

# Addressing incomplete lineage sorting and paralogy in the inference of uncertain salmonid phylogenetic relationships

Matthew A. Campbell<sup>1</sup>, Thaddaeus J. Buser<sup>2</sup>, Michael E. Alfaro<sup>3</sup> and J. Andrés López<sup>1,4</sup>

<sup>1</sup> University of Alaska Museum, University of Alaska—Fairbanks, Fairbanks, AK, USA

<sup>2</sup> Department of Fisheries and Wildlife, Oregon State University, Corvallis, OR, USA

<sup>3</sup> Department of Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, CA, USA

<sup>4</sup> College of Fisheries and Ocean Sciences, University of Alaska—Fairbanks, Fairbanks, AK, USA

## ABSTRACT

Recent and continued progress in the scale and sophistication of phylogenetic research has yielded substantial advances in knowledge of the tree of life; however, segments of that tree remain unresolved and continue to produce contradicting or unstable results. These poorly resolved relationships may be the product of methodological shortcomings or of an evolutionary history that did not generate the signal traits needed for its eventual reconstruction. Relationships within the euteleost fish family Salmonidae have proven challenging to resolve in molecular phylogenetics studies in part due to ancestral autopolyploidy contributing to conflicting gene trees. We examine a sequence capture dataset from salmonids and use alternative strategies to accommodate the effects of gene tree conflict based on aspects of salmonid genome history and the multispecies coalescent. We investigate in detail three uncertain relationships: (1) subfamily branching, (2) monophyly of *Coregonus* and (3) placement of *Parahucho*. Coregoninae and Thymallinae are resolved as sister taxa, although conflicting topologies are found across analytical strategies. We find inconsistent and generally low support for the monophyly of *Coregonus*, including in results of analyses with the most extensive dataset and complex model. The most consistent placement of *Parahucho* is as sister lineage of *Salmo*.

**Subjects** Aquaculture, Fisheries and Fish Science, Biodiversity, Evolutionary Studies

**Keywords** *Coregonus*, Gene tree and species tree conflict, Incomplete lineage sorting, *Parahucho*, Salmonidae, Segmental polyploidy, Phylogeny, Ohnolog resolution

## INTRODUCTION

Within ray-finned fishes (Actinopterygii), there are numerous relationships that are challenging to resolve. Historically, anatomical characteristics have provided great insight into the evolutionary relationships of some groups, but failed to confidently resolve others (*Betancur-R et al., 2017*). In the past few decades, DNA sequence data has emerged as a method to gather numerous characters for phylogenetic inference. Sequence data generation in recent years has undergone dramatic and rapid evolution with respect to the cost and scale of DNA sequence determination and has fostered the production of

Submitted 3 February 2020

Accepted 28 May 2020

Published 3 July 2020

Corresponding author

Matthew A. Campbell,  
maccampbell2@alaska.edu

Academic editor

Agnes Dettai

Additional Information and  
Declarations can be found on  
page 19

DOI 10.7717/peerj.9389

© Copyright

2020 Campbell et al.

Distributed under

Creative Commons CC-BY 4.0

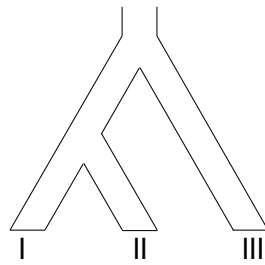
**OPEN ACCESS**

genomic-scale datasets for phylogenetic estimation. The theory and implementations of methods for analyzing these data have increased in complexity and performance broadly in parallel with increased data generation (Muir *et al.*, 2016). The sheer quantity of data contained within genome-scale datasets has contributed to expectations that molecular phylogenetics may provide conclusive results regarding previously elusive evolutionary relationships. While additional anatomical investigation and the application of molecular phylogenetics has aided the resolution of some groups, such as the Scorpaenoidea (Smith, Everman & Richardson, 2018), counterexamples certainly remain. In particular, there are parts of the ray-finned fish tree of life that remain poorly resolved even with the application of high-throughput sequencing data sets of substantial numbers of loci. These persistently challenging-to-resolve relationships include those among early-branching euteleost lineages (Campbell *et al.*, 2017) and the intra-clade relationships of Pelagiaria (Campbell *et al.*, 2018; Friedman *et al.*, 2019; Miya *et al.*, 2013).

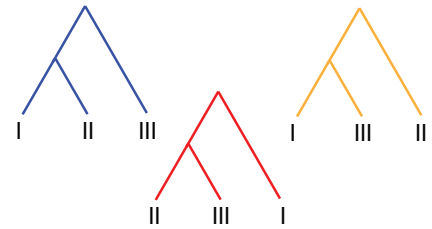
While conclusive resolution is desirable, evolutionary relationships can be expected to remain unresolved with molecular data under certain scenarios. For example, when the characteristics of the evolutionary process being investigated generated signals with ambiguous origins, prevented the formation and preservation of a phylogenetic signal, or when the methodological framework used to investigate that process is poorly suited for the data type. Discordance between gene trees or gene tree–species tree (GT–ST) conflict (Fig. 1) represents a special challenge in the inference of recalcitrant relationships because it may be the combined product of conflicting histories and functionally undetectable phylogenetic signal with difficult to disentangle contributions from each factor.

One predicted cause of GT–ST conflict is incomplete lineage sorting (ILS), which may be defined as the failure for two or more allelic lineages to coalesce within a population (Degnan & Rosenberg, 2009). Incompletely sorted diversity can lead to conflicting gene tree topologies if the time between speciation events is small and/or the population sizes are large (Fig. 2). The continued investigation of natural systems has emphasized the pervasiveness of hybridization (Mallet, 2005), and divergence times before or after hybridization may result in ILS and gene tree incongruence signatures identical to patterns exhibited by hybridization alone (Yu *et al.*, 2011). In tree-based evolutionary analyses, hybridization events are very easily overlooked, and GT–ST conflict will be considered a consequence of ILS. Alternatively, in phylogenetic network analysis, such reticulations are largely taken as hybridization (Yu *et al.*, 2011). Another potential source of GT–ST conflict is gene or genome duplication events. Genes that are duplicated, particularly those arising from whole-genome duplication (WGD) are frequently lost (Berthelot *et al.*, 2014; Scannell *et al.*, 2007). However duplication of a gene followed by loss of alternative copies in different lineages may lead to GT–ST conflict (Fig. 3). While massive gene loss is typical following WGD as polyploid organisms revert to a normal ploidy level (Gerstein & Otto, 2009), many genes in the genomes of lineages with ancestral genome duplications exist in a continuum from polyploidy to normal ploidy, a condition known as segmental polyploidy. In lineages where genomes exhibit segmental polyploidy, ILS can be expected

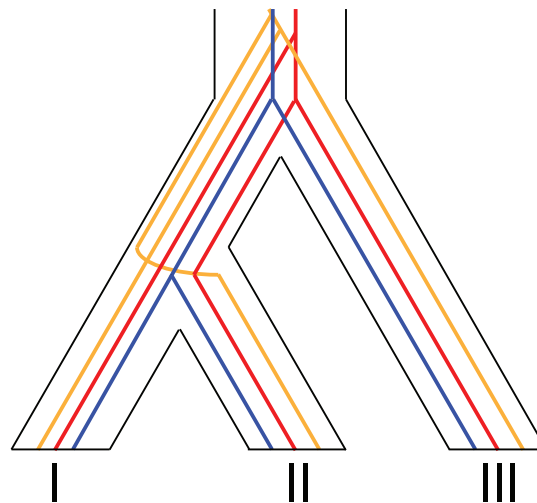
A. Species Relationships: ((I, II), III)



B. Possible Gene Trees



C. Different Gene Trees within Species Tree

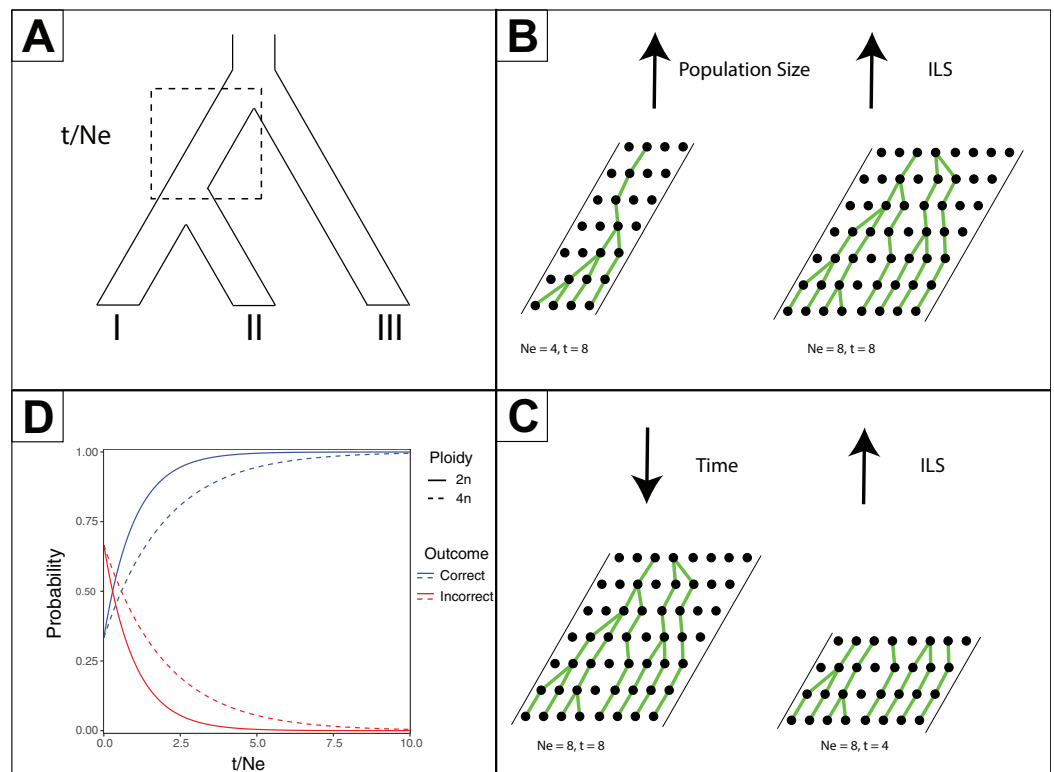


**Figure 1** Depiction of gene tree and species tree conflict. (A) Hypothetical species relationships between three species, I, II and III. (B) Depiction of three possible gene trees. (C) Combined species tree and gene trees. Only one of the gene trees (blue) follows the species tree relationships.

Full-size  DOI: [10.7717/peerj.9389/fig-1](https://doi.org/10.7717/peerj.9389/fig-1)

to be more pronounced in the tetrasomic regions of genome compared to the disomic regions as their respective population sizes differ (Figs. 2B and 2D).

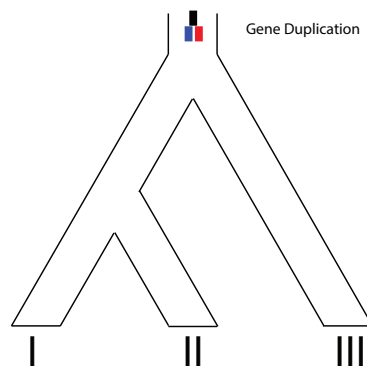
Whole-genome duplication events are known from across many eukaryotic lineages (Campbell *et al.*, 2016) and some lineages have experienced successive rounds of WGD producing nested levels of ancestral polyploidy in the genomes of descendant lineages (Meyers, Levin & Geber, 2006; Wendel, 2015). Genomes of extant fishes are partly shaped by two rounds of WGD (commonly noted as 1R and 2R; Fig. S1) which are inferred to have taken place 450–600 million years ago (mya) (Dehal & Boore, 2005). A third well-supported WGD (3R) is inferred in the common ancestor of teleosts at around 350 mya (Meyer & Van de Peer, 2005). Additional WGD events are scattered through lineages in the fish tree of life (Leggatt & Iwama, 2003). A WGD occurred in the common ancestor of the Salmoniformes some time after the lineage diverged from the Esociformes. This WGD event occurred ~88 mya ago and it is known as the Ss4R (salmonid-specific fourth round) (Berthelot *et al.*, 2014; Macqueen & Johnston, 2014).



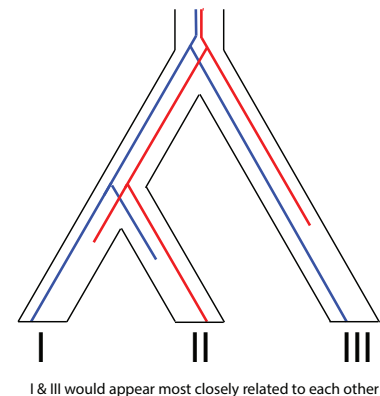
**Figure 2** Effects of time and effective population size ( $N_E$ ) on incomplete lineage sorting (ILS). (A) Species tree depicting relationships between species I, II and III. The branch leading to the common ancestor of I and II is indicated in a box as where the ratio of time ( $t$ , in generations) versus  $N_E$  is considered for the other subpanels. (B) Depiction of how increased population sizes increase ILS. A single coalescent event is depicted for each generation with population sizes of eight and four. (C) Depiction of decreased time increasing ILS. For a population of eight, four or eight generations are shown. (D) For a three-taxon species tree (as in Subpanel A), the probability of inferring a correct or incorrect gene tree is plotted as a function of  $t/N_E$ . As there are three possible outcomes, and only one correct one, the initial probability of inferring an incorrect gene tree is  $2/3$  at  $t/N_E = 0$ . The effect of different ploidy levels is shown, with increased ploidy increasing ILS. [Full-size !\[\]\(5f471a71b78d7676bc356df190b88ab4\_img.jpg\) DOI: 10.7717/peerj.9389/fig-2](https://doi.org/10.7717/peerj.9389/fig-2)

Salmonids are a group of 228 fish species divided among three subfamilies: Coregoninae (whitefishes, ciscos, and innocu: 86 species), Thymallinae (graylings: 18 species), and Salmoninae (salmons, trouts, charrs, taimens: 124 species) (Fricke, Eschmeyer & Fong, 2019; Nelson, 2006). Despite the commercial and cultural importance of many salmonids, and intensive research on many of the commercially exploited species, questions about their evolutionary relationships remain unanswered. While monophyly of each of the three subfamilies is generally accepted, relationships between these subfamilies remain contentious. A sister-group relationship between Salmoninae and Thymallinae (S + T) has been supported by morphological (Sanford, 1990; Wilson & Williams, 2010) and molecular evidence (Betancur-R *et al.*, 2013; Shedko, Miroshnichenko & Nemkova, 2012; Yasuike *et al.*, 2010). In contrast, several molecular phylogenetic studies have offered support for a clade comprising Coregoninae and Salmoninae (C + S) (Alexandrou *et al.*, 2013; Crête-Lafrenière, Weir & Bernatchez, 2012; Near *et al.*, 2012). Finally, Coregoninae sister

A. Species Relationships: ((I, II), III)



B. Duplication Followed by Loss



**Figure 3** Example of gene duplication and loss misleading phylogenetic inference. (A) Species tree of three species depicting a gene duplication event in the common ancestor of all three species. (B) Example of loss of paralogs in different lineages. Each species exhibits a single copy of the formerly duplicated gene; however, construction of a phylogeny from those sequences would make it appear that species I and species III would be most closely related to each other. [Full-size !\[\]\(74d4806277d7e73349d8e8c0897931e9\_img.jpg\) DOI: 10.7717/peerj.9389/fig-3](https://doi.org/10.7717/peerj.9389/fig-3)

to Thymallinae (C + T) has been indicated by several analyses ([Burridge et al., 2012](#); [Campbell et al., 2013](#); [Horreo, 2017](#); [Li et al., 2010](#); [Macqueen & Johnston, 2014](#); [Robertson et al., 2017](#)). Thus, each of the three possible arrangements of the subfamilies has found some support in the literature. Another point of contention is the relationship between the coregonine genera *Coregonus* and *Stenodus*. Previous studies have variously placed *Stenodus* as sister to ([Bodaly et al., 1991](#); [Horreo, 2017](#); [Sajdak & Phillips, 1997](#); [Vuorinen et al., 1998](#)) or nested within *Coregonus* ([Bernatchez, Colombani & Dodson, 1991](#); [Crête-Lafrenière, Weir & Bernatchez, 2012](#)). A final point of debate is the placement of *Parahucho* within Salmoninae. Previous studies have variously placed *Parahucho* in a polytomy with *Oncorhynchus* + *Salmo* + *Salvelinus* ([Shedko, Miroshnichenko & Nemkova, 2012](#)) or as sister to: the genus *Salmo* ([Crespi & Fulton, 2004](#); [Lecaudey et al., 2018](#); [Matveev, Nishihara & Okada, 2007](#); [Oakley & Phillips, 1999](#)); the genus *Salvelinus* ([Campbell et al., 2013](#); [Crête-Lafrenière, Weir & Bernatchez, 2012](#); [Horreo, 2017](#)); a clade composed of *Oncorhynchus* + *Salvelinus* ([Crête-Lafrenière, Weir & Bernatchez, 2012](#)); or a clade composed of *Salmo* + *Salvelinus* + *Oncorhynchus* ([Alexandrou et al., 2013](#)).

While phylogenetic uncertainty is not unique to salmonids, the numerous hypotheses of salmonid intrarelationships proposed in recent literature suggest serious challenges with molecular phylogenetic inference of the group and WGD may be a contributing factor ([Allendorf & Thorgaard, 1984](#); [Berthelot et al., 2014](#); [Lien et al., 2016](#); [Macqueen & Johnston, 2014](#); [Ohno, 1970](#)). Well-characterized genomes of salmonids reveal that 10–15% of their genomes are tetrasomically inherited, with high levels of overall genomic similarity throughout as a result of the Ss4R ([Lien et al., 2016](#); [Pearse et al., 2019](#)). Within rainbow trout (*Oncorhynchus mykiss*), for example, these duplicated regions contain 2,278 tetrasomically inherited pairs of genes ([Campbell et al., 2019](#); [Pearse et al., 2019](#)), meaning that 10.62% of the 42,884 annotated protein-coding genes in rainbow trout

are tetrasomically inherited. Comparative genome mapping shows broad conservation of partial tetrasomy across Salmoninae in eight chromosome arms, the “Magic Eight” with homologous regions in *Salmo*, *Salvelinus* and *Oncorhynchus* species indicated to be tetrasomic (Brieuc et al., 2014; Campbell et al., 2019; Kodama et al., 2014; Lien et al., 2011; Sutherland et al., 2016). Further analyses of recently produced genome assemblies and linkage mapping has continued to support the idea of the same genomic regions maintaining tetrasomy across salmonids (Blumstein et al., 2019). Genomic comparisons support that there are at least seven tetrasomically pairing chromosomes in salmonids, with possible intermediate pairs falling between tetrasomy and disomy present, and one tetrasomically pairing set identified only by linkage mapping. Tetrasomic recombination results in larger population sizes and duplicated loci for some genes and genome regions and may contribute to GT–ST conflict via ILS (Figs. 2B and 2D). While the path to rediploidization in salmonids has conserved patterns across species where rediploidization occurred prior to speciation, continued rediploidization within lineages after speciation has resulted in unique signatures—lineage-specific rediploidization or lineage-specific ohnolog resolution (LORe). These signatures are elevated in tetrasomically pairing regions of the salmonid genome (Robertson et al., 2017). Due to LORe, loci from tetrasomically pairing genomic regions of salmonid genomes are likely to have problematic orthology assignment. With a complex genome structure riddled with duplications and high-similarity regions that are undergoing independent rediploidization, the potential for the accidental inclusion of paralogous loci in phylogenetic analyses of this group is concerning (i.e., Fig. 3). Together both tetrasomy and lineage-specific rediploidization processes may be responsible for widespread GT–ST conflict in salmonids.

In this article we analyze a DNA sequence dataset from 500 Ultra Conserved Element (UCE) loci. These UCE loci were designed to be exchangeable across data sets and to be single copy across actinopterygian fishes (Faircloth et al., 2013). One particular property of UCE loci that is, notable is that variation increases away from the core conserved region, providing variable sites across time scales and higher phylogenetic informativeness compared to protein-coding gene loci (Gilbert et al., 2015). The generous starting pool of 500 loci allows us to filter for paralogs or duplicates originating from the Ss4R while retaining relatively informative loci for phylogenetic inference. With UCE loci from representative salmonids, we test the following phylogenetic hypotheses while considering particularities of salmonid genome evolution: (1), the relationships of salmonid subfamilies, (2), the monophyly of the coregonine genus *Coregonus* and (3), the placement of the enigmatic salmonine genus *Parahucho*. To identify clear orthologs for analysis within salmonids from markers designed from diploid species, we filter loci based on assembled duplicates and on genomic location in known tetrasomic regions of salmonid genomes. We then test the effects of missing data and tetrasomic inheritance in phylogenetic analyses by generating alternative subsamples of the full dataset and compared the resulting inferred relationships with and without the multispecies coalescent.

## METHODS

### Sequence data acquisition and processing

We obtained sequence data from representatives of all major salmonid genera ( $n = 15$ ), as well as from representatives of the two extant esociform families Esocidae and Umbridae, which are the closest diploid relatives of salmonids (Ishiguro, Miya & Nishida, 2003; López, Chen & Ortí, 2004). We generated genetic sequence data for phylogenetic analysis by sequence capture of 500 UCE loci following the protocol outlined for the 500 UCE actinopterygian probe set (Faircloth et al., 2013) modified according to experimental conditions described in Campbell et al. (2017). We sequenced these loci for 11 taxa and incorporated data for six additional taxa from previous studies described in Table S1 (Campbell et al., 2017; Faircloth et al., 2013).

Demultiplexed reads were trimmed of low-quality bases and filtered for minimum length with Trimmomatic version 0.32 using the following command line specifications and adapter fasta file available for download with the program: ILLUMINACLIP:TruSeq3-PE-splitAdapter.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36 (Bolger, Lohse & Usadel, 2014). Sequences from a subset of samples were assembled with Velvet version 1.2.10 (Zerbino & Birney, 2008) to establish a range of suitable k-mer lengths for sequence assembly. We initially attempted optimal assemblies using various criteria with VelvetOptimiser version 2.2.5 (Gladman & Seemann, 2012). However, because some regions of the salmonid genome are tetrasomic, and thus do not meet the ploidy assumptions made in VelvetOptimiser, we established a range of k-mer lengths, 57–83 sites representing between 1/2 and 2/3 of total read length, and assembled salmonid sequences from all odd k-mer values within that range. We evaluated each of the resulting assemblies for the presence of UCEs using the PHYLUCE pipeline (Faircloth, 2015) and retained only the assembly that contained the maximum number of UCE loci, following the approach described in Campbell et al. (2017).

