition homogeneity tests (Farris et al., 1994; Swofford, 2002) were performed on the single individual datasets to determine the validity of using a single concatenated matrix for the mtDNA, nDNA, and allDNA partitions. These tests were implemented in PAUP* v.4.0b10 (Swofford, 2002) with 1000 replicates using a heuristic search with 10 replicates of random sequence addition and TBR branch swapping.

We also used Shimodaira–Hasegawa (SH) tests (Shimodaira and Hasegawa, 1999) in cases of conflict for the placement of particular taxa, and also to compare the molecular phylogeny of the Osmeridae to other hypotheses. The SH test compares the likelihoods of alternative topologies for a given dataset and was used to evaluate whether some topologies were significantly better than others (Felsenstein, 2004). This test was implemented in PAUP* v.4.0b10 (Swofford, 2002) using the likelihood model selected by the AIC method in Modeltest v3.6 (Posada and Crandall, 1998) for the corresponding dataset, with 10,000 RELL (re-estimated log-likelihood) bootstrap replicates.

SH tests were conducted on the mtDNA, nDNA, and allDNA datasets, using *Plecoglossus altivelis* as the outgroup, to compare alternative placements of *Mallotus villosus* in the osmerid phylogeny and conflicting relationships within the polytypic *Osmerus*, *Spirinchus*, and *Hypomesus*. Alternative outgroups were also used (*Retropinna retropinna* for *M. villosus*, *M. villosus* for *Osmerus*, and *Allosmerus* for *Spirinchus*).

To compare our molecular phylogeny of the Osmeridae with eight previous morphological hypotheses of their interrelationships, we compared the molecular topology generated from Bayesian and ML analysis of the allDNA dataset to the topologies shown in Fig. 1. Because four of the phylogenies in Fig. 1 did not include *Plecoglossus altivelis*, this taxon was removed from the molecular phylogeny for those comparisons. In the phylogeny of Johnson and Patterson (1996), the Salangidae were embedded within the Osmeridae. To test the molecular phylogeny with the Salangidae and Plecoglossidae as sister taxa, against their topology, we used

a dataset consisting of concatenated cyt*b*, 16S, 12S, and RAG1 sequences. Data for these gene regions, but not ITS2 or S71, were available from GenBank for the salangid *Salangichthys microdon*. We also tested these topologies with only RAG1 sequence data because the mitochondrial data provided little phylogenetic resolution (see Section 3).

#### 2.3.4. Species tree estimation

Phylogenetic analyses were also performed to estimate a species tree from the joint posterior distribution of gene trees from the three DNA and three nDNA gene regions. This analysis was performed using BEST version 1.6 (Liu and Pearl, 2007; Liu et al., 2008), which uses a Bayesian hierarchical model to calculate the weights for individual gene trees to produce a species tree. The BEST analysis was conducted on the reduced individual allDNA data partition. The same substitution models were used as for Bayesian phylogenetic inference. A single run with one chain was conducted for $5 \times 10^6$ generations with trees sampled every 1000 generations. We assessed convergence of the chains by examining stationarity of log-likelihood values against MCMC generation; of the 5000 gene trees generated in each analysis, 20% were excluded as burn-in for the second part of the BEST analysis, which calculates the posterior distribution of species trees. A species tree was generated from a 50% majority-rule consensus of the estimated posterior distribution of species trees.

### 2.4. Character mapping

Under the assumption that the molecular phylogeny is the "true" osmerid phylogeny, we examined character evolution by mapping the morphological characters used by McAllister (1963, 1966), Begle (1991), Wilson and Williams (1991), Johnson and Patterson (1996), and Patterson and Johnson's (1997) correction of Begle's (1991) matrix onto the molecular topology using MacClade (Maddison and Maddison, 2003). Johnson and Patterson's (1996) analysis included a hypothetical outgroup with all states coded as '0' and considered some multistate characters as ordered. For our analysis the outgroup was coded as '?' for characters from other studies and we examined the effect of maintaining and removing ordering on the pattern of character evolution. Begle (1991) and Johnson and Patterson (1996) included additional taxa not thoroughly sampled in the current study (e.g. retropinnids and galaxiids), therefore, their matrices were scanned for characters that showed variation among the Osmeridae and Plecoglossidae species sequenced for the molecular analysis. The most comprehensive morphological assessment conducted to date is that of Johnson and Patterson (1996), who noted differences between their analysis and those of McAllister (1963, 1966), Howes and Sanford (1987), Begle (1991), and Wilson and Williams (1991) in the coding of some characters. In places of disagreement among studies, we included Johnson and Patterson's (1996) and Patterson and Johnson's (1997) character state designations as these are the most thorough analyses. Further, because Johnson and Patterson's (1996) phylogeny included the Salangidae and Plecoglossidae nested within the Osmeridae, we separately reviewed the characters used in their 1996 and 1997 (Patterson and Johnson) studies with a phylogeny that includes the Salangidae. Supplementary Tables 3 and 4 contain the data matrix for the entire character dataset and the characters and their state names, respectively.

## 3. Results

### 3.1. Sequences

Due to ambiguous sequence alignments in the 16S, ITS2, and S71 matrices, 28, 242, and 65 characters were excluded from the respective datasets with all sequences; however, Bayesian analysis conducted on the single individual allDNA dataset with all characters included yielded an identical topology with almost identical posterior probabilities, suggesting exclusion of sites did not impact the phylogenetic conclusions (data not shown). Because 12S sequences were not available for all individuals used in the reduced analyses based on 2–3 individuals per species, combined mtDNA and allDNA partitions included only *cytb* and 16S sequences; however, the further reduced partitions with only a single individual representing each species included 12S. Including indels, the reduced mtDNA (*cytb* + 16S), nDNA, and mtDNA + nDNA datasets contained 959, 2597, and 3556 characters, respectively, while the mtDNA (cyt*b* + 16S + 12S), nDNA, and allDNA datasets based on a single individual contained 1352, 2597, and 3949 characters, respectively. The number of variable and parsimony-informative sites for each gene region is listed in Supplementary Table 2. Additional data, such as distance matrices and transition/transversion ratios can be obtained from the authors. Sequences generated for this study have been submitted to GenBank (Table 1) and phylogenetic trees from analysis of the mtDNA, nDNA, and allDNA datasets have been submitted to TreeBase (Accession Nos. SN4169-20298, SN4169-20299, and SN4169-20300, respectively).

