Comparative Analysis of the Shared Sex-Determination Region (SDR) among Salmonid Fishes

Joshua J. Faber-Hammond[1,2], Ruth B. Phillips[2,3], and Kim H. Brown[1,*]

[1]Department of Biology, Portland State University

[2]School of Biological Sciences, Washington State University Vancouver

[3]Center for Reproductive Biology, Washington State University, Pullman

*Corresponding author: E-mail: kibr2@pdx.edu.

Data deposition: This project has been deposited in NCBI GenBank under the accession numbers KJ851798, KJ851799, and KJ908737.

Accepted: June 19, 2015

Abstract

Salmonids present an excellent model for studying evolution of young sex-chromosomes. Within the genus, *Oncorhynchus*, at least six independent sex-chromosome pairs have evolved, many unique to individual species. This variation results from the movement of the sex-determining gene, *sdY*, throughout the salmonid genome. While *sdY* is known to define sexual differentiation in salmonids, the mechanism of its movement throughout the genome has remained elusive due to high frequencies of repetitive elements, rDNA sequences, and transposons surrounding the sex-determining regions (SDR). Despite these difficulties, bacterial artificial chromosome (BAC) library clones from both rainbow trout and Atlantic salmon containing the *sdY* region have been reported. Here, we report the sequences for these BACs as well as the extended sequence for the known SDR in Chinook gained through genome walking methods. Comparative analysis allowed us to study the overlapping SDRs from three unique salmonid Y chromosomes to define the specific content, size, and variation present between the species. We found approximately 4.1 kb of orthologous sequence common to all three species, which contains the genetic content necessary for masculinization. The regions contain transposable elements that may be responsible for the translocations of the SDR throughout salmonid genomes and we examine potential mechanistic roles of each one.

Key words: sex-chromsome, transposable element, retrotransposition, comparative genomics, sex-determination, salmonid.

Introduction

It has long been known that salmonid fishes, including trout and salmon, possess an XX/XY sex determination system with some populations having morphologically distinct sex chromosomes (Thorgaard 1977, 1978). Despite the close relationships among species within this group, the specific chromosome responsible for sex determination is variable across the lineage (Phillips et al. 2001; Woram et al. 2003). This wide variety of sex chromosomes among salmonids is believed to have originated as a result of the sex determining region's ability to translocate between chromosomes throughout the related genomes. Supporting this hypotheses, recent work has identified the putative sex determining gene in salmonid fishes (*sdY*) and shown that it is shared across all salmon and trout examined (Yano et al. 2012). While *sdY* appears to be necessary and sufficient for sex-determination in salmonids, it is not the only shared sex-linked genetic marker. Other markers, including *OmyY1, OtY2, OtY3,* and *GH-*Ψ, appear as noncoding male-specific markers in many, but not all, species within the genus *Oncorhynchus* (Du et al. 1993; Brunelli et al. 2008). Based on these regions' similarity between species, it is clear that the sex determining regions in salmonids have similar content, although the size and specific shared genetic content is largely unknown.

This interspecific variation among closely related species presents a unique opportunity to study the evolution of sex chromosomes, given their wide variety and recent genesis. For example, within the genus *Oncorhynchus* there are at least six unique sex-linkage groups (Davidson et al. 2009; Phillips 2013), most of which have their origins within the last 6–8 Myr (Wilson and Turner 2009). The mechanism of *sdY* translocation between species has yet to be determined, but the gene's movement throughout salmonid genomes may be correlated with speciation events based on the number of unique sex chromosome pairs among species. It has been proposed that speciation among salmonids is accompanied by extensive

© The Author(s) 2015. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

genomic rearrangements, which may be accelerated through sudden and widespread bursts of transposon activation (Murata et al. 1996; de Boer et al. 2007). Such sudden releases of transposon activity may be jointly responsible for genome rearrangements and the transposition of the *sdY*, with both contributing to rapid speciation.

Sequences surrounding *OtY3* from Chinook (*Oncorhynchus tshawytscha*) and *OmyY1* from rainbow trout (*Oncorhynchus mykiss*) have been publically available since 2006 and 2008 respectively (Brunelli et al. 2008; GB:DQ393568; GB:EU081756). Although there is significant sequence identity shared between the regions, the lengths of the sequences fail to show the entire shared orthology. This is based on the observation that the contigs both contain nonaligning sequence on the 5′-end while the alignment appears to continue beyond the 3′-ends of the sequences. Despite having common genetic markers, it has been notoriously difficult to retrieve clones for, sequence, and assemble the sex determining region in salmonids due to the high density of repetitive sequence, ribosomal DNA, and transposable elements (TEs). An 800 kb rainbow trout bacterial artificial chromosome (BAC) contig containing *OmyY1* and *sdY* has been sequenced and characterized, identifying a long list of genes in the sex determining region, including many TEs which are candidates for relocating *sdY* among species (Phillips et al. 2013). Despite this, it was unknown which elements are uniquely sex-linked in rainbow trout or shared on the Y chromosome across species (if any) given the unknown boundaries of the shared sex determination region.

With the recent discovery of *sdY*, attempts were made to use the gene to probe clonal BAC libraries. Three clones in the Atlantic salmon (*Salmo salar*) library CHORI-214 were found and confirmed to have *sdY* sequence (Palibroda et al. 2013). In this study, we sequenced these clones and compared the assembly to the publically available *OtY3* contig and the assembly from an improved rainbow trout SDR (sex-determining region) sequencing effort with the goal of estimating the size of the shared SDR and finding its discrete borders. In addition, we extended the known *OtY3*-surrounding Chinook SDR sequence using a combination of uncloned genome walking polymerase chain reaction (PCR) methods, which allowed us to capture the entire shared SDR orthology between all three species containing all functional elements of *sdY*. Examining multiple salmonid Y chromosomes offers a more direct method for determining SDR borders than comparing the X and Y chromosomes for any given species. This is because sequence content outside the SDR is highly repetitive, the X chromosome is expected to be highly dimorphic, and draft salmonid genomes are incomplete in this region making them difficult to probe for the site of *sdY* insertion. Using comparative Y chromosome information, we were able to look for common TEs in the region possibly responsible for the translocation of the SDR, identify a candidate element, and propose a model for translocation of the SDR cassette throughout salmonid genomes.

Materials and Methods

BAC Library Screening and DNA Extraction

A rainbow trout BAC library was developed using EcoR1 digested genomic DNA from the YY Swanson clonal line (produced by androgenesis) (Young et al. 1998; Palti et al. 2004). The library was screened for *OmyY1* as described in Phillips et al. (2013) and OmyY1 positive clones were obtained from the National Center for Cool and Cold Water Aquaculture, ARS-USDA as stab cultures. Two clones (223F06, 143K08) were cultured in 300 ml LB broth +chloramphenicol (12.5 μg/ml) overnight at 37 °C with shaking at 250 revolutions per minute, and DNA was isolated using a BACMAX extraction kit (Epicentre). Samples were confirmed through PCR to contain *OmyY1* and *sdY*.

The CHORI-214 Atlantic Salmon BAC library was created as described in Thorsen et al. (2005). Identification numbers for *sdY*-containing clones were reported on at the Plant and Animal Genome Conference in 2013 (Palibroda et al. 2013). Three clones (367C1, 227J24, 524M13) were ordered from the BACPAC Resources Center at the Children's Hospital Oakland Research Institute (CHORI) as stabs. These clones were cultured and DNA was extracted as described above. *OmyY1* is deleted in Atlantic salmon but *sdY* was PCR confirmed in two out of three samples.

BAC Sequencing and Assembly

Samples were sent to the W.M. Keck Center for Comparative and Functional Genomics at the University of Illinois at Urbana-Champaign for 454 sequencing. Clones were barcoded and multiplexed to retrieve clone-specific sequence. Sequencing yielded 30,744 raw reads covering approximately 29 Mb (N50=1,086) for rainbow trout BACs and 214,516 raw reads covering approximately 204Mb (N50=1,010) for Atlantic salmon BACs.

Bowtie2 was used to align and remove vector (*E. coli*) DNA sequences from the raw data sets. MIRA (version 3.4.0.1) was used to assemble sequences from each clone individually using the following job parameters denovo, genome, accurate and 454 (Chevreux et al. 1999). Primary assemblies were then grouped by species (367C1 + 227J24 + 524M13 and 223F06 + 143K08) and reassembled to obtain longer species-specific contigs. This produced 857 rainbow trout contigs covering approximately 1 Mb (N50 = 1,193), and 47 Atlantic salmon contigs covering approximately 170 kb (N50 = 5,808). Coverage was 33 × in the rainbow trout assembly and 9 × in the Atlantic salmon assembly, considering only contigs greater than 5 kb.

To improve assemblies surrounding repetitive sequences, samples from single BAC clones per species (143K08, 227J24) were sent to the Washington State University Molecular Biology and Genomics Core for Pacific Biosciences sequencing. Two micrograms of DNA was gently sheared to

an average size of 12 kb using a Covaris g-tube and used for PacBio library construction. Size selection was performed by two rounds of 0.45 × AMPure XP (Beckmann Coulter Genomics) purification. The resulting libraries were bound to PacBio P4 DNA polymerase and sequenced using C3 chemistry on two SMRT cells per sample. Clones 227J24 and 143K08 generated 210 and 280 Mb of post filtered data, respectively, that was fed to the HGAP assembly pipeline as implemented within the SMRT analysis pipeline v. 2.3.0. Clone 143K08 assembled into ten contigs covering approximately 200 kb (N50 = 33,379) and 227J24 into 15 contigs covering approximately 249 kb (N50 = 61,122). Coverage was 574 × in the 143K08 assembly and 196 × in the 227J24 assembly.

Genome Walking in Chinook

Using a recently published Chinook salmon complete *sdY* mRNA sequence (GB: KC756279), we were able to hypothesize a tiled alignment with the published *OTY3* contig (GB: DQ393568.1). This allowed us to extend the known Chinook SDR by 2,418 bp, and this additional sequence aligned to the rainbow trout SDR contig without introns. Next, we used genome walking methods as described by Siebert et al. (1995) and Rondeau et al. (2013) to extend the SDR sequence further. Using DNA extracted (DNeasy Blood and Tissue Kit; Qiagen) from fin clips from seven phenotypically adult male Chinook salmon collected from Warm Springs Hatchery (Oregon) during the summer of 2011, we confirmed genetic sex using PCR amplification of both *OtY2* and *GH-Ψ*. Using the extended Chinook sequences as a template we designed primers (table 1) for genome walking. First, 1 μg of pooled genomic DNA was digested by EcoRI (NEB) according to manufacturer's protocol. Following denaturation of the restriction enzyme digest, DNA was purified with the UltraClean PCR Clean-Up Kit (MoBio). The walking adaptor was ligated to digested samples using T4 ligase (NEB) overnight at 16 °C. The EcoRI walking adaptor library was diluted to a total volume of 200 μl.