We identified homologous UCE loci and prepared them for alignment using the PHYLUCE pipeline. The PHYLUCE package screens out reciprocally duplicate enriched loci (i.e., potential paralogs) and performs a multiple sequence alignment of each of the remaining loci using MAFFT version 7.130b (Katoh et al., 2002; Katoh & Standley, 2013; Katoh & Toh, 2008). We screened tetrasomic loci in our dataset by searching previously published assemblies of Atlantic salmon (*Salmo salar*) (Lien et al., 2016) and rainbow trout (Pearse et al., 2019) for the UCEs sequenced in our study. Within Atlantic salmon, nine pairs of chromosome arms were indicated by Lien et al. (2016) that may be tetrasomic, seven of which were identified LORe regions by Robertson et al. (2017). From rainbow trout, seven chromosome pairs that were clearly tetrasomic in Pearse et al. (2019) were screened constituting homeologs to the seven LORe pairs screened in Atlantic salmon (Blumstein et al., 2019). BLASTN (Altschul et al., 1990) alignments of assembled UCEs that presented two alignments of >95% overlap and >97% similarity in the tetrasomic regions of either Atlantic salmon or rainbow trout were considered tetrasomic. UCEs that were not assembled and found to be single copy for all targeted species were also removed to create a “100% coverage alignment.”

To understand the effects of relaxing strict filtering of loci on phylogenetic inference, we generated four additional datasets: (1) a “95% coverage alignment” allowing one missing taxon per locus, (2) a “93% coverage alignment” allowing up to two missing taxa per locus, (3) a “75% coverage alignment” allowing up to five missing taxa per locus, (4) a “tetrasomic loci alignment” including only tetrasomically inherited loci permitting 5 missing taxa per locus.

### **Phylogenetic inference with the multispecies coalescent**

We conducted joint GT–ST estimation of each of the five datasets with the StarBEAST2 Bayesian framework (Ogilvie, Bouckaert & Drummond, 2017) as implemented in BEAST version 2.4.8 (Bouckaert et al., 2014). For analysis in StarBEAST2, our data sets were partitioned by each UCE locus and the Akaike Information Criterion (AIC) was applied to select the best-fitting model of nucleotide substitution for each locus in each of our datasets using ModelGenerator version 0.85 (Keane et al., 2006). We ran several iterations of BEAST with varying length and sampling frequency and evaluated convergence of the runs and effective sample sizes (ESS) with Tracer version 1.6 (Rambaut et al., 2018) to determine proper sampling frequency and MCMC chain run length. We required that ESS values be 200 or greater to assure an adequate sample size for our posterior parameter estimations. Analysis of our initial BEAST runs showed that the estimates of parameter values failed to converge for datasets with more than 13 UCEs even after 300 million to 2 billion MCMC generations and showed evidence of over-parameterization, such as a lack of convergence in each chain and low consistency in LnL/prior values. To account for this lack of convergence in the larger datasets and test the effects of using a simpler nucleotide substitution model, we conducted an additional set of phylogenetic inferences in BEAST of all of our datasets and partition schemes thereof but specified the HKY+ $\Gamma$  nucleotide substitution model with four categories of gamma distributed rate variation and empirical base frequencies for each partition. For each of our analyses, we combined the independent chains and applied a burnin with LogCombiner version 2.1.3, and generated a maximum clade credibility tree using TreeAnnotator version 2.1.2, specifying a posterior probability (PP) limit of 0.50 and median node heights.

In addition to the joint GT–ST estimation conducted in BEAST, we also estimated the ML tree of each UCE locus and used summary coalescence analyses to estimate a species tree from the resulting set of gene trees a posteriori. We conducted this analysis in addition to the joint GT and ST analysis conducted in StarBEAST2 because this approach can show high accuracy and is computationally more tractable than a full probabilistic approach (Mallo & Posada, 2016). For each of our datasets, we estimated the ML tree of each UCE locus using RAxML version 8.0.19, specifying the GTR+ $\Gamma$  model of sequence evolution. We passed these trees to ASTRAL version 5.5.9 and MP-EST version 2.2.0, which each estimated a species tree from the collection of UCE loci.

### **Concatenated analyses**

A concatenated maximum likelihood (ML) analysis was conducted with each data set (100%, 95%, 93%, 75% coverage alignments and the tetrasomic loci alignment).



We conducted both partitioned (i.e., by gene, or objectively defined) and un-partitioned analyses under the GTR+ $\Gamma$  model of sequence evolution using RAxML version 8.2.10 (Stamatakis, 2014) with rapid bootstrap stopping. We used the program PartitionFinder version 1.1.1 (Lanfear et al., 2012) to generate an objective optimal partitioning strategy with partitions of individual UCE loci specified with the following options: the GTR+ $\Gamma$  model, greedy algorithm (Lanfear et al., 2014) and likelihoods estimated by RAxML (-raxml option). If the objective partition strategy differed from either unpartitioned or partitioned strategies, a third analyses of that data set was undertaken with RAxML.

### Salmonid subfamily relationships

Though the composition and monophyly of salmonid subfamilies is well-established (Norden, 1961; Sanford, 1990), the relationships between them remain contentious. However, if we assume that each subfamily is truly monophyletic, and that their relationships take the form of a bifurcating tree, then there are only three possible arrangements of the salmonid subfamilies. This tractable number of potential combinations allows us to consider each possible arrangement of the taxa explicitly and test for the most preferred arrangement given our data. To avoid test-specific biases, we evaluated the support for each possible relationship of the subfamilies using four different approaches: (1) triplet analysis of independently estimated gene trees, (2), Bayes factors computed from the posterior trees of StarBEAST2 analyses, (3), a test under the multi-species coalescent with the program MP-EST, and, (4), approximately unbiased (AU) tests of concatenated trees.

#### Triplet analysis

A triplet analysis is based around a three-taxon rooted tree. In this manuscript, the three salmonid subfamilies form the basis of the test. Expectations with triplet analyses are that given the three possible relationships present with a rooted three-taxon case, the correct relationship should predominate over roughly equal portions of the alternative hypotheses (Campbell, Chen & López, 2014; Cranston, 2010; Pamilo & Nei, 1988). We performed triplet analysis using only the most stringent (i.e., no tetrasomic loci, no missing taxa) dataset, the 100% coverage alignment. For each UCE locus, a salmonine, coregonine, and thymalline were chosen at random with *Esox lucius* as a fixed outgroup. We inferred a ML for each set of taxa with RAxML and the tree stored. We repeated the process of random draws of three taxon sets 1,000 times for each locus and calculated the final frequency of relationships among the three salmonid subfamilies.

#### Bayes factors

We compared the strength of evidence for different salmonid subfamily relationships using Bayes factors calculated from the StarBEAST2 post-burnin trees (Jeffreys, 1935; Kass & Raftery, 1995). We assumed a uniform prior for these calculations, that is, each of the three subfamily relationships were equally likely. A Bayes factor is then the ratio of the posterior probabilities between two alternatives, and here, we present ratios of posterior probabilities of C+T/C+S, C+T/S+T and C+S/S+T.

**Table 1 Characteristics of UCE alignments.** For the five alignments generated for the study, the amount of missing data allowed in terms of taxa/locus is given, the number of resulting UCEs generated, and the number that were considered tetrasomic are presented. The resulting alignment length, unique alignment patterns and percentage of gaps and undetermined portions are then provided.

Alignment name	Missing data threshold	Number of UCEs	Number of tetrasomic loci	Number of loci analyzed	Alignment length	Unique alignment patterns	Gaps and undetermined (%)
100% Coverage	No missing taxa/locus	8	2	6	2,331	645	15.63
95% Coverage	1 taxon/locus	19	6	13	5,297	1,261	21.18
93% Coverage	2 taxa/locus	42	13	28	10,232	2,382	21.89
75% Coverage	5 taxa/locus	138	33	105	36,835	8,688	33.83
Tetrasomic loci	5 taxa/locus	138	33	33	12,758	2,661	33.13

### **Tests under multispecies coalescent**

We used a likelihood ratio test, incorporating the multispecies coalescent with MP-EST version 2.2.0 (Liu, Yu & Edwards, 2010) and the R package phybase version 2.0 (Liu & Yu, 2010) available at <https://github.com/lliu1871/phybase>. We supplied unconstrained gene trees from our summary coalescence analyses to MP-EST and constrained the species tree topologies but allowed the branch lengths to be optimized to best fit the data presented by the gene trees. We modified the test2.sptree function of phybase (provided in Data Supplement, <https://doi.org/10.25338/B8DC81>) to enable all three subfamily arrangements to be compared as described in Liu et al. (2019). Likelihood ratio test statistics ( $t$ ) were calculated between pairs of alternative trees and we generated a distribution of the test statistic by bootstrap sampling gene trees 100 times to estimate the null distribution for  $t$  and provide a measure of statistical significance ( $p$ -value).

### **Approximately unbiased tests**

We conducted topological tests of the three possible relationships of salmonid subfamilies in a concatenated framework by constraining subfamily relationships for different alignments and partitioning schemes and passing per site likelihoods from RAxML version 8.2.11 to the program CONSEL version 0.20. We used CONSEL to compute the AU test from these per site likelihoods (Shimodaira & Hasegawa, 2001).

## **RESULTS**

### **Sequence alignment characteristics**

Of the 500 UCE loci that we targeted, the number of total captured loci and the number of putatively single copy loci sequenced from each taxon varied (Table S1). For diploid outgroups, *Esox* and *Umbra*, 415 and 409 of the captured UCEs, respectively, were determined to be single-copy. Within salmonids, 166–275 of the captured loci were indicated to be single copy. Overall, we retained only eight single-copy UCE loci across all sampled taxa. Two of these loci were identified as coming from tetrasomically-inherited regions of the salmonid genome, and so were removed. Thus only six out of 500 candidate UCE loci passed the most stringent screening process and were present in all targeted taxa for a total of 2,331 aligned nucleotide sites (Table 1). The datasets that we generated

when allowing for missing data yielded more retained UCE loci, ranging between 13 and 105, with aligned lengths between 5,297 and 36,835 aligned nucleotide sites (Table 1). We identified 33 tetrasomically inherited loci for analysis in the “tetrasomic loci alignment” (see Table 1) and this dataset contains a total of 12,758 aligned nucleotide sites. Across all datasets, the proportion of gaps and other ambiguously alignment regions ranged from 15.63% to 33.83%.

### Model selection and MCMC convergence

For each locus in the “100% coverage” alignment (see Table 1), we initially specified the model of nucleotide substitution that received the lowest AIC score from ModelGenerator (Table S2). We obtained sufficient ESS and convergence of the parameter estimates with three independent MCMC chains of 400 million generations, sampled every 40,000 generations, with a 10% burn-in. For the “95% coverage” dataset, we specified nucleotide substitution models for each locus which received the lowest AIC score in our analysis using ModelGenerator. We obtained sufficient ESS and convergence with two chains of 50 million generations, sampled every 5,000 generations, with a 10% burn-in. For all remaining datasets, analyses with models of nucleotide substitution specified by ModelGenerator failed to generate sufficient ESS and convergence of parameter estimates in our MCMC chains. In response, we broadly reduced parameterization by simplifying nucleotide substitution models.

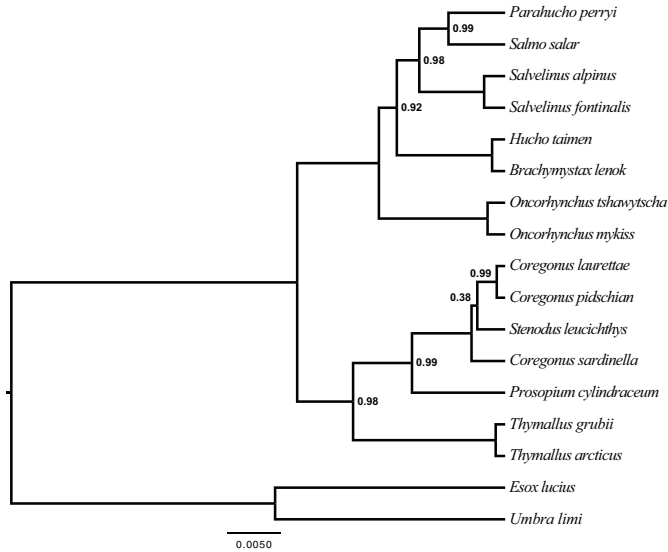
Analyses of the “93% coverage” alignment, the “75% coverage”, and “tetrasomic loci” alignments generated sufficient ESS and convergence of the parameter estimates with two independent MCMC chains of 200 million generations sampled every 20,000 generations, three chains of 800 million generations sampled every 80,000 generations, and four chains of 200 million generations sampled every 20,000 generations, respectively, and in each case we applied a 10% burn-in.

### Phylogenetic inference

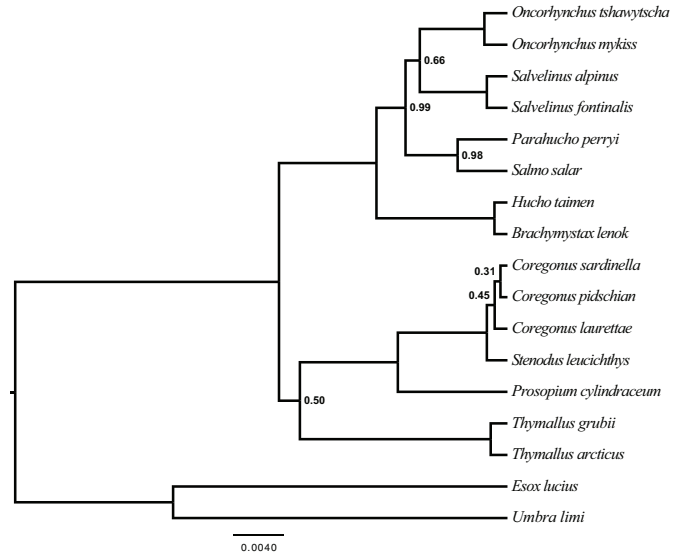
The results of joint GT-ST analyses are shown in Figs. 4 and 5 and summarized in Table 2. The clade Coregoninae + Thymallinae (C + T) is found in our 100% coverage alignment with ModelGenerator-specified nucleotide substitution models with strong support (0.98 PP) and in all analyses with reduced coverage datasets with very weak support (0.34–0.52 PP; see Table 2). Analysis of the tetrasomic loci alignment strongly supports Coregoninae + Salmoninae (C + S, 0.99 PP). The monophyly of *Coregonus* is not supported in the results of the 100% coverage species-tree, but is weakly supported (0.43–0.45 PP) in two of three of our reduced coverage analyses and strongly supported in our tetrasomic loci species-tree (1.00 PP). *Salmo* is inferred as the sister group of *Parahucho* in the 100% coverage species-tree with strong support (0.99 PP) and in all the reduced coverage analyses (0.98–0.99 PP). In contrast, the species tree inferred from the tetrasomic loci alignment places *Oncorhynchus* and *Parahucho* as sister lineages (0.38 PP).

The results of the summary coalescent analyses conducted in ASTRAL support a sister group relationship between Salmoninae and Thymallinae (4/5 analyses), but with low

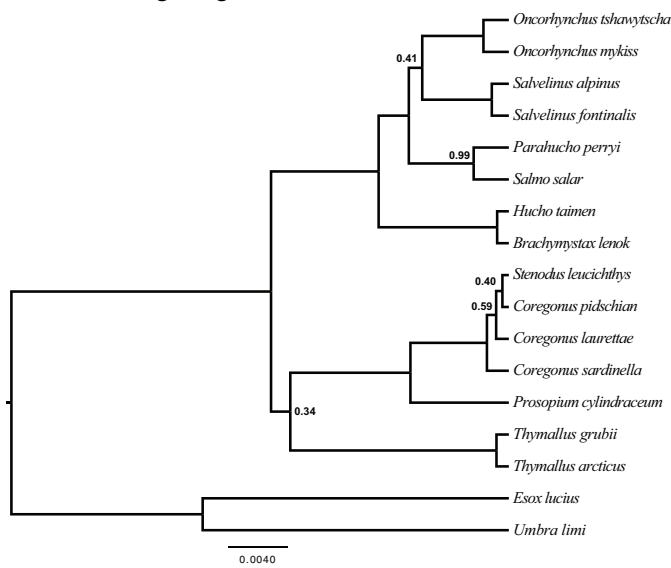
## A. 100% Coverage Alignment



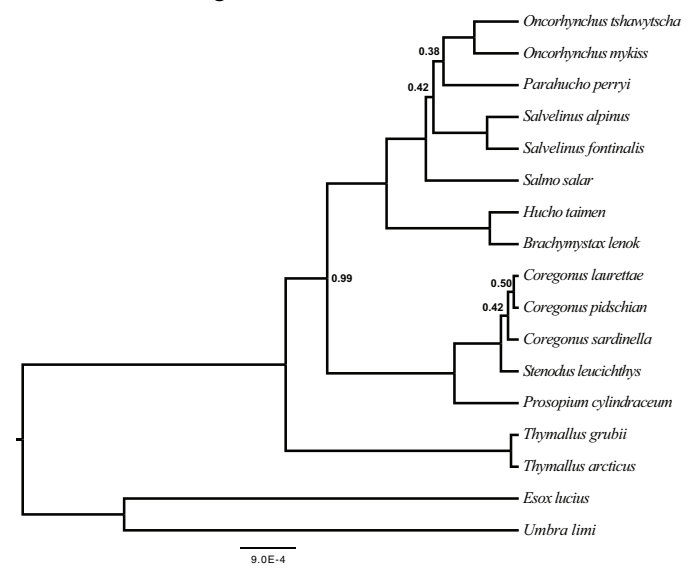
## B. 95% Coverage Alignment



## C. 93% Coverage Alignment

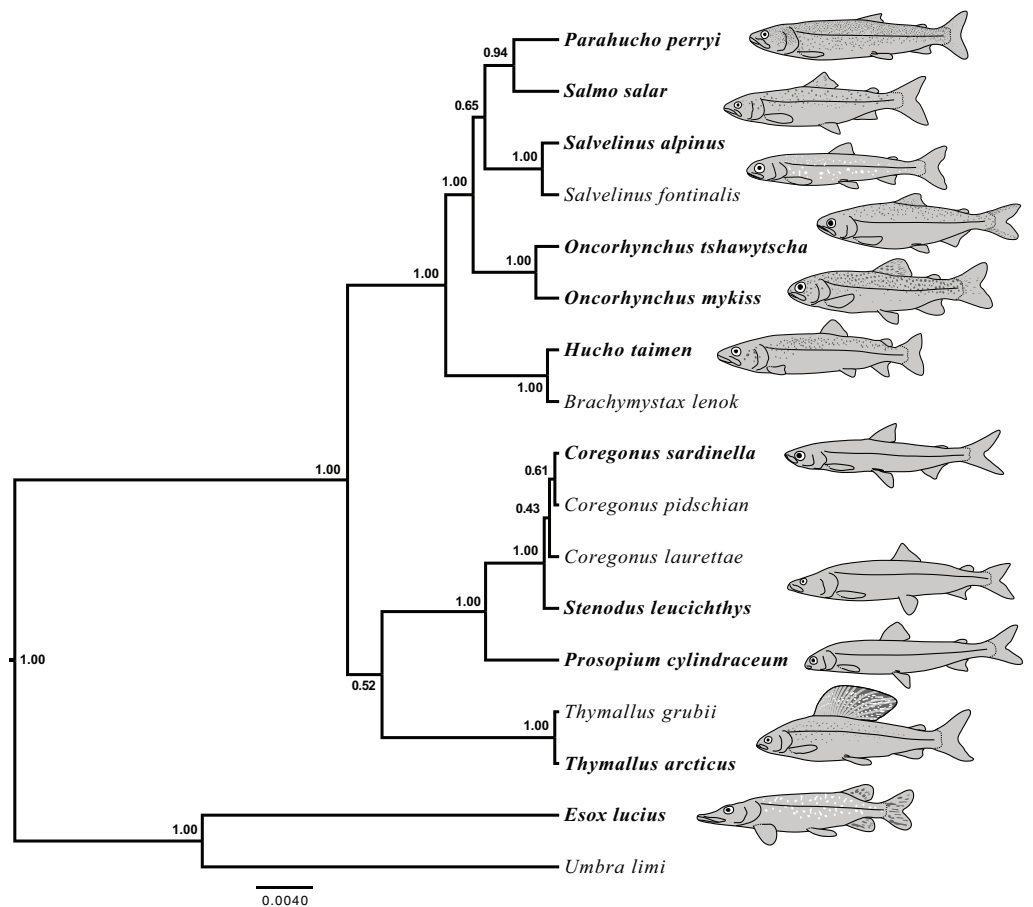


## D. Tetrasomic Loci Alignment



**Figure 4** Species trees generated in StarBEAST2. A joint estimation of species trees and gene trees is presented with posterior probability presented at nodes for four separate alignments. Posterior probabilities equal to 1.00 are not shown. (A) A total of 100% coverage alignment. (B) A total of 95% coverage alignment. (C) A total of 93% coverage alignment. (D) Tetrasomic loci alignment. [Full-size !\[\]\(bd23efe2a48a7d49c4d407ec14eec8e3\_img.jpg\) DOI: 10.7717/peerj.9389/fig-4](https://doi.org/10.7717/peerj.9389/fig-4)

support (PP ranges from 0.38 to 0.57), and, from tetrasomic loci, Coregoninae + Salmoninae (0.91 PP, see [Table 2](#)). Monophyly of *Coregonus* is supported in the results of 3/5 of the analyses, though with only weak support (PP ranges from 0.68 to 0.74). Within Salmoninae, *Parahucho* is variously supported as sister to several different taxa with weak support (BPP < 0.70), and to a *Oncorhynchus* + *Salvelinus* clade with moderate support from tetrasomic loci (0.85 BPP, see [Table 2](#)). The results of the summary coalescent analyses conducted in MP-EST are inconsistent and reported in [Table 2](#).



**Figure 5** Species tree generated in StarBEAST2 from 75% coverage alignment. A joint estimation of species trees and gene trees is presented with posterior probability presented at nodes. Depictions of species in the phylogeny are presented with bold species names indicating the represented species. *Hucho taimen* is represented by a drawing of *Hucho hucho*. Drawings by Thaddeus Buser.

Full-size DOI: 10.7717/peerj.9389/fig-5

## Concatenated analyses

Concatenated ML analyses supported either Coregoninae + Salmoninae or Salmoninae + Thymallinae relationships depending on the number of loci and partitioning strategy (Table S3, trees provided in Data Supplement). Partitioning strategy of alignments only altered inferred subfamily relationships in one case (95% coverage alignment), but caused substantial changes in bootstrap support (BS) values. Monophyly of *Coregonus* was only supported in three of thirteen analyses (BS range 70–100%), and the sister of *Parahucho* varied from *Salmo* (5/13 analyses, BS range 24–100%), to *Oncorhynchus* (5/13 analyses, BS range 50–99%) to *Salvelinus* (3/13 analyses, BS range 54–67%).

## Salmonid subfamily relationships

A majority (50.79%) of trees generated in the triplet analyses support a Coregoninae + Thymallinae clade. The Salmoninae + Coregoninae (31.31%) and Salmoninae + Thymallinae (17.90%) clades are found in smaller but substantial proportions of the

**Table 2 Outcomes of phylogenetic analyses incorporating the multispecies coalescent.** For the three methods employing the multispecies coalescent, (A) StarBEAST2, (B) ASTRAL, and (C) MP-EST, summaries of the results are presented for each of the five alignments. The three key relationships examined in the study, subfamily relationships, monophyly of *Coregonus*, and the sister lineage of *Parahucho* are presented along with support values if available.

Alignment	Number of independently modeled loci	Sister subfamilies	Sister subfamilies posterior probability	Monophyly of <i>Coregonus</i>	Support for monophyly of <i>Coregonus</i>	Sister of <i>Parahucho</i>	Support for sister of <i>Parahucho</i>
(A) StarBEAST2 analyses							
100% Coverage	6	Coregoninae + Thymallinae	0.98	No	–	<i>Salmo</i>	0.99
95% Coverage	13	Coregoninae + Thymallinae	0.50	Yes	0.45	<i>Salmo</i>	0.98
93% Coverage	28	Coregoninae + Thymallinae	0.34	No	–	<i>Salmo</i>	0.99
75% Coverage	105	Coregoninae + Thymallinae	0.52	Yes	0.43	<i>Salmo</i>	0.98
Tetrasomic loci	33	Coregoninae + Salmoninae	0.99	Yes	1.00	<i>Oncorhynchus</i>	0.38
(B) ASTRAL analyses							
100% Coverage	6	Salmoninae + Thymallinae	0.42	Yes	0.74	( <i>Salmo</i> + ( <i>Oncorhynchus</i> + <i>Salvelinus</i> ))	0.69
95% Coverage	13	Salmoninae + Thymallinae	0.38	Yes	0.68	( <i>Salmo</i> + ( <i>Oncorhynchus</i> + <i>Salvelinus</i> ))	0.68
93% Coverage	28	Salmoninae + Thymallinae	0.57	No	–	<i>Salmo</i>	0.46
75% Coverage	105	Salmoninae + Thymallinae	0.50	No	–	( <i>Oncorhynchus</i> + <i>Salvelinus</i> )	0.54
Tetrasomic loci	33	Coregoninae + Salmoninae	0.91	Yes	0.53	( <i>Oncorhynchus</i> + <i>Salvelinus</i> )	0.85
(C) MP-EST analyses							
100% Coverage	6	Coregoninae + Salmoninae		Yes		<i>Oncorhynchus</i>	
95% Coverage	13	Salmoninae + Thymallinae		Yes		( <i>Salmo</i> + <i>Oncorhynchus</i> )	
93% Coverage	28	Salmoninae + Thymallinae		No		<i>Salmo</i>	
75% Coverage	105	Coregoninae + Thymallinae		No		<i>Oncorhynchus</i>	
Tetrasomic loci	33	Coregoninae + Salmoninae		No		( <i>Oncorhynchus</i> + <i>Salvelinus</i> )	

remaining trees. Bayes factors from the 100% coverage alignment indicate that a sister group relationship of Coregoninae and Thymallinae should be strongly preferred. Bayes factors for all other datasets when comparing C + T to alternatives do not provide strong

**Table 3 Hypothesis testing with the multispecies coalescent.** The testing of subfamily relationships either through (A) Bayes factor (StarBEAST2), or, (B) the *t*-statistic (MP-EST, described in text) are shown. Partitioning and modeling of each of the five alignments is summarized along with the number of loci analyzed. Significance values are presented along with the *t*-statistic and are in bold face if significant.

(A)	StarBEAST2 Analyses	Partitioning and modeling scheme	Number of independently modeled loci	Bayes factor C+T vs C+S	Bayes factor C+T vs S+T	Bayes factor C+S vs S+T			
	100% Coverage alignment	By UCE, ModelGenerator	6	131.47	108.84	0.82			
	95% Coverage alignment	By UCE, ModelGenerator	13	1.82	2.21	1.21			
	93% Coverage alignment	By UCE, HKY+Γ	28	1.00	1.01	1.00			
	75% Coverage alignment	By UCE, HKY+Γ	105	1.23	9.33	7.57			
	Tetrasomic loci alignment	By UCE, HKY+Γ	33	0.00	0.81	248.20			
(B)	MP-EST analyses	Partitioning and modeling scheme	Number of independently modeled loci	<i>t</i> -statistic CT vs CS	<i>p</i> -value	<i>t</i> -statistic CT vs ST	<i>p</i> -value	<i>t</i> -statistic CS vs ST	<i>p</i> -value
	100% Coverage alignment	By UCE	6	14.80	0.27	11.69	0.37	3.10	0.28
	95% Coverage alignment	By UCE	13	5.60 × 10 <sup>-5</sup>	0.49	0.61	0.47	0.61	0.41
	93% Coverage alignment	By UCE	28	4.74	0.39	2.26	0.27	7.01	0.27
	75% Coverage alignment	By UCE	105	11.25	0.28	10.58	0.18	0.67	0.36
	Tetrasomic loci alignment	By UCE	33	130.18	0.06	8.10 × 10 <sup>-4</sup>	0.45	130.18	<b>0.05</b>

support for C + T (Table 3). MP-EST analysis indicated that a sister group relationship for C + S was significantly supported via likelihood ratio tests for the tetrasomic loci; no other data sets could distinguish among alternative topologies. The concatenated AU test results were not significant with all 100% and 95% coverage alignment tests, indicating that the alternative C + T hypothesis could not be rejected in analyses of the reduced data sets (Table S3).

## DISCUSSION

### Salmonid subfamily relationships

A sister group relationship of Coregoninae and Thymallinae (C + T) is broadly preferred to alternative topologies in our analyses (e.g., Fig. 5). This relationship is recovered in the joint GT–ST estimations of all data sets that exclude tetrasomically-inherited loci (PP = 0.34–0.98) and is also the preferred hypothesis indicated by Bayes factors and the triplet analysis. While no concatenated results support C + T, this result is obtained with as few as six UCE loci in joint GT–ST estimation with declining support as the number of loci increased. A Coregoninae + Thymallinae clade has been identified in previous molecular studies that either implicitly or explicitly account for potential polyploidy in their data. The C + T clade was recovered by analyses of mitochondrial genomes, which are not affected by issues of paralogy and can be expected to be less prone to ILS due to their haploid mode of transmission (Campbell *et al.*, 2013; Horreo, 2017). Likewise, constructing phylogenies with both copies of Ss4R duplicated genes, an explicit approach, leads to support of a sister Coregoninae + Thymallinae (Macqueen & Johnston, 2014; Robertson *et al.*, 2017).

Our results do not support the conclusions of previous studies that examined morphological characters of the salmonids and unanimously placed Thymallinae as the sister group to Salmoninae (*Kendall & Behnke, 1984; Sanford, 1990; Stearley & Smith, 1993; Wilson & Li, 1999*). Furthermore, our results combined with mitogenomic and other studies explicitly addressing polyploidy strongly support C + T over the two alternative hypotheses apparent in molecular phylogenetic studies of salmonids (*Alexandrou et al., 2013; Crête-Lafrenière, Weir & Bernatchez, 2012; Shedko, Miroshnichenko & Nemkova, 2012*). In light of growing molecular evidence conflicting with the morphology-based hypothesis, a critical reassessment of the morphological evidence that supports that relationship is needed but is beyond the scope of the present study.

Though C + T is the preferred grouping in our most stringent analyses, not all of the inference strategies implemented here supported this relationship. Analysis of the concatenated dataset and GT-ST estimation using ASTRAL produced inconsistent results and generally did not support the C + T clade. The inconsistent results between summary coalescence approach and joint estimation of the species tree and gene trees may indicate loss of a weak phylogenetic signal in the summarization step. Alternatively, the phylogenetic signal may have been blurred as a result of treating all UCE loci as a single partition evolving under a common substitution mode. The triplet analysis did not meet our prediction that, given three possible topologies, the true relationship should predominate and the alternative, incorrect arrangements should receive approximately equal portions of support. The triplet analyses yielded the putatively incorrect topologies at unequal frequencies (31.30% and 17.90%). This could be a result of the relatively small size of our “100% coverage” alignment or be a product of resampling selecting certain individuals frequently due to few representatives in some subfamilies (two thymalline species are in the dataset).

### Monophyly of *Coregonus*

Our results inconsistently support the monophyly of *Coregonus* (e.g., *Figs. 4 and 5*). Though the joint GT-ST analysis of tetrasomic loci and the ML analysis of the concatenated 100% coverage data set shows strong support for a monophyletic *Coregonus*, all other analyses that recovered the clade did so with weak support (PP < 0.75, BS = 71%, see *Table 2*). Even within a given analytical framework, support for monophyly of this genus varied. For example, MP-EST analyses showed a monophyletic *Coregonus* for the two most stringent datasets, but not in the analysis of any others. The uncertainty in this relationship within our data set may be derived from the relatively young age (i.e., ~10 mya, *Horreo, 2017*) of the clade that includes *Stenodus* and *Coregonus* resulting in low variation in the conserved UCE loci that were identified to be single copy across a broad range of salmonid species. Additionally, during the analytical study design of this manuscript, a chromosomal-level assembly of a coregonine was unavailable leading to uncertainty in the tetrasomic status of loci in this subfamily based off salmonine genome assemblies.

Previous studies examining the monophyly of *Coregonus* with respect to *Stenodus* have yielded conflicting results (*Bernatchez, Colombani & Dodson, 1991; Bodaly et al., 1991;*



Crête-Lafrenière, Weir & Bernatchez, 2012; Horreo, 2017; Sajdak & Phillips, 1997; Vuorinen et al., 1998). Most recently, analysis of mitogenomic data has shown *Coregonus* to be monophyletic, with *Stenodus* as its sister taxon (Horreo, 2017). However, the phylogenetic placement of the enigmatic *Coregonus huntsmani*, a divergent member of the genus, remains untested in any mitogenomic or phylogenomic study.

Crête-Lafrenière, Weir & Bernatchez (2012) found *C. huntsmani* to be the sister lineage to *Stenodus* and all other *Coregonus* species and earlier isozyme and mitochondrial restriction data show *C. huntsmani* to be distinct from the other main *Coregonus* subgenera (*Coregonus* and *Leucichthys*) but did not include *Stenodus* in the sampling (Bernatchez et al., 1991). We examined publicly available COI data from coregonines including two haplotypes from *C. huntsmani* and found that species to be sister group to a monophyletic *Coregonus* with low support (BS = 58%, alignment tree and methods are supplied in the [Data Supplement](#)). At present molecular data from *C. huntsmani* are very limited, however including this taxon in future datasets will be critical for conclusively testing the monophyly of *Coregonus* and its relationship with *Stenodus*. Further investigations of phylogenetic relationships of *Coregonus* and *Stenodus* should also consider the role and prevalence of hybridization within *Coregonus* and between *Stenodus* and *Coregonus*, (McPhail, 2007) as the signatures of hybridization and ILS are difficult to disentangle and our analyses assumed such signatures derived from ILS (Yu et al., 2011).

### Placement of *Parahucho*

The Sakhalin taimen *Parahucho perryi* is a rare species and the sole member of its genus (Rand, 2006; Vladykov & Gruchy, 1972). This critically endangered fish occupies a limited geographic range and has a narrow range of suitable habitats in Northern Japan and the Russian Far East (Fukushima et al., 2011; Kimura, 1966; Rand, 2006). The placement of *Parahucho* as sister to *Salmo* is strongly supported by all joint GT–ST analyses of non-tetrasomic loci (4 of 4, 0.98–0.99 PP) and by concatenated analysis of our 100% and 95% coverage alignment (BS = 66–100%). The original description of Sakhalin taimen placed it as a congener of Atlantic salmon (*Salmo salar*), as *Salmo perryi* (Brevoort, 1856). Subsequently, the similarities of Sakhalin taimen to the European huchen *Hucho hucho* were noted, and these two species were placed in the same genus, *Hucho* (Günther, 1868; Jordan & Snyder, 1902). *Hucho* and *Parahucho* are both large bodied piscivores, but differ in that the latter has 10 less vertebrae, basibranchial teeth, a median set of teeth on the supralingual, and 70–100 less scales in the lateral line (Vladykov, 1963; Vladykov & Gruchy, 1972). Indeed, the two genera are not closely allied in any of the results presented here, nor has a close relationship of the two genera been reported in any other molecular phylogenetic studies (Campbell et al., 2013; Crespi & Fulton, 2004; Matveev, Nishihara & Okada, 2007; Oakley & Phillips, 1999). Though there has been considerable disagreement in the placement of *Parahucho*, the emerging consensus from nuclear gene sequences supports a *Parahucho* + *Salmo* sister relationship (Lecaudey et al., 2018), as is recovered here.

## Phylogenetic results from disomic and tetrasomic loci

Resolving subfamily relationships was a primary aim of this study, but our results repeatedly showed a conflict in this segment of the phylogeny between analyses of the most stringently assembled dataset (i.e., 100% coverage) and of the remaining alignments, and particularly of the tetrasomic loci. For example, in GT–ST analyses conducted with StarBEAST2 of the 100%, 95%, 93% and 75% coverage alignments, Coregoninae and Thymallinae are supported as sister lineages. Analyses of the 100% coverage data set yield very strong (0.98 PP) support for that grouping, which declines substantially in analyses of the other datasets (0.34–0.52 PP). On the other hand, the 33 tetrasomic loci we removed strongly supported a Coregoninae + Salmoninae clade (0.99 PP). Likewise, we observed strong support for a sister relationship between *Salmo* and *Parahucho* in the more reduced datasets (0.98–1.00 PP), and less support in the more extensive 75% alignment (0.94 PP). The topology of the species tree inferred from StarBEAST2 analysis of tetrasomic loci indicates that *Salmo* and *Parahucho* are not closely related (Fig. 4D). In summary, two of three sets of alternative placements we considered in detail (i.e., subfamily relationships, *Coregonus* monophyly, and placement of *Parahucho*) in this study yield conflicting results when comparing datasets of loci with differing ploidy levels. We also find that support values for the C + T clade decline with greater proportions of missing data in the underlying dataset.

We attempted to identify tetrasomic loci in the sequenced dataset through the assembly of a single contig matching a UCE locus that was placed within genomic locations of known tetrasomy in rainbow trout and Atlantic salmon. While there is evidence for broad conservation of homologous tetrasomic genomic regions in salmonids (Brieuc et al., 2014; Kodama et al., 2014; Lien et al., 2011; Sutherland et al., 2016), there is also evidence of substantial lineage-specific rediploidization in salmonids and polyploids in general taxa (Robertson et al., 2017; Spoelhof, Soltis & Soltis, 2017). Our definition of tetrasomic regions from two salmonines is taxonomically restrictive and may have allowed the inclusion of regions of tetrasomic inheritance in other species sequenced. Consequently, increasing the proportion of missing data may have introduced paralogous loci into our analyses due to lineage-specific rediploidization processes (e.g., Fig. 3B).

## CONCLUSION

In salmonids, ancestral autopolyploidy and the resulting residual tetrasomy increases the potential for conflicts between gene trees and species trees due to a greater likelihood for incomplete lineage sorting. Lineage-specific rediploidization processes also are known to be concentrated in the same tetrasomically pairing chromosome regions contributing to the analysis of paralogous loci. Applying stringent criteria aimed at minimizing these problems yielded a highly reduced dataset comprising six UCE loci. These loci show clear support for a Coregoninae + Thymallinae clade and for the placement of *Parahucho* sister to *Salmo*, but do not support the monophyly of *Coregonus* in joint GT and ST analysis. Expanding the number of aligned loci by lowering stringency of filtering resulted in generally reduced confidence in clade makeup and support for conflicting phylogenetic

relationships while increasing support for the monophyly of *Coregonus*. The use of tetrasomic loci likewise resulted in an alternative topology from that of our 100% coverage analysis, and contradictory relationships may exhibit high support values. The accurate inference of salmonid phylogenies is challenged not only by limits of data and the sophistication of analyses, but also in part due to the evolutionary history and patterns of the lineage.

## ACKNOWLEDGEMENTS

We would like to thank Scott Edwards and Lang Liu for providing code and a manuscript in advance of publication along with helpful advice for hypothesis testing under the multispecies coalescent. We would also like to thank Max Belasco for his assistance generating the sequence data used in this study. Michio Fukushima kindly allowed us to use a picture of *Parahucho perryi*, Armando Piccinini and Stefano Porcellotti who allowed us to use a picture of *Hucho hucho*, and Beatrice Smith who allowed us to use a picture of *Stenodus leucichthys* to create a depiction of each of these species for our Fig. 5. We would like to recognize Kevin McCracken (University of Miami) for his support in promoting Next-Generation Sequencing and Bioinformatics and the University of Alaska Fairbanks from which this data set was generated. Tissues in this study were provided by the University of Alaska Museum of the North, the Burke Museum (University of Washington) and the Academy of Natural Sciences of Drexel University.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

This study was funded by a Genomics Seed Grant from the Alaska Idea Network for Biomedical Research (INBRE) awarded to Matthew A. Campbell and J. Andrés López. Research reported in this publication was supported by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under grant number P20GM103395. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Grant Disclosures

The following grant information was disclosed by the authors:  
Genomics Seed Grant from the Alaska Idea Network for Biomedical Research (INBRE).  
Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health: P20GM103395.

### Competing Interests

The authors declare that they have no competing interests.

### Author Contributions

- Matthew A. Campbell conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.

- Thaddaeus J. Buser analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Michael E. Alfaro conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- J. Andrés López conceived and designed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.

### DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

The sequence data are available at GenBank under BioProject [PRJNA603774](#) as a Targeted Locus Study: KDUN00000000, KDUO00000000, KDUP00000000, KDUQ00000000, KDUR00000000, KDUS00000000, KDUT00000000, KDUU00000000, KDUV00000000, KDUW00000000, KDUX00000000, KDUY00000000, KDUZ00000000, KDVA00000000, KDVB00000000.

### Data Availability

The following information was supplied regarding data availability:

Assembled sequence data matching UCE loci are available both from NCBI: [PRJNA603774](#) and at Dryad: Campbell, Matthew; Buser, Thaddaeus; Alfaro, Michael; Lopez, J. Andres (2020), Data from: Addressing incomplete lineage sorting and paralogy in the inference of uncertain salmonid phylogenetic relationships, v3, UC Davis, Dataset, [DOI 10.25338/B8DC81](#).

Alignments, tree files, scripts used in the triplet analysis and hypothesis testing under the multispecies coalescent are available in the [Supplemental Files](#).

Accession numbers and originating tissue collection are available in [Table S4](#). Specimens are housed at ANSP, FMNH, UAMN or UWFC.

Phybase is available at GitHub: <https://github.com/lliu1871/phybase>. The altered function is available as a [Supplemental File](#).

### Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.9389#supplemental-information>.

## REFERENCES

- Alexandrou MA, Swartz BA, Matzke NJ, Oakley TH. 2013. Genome duplication and multiple evolutionary origins of complex migratory behavior in Salmonidae. *Molecular Phylogenetics and Evolution* **69**(3):514–523 DOI [10.1016/j.ympev.2013.07.026](#).
- Allendorf FW, Thorgaard GH. 1984. Tetraploidy and the evolution of salmonid fishes. In: Turner BJ, ed. *Evolutionary Genetics of Fishes, Monographs in Evolutionary Biology*. Boston: Springer, 1–53.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**(3):403–410 DOI [10.1016/S0022-2836\(05\)80360-2](#).

- Bernatchez L, Colombani F, Dodson JJ. 1991.** Phylogenetic relationships among the subfamily Coregoninae as revealed by mitochondrial DNA restriction analysis. *Journal of Fish Biology* 39:283–290 DOI [10.1111/j.1095-8649.1991.tb05091.x](https://doi.org/10.1111/j.1095-8649.1991.tb05091.x).
- Bernatchez L, Edge TA, Dodson JJ, Qadri SU. 1991.** Mitochondrial DNA and isozyme electrophoretic analyses of the endangered Acadian whitefish, *Coregonus huntsmani* Scott, 1987. *Canadian Journal of Zoology* 69(2):311–316 DOI [10.1139/z91-050](https://doi.org/10.1139/z91-050).
- Berthelot C, Brunet F, Chalopin D, Juanchich A, Bernard M, Noel B, Bento P, Da Silva C, Labadie K, Alberti A, Aury J-M, Louis A, Dehais P, Bardou P, Montfort J, Klopp C, Cabau C, Gaspin C, Thorgaard GH, Boussaha M, Quillet E, Guyomard R, Galiana D, Bobe J, Volff J-N, Genet C, Wincker P, Jaillon O, Roest Crollius H, Guiguen Y. 2014.** The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nature Communications* 5(1):3657 DOI [10.1038/ncomms4657](https://doi.org/10.1038/ncomms4657).
- Betancur-R R, Broughton RE, Wiley EO, Carpenter K, López JA, Li C, Holcroft NI, Arcila D, Sanciangco M, Ii JCC, Zhang F, Buser T, Campbell MA, Ballesteros JA, Roa-Varon A, Willis S, Borden WC, Rowley T, Reneau PC, Hough DJ, Lu G, Grande T, Arratia G, Ortí G. 2013.** The tree of life and a new classification of bony fishes. *PLOS Currents* 5:ecurrents.tol.53ba26640df0ccae75bb165c8c26288 DOI [10.1371/currents.tol.53ba26640df0ccae75bb165c8c26288](https://doi.org/10.1371/currents.tol.53ba26640df0ccae75bb165c8c26288).
- Betancur-R R, Wiley EO, Arratia G, Acero A, Bailly N, Miya M, Lecointre G, Ortí G. 2017.** Phylogenetic classification of bony fishes. *BMC Evolutionary Biology* 17(1):162 DOI [10.1186/s12862-017-0958-3](https://doi.org/10.1186/s12862-017-0958-3).
- Blumstein DM, Campbell MA, Hale MC, Sutherland BJB, McKinney GJ, Stott W, Larson WA. 2019.** Comparative genomic analyses and a novel linkage map for cisco (*Coregonus artedii*) provides insight into chromosomal evolution and rediploidization across salmonids. *bioRxiv* 834937 DOI [10.1101/834937](https://doi.org/10.1101/834937).
- Bodaly RA, Vuorinen J, Ward RD, Luczynski M, Reist JD. 1991.** Genetic comparisons of New and Old World coregonid fishes. *Journal of Fish Biology* 38(1):37–51 DOI [10.1111/j.1095-8649.1991.tb03089.x](https://doi.org/10.1111/j.1095-8649.1991.tb03089.x).
- Bolger AM, Lohse M, Usadel B. 2014.** Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120 DOI [10.1093/bioinformatics/btu170](https://doi.org/10.1093/bioinformatics/btu170).
- Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014.** BEAST 2: a software platform for Bayesian evolutionary analysis. *PLOS Computational Biology* 10(4):e1003537 DOI [10.1371/journal.pcbi.1003537](https://doi.org/10.1371/journal.pcbi.1003537).
- Brevoort JC. 1856.** Notes on some figures of Japanese fish taken from recent specimens by the artists of the U.S. Japan Expedition. In: Narrative of the Expedition of an American Squadron to the China Seas and Japan, Performed in the Years 1852, 1853, and 1854 under the Command of Commodore M.C. Perry, United States Navy, by Order of the Government of the United States, U.S. Senate Ex. Doc. No. 79, 33rd Congress, 2nd Session. Beverley Tucker, Washington, D.C., 253–288.
- Brieuc MSO, Waters CD, Seeb JE, Naish KA. 2014.** A dense linkage map for Chinook salmon (*Oncorhynchus tshawytscha*) reveals variable chromosomal divergence after an ancestral whole genome duplication event. *G3 GenesGenomesGenetics* 4(3):447–460 DOI [10.1534/g3.113.009316](https://doi.org/10.1534/g3.113.009316).
- Burridge CP, McDowall RM, Craw D, Wilson MV, Waters JM. 2012.** Marine dispersal as a pre-requisite for Gondwanan vicariance among elements of the galaxiid fish fauna. *Journal of Biogeography* 39(2):306–321 DOI [10.1111/j.1365-2699.2011.02600.x](https://doi.org/10.1111/j.1365-2699.2011.02600.x).

- Campbell MA, Alfaro ME, Belasco M, López JA. 2017.** Early-branching euteleost relationships: areas of congruence between concatenation and coalescent model inferences. *PeerJ* 5(Part A):e3548 DOI 10.7717/peerj.3548.
- Campbell MA, Chen W-J, López JA. 2014.** Molecular data do not provide unambiguous support for the monophyly of flatfishes (Pleuronectiformes): a reply to Betancur-R and Ortí. *Molecular Phylogenetics and Evolution* 75:149–153 DOI 10.1016/j.ympev.2014.02.011.
- Campbell MA, Ganley ARD, Gabaldón T, Cox MP. 2016.** The case of the missing ancient fungal polyploids. *American Naturalist* 188(6):602–614 DOI 10.1086/688763.
- Campbell MA, Hale MC, McKinney GJ, Nichols KM, Pearse DE. 2019.** Long-term conservation of ohnologs through partial tetrasomy following whole-genome duplication in Salmonidae. *G3 Genes Genomes Genetics* 9:g3.400070.2019 DOI 10.1534/g3.119.400070.
- Campbell MA, López JA, Sado T, Miya M. 2013.** Pike and salmon as sister taxa: detailed intraclade resolution and divergence time estimation of Esociformes + Salmoniformes based on whole mitochondrial genome sequences. *Gene* 530(1):57–65 DOI 10.1016/j.gene.2013.07.068.
- Campbell MA, Sado T, Shinzato C, Koyanagi R, Okamoto M, Miya M. 2018.** Multi-locus phylogenetic analysis of the first molecular data from the rare and monotypic Amarsipidae places the family within the Pelagia and highlights limitations of existing data sets in resolving pelagian interrelationships. *Molecular Phylogenetics and Evolution* 124(1):172–180 DOI 10.1016/j.ympev.2018.03.008.
- Cranston K. 2010.** Summarizing gene tree incongruence at multiple phylogenetic depths. In: Knowles LL, Kubatko LS, eds. *Estimating Species Trees: Practical and Theoretical Aspects*. Hoboken: John Wiley & Sons, Inc., 129–142.
- Crespi BJ, Fulton MJ. 2004.** Molecular systematics of Salmonidae: combined nuclear data yields a robust phylogeny. *Molecular Phylogenetics and Evolution* 31(2):658–679 DOI 10.1016/j.ympev.2003.08.012.
- Crête-Lafrenière A, Weir LK, Bernatchez L. 2012.** Framing the Salmonidae family phylogenetic portrait: a more complete picture from increased taxon sampling. *PLOS ONE* 7(10):e46662 DOI 10.1371/journal.pone.0046662.
- Degnan JH, Rosenberg NA. 2009.** Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution* 24(6):332–340 DOI 10.1016/j.tree.2009.01.009.
- Dehal P, Boore JL. 2005.** Two rounds of whole genome duplication in the ancestral vertebrate. *PLOS Biology* 3(10):e314 DOI 10.1371/journal.pbio.0030314.
- Faircloth BC. 2015.** PHYLUCES is a software package for the analysis of conserved genomic loci. *Bioinformatics* 32(5):786–788 DOI 10.1093/bioinformatics/btv646.
- Faircloth BC, Sorenson L, Santini F, Alfaro ME. 2013.** A phylogenomic perspective on the radiation of ray-finned fishes based upon targeted sequencing of Ultraconserved Elements (UCEs). *PLOS ONE* 8(6):e65923 DOI 10.1371/journal.pone.0065923.
- Fricke R, Eschmeyer W, Fong JD. 2019.** Catalog of fishes. Available at <http://researcharchive.calacademy.org/research/ichthyology/catalog/SpeciesByFamily.asp>.
- Friedman M, Feilich KL, Beckett HT, Alfaro ME, Faircloth BC, Černý D, Miya M, Near TJ, Harrington RC. 2019.** A phylogenomic framework for pelagiarian fishes (Acanthomorpha: Percomorpha) highlights mosaic radiation in the open ocean. *Proceedings of the Royal Society B: Biological Sciences* 286(1910):20191502 DOI 10.1098/rspb.2019.1502.
- Fukushima M, Shimazaki H, Rand PS, Kaeriyama M. 2011.** Reconstructing Sakhalin taimen *Parahucho perryi* historical distribution and identifying causes for local extinctions. *Transactions of the American Fisheries Society* 140:1–13 DOI 10.1080/00028487.2011.544999.

- Gerstein AC, Otto SP. 2009.** Ploidy and the causes of genomic evolution. *Journal of Heredity* **100**(5):571–581 DOI [10.1093/jhered/esp057](https://doi.org/10.1093/jhered/esp057).
- Gilbert PS, Chang J, Pan C, Sobel EM, Sinsheimer JS, Faircloth BC, Alfaro ME. 2015.** Genome-wide ultraconserved elements exhibit higher phylogenetic informativeness than traditional gene markers in percomorph fishes. *Molecular Phylogenetics and Evolution* **92**:140–146 DOI [10.1016/j.ympev.2015.05.027](https://doi.org/10.1016/j.ympev.2015.05.027).
- Gladman S, Seemann T. 2012.** VelvetOptimiser. Available at <https://github.com/Victorian-Bioinformatics-Consortium/VelvetOptimiser>.
- Günther ACLG. 1868.** *Catalogue of the Fishes in the British Museum, Catalogue of the Physostomi, containing the families Salmonidae, Percopsidae, Galaxidae, Mormyridae, Gymnarchidae, Esocidae, Umbridae, Scombresocidae, Cyprinodontidae, in the collection of the British Museum.* London: British Museum of Natural History.
- Horreo JL. 2017.** Revisiting the mitogenomic phylogeny of Salmoninae: new insights thanks to recent sequencing advances. *PeerJ* **5**:e3828 DOI [10.7717/peerj.3828](https://doi.org/10.7717/peerj.3828).
- Ishiguro NB, Miya M, Nishida M. 2003.** Basal euteleostean relationships: a mitogenomic perspective on the phylogenetic reality of the “Protacanthopterygii”. *Molecular Phylogenetics and Evolution* **27**(3):476–488 DOI [10.1016/S1055-7903\(02\)00418-9](https://doi.org/10.1016/S1055-7903(02)00418-9).
- Jeffreys H. 1935.** Some tests of significance, treated by the theory of probability. In: *Mathematical Proceedings of the Cambridge Philosophical Society*, Cambridge: Cambridge University Press, 203–222.
- Jordan DS, Snyder JO. 1902.** A review of the salmonoid fishes of Japan. *Proceedings of the United States National Museum* **24**(1265):567–593 DOI [10.5479/si.00963801.24-1265.567](https://doi.org/10.5479/si.00963801.24-1265.567).
- Kass RE, Raftery AE. 1995.** Bayes factors. *Journal of the American Statistical Association* **90**(430):773–795 DOI [10.1080/01621459.1995.10476572](https://doi.org/10.1080/01621459.1995.10476572).
- Katoh K, Misawa K, Kuma K, Miyata T. 2002.** MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**(14):3059–3066 DOI [10.1093/nar/gkf436](https://doi.org/10.1093/nar/gkf436).
- Katoh K, Standley DM. 2013.** MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**(4):772–780 DOI [10.1093/molbev/mst010](https://doi.org/10.1093/molbev/mst010).
- Katoh K, Toh H. 2008.** Recent developments in the MAFFT multiple sequence alignment program. *Briefings in Bioinformatics* **9**(4):286–298 DOI [10.1093/bib/bbn013](https://doi.org/10.1093/bib/bbn013).
- Keane TM, Creevey CJ, Pentony MM, Naughton TJ, McInerney JO. 2006.** Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evolutionary Biology* **6**(1):29 DOI [10.1186/1471-2148-6-29](https://doi.org/10.1186/1471-2148-6-29).
- Kendall AW Jr, Behnke RJ. 1984.** Salmonidae: development and relationships. In: Moser HG, Richards WJ, Cohen DM, Fahay MP, Kendall AW Jr, Richards SL, eds. *Ontogeny and Systematics of Fishes*. Lawrence: American Society of Ichthyologists and Herpetologists, 142–149.
- Kimura S. 1966.** On the life history of the salmonid fish, *Hucho perryi* (Brevoort), found in Nemuro. *Hokkaido Japan Journal of Ichthyol* **14**:17–25 2.
- Kodama M, Briec MSO, Devlin RH, Hard JJ, Naish KA. 2014.** Comparative mapping between coho Salmon (*Oncorhynchus kisutch*) and three other salmonids suggests a role for chromosomal rearrangements in the retention of duplicated regions following a whole genome duplication event. *G3 Genes Genomes Genetics* **4**(9):1717–1730 DOI [10.1534/g3.114.012294](https://doi.org/10.1534/g3.114.012294).

- Lanfear R, Calcott B, Ho SYW, Guindon S. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution* 29(6):1695–1701 DOI 10.1093/molbev/mss020.
- Lanfear R, Calcott B, Kainer D, Mayer C, Stamatakis A. 2014. Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Evolutionary Biology* 14(1):1–14 DOI 10.1186/1471-2148-14-82.
- Lecaudey LA, Schlieven UK, Osinov AG, Taylor EB, Bernatchez L, Weiss SJ. 2018. Inferring phylogenetic structure, hybridization and divergence times within Salmoninae (Teleostei: Salmonidae) using RAD-sequencing. *Molecular Phylogenetics and Evolution* 124:82–99 DOI 10.1016/j.ympev.2018.02.022.
- Leggatt RA, Iwama GK. 2003. Occurrence of polyploidy in the fishes. *Reviews in Fish Biology and Fisheries* 13(3):237–246 DOI 10.1023/B:RFBF.0000033049.00668.fe.
- Li J, Xia R, McDowall RM, López JA, Lei G, Fu C. 2010. Phylogenetic position of the enigmatic *Lepidogalaxias salamandroides* with comment on the orders of lower euteleostean fishes. *Molecular Phylogenetics and Evolution* 57(2):932–936 DOI 10.1016/j.ympev.2010.07.016.
- Lien S, Gidskehaug L, Moen T, Hayes BJ, Berg PR, Davidson WS, Omholt SW, Kent MP. 2011. A dense SNP-based linkage map for Atlantic salmon (*Salmo salar*) reveals extended chromosome homeologies and striking differences in sex-specific recombination patterns. *BMC Genomics* 12(1):615 DOI 10.1186/1471-2164-12-615.
- Lien S, Koop BF, Sandve SR, Miller JR, Kent MP, Nome T, Hvidsten TR, Leong JS, Minkley DR, Zimin A, Grammes F, Grove H, Gjuvslund A, Walenz B, Hermansen RA, von Schalburg K, Rondeau EB, Di Genova A, Samy JKA, Olav Vik J, Vigeland MD, Caler L, Grimholt U, Jentoft S, Inge Våge D, De Jong P, Moen T, Baranski M, Palti Y, Smith DR, Yorke JA, Nederbragt AJ, Tooming-Klunderud A, Jakobsen KS, Jiang X, Fan D, Hu Y, Liberles DA, Vidal R, Iturra P, Jones SJM, Jonassen I, Maass A, Omholt SW, Davidson WS. 2016. The Atlantic salmon genome provides insights into rediploidization. *Nature* 533:200–205 DOI 10.1038/nature17164.
- Liu L, Anderson C, Pearl D, Edwards SV. 2019. Modern phylogenomics: building phylogenetic trees using the multispecies coalescent model. In: Anisimova M, ed. *Evolutionary Genomics: Statistical and Computational Methods*. Berlin: Springer.
- Liu L, Yu L. 2010. Phybase: an R package for species tree analysis. *Bioinformatics* 26(7):962–963 DOI 10.1093/bioinformatics/btq062.
- Liu L, Yu L, Edwards SV. 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC Evolutionary Biology* 10(1):302 DOI 10.1186/1471-2148-10-302.
- López JA, Chen WJ, Ortí G. 2004. Esociform phylogeny. *Copeia* 2004(3):449–464 DOI 10.1643/CG-03-087R1.
- Macqueen DJ, Johnston IA. 2014. A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. *Proceedings of the Royal Society B: Biological Sciences* 281(1778):20132881 DOI 10.1098/rspb.2013.2881.
- Mallet J. 2005. Hybridization as an invasion of the genome. *Trends in Ecology & Evolution* 20(5):229–237 DOI 10.1016/j.tree.2005.02.010.
- Mallo D, Posada D. 2016. Multilocus inference of species trees and DNA barcoding. *Philosophical Transactions of the Royal Society B: Biological Sciences* 371(1702):20150335 DOI 10.1098/rstb.2015.0335.
- Matveev V, Nishihara H, Okada N. 2007. Novel SINE families from salmons validate Parahucho (Salmonidae) as a distinct genus and give evidence that SINES can incorporate LINE-related



- 3'-tails of other SINEs. *Molecular Biology and Evolution* **24(8)**:1656–1666  
DOI [10.1093/molbev/msm083](https://doi.org/10.1093/molbev/msm083).
- McPhail JD. 2007.** *The freshwater fishes of British Columbia*. University of Alberta Press.
- Meyer A, Van de Peer Y. 2005.** From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). *BioEssays* **27(9)**:937–945 DOI [10.1002/bies.20293](https://doi.org/10.1002/bies.20293).
- Meyers LA, Levin DA, Geber M. 2006.** On the abundance of polyploids in flowering plants. *Evolution* **60(6)**:1198–1206 DOI [10.1111/j.0014-3820.2006.tb01198.x](https://doi.org/10.1111/j.0014-3820.2006.tb01198.x).
- Miya M, Friedman M, Satoh TP, Takeshima H, Sado T, Iwasaki W, Yamanoue Y, Nakatani M, Mabuchi K, Inoue JG, Poulsen JY, Fukunaga T, Sato Y, Nishida M. 2013.** Evolutionary origin of the Scombridae (Tunas and Mackerels): Members of a Paleogene adaptive radiation with 14 other pelagic fish families. *PLOS ONE* **8(9)**:e73535 DOI [10.1371/journal.pone.0073535](https://doi.org/10.1371/journal.pone.0073535).
- Muir P, Li S, Lou S, Wang D, Spakowicz DJ, Salichos L, Zhang J, Weinstock GM, Isaacs F, Rozowsky J, Gerstein M. 2016.** The real cost of sequencing: scaling computation to keep pace with data generation. *Genome Biology* **17(1)**:53 DOI [10.1186/s13059-016-0917-0](https://doi.org/10.1186/s13059-016-0917-0).
- Near TJ, Eytan RI, Dornburg A, Kuhn KL, Moore JA, Davis MP, Wainwright PC, Friedman M, Smith WL. 2012.** Resolution of ray-finned fish phylogeny and timing of diversification. *Proceedings of the National Academy of Sciences of the United States of America* **109(34)**:13698–13703 DOI [10.1073/pnas.1206625109](https://doi.org/10.1073/pnas.1206625109).
- Nelson JS. 2006.** *Fishes of the world*. Fourth Edition. New York: John Wiley and Sons, Inc.
- Norden CR. 1961.** Comparative osteology of representative salmonid fishes, with particular reference to the grayling (*Thymallus arcticus*) and its phylogeny. *Journal of the Fisheries Research Board of Canada* **18(5)**:679–791 DOI [10.1139/f61-052](https://doi.org/10.1139/f61-052).
- Oakley TH, Phillips RB. 1999.** Phylogeny of salmonine fishes based on growth hormone introns: Atlantic (*Salmo*) and Pacific (*Oncorhynchus*) salmon are not sister taxa. *Molecular Phylogenetics and Evolution* **11(3)**:381–393 DOI [10.1006/mpev.1998.0599](https://doi.org/10.1006/mpev.1998.0599).
- Ogilvie HA, Bouckaert RR, Drummond AJ. 2017.** StarBEAST2 rings faster species tree inference and accurate estimates of substitution rates. *Molecular Biology and Evolution* **34(8)**:2101–2114 DOI [10.1093/molbev/msx126](https://doi.org/10.1093/molbev/msx126).
- Ohno S. 1970.** *Evolution by gene duplication*. New York: Springer.
- Pamilo P, Nei M. 1988.** Relationships between gene trees and species trees. *Molecular Biology and Evolution* **5**:568–583.
- Pearse DE, Barson NJ, Nome T, Gao G, Campbell MA, Abadía-Cardoso A, Anderson EC, Rundio DE, Williams TH, Naish KA, Moen T, Liu S, Kent M, Moser M, Minkley DR, Rondeau EB, Briec MSO, Sandve SR, Miller MR, Cedillo L, Baruch K, Hernandez AG, Ben-Zvi G, Shem-Tov D, Barad O, Kuzishchin K, Garza JC, Lindley ST, Koop BF, Thorgaard GH, Palti Y, Lien S. 2019.** Sex-dependent dominance maintains migration supergene in rainbow trout. *Nature Ecology & Evolution* **3(12)**:1731–1742 DOI [10.1038/s41559-019-1044-6](https://doi.org/10.1038/s41559-019-1044-6).
- Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018.** Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology* **67(5)**:901–904 DOI [10.1093/sysbio/syy032](https://doi.org/10.1093/sysbio/syy032).
- Rand PS. 2006.** *Hucho perryi*. The IUCN Red List of Threatened Species 2006. e.T61333A12462795. Available at <https://dx.doi.org/10.2305/IUCN.UK.2006.RLTS.T61333A12462795.en>.
- Robertson FM, Gundappa MK, Grammes F, Hvidsten TR, Redmond AK, Lien S, Martin SAM, Holland PWH, Sandve SR, Macqueen DJ. 2017.** Lineage-specific rediploidization is a

- mechanism to explain time-lags between genome duplication and evolutionary diversification. *Genome Biology* **18**(1):111 DOI [10.1186/s13059-017-1241-z](https://doi.org/10.1186/s13059-017-1241-z).
- Sajdak SL, Phillips RB. 1997.** Phylogenetic relationships among *Coregonus* species inferred from the DNA sequence of the first internal transcribed spacer (ITS1) of ribosomal DNA. *Canadian Journal of Fisheries and Aquatic Sciences* **54**(7):1494–1503 DOI [10.1139/f97-057](https://doi.org/10.1139/f97-057).
- Sanford CJ. 1990.** The phylogenetic relationships of salmonoid fishes. *Bulletin of the British Museum* **56**:145–153.
- Scannell DR, Frank AC, Conant GC, Byrne KP, Woolfit M, Wolfe KH. 2007.** Independent sorting-out of thousands of duplicated gene pairs in two yeast species descended from a whole-genome duplication. *Proceedings of the National Academy of Sciences of the United States of America* **104**(20):8397–8402 DOI [10.1073/pnas.0608218104](https://doi.org/10.1073/pnas.0608218104).
- Shedko SV, Miroshnichenko IL, Nemkova GA. 2012.** Phylogeny of salmonids (Salmoniformes: Salmonidae) and its molecular dating: Analysis of nuclear RAG1 gene. *Russian Journal of Genetics* **48**(5):575–579 DOI [10.1134/S1022795412050201](https://doi.org/10.1134/S1022795412050201).
- Shimodaira H, Hasegawa M. 2001.** CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* **17**(12):1246–1247 DOI [10.1093/bioinformatics/17.12.1246](https://doi.org/10.1093/bioinformatics/17.12.1246).
- Smith WL, Everman E, Richardson C. 2018.** Phylogeny and taxonomy of flatheads, scorpionfishes, sea robins, and stonefishes (Percomorpha: Scorpaeniformes) and the evolution of the lachrymal saber. *Copeia* **106**(1):94–119 DOI [10.1643/CG-17-669](https://doi.org/10.1643/CG-17-669).
- Spoelhof JP, Soltis PS, Soltis DE. 2017.** Pure polyploidy: closing the gaps in autopolyploid research. *Journal of Systematics and Evolution* **55**(4):340–352 DOI [10.1111/jse.12253](https://doi.org/10.1111/jse.12253).
- Stamatakis A. 2014.** RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**(9):1312–1313 DOI [10.1093/bioinformatics/btu033](https://doi.org/10.1093/bioinformatics/btu033).
- Stearley RF, Smith GR. 1993.** Phylogeny of the Pacific trouts and salmons (*Oncorhynchus*) and genera of the family Salmonidae. *Transactions of the American Fisheries Society* **122**:1–33 DOI [10.1577/1548-8659\(1993\)122<0001:POTPTA>2.3.CO;2](https://doi.org/10.1577/1548-8659(1993)122<0001:POTPTA>2.3.CO;2).
- Sutherland BJB, Gosselin T, Normandeau E, Lamothe M, Isabel N, Audet C, Bernatchez L. 2016.** Salmonid chromosome evolution as revealed by a novel method for comparing RADseq linkage maps. *Genome Biology and Evolution* **8**:3600–3617 DOI [10.1093/gbe/evw262](https://doi.org/10.1093/gbe/evw262).
- Vladykov VD. 1963.** Review of salmonid genera and their broad geographical distribution. *Transactions of the Royal Society of Canada* **1**:459.
- Vladykov VD, Gruchy CG. 1972.** Comments on the nomenclature of some subgenera of Salmonidae. *Journal of the Fisheries Research Board of Canada* **29**(11):1631–1632 DOI [10.1139/f72-251](https://doi.org/10.1139/f72-251).
- Vuorinen JA, Bodaly RA, Reist JD, Luczynski M. 1998.** Phylogeny of five *Prosopium* species with comparisons with other Coregonine fishes based on isozyme electrophoresis. *Journal of Fish Biology* **53**:917–927.
- Wendel JF. 2015.** The wondrous cycles of polyploidy in plants. *American Journal of Botany* **102**(11):1753–1756 DOI [10.3732/ajb.1500320](https://doi.org/10.3732/ajb.1500320).
- Wilson MVH, Li G-Q. 1999.** Osteology and systematic position of the Eocene salmonid †*Eosalmo driftwoodensis* Wilson from western North America. *Zoological Journal of the Linnean Society* **125**(3):279–311 DOI [10.1111/j.1096-3642.1999.tb00594.x](https://doi.org/10.1111/j.1096-3642.1999.tb00594.x).

- Wilson MVH, Williams RRG. 2010.** Salmoniform fishes: key fossils, supertree, and possible morphological synapomorphies. In: Nelson JS, Schultze HP, Wilson MVH, eds. *Origin and Phylogenetic Interrelationships of Teleosts*. München: Dr. Friedrich Pfeil, 379–409.
- Yasuike M, Jantzen S, Cooper GA, Leder E, Davidson WS, Koop BF. 2010.** Grayling (Thymallinae) phylogeny within salmonids: complete mitochondrial DNA sequences of *Thymallus arcticus* and *Thymallus thymallus*. *Journal of Fish Biology* **76(2)**:395–400  
[DOI 10.1111/j.1095-8649.2009.02494.x](https://doi.org/10.1111/j.1095-8649.2009.02494.x).
- Yu Y, Than C, Degnan JH, Nakhleh L. 2011.** Coalescent histories on phylogenetic networks and detection of hybridization despite incomplete lineage sorting. *Systematic Biology* **60(2)**:138–149  
[DOI 10.1093/sysbio/syq084](https://doi.org/10.1093/sysbio/syq084).
- Zerbino DR, Birney E. 2008.** Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* **18(5)**:821–829 [DOI 10.1101/gr.074492.107](https://doi.org/10.1101/gr.074492.107).