### 3.2. Species monophyly

Analyses of all datasets showed that all species were reciprocally monophyletic for each gene region in most cases (Supplementary Figs. 1–6). *Hypomesus pretiosus* and *H. transpacificus* were not reciprocally monophyletic for any genes apart from S71, and some other species were not monophyletic for 16S (*Thaleichthys pacificus*), 12S (*T. pacificus*), ITS2 (*O. dentex*, *O. mordax*, *S. lanceolatus*), and S71 (*S. starski*, *S. thaleichthys*) (Supplementary Figs. 1–6). Phylogenies from the complete and reduced datasets (Supplementary Figs. 1–6, Figs. 3 and 4, respectively) generally recovered the same topologies, with some disagreement at poorly supported nodes. Consequently, it is unlikely that sequence choice for the reduced datasets had a substantial effect on the main results and conclusions from analysis of reduced datasets.

### 3.3. Individual gene analyses

Analysis of the individual genes generally resulted in poorly resolved topologies (Supplementary Figs. 1–6). Of the six genes, topologies from the nuclear S71 and RAG1 regions (Supplementary Figs. 5 and 6, respectively) most closely matched those of the combined dataset analyses. Although the intergeneric relationships were not fully resolved, the monophyly of the three polytypic genera, *Hypomesus*, *Osmerus*, and *Spirinchus*, was supported by multiple gene regions (Supplementary Figs. 1–3, 5, and 6). Analyses of individual genes also showed apparent conflicts within these polytypic genera, and in the position of *Mallotus villosus*. Analysis of ITS2 yielded relationships not seen in any other analyses, although with very low levels of support (Supplementary Fig. 4).

### 3.4. Combined dataset analyses

Partition homogeneity tests showed no significant conflict among the mtDNA genes ($P = 0.54$) or the nDNA ($P = 0.41$) gene regions, but did suggest conflict between the mtDNA and nDNA datasets ($P = 0.01$). Despite this significant difference, we considered it acceptable to combine all of the data for analysis for two main reasons: there was little phylogenetic resolution from the mtDNA dataset and the nDNA and allDNA data resulted in the same topology apart from poorly supported relationships in *Osmerus* (discussed below), which suggests a significance level at $P = 0.05$ is too stringent for these data (Wiens, 1998).

The phylogenies from the combined mtDNA (cyt*b*, 16S, 12S), nDNA and allDNA datasets are shown in Figs. 3A, 3B, and 4, respectively. Analysis of the single individual/species mtDNA dataset did not produce a fully resolved phylogeny, but showed some support for the monophyly of *Hypomesus* and *Osmerus* (Fig. 3A). By contrast, the nDNA and allDNA phylogenies are well-supported (Figs. 3B and 4).

The different methods of phylogenetic reconstruction for the nDNA and allDNA datasets yielded the same intergeneric relationships, apart from the interesting exception of *Mallotus villosus*. For both of these partitions, *M. villosus* was sister to the (*Osmerus*, (*Thaleichthys*, (*Allosmerus*, *Spirinchus*))) clade [hereby designated as OTAS] using NJ, ML, and Bayesian methods (Figs. 3B and 4A), but was sister to all genera with parsimony reconstruction (Fig. 4B). All phylogenetic reconstruction methods of the nDNA and allDNA datasets show *Allosmerus* and *Spirinchus* as the most recently diverged sister genera, which are sister to *Thaleichthys* (Figs. 3B and 4). *Osmerus* was sister to this latter clade and *Hypomesus* was sister to the rest of the genera (Figs. 3B and 4A). Support values for these relationships were generally high across reconstruction methods, with Bayesian posterior probabilities of 0.92–1.0 and ML bootstrap values 84–100% for the nDNA dataset (Fig. 3B), and 0.97–1.0 and 88–100% for the allDNA dataset (Fig. 4A).

### 3.5. Osmerid species tree

The species tree estimated from the reduced allDNA dataset by BEST analysis (Liu and Pearl, 2007; Liu et al., 2008) did not contradict the results of the concatenated sequences, although it was not fully resolved (Supplementary Fig. 8). *Mallotus* appeared as sister to the OTAS clade with very low (53%) support, and the placement of *Spirinchus* was unresolved (Supplementary Fig. 8).

### 3.6. Placement of Mallotus villosus and relationships within Osmerus, Spirinchus, and Hypomesus

The placement of *Mallotus villosus* as either the sister to the OTAS clade or to all of the genera is unclear. SH tests on the nDNA and allDNA datasets using *Plecoglossus altivelis* as the outgroup showed that the topology with *M. villosus* sister to the OTAS clade had a higher likelihood ($-\ln L = 7484.8$ vs. 7484.9; $-\ln L = 11825.8$ vs. 11829.4, respectively), whereas the allDNA dataset using *Retropinna retropinna* as the outgroup, the alternative topology had a higher likelihood ($-\ln L = 12516.0$ vs. 12516.7). Not surprisingly, the likelihood values did not significantly differ ($0.13 \leqslant P \leqslant 0.61$).

Relationships among the three *Osmerus* and *Spirinchus* species were ambiguous; different gene regions and data partitions yielded different relationships within each genus. For *Osmerus*, an *O. dentex–O. eperlanus* relationship was well-supported by ML and parsimony analysis of the allDNA dataset (Fig. 4A). Similarly, in *Spirinchus*, an *S. starksi–S. thaleichthys* relationships was moderately to highly supported in the S71, nDNA and allDNA partitions (Supplementary Fig. 5, Figs. 3B, 4A, respectively). SH tests among the alternative topologies for each genus, however, did not find significant differences ($P > 0.05$ in all comparisons).

With respect to *Hypomesus*, a previous phylogenetic analysis (Ilves and Taylor, 2008) was unable to resolve the placement of *H. olidus* as either sister to *H. nipponensis* or to the (*H. japonicus*, (*H. pretiosus*, *H. transpacificus*)) clade. This uncertainty also appears to extend to the family-level phylogeny, where different partitions supported both arrangements. An *H. olidus–H. nipponensis* sister relationship was poorly-supported (Fig. 3A; Supplementary Figs. 3 and 6), while there was mixed support for a topology with *H. nipponensis* sister to the other *Hypomesus* species (Figs. 3B, 4; Supplementary Fig. 5). The latter topology had a higher likelihood with the nDNA and allDNA partitions, while the former topology had a higher likelihood under the mtDNA data; however, none of the likelihoods were significantly different (SH tests, $P > 0.05$ in all comparisons).

### 3.7. Comparison of molecular phylogeny and previous hypotheses

Results of the SH tests showed that the molecular phylogeny had a significantly higher likelihood than all eight previous morphological hypotheses of Osmeridae interrelationships (Table 3). When Johnson and Patterson's (1996) and the molecular phylogenies were compared with only RAG1 sequences, support for the molecular topology increased ($P = 0.032$). Excluding *Mallotus villosus*, the molecular placement of which was uncertain, yielded the same results (data not shown).

### 3.8. Character mapping

An initial matrix consisting of 155 morphological characters was generated from McAllister (1963, 1966), Begle (1991), Wilson and Williams (1991), Johnson and Patterson (1996), and Patterson and Johnson (1997). After scanning this matrix for duplicate characters and incorporating the character deletions and coding changes suggested by Johnson and Patterson (1996) and Patterson and Johnson (1997), a final matrix of 114 characters was mapped onto the molecular phylogeny (Figs. 3B and 4). A trace of all characters with unambiguous changes onto the molecular tree was 204 steps (Fig. 5), and 193 steps for the Johnson and Patterson (1996) topology (Fig. 6) [Salangidae data not available for most characters]. Considering all characters from Johnson and Patterson (1996) unordered had no effect on these tree lengths.

Assuming that the molecular topology is the "true" species tree for the Osmeridae, our analysis revealed numerous homoplasies in the morphological characters studied (Fig. 5); however, there was also morphological support for most of the relationships in the molecular phylogeny. Two shared character states unambiguously united the (*Mallotus*, (*Osmerus*, (*Thaleichthys*, (*Allosmerus*, Spirinchus)))) clade, 20 character states defined the (*Osmerus*, (*Thaleichthys*, (*Allosmerus*, *Spirinchus*))) clade, and five character states were shared by *Thaleichthys*, *Allosmerus*, and *Spirinchus* (Fig. 5). Of the 114 characters, none exclusively defined the *Allosmerus–Spirinchus* sister relationship, although they share the 'no cucumber odor' state of character 64, along with *Plecoglossus* and the salangids, and have a head length of 4.7 or greater than standard length (character 100), which is also shared by at least one species of *Osmerus* (state of this character is missing for *P. altivelis*). *Osmerus* apparently lacks unique character states in this dataset.

## 4. Discussion

### 4.1. Relationships of the Osmeridae and related families

Phylogenetic analysis of RAG1 sequences supported a previous result (López et al., 2004) that the Plecoglossidae and Salangidae are not contained in the Osmeridae 'proper' (Supplementary Fig. 7B), while analysis of combined mtDNA data nested these two families within the Osmeridae, although with very low support (Supplementary Fig. 7A). The sister relationship between the Plecoglossidae and Salangidae (López et al., 2004) was also generally supported by our analyses; however, parsimony analysis of the mtDNA and RAG1 genes indicated a poorly supported sister relationship between the Salangidae and Osmeridae (data not shown). Additional nuclear sequence data for the salangids may further clarify their position relative to *Plecoglossus* and the osmerids.

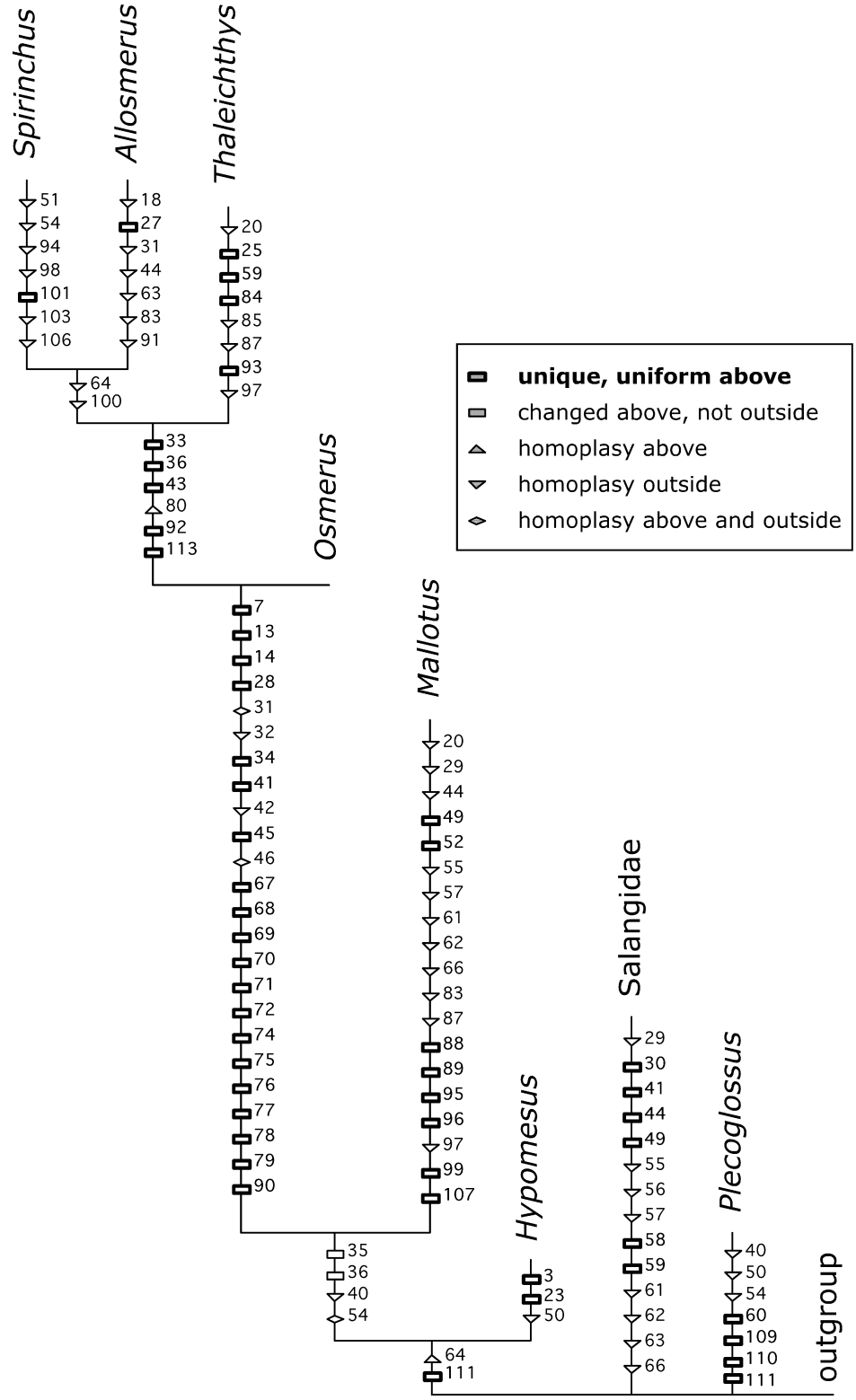


**Fig. 5.** Molecular phylogeny of osmerid genera with characters from McAllister (1963, 1966) [67–108], Begle (1991) [22–27], Wilson and Williams (1991) [1–21], Johnson and Patterson (1996) [28–66, 114] and Patterson and Johnson (1997) [109–113] mapped. Branch lengths are proportional to unambiguous changes. Numbers refer to characters listed in Supplementary Tables 3 and 4. Outgroup from Johnson and Patterson (1996) with all states coded as 0. 'Above' refers to all taxa that share a common ancestor at the designated node. 'Outside' refers taxa that are not monophyletic at the designated node.

### 4.2. Molecular systematics of the Osmeridae

#### 4.2.1. Intergeneric relationships

All analyses of the combined nDNA and allDNA datasets produced a highly supported phylogeny of the Osmeridae (Figs. 3B and 4A, respectively). The lack of resolution from the mitochondrial gene regions is possibly due to saturation in cyt*b*, where uncorrected distances between in-group species ranged from ~3% to 18%, and low substitution rates in the 12S and 16S regions with uncorrected distances between ~1% and 5%.

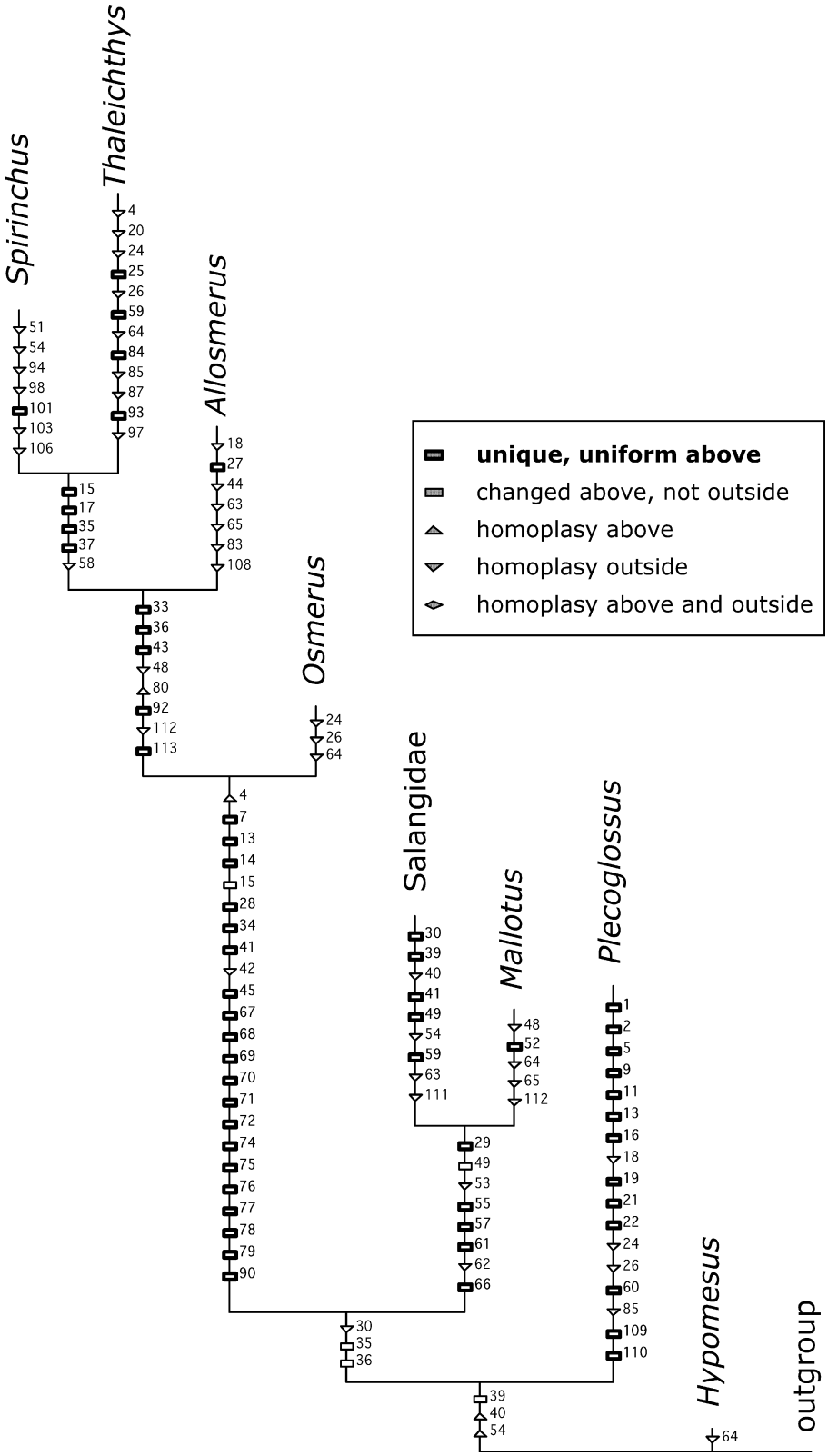


**Fig. 6.** Osmeridae, Salangidae, and Plecoglossidae interrelationships from Johnson and Patterson (1996), with morphological characters from McAllister (1963, 1966) [67–108], Begle (1991) [22–27], Wilson and Williams (1991) [1–21], Johnson and Patterson (1996) [28–66, 114] and Patterson and Johnson (1997) [109–113] mapped. Branch lengths are proportional to unambiguous changes. Numbers refer to characters listed in Supplementary Tables 3 and 4. Outgroup from Johnson and Patterson (1996) with all states coded as 0. 'Above' refers to all taxa that share a common ancestor at the designated node. 'Outside' refers taxa that are not monophyletic at the designated node.

The placement of *Mallotus villosus* differs between the phylogenies resulting from parsimony, where it appears as sister to all other genera, and the other reconstruction methods, where it is sister to the (*Osmerus*, (*Thaleichthys*, (*Allosmerus*, *Spirinchus*)) [OTAS] clade (Fig. 4). A possible explanation for this difference may be long branch attraction, where the high substitution rate along the *M.*

*villosus* lineage is interpreted by the parsimony reconstruction method as indicative of a closer relationship between this taxon and the outgroup *Plecoglossus altivelis* (Felsenstein, 2004). SH tests tend to be conservative (e.g. Buckley, 2002), meaning that a non-significant result may not indicate that there is no difference between the likelihoods of alternative topologies. SH tests failed to find a significant difference between the two topologies with different placements of *Mallotus*; however, based on high Bayesian and ML bootstrap support values for the nDNA and allDNA combined datasets, the combined evidence points more strongly towards a phylogeny where *Mallotus* is sister to the OTAS clade (Figs. 3B and 4A, respectively).

#### 4.2.2. Relationships within polytypic genera

While the intergeneric relationships within the Osmeridae are generally highly supported, the same cannot be said of the relationships within the three polytypic genera *Spirinchus*, *Osmerus*, and *Hypomesus*. Although historical hybridization between species could be a cause of unresolved intrageneric relationships, this seems unlikely in our case because the unexpected phylogenies were not restricted to mtDNA, which, lacking recombination, would be expected to retain any signal of historical hybridization longer than at nuclear loci (e.g. Arnold, 1997). Alternatively, a possible explanation for the discrepancies is that the times between divergences within these genera may have occurred in a relatively short span of time, indicated by the short internodes (Figs. 3B and 4), thereby obscuring the relationships among the species. These short speciation time internodes would have lessened the probability of the gene trees being equal to the species tree due to coalescent variance in the ancestral populations (Maddison, 1997; Degnan and Rosenberg, 2006). Increased geographic sampling of widely distributed *Spirinchus*, *Osmerus*, and *Hypomesus* species may help capture more of the genetic variation in these species; however, sequencing of additional loci will be necessary to overcome difficulties in determining branching relationships among species with short internodes.

The relationships within the genus *Hypomesus* are partially obscured with regards to whether *H. olidus* and *H. nipponensis* are sister taxa or whether *H. nipponensis* is sister to the rest of the *Hypomesus* species. Although SH tests failed to find a significant difference between the two topologies (Section 3.6), the former arrangement is poorly supported when it appears, while the latter is highly supported by Bayesian (1.0) and ML (89%) analysis of the nDNA dataset (Fig. 3B) and by Bayesian analysis (0.99) of the allDNA partition (Fig. 4A). Therefore, we consider the topology with *H. nipponensis* as sister to the rest of *Hypomesus* to be the best hypothesis of the relationships within this genus based on currently available data.

### 4.3. Characters used in previous systematic studies

Johnson and Patterson's (1996) morphological analysis has been the most comprehensive treatment of osmerid systematics to date, yet our molecular results conflict with it and all previous hypotheses. Their hypothesis (Figs. 1 and 6) is incongruent with the molecular phylogeny (Figs. 4A and 5) in three respects: (1) They place *Hypomesus* as the most basal taxon as opposed to *Plecoglossus* and the Salangidae, (2) In their analysis, the Salangidae are sister to *Mallotus* not to *Plecoglossus*, and (3) They consider *Allosmerus* and *Thaleichthys* as sister taxa, as opposed to an *Allosmerus–Spirinchus* relationship from the molecular analyses. Again assuming that the molecular topology represents the "true" osmerid phylogeny, a brief review of the morphological characters used by Johnson and Patterson (1996) may identify possible reasons for the incongruence between the current molecular results and all previous morphological hypotheses of relationships among osmerid genera.

There are four basic interrelated issues involved in the use of morphological characters in smelt systematics that could cause the incongruence with molecular data: character state polarity, weighting gains vs. losses, parallel evolution, and character coding issues. First, the polarity of morphological traits may be difficult to determine due to variability within different character states and/or difficulties in determining appropriate outgroup taxa. For example, the pattern of endopterygoid teeth (character 39; Supplementary Table 4), one of the 'primitive' characters Johnson and Patterson (1996) described as supporting *Hypomesus* as the most basal taxon, was assigned three ordered states by Johnson and Patterson (1996). The intermediate state of this character, however, showed a large amount of variation, and further, both ingroup (salangids) and outgroup (salmonids) taxa lack these teeth, leaving only *Hypomesus*, retropinnids, and a hypothetical outgroup sharing the 'ancestral' state. Given these issues, the polarity assigned to this character may be questionable and may therefore not be appropriate for placing *Hypomesus* basal to all osmerid taxa.

Next are the two related issues of how to weight character gains vs. losses and how to deal with homoplasious characters. If the molecular tree represents the "true" osmerid phylogeny, our analysis suggests that numerous morphological characters used in previous systematic studies of the Osmeridae are homoplasious (Figs. 5 and 6). This will not come as a surprise to anyone who has studied osmeroid fishes, as repeated losses and gains of particular character states have been noted by many authors (Johnson and Patterson, 1996, and references therein; Waters et al., 2000, 2002). The manner in which gains and losses are weighted can affect phylogenetic conclusions. For instance, with respect to two of the four 'primitive' character states displayed by *Hypomesus* (40 and 54; Supplementary Table 4) in Johnson and Patterson's (1996) analysis, the molecular phylogeny suggests independent evolution of the derived state while their phylogeny favors reversals. Although there is often good reason to differentially weight gains and losses (e.g. Wiens, 1999; Waters et al., 2002), characters 40 and 54 involve either a change in position or in articulation point, not a strict gain or loss of function. Thus, lacking a functional morphology-based justification, both the molecular (this study) and morphological (Johnson and Patterson, 1996) topologies may be considered equally parsimonious with respect to the evolution of these characters. Waters et al. (2002) reduced incongruence between molecular and morphological analyses of the Southern Hemisphere smelts by differentially weighting gains and losses. Homoplasious characters may also explain the strong morphological support found by Johnson and Patterson (1996) for a sister relationship between *Mallotus* and the Salangidae. There are five apparent synapomorphies that supported a sister relationship between these taxa (Fig. 6). For four of these traits (characters 61, 66, 55, and 57; Supplementary Tables 3 and 4), the *Mallotus*/salangid character state is shared by other fishes (Matsuoka and Iwai, 1983; Sazanov, 1986, both cited in Johnson and Patterson, 1996), which leaves open the possibility that the character states are also homoplasious among osmerid taxa. Homoplasious traits may also explain the conflict between Johnson and Patterson's (1996) placement of *Spirinchus* and *Thaleichthys* as sister taxa and our molecular phylogeny with an *Allosmerus–Spirinchus* sister relationship. *Osmerus* and members of the Galaxiidae share one of the traits that define a *Spirinchus–Thaleichthys* relationship (34; Supplementary Tables 3 and 4). *Allosmerus* and *Spirinchus* share two homoplastic traits: lack of cucumber odor (64, Fig. 5) and head length greater than 4.7 in standard length (100, Fig. 5). As stated by Johnson and Patterson (1996), the cucumber odor could be subjectively quantified and furthermore, could have been missed because it is only detectable in fresh specimens. Information about relative head length was unavailable for *Plecoglossus*; however, relatively small head sizes are also shared by species of *Osmerus*.

On a more positive note, even though homoplasy appears to have complicated the inference of osmerid relationships from morphology, the combined data from multiple studies showed that 28 characters unambiguously supported the intergeneric relationships inferred from molecular analysis and a further 26 characters were autapomorphic for particular genera (Fig. 5).

A final issue that may help explain incongruence between the molecular and morphological osmerid phylogenies is problems with character coding. Waters et al. (2002) suggested that some studies may have been inadvertently biased by individual systematists who sought characters to define anticipated relationships among particular taxa and we noticed instances in Johnson and Patterson's (1996) coding of polytypic genera where the method by which coding was decided was unclear. For example, Johnson and Patterson (1996) coded *Hypomesus* as polymorphic for both ancestral and derived states for character 47 (Supplementary Tables 3 and 4) although only *H. transpacificus* displayed the ancestral condition. This species is one of the most recently diverged in the genus (Fig. 4; Ilves and Taylor, 2008), thus, an alternative interpretation is that the 'derived' state is shared by all osmerids and a reversal occurred along the lineage leading to *H. transpacificus*. A similar issue was apparent with respect to how Johnson and Patterson (1996) coded character 57 (Supplementary Tables 3 and 4) for the Salangidae. The family was assigned the 'derived' state along with *Mallotus* even though two genera displayed the 'ancestral' condition, adding a synapomorphy to the salangid—*Mallotus* sister relationship found by Johnson and Patterson (1996). To avoid uncertainty about how polytypic taxa were coded, studies of higher taxa should include a description of what criteria were used to determine character states for polytypic taxa.

Johnson and Patterson (1996) discussed their confidence in different parts of their osmerid phylogeny. Their data strongly supported *Allosmerus*, *Thaleichthys*, and *Spirinchus* as derived taxa, and *Hypomesus* as a basal osmerid, conclusions also supported by our data (Figs. 3B, 4; Supplementary Figs. 1 and 4–6). Johnson and Patterson (1996) were less certain, however, about the arrangement of *Osmerus*, *Mallotus*, and the Salangidae because exchanging the positions of *Osmerus* and (*Mallotus* + Salangidae) only increased the tree length by a small amount. Four synapomorphies (characters 28, 34, 41, and 45) from Johnson and Patterson's (1996) data define the clade containing *Osmerus*, *Thaleichthys*, *Allosmerus*, and *Spirinchus* clade (Figs. 5 and 6). Including characters from other studies further increases the number of shared character states to 20 (Figs. 5 and 6) and exchanging the positions of *Osmerus* and *Mallotus* on this phylogeny increases the tree length from 193 to 220 steps (data not shown). Thus, while there may be limited support for the monophyly of the (*Osmerus*, (*Thaleichthys*, (*Allosmerus*, *Spirinchus*))) clade from Johnson and Patterson's (1996) dataset, this grouping is strongly supported by our molecular analyses and numerous shared character states from McAllister (1963), and Wilson and Williams (1991) (Figs. 5 and 6).

### 4.4. Taxonomic considerations

All genera and species are monophyletic from the combined analysis; therefore, taxonomic changes within the Osmeridae are unnecessary. Furthermore, debate about whether or not the Salangidae and Plecoglossidae should be designated as subfamilies within the Osmeridae (e.g. Fu et al., 2005; Nelson, 2006) becomes unnecessary if an unranked nomenclature system, such as the Phylocode (de Queiroz and Gauthier, 1990, 1992, 1994; Cantino and de Queiroz, 2006) is implemented. The Osmeridae, Salangidae, and Plecoglossidae are reciprocally monophyletic according to current data, therefore, they would all retain their respective names. In the future, additional names could be attached to the Plecoglossidae–Salangidae node and the node joining this clade to the Osmeridae, with sufficient evidence to support these relationships; however, because names are rank-free and do not require particular suffixes, the level of classification does not need to be taken into consideration.

## 5. Conclusions

Molecular phylogenetic analysis of concatenated mitochondrial and nuclear gene sequences of all species of the Osmeridae yielded a well-resolved phylogeny of the genera. Our molecular hypothesis conflicts with all previous morphological hypotheses, likely due mainly to homoplasious traits used in the construction of the earlier phylogenies. Species tree analysis did not contradict the results from the concatenated data, although not all intergeneric relationships were resolved. Additional sequencing of nuclear genes may help better determine the position of the problematic *Mallotus villosus* and uncertainties in the relationships among the species of the polytypic *Spirinchus*, *Osmerus*, and *Hypomesus*. Relationships among the Plecoglossidae, Salangidae, and Osmeridae are not yet clear; however, analysis of the nuclear RAG1 gene suggests the Plecoglossidae and Salangidae together are sister to the Osmeridae (Supplementary Fig. 7B; López et al., 2004). Additional nuclear gene sequences, particularly from salangid species, could aid in further resolving these systematic relationships.

The biogeography of the osmerids, and of the Holarctic region in general has been of great interest (e.g. McAllister, 1963; Briggs, 1974; Wilson and Williams, 1991; Taylor and Dodson, 1994). The availability of a resolved osmerid phylogeny will aid further studies of these, and other, fishes, including the evolution of morphological traits and the biogeographic history of Holarctic faunas.

## Acknowledgments

This research would not have been possible without the generous sample donations from many people across the Northern Hemisphere: R. Gross, D. Hay, K. Hieb, J. Jurvelius, E. Lowery, K. Maslenikov, J. McPhail, S. Mori, A. Morton, P. Moyle, L. Plough, T. Quinn, K. Takata, P. Tamkee, J. Taylor, M. Thompson, B. Urbain, and P. Westley. K. Maslenikov at the University of Washington Fish Collection was particularly helpful. Fieldwork collecting longfin smelt (*Spirinchus thaleichthys*) from Harrison and Pitt Lakes was conducted with the assistance of J. Hume and S. MacLellan from the Department of Fisheries and Oceans, G. Scholten, G. Wilson, and P. Woodruff from the BC Ministry of Environment, and Mike Stamford. Many thanks also to J. Gow, B. Davis, W. Maddison, S. Otto, M. Berbee, C. Stephens, L. Harris, J. McLean, J. Mee, S. Northrup, J. Weir, and P. Woodruff for discussion and support throughout this project. Early versions of this manuscript were greatly improved by comments from C.P. Burridge, W.K. Savage, M.J. Hickerson and one anonymous reviewer. This research was financially supported by a UGF grant from the University of British Columbia awarded to K.L.I. and a Natural Resources and Engineering Research Council of Canada Discovery Grant awarded to E.B.T.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2008.10.021.

## References

Arnold, M.L., 1997. Natural Hybridization and Evolution. Oxford University Press, New York.

Begle, D.P., 1991. Relationships of osmeroid fishes and the use of reductive characters in phylogenetic analysis. Syst. Zool. 40, 33–53.

Briggs, J.C., 1974. Marine Zoogeography. McGraw-Hill, Inc., Toronto.

Buckley, T.R., 2002. Model misspecification and probabilistic tests of topology: evidence from empirical data sets. Syst. Biol. 51, 509–523.
Cantino, P.D., de Queiroz, K., 2006. PhyloCode. Available from: <http://ohiou.edu/phylocode>.
Chapman, W.M., 1941. The osteology and relationships of the osmerid fishes. J. Morphol. 69, 279–301.
Chow, S., Hazama, K., 1998. Universal PCR primers for S7 ribosomal protein introns in fish. Mol. Ecol. 7, 1247–1263.
Cuvier, G., 1817. Le regne Animal. Deterville, Paris.
de Queiroz, K., Gauthier, J., 1990. Phylogeny as a central principle in taxonomy: phylogenetic definitions of taxon names. Syst. Zool. 39, 307–322.
de Queiroz, K., Gauthier, J., 1992. Phylogenetic taxonomy. Annu. Rev. Ecol. Syst. 23, 449–480.
de Queiroz, K., Gauthier, J., 1994. Toward a phylogenetic system of biological nomenclature. Trends Ecol. Evol. 9, 27–31.
Degnan, J.H., Rosenberg, N.A., 2006. Discordance of species trees with their most likely gene trees. PLoS Genet. 2, e68.
Eschmeyer, W.N., 2006. Catalogue of Fishes. California Academy of Sciences, San Francisco. Available from: <http://www.calacademy.org/research/ichthyology/catalog>.
Farris, J.S., Källersjö, M., Kluge, A.G., Bult, C., 1994. Testing significance of congruence. Cladistics 10, 315–319.
Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.
Felsenstein, J., 2004. Inferring Phylogenies. Sinauer Associates, Inc., Sunderland, MA.
Fu, C., Luo, J., Wu, J., López, J.A., Zhong, Y., Lei, G., Chen, J., 2005. Phylogenetic relationships of salangid fishes (Osmeridae, Salanginae) with comments on phylogenetic placement of the salangids based on mitochondrial DNA sequences. Mol. Phylogenet. Evol. 35, 76–84.
Greenwood, P.H., Rosen, D.E., Weitzman, S.H., Myers, G.S., 1966. Phyletic studies of teleostean fishes, with a provisional classification of living forms. Bull. Am. Mus. Nat. Hist. 131, 339–456.
Hillis, D.M., Bull, J.J., 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42, 182–192.
Howes, G.J., Sanford, C.P.J., 1987. The phylogenetic position of the Plecoglossidae (Teleostei, Salmoniformes), with comments on the Osmeridae and Osmeriodei. In: Proceedings of the Congress of European Ichthyologists, 5th, Stockholm, pp. 17–30.
Hubbs, C.L., 1925. A revision of the osmerid fishes of the North Pacific. Proc. Biol. Soc. Wash. 38, 49–56.
Huelsenbeck, J.P., Ronquist, F., 2001. MrBayes: Bayesian inference of phylogeny. Bioinformatics 17, 754–755.
Ilves, K.L., Taylor, E.B., 2007. Are *Hypomesus chishimaensis* and *H. nipponensis* (Pisces: Osmeridae) distinct species? A molecular assessment using comparative sequence data from five genes. Copeia 2007, 180–185.
Ilves, K.L., Taylor, E.B., 2008. Evolutionary and biogeographic patterns within the smelt genus *Hypomesus* (Pisces: Osmeridae) in the North Pacific Ocean. J. Biogeogr. 35, 48–64.
Johnson, G.D., Patterson, C., 1996. Relationships of lower Euteleostean fishes. In: Stiassny, M., Parenti, L.J., Johnson, L.R., David, G. (Eds.), Interrelationships of Fishes. Academic Press, San Diego, pp. 251–332.
Kendall, W.C., 1927. The smelts. Bull. U.S. Bur. Fish. 42, 217–375.
Klyukanov, V.A., 1969. Morphological bases of classification of smelts of the genus *Osmerus* (Osmeridae). Zool. J. 48, 99–109.
Klyukanov, V.A., 1970. Morphological basis of the classification of smelts of the genus *Hypomesus*. Zool. J. 49, 1534–1542.
Klyukanov, V.A., 1975. The systematic position of the Osmeridae in the order Salmoniformes. J. Ichthyol. 15, 1–17.
Kocher, T.D., Thomas, W.K., Meyer, A., Edwards, S.V., Pääbo, S., Villablanca, F.X., Wilson, A.C., 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc. Natl. Acad. Sci. USA 86, 6196–6200.
Linnaeus, C., 1758. Systema Naturae, 10th ed. Stockholm. Holmiae: Impensis Direct. Laurentii Salvi, 1758–1759.
Liu, L., Pearl, D.K., 2007. Species trees from gene trees: reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. Syst. Biol. 56, 504–514.
Liu, L., Pearl, D.K., Brumfield, R.T., Edwards, S.V., 2008. Estimating species trees using multiple allele data. Evolution 62, 2080–2091.
López, J.A., Chen, W., Ortí, G., 2004. Esociform phylogeny. Copeia 2004, 449–464.
Luey, J.E., Krueger, C.C., Schreiner, D.R., 1982. Genetic relationships among smelt, genus *Osmerus*. Copeia 1982, 725–728.
Maddison, D.R., Maddison, W.P., 2003. MacClade 4.06: Analysis of Phylogeny and Character Evolution. Sinauer Associates, Sunderland, MA.
Maddison, W.P., 1997. Gene trees in species trees. Syst. Biol. 46, 523–536.
Matsuoka, M., Iwai, T., 1983. Adipose fin cartilage found in some teleostean fishes. Jpn. J. Ichthyol. 30, 37–46.
McAllister, D.E., 1963. A revision of the smelt family, Osmeridae. Bull. Nat. Mus. Can. 191, 1–53.
McAllister, D.E., 1966. Numerical taxonomy and the smelt family, Osmeridae. Can. Field. Nat. 80, T227–T238.
McAllister, D.E., Parker, B., Couvillion, K.E., 1980. *Osmerus mordax* (Mitchill), rainbow smelt. In: Lee, D.S., Gilbert, C.R., Hocutt, C.H., Jenkins, R.E., McAllister, D.E., Stauffer, J.R. (Eds.), Atlas of North American Freshwater Fishes. North Carolina State Museum of Natural History, Raleigh, NC, p. 124.
Nellbring, S., 1989. The ecology of smelts (genus *Osmerus*): a literature review. Nordic J. Freshw. Res. 65, 116–145.
Nelson, J.S., 2006. Fishes of the World, fourth ed. John Wiley and Sons, Hoboken, NJ.
Parenti, L.R., 1986. The phylogenetic significance of bone types in euteleost fishes. Zool. J. Linn. Soc. 87, 37–51.
Patterson, C., Johnson, G.D., 1997. The data, the matrix, and the message: comments on Begle's "Relationships of osmeroid fishes". Syst. Biol. 46, 358–365.
Posada, D., Buckley, T.R., 2004. Model selection and model averaging in phylogenetics: advantages of Akaike Information Criterion and Bayesian approaches over likelihood ratio tests. Syst. Biol. 53, 793–808.
Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14, 817–818.
Presa, P.B., Pardo, G., Martinez, P., Bernatchez, L., 2002. Phylogeographic congruence between mtDNA and rDNA ITS markers in Brown Trout. Mol. Biol. Evol. 19, 2161–2175.
Quenouille, B., Bermingham, E., Planes, S., 2004. Molecular systematics of the damselfishes (Teleostei: Pomacentridae): Bayesian phylogenetic analyses of mitochondrial and nuclear DNA sequences. Mol. Phylogenet. Evol. 31, 66–88.
Rambaut, A., 1996. Se-Al: A Manual Sequence Alignment Editor, v.2.0a11. Available from: <http://tree.bio.ed.ac.uk/software/seal/>.
Roberts, T.R., 1984. Skeletal anatomy and classification of the neotenic asian salmoniform superfamily Salangoidea (icefishes or noodlefishes). Proc. Cal. Acad. Sci. 43, 179–220.
Saruwatari, T., López, J.A., Pietsch, T.W., 1997. A revision of the osmerid genus *Hypomesus* Gill (Teleostei: Salmoniformes), with the description of a new species from the Southern Kuril Islands. Species Divers. 2, 59–82.
Sazanov, Y.I., 1986. Morphology and classification of the fishes of the family Platytroctidae (Salmoniformes, Alepocephaloidei). Tr. P. P. Shirshov Inst. Oceanol. 121a, 51–96 (in Russian).
Scott, W.B., Crossman, E.J., 1973. Freshwater fishes of Canada. Fish. Res. Board. Canada Bull. 184.
Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 1114–1116.
Sidorov, L.K., Pichugin, Y., 2004. Morphological traits of lacustrine forms of smelts of the genus *Hypomesus* (Salmoniformes) from the Southern Kurils. J. Ichthyol. 44, 433–443.
Swofford, D.L., 2002. PAUP*: Phylogenetic Analysis using Parsimony. Ver. 4.0b10. Sinauer Associates, Sunderland, MA.
Taylor, E.B., Dodson, J.J., 1994. A molecular analysis of relationships and biogeography within a species complex of Holarctic fish (genus *Osmerus*). Mol. Ecol. 3, 235–248.
Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The ClustalX-Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.
Waters, J.M., López, J.A., Wallis, G.P., 2000. Molecular phylogenetics and biogeography of galaxiid fishes (Osteichthyes: Galaxiidae): dispersal, vicariance, and the position of *Lepidogalaxias salamandroides*. Syst. Biol. 49, 777–795.
Waters, J.M., Saruwatari, T., Kobayashi, T., Oohara, I., McDowall, R.M., Wallis, G.P., 2002. Phylogenetic placement of retropinnid fishes: data set incongruence can be reduced by using asymmetric character state transformation costs. Syst. Biol. 51, 432–449.
Wiens, J.J., 1998. Combining data sets with different phylogenetic histories. Syst. Biol. 47, 568–581.
Wiens, J.J., 1999. Polymorphism in systematics and comparative biology. Annu. Rev. Ecol. Syst. 30, 327–362.
Wilson, M.V.H., Williams, R.R.G., 1991. New Paleocene genus and species of smelt (Teleostei: Osmeridae) from freshwater deposits of the Paskapoo Formation, Alberta, Canada, and comments on osmerid phylogeny. J. Vertebr. Paleontol. 11, 434–451.