Two nested PCR reactions were performed with sequence-specific forward primers and reverse primers complementary to the adaptor sequence (table 1). For the first PCR, each reaction contained the following reagents: 1 μl EcoRI library, 1.25 μl forward (CKSDR_F12) and reverse (AP1) primers at 4 nM concentration, 3.25 μl nuclease free H2O, and 6.25 μl LongAmp Hot Start Taq 2 × Master Mix (NEB). PCR conditions were as follows: 65 °C for 5 min, 94 °C for 30 s, 40 cycles of 94 °C for 30 s, 62 °C for 30 s (with −0.5 °C increments per cycle until a lower limit of 52 °C), and 65 °C for 6 min, followed by one final extension of 65 °C for 10 min and a 4 °C hold. Two microliters of PCR product was diluted in H_2O to a final volume of 250 μl, and 1 μl of this dilution was used for a second nested PCR with different forward primers (CKSDR_F8, F11, F13-F15) and reverse primer AP2 (table 1).

Table 1
List of Oligos Designed/Used for Genome Walking in Chinook Salmon

Primer Name	Location in Chinook SDR (bp)	Sequence
CKSDR F12	14,557	ATGTGTATGGGGCCTTACTTACTTAC
CKSDR F8	14,632	GGGCTCAGCAGCTATTCAAG
CKSDR F14	14,768	CCAAACTTGTACCCTAATCACCTCT
CKSDR F15	14,788	CCTCTCCTCAACCTTTTCACCTAT
CKSDR F11	14,799	CCTTTTCACCTATTTTCATTCCCTAT
CKSDR F13	14,943	TCAGATTAATGCTTACTTCATGCTG
AP1	N/A	CTAATACGACTCACTATAGGGC
AP2	N/A	TCGAGCGGCCGCCCGGGCAGGT
GH1 R1	16,343	AAACCGTCCAGTCTTCAACC
GH1 R2	16,369	CAGTTCGGGACACAAAACCT
GH1 R4	17,081	CACTCTCCAATCACCACGTC
CKSDR R1	N/A	ACAAGACAGTTGGCGCTCTT
CKSDR R2	N/A	GGATCGAACCAGGGTCTACA
CKSDR R3	N/A	TCCGCTCGATAGGAGACATT
CKSDR R4	N/A	CTAACGTCCGGTCCACTTGT

Conditions for this second PCR were identical to the first except the initial 65 °C step was omitted.

PCR products were visualized on an agarose gel for proper amplification (supplementary file S1*A*, Supplementary Material online), and samples were prepared for Sanger sequencing on an ABI3130xl. The samples were cleaned using Exo/FastAP using manufacturer protocol (ThermoFisher). PCR sequencing was performed with the BigDye Terminator v3.1 kit and protocol from Applied Biosystems and purified with Agencourt CleanSeq Dye Terminator Removal beads. Samples were sent to Oregon Health and Science University DNA Services Core for sequencing with approximately 1.2 kb of additional sequence obtained on the 3′-end of the contig.

Examining the additional 3′-end sequence on the Chinook SDR, we identified repetitive sequence alignment matching part of a sequenced Chinook clone containing *growth hormone 1* (*GH1*) (GB: EU621900.1). We designed predicted downstream reverse primers (GH1 R1-R7) based on sequence from this clone to use with previously designed CKSDR forward primers and performed long-range PCR on genomic Chinook DNA (table 1; supplementary file S1*B*, Supplementary Material online). Samples were prepared and sent for sequencing as previously described and an additional approximately 900 bp of sequence was obtained. Genome walking sequences generated in this study are available in GenBank under accession KJ908737, and tiled alignments of all sequences used to generate the Chinook salmon SDR are displayed in figure 1.

We designed reverse primers at various distances beyond the 3′-shared SDR orthology with rainbow trout using the rainbow trout sequence as a template (CKSDR R1-R4) (table 1). We then performed long-range low stringency PCR on both rainbow trout and Chinook salmon genomic

DNA, pairing each reverse primer with each of 3 CKSDR forward primers. Positive bands in rainbow trout and negative bands in Chinook salmon suggest that the shared SDR sequence orthology does not continue beyond the repetitive Chinook *GH1* clone sequence alignment at the 3′-end (supplementary file S1C, Supplementary Material online).

Bioinformatics

BLAST was used to identify contigs from the Atlantic salmon and rainbow trout assemblies containing *sdY* and other verified male-specific markers (*OmyY1, OtY2, OtY3, GH-Ψ*) in BAC assemblies. This resulted in a single contig from each assembly, each of which was used to scan the assemblies a second time for additional significant comparative sequence identity. We used MULAN (Ovcharenko et al. 2005) to perform threaded blockset aligner (TBA) analysis with a minimum sequence identity cutoff of 50% using the rainbow trout, Atlantic salmon, and extended Chinook salmon (*OtY3*) contigs, all containing *sdY*. For all candidate evolutionary conserved region (ECR) blocks identified by MULAN, we performed phylogenetic analysis using ClustalW (Larkin et al. 2007) to determine whether sequences in SDR contigs were more related to each other than to other similar sequences found throughout the rainbow trout (Miller M, unpublished data) and Atlantic salmon draft genome assemblies (GB: AGKD02000000). Using only ECR blocks exhibiting direct phylogenetic relatedness, we quantified the total length of alignment across each contig, the total length of alignment not including species-specific insertions, and the percent identity of the entire aligned region.

We performed multiple database searches to look for signatures of TEs. First, BLASTx was run for each *sdY*-containing contig to check for TE protein alignments in the NCBI nr protein database using default parameters. We also searched in repetitive element-specific databases GyDB and Dfam using BLAST and hidden Markov model algorithms (Lloréns et al. 2008; Wheeler et al. 2013). Next, we checked each *sdY*-containing contig for transcript alignments within the cGRASP EST (expressed sequence tag) cluster database using BLASTn default parameters to identify additional regions of the SDR contigs that are potentially transcribed but may not code for proteins. The resulting hits are reported as candidate TEs responsible for relocation of the SDR cassette. Each of these candidate TEs and flanking sequences were analyzed for functional motifs using GenomeNet Motif (http://www.genome.jp/tools/motif/), directionality of transcription, signature of transposition events (i.e., target or insertion sites), and amino acid sequence identity within the SDR contigs (local BLASTx).

We also calculated phylogenetic relatedness of elements between species (ClustalW) using related sequences in the rainbow trout and Atlantic salmon genomes. Neighbor-joining trees were generated using candidate elements within each SDR contig and the top ten alignments in the rainbow trout and Atlantic salmon genomes to each SDR-related sequence. We attempted to identify potential outgroup sequences in a number of different ways including NCBI nr database searches for sequences similar to candidate TEs in nonsalmonid species and searches for nonsalmonid outgroup sequences using translated amino acid sequence intermediaries. The resulting sequences were either too similar to candidate TEs to properly root the tree or too dissimilar to properly align. This is most likely due to the nature of TEs which have a complex evolutionary history within a single species' genome, where related sequences within salmonid genomes can be more dissimilar to candidate TEs than similar sequences from other species. For these reasons, we generated unrooted phylogenetic trees to observe potential clustering of candidate TEs within the SDR using Dendroscope v3.2.10 (Huson and Scornavacca 2012).

Results and Discussion

SDR Comparative Analysis

Assembly of the *sdY*-containing BAC clones in Atlantic salmon and rainbow trout provided SDR contigs of 32,566 bp (GB: KJ851799) and 31,198 bp (GB: KJ851798), respectively. The Chinook salmon SDR was extended to 16,722 bp (supplementary file S2, Supplementary Material online), compared with the previously published 12,567 bp *OtY3* contig. Atlantic salmon is the most distantly related species of the three salmonid fishes examined (Oakley and Phillips 1999), and as such shares lower orthology with Chinook salmon and rainbow

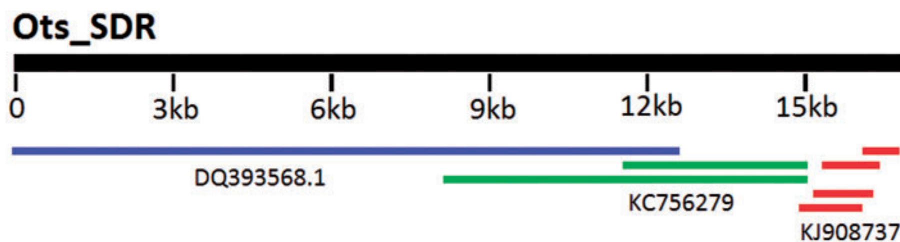

Fig. 1.—Sequence alignments of GenBank accessions DQ393568.1 (*OtY3*), KC756279 (Chinook salmon *sdY*), and KJ908737 along the predicted Ots_SDR scaffold (16,722 bp). KJ908737 represents novel sequence generated by genome-walking methods in this study.

trout than Chinook and rainbow share with each other, both in percent identity and alignment length (table 2). The Atlantic salmon SDR contig aligns with Chinook salmon and rainbow trout across 4.1 kb (minus gaps) with 83.2% and 84.3% sequence identity, respectively. The shared orthology between rainbow trout and Chinook salmon covers approximately 11 kb of sequence (minus gaps) with 96% sequence identity to each other (fig. 2).

It is likely the boundaries of shared orthology at the 5′- and 3′-ends of the SDR contigs represent the true boundaries of the contiguous SDR cassette between *Oncorhynchus* and *Salmo* (i.e., the unique sequence at each end likely does not fall within large gaps and shared orthology does not continue beyond the sequence ends). We base this assessment on local TBA multisequence alignments conducted between entire rainbow trout and Atlantic salmon BAC clone assemblies. This analysis included multiple contigs flanking both sides of the SDR that failed to assemble with the SDR contigs due to the high frequency of repetitive sequences in the region. Other than discrete repetitive sequences matching between libraries, no evidence of continuous orthology from shared ancestry was observed outside the SDR contigs containing *sdY*.

For Chinook salmon, a complete assembly of the SDR contig flanking regions was not available, therefore sequence beyond the 5′- and 3′-ends of the SDR contig are unknown. However, the observed lengths of the Chinook-specific sequences flanking the shared SDR are longer than any internal gap in observed alignment with rainbow trout. PCR tests further suggest that Chinook salmon and rainbow trout orthology ends at the 3′-end of the contig (supplementary file S1C, Supplementary Material online), with primers throughout the 3′-region in the rainbow trout scaffold not amplifying a product in Chinook. It is important to note that shared SDR alignment is expected to extend significantly between Chinook salmon and other related Pacific salmon with different sex-chromosomes. This is based on the observation that neither Atlantic salmon nor rainbow trout have the Y-linked *GH-Ψ* marker, although this marker does amplify in males in many other species within *Oncorhynchus* (Du et al. 1993; Forbes et al. 1994). This may be due to deletions of *GH-Ψ* during transposition of the SDR cassette in certain species or a genesis of the sex-linked marker following rainbow trout's evolutionary split with Pacific salmon. Although a Chinook *GH-Ψ* Cosmid clone has been sequenced and assembled (Devlin RH, unpublished data), connecting it to the Chinook SDR is difficult given the high frequency of repetitive elements in the region and the unknown genetic distance and orientation between the two contigs.

The alignment between the Chinook SDR and Atlantic salmon SDR mirrors the alignment between SDRs from rainbow trout and Atlantic salmon, which is expected given the close phylogenetic relationship of Chinook salmon and rainbow trout. We can draw two conclusions from this; 1) the shared orthology observed between Atlantic salmon (genus *Salmo*) and both *Oncorhynchus* species likely represents the complete SDR orthology shared between genera, 2) all functional elements of the *sdY* gene necessary for masculinization are found within the approximately 4.1 kb of shared orthology. The alignment between the rainbow trout and Chinook salmon SDR is much more extensive, and it is likely that the size and shape of the SDR may differ in each pairwise comparison within the genus. Future work will focus on sequencing SDR contigs from Pacific salmon species with different sex-chromosomes pairs.

Characterization of Candidate Transposons

TC1-like Transposase

This transposon is found within the shared SDR orthology of all three species (fig. 3*A*). While widespread and highly variable in salmonids, *TC1* transposons were first discovered, and therefore best characterized, in *Caenorhabditis elegans* (Goodier and Davidson 1994; Vos and Plasterk 1994; Krasnov et al. 2005). *TC1* and related mariner transposons are often between 1.3 and 2.4 kb, contain a single open reading frame (ORF) for transposase, are flanked by terminal inverted repeats (TIR) (Bessereau 2006) and appear to insert themselves randomly across genomic loci (Vigdal et al. 2002). Analysis of the sequence within the SDR revealed stretches of strong identity to a partial *TC1-like transposase* amino acid sequence in the common carp (*Cyprinus carpio*, GB: AET85182.1). This alignment is found in multiple ORFs (likely due to in/dels occurring since transposition) which would transcribe in the 5′- to 3′-direction relative to all SDR scaffolds. By conducting a local BLASTn analysis with each SDR contig against itself, several short inverted homologies became apparent, and one 115 bp inverted repeat in-particular flanked the *TC1-like transposase* in all species. The distance between inverted repeats ranged from 1,273 bp and 1,477 bp, falling within the expected size range for functional *TC1* elements. Of the three major TEs located within multiple SDR contigs, this is the only element that does not replicate itself during transposition (Vos and Plasterk 1994; Bessereau 2006).

TC1-like transposons are found between *sdY* exons in all three species (fig. 4), therefore the sequences are transcribed along with preprocessed *sdY* RNA molecules. *TC1-like* elements typically transpose via double strand DNA breaks, and it may be possible that a transposase enzyme targeted the element for transposition but instead moved the larger SDR flanking regions throughout the genome. Phylogenetic relationships show the *TC1-like* elements within examined *Oncorhynchus* SDRs have shared ancestry, but they are unique from the element in *S. salar* (fig. 5). It appears the *TC1-like transposon* elements inserted themselves during independent events in *Salmo* and *Oncorhynchus,* and are therefore not necessarily common across salmonids in general.

Despite the independent origin of the element between genera, they may have functioned similarly to move the SDR, but that would have required *TC1-like transposase* to transpose a region much larger than 10 kb, which is undocumented in current literature for a transposon of this type. If the *TC1-like transposon* excised the entire region and moved it to a different chromosome, we would expect to find TIRs flanking the entire shared SDR region as a result of nonrecognition of the apparent TIRs on both the 5′- and 3′-end of the transposon. Although the sequences may have since been deleted or degenerated, there are no such additional inverted repeats found within any SDR contig. If the cut-and-paste mechanism of transposition of the region were viable though, it could serve to explain why introns remain intact in the SDR cassette, as opposed to moving via a processed RNA intermediary.

Table 2
Statistics for Pairwise Alignments between SDR Contigs

		Scaffold Species		
		O. mykiss	***O. tshawytscha***	***S. salar***
Aligned species	*O. mykiss*	—	10,978 bp[a]	4,143 bp
			13,497 bp[b]	4,187 bp
			96%[c]	84.3%
	O. tshawytscha	10,978 bp	—	4,095 bp
		12,088 bp		4,187 bp
		96%		83.2%
	S. salar	4,143 bp	4,095 bp	—
		5,129 bp	6,234 bp	
		84.3%	83.2%	

[a]First statistic reported is the length of alignment excluding gaps.
[b]Second statistic is total length of alignment including gaps.
[c]Third statistic is the percent sequence identity for confirmed ERCs.

Pol-like Protein (nonlong terminal repeat retrotransposon)

These retrotransposon sequences are found downstream of shared SDR orthology on both the Atlantic salmon and rainbow trout contigs. In Atlantic salmon, there is strong alignment to nearly the full length of *pol-like protein* (GB: BAC82624.1) 3.9 kb downstream from the closest shared SDR orthology. Similarly, in rainbow trout, we find strong alignment to nearly the full length of *pol-like protein* located 3.8 kb downstream from shared orthology (fig. 4). These *pol-like protein* sequences contain domains from the non-LTR (nonlong terminal repeat) retrotransposon superfamily (fig. 3*B*). There are a range of characteristics found in non-LTR retrotransposons, the simplest of which are R2 elements

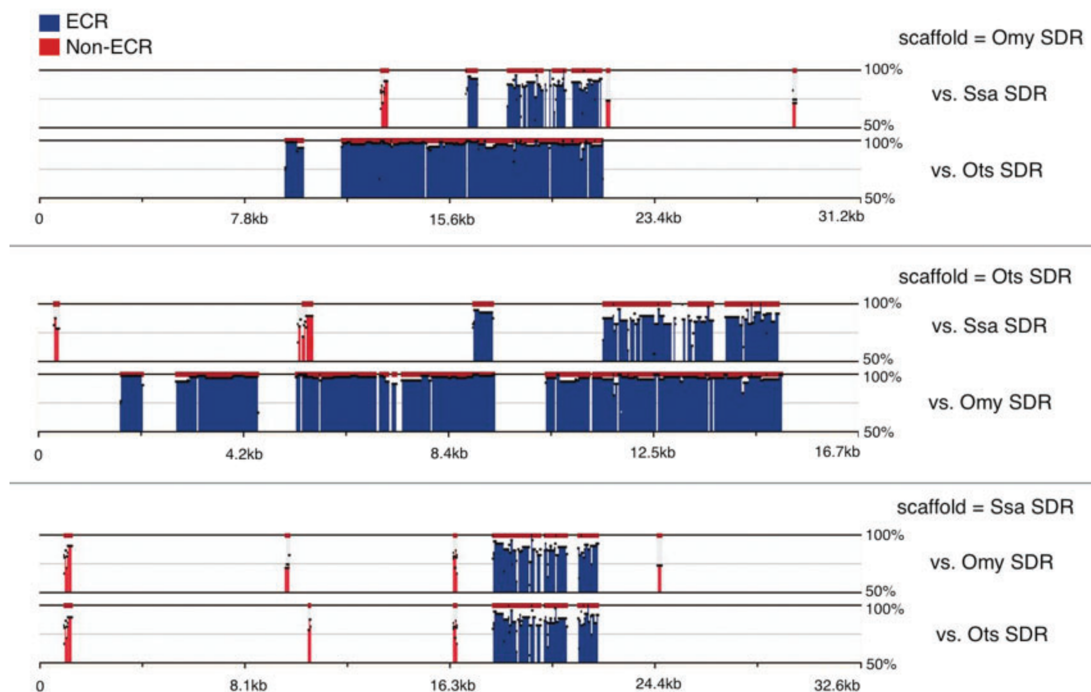

Fig. 2.—Pairwise sliding window alignments between each SDR contig. Evolutionary Conserved Regions (ECRs) were confirmed independently through phylogenetic analysis and are reported in blue. Red alignment blocks have greater than 50% sequence identity, however are not related by descent. Omy, *O. mykiss*, Ots, *O. tshawytscha*, Ssa, *S. salar*. The *sdY* mRNA sequences are almost entirely contained within the orthologous sequence shared by all three species in each comparison.

(originally described in *Bombyx mori*) that contain a single ORF coding for reverse transcriptase and a restriction enzyme-like domain (Burke et al. 1999; Kojima and Fujiwara 2004). More complex non-LTR retrotransposons contain multiple ORFs, and may contain 5′-leucine zippers, nucleic acid binding domains, zinc-finger domains, and CCHH/CCHC domains (Han 2010). No non-LTR element observed codes for all proteins necessary to complete retrotransposition alone, so it is believed that host-encoded factors are necessary to facilitate the process (Han 2010). Therefore, functional non-LTR elements may not have all expected domains. These specific retrotransposons most resemble ORF2 of a LINE-1 element, which normally contains an endo/exonuclease domain at the C-terminus, a reverse transcriptase, and a zinc-finger domain at the N-terminus (Han 2010). ORF2 in the SDR shows domains for endo/exonuclease and reverse transcriptase in both Atlantic salmon and rainbow trout (fig. 3*B*).

Phylogenetic relationships show these two *pol-like protein* sequences do not have direct shared ancestry, therefore it is an unlikely candidate TE responsible for the relocation of the SDR cassette in all salmonid genomes (fig. 6). Additionally, the retrotransposons are found in opposite orientations in Atlantic salmon and rainbow trout. In Atlantic salmon, the *pol-like protein* has a plus/minus alignment relative to the SDR scaffold, but in rainbow trout the gene has a plus/plus orientation. All evidence suggests the two separate *pol-like protein* retrotransposons are coincidentally found in close proximity to the SDR in multiple salmonids. The position and orientation of the element in Atlantic salmon makes it a potential candidate for movement of the SDR via retrotransposon-mediated transduction (Xing et al. 2006), however this is not likely the mechanism in either *Oncorhynchus* species examined unless the signature following transduction was completely lost.

Unnamed Protein Products

In all three SDR contigs examined, there are sequences with adjacent BLASTx hits to multiple *O. mykiss* unnamed protein products (emb: CDQ68591.1, CDR00042.1). Although there is neither specific annotation information nor gene ontology information on these proteins, they appear to be highly repetitive in the Atlantic salmon draft genome, rainbow trout draft

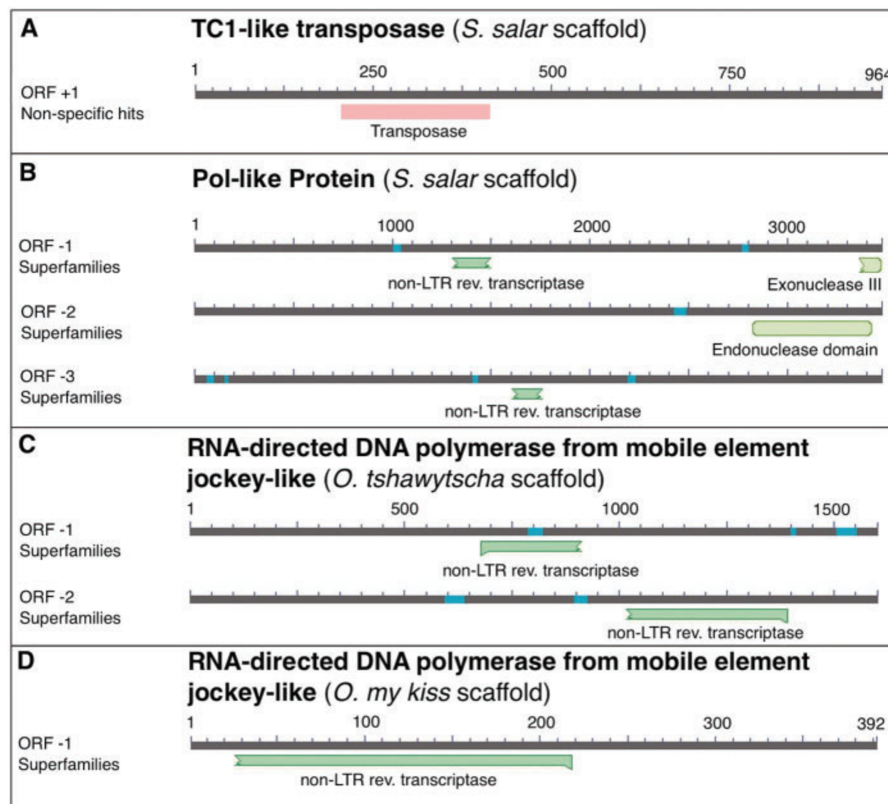

Fig. 3.—Major conserved domains of three candidate TEs described in the text. The shared repetitive element containing several unnamed proteins did not retrieve any conserved domains in BLASTx searches. In panels *A*, *B*, and *C*, only scaffolds from species with the highest BLASTx score to the respective gene are represented. The scaffold lengths correspond to the length of alignment to the TE of interest. In panel *D*, rainbow trout was used as a scaffold to show the difference in alignment length on the 5′-end of the sequence, which may be explained by incomplete reverse transcription. In panels *B* and *C*, there are domains found in different ORFs. Light blue regions of scaffolds have been masked by low complexity filter SEG.

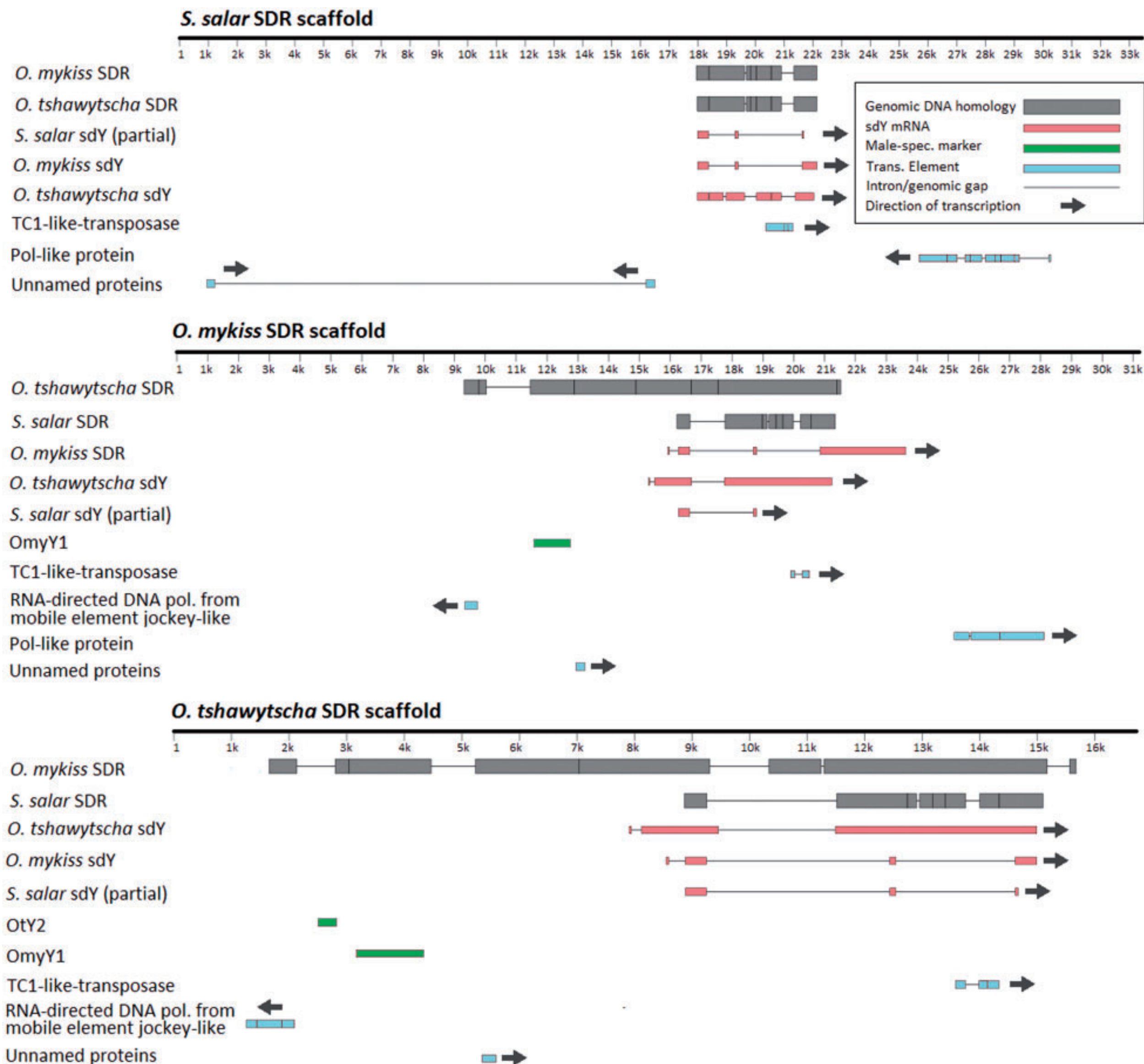

Fig. 4.—Comparative homology among three SDR scaffolds. Important genetic elements within or surrounding the shared SDR are mapped. This includes sex-linked genetic markers, *sdY* mRNAs from each species, and TEs possibly involved in transposition of the region. Accession numbers for the TEs are found in the text. Accession numbers for other sex-linked elements are as follows*: S. salar sdY* (GB:JF826020), *O. mykiss sdY* (GB:NM_001281416), *O. tshawytscha sdY* (GB:KF006343), OmyY1 (GB:JQ995497), OtY2 (GB:GU181208). OmyY1 and OtY2 are both found in Chinook salmon, however OtY2 is not found in rainbow trout and neither marker is found in Atlantic salmon. This explains the well-documented success of the particular markers to accurately amplify in males in the respective species.

genome, and in the cGRASP EST database. The sequence contains two 2Fe-2S ferredoxin-type iron-sulfur binding regions, a flagellar motor protein, an incomplete motif for *DNA polymerase III* subunits gamma and tau, and additional plasmid-related motifs. Taken together, evidence suggests prokaryotic origin and replicative function for these sequences.

The location of the sequences within the three contigs is inconsistent in relation to shared orthology. In Chinook salmon and rainbow trout the sequences are found inside the SDR shared orthology, whereas in Atlantic Salmon the sequence is duplicated with one copy approximately 17 kb and the other copy approximately 1.5 kb upstream of shared orthology, the latter in opposite orientation relative to *Oncorhynchus* (fig. 4). Phylogenetic analysis shows that the sequences in the *Oncorhynchus* SDRs share an evolutionary origin separate from those in the Atlantic salmon SDR (fig. 7).

Although these sequences may play a role in transposition of the SDR throughout salmonid genomes, there is insufficient information in the literature regarding protein function to make such a conclusion. The location, orientation, and

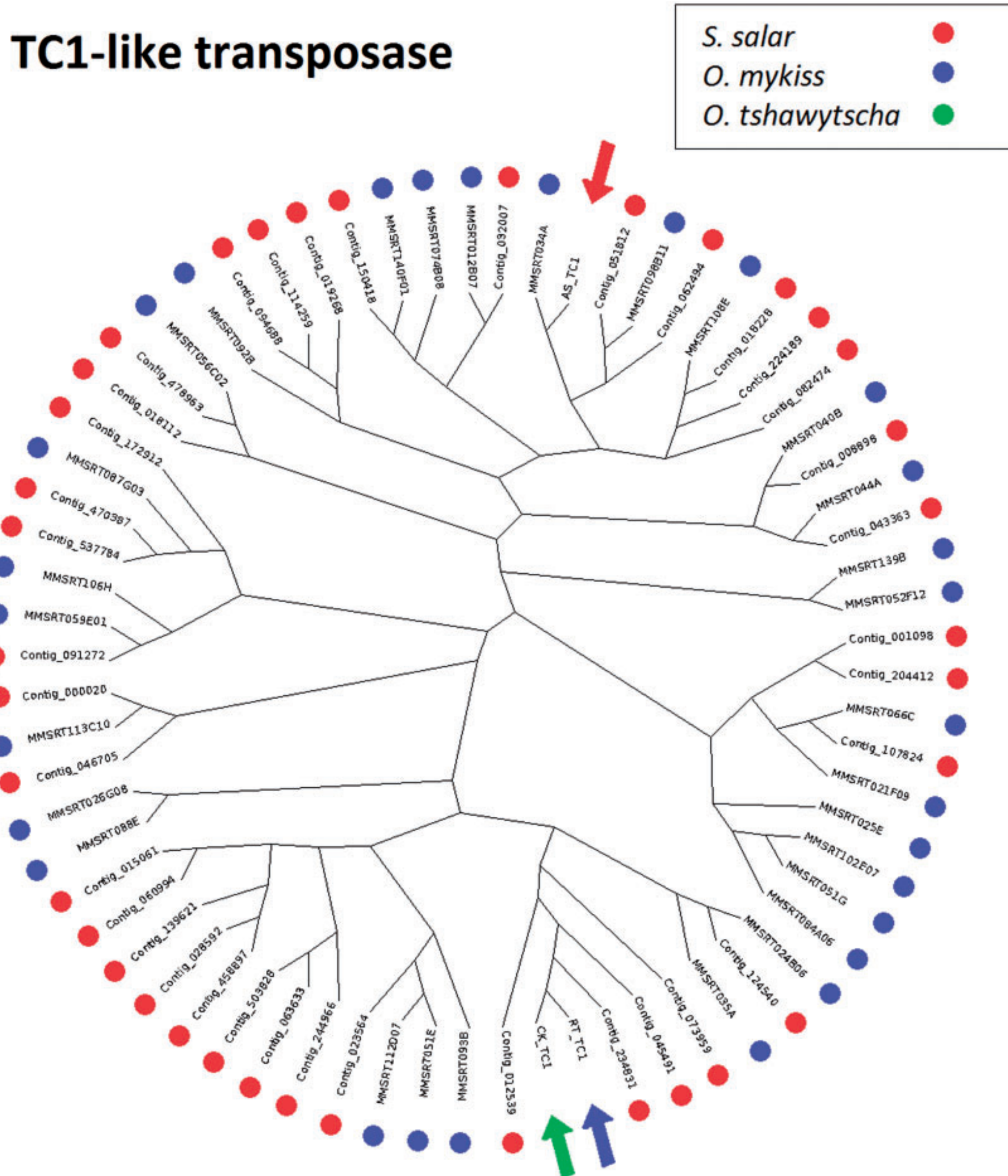

Fig. 5.—Neighbor-joining tree for *TC1-like transposase*. Tree was generated from multiple sequence alignments performed by ClustalW. Input sequences include elements within each SDR contig, and the top ten alignments in the rainbow trout and Atlantic salmon genomes to each SDR-related sequence. Colored circles represent the source species for genomic sequences and colored arrows represent the elements found in the three SDRs.

evolutionary origin of the sequences differ between genera therefore the exact mechanism of transposition would also differ if it were involved in the process.

RNA-Directed DNA Polymerase from Mobile Element Jockey-like (non-LTR retrotransposon)

This transposon resembles *RNA-directed DNA polymerase from mobile element jockey-like* in *Orechromis niloticus* (GB: XP_005471658) and is shared between SDR contigs from both Chinook salmon and rainbow trout. Although this retrotransposon was originally recognized only in Chinook salmon as having similarity to *ReO6* in *Danio rerio* (Brunelli et al. 2008), based on more recent GenBank submissions we see that the described element is part of a larger retrotransposon found in both *Oncorhynchus* species examined. Jockey is a typical LINE element (non-LTR retrotransposon) preferring template RNA over DNA, originally described in *Drosophila* (Ivanov et al. 1991). An independently functional version of this retrotransposon should have the structures listed previously for LINE elements, although lack of certain structures

may be compensated for by other host-encoded factors. Phylogenetic analysis confirms that these sequences share an evolutionary origin and are true orthologs (fig. 8). In Chinook salmon, it appears to only code for reverse transcriptase across two ORFs and is missing most other expected structures (fig. 3*C*). The broken reverse transcriptase template contains enzymatic active sites, NTP-binding sites, and one nucleic acid binding site. In rainbow trout, the length of alignment to the element is shorter, and it only has one ORF coding for the 3′-end (relative to the SDR contig) of reverse transcriptase (fig. 3*D*). Most intriguing is its position in the comparative SDRs; it resides at the 5′-boundary shared between rainbow trout and Chinook salmon, exactly where it would be expected in relation to a predicted SDR transcript if it were to insert and reverse transcribe the SDR cassette at a new genomic locus (fig. 4).

Based on the orientation of alignment to the predicted protein structures, transcription of this element would occur in the 3′- to 5′-direction relative to the SDR scaffolds if it were functional. The gene sits in the same orientation in both species, so that if an RNA copy of the SDR were transcribed in the 3′- to 5′-direction (again, relative to the genomic scaffolds) this non-LTR element could serve as a terminator sequence for transcription, a site of insertion of the SDR sequence into a genomic target, and a site of initiation for reverse transcription of the SDR (fig. 9). Even if the non-LTR element could not produce a functional reverse transcriptase enzyme, it is possible that fully functional retrotransposon machinery from elsewhere in the genome recognized a still-intact promoter and target sequence for retrotransposition of the SDR. Similar phenomena have been observed in vitro for other types of mobile elements (Casacuberta and Santiago 2003; Bolton et al. 2005). It is also possible that the translated retrotransposon

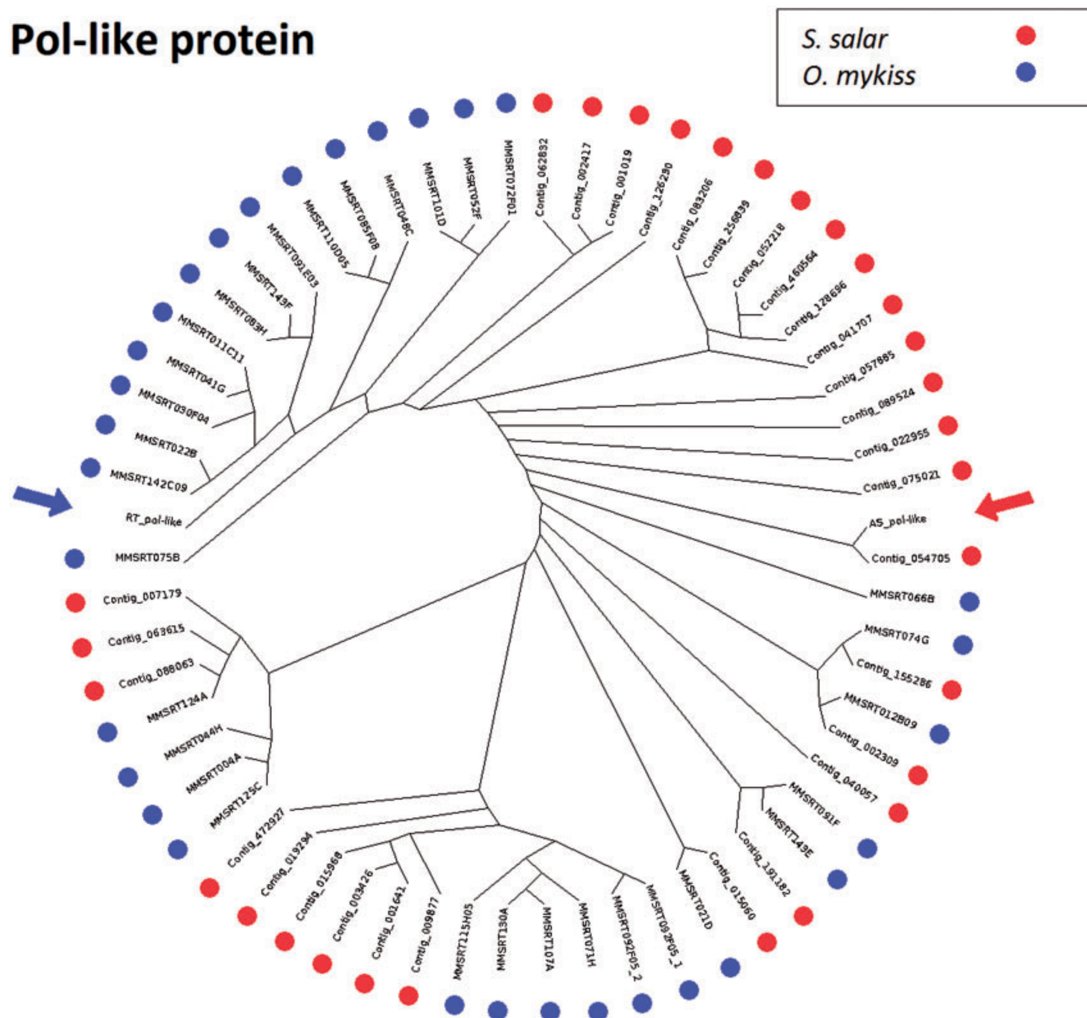

Fig. 6.—Neighbor-joining tree for *pol-like protein*. Tree was generated from multiple sequence alignments performed by ClustalW. Input sequences include elements within each SDR contig, and the top ten alignments in the rainbow trout and Atlantic salmon genomes to each SDR-related sequence. Colored circles represent the source species for genomic sequences and colored arrows represent the elements found in the three SDRs.

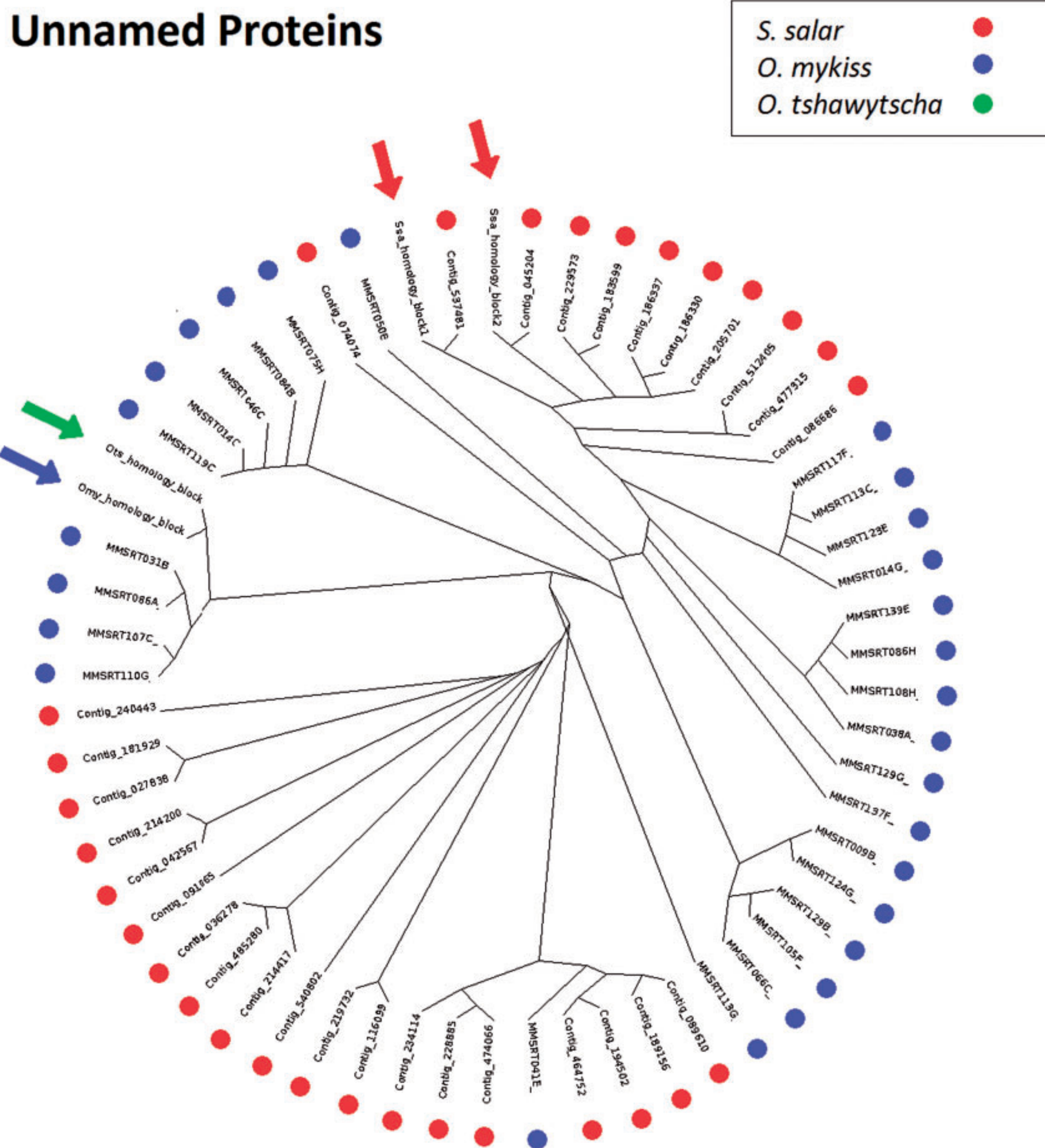

Fig. 7.—Neighbor-joining tree for *Unnamed proteins*. Tree was generated from multiple sequence alignments performed by ClustalW. Input sequences include elements within each SDR contig, and the top ten alignments in the rainbow trout and Atlantic salmon genomes to each SDR-related sequence. Colored circles represent the source species for genomic sequences and colored arrows represent the elements found in the three SDRs.

had partial functionality, lacking of endo/exonuclease activity which was compensated for by other cis or trans elements (Gilbert et al. 2002, de Boer et al. 2007).

It is important to note that retrotransposed gene copies are typically processed prior to insertion in a new genomic locus, and rarely is flanking retrotransposon sequence found at the insertion site (Ewing et al. 2013; Richardson et al. 2014). This phenomenon is primarily studied in humans and other mammals which have been shown to have a single ancestral lineage of L1 families, whereas fish have more than 30 separate lineages, many of which are still active in the genome (Furano et al. 2004). Therefore it is likely that retrotransposon functionality would vary in fish according to this diversity, and errors in reverse transcription of element/transcribed-gene constructs may leave signatures of partial LINE element sequence. Looking at conserved domains for *RNA-directed DNA polymerase from mobile element jockey-like*, we see that only the 3′-end of the element is conserved right at the border of shared orthology in rainbow trout compared with Chinook salmon, potentially resulting from original insertion of the complete element near the Chinook SDR and reverse transcription error during insertion of the SDR on the trout Y-chromosome (figs. 3 and 4). Additionally, in this scenario, the SDR genomic template would contain the reverse

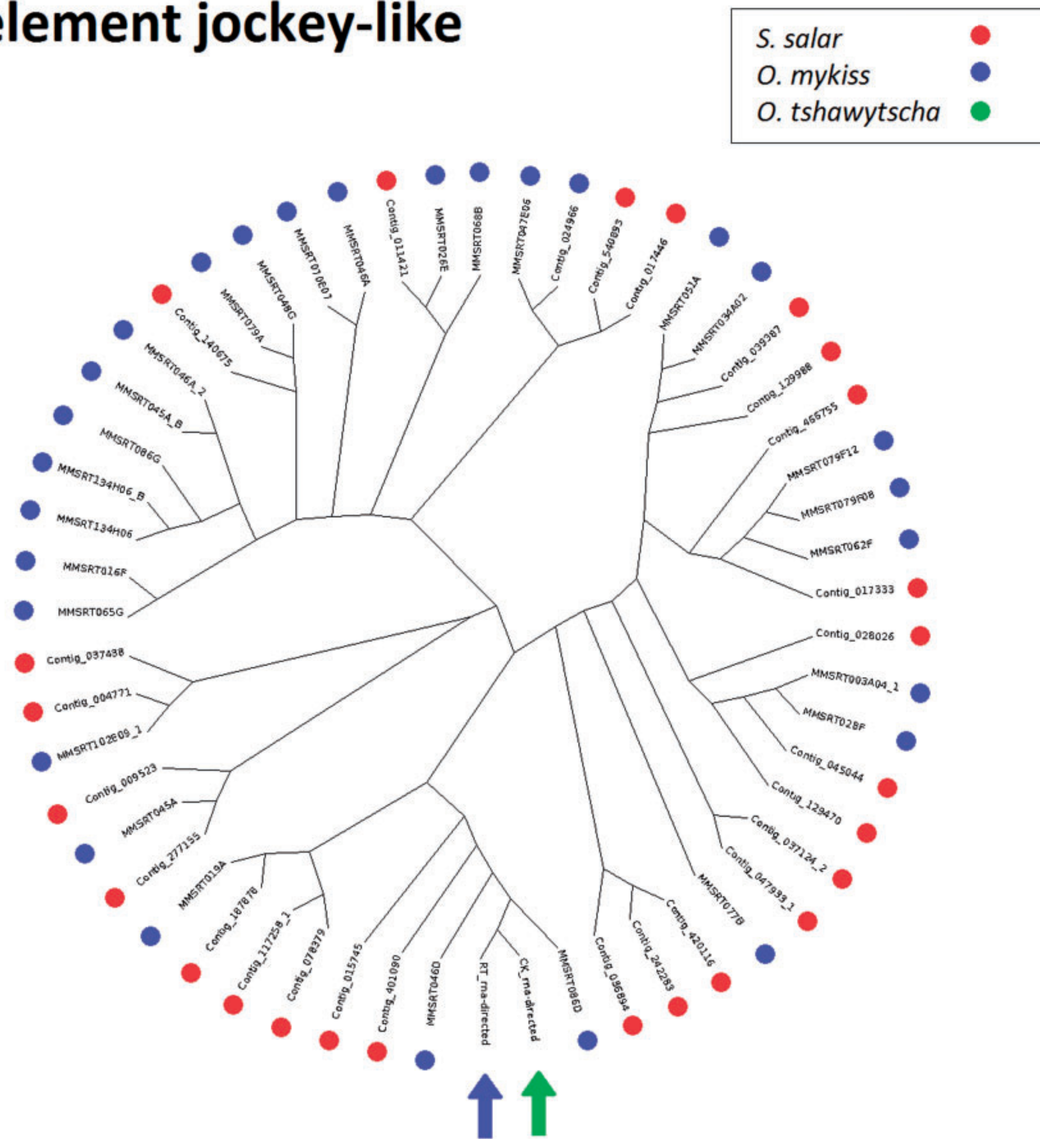

Fig. 8.—Neighbor-joining tree for *RNA-directed DNA polymerase from mobile element jockey-like*. Tree was generated from multiple sequence alignments performed by ClustalW. Input sequences include elements within each SDR contig, and the top ten alignments in the rainbow trout and Atlantic salmon genomes to each SDR-related sequence. Colored circles represent the source species for genomic sequences and colored arrows represent the elements found in the three SDRs.

complement of the *sdY* gene, and the splicing signals would go unrecognized, possibly explaining why the orthologous SDRs share some intronic sequence.

Interestingly, the SDR most often appears at the telomeres of respective salmonid chromosomes (Phillips 2013), and studies have identified an endonuclease-independent (EN_i) mechanism of LINE retrotransposition shown to specifically target telomeric regions in mammals (Morrish et al. 2002, 2007). The researchers found 83% of EN_i retrotransposition events of engineered LINE constructs inserted themselves at telomeric locations, and they note similarities between the mechanisms of EN_i retrotransposition and telomerase function. Given the greater diversity of LINE elements in fish genomes compared with mammals (Furano et al. 2004), this mechanism may be viable in salmonids as well. It may explain why various telomeric SDR loci in salmonids are targeted for insertion by non-fully functional LINE elements, and why SDRs are so often found relocated to other telomeres.

While this retrotransposon is a strong candidate in the region for having a role in relocating the SDR cassette

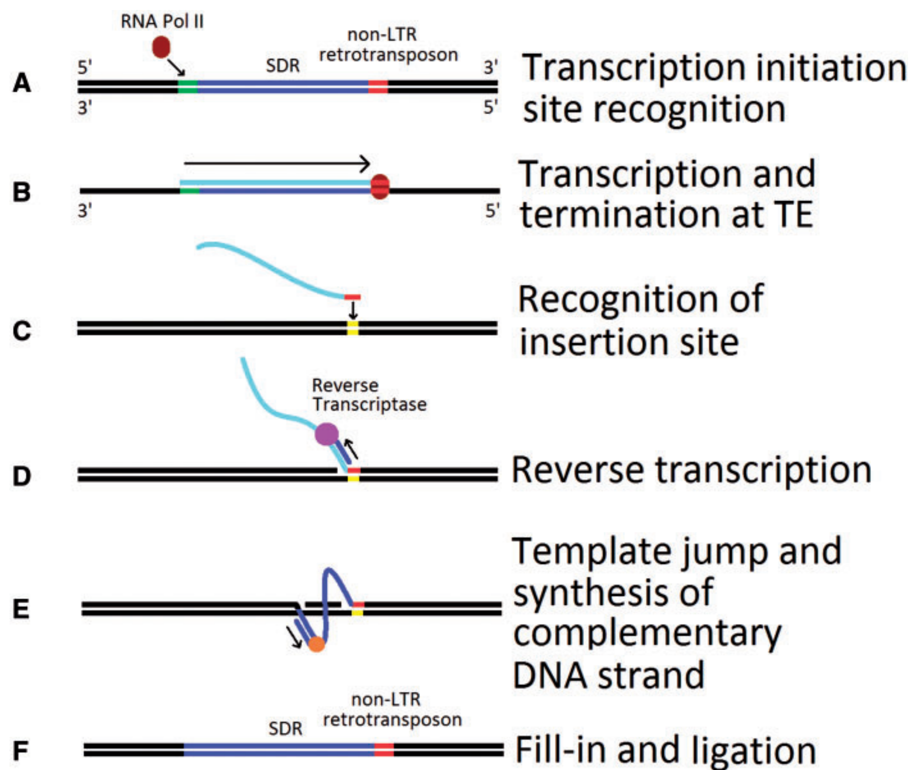

Fig. 9.—A simplified hypothetical model of retrotransposition of the SDR facilitated by RNA-directed DNA polymerase from mobile element jockey-like. (*A*) A promoter upstream of *sdY* acts as recognition site for an RNA polymerase, and (*B*) transcription occurs until a termination site in the retrotransposon is recognized. All internal transcripts in the SDR are naturally transcribed in the opposite direction. (*C*) The retro-TE flanks the SDR transcript and recognizes a target insertion site elsewhere in the genome. (*D*) The SDR is reverse transcribed, (*E*) a template jump occurs, and the complementary DNA strand is synthesized. (*F*) The nonhomologous flaps of DNA are removed, gaps are filled in and ligated by an unknown series of enzymes. The result is duplication of the SDR with the flanking retrotransposon, and a formation of a new Y chromosome. Males with two Y chromosomes may pass one, both, or neither to offspring with a 25% chance of normal XY male offspring with a new sex-chromosome pair.

Table 3

Sequences that May Be Involved in Initiation of Transcription of the Shared SDR Region in Each of the Three Salmonids Examined

Potential Transcription Initiation Sequence	Locus[a] in *O. mykiss*[b]	Locus[a] in *O. tshawytscha*[b]	Locus[a] in *S. salar*[b]
>Repetitive SDR 1	21,687–21,724	15,136–15,168	24,330–24,361
TTTTACATTGTAGTAATTTAGCAGACACTCTTA			
>Repetitive SDR 2	21,376–21,408	13,875–13,916	24,539–24,501[c]
GCGCCGTACAATTTGGCCCAGCGTCGTCCGGGTTAGGGGAGG			
>Repetitive SDR 3	24,668–24,708	—	—
ATCTCCTTTTTCAGTAGTGTGAGGCAGCTTGATATACAAGT			
>Repetitive SDR 4	21,951–22,038	—	—
AACTGCATTGTTGGTTAAGGGCTTGTAAGTAAGCATTTCACTGTAA GATCTACACCTGTTGTATTCGGCGCATGTGACAAATACAATT			
>Repetitive SDR 5	21,687–21,724	—	—
TCATTGTAAATAACAATCTGTTATTAACTGACTTGCCT			
>Repetitve SDR 6	—	—	24,945–25,015
GGTAGATCAGCTTAATATTGCAGATAGATTGTAA CTTCCATCAATGTAATTGTCTGCATCACTTCCAATCC			

[a]The loci reported are the base pair location relative to the sequenced SDR scaffolds.
[b]The most 3′-border of shared SDR homology in the scaffolds of *O. mykiss*, *O. tshawytscha,* and *S. salar* are 21,408, 15,154, and 22,136 bp, respectively.
[c]All sequences align in a plus/plus orientation to the SDR scaffolds except "Repetitive SDR 2" which is plus/minus in *S. salar*.

throughout the salmonid genome due to its position and orientation in two of the three SDR contigs examined, the model is only speculative (fig. 4). It would additionally require the initiation of transcription of the entire region containing *sdY* downstream of the shared orthology, then occurring in the 3′- to 5′-direction without termination until RNA polymerase reaches this retrotransposon. Throughout the shared SDR, there are no obvious genes that sit in an orientation where a terminator sequence would be recognized short of transcribing *RNA-directed DNA polymerase from mobile element jockey-like*, which is consistent with our proposed model of transposition.

We searched for possible shared candidate sites of transcription initiation located nearby or downstream of the 3′-borders of shared SDR orthology by searching for sequence alignments within the cGRASP EST cluster database (table 3). We found six short sequences that are highly repetitive within the transcriptomes of the three study species and in all teleost transcriptomes in the database, indicating they are often transcribed as part of larger ESTs and appear to somehow replicate themselves. In the SDRs, they may be directly involved in transcription as a promoter or enhancer, or indirectly involved as part of a larger unknown transcript. Repetitive SDR 1 is found in all species and sits in the same orientation at the 3′-end of the shared orthology in rainbow trout and Chinook salmon. In Atlantic salmon it is located further downstream of the shared SDR, but remains a viable candidate for involvement in retrotransposition. However, when a non-LTR retrotransposon inserts itself in a new location in a genome, the 5′-end of the sequence can be lost due to incomplete reverse transcription (Gilbert et al. 2002). This is a more likely outcome with a long transcript containing the entire SDR, therefore it is possible the true initiation site for transcription of the SDR is lost during each transposition event and there would be no signature in current genomes. Additional work must be done to establish a possible mechanistic role of these repetitive sequences in transposition.

Additional Considerations

LINE elements have been implicated in retrotransposition of gene transcripts in a wide variety of species resulting in gene duplication which may be evolutionarily beneficial if the target is highly expressed (Ewing et al. 2013; Wong et al. 2013; Richardson et al. 2014). Other duplicate gene fates may include neofunctionalization, subfunctionalization, or pseudogenization to mitigate possible negative effects of duplicate copies (Prince and Pickett 2002; Ewing et al. 2013). In the case of the SDR, it is likely detrimental to have multiple copies of the sex-determining gene on different Y chromosomes. In a male experiencing germ-line retrotransposition of the *sdY* region, however, one-quarter of the sperm produced (and therefore offspring) would possess only the new Y chromosome with the ancestral X chromosome based on Mendelian segregation patterns. Such individuals would develop normally as males passing on the new Y chromosome to half of their offspring (males). If males inheriting both duplicated SDR copies experienced either lower fitness or abnormal development, this would account for rapid elimination of fish with duplicate *sdY* regions. If having duplicate *sdY* copies is lethal then inheriting two copies would only be a problem for the F1 generation, as only normal males would grow up to contribute to the F2 generation with one X and either the ancestral or new Y chromosome. Furthermore, if having multiple *sdY* copies is neutral, they would only segregate together one-quarter of the time in meiosis and that frequency compounded across multiple generations would quickly approach zero.

This phenomenon may be occurring in Tasmanian Atlantic salmon, for example, where the sex-determining locus has been mapped to three separate chromosomes (Eisbrenner et al. 2014). Even though it is expected that the relocation of the SDR is recent since stocks were originally imported in 1965–1967, in each of the 58 families tested the SDR mapped to only one of the three loci. The sudden occurrence of multiple Y chromosomes in separate males in a single population could conceivably contribute to speciation as X chromosomes coevolve and hybrids with multiple sex-chromosome pairs see a potential lower fitness. It would be interesting to test fitness-related hypotheses as Tasmanian Atlantic salmon sex chromosomes continue to evolve.

Although we have proposed non-LTR retrotransposition as one viable model for movement of SDR cassettes, no signature of *RNA-directed DNA polymerase from mobile element jockey-like* in Atlantic salmon was observed, therefore the model may not apply in *S. salar*. It is feasible, however, that the signature for the transposon has been lost in Atlantic salmon through mutation or deletion over evolutionary time and that a separate element has performed a similar function in Tasmanian stocks. TEs tend to accumulate on sex chromosomes (Charlesworth 1991; Ferreira and Martins 2008) and if the implicated TE is deleted, another TE may coincidentally move to that locus and continue to transpose the region. *Pol-like protein* in the Atlantic salmon SDR is also a non-LTR element and it appears to have a unique origin compared with other elements in the region, therefore it may play a mechanistic role in the retrotransposition of the SDR in Atlantic salmon only, or the signature has been lost in *Oncorhynchus*.

It is also possible that the proposed mechanism is unique to the genus *Oncorhynchus* and that different mechanisms or combinations of mechanisms are at work among the different genera. Alternatively, there may be an as-of-yet undescribed mechanism of transposition at work in all salmonids that is undetectable by current database searches, just as the *ReO6*-related retrotransposon was originally unreported in the rainbow trout SDR due to database incompleteness during publication (Brunelli et al. 2008). Despite the evidence presented

here, it may be the case that transposition may not play a role at all, and instead some form of nonhomologous recombination or chromosomal translocation may explain the movement of the region (Woram et al. 2003). Analysis of additional salmonid SDRs should provide more clarity to the evolutionary history of the region.

Conclusions

Sex-determination in salmonids has been a topic of considerable interest in the scientific community for a long time. The recent discovery of *sdY* paved the way for our comparative species analysis, in which we observe the size and content of the shared SDR for the first time. This information will aid in studying the mechanism of salmonid sex determination by narrowing the scope to a discrete genomic region. Additionally, it offers evolutionary insights implicating transposition as a catalyst for young sex-chromosome formation and possibly as a mechanism of speciation. To enhance our understanding of salmonid sex chromosome evolution, characterizing the shared SDR from more salmonid species should be a priority. This work will facilitate more advances and evolutionary insights toward this unique and interesting biological system.

Acknowledgments

The authors would like to thank the Oregon Health & Science University DNA Services Core, Washington State University Molecular Biology and Genomics Core, and the W.M. Keck Center for Comparative and Functional Genomics at the University of Illinois at Urbana-Champaign for Sanger, PacBio, and 454 sequencing. Specifically, the authors would like to thank M. Reeves, M. Wildung and D. Pouchnik, and J. Han at those respective institutions for their support. They also thank M. Miller for providing a copy of the unpublished rainbow trout draft genome, J. DeKoning, J. Brunelli, and G. Thorgaard for their previous work on this project which paved the way for this comparative study, and the anonymous reviewers who offered many helpful suggestions to improve this manuscript. This study was supported by grants from the Agriculture and Food Research Initiative 2009-35205-05067 to R.B.P. and National Institute of Environmental Health Sciences (R00ES018892) to K.H.B. Research conducted at Portland State University and Washington State University Vancouver. Finally, this work is dedicated to the life and memory of Brian Hammond.

Supplementary Material

Supplementary files S1 and S2 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Literature Cited

Bessereau JL. 2006. Transposon in *C. elegans*. In: WormBook: The Online Review of *C. elegans* Biology [Internet]. Pasadena (CA): WormBook; 2005-. Available from: http://www.ncbi.nlm.nih.gov/books/NBK19727/.

Bolton EC, Coombes C, Eby Y, Cerdell M, Boeke JD. 2005. Identification and characterization of critical cis-acting sequences within the yeast Ty1 retrotransposon. RNA 11(3):308–322.

Brunelli JP, Wertzler KJ, Sundin K, Thorgaard GH. 2008. Y-specific sequences and polymorphisms in rainbow trout and Chinook salmon. Genome 51:739–748.

Burke WD, Malik HS, Jones JP, Eickbush TH. 1999. The domain structure and retrotransposition mechanism of R2 elements are conserved throughout arthropods. Mol Biol Evol. 16(4):502–511.

Casacuberta JM, Santiago N. 2003. Plant LTR-retrotransposons and MITEs: control of transposition and impact on the evolution of plant genes and genomes. Gene 311:1–11.

Charlesworth B. 1991. The evolution of sex chromosomes. Science 251(4997):1030–1033.

Chevreux B, Wetter T, Suhai S. 1999. Genome sequence assembly using trace signals and additional sequence information. In: Computer Science Biology: Proceedings of the German Conference on Bioinformatics GCB'99 Hannover (Germany): GCB. p. 45–56.

Davidson WS, Huang TK, Fujiki K, Von Schalburg KR, Koop BF. 2009. The sex determining loci and sex chromosomes in the family salmonidae. Sex Dev. 3(2-3):78–87.

de Boer JG, Yazawa R, Davidson WS, Koop BF. 2007. Bursts and horizontal evolution of DNA transposons in the speciation of pseudotetraploid salmonids. BMC Genomics 8(1):422.

Du SJ, Devlin RH, Hew CL. 1993. Genomic structure of growth hormone genes in chinook salmon (*Oncorhynchus tshawytscha*): presence of two functional genes, GH-I and GH-II, and a male-specific pseudogene, GH-Ψ. DNA Cell Biol. 12(8):739–751.

Eisbrenner WD, et al. 2014. Evidence for multiple sex-determining loci in Tasmanian Atlantic salmon (*Salmo salar*). Heredity 113:86–92.

Ewing A, et al. 2013. Retrotransposition of gene transcripts leads to structural variation in mammalian genomes. Genome Biol. 14:R22.

Ferreira IA, Martins C. 2008. Physical chromosome mapping of repetitive DNA sequences in Nile tilapia *Oreochromis niloticus*: evidences for a differential distribution of repetitive elements in the sex chromosomes. Micron 39(4):411–418.

Forbes SH, Knudson KL, North TW, Allendorf FW. 1994. One of two growth hormone genes in coho salmon is sex linked. Proc Natl Acad Sci U S A. 91:1628ñ1631.

Furano AV, Duvernell DD, Boissinot S. 2004. L1 (LINE-1) retrotransposon diversity differs dramatically between mammals and fish. Trends Genet. 20(1):9–14.

Gilbert N, Lutz-Prigge S, Moran JV. 2002. Genomic deletions created upon LINE-1 retrotransposition. Cell 110(3):315–325.

Goodier JL, Davidson WS. 1994. Tc1 transposon-like sequences are widely distributed in salmonids. J Mol Biol. 241(1):26–34.

Han JS. 2010. Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions. Mobile DNA 1:15.

Huson DH, Scornavacca C. 2012. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. Syst Biol. 61(6):1061–1067

Ivanov VA, Melnikov AA, Siunov AV, Fodor II, Ilyin YV. 1991. Authentic reverse transcriptase is coded by jockey, a mobile *Drosophila* element related to mammalian LINEs. EMBO J. 10(9):2489.

Kojima KK, Fujiwara H. 2004. Cross-genome screening of novel sequence-specific non-LTR retrotransposons: various multicopy RNA genes and microsatellites are selected as targets. Mol Biol Evol. 21(2):207–217.

Krasnov A, Koskinen H, Afanasyev S, Mölsä H. 2005. Transcribed Tc1-like transposons in salmonid fish. BMC Genomics 6(1):107.

Larkin MA, et al. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23(21) 2947–2948.

Lloréns C, Futami R, Bezemer D, Moya A. 2008. The Gypsy Database (GyDB) of mobile genetic elements. Nucleic Acids Res. 36(Suppl 1):D38–D46.

Morrish TA, et al. 2002. DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. Nat Genet. 31(2):159–165.

Morrish TA, et al. 2007. Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres. Nature 446(7132):208–212.

Murata S, Takasaki N, Saitoh M, Tachida H, Okada N. 1996. Details of retropositional genome dynamics that provide a rationale for a generic division: the distinct branching of all the Pacific salmon and trout (*Oncorhynchus*) from the Atlantic salmon and trout (*Salmo*). Genetics 142(3):915–926.

Oakley TH, Phillips RB. 1999. Phylogeny of salmonine fishes based on growth hormone introns: Atlantic (*Salmo*) and Pacific (*Oncorhynchus*) salmon are not sister taxa. Mol Phylogenet Evol. 11(3):381–393.

Ovcharenko I, et al. 2005. Mulan: multiple-sequence local alignment and visualization for studying function and evolution. Genome Res. 15:184–194.

Palibroda EI, Lubieniecki KP, Lai YYY, Eisbrenner WD, Davidson WS. 2013. The Organization of The Atlantic Salmon (Salmo salar) sdY Gene and Its Potential Genomic Environments; (W056 Aquaculture 2). Presented at the Plant and Animal Genome XXI Conference, January 13, 2013, San Diego, CA. https://pag.confex.com/pag/xxi/webprogram/Paper6852.html.

Palti Y, Gahr SA, Hansen JD, Rexroad CE. 2004. Characterization of a new BAC library for rainbow trout: evidence for multi-locus duplication. Anim Genet. 35(2):130–133.

Phillips RB. 2013. Evolution of the sex chromosomes in salmonid fishes. Cytogenet Genome Res. 141(2-3):177–185.

Phillips RB, et al. 2013. Characterization of the OmyY1 region on the rainbow trout Y chromosome. Int J Genomics. Article ID 261730; doi:10.1155/2013/261730.

Phillips RB, Konkol NR, Reed KM, Stein JD. 2001. Chromosome painting supports lack of homology among sex chromosomes in *Oncorhynchus, Salmo*, and *Salvelinus* (*Salmonidae*). Genetica 111:119–123.

Prince VE, Pickett FB. 2002. Splitting pairs: the diverging fates of duplicated genes. Nat Rev Genet. 3(11):827–837.

Richardson SR, Salvador-Palomeque C, Faulkner GJ. 2014. Diversity through duplication: whole-genome sequencing reveals novel gene retrocopies in the human population. Bioessays 36:475–481.

Rondeau EB, et al. 2013. Genomics of sablefish (*Anoplopoma fimbria*): expressed genes, mitochondrial phylogeny, linkage map and identification of a putative sex gene. BMC Genomics 14(1):452.

Siebert PD, Chenchik A, Kellogg DE, Lukyanov KA, Lukyanov SA. 1995. An improved PCR method for walking in uncloned genomic DNA. Nucleic Acids Res. 23(6):1087.

Thorgaard GH. 1977. Heteromorphic Sex Chromosomes in Male Rainbow Trout. Science 196(4292):900–902.

Thorgaard GH. 1978. Sex chromosomes in the sockeye salmon: a Y-autosome fusion. Can J Genet and Cytol 20(3):349–354.

Thorsen J, et al. 2005. A highly redundant BAC library of Atlantic salmon (*Salmo salar*): an important tool for salmon projects. BMC Genomics 6(1):50.

Vigdal TJ, Kaufman CD, Izsvák Z, Voytas DF, Ivics Z. 2002. Common physical properties of DNA affecting target site selection of sleeping beauty and other Tc1/mariner transposable elements. J Mol Biol. 323(3):441–452.

Vos JC, Plasterk RH. 1994. Tc1 transposase of *Caenorhabditis elegans* is an endonuclease with a bipartite DNA binding domain. EMBO J. 13(24):6125.

Wheeler TJ, et al. 2013. Dfam: a database of repetitive DNA based on profile hidden Markov models. Nucleic Acids Res. 41(D1):D70–D82.

Wilson WD, Turner TF. 2009. Phylogenetic analysis of the Pacific cutthroat trout (*Oncorhynchus clarki* ssp.: *Salmonidae*) based on partial mtDNA ND4 sequences: a closer look at the highly fragmented inland species. Mol Phylogenet Evol. 52(2):406–415.

Wong K, Adams DJ, Keane TM. 2013. Go retro and get a GRIP. Genome Biol. 14.3:108.

Woram RA, et al. 2003. Comparative genome analysis of the primary sex-determining locus in salmonid fishes. Genome Res. 13:272–280.

Xing J, et al. 2006. Emergence of primate genes by retrotransposon-mediated sequence transduction. Proc Natl Acad Sci U S A. 103(47):17608–17613.

Yano A, et al. 2012. An immune-related gene evolved into the master sex-determining gene in rainbow trout, *Oncorhynchus mykiss*. Curr Biol. 22(15):1423–1428.

Young WP, Wheeler PA, Coryell VH, Keim P, Thorgaard GH. 1998. A detailed linkage map of rainbow trout produced using doubled haploids. Genetics 148(2):839–850.

Associate editor: Judith Mank