



Inferring phylogenetic structure, hybridization and divergence times within Salmoninae (Teleostei: Salmonidae) using RAD-sequencing

Laurène A. Lecaudey^a, Ulrich K. Schliewen^b, Alexander G. Osinov^c, Eric B. Taylor^d, Louis Bernatchez^e, Steven J. Weiss^{a,*}

^a Department of Zoology, University of Graz, Universitätsplatz 2, A-8010 Graz, Austria

^b Department of Ichthyology, SNSB-Bavarian State Collection of Zoology, D-81247 Munich, Germany

^c Biological Faculty, Lomonosov Moscow State University, Moscow 119991, Russia

^d Department of Zoology, Biodiversity Research Centre and Beaty Biodiversity Museum, University of British Columbia, 6270 University Ave., Vancouver BC V6T 1Z, Canada

^e Département de Biologie, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, QC G1V 0A6, Canada

ARTICLE INFO

Keywords:

Salmonidae
Salvelinus
Salmo
Next-generation sequencing
Ancient introgression
SNPs

ABSTRACT

Phylogenetic studies focusing on Salmonidae have revealed significant obstacles in trying to clarify some interspecific relationships within the Salmoninae subfamily, due to a limited number of markers typed, conflicting phylogenetic signals and ancient hybridization events. To infer reliable phylogenetic relationships, evaluate several putative scenarios of ancient hybridization, and estimate divergence times within Salmoninae, we applied restriction-site associated DNA sequencing (RAD-seq) to 43 samples, including 26 genetic lineages across 21 species, largely representing the subfamily, with an emphasis on the genus *Salvelinus*. We identified 28,402 loci and 28,363 putatively unlinked SNPs, which were used in downstream analyses. Using an iterative k-means partitioned dataset and a Maximum Likelihood approach; we generated a well-supported phylogeny, providing clear answers to several previous phylogenetic uncertainties. We detected several significant introgression signals, presumably ancient, in the genus *Salvelinus*. The most recent common ancestor of Salmonidae dates back to approximately 58.9 MY ago (50.8–64 MY) and the crown age of Salmoninae was estimated to be 37.7 MY (35.2–40.8 MY) using a Bayesian molecular dating analysis with a relaxed molecular clock. The divergence among genera of the subfamily occurred between the late Eocene and middle of the Miocene (\approx 38–11 MY) such as the divergence between the genus *Oncorhynchus* and *Salvelinus*, which we estimated to 21.2 MY ago (95% HPD: 19.8–23.0 MY), while species diversification took place mainly during the Neogene (\approx 22–1.5 MY), with more than half of these events occurring in the last 10 MY.

1. Introduction

The Salmonidae family, consisting of salmon, trout, charr, grayling, whitefishes and their relatives, is a very important group of temperate freshwater fishes in terms of both economic and ecological value; combined with their tetraploid ancestry, life-history diversity and rates of diversification, they have attracted considerable interest from the research community. The family includes 11 extant genera divided into three monophyletic subfamilies: Coregoninae, Thymallinae and Salmoninae (Nelson, 2006). Salmoninae, the most speciose subfamily, contains seven genera: *Brachymystax*, *Hucho*, *Oncorhynchus*, *Parahucho*,

Salmo, *Salvelinus* and *Salvethymus*. Salmonid species offer valuable opportunities to investigate mechanisms of speciation and adaptation within an ecological and evolutionary framework. More specifically, they provide the possibility to study the effect of hybridization and genome duplication on species evolution. Indeed, one of the most remarkable features of salmonid evolutionary history is their autopolyploid origin (Allendorf and Thorgaard, 1984; Svärdson, 1945). They descend from a single tetraploid ancestor resulting from a whole genome duplication event (WGD) known as Ss4R (Lien et al., 2016), which took place around 95MY ago (88 - 103MY) based on the latest estimates (Macqueen and Johnston, 2014). However, since the Ss4R,

Abbreviations: RAD-seq/RAD-sequencing, Restriction-site associated DNA sequencing; SNPs, single nucleotide polymorphisms; MY, million years; RAxML, Randomized Axelerated Maximum Likelihood; BIC, Bayesian Information Criterion; IC, internode certainty; ML, Maximum Likelihood; sd, standard deviation; MCMC, Markov Chain Monte Carlo; HPD, Highest Posterior Density; BS, Bootstrap Support; BI, Bayesian Inference; BER, Bering clade; SIB, Siberian clade; ACD, Acadian clade; ATL, Atlantic clade; ARC, Arctic clade; OKH, Okhotsk Sea clade; NORs, nucleolus organizer regions

* Corresponding author.

E-mail address: steven.weiss@uni-graz.at (S.J. Weiss).

<https://doi.org/10.1016/j.ympev.2018.02.022>

Received 30 August 2017; Received in revised form 11 January 2018; Accepted 20 February 2018

1055-7903/© 2018 Elsevier Inc. All rights reserved.

salmonids have been through a process of rediploidization, by means of genomic reorganizations driven by selection, retaining only part of the ancestral tetraploid genome. It is estimated that up to 25% of the salmonid genome went through delayed rediploidization (Robertson et al., 2017) and around 10% still retains residual tetrasomy (Allendorf et al., 2015; Lien et al., 2016). WGD has an essential role in long-term evolutionary success; it is a key mechanism driving the development of new expression patterns and gene functions providing lineage-specific physiological adaptations, such as anadromy, therefore potentially promoting evolutionary diversification and facilitating speciation (Robertson et al., 2017). The partially delayed rediploidization is thought to have slowed down functional divergence, explaining the delay of at least 30MY between the Ss4R and lineage divergence (Macqueen and Johnston, 2014; Robertson et al., 2017).

There have been numerous comprehensive attempts to evaluate phylogenetic relationships among salmonids, using molecular methods (Crespi and Fulton, 2004; Osinov and Lebedev, 2004; Wang et al., 2011; Wilson and Turner, 2009; Yasuike et al., 2010). Shed'ko et al. (2013, 2012) provided extensive taxon coverage but was limited to mtDNA markers, and several other studies have extended this approach with whole mitogenomes (Campbell et al., 2013; Ma et al., 2015; Macqueen and Johnston, 2014; Sahoo et al., 2015). Other comprehensive studies included multiple nuclear and mitochondrial genes, such as Alexandrou et al. (2013) which focused on the dating of anadromy, while incorporating ancestral character simulation; and Crête-Lafrenière et al. (2012) who have so far provided the most extensive taxon coverage. Macqueen and Johnston (2014) were the first to estimate salmonid subfamily relationships using a large dataset of nuclear genes with a strict 1:1 orthology, which provided strong support for a sister relationship between Coregoninae and Thymallinae. The same authors also provided the first direct estimate of the timing of the whole-genome duplication event of salmonids. Collectively, these efforts have provided considerable clarifications on the phylogenetic relationships among salmonid taxa. Nonetheless, in spite of the substantial research contributions directed toward investigating phylogenetic relationships within Salmoninae, some knowledge gaps persist, presumably due to partially incomplete taxon coverage, limited number of markers, conflicting phylogenetic signals of different genomic regions and potentially ancient hybridization events. Additionally, the contrasting rates of rediploidization of different regions of the genome, following the WGD, has only recently been demonstrated (Lien et al., 2016; Robertson et al., 2017), and therefore its impact on phylogenetic signals within salmonids has been largely neglected.

Due to these various factors, some critical points of salmonid phylogeny remain unsettled, such as the exact position of certain species within the phylogenetic tree, as well as the placement of the two monotypic genera: *Parahucho* and *Salvelinus*. For instance, The Sakhalin taimen, *Parahucho perryi*, was formally included in the genus *Hucho*, despite the lack of morphological support for this designation (Sanford, 2000), but multiple molecular studies support the taxon as constituting a separate and monotypic genus (Crespi and Fulton, 2004; Matveev et al., 2007; Oakley and Phillips, 1999; Osinov, 1991), although its phylogenetic position within Salmoninae is still unclear. Within the genus *Salmo*, two taxa have also undergone systematic revision based on genetic information, namely *Salmo ohridanus* (formerly in the monotypic genus *Acantholingua*) and softmouth trout *Salmo obtusirostris* (formerly *Salmothymus*) (Snoj et al., 2002), but not without controversy. Hybridization has played a role in the evolution of softmouth trout (Sušnik et al., 2007), and despite molecular evidence supporting its inclusion in the genus *Salmo* (Snoj et al., 2002; Sušnik et al., 2007), some authors still question whether or not its unique behavior and morphology could underscore a hybridization event with a more distant taxon (Esteve et al., 2014). The genus *Salvelinus* has been shown to comprise multiple taxa with a history of interspecific hybridization (Baxter et al., 1997; Bernatchez et al., 1995; Gross et al., 2004; Radchenko, 2004; Redenbach and Taylor, 2002; Wilson and

Bernatchez, 1998; Wilson and Hebert, 1993; Yamamoto et al., 2006). Additionally, the long-finned charr, endemic to the Lake El'gygytyn in the Russian Far East (Siberia), is characterized by a unique and highly distinct morphology, and was thought to represent an ancestral form of charr, and was therefore placed in a new genus (*Salvelinus*) (Chereshnev and Skopets, 1990). However, subsequent phylogenetic studies placed it clearly within the genus *Salvelinus* and identified it as the sister-group to the *S. alpinus*–*S. malma* complex (Brunner et al., 2001; Crête-Lafrenière et al., 2012; Osinov et al., 2015; Shed'ko, 2002; Shubina et al., 2013), but this placement has not yet prompted taxonomic change. Thus, there are series of questions and uncertainties concerning the evolution and systematics of salmonids that likely involved various degrees of hybridization or require significantly increased resolution to address and resolve.

Restriction-site associated DNA sequencing (RAD-seq) (Baird et al., 2008; Miller et al., 2007; Rowe et al., 2011) produces large datasets with millions of genome-wide short sequences with deep coverage; and therefore is increasingly used to detect single nucleotide polymorphisms (SNPs) across a large number of loci in phylogenetic studies (Cruaud et al., 2014; Díaz-Arce et al., 2016; Eaton and Ree, 2013; Rubin et al., 2012). RAD-seq largely overcomes the limitation of traditional methods by drastically improving locus sampling across the genome in a single sequencing run, and yielding a much more reliable dataset of sequences and SNPs. This method is promising for systematic studies of closely related taxa, as it also allows the detection of introgression. RAD-seq relies on the retention of enzyme restriction sites across samples in order to obtain homologous sequences. Therefore, when using this method for phylogenetic inference, the age of the family or subfamily of interest is a critical parameter for locus recovery across species, since the number of shared loci is expected to be directly linked to evolutionary rates and divergence, due to a higher number of mutations between more distantly related species. This issue is exacerbated in the case of longer enzyme restriction sites. However, although the number of shared loci in a RAD-seq dataset decreases with the increasing phylogenetic distance between taxa, inadequate or unequal coverage can produce comparable proportions of missing data (Eaton et al., 2017). RAD-sequencing is most useful for resolving shallow phylogenetic questions, but with adequate taxa sampling, good quality DNA samples, increased coverage and accurate sample normalization during library preparation, a sufficient number of orthologous loci can be generated for precise phylogenetic inferences of clades as old as 60 to 80 MY (Cariou et al., 2013; Eaton et al., 2017; Herrera and Shank, 2016; Rubin et al., 2012).

The aim of this study is to investigate and more fully resolve the phylogenetic relationships among salmonid fish species within the Salmoninae subfamily, with a focus on the genus *Salvelinus*; as well as detect putative ancient hybridization events. We focus on clarifying some of the remaining uncertainties and controversial points of Salmoninae systematics using a RAD-seq dataset, including the main representatives of the subfamily, to produce a reliable phylogenetic hypothesis. Additionally, we estimate the divergence time between the different clades and genera.

2. Material & methods

2.1. Taxon sampling

This dataset includes representatives of the 7 genera of the Salmoninae subfamily and a subset of 21 species among 122 extant species of Salmoninae (98 species, > 80%, belong to the combined genera *Salvelinus* and *Salmo*) (Froese and Pauly, 2017; “GBIF: The Global Biodiversity Information Facility,” 2016; Kottelat and Freyhof, 2007); however, the exact number of extant species remains a topic of debate. More precisely, the dataset consist of 43 individuals: one *Brachymystax* species, one *Hucho* species, five *Oncorhynchus* species, five *Salmo* species, seven *Salvelinus* species, one *Thymallus* species and two

Download English Version:

<https://daneshyari.com/en/article/8648863>

Download Persian Version:

<https://daneshyari.com/article/8648863>

[Daneshyari.com](https://daneshyari.com)