Salmonid chromosome evolution as revealed by a novel method for comparing RADseq linkage maps

Ben J. G. Sutherland[1], Thierry Gosselin[1], Eric Normandeau[1], Manuel Lamothe[2], Nathalie Isabel[2], Céline Audet[3] and Louis Bernatchez[1]

[1] Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, QC G1V 0A6, Canada

[2] Centre de Foresterie des Laurentides, Ressources Naturelles Canada, Québec, QC G1V 4C7, Canada

[3] Institut des Sciences de la Mer de Rimouski, Université du Québec à Rimouski, Rimouski, QC G5L 3A1, Canada

* Author for Correspondence: Ben Sutherland, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, QC, Canada

Phone: 1 (418)-656-2131 (x8455)

Email: ben.sutherland.1@ulaval.ca (BS)

Data Deposition: Raw sequence data has been uploaded to SRA under BioProject PRJNA308100 and accession SRP068206.

Abstract

Whole genome duplication (WGD) can provide material for evolutionary innovation. Assembly of large, outbred eukaryotic genomes can be difficult, but structural rearrangements within such taxa can be investigated using linkage maps. RAD sequencing provides unprecedented ability to generate high-density linkage maps for non-model species, but can result in low numbers of homologous markers between species due to phylogenetic distance or differences in library preparation. Family Salmonidae is ideal for studying the effects of WGD as the ancestral salmonid underwent WGD relatively recently, around 65 million years ago, then rediploidized and diversified. Extensive synteny between orthologous chromosomes occurs in extant salmonids, but each species has both conserved and unique chromosome arm fusions and fissions. Here we generate a high-density linkage map (3826 markers) for the *Salvelinus* genera (Brook Charr *S. fontinalis*), and then identify orthologous chromosome arms among the other available salmonid high-density linkage maps, including six species of *Oncorhynchus*, and one species for each of *Salmo* and *Coregonus*, as well as the sister group for the salmonids, *Esox lucius* for homeolog designation. To this end, we developed MapComp, a program that identifies identical and proximal markers between linkage maps using a reference genome of a related species as an intermediate. This approach increases the number of comparable markers between linkage maps by 5-fold, enabling a characterization of the most likely history of retained chromosomal rearrangements post-WGD, and identifying several conserved chromosomal inversions. Analyses of RADseq-based linkage maps from other taxa will also benefit from MapComp, available at: https://github.com/enormandeau/mapcomp/

Keywords: comparative genomics; linkage map; RADseq; salmon; whole genome duplication

Introduction

Whole genome duplication (WGD) can provide the raw material for evolutionary innovation by generating redundant copies of all chromosomes (i.e. producing homeologous chromosome pairs). After WGD, the genome can then undergo rediploidization while retaining all homeologous chromosomes, thereby doubling the pre-duplication chromosome number. After rediploidization, homeologous pairs diverge through several mechanisms. Gene copies can evolve new functions, sub-functionalize the original function between the two copies or, most frequently, accumulate mutations that disrupt functionality of one copy (Ohno 1970; Force et al. 1999; Brunet et al. 2006). Cross-taxa analyses suggest that during rediploidization, genes that only keep one copy are often retained within the same homeologous chromosome (Sankoff et al. 2010). However, this non-random retention is not always observed (Berthelot et al. 2014). Interestingly, rediploidization does not always complete. For example, some homeologous chromosome arms in the salmonids continue recombining, which is referred to as residual tetrasomy (Allendorf et al. 2015).

Eukaryotic genomes with ancestral WGD (e.g. pseudotetraploid genomes) are challenging to assemble (Davidson et al. 2010). Linkage maps can be highly useful for comparing chromosomal evolution among lineages using shared markers between maps to identify corresponding chromosomes (i.e. orthologous chromosomes) between species (Naish et al. 2013; Kodama et al. 2014). Furthermore, high quality, dense linkage maps are valuable for validating and orienting genomic scaffolds (Fierst 2015; Mascher & Stein 2014), especially for cases of residual polyploidy, large genome size, and high repeat content (Ming & Man Wai 2015; Amores et al. 2014). Recent advances in sequencing, such as through reduced-representation library sequencing (e.g. RADseq) (Baird et al. 2008; Elshire et al. 2011; Andrews

et al. 2016), have made high-density linkage maps increasingly easy to produce. These methods provide thousands of markers without requiring marker design effort. They can also generate haplotype loci (i.e. loci with more than one SNP) to improve mapping resolution by increasing ability to assign alleles to a parent (Catchen et al. 2011). RADseq-based SNP markers are contained in short sequence fragments, which allow for mapping against a genome to identify nearby genes or physical distances between markers (Amores et al. 2011; Henning et al. 2014). RADseq also enables comparative genomics through the use of direct marker-to-marker comparisons to find homologous (i.e. nearly identical markers between species) markers between linkage maps (Kodama et al. 2014). The ability to identify orthologous chromosomes between maps is dependent on being able to identify enough shared markers between species, and this decreases with phylogenetic distance due to sequence divergence (Gonen et al. 2015). This issue is compounded further when different protocols or restriction enzymes are used for library generation. Due to this, it has been suggested to use a common enzyme and protocol to ensure compatibility of maps (Larson et al. 2015), but this may not always be possible or desirable. Here we developed a method to use an intermediate reference genome for integrating linkage maps of different species by pairing both homologous and proximal markers in order to investigate orthologous and syntenic relationships among species.

Salmonids are a highly relevant study system for investigating the effects of WGD. The ancestor of modern day salmonids experienced a relatively recent salmonid-specific (4R) WGD, and subsequently underwent rediploidization (Allendorf & Thorgaard 1984; Davidson et al. 2010). Post-WGD, the salmonid lineage diversified into three subfamilies, 11 genera and more than 60 described species (Crête-Lafrenière et al. 2012), although this diversification was likely due to environmental factors rather than being caused by WGD (Macqueen & Johnston 2014).

Analysis of the Atlantic Salmon genome suggests that the rediploidization process was rapid and that two classes of homeolog similarity exist: immediately rediploidized homeologs and those remaining in residual tetrasomy (see Figure 3b in Lien et al. 2016). Although much remains to be understood about this process in salmonids, fundamental work on chromosomal evolution has been conducted using cytogenetics and genetic maps (Phillips & Ráb 2001; Naish et al. 2013). From linkage map comparisons using homologous markers, it is known that the same eight pairs of orthologous homeologs are residually tetrasomic in Chinook Salmon (Brieuc et al. 2014), Coho Salmon (Kodama et al. 2014) and Sockeye Salmon (Larson et al. 2015).

Chromosomal evolution within family Salmonidae (i.e. whitefish, trout, charr and salmon) is typified by centric Robertsonian fusions, whereby two acrocentric chromosomes fuse into one larger metacentric chromosome, retaining the total number of chromosome arms (*nombre fondamental* (NF) = 100) (Phillips & Ráb 2001). Fissions and whole arm translocations can also occur, subsequently separating the fused metacentric chromosomes. Cytogenetic research has identified the presence of two major karyotype groups in salmonids that differ in the number of retained chromosome fusion events. Type A species (2n = ~80 chromosomes) have more acrocentric than metacentric chromosomes, whereas Type B species (2n = ~60 chromosomes) have more metacentric than acrocentric chromosomes (Phillips & Ráb 2001). Adaptive mechanisms or selective forces driving these rearrangements and correlation with habitat or species biology remain generally unknown (Phillips & Ráb 2001). Among species, highly retained micro and macro synteny is expected between orthologous chromosome arms (Kodama et al. 2014), and within a species collinearity is observed between homeologs (Berthelot et al. 2014). Conservation of chromosome fusions has been partially explored between Chinook *O. tshawytscha* and Coho Salmon *O. kisutch* with Atlantic Salmon, allowing for the

phylogenetic timing of rearrangements in these species (Kodama et al. 2014). However, this characterization has not been examined across the salmonid lineage using all available high-density maps. Furthermore, some genera do not yet have high-density genetic maps available.

The fusions of acrocentric to metacentric chromosomes may have important implications for the rediploidization process in salmonid evolution. Residual tetrasomy is thought to always involve at least one homeolog in a metacentric fusion (Wright et al. 1983; Allendorf et al. 2015), indicating that fusions can affect rediploidization efficiency. Metacentric chromosomes are more likely to form tetravalents at meiosis, permitting recombination between homeologs and preventing sequence divergence particularly in male salmonid telomeric regions (Allendorf & Danzmann 1997; Waples et al. 2016; Allendorf et al. 2015; May & Delany 2015). Salmonid species with few fusions may therefore provide new insight on the rediploidization process. This highlights the importance of characterizing orthologous relationships between chromosome arms and the acrocentric or metacentric states of chromosomes in order to gain information regarding expectations for levels of homeolog differentiation within each species.

High-density linkage maps have been constructed for Lake Whitefish *Coregonus clupeaformis* (Gagnaire et al. 2013), Atlantic Salmon *S. salar* (Lien et al. 2011; Gonen et al. 2014) and members of *Oncorhynchus* including Rainbow Trout *O. mykiss* (Miller et al. 2012; Palti et al. 2015), Chinook Salmon *O. tshawytscha* (Brieuc et al. 2014), Coho salmon *O. kisutch* (Kodama et al. 2014), Pink Salmon *O. gorbuscha* (Limborg et al. 2014), Chum Salmon *O. keta* (Waples et al. 2016) and Sockeye Salmon *O. nerka* (Everett et al. 2012; Larson et al. 2015). No high-density maps exist for members of *Salvelinus*, but low-density microsatellite-based maps exist for Arctic Charr *S. alpinus* and Brook Charr *S. fontinalis* (Woram et al. 2004; Timusk et al. 2011), as well as a low-density (~300 marker) EST-derived SNP map for *S. fontinalis* (Sauvage

et al. 2012a). Genome assemblies exist for Rainbow Trout (Berthelot et al. 2014) and Atlantic Salmon (Lien et al. 2016). A genome assembly and low-density genetic map are also available for Northern Pike *Esox lucius*, a sister species to the salmonid WGD (Rondeau et al. 2014). With these resources available, it becomes especially valuable to integrate the information from all of the maps to detail the chromosomal evolution of the salmonids.

In this study, we use a mapping family previously used to generate a low-density EST-derived SNP linkage map (Sauvage et al. 2012a) to produce the first high-density RADseq map for the genus *Salvelinus*, the Brook Charr *S. fontinalis*. Brook Charr is a species of importance for conservation, aquaculture and fisheries, and an underrepresented lineage of Salmonidae in terms of genomic resource availability.

Further, we developed MapComp, a program to enable comparisons of genetic maps built from related species with or without the same RADseq protocol using an intermediate reference genome. MapComp follows earlier proposed approaches to integrate non-model maps with model species genomes (Sarropoulou et al. 2008). It identifies on average 5-fold more marker pairs between linkage maps than methods relying on identical markers only, and creates pairwise comparison plots for data visualization. MapComp enabled a detailed characterization of the orthologous and homeologous chromosome arms representing all main genera comprised within the salmonid family. This characterization enabled the identification of the most likely historical chromosomal rearrangements occurring at different levels of the salmonid phylogeny, including some potential inversion events. This comprehensive view provides new insight on the post-WGD chromosome evolution of Family Salmonidae.

Materials and Methods

Brook Charr genetic map

Animals

Full details regarding the experimental mapping family were reported previously (Sauvage et al. 2012a; 2012b). The F_0 individuals were from a domestic population used in Québec aquaculture for 100 years, supplied here from the Pisciculture de la Jacques-Cartier (Cap-Santé, Québec), and a wild anadromous population from Laval River (near Forestville, Québec) that have been kept in captivity for three generations at the Station aquicole de l'ISMER (Rimouski, Québec). Three biparental crosses of F_1 individuals produced three F_2 families, and the family with the largest number of surviving offspring was chosen to be the mapping family (n = 192 full-sib F_2 offspring).

DNA extraction, sampling preparation and sequencing

DNA was extracted from the fin of F_2 offspring and F_1 parents by high salt extraction (Aljanabi & Martinez 1997) with an additional RNase A digestion step (QIAGEN), as previously reported (Sauvage et al. 2012a). Quality of the extracted genomic DNA was quality validated by gel electrophoresis and quantified using Quant-iT PicoGreen double-stranded DNA Assay (Life Technologies) using a Fluoroskan Ascent FL fluorometer (Thermo LabSystems).

Double-digest RADseq (Baird et al. 2008) was performed as per methods previously outlined (Elshire et al. 2011) and described in full elsewhere (Poland et al. 2012). Briefly, two restriction enzymes were used (*Pst*I and *Msp*I) to digest genomic DNA. Digested DNA was then ligated with adapters and barcodes for individual identification then amplified by PCR. For the offspring, uniquely barcoded individuals were then combined in equimolar proportions into eight pools, each pool containing 25 individuals. Pools were each sequenced on a single lane on a

HiSeq2000 at Génome Québec Innovation Centre (McGill University, Montréal). In order to obtain deeper sequencing of the parents, each parent individual was sequenced using an Ion Torrent at the sequencing platform at IBIS (Institut de Biologie Intégrative et des Systèmes, Université Laval, Québec City). This platform change between F_1 and F_2 individuals occurred due to equipment availability, but extra precaution was taken to ensure proper correspondence of loci (*see below*).

Bioinformatic pipeline and reduced genome de novo *assembly*

Raw reads were inspected for overall quality and presence of adapters with fastqc (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Adapters were removed and raw reads were truncated to 80pb using CUTADAPT v.1.9 Dev.0 (Martin 2011). Reads were de-multiplexed, by barcodes, and quality trimmed to 80bp using the STACKS v.1.32 (Catchen et al. 2011; 2013) *process_radtags* module. The ploidy-informed empirical procedure was used (Ilut et al. 2014) to optimize *de novo* assembly. Sequence similarity was explored to find the optimum clustering threshold, which is highly important for pseudotetraploid salmonid *de novo* assembly (see Additional File S1 for pipeline parameters). Data from each individual were grouped into loci, and polymorphic nucleotide sites were identified with the *ustacks* module. The catalog construction used all loci identified across the parents. Differentially fixed loci (i.e. monomorphic loci among parents) were allowed to merge as a single locus when no mismatches were found (*cstacks*). Loci from parents and offspring were matched against the parental catalog to determine the allelic state at each locus in each individual in *sstacks*. To improve the quality of the *de novo* assemblies produced in STACKS and to reduce the risk of generating nonsensical loci with repetitive sequences and paralogs, we used the correction module *rxstacks*. The log-likelihood threshold for *rxstacks* was chosen based on the distribution of mean and median log-

likelihood values. After the correction module, the catalog and individuals' matches were rebuilt with the corrected individuals files. The *genotypes* module of STACKS was used to output markers along with their allelic state and raw genotypes. The markers were translated using the function *genotypes_summary.R* of stackr (v.0.2.1; Gosselin & Bernatchez, 2015) into fully- or semi-informative markers types, specifically the four types of markers that permitted in the outbreeding design: *ab* x *ac*, *ab* x *ab*, *ab* x *aa* and *aa* x *ab* (Wu et al. 2002).

Pre-mapping quality control

Several steps of quality control were performed based on recommendations (van Ooijen & Jansen 2013). Pre-mapping quality control consisted of excluding individuals with > 30% missing data (22 progeny), monomorphic loci and loci with an incomplete segregation pattern inferred from the parents (i.e. missing alleles) using the *genotypes_summary.R* function. This function was also used to filter errors in the phenotype observations of markers with a segregation distortion filter using a chi-square goodness-of-fit test (*filter.GOF*). With heterozygous parents, not all of the markers contribute equally to the construction of the map, because linkage phases change across loci (van Ooijen & Jansen 2013). Therefore, tolerance for genotyping errors (goodness-of-fit threshold: 12 to 20) and missing genotypes (50% to 90% thresholds) were also explored with *genotypes_summary.R*.

Linkage mapping and post-mapping quality control

The linkage map was first built in JOINMAP v4.1; (van Ooijen 2006) using the pseudo-testcross approach (Grattapaglia & Sederoff 1994; van Ooijen & Jansen 2013) that only uses the markers segregating in a uni-parental configuration (i.e. *ab* x *ac* and *ab* x *ab* markers are excluded at this step). Then a consensus map including the bi-parental configuration markers was produced with the joint data analyses as a CP population type (cross pollinator, or full-sib family) using the

multipoint maximum likelihood mapping algorithm for marker order (van Ooijen 2011; van Ooijen & Jansen 2013). The pseudo-testcross maps were used for confirmation. Separate maximum likelihood maps were generated for each parent, and only the female map was retained, as is typical for salmonid mapping studies (Kodama et al. 2014). Markers were grouped with the independent LOD option of JOINMAP with a range of 15 to 40 LOD, for the minimum and maximum threshold, respectively. A total of 42 linkage groups (LGs) were defined by evaluating stability of marker numbers over increasing consecutive LOD values. This number of LGs corresponds to the expected chromosome number of Brook Charr (2n = 84). During mapping, the stabilization criterion was monitored in the session log with the sum of recombination frequencies of adjacent segments and the mean number of recombination events. Default mapping parameters usually performed well with the smaller LG, but for larger LG the stabilization was not always reached, so more EM cycles and longer chains per cycle were used. For full details of parameters used in JOINMAP, see Additional File S1.

Problematic markers, unlinked markers and small linkage groups were inspected and tested by using several JOINMAP features, including *crosslink*, *genotype probabilities*, *fit* and *stress*. As recommended by van Ooijen and Jansen (2013), detection of errors in ordering and genotyping along with marker exclusion followed these criteria: (i) oversized LG, which can occur with high marker numbers; (ii) incidence of improbable genotypes e.g. double recombinants (Henning et al. 2014), also inspected using the *countXO* function of R/qtl (Broman et al. 2003); (iii) drastic changes of orders; and (iv) low levels of fit or high levels of stress. Maps were inspected for distortion before and after manual exclusion of markers. Mapping distances (cM) were calculated using the Haldane mapping function.

MapComp

Map comparison through intermediate reference genome

In order to compare the Brook Charr map to other salmonid maps, published linkage map datasets were collected (see the Code and Pipeline Availability section), including information on marker name, sequence, linkage group and cM position. Comparisons of linkage group orthology and order were investigated for Chinook Salmon (Brieuc et al. 2014), Pink Salmon (Limborg et al. 2014), Chum Salmon (Waples et al. 2016), Rainbow Trout (Miller et al. 2012; Palti et al. 2015), Sockeye Salmon (Everett et al. 2012; Limborg et al. 2015), Coho Salmon (Kodama et al. 2014), Atlantic Salmon (Lien et al. 2011), Lake Whitefish (Gagnaire et al. 2013), and the salmonid WGD sister outgroup Northern Pike (Rondeau et al. 2014) (Table 1).

The basic workflow of MapComp is shown in Figure 1. First, all marker sequences were combined into a single fasta file and mapped to a reference genome, here the Rainbow Trout scaffolds (http://www.genoscope.cns.fr/trout/data/ added 28-Apr-2014) (Berthelot et al. 2014) and the Atlantic Salmon assembly ICSASG_v2 (AGKD00000000.4; Lien et al. 2016) using *BWA mem* (Li & Durbin 2009). Matches to the reference were only retained when mapping quality score (MAPQ) was ≥ 10 and a single match was found in the target genome. When two markers (i.e. one from each species) mapped to the same reference genome scaffold or contig, the two closest markers were taken as a marker pair. Markers were paired without replacement (i.e. once the closest marker pair was selected, other markers also pairing with the marker that has now been paired were then discarded). Each marker pair was then added to an Oxford grid. The pipeline developed for MapComp is available at https://github.com/enormandeau/mapcomp/.

Table 1. Overview of compared species. The common and scientific name for each species in the analysis are displayed along with the source of the genetic map, the type of map and number of markers, the chromosome number for the species, and expected genome size (C-value and Gbp) obtained from (Gregory 2016).

Common and Scientific Name with References	**Map Type (Num. markers)**	**Number of Chr. (1n)**	**Exp. Genome Size (C-value)** (Gregory 2016)	**Exp. Genome Size (Gbp)**
Northern Pike* *Esox lucius* (Rondeau et al. 2014)	EST-based microsatellite (524)	25	0.85-1.40	0.8-1.4
Lake Whitefish *Coregonus clupeaformis* (Gagnaire et al. 2013)	RADseq with *Sbf*I (3438)	40	2.44-3.44	2.4-3.4
Atlantic Salmon *Salmo salar* (Lien et al. 2011)	EST-based SNP chip (5650)	29	2.98-3.27	2.8-3.2
Brook Charr *Salvelinus fontinalis*	RADseq with *Pst*I and *Msp*I (3826)	42	2.86-3.50	2.8-3.4
Rainbow Trout *O. mykiss* (Palti et al. 2015)	RADseq with *Sbf*I (955)	29	1.87-2.92	1.8-2.9
Coho Salmon *O. kisutch* (Kodama et al. 2014)	RADseq with *Sbf*I (5377)	30	2.60-3.05	2.5-3.0
Chinook Salmon *O. tshawytscha* (Brieuc et al. 2014)	RADseq with *Sbf*I (6352)	34	2.45-3.30	2.4-3.2
Pink Salmon *O. gorbuscha* (Limborg et al. 2014)	RADseq with *Sbf*I (7035)	26	2.23-2.57	2.2-2.5
Chum Salmon *O. keta* (Waples et al. 2016)	RADseq with *Sbf*I (6119)	37	2.49-2.76	2.4-2.7
Sockeye Salmon *Oncorhynchus nerka* (Larson et al. 2015)	RADseq with *Sbf*I (6262)	28**	2.77-3.04	2.7-3.0

*sister species to salmonid WGD; **two different LG09 in Sockeye, one was instead labeled as LG29 in Table 2.

To identify homeology relationships, where two chromosome arms originate from the same pre-duplicated chromosome arm, comparisons between Northern Pike *E. lucius* and the other salmonids were conducted. This required changing some parameters in MAPCOMP to allow for multiple hits from the non-duplicated Northern Pike map against the Atlantic Salmon and Rainbow Trout reference genome intermediates, as each marker could be present in at least duplicate in the salmonid genome. Specifically, the mapping quality threshold was lowered (MAPQ ≥ 2) and mapping against more than one locus in the reference genome was permitted.

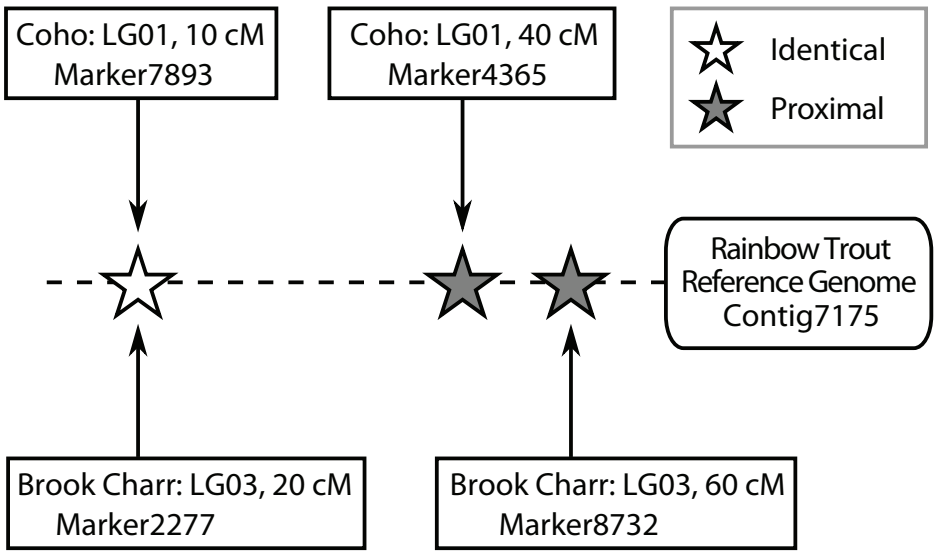

Fig. 1. Schematic of MapComp using a reference genome to pair markers. MapComp compares genetic maps from two different species by mapping marker sequences against a reference genome, then retaining high quality mappings that only hit against one locus in the genome. Markers from each species are paired if they hit against the same contig/scaffold by taking the closest two markers together as each pair. Each marker is paired without replacement, and so any other marker that was second closest to the now-paired marker is discarded. This method captures identical markers (white star in image) and non-identical markers (grey stars). Finally, the linkage group and cM position of each marker is plotted in an Oxford grid. Note that the marker names and contig ID in the schematic are for demonstration purposes only and do not reflect actual pairings.

Characterization of orthology and homeology between chromosome arms

Orthology of chromosome arms between Chinook Salmon and Coho Salmon maps identified previously using homologous markers (Kodama et al. 2014) was confirmed using MapComp. Chinook Salmon and Coho Salmon were then individually compared with Brook Charr to identify corresponding chromosome arms in Brook Charr. Once these orthology relationships were obtained, the Brook Charr map was compared with Pink Salmon, Sockeye Salmon, Chum Salmon, Rainbow Trout and Atlantic Salmon. Orthology was identified in Lake Whitefish using a consensus approach, where results from comparisons of Lake Whitefish with multiple different species were considered for unambiguous determination of orthology. Homeologs were also identified using a consensus approach, and the original Northern Pike linkage groups were given

.1 or .2 designations to represent the duplicated chromosomes. These results were compared to the results of Rondeau et al. (2014), in which BLAST was used with Atlantic Salmon linkage groups against the Northern Pike genome to identify salmonid WGD homeologs.

Identification of putative inversions

Plots from MAPCOMP were visually inspected for inversions. During linkage mapping, when markers do not fit in the linkage group, they can be placed at the distal ends of the linkage group (Henning et al. 2014). Therefore, to avoid the erroneous identification of inversions, evidence for inversions was only considered when non-inverted regions flanked the inverted region. As the analysis is based on linkage maps and not assembled genomes, all inversions were considered putative. Furthermore, phylogenetic relationships and inversion conservation across species were considered (i.e. when an inversion was identified within multiple species within a lineage). Centromere locations were obtained from Chinook Salmon (Brieuc et al. 2014) to allow the characterization of selected inversions as either pericentric (involving the centromere) or paracentric (not involving the centromere).

Conservation of rearrangements and identification of full coverage of linkage groups

The conservation of chromosomal rearrangements among the salmonids was analyzed by using the most taxonomically complete phylogeny of the salmonids (Crête-Lafrenière et al. 2012) but with Rainbow Trout as the outgroup to the other *Oncorhynchus* clades as reported previously (Kinnison & Hendry 2004), and with the still-debated clade containing Pink Salmon, Chum Salmon and Sockeye Salmon arranged in the most parsimonious phylogeny in terms of the numbers of required fusions/fissions. The analysis of metacentric conservation was based on the analysis of conservation in Coho, Chinook, Rainbow Trout and Atlantic Salmon (Kodama et al. 2014), but re-analyzed using MAPCOMP and additional maps in the present study (i.e. Pink

Salmon, Chum Salmon, Sockeye Salmon, Brook Charr and Lake Whitefish). To confirm that a metacentric chromosome was completely present, for conserved metacentric identification we required evidence from both sides of the centromere.

Results

Generation of a Brook Charr linkage map

On average, 10M single-end reads were obtained for each parent and 5M for each individual offspring. Using STACKS v1.32 (Catchen et al. 2011), 6264 segregating markers were identified, each containing one to five SNPs. Missing data per marker followed a heavy-tailed distribution, having a mode of 10 individuals genotypes missing for ~700 markers. Female, male and consensus genetic maps were generated, but the female-specific map (n = 3826 markers) was retained as the final map, as is typical for salmonids due to low recombination rate (e.g. Kodama et al. 2014; Brieuc et al. 2014) and increased residual tetrasomy in males (Allendorf et al. 2015).

A total of 42 linkage groups were characterized in the female map (Figure 2), corresponding to the expected haploid chromosome number for Brook Charr (Phillips & Ráb 2001). On average, metacentric linkage groups were 270 cM (range = 185-342 cM) containing 126 markers (range = 107-175 markers), whereas acrocentric linkage groups were 156 cM (range = 65-230 cM) containing 83 markers (range = 33-134). The total length of the female map was 7453.9 cM. Descriptive statistics for the linkage groups are in Additional File S2. This size is in the range of other high-density salmonid maps, such as the Coho Salmon linkage map (6596.7 cM) (Kodama et al. 2014), although is larger than the Chinook Salmon map (4164 cM) (Brieuc et al. 2014). The female map contains 3826 markers with the following marker types, as defined by Wu et al. (2002): 254 fully-informative (*ab* x *ac*), 954 semi-informative (*ab* x *ab*) and 2618 fully informative in female parent (*ab* x *aa*). The female map is in Additional File S3. The

consensus map contained an additional 2385 markers that were informative in the male parent (*aa* x *ab*) but although these markers were in the correct linkage groups, they did not position well within the linkage group. This is most likely due to the low recombination rate known to occur in male salmonids, as almost complete crossover interference can occur within male salmonids during meiosis (Naish et al. 2013).

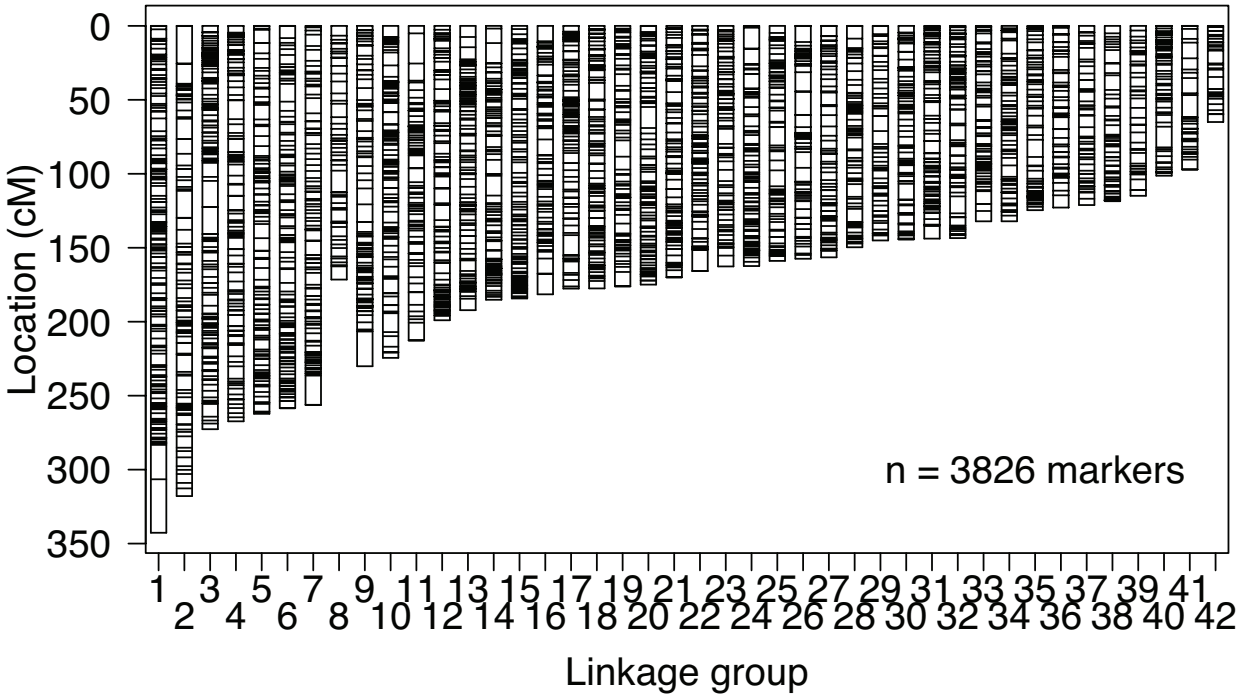

Fig. 2. Brook Charr *Salvelinus fontinalis* linkage map. Eight metacentric (LG1-8) and 34 acrocentric linkage groups (LG9-42) were identified in the female map. Horizontal lines within each linkage group are markers (total = 3826 markers).

Identification of orthologous chromosome arms among the salmonids

Assignment of linkage groups to chromosome arms has been performed using fluorescence *in situ* hybridization with BAC probes for Atlantic Salmon (Phillips et al. 2009), and orthology has been designated using homologous microsatellite and RADseq markers (using the same library preparation protocols) among Chinook Salmon, Coho Salmon, Rainbow Trout and Atlantic

Salmon (Danzmann et al. 2008; Phillips et al. 2009; Naish et al. 2013; Kodama et al. 2014; Brieuc et al. 2014) and recently Sockeye Salmon (Larson et al. 2015). A full comparison across all existing maps has yet to be completed. The low-density linkage map of the Northern Pike *E. lucius* has been compared with Atlantic Salmon (Rondeau et al. 2014), but not yet with the rest of the salmonids. Details on the linkage maps and species used in this analysis, including expected genome sizes (Gregory 2016) are provided in Table 1.

To begin orthology designation of linkage groups, Chinook Salmon and Coho Salmon linkage maps were used to compare with the map of Brook Charr using MapComp pairing markers through the Rainbow Trout genome (Berthelot et al. 2014) (see MapComp schematic in Figure 1, and Methods for full details). All chromosome arms (NF = 50) were identified unambiguously in Brook Charr (Figure 3; Table 2). The Brook Charr linkage map was then compared with linkage maps of Sockeye Salmon, Chum Salmon, Pink Salmon, Rainbow Trout and Atlantic Salmon (Table 2). In a few rare cases where orthology with Brook Charr was not obvious, species were also compared to Chinook Salmon or others to clearly indicate the corresponding chromosome arm. One chromosome arm in Rainbow Trout required the use of a second Rainbow Trout linkage map to unambiguously identify orthology (see Table 2; Miller et al. 2012). Most arms were also identified in the more distantly related Lake Whitefish, but seven arms remained unidentifiable.

Table 2. Orthologous chromosome arms across the salmonids. All orthologous relationships among salmonids and the pre-duplicated Northern Pike are displayed as identified by MapComp. Bold/grey-shaded species are those with studies that specifically tested for residual tetrasomy using duplicate markers or sequence similarity, and within these species, the bold/grey-shaded homeologs are those with evidence of residual tetrasomy in the original studies (see Table 1 for references). Note that Northern Pike only has 25 chromosomes, but here each chromosome is listed twice to accommodate the duplicate orthologs in the other species.

Northern Pike	Lake Whitefish	Atlantic Salmon	Brook Charr	Rainbow Trout	Coho Salmon	Chinook Salmon	Pink Salmon	Chum Salmon	Sockeye Salmon
1.1	Cclu28	Ssa20b	Sf25	Omy27	Co15b	Ots13q	Og13b	Ok18	On11a
1.2	Cclu35	Ssa09c	Sf38	Omy24	Co18a	Ots14q	Og19b	Ok01a	On14b
2.1	*Cclu04a?*	**Ssa26**	Sf06a	Omy06b	**Co03b**	**Ots04q**	**Og04b**	**Ok14b**	**On27a**
2.2	*Cclu04a?*	**Ssa11a**	Sf28	Omy26	**Co08b**	**Ots12q**	**Og05b**	**Ok02b**	**On28a**
3.1	Cclu25	Ssa14a	Sf22	Omy08b	Co30	Ots10q	Og22	Ok23	On29
3.2	Cclu26	Ssa03a	Sf11	Omy28	Co27	Ots28	Og08a	Ok15	On20b
4.1	Cclu16	Ssa09b	Sf33	Omy25a	Co15a	Ots08q	Og23a	Ok30a	On05a
4.2	Cclu29	Ssa05a	Sf07b	Omy14b	Co19b	Ots21	Og13a	Ok04b	On06b
5.1	Cclu05a	Ssa19b	Sf01a	Omy16a	Co20a	Ots24	Og11b	Ok03	On23b
5.2	Cclu15	Ssa28	Sf27	Omy20	Co25	Ots25	Og07a	Ok24	On16
6.1	*Cclu05b?*	Ssa01b	Sf01b	Omy23	**Co11a**	**Ots01q**	**Og06b**	**Ok02a**	**On28b**
6.2	*Cclu05b?*	Ssa18a	Sf36	Omy01b	**Co04b**	**Ots06q**	**Og08b**	**Ok36**	**On10a**
7.1	Cclu13	Ssa13b	Sf08b	Omy12a	Co06a	Ots09p	Og16b	Ok12	On22b
7.2	Cclu08	Ssa04b	Sf09	Omy10a	Co28	Ots30	Og02b	Ok11	On10b
8.1	Cclu36	Ssa23	Sf04a	Omy04a	Co10a	Ots01p	Og06a	Ok26	On17
8.2	Cclu06a	Ssa10a	Sf17	Omy05b	Co13a	Ots05q	Og15a	Ok01b	On24a
9.1	Cclu06b	**Ssa02b**	Sf42	Omy13a	**Co20b**	**Ots32**	**Og26a**	**Ok32a**	**On21a**
9.2	Cclu38b	**Ssa12a**	Sf03b	Omy17b	**Co01b**	**Ots02q**	**Og18a**	**Ok10b**	**On15b**
10.1	Cclu10	Ssa27	Sf23	Omy18b	Co17a	Ots13p	Og24a	Ok20	On04a
10.2	Cclu24a	Ssa14b	Sf34	Omy14a	Co14b	Ots31	Og09b	Ok19a	On07a
11.1	Cclu18	**Ssa06a**	Sf14	Omy13b	**Co10b**	**Ots27**	Og26b	**Ok32b**	**On21b**
11.2	*missing*	**Ssa03b**	Sf08a	Omy12b^	**Co06b**	**Ots09q**	Og16a	**Ok05**	**On26**
12.1	Cclu27	Ssa13a	Sf18	Omy16b	Co24	Ots22	Og10a	Ok27	On20a
12.2	Cclu14	Ssa15b	Sf30	Omy09b	Co17b	Ots16q	Og03a	Ok28a	On07b
13.1	Cclu34	Ssa24	Sf06b	Omy06a	Co03a	Ots04p	Og04a	Ok14a	On27b
13.2	Cclu37	Ssa20a	Sf40	Omy11a	Co08a	Ots12p	Og05a	Ok25	On13b
14.1	Cclu04b	Ssa01c	Sf13	Omy05a	Co23	Ots20	Og11a	Ok09	On04b
14.2	Cclu33	Ssa11b	Sf10	Omy29^	Co29	Ots33	Og20	Ok06	On19
15.1	Cclu31	Ssa09a	Sf35	Omy25b	Co14a	Ots08p	Og10b	Ok35a	On12a
15.2	Cclu22	Ssa01a	Sf12	Omy19b	Co07b	Ots11q	Og17a	Ok29a	On18a
16.1	Cclu02+03	Ssa21	Sf26	Omy22	Co26	Ots26	Og01b	Ok07	On01
16.2	Cclu32	Ssa25	Sf24	Omy03b	Co02b	Ots03q	Og09a	Ok34a	On03a
17.1	Cclu38	Ssa12b	Sf03a	Omy17a	Co01a	Ots02p	Og18b	Ok10a	On15a
17.2	Cclu21	Ssa22	Sf21	Omy07b	Co05b	Ots07q	Og03b	Ok21	On02b

18.1	Cclu40	Ssa15a	Sf19	Omy08a	Co12a	Ots05p	Og12a	Ok19b	On24b
18.2	Cclu17	Ssa06b	Sf31	Omy04b	Co21	Ots18	Og14a	Ok08	On13a
19.1	Cclu30	Ssa10b	Sf15	Omy02b	Co22	Ots19	Og01a	Ok17	On09
19.2	Cclu11	Ssa16a	Sf20	Omy01a	Co04a	Ots06p	Og02a	Ok22	On25
20.1	*Cclu01a^?*	**Ssa05b**	Sf07a	Omy02a	**Co13b**	**Ots23**	Og15b	**Ok31**	**On14a**
20.2	*Cclu01a^?*	**Ssa02a**	Sf29	Omy03a	**Co02a**	**Ots03p**	Og19a	**Ok34b**	**On03b**
21.1	Cclu12	Ssa29	Sf05b	Omy15a	Co11b	Ots29	Og07b	Ok04a	On06a
21.2	Cclu39	Ssa19a	Sf16	Omy11b	Co18b	Ots16p	Og12b	Ok28b	On22a
22.1	*Cclu19?*	**Ssa17a**	Sf39	Omy07a	**Co05a**	**Ots07p**	Og21b	**Ok37**	**On02a**
22.2	*Cclu19?*	**Ssa16b**	Sf05a	Omy18a	**Co16b**	**Ots14p**	Og23b	**Ok30b**	**On05b**
23.1	Cclu02^	**Ssa07b**	Sf02b	Omy21a	**Co09a**	**Ots15p**	Og25a	**Ok13b**	**On08b**
23.2	Cclu01b^	**Ssa17b**	Sf37	Omy15b	**Co19a**	**Ots17**	Og14b	**Ok33**	**On23a**
24.1	Cclu24b	Ssa07a	Sf02a	Omy21b	Co09b	Ots15q	Og25b	Ok13a	On08a
24.2	Cclu23	Ssa18b	Sf32	Omy09a	Co16a	Ots10p	Og21a	Ok35b	On12b
25.1	*missing*	**Ssa04a**	Sf04b	Omy10b	**Co12b**	**Ots34**	Og24b	**Ok16**	**On11b**
25.2	*missing*	**Ssa08**	Sf41	Omy19a	**Co07a**	**Ots11p**	Og17b	**Ok29b^**	**On18b**

Notation: ^ = evidence for orthology is weak; italics in Lake Whitefish indicate problems resolving orthology.

In Brook Charr, a total of eight metacentric and 34 acrocentric chromosomes were expected from salmonid cytogenetics (Phillips & Ráb 2001; Lee & Wright 1981) and all were identified here (Table 2), increasing the resolution of the Brook Charr linkage maps from existing microsatellite and low density SNP-based linkage maps (Sauvage et al. 2012a; Timusk et al. 2011). Since Brook Charr has the fewest metacentric chromosomes of the species characterized here, often two acrocentric chromosomes in Brook Charr correspond to two fused arms on the same metacentric chromosome in another species. In some cases, due to tandem chromosome fusions observed in Atlantic Salmon (Phillips et al. 2009) three linkage groups in Brook Charr correspond to one linkage group in Atlantic Salmon. For example, Brook Charr Sf33, Sf35, Sf38 are in tandem fusions in Atlantic Salmon Ssa09. We compared orthology identified with MapComp between Coho Salmon and Chinook Salmon with an analyses that used homologous markers (Kodama et al. 2014), and found the same correspondence. We identified discordant results for five putative orthologous chromosomes between Chinook Salmon and Atlantic Salmon, and for two putative orthologous chromosomes between Chinook and Rainbow

Trout that were based on earlier studies. The rest of the results among these species corresponded between the studies (total = 50 orthologous relationships in four species). Additionally, orthology was determined for Pink Salmon, Chum Salmon, Sockeye Salmon and Lake Whitefish.

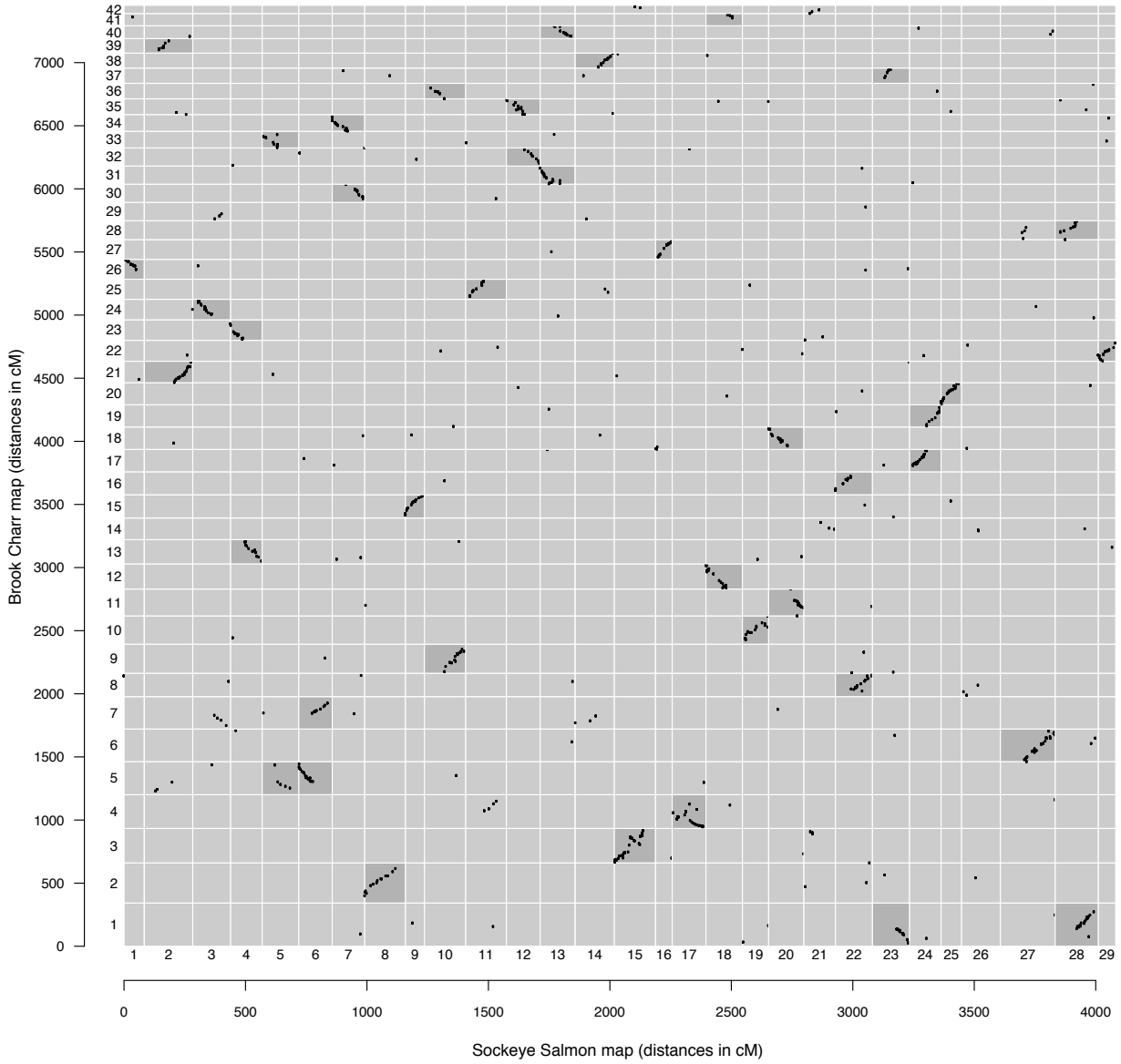

Fig. 3. MAPCOMP determination of orthologous chromosome arms. Brook Charr compared with Sockeye Salmon with markers paired through the Rainbow Trout genome identifies orthology between chromosome arms. A putative inversion can be seen between Brook Charr LG03 and Sockeye Salmon LG15.

Homeologous chromosome identification

To identify homeologous chromosomes (i.e. chromosome arms originating from the same pre-duplicated chromosome), the genetic map of Northern Pike was compared with the maps of all species using MapComp as described in the Methods. All homeologous pairs in Atlantic Salmon identified by MapComp using the Rainbow Trout intermediate reference genome were concurrent with those originally identified (Rondeau et al. 2014), but here were also extended to all other species (Table 2).

Homeologous chromosomes that undergo residual tetrasomy can be identified by mapping isoloci (i.e. markers recombining between homeologous chromosomes) in haploid crosses, as has been done for Atlantic Salmon, Coho Salmon, Chinook Salmon, Sockeye Salmon and Chum Salmon (see Table 1 for references). Sequence similarity between homeologous chromosomes has also been used to identify residual tetrasomic pairs in the analysis of the Atlantic Salmon genome (Lien et al. 2016). As can be seen from the orthologous chromosomes in Table 2, the same eight pairs of homeologous chromosomes are residually tetrasomic in the *Oncorhynchus* species, and seven of these are still residually tetrasomic in Atlantic Salmon. Without a haploid cross for our Brook Charr map, here we cannot specify whether any homeologs still undergo residual tetrasomy in this species. The homeologs identified using Northern Pike with MapComp were compared with those identified using isoloci in mapping crosses. All eight pairs of homeologs identified in Chinook Salmon (Brieuc et al. 2014) were confirmed as originating from the same ancestral chromosome. The eight pairs of homeologs identified in Coho Salmon (Kodama et al. 2014) and Sockeye Salmon (Larson et al. 2015) were also confirmed here. In addition, all other homeologies (total = 25 pairs) in all evaluated species were also identified using MapComp, with the exception of the aforementioned unidentifiable chromosome arms of Lake Whitefish (Table 2). Interestingly, all of the missing orthology

designations in Lake Whitefish are, without exception, those with residual tetrasomy in *Oncorhynchus* (Table 2).

Conserved and species-specific chromosome rearrangements

Shared rearrangements among species in a clade (e.g. fusion events) are likely to have occurred prior to the diversification of the clade, as demonstrated for nine metacentric fusions in Coho Salmon and Chinook Salmon (Kodama et al. 2014). Here, chromosome arm orthology designation furthered this analysis, allowing the inclusion of Brook Charr and Lake Whitefish as well as the clade containing Sockeye Salmon, Pink Salmon and Chum Salmon within *Oncorhynchus.* We identified 16 different fusion events conserved in at least two species, five fission events conserved in at least two species, 87 species-specific fusion events, and five species-specific fission events (Figure 4). For clarity, when discussing fusions and fissions here we use chromosome names from the Northern Pike chromosomes to refer to chromosome arms, including the duplicate designation (e.g. 1.1/1.2 or 3.1/3.2), as shown in Table 2. The phylogeny in Figure 4 was adapted from previous literature (Stearley & Smith 1993; Kinnison & Hendry 2004; Crête-Lafrenière et al. 2012). For the clade containing Pink, Sockeye and Chum Salmon, in which the sister relationships remain unclear (Kinnison & Hendry 2004), we present the most parsimonious phylogenetic relationship in terms of required number of fusion/fission events. With Pink Salmon as the sister species to the Chum and Sockeye Salmon clade, this requires three fewer fission or fusion events, and one fewer fission of a conserved metacentric chromosome, the 7.1-11.2 fusion.

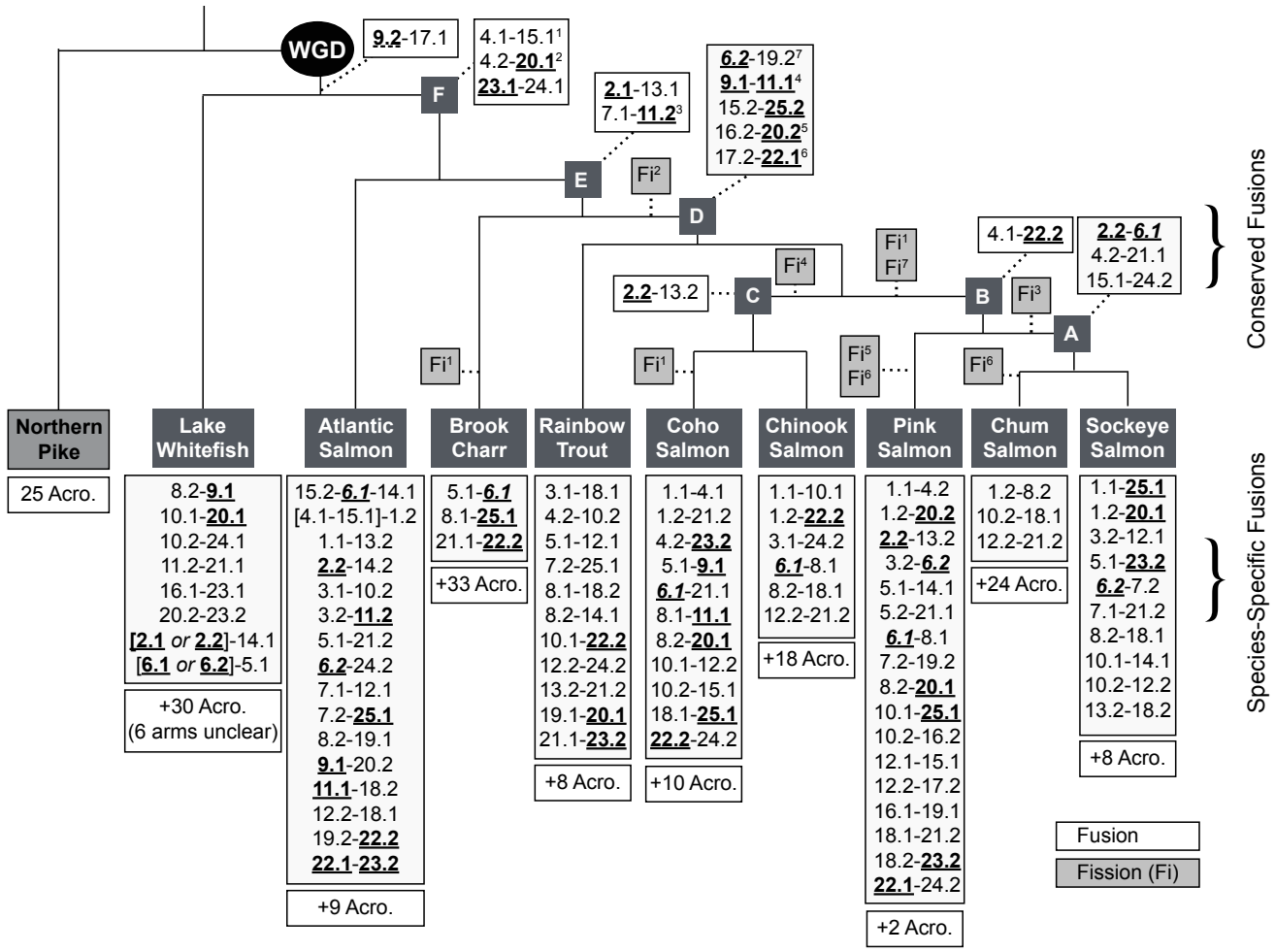

Fig. 4. Fusions and fissions across the salmonid lineage. Different fusions and fissions have occurred during the evolution of the salmonids. White boxes display the fusion events, where the orthologous chromosomes for all species are named according to the corresponding Northern Pike linkage group ID, with .1 or .2 to correspond with the post-duplicated salmonid chromosomes. Bold and underlined chromosome numbers are the homeologous pairs that still undergo residual tetrasomy, and the italicized chromosome (6.1/6.2) is not residually tetrasomic in Atlantic Salmon but is in the Pacific salmonids. Above the species names are conserved fusions, whereas below are the species-specific fusions. Also shown are fissions in light grey with the notation (Fi^x), where *x* corresponds to the superscript in the original fusion (e.g. 4.2-20.1 at point (F) in the phylogeny is divided at Fi^2 prior to point (D)). The phylogeny is adapted from (Stearley & Smith 1993; Kinnison & Hendry 2004; Crête-Lafrenière et al. 2012), with minor modifications to the relationships within the Pink, Chum and Sockeye Salmon clade, as described in the results. Branch lengths are not to scale and are for illustrative purposes of relationships between species only.

The oldest identified rearrangement is the 9.2-17.1 fusion event that is conserved in all species investigated (see Figure 4). Conserved in all species except Lake Whitefish is the 23.1-24.1 fusion (see F in Figure 4). Another metacentric fusion event at this same point in the phylogeny was also identified (4.2-20.1) that is still present in both Atlantic Salmon and Brook Charr, but not in any members of *Oncorhynchus*, suggesting that the a fission occurred prior to the speciation of any members of *Oncorhynchus* (Fi^2 in Figure 4). One fusion (2.1-13.1) is present in Brook Charr and all *Oncorhynchus* spp. (fused at E in Figure 4). Another fusion at this same point (7.1-11.2; E in Figure 4) is found in all descendants except for the Chum and Sockeye Salmon clade (fission Fi^3 in Figure 4).

More recent rearrangements include five fusions prior to the speciation of the *Oncorhynchus* clade (D in Figure 4). One is present in all *Oncorhynchus* species (15.2-25.2), another is present in all *Oncorhynchus* species except Pink Salmon (16.2-20.2; see Fi^5 in Figure 4), another underwent fission in the Chinook/Coho lineage (9.1-11.1; see Fi^4) and one underwent fission in Pink and Chum Salmon (17.2-22.1; see Fi^6). Conserved fusions were found within the *Oncorhynchus* lineage as well, including one fusion in the Chinook/Coho lineage, (2.2-13.2; C in Figure 4) and three fusions prior in the Sockeye/Chum lineage (A in Figure 4). Each species also has had species-specific fusions, ranging in number from only three fusions in Chum Salmon and three in Brook Charr to up to 17 in Pink Salmon and 18 in Atlantic Salmon (Figure 4).

Some rearrangements are more complex and thus it is more difficult to unambiguously describe their history. For example, a three chromosome arm fusion in Atlantic Salmon occurred through a single fusion (4.1-15.1) that either fused once prior to the divergence of Atlantic Salmon and underwent three different fission events (Fi^1 in Figure 4), or fused three independent times with the same fusion partner. It is not clear which of these possibilities is correct, but in

Figure 4 we display the first and more parsimonious scenario. In Atlantic Salmon, after the metacentric fusion, an additional fusion occurred, adding a third chromosome arm (1.2 with [15.1-4.1]). Three other different fusions appeared to have occurred two independent times: 8.2-18.1 in Chinook and Sockeye; 12.2-21.2 in Chinook and Chum; and 7.2-25.1 in Atlantic Salmon and Rainbow Trout. For each of these multiple independent origins, the alternate explanations are possible but less parsimonious. Although these few independent origin cases are not entirely clear, we display the most parsimonious rearrangements, requiring the fewest independent fusions/fissions in Figure 4.

Putative lineage-specific inversions

Several inversions flanked by non-inverted regions were revealed between linkage maps, suggesting the presence of chromosomal segment inversions (Figure 5). These putative inversions are more supported when phylogenetically conserved. Future genome assemblies for the species involved will be valuable for further inversion identification. A striking putative inversion was identified in one of the metacentric chromosomes conserved across all evaluated salmonids (9.2-17.1; Figure 4). An inversion near the center of the linkage group is present in only the Pink, Chum and Sockeye Salmon lineage. This inversion is visible in the Oxford grids between these species and Coho, Chinook Salmon and Brook Charr (see Figure 5a-b). As a result, the conformation observed in Pink, Chum and Sockeye Salmon is likely the derived form. Rainbow Trout does not indicate the inversion against the ancestral conformation, but also does not indicate the inversion with Pink, Chum and Sockeye Salmon, as there is a gap with no marker pairs available at the inverted locus in the Rainbow Trout linkage group.

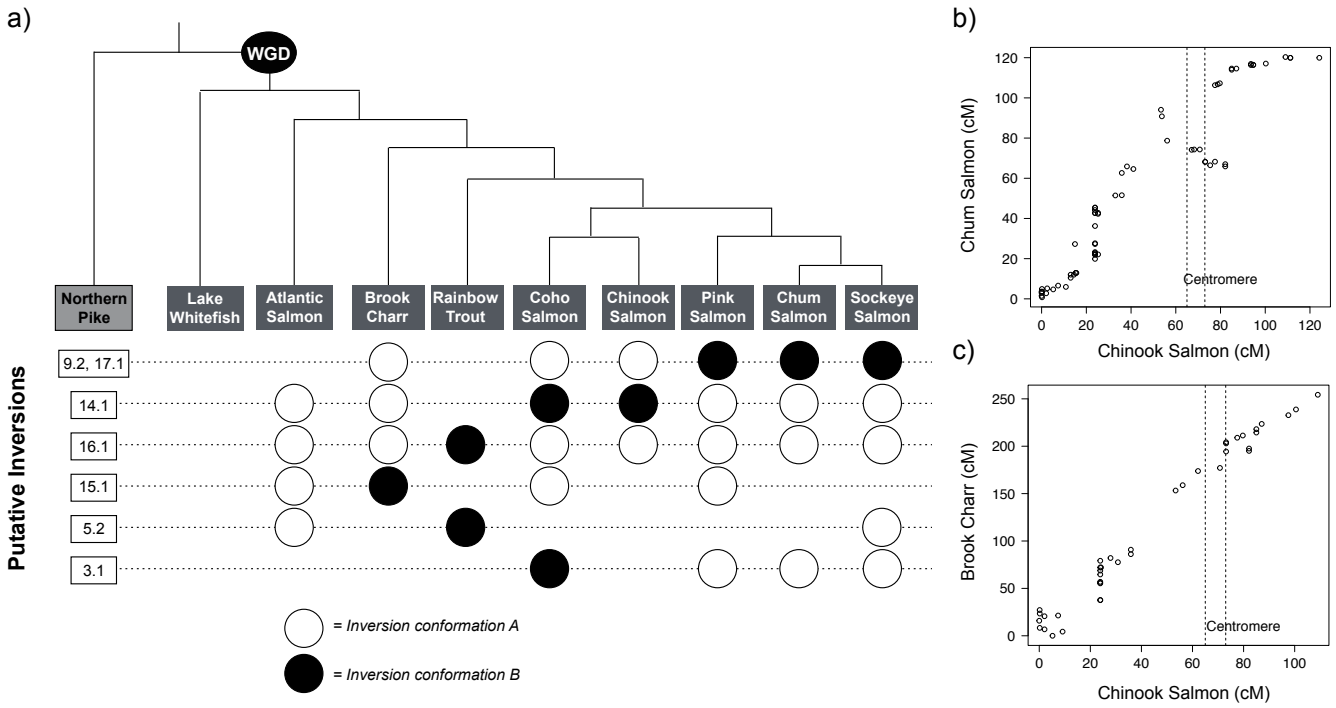

Fig. 5. Putative conserved and species-specific inversions. a) The salmonid phylogeny is shown highlighting six different inversion events, each listed according to the Northern Pike chromosome and represented on a dotted line below the phylogeny. Per line (inversion event), white circles indicate the more common or likely ancestral inversion conformation, and black circles the less common and more likely derived inversion conformation. If no circle is present, the inversion was not visible in the linkage map of that species. The putative pericentric inversion across the fusion 9.2-17.1 is displayed in (b) showing two species (Chum and Chinook Salmon) with different inversion conformations, and in (c) for two species with the same conformation (Brook Charr and Chinook Salmon). Predicted centromere positions previously identified in Chinook Salmon (Brieuc et al. 2014) are also shown in (b-c). Full names for species are defined in Table 1, the phylogeny is as described in the Results and Figure 4, and probable genes within the 9.2-17.1 inversion are shown in Additional file S4.

To further characterize the 9.2-17.1 inversion, centromere locations obtained from Chinook Salmon (Brieuc et al. 2014) were compared to the location of the inversion. The inverted region corresponds to the Chinook Salmon linkage group Ots02 between ~49-82 cM (see Figure 5b) and the centromere for this linkage group was estimated to be between 65-73 cM. This inversion therefore most likely contains the centromere (i.e. a pericentric inversion). This inverted region between Chinook Salmon and Chum Salmon was visible when using either the

Rainbow Trout or Atlantic Salmon genome as the intermediate reference genome, and had evidence from 12 markers mapping through two different scaffolds. To identify the genes that may be contained within this inverted region, the mapped locations on the Atlantic Salmon genome of the markers at the distal ends of the inverted region were taken. As the Atlantic Salmon genome has been annotated (Lien et al. 2016), this region of the genome (Ssa12 between 37,048,324-44,754,074 bp) was inspected for gene content. This region (~7.7 Mb) putatively contains 11 genes (based on alignment evidence, here we do not include predicted genes), including *cytokine-like protein 1*, *solute carrier family 2, facilitated glucose transporter member 9* and *cd8 beta*, among others (see Additional File S4). The exact genes found in this region in the species with the derived conformation of the inversion will require more genomes to be available before exploring further, including the actual breakpoints of the inversion and whether these occur within coding genes. In summary, 9.2 fused with 17.1 in the ancestor of all salmonids investigated here, then an inversion of a segment ~7.7 Mb containing coding sequences occurred across the centromere specifically in the Pink, Chum and Sockeye Salmon lineage. Other inversions were also visible (Figure 5a; Additional File S5). More information on these and potentially new inversions will be obtained as more assembled genomes become available.

Benefits of MapComp *versus direct marker comparison and effect of intermediate reference genome*

Linkage group orthology between species are typically identified by finding homologous markers using reciprocal best-hit BLAST (Kodama et al. 2014). The method implemented in MapComp, where we accept both identical and proximal markers, leads to a far greater number of retained marker pairs (on average 5-fold; Table 3). For example, between Brook Charr and

Chinook Salmon, 907 marker pairs were identified using MAPCOMP, whereas direct mapping identified 190 pairs.

MAPCOMP was tested on both the Rainbow Trout and the Atlantic Salmon genome as intermediate references for pairing markers between maps (Table 3). Results using either genome were highly concordant, having only minor differences in the number of markers mapped and paired. Although the Rainbow Trout genome as an intermediate provided slightly more mapped markers (on average 1.2-fold more than Atlantic Salmon), the Atlantic Salmon genome provided more marker pairs (on average 1.4-fold). This could be due to slight differences in contiguity of the two genomes. As the Atlantic Salmon linkage map sequence information was obtained in EST format, the number of markers mapping from the map to the genome was lower than expected relative to the shorter reads, probably due to the longer sequences and because BWA mem is not a splice-aware aligner.

Table 3. MAPCOMP results using two different intermediate genomes and comparison with results from reciprocal BLAST. The number and percentage of markers from each species that map to each genome are shown, along with the number of markers pairs between each species and Brook Charr identified by MAPCOMP. Also shown is the number of homologous markers that would have been found between each species and Brook Charr with a reciprocal BLAST approach. Numbers of mappings and pairs were similar when tested on the Rainbow Trout or Atlantic Salmon genome assemblies. *N/A* values are present as Brook Charr is not paired against itself.

Species	Total Markers	Map to Rainbow Trout genome: No. (%)	No. Marker Pairs	Map to Atlantic Salmon genome: No. (%)	No. Marker Pairs	Recip. best-hit BLAST
Lake Whitefish	3438	1156 (33%)	346	1185 (34%)	609	111
Atlantic Salmon	5650*	2776 (49%)	619	3434 (60%)	1041	208
Brook Charr	3826	2321 (60%)	N/A	2454 (64%)	N/A	N/A
Rainbow Trout	955	837 (87%)	300	626 (65%)	411	30
Coho Salmon	5377	3873 (72%)	813	2856 (53%)	1068	182
Chinook Salmon	6352	4663 (73%)	907	3472 (54%)	1162	190
Pink Salmon	7035	4544 (65%)	841	3481 (49%)	1129	210
Chum Salmon	6119	4150 (67%)	795	3139 (51%)	1049	205
Sockeye Salmon	6262	4034 (64%)	771	3138 (50%)	1061	209

*from EST sequences

MapComp parameters can be adjusted for the maximum distance allowed between paired markers. Here we used a maximum distance of 10 Mbp, but most paired markers were at a much smaller distance than this (see marker distance distribution examples in Additional File S6). With a greater phylogenetic distance between species, fewer identical markers were found. For example, the number of identical markers for Chinook and Coho Salmon are high, but the Chinook Salmon comparison with Brook Charr depended more on non-identical markers, as did the comparison between Lake Whitefish and Brook Charr. These parameters can be easily tested by the user, allowing the identification of the optimal settings for individual datasets to permit the greatest number of comparisons without increasing noise in the Oxford grids. Here we found that 10 Mbp provided more markers without a substantial increase in noise.

Discussion

Linkage maps have many applications, including QTL analysis, assisting genome assembly and comparative genomics. With advances in sequencing technology and techniques (Baird et al. 2008), high quality and dense linkage maps are increasingly available for many species, including non-model species. Dense linkage maps are highly useful for anchoring genome scaffolds to chromosomes (Ming & Man Wai 2015) or for comparative genomics, allowing for information transfer from model to related non-model organisms (Naish et al. 2013). They are also useful for cross-species QTL comparisons (Larson et al. 2015) to understand genome function and evolution, such as that after a whole genome duplication (Kodama et al. 2014).

Salmonids are a valuable taxon for studying genome duplication. Recently, Kodama et al. (2014) characterized several chromosomal fusions and positioned them in the salmonid phylogeny based on the conservation of the fusion across the investigated lineages. This has indicated that structural rearrangements have occurred throughout the evolution of the salmonids,

with rearrangements retained from different points in evolutionary history. Here, we further demonstrate the diversity of these rearrangements by identifying all orthologous arms in additional species and genera, and evaluate the most likely timings of rearrangements throughout salmonid evolutionary history.

Residual tetrasomy, or recombination between homeologous chromosomes, occurs through tetravalent formation at meiosis and probably requires that at least one of the homeologous chromosome arms involved is in a metacentric fusion (Wright et al. 1983; Allendorf et al. 2015). As evidence of this, fewer tetravalents form at meiosis when fewer metacentrics chromosomes are present in a species. For example, fewer tetravalents have been observed in comparisons between Rainbow Trout (1n = 20 metacentrics) and Brown Trout *Salmo trutta* (1n = 10 metacentrics) (May & Delany 2015). This is important to consider for salmonid genome assembly. If a homeologous chromosome pair that is known to be residually tetrasomic in other salmonid species (see Table 2) occurs as two acrocentric chromosomes in another species, this pair may not undergo residual tetrasomy and therefore may have increased sequence divergence between homeologs. Therefore, species with few metacentric chromosomes may offer additional information regarding rediploidization. Even though Brook Charr has fewer metacentrics than other salmonids, all of the known residually tetrasomic homeolog pairs from the other salmonids have one homeolog present in a metacentric chromosome in Brook Charr (Table 2; n = 8 metacentrics).

As metacentric formation is thought to be important for residual tetrasomy, the timing of fusion events may provide additional insight into the rediploidization process in salmonids. From the present study, it is interesting to note that many of the residually tetrasomic pairs have at least one homeolog involved in an ancient conserved fusion (Figure 4). The second homeolog varies

more in its fusion partner across the lineage, or can be present as an acrocentric. For example, 9.2 of the 9.1/9.2 residually tetrasomic homeolog pair is fused with 17.1 in all assessed species (fused at F in Figure 4). In contrast, 9.1 varies more in its binding partner and sometimes is acrocentric in extant salmonids. Similarly, the residually tetrasomic 23.1 is fused with 24.1 in all assessed species except Lake Whitefish (10.2-24.1), whereas 23.2 is more variable and occasionally acrocentric. These ancient fusions may be informative about mechanisms that have prevented rediploidization in salmonids.

The fusion history of the other residually tetrasomic pairs are not as simple as the above two examples. Within a residually tetrasomic homeolog pair, it is not always the same homeolog in a metacentric fusion across species. This agrees with previous indications that only one of the homeologs must be bound in a metacentric to prevent rediploidization. For example, in Atlantic Salmon, 2.2 is metacentric and 2.1 is acrocentric, whereas in Brook Charr and all *Oncorhynchus* spp., 2.1 is the conserved metacentric (fused at E in Figure 4). Another example of this differing metacentric binding occurs for 20.1/20.2, where in Atlantic Salmon and Brook Charr 20.1 is in a conserved metacentric fusion and 20.2 is acrocentric, whereas in all *Oncorhynchus* spp. 20.2-16.2 is the conserved fusion. Therefore, even though these metacentrics may be required to retain residual tetrasomy, the homeolog bound in the metacentric chromosome can differ among the species. Further characterization of this will be facilitated with the production of high-density linkage maps for more species from salmonid genera outside of *Oncorhynchus* that are represented by only one species (e.g. *Salvelinus*, *Coregonus*, *Salmo*) or none (e.g. *Thymallus*).

Although rediploidization of the salmonids may have generally occurred prior to the salmonid radiation (Lien et al. 2016), the rediploidization process has apparently continued since the speciation of Atlantic Salmon, at least in one homeologous pair (i.e. 6.1/6.2). This pair has

rediploidized in *Salmo salar* as demonstrated by sequence similarity between homeologous chromosomