Molecular Phylogenetics and Evolution
Vol. 11, No. 3, April, pp. 381–393, 1999
Article ID mpev.1998.0599, available online at http://www.idealibrary.com on IDEAL®

Phylogeny of Salmonine Fishes Based on Growth Hormone Introns: Atlantic (*Salmo*) and Pacific (*Oncorhynchus*) Salmon Are Not Sister Taxa

Todd H. Oakley[1] and Ruth B. Phillips

Department of Biological Sciences, University of Wisconsin-Milwaukee, PO Box 413, Milwaukee, Wisconsin 53201

Received February 10, 1998; revised September 14, 1998

Though salmonid fishes are a well-studied group, phylogenetic questions remain, especially with respect to genus-level relationships. These questions were addressed with duplicate growth hormone (GH) introns. Intron sequences from each duplicate gene yielded phylogenetic trees that were not significantly different from each other in topology. Statistical tests supported validity of the controversial monotypic genus *Parahucho*, monophyly of *Oncorhynchus*, and inclusion of *Acantholingua ohridana* within *Salmo*. Suprisingly, GH1 intron C (GH1C) did not support the widely accepted hypothesis that *Oncorhynchus* (Pacific salmon and trout) and *Salmo* (Atlantic salmon and trout) are sibling genera; GH2C was ambiguous at this node. Previously published data were also examined for support of *Salmo* and *Oncorhynchus* as sister taxa and only morphology showed significant support. If not sister taxa, the independent evolution of anadromy—the migration to sea and return to freshwater for spawning—is most parsimonious. While there was incongruence with and among published data sets, the GH1C intron phylogeny was the best hypothesis, based on currently available molecular data.

Key Words: **salmonid phylogeny; growth hormone; intron; *Acantholingua; Oncorhynchus; Parahucho; Salmo;* parametric bootstrapping.**

INTRODUCTION

Salmonid fishes, including salmon, trout, and char, are a commercially important and biologically interesting family of fishes. The entire family is tetraploid, a rare situation among animals, resulting from ancestral genome duplication (reviewed in Allendorf and Thorgaard, 1984). Tetraploidization in salmonids has led to complex patterns of chromosomal evolution (reviewed in Hartley, 1987) and provides an unusual situation valuable for study of gene duplication (Li, 1980). Salmonids also exhibit complex migratory and mating behaviors. For these reasons, the group has been studied extensively, including many systematic and phylogenetic investigations (reviewed in Stearley and Smith, 1993; Phillips and Oakley, 1997).

Despite an unusually large number of phylogenetic studies conducted on some groups of salmonids—especially *Oncorhynchus* (Domanico *et al.*, 1997; Kitano *et al.*, 1997; Oohara *et al.*, 1997; reviewed in Phillips and Oakley, 1997)—other important groups are understudied. The result has been a paucity of information about deeper divergence within Salmonidae. Although several studies based on morphology have been performed (Norden, 1961; Kendall and Behnke, 1984; Dorofeyeva, 1989; Stearley, 1992), all propose different relationships among genera (Fig. 1). Also, while multiple nucleotide sequences have been determined for representatives of various genera, no molecular study has yet focused on genus-level systematics. Our study focused on genus-level relationships by using sequences of duplicate growth hormone (GH) introns. Because GH introns produced a suprising result by not grouping *Oncorhynchus* and *Salmo* as sister genera, previously published data were rigorously analyzed in light of this finding. Various other taxonomic proposals were also tested with GH intron data, including the status of genera *Parahucho*, *Acantholingua*, and *Oncorhynchus*. To facilitate understanding of competing hypotheses, a brief review of salmonid taxonomy is presented (see also Stearley and Smith, 1993; Phillips and Oakley, 1997).

Genus-Level Systematics of Salmonid Fishes

The subfamily Salmoninae contains at least five genera and as many as nine have been proposed. However, some of the proposed genera may not represent monophyletic groups or may cause another genus to be paraphyletic (Phillips and Oakley, 1997).

Parahucho is a controversial genus and its status hinges on the relationships of three species: *Brachy-*

[1] To whom correspondence should be addressed at Present Address: Zoology Department, Duke University, Box 90325, Durham, NC 27708-0325. Fax: (919) 684-6168. E-mail: tho@acpub.duke.edu.

1055-7903/99 $30.00

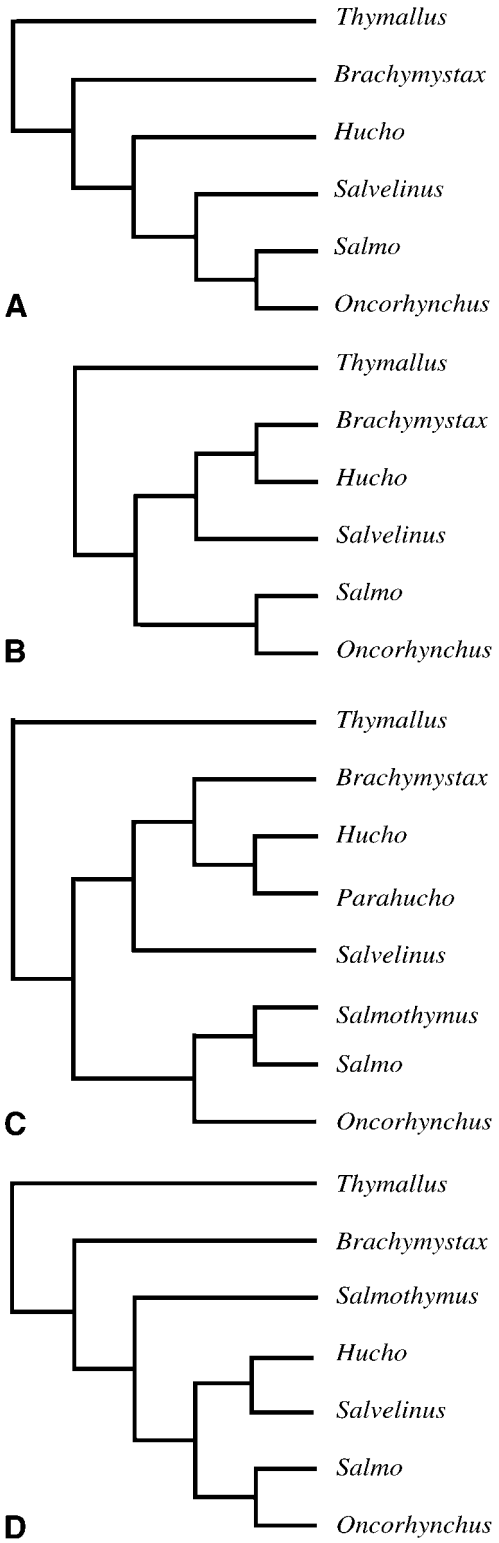

FIG. 1. Various phylogenetic hypotheses based on morphological data. (A) Norden (1961); (B) Kendall and Behnke (1984); (C) Dorofeyeva (1989); (D) Stearley and Smith (1993).

mystax lenok, *Hucho hucho*, and the Sakhalin taimen, *Hucho perryi* or *Parahucho perryi* (because of the ambiguity in genus name, the specific name '*perryi*' alone will be used to refer to the fish in the following discussion). If *Hucho* (including *perryi*) is monophyletic, there is no reason for *Parahucho* to exist (Stearley and Smith, 1993). If, however, including *perryi* in *Hucho* makes the genus paraphyletic, the status of *Parahucho* as a monotypic genus is warranted (Phillips *et al.*, 1995; Phillips and Oakley, 1997). Paraphyly of *Hucho* is the case if *B. lenok* and *H. hucho* are sister species and *perryi* is included in *Hucho*.

Hucho contains up to three other species, including *H. hucho* with two subspecies (*H. h. taimen* and *H. h. hucho*). Two additional species, *H. bleekeri* and *H. ishikawai*, are discussed in the literature (Holcik, 1988), but were not available for this study. Both species are rare. *H. bleekeri* is restricted to the upper Yangtze River in China and *H. ishikawai* to the Yalu River in Korea (Holcik, 1988). Neither has been included in previous phylogenetic analyses, and Holcik (1988) suggested that *H. ishikawai* is another subspecies of *H. hucho*.

Relationships among *Salmothymus* and *Acantholingua* are uncertain. Hadzisce (1961) proposed that each genus contains a single species, *S. obtusirostis* and *A. ohridana*. Svetovidov (1975) placed both species in *Salmothymus* and assigned subgeneric rank to *Acantholingua*. Behnke (1968) argued that both *Salmothymus* and *Acantholingua* be considered subgenera of *Salmo*. Stearley and Smith (1993) assigned several species to *Salmothymus*, including *S. obtusirostris* and *S.* (*Platysalmo*) *platycephalus*. They also believe that *Salmo ishchan* and *Salmo carpio* may be members of *Salmothymus*. Their morphological analysis provided ambiguous results with regard to placement of *Acantholingua ohridana*. Two equally most parsimonious trees differed in placement of *A. ohridana*. Only *A. ohridana* was available for the current study, which allows testing of the hypothesis that the species is a member of *Salmo* (Behnke, 1968).

Other salmonid genera, *Brachymystax*, *Salvelinus*, and *Oncorhynchus*, are generally considered monophyletic groups. *Brachymystax* contains a single species, *B. lenok*, with two morphotypes: a sharp-snouted form and a blunt-snouted form. Representatives of both morphotypes were used in the current study. Osinov (1991) argued that the morphotypes are separate species and demonstrated evidence for reproductive isolation between types.

Chars of genus *Salvelinus* contain six commonly recognized species and several subspecies (Behnke, 1980; Cavender, 1980). Taxonomy and relationships among chars are a matter of considerable debate (Phillips and Oakley, 1997). Because of this uncertainty, three different species were chosen to represent *Salvelinus* for the current study. Phylogenetic analysis of

ribosomal DNA sequences (Pleyte *et al.*, 1992) has suggested that the genus contains three pairs of sibling species. One species from each of these pairs (*S. alpinus*, *S. namaycush*, and *S. confluentus*) was chosen to represent *Salvelinus*. These species also represent each of two groups of *Salvelinus* proposed by Savvaitova (1980).

Oncorhynchus was recently revised and now contains 'Pacific trouts' (*O. mykiss*, *O. clarki* used in this study) previously considered members of *Salmo* (Smith and Stearley, 1989). There is consensus that *Salmo* has at least two species, *S. trutta* and *S. salar* (both used in this study). In addition, *S. trutta* contains a number of distinct geographical and ecological forms, sometimes considered species or subspecies (Giuffra *et al.*, 1994; Patarnello *et al.*, 1994; Bernatchez and Osinov, 1995). As discussed earlier, *A. ohridana* and *Salmothymus obtusirostris* may also be members of *Salmo* (Behnke, 1968).

Growth hormone intron sequences were chosen to address questions in salmonid systematics. An important consideration when choosing DNA sequence for phylogenetic analysis is the evolutionary rate of the region in question. Previous studies have suggested that GH introns are evolving at a rate suitable for intergeneric or interspecific questions in salmonid fishes (Blackhall, 1994; McKay *et al.*, 1996; Domanico *et al.*, 1997). In addition, the duplicate nature of the gene is valuable for molecular evolutionary investigation.

Gene Duplication and Salmonid GH

Gene duplication is known to be an important source of evolutionary innovation (Ohno, 1970; Kimura and Ohta, 1974). Tetraploid animals such as *Xenopus laevis* (Hughes and Hughes, 1993; Hughes, 1994) and *Cyprinus carpio* (Larhammar and Risinger, 1994) have been used to study gene evolution following duplication. While studies of single tetraploid species have been instructive, salmonids provide an additional advantage in that an entire family has radiated following a tetraploidization event (reviewed in Allendorf and Thorgaard, 1984).

There is strong evidence that GH was duplicated in the salmonid tetraploidization event, as salmonids contain two active GH genes and most other vertebrates have only one (Agellon *et al.*, 1988; Rentier-Delrue *et al.*, 1989). In addition, Devlin (1993) performed a phylogenetic analysis on GH coding sequences from four species. GH1 formed a monophyletic group separate from GH2 sequences of the same species, further supporting the duplication by tetraploidy hypothesis.

Our study utilized duplicate GH intron sequences from 13 species, subspecies, or morphotypes. Since duplication of the intron preceded speciation, a reasonable null hypothesis is that each duplicate should yield an identical phylogeny with respect to both topology and branch lengths. Deviations from this null hypothesis have interesting implications for evolution of duplicate introns. For example, if lengths of corresponding branches are significantly different between duplicate phylogenies, evolutionary rates are not equal, implying either mutation rate differences or differences in selective constraints.

MATERIALS AND METHODS

Determination of Nucleotide Sequences

GenBank accession numbers for previously published sequences, as well as collection information associated with sequences determined for the current study are listed in Table 1. A single individual from each species was used for sequencing. The third intron of the growth hormone gene was amplified using the polymerase chain reaction (PCR). Primers were designed to anneal to conserved regions of flanking GH exons from trout and salmon sequences available in GenBank as described by Domanico (1994). For growth hormone 2 intron C (GH2C), the primer GH2CFB (5′-ATCGTGAGCCCAATCGACAAGCAG-3′), which anneals to exon 3 was used in conjuction with GH2CR (5′-GGGTACTCCCAGGATTCAATCAGG-3′), which anneals to exon 4. For growth hormone 1 intron C (GH1C), the primer GH1L (5′-ATCGTGAGCCCAATCGACAAGCAC-3′), which anneals to exon 3 was used in conjuction with GH2CR. Primers for GH2C amplified introns from all species attempted. GH1C could not be amplified from *Salmo trutta* or *Acantholingua orhidana* due to nonspecificity of the 5′ primer. Since relationships of these taxa were clear based on GH2C data, new GH1C primers were not designed.

The PCRs contained approximately 50 ng of template DNA, 150 ng of each primer, 1× Thermal Buffer (Promega, Madison, WI), 200 µM each deoxynucleotide triphosphate, and 0.015 µM $MgCl_2$ in 50 µl total volume. The reactions were cycled in a Coy Model 60 thermal cycler at the following temperatures: an initial 94°C denaturation for 3 min, and then 35 cycles of 94°C denaturation for 1.5 min, 55°C annealing for 2 min, 72°C extension for 3 min, and a final 72°C extension for 7 min. GH2C primers amplified only GH2 sequence, while GH1C primers amplified both GH1 and GH2, which were quite different in size in all species and therefore could be separated using electrophoresis. There was no ambiguity in determining GH1 introns from GH2 introns as the sequences are quite diverged and size of GH1C was always larger than GH2C. Amplified DNA from both genes was fractionated by electrophoresis through 2% low-melting agarose gels, cut from the gels with sterile razor blades and purified using Wizard PCR mini-columns (Promega, Madison, WI) according to manufacturer's instructions.

Nucleotide sequences of GH introns were determined

TABLE 1

Collection Data and GenBank Accession Numbers for Growth Hormone Intron C Sequences Used in this Study

Species	Collector	Collection/stock information[b]	Accession no. GH1C	Accession no. GH2C
B. lenok—blunt	A. Osinov	River Kohr, Russia	AF005918	AF005916
B. lenok—sharp	A. Osinov	River Kohr, Russia	AF005919	AF005917
H. h. hucho	P. Rab	April 28, 1994 from the Turiec River in Slovakia, a tributary of the Upper Danube	N/A	AF005906
H. h. taimen	D. Proebstel	Amur River in Eastern Russia	N/A	AF005907
P. perryi	F. Yamazaki and A. Goto	Hokkaido Fish Hatchery, Hokkaido, Japan	AF005920	AF005908
S. a. alpinus		Norwegian stock reared at The Freshwater Institute in Manitoba, Canada	AF005921	AF005909
S. namaycush		National Fish Hatchery, Iron River, WI, Green Lake stock	AF005922	AF005910
S. confluentus		Arrow Lake, WI	N/A	AF005911
S. salar			X61938	M21573
S. trutta	S. Hartley	Scotland	N/A	AF005912
O. mykiss		Evergreen hatchery, Pound, WI	AF005923	J03797[a]
O. clarki	P. Bentzen		AF005924	AF005913
O. kisutch		Kettle Moraine Hatchery, WI	AF005925	U04930[a]
O. tshawytscha	M. Dorschner	Milwaukee River, Estabrook Park, Milwaukee WI	Du et al (1993)[a]	AF005914
O. gorbuscha		Alaska	AF005926	Domanico (1994)[a]
O. keta	B. Evrett	Yukon R., Alaska	AF005927	L04688[a]
A. ohridana		Lake Ohrid, Bosnia-Herzegovina	N/A	AF005915
O. nerka			U14551[a]	U14555[a]

[a] Sequences were obtained from GenBank or the literature.
[b] Collection information is only for sequences determined for this study.

using 1 μg of purified DNA, 3.2 pmol sequencing primer, and labeled di-deoxynucleotides with an ABI 373 automated sequencer according to manufacturer's instructions. Sequences of both strands of both GH1C and GH2C introns were determined using the same primers as used for PCR. Because the length of GH1C introns was too great to allow for sufficient overlap of opposite strands when using only two sequencing primers, three additional primers were designed to anneal to conserved regions of the intron (based on sequences in GenBank, and preliminary data): GH300F (3′-CATTTACAATGACAAGCAGAG-5′), GH300R (3′-CTCTGCTTGTCATTGTAAATG-5′), and GH600R (3′-GGTTTCCTGGACAGAAAAGCAC-5′).

For analysis of branching order of genera using additional data sets, the nuclear ribosomal DNA sequence of the internal transcribed spacer (ITS1) of *Salmo salar* was determined by amplifying the gene as described by Domanico *et al.* (1997), cloning into a 'T vector' prepared as described by Marchuk *et al.* (1991), and sequencing with an ABI 373 sequencer.

Raw sequence data were analyzed using the computer program Sequencher (Gene Codes, Ann Arbor, MI) by removing unreliable sequence regions (flanking sequences with greater than two ambiguous bases per 25), and aligning trimmed sequences into a contiguous sequence. Discrepancies between opposite strands or same-direction extensions of different primers were resolved by visual comparison of chromatogram data.

Sequence Alignment

Salmonid GH introns are variable in length; so alignment (see Acknowlegments for alignment availability) required large gaps. Initially, the computer programs ClustalW (Thompson *et al.*, 1994) and PileUp (GCG, Madison, WI) were used to align sequences, but the presence of both large and small indels proved problematic for the alignment algorithms, and unacceptable alignments resulted for all of a variety of different parameters. For example, a deletion of 81 base pairs (bp) that occurs in all species except *B. lenok* and *H. hucho* was not identified by the computer programs using default parameters, even though the deletion is flanked by sequences of 20 and 50 bp with over 95% identity. Greatly decreasing the relative gap extension penalty in ClustalW allowed the program to identify the large gap, but other areas of the alignment then suffered, as unnecessary gaps were introduced elsewhere. Therefore, alignments that were used in further analyses were generated by eye using the computer programs SeqApp (D. Gilbert, University of Indiana) and Node Zero (Oakley, unpublished) in order to maximize hypothesized homology of characters. Phylogenetic analyses performed by ignoring any site contain-

ing a gap were very similar to analyses including such sites. Furthermore, using the default parameters of ClustalW for alignment of both introns, followed by parsimony analyses, yielded trees consistent with parsimony analyses based on our preferred alignment. These analyses suggested that specifics of alignment were not crucial for the results of this study; so only results of our preferred alignment are presented.

A highly variable region (bases 228–618, most of which is a large insert in *O. kisutch*) of GH1C was not included in subsequent phylogenetic analyses because it could not be aligned unambiguously. Since the focus of this study was on intergeneric comparisons, intraspecific variation was not examined in detail; however, we know that some intraspecific variation exists here. Intraspecific differences occurred within the excluded region in *B. lenok* (each morphotype had a different sequence) and intraindividual variation (three alleles consisting of different numbers of 31 nt tandem repeats) has been reported in this region in *O. kisutch* (Forbes *et al.*, 1994). Preliminary sequence data for *H. hucho* indicate that a microsatellite repeat $(AT)^N$ also occurs in the same region in both subspecies. We hypothesize that individuals were heterozygous for alleles of different length, making direct sequencing impossible; so GH1C from *H. hucho* is not included in this study. Because of this intraspecific variation in multiple species, the region of GH1C was excluded from our intergeneric analysis.

Phylogenetic Analysis of GH Introns

Phylogenetic trees were constructed based on separately aligned GH1C and GH2C sequences. The branch and bound option of PAUP 3.1.1 (Swofford, 1993) was used for maximum parsimony analysis. PAUP 3.1.1 was also used for nonparametric bootstrap analysis (Felsenstein, 1985) using the heuristic search option. Alignment gaps were coded as binary characters, but including gap characters did not significantly affect tree topology for either intron (data not shown). Trees are shown rooted along the branch leading to *Brachymystax* and *Hucho*. Such a root is supported by chromosome data (Hartley, 1987). GH data from a close outgroup to the salmonines, *Thymallus arcticus* (a member of the subfamily Thymallinae), was not included because huge deletions caused ambiguous alignment.

Maximum likelihood analyses were performed using a test version of PAUP* ver. 4d55 (Swofford, 1997). First, the optimal model was determined for each data set. Data analyzed using a model with more parameters always have an equal or better likelihood score than the same data analyzed with a less parameter-rich model. The change in likelihood scores from one model to the next can be compared to a χ^2 distribution for statistical significance (Goldman, 1993). If the change in likelihood is significant, the model better fits the data; otherwise, the parameters introduced are considered superfluous. For this study, models were tested for significant improvement in the following order: JC69, F81, HKY85, GTR (abbreviations described in Table 2). A parameter (discrete Γ) describing rate heterogeneity among sites (Yang, 1993) was then added to the most parameter-rich model that significantly increased likelihood and tested for significant improvement of model fit. The optimal model for each data set was then used in a heuristic search for the best tree. As with parsimony analysis, GH1C and GH2C were considered separately. A combined analysis was also performed using species for which both intron sequences were available.

A likelihood ratio test of congruence (Huelsenbeck and Bull, 1996) was used to test for significant incongruence between GH1C and GH2C tree topologies. The null hypothesis, that the ML tree based on combined intron data underlies both introns, was tested. First, the likelihood values of the best tree were compared with the likelihood obtained when assuming the null hypothesis (δ). A null distribution for this statistic was determined by simulating 100 data sets with Seq-Gen (Rambaut and Grassly, 1997) using the optimal model and parameters estimated assuming the combined tree. Values of δ determined for each simulated replicate formed the null distribution for the test.

The hypothesis that corresponding branch lengths were identical between trees based on GH1C and GH2C was also tested with a likelihood ratio test (Huelsenbeck *et al.*, 1997). The likelihood under the null hypothesis (branch lengths constrained to be equal) was compared to the likelihood under the alternative hypothesis (branch lengths not constrained). To determine significance, the test statistic (δ) was compared to

TABLE 2

Likelihood Model Choice for GH1 and GH2 Intron C

Data set	Model	log L_0	log L_1	δ
GH1C	JC69 vs F81	−1766.75	−1749.36	34.79[a]
	F81 vs HKY85	−1749.36	−1739.42	19.89[a]
	HKY85 vs HKY85 + Γ	−1739.42	−1736.38	6.07[b]
	HKY85 vs GTR	−1739.42	−1738.78	1.28
GH2C	JC69 vs F81	−1132.03	−1012.86	238.34[a]
	F81 vs HKY85	−1012.86	−1011.62	2.48
	F81 vs F81 + Γ	−1012.86	−1011.52	2.68
	F81 vs GTR	−1012.86	−1010.44	4.84

Note. Model abbreviations follow Swofford *et al.* (1996): JC69 = multiple substitution (Jukes and Cantor, 1969); F81 = multiple substitution, empirical base frequency (Felsenstein, 1981); HKY85 = multiple substitution, empirical base frequency, transition–transversion parameters (Hasegawa *et al.*, 1985); GTR = General time-reversible (Lanave *et al.*, 1984; Tavare, 1986; Rodriguez *et al.*, 1990); Γ = rate heterogeneity among sites estimated with discrete approximation of Γ distribution.

[a] Significant improvement based on χ^2 distribution ($P < 0.0001$).

[b] Significant improvement based on χ^2 distribution ($P < 0.05$).

a χ^2 distribution with one degree of freedom (Huelsenbeck *et al.*, 1997).

Three null hypotheses concerning salmonid relationships were tested using GH intron data. While these could have been addressed with a T-PTP test, this approach was not used because Swofford *et al.* (1996) showed that the test 'does not test for monophyly'. Instead, a parsimony parametric bootstrapping procedure (Hillis *et al.*, 1996) was used. Parsimony was used instead of likelihood because the analogous likelihood tests would have required an unreasonable amount of computer time.

To illustrate the approach used, consider the first null hypothesis, that *perryi* and *Hucho* are monophyletic. Using GH2C sequence data, the most parsimonious tree consistent with the null hypothesis is 12 steps longer than the most parsimonious tree. To determine statistical significance of this difference, parametric bootstrapping was used (Huelsenbeck *et al.*, 1995; Hillis *et al.*, 1996) to generate a null distribution. Parameter values for the optimal model were estimated with likelihood on the most parsimonious tree consistent with *perryi*/*Hucho* monophyly using a test version of PAUP* ver. 4d55 (Swofford, 1997). Parameter values and tree data were then entered into Seq-Gen (Rambaut and Grassly, 1997), which was used to simulate 100 different data sets assuming the optimal model of DNA evolution and *perryi*/*Hucho* monophyly. The difference in steps between the most parsimonious tree and the most parsimonious tree consistent with *perryi*/*Hucho* monophyly was determined for each simulated data set, and these difference values constituted the null distribution for the test. The actual difference in number of steps between trees was then compared to the randomly generated null distribution to determine a '*P*' value for the test.

Two other taxonomic proposals were also addressed using the described bootstrapping procedure. The null hypothesis that *A. ohridana* does not form a monophyletic group with *Salmo* was tested in a similar fashion using GH2C data. Polyphyly of *Oncorhynchus* was tested using both GH1C and GH2C data.

Are Oncorhynchus and Salmo Sister Genera?

This question was examined using a combined data set including GH1C, GH2C, ITS1, ITS2, cyt-*b*, and morphology (see Table 3 for sources of data). For this comparison one species from each of four genera was used. For all data sets except cyt-*b*, the same species was used to represent each genus—*P. perryi*, *O. mykiss*, *Salmo salar*, and *Salvelinus namaycush*. For cyt-*b*, the closely related species *Salvelinus fontinalis* was used since *S. namaycush* sequence was not available. Similarly, *Salmo trutta* was used for cyt-*b* and not *S. salar*. All data sets were analyzed with parsimony, and sequence data were also analyzed with maximum likelihood.

TABLE 3

Number of Synapomorphies that Support Each Topology (see Fig. 6) for Each of Six Data Sets

Data set	Topology A	Topology B	Topology C
Morphology**	11	2	2
ITS2	8	1	7
ITS1	8	9	3
GH2C	0	1	0
Cytochrome-*b*	14	12	18
GH1C**	0	0	14
Combined	41	25	44

Note. Two data sets (**) show a significant difference from a null hypothesis of equal support for each topology, indicating presence of phylogenetic signal. Morphology data from Stearley and Smith (1993); ITS1/2 = Internal transcribed spacer 1 and 2 (Domanico, 1994; Domanico *et al.*, 1997); GH2C = Growth hormone 2 intron C; cyt-*b* = mitochondrial cytochrome *b* (GenBank Accession Nos. D58396–D58403); GH1C = Growth hormone 1 intron C (Table 1). Topology A–C refers to Fig. 6.

Parsimony analysis was performed on the four-taxon sets by counting the number of synapomorphies in each data set that supported each of the three possible topologies. Statistical significance of phylogenetic signal was determined by calculating the probability of the observed synapomorphies using a binomial distribution and assuming that, in the complete absence of phylogenetic signal, each tree would be supported by an equal number of synapomorphies.

Likelihood analysis was also performed on the four-taxon sequence data. Optimal models were determined for each data set as described earlier for GH introns. Likelihood values were calculated separately for each possible topology using the optimal model for each data set.

Incongruence length difference (ILD) tests (Farris *et al.*, 1994) were performed using the partition heterogeneity test of PAUP* 4d56 (Swofford, 1997). Invariant sites were excluded and 1000 randomized replicates were used for each test. The ILD test was used to determine whether or not groups of data partitions were significantly incongruent.

RESULTS

Phylogenetic Analysis of GH Intron C

Analysis of GH1C and GH2C sequences yielded highly congruent tree topologies (Figs. 2–4). Surprisingly, neither intron supported the conventional hypothesis of *Oncorhynchus* and *Salmo* as sister taxa. GH2C data produced a polychotomous node, while GH1C showed unequivocal support for *Salvelinus*/*Oncorhynchus* forming a monophyletic group. A likelihood ratio

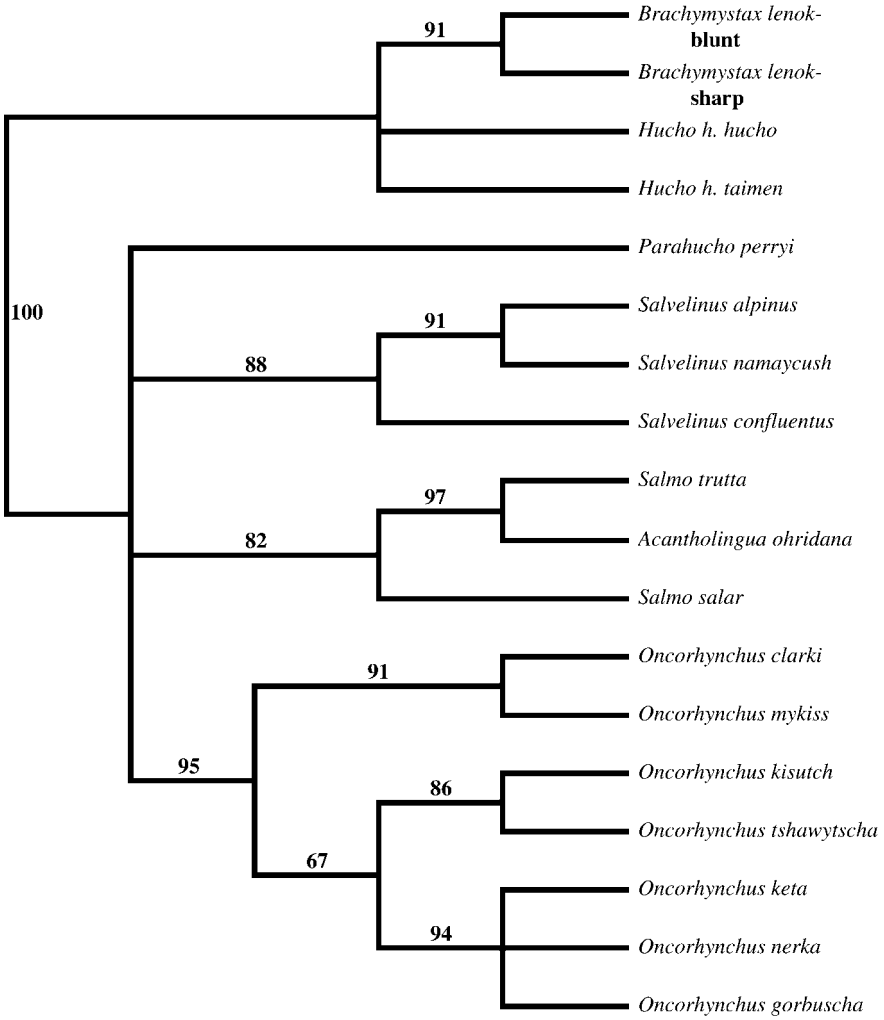

FIG. 2. Results of phylogenetic analysis of GH2C using parsimony. Strict consensus of 78 MP trees by treating gaps as missing. Numbers above nodes are nonparametric bootstrap percentages based on 1000 replications.

test (Huelsenbeck and Bull, 1996) failed to indicate significant incongruence between GH1C and GH2C data ($\alpha = 0.05$; $P = 0.19$). A separate likelihood ratio test did refute the hypothesis that corresponding branch lengths are equal between introns ($P < 0.0001$). This is also illustrated by the widely different number of informative sites (14:1) when considering four genera (Table 3).

Bootstrapping was useful for addressing taxonomic hypotheses. Based on GH2C, the node including *perryi, Salvelinus, Oncorhynchus,* and *Salmo* (and not *Hucho*) was supported by a 100% nonparametric bootstrap value and had the most synapomorphies of any node on the tree (13). Parametric bootstrapping (Fig. 5) also rejected the null hypothesis of *Hucho*/*perryi* monophyly ($P < 0.01$). Analysis of GH2C supported inclusion of *ohridana* in *Salmo* with a 82% nonparametric bootstrap value and rejection of nonmonophyly by a parametric bootstrapping test ($P < 0.01$). Monophyly of *Oncorhynchus* was also supported by high nonparametric bootstrap values, and parametric bootstrapping (Fig. 5) rejected nonmonophyly ($P < 0.01$ for both introns).

Are Oncorhynchus and Salmo Sibling Taxa?

Six four-taxon data sets were available to assess support for grouping *Oncorhynchus* and *Salmo* together as sister genera (Topology A of Fig. 6). Combining all data resulted in 44 synapomorphies favoring topology C—*Oncorhynchus*/*Salmo* not sister taxa, and 41 supporting A—*Oncorhynchus*/*Salmo* as sister taxa (Table 3). Most of the support for a sister relationship came from morphological data while GH1C and cyt-*b* most supported a different grouping. Based on an ILD test, significant incongruence was observed when all six data partitions were combined ($P = 0.001$). Since

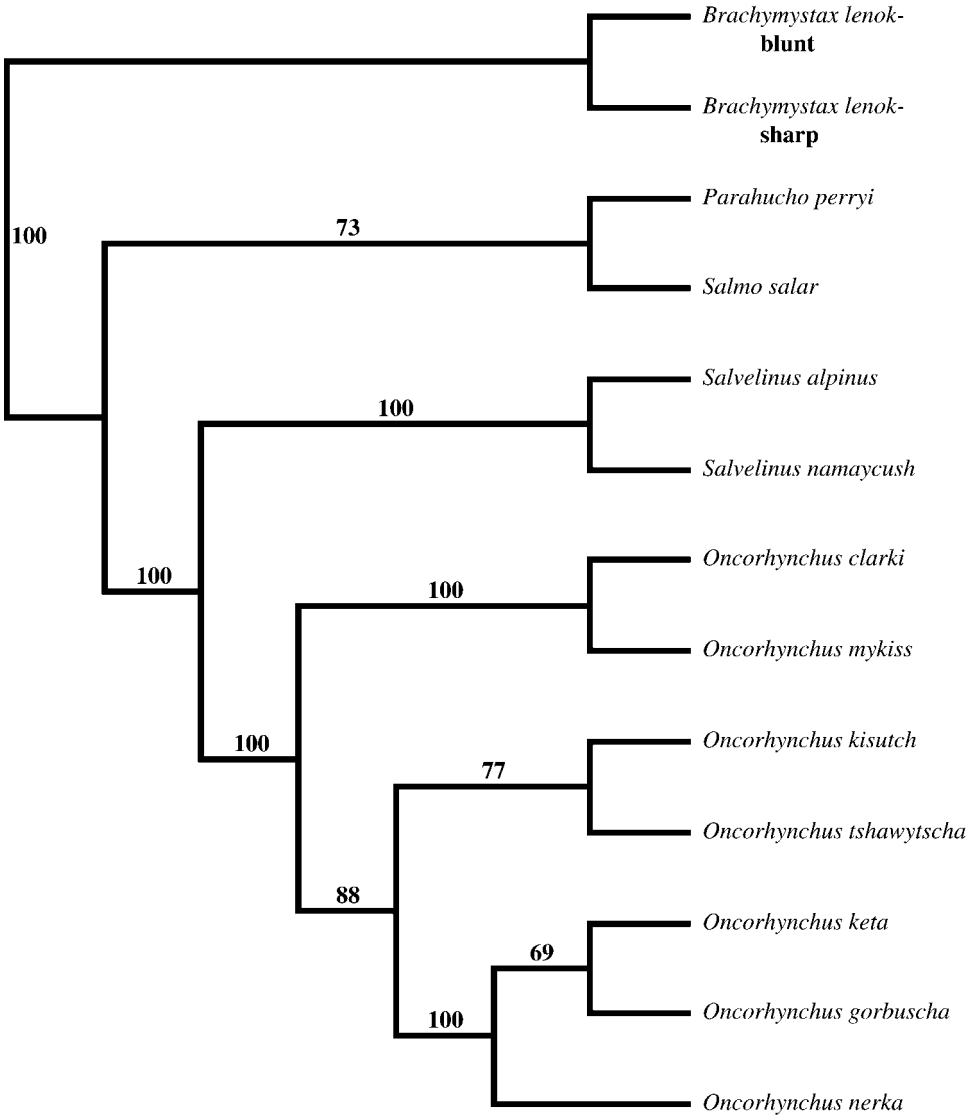

FIG. 3. The single most parsimonious tree based on analysis of GH1C. Numbers above nodes are nonparametric bootstrap percentages based on 1000 replications.

morphological data and GH1C data were the only data partitions with significant phylogenetic signal (Table 3) and since they supported different topologies, both were removed from the four-taxon analysis and incongruence was tested. The remaining four data sets were not significantly incongruent ($P = 0.205$), but parsimony and likelihood gave different answers: *Oncorhynchus*/*Salmo* as sister groups was favored using parsimony but not using likelihood (Tables 3 and 4). Using likelihood, two of four data partitions as well as all combined data indicated that a sister relationship was not the most likely result (Table 4). It is also interesting that, based on ITS2 data, parsimony suggested a sister relationship but likelihood analysis did not.

DISCUSSION

Phylogenetic Analysis of GH Intron C

Analysis of morphological data by different authors has produced conflicting taxonomic hypotheses in salmonid fishes. Analysis of GH introns provided clear evidence refuting some of these confounding hypotheses. Parsimony analyses of GH2C, for example, strongly supported paraphyly of *Hucho* when *perryi* is included in the genus. Both parametric and nonparametric bootstrapping support this. In addition, separate analysis of GH2C insertion–deletion events (indels) supports the same conclusion, and a large deletion is present in all species except *Brachymystax lenok* and

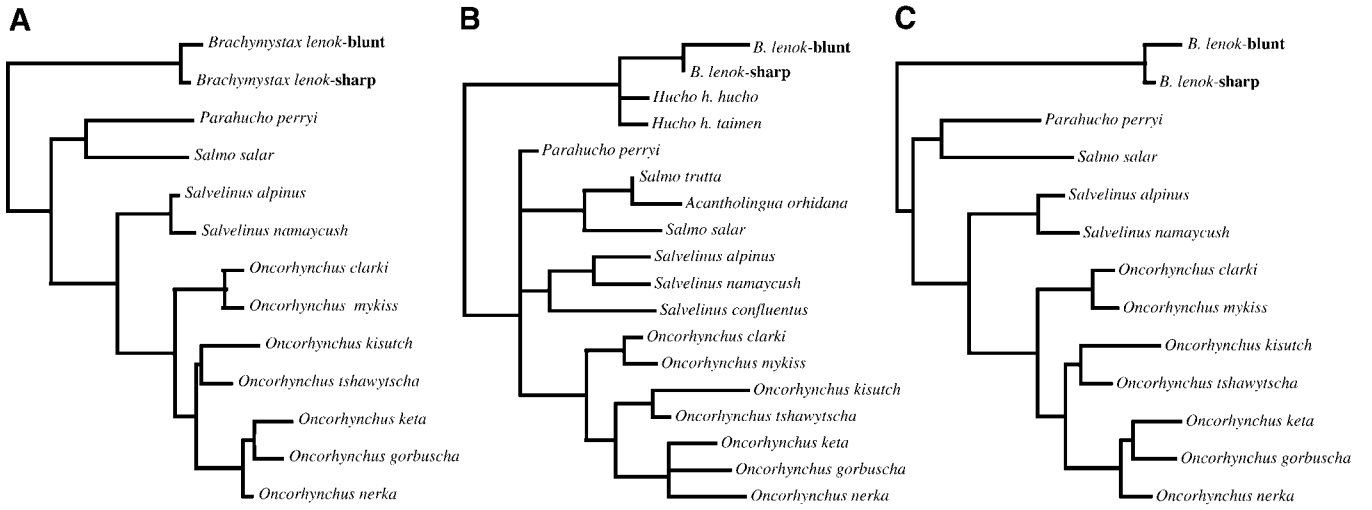

FIG. 4. (A) Maximum likelihood (ML) tree obtained from GH1C data using the HKY85 + Γ model of evolution. (B) ML tree from GH2C using F81 model of evolution. (C) ML tree of GH1C + GH2C data using the HKY85 + Γ model. Models used were determined to be optimum for the given data set (Table 2).

Hucho hucho. These data clearly show that the monotypic genus *Parahucho* is necessary to prevent paraphyly of *Hucho* in agreement with earlier studies (Phillips *et al.*, 1995; Phillips and Oakley, 1997).

Phylogenetic placement of *Acantholingua ohridana* based on morphological features has been ambiguous (Stearley and Smith, 1993). The GH2C sequence in this report is the first molecular data available for the fish and analysis of the intron indicates a close relationship to *Salmo trutta.* The position as a sister species to *S. trutta* with *S. salar* basal to the pair rejects subgeneric status of *Acantholingua.* The molecular data therefore support a change in genus and the species name *Salmo ohridana* and indicate that the taxon name *Acantholingua* is not necessary.

In agreement with morphology (Smith and Stearley, 1989; Stearley and Smith, 1993) and other molecular data (Devlin, 1993; Blackhall, 1994; McKay *et al.*, 1996) analyses of both GH introns (GH1C and GH2C) support monophyly of *Oncorhynchus.* The genus was recently revised to include Pacific trouts as well as Pacific salmon (Smith and Stearley, 1989), and both intron data sets studied here included two Pacific trouts (*O. mykiss* and *O. clarki*) as well as five Pacific salmon.

While several questions of genus monophyly were clearly answered by GH intron data, only GH1C resolved the branching order of the genera, as GH2C data produced a polychotomous node between *Parahucho, Salmo, Salvelinus,* and *Oncorhynchus.* Although the topologies were not significantly different, the corresponding branch lengths were, suggesting that the introns are evolving at different rates. This difference in branch length is obvious when comparing maximum likelihood trees of GH1C (Fig. 4A) and GH2C (Fig. 4B).

Relationships suggested by GH1C (Figs. 3 and 4A) and by GH1C and GH2C combined (Fig. 4C) did not support the widely accepted hypothesis of a monophyly of *Salmo* and *Oncorhynchus,* instead pairing *Parahucho* with *Salmo* and *Salvelinus* with *Oncorhynchus.* Analysis of previously published data allowed for a comparison of amounts of support for these different hypotheses.

Are Oncorhynchus and Salmo Sister Taxa?

How to proceed with multiple data sets is currently one of the most debated issues in systematics (Kluge, 1989; Miyamoto and Fitch, 1995; Huelsenbeck *et al.*, 1996). One option is to combine all data, whether or not conflicting signal exists (Kluge, 1989). Another approach is conditional combination (Bull *et al.*, 1993), in which partitions are combined only if there is not significant conflict. If conflict is present, the model of phylogenetic analysis is adjusted in the hopes of reducing conflicting signal so that data can then be combined.

Although somewhat ambiguous, the majority of data currently supports topology C of Fig. 6—that *Oncorhynchus* and *Salmo* are not sister taxa. Total evidence separates *Oncorhynchus* and *Salmo,* but a difference of only three synapomorphies (41 vs 44) indicates the presence of conflicting phylogenetic signal. This conflict was significant, based on an ILD test using all six data partitions. After removing GH1C and morphology data, the four remaining partitions were not significantly incongruent, but parsimony and likelihood gave different answers for the best tree using the remaining four data sets. Though parsimony favored grouping *Oncorhynchus* and *Salmo* by two synapomorphies, likeli-

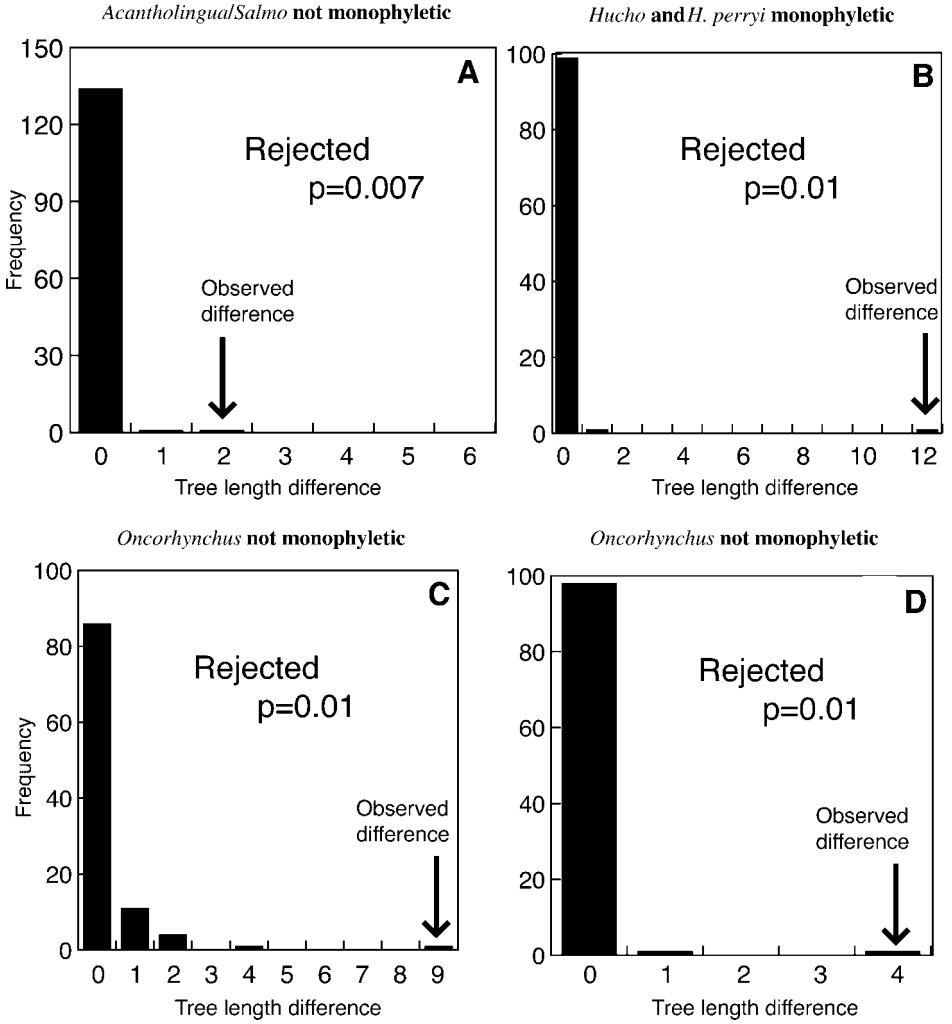

FIG. 5. Null distributions based on parametric bootstrap analysis of four taxonomic hypotheses. Data were simulated with likelihood estimates of parameters and distributions were determined by measuring the difference in tree length between null tree and best tree. (A) The null hypothesis that *Acantholingua ohridana* is separate from *Salmo* species is rejected using GH2C data. (B) The hypothesis that *perryi* is monophyletic within *Hucho* was rejected using GH2C data. (C) and (D) Nonmonophyly of *Oncorhynchus* was rejected with GH1C and GH2C data.

hood using models chosen statistically did not pair the genera. More data may further clarify this node, but since neither total evidence nor data analyzed with optimal likelihood models group the two genera, they cannot be considered sister taxa based on available data.

If not sister genera, the proposed phylogeny supports the hypothesis of Stearley (1992) that anadromy evolved independently in each genus. *Oncorhynchus* and *Salmo* contain the only salmonine species that exhibit anadromy—individuals return to their natal freshwater streams to spawn after spending time living at sea. If loss and gain of anadromy are equally likely, it is more parsimonious to infer independent origin than a single origin and multiple losses of anadromy based on the phylogeny in Fig. 4C. This assumes that gain and loss of anadromy are equiprobable (Omland, 1997; Cunningham *et al.*, 1998), but exploring this assumption is beyond the scope of this paper. Besides phylogenetic evidence, Stearley (1992) also presented nonphylogenetic data that supports the independent anadromy hypothesis.

CONCLUSIONS

The phylogenetic trees based on GH intron data (Figs. 2–4) appear among the best hypotheses of salmonid relationships currently available. Trees based on these introns have well-supported nodes and are highly congruent, and the only area in which there was not exact corroboration was well-studied using several other data sets. These data clearly support several controversial taxonomic proposals. *Parahucho* is a valid, monotypic genus, *Oncorhynchus* is monophyletic, and

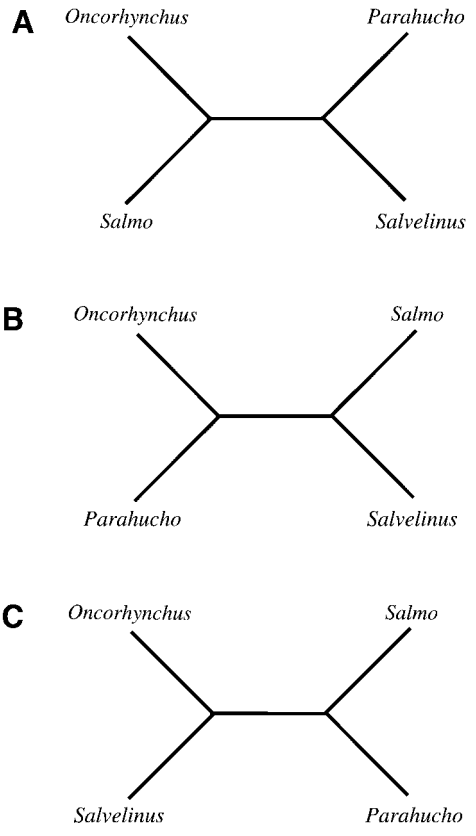

FIG. 6. The three possible tree topologies for four salmonid genera that are referred to in the text and tables as A, B, and C.

Acantholingua should be included within *Salmo*. Finally, the commonly accepted pairing of *Salmo* and *Oncorhynchus* as sister genera is not supported; however, data from this node is somewhat ambiguous and would benefit from further study.

ACKNOWLEDGMENTS

Many thanks to all the collectors listed in Table 1 for providing specimens. Also thanks to C. Cunningham, S. Hoot, J. Marcus, J. Wares, and two anonymous reviewers for comments that improved earlier versions of the manuscript. This work was supported by a grant from the National Science Foundation to R.B.P. The Milwaukee County Zoological Society provided funds to T.H.O. for necessary computer software. All sequence alignments are available at http://www.duke.edu/~tho/GHdata.html or from the authors by request.

TABLE 4

Likelihood Analysis of *Oncorhynchus, Salmo, Parahucho,* and *Salvelinus* with Multiple Data Sets

Data set (model)	Topology A	Topology B	Topology C
cyt-*b* (GTR + Γ)	−2808.21	−2807.88	−2805.46[a]
ITS1 (GTR + Γ)	−1516.84	−1514.09[a]	−1517.83
ITS2 (GTR + Γ)	−789.24	−790.01	−787.647[a]
GH1C (HKY85)	−1356.04	−1356.04	−1330.01[a]
GH2C (F81)	−756.75	−756.00[a]	−756.75
Combined	−7227.08	−7224.02	−7197.70[a]

Note. Model abbreviations as in Table 2. Topology A–C refers to Fig. 6.

[a] Best tree for given data set.

REFERENCES

Agellon, L. B., Davies, S. L., Lin, C. M., Chen, T. T., and Powers, D. A. (1988). Rainbow trout has two genes for growth hormone. *Mol. Reprod. Dev.* **1:** 11–17.

Allendorf, F. W., and Thorgaard, G. H. (1984). Tetraploidy and the evolution of salmonid fishes. *In* "Evolutionary Genetics of Fishes" (B. J. Turner, Ed.), pp. 1–53. Plenum, New York.

Behnke, R. J. (1968). A new subgenus and species of trout: *Platysalmo platycephalus* from southcentral Turkey with comments on the classification of the subfamily Salmoninae. *Mitt. Hambg. Zool. Mus. Inst.* **66:** 1–15.

Behnke, R. J. (1980). A systematic review of the genus *Salvelinus*. *In* "Charrs: Salmonid Fishes of the Genus *Salvelinus*" (E. K. Balon, Ed.), pp. 441–481. Junk, The Hague.

Bernatchez, L., and Osinov, A. (1995). Genetic diversity of trout (genus *Salmo*) from its most eastern native range based on mitochondrial DNA and nuclear gene variation. *Mol. Ecol.* **4:** 285–297.

Blackhall, W. J. (1994). "A Molecular Study of Introgression between Westslope Cutthroat Trout and Rainbow Trout," M.S. Thesis. Univ. of Alberta, Edmonton, Alberta, Canada.

Bull, J. J., Cunningham, C. W., Swofford, D. L., and Waddell, P. J. (1993). Partitioning and combining data in phylogenetic analysis. *Syst. Biol.* **42:** 384–397.

Cavender, T. M. (1980). Systematics of *Salvelinus* from the North Pacific Basin. *In* "Charrs: Salmonid Fishes of the Genus *Salvelinus*" (E. K. Balon, Ed.), pp. 295–322. Junk, The Hague.

Cunningham, C. W., Omland, K. E., and Oakley, T. H. (1998). Reconstructing ancestral character states: A critical reappraisal. *Trends Ecol. Evol.*, in press.

Devlin, R. H. (1993). Sequence of sockeye type 1 and type 2 growth hormone genes and the relationships of rainbow trout with Atlantic and Pacific salmon. *Can. J. Fish. Aquat. Sci.* **50:** 1738–1748.

Domanico, M. J. (1994). "Phylogenetic Analysis of the Pacific Salmon Using Nuclear and Mitochondrial DNA Sequences," Ph.D. Thesis. Univ. of Wisconsin-Milwaukee, WI.

Domanico, M. J., Phillips, R. B., and Oakley, T. H. (1997). Phylogenetic analysis of the Pacific salmon (genus *Oncorhynchus*) using nuclear and mitochondrial DNA sequences. *Can. J. Fish. Aquat. Sci.* **54:** 1865–1872.

Dorofeyeva, E. A. (1989). The basic principles of classification and phylogeny of the salmonid fishes (Salmoniformes: Salmonoidei: Salmonidae). *In* "Biology and Phylogeny of Fishes" (V. M. Korovinoi, Ed.), pp. 5–16. USSR Academy of Sciences, Proceedings of the Zoological Institute 201, St. Petersburg.

Farris, J. S., Källersjö, M., Kluge, A. G., and Bult, C. (1994). Testing significance of incongruence. *Cladistics* **10:** 315–319.

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: A maximum likelihood approach. *J. Mol. Evol.* **17:** 368–376.

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

Forbes, S. H., Knudsen, K. L., North, T. W., and Allendorf, F. W. (1994). One of two growth hormone genes in coho salmon is sex-linked. *Proc. Natl. Acad. Sci. USA* **91:** 1628–1631.

Giuffra, E., Bernatchez, L., and Guyomard, R. (1994). Mitochondrial control region and protein coding genes sequence variation among

phenotypic forms of brown trout *Salmo trutta* from northern Italy. *Mol. Ecol.* **3:** 161–171.

Goldman, N. (1993). Statistical tests of models of DNA substitution. *J. Mol. Evol.* **36:** 182–198.

Hadzisce, S. (1961). Zur Kentnis des *Salmothymus orhidana* (Steindachner) (Pisces, Salmonidae). *Int. Ver. Theor. Angew. Limnol. Verh.* **14:** 785–791.

Hartley, S. E. (1987). Chromosomes of salmonid fishes. *Biol. Revue* **62:** 197–214.

Hasegawa, M., Kishino, H., and Yano, T. (1985). Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **21:** 160–174.

Higgins, D. G., Bleasby, A. J., and Fuchs, R. (1992). CLUSTAL V: Improved software for multiple sequence alignment. *Comput. Appl. Biosci.* **8:** 189–191.

Hillis, D. M., Mable, B. K., and Moritz, C. (1996). Applications of molecular systematics: The state of the field and a look to the future. *In* "Molecular Systematics" (D. M. Hillis, C. Moritz, and B. K. Mable, Eds.), pp. 515–543. Sinauer, Sunderland, MA.

Holcik, J. (1988). "The Eurasian Huchen, *Hucho hucho*," Junk, Dordrecht, The Netherlands.

Huelsenbeck, J. P., and Bull, J. J. (1996). A likelihood ratio test to detect conflicting phylogenetic signal. *Syst. Biol.* **45:** 92–98.

Huelsenbeck, J. P., Bull, J. J., and Cunningham, C. W. (1996). Combining data in phylogenetic analysis. *Trends Ecol. Evol.* **11:** 152–158.

Huelsenbeck, J. P., Hillis, D. M., and Jones, R. (1995). Parametric bootstrapping in molecular phylogenetics: Applications and performance. *In* "Molecular Zoology: Strategies and Protocols" (J. Ferraris and S. Palumbi, Eds.), pp. 19–45. Wiley, New York.

Huelsenbeck, J. P., Rannala, B., and Yang, Z. (1997). Statistical tests of host–parasite cospeciation. *Evolution* **51:** 410–419.

Hughes, A. L. (1994). The evolution of functionally novel proteins after gene duplication. *Phil. Trans. R. Soc. B* **256:** 119–124.

Hughes, M. K., and Hughes, A. L. (1993). Evolution of duplicated genes in a tetraploid animal, *Xenopus laevis*. *Mol. Biol. Evol.* **10:** 1360–1369.

Jukes, T. H., and Cantor, C. R. (1969). Evolution of protein molecules. *In* "Mammalian Protein Metabolism" (H. N. Munro, Ed.), pp. 21–132. Academic Press, New York.

Kendall, A. W., Jr., and Behnke, R. J. (1984). Salmonidae: Development and relationships. *In* "Ontogeny and Systematics of Fishes" (H. G. Moser, Ed.), pp. 142–149. Am. Soc. of Icthyol. Herpetol. Spec. Publ. 1.

Kimura, M., and Ohta, T. (1974). On some principles governing molecular evolution. *Proc. Natl. Acad. Sci. USA* **71:** 2848–2852.

Kitano, T., Matsuoka, N., and Saitou, N. (1997). Phylogenetic relationship of the genus *Oncorhynchus* species inferred from nuclear and mitochondrial markers. *Genes Genet. Syst.* **72:** 25–34.

Kluge, A. G. (1989). A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). *Syst. Zool.* **18:** 1–32.

Lanave, C., Preparata, G., Saccone, C., and Serio, G. (1984). A new method for calculating evolutionary substitution rates. *J. Mol. Evol.* **20:** 86–93.

Li, W. H. (1980). Rate of gene silencing at duplicate loci: A theoretical study and interpretation of data from tetraploid fishes. *Genetics* **95:** 237–258.

Marchuk, D., Drumm, M., Saulino, A., and Collins, S. (1991). Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. *Nucleic Acids Res.* **19:** 1154.

McKay, S. J., Devlin, R. H., and Smith, M. J. (1996). Phylogeny of Pacific salmon and trout based on growth hormone type-2 (GH2) and mitochondrial NADH dehydrogenase subunit 3 (ND3) DNA sequences. *Can. J. Fish. Aquat. Sci.* **53:** 1165–1176.

Miyamoto, M. M., and Fitch, W. M. (1995). Testing species phylogenies and phylogenetic methods with congruence. *Syst. Biol.* **44:** 64–76.

Norden. (1961). Comparative osteology of representative salmonid fishes, with particular reference to the grayling (*Thymallus arcticus*) and its phylogeny. *J. Fish. Res. Board Can.* **8:** 679–791.

Ohno, S. (1970). "Evolution by Gene Duplication," Springer-Verlag, New York.

Omland, K. E. (1997). Examining two standard assumptions of ancestral reconstructions: Repeated loss of dichromatism in dabbling ducks (Anatini). *Evolution* **51:** 1636–1646.

Oohara, I., Sawano, K., and Okazaki, T. (1997). Mitochondrial DNA sequence analysis of the Masu salmon—phylogeny in the genus *Oncorhynchus*. *Mol. Phylogenet. Evol.* **7:** 71–78.

Osinov, A. G. (1991). Counter current dispersion and speciation in lenoks of the genus *Brachymystax*. *Genetika* **29:** 654–669.

Patarnello, T., Bargelloni, L., Caldara, F., and Colombo, L. (1994). Cytochrome *b* and 16S rRNA sequence variation in the *Salmo trutta* (Salmonidae, Teleostei) species complex. *Mol. Phylogenet. Evol.* **3:** 69–74.

Phillips, R. B., and Oakley, T. H. (1997). Phylogenetic relationships among the Salmoninae based on nuclear and mitochondrial DNA sequences. *In* "Molecular Systematics of Fishes" (T. D. Kocher and C. A. Stepien, Eds.), pp. 145–162. Academic Press, San Diego.

Phillips, R. B., Oakley, T. H., and Davis, E. L. (1995). Evidence for the paraphyly of *Hucho* (Salmonidae) based on rDNA RFLP's. *J. Fish Biol.* **47:** 956–961.

Pleyte, K. A., Duncan, S. D., and Phillips, R. B. (1992). Evolutionary relationships of the salmonid genus *Salvelinus* inferred from DNA sequences of the first internal transcribed spacer (ITS 1) of ribosomal DNA. *Mol. Phylogenet. Evol.* **1:** 223–230.

Rambaut, A., and Grassly, N. C. (1997). Seq-Gen: An application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Comput. Appl. Biosci.* **13:** 235–238.

Rentier-Delrue, F., Swennen, D., Mercier, L., Lion, M., Benrubi, O., and Martial, J. A. (1989). Molecular cloning and characterization of two forms of trout growth hormone cDNA: Expression and secretion of tGH-II by *Escherichia coli*. *DNA* (*N.Y.*) **8:** 271–278.

Rodriguez, F., Oliver, J. L., Marin, A., and Medina, J. R. (1990). The general stochastic model of nucleotide substitution. *J. Theor. Biol.* **142:** 485–501.

Savvaitova, K. A. (1980). Taxonomy and biogeography of charrs in the Paleoarctic. *In* "Charrs, Salmonid Fishes of the Genus *Salvelinus*" (E. K. Balon, Ed.), pp. 281–294. Junk, The Hague.

Smith, G. R., and Stearley, R. F. (1989). The classification and scientific names of rainbow and cutthroat trouts. *Fisheries* **14:** 4–10.

Stearley, R. F. (1992). Historical ecology of Salmoninae, with special reference to *Oncorhynchus*. *In* "Systematics, Historical Ecology, and North American Freshwater Fishes" (R. L. Mayden, Ed.), Stanford Univ. Press, Stanford, CA.

Stearley, R. F., and Smith, G. R. (1993). Phylogeny of the Pacific trouts and salmons (*Oncorhynchus*) and genera of the family Salmonidae. *Trans. Am. Fish. Soc.* **122:** 1–33.

Svetovidov, A. (1975). Comparative osteological study of the Balkan endemic genus *Salmothymus* in relation to its classification. *Zool. Zh.* **54:** 1174–1190.

Swofford, D. L. (1993). PAUP 3.1.1: Phylogenetic analysis using parsimony, Illinois Natural History Survey., Champaign, Illinois.

Swofford, D. L. (1997). PAUP* 4.0x: Phylogenetic analysis using parsimony and other methods 4.0, Smithsonian Institution, Washington D.C.

Swofford, D. L., Thorne, J. L., Felsenstein, J., and Wiegmann, B. M. (1996). The topology-dependent permutation test for monophyly does not test for monophyly. *Syst. Biol.* **45:** 575–579.

Swofford, D. L., Olsen, G. J., Waddell, P. J., and Hillis, D. M. (1996). Phylogenetic inference. *In* "Molecular Systematics" (D. M. Hillis, C. Moritz and B. K. Mable, Eds.). Sinauer, Sunderland, MA.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22:** 4673–4680.

Tavare, S. (1986). Some proabilistic and statistical problems on the analysis of DNA sequences. *Lec. Math. Life Sci.* **17:** 57–86.

Yang, Z. (1993). Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Mol. Biol. Evol.* **10:** 1396–1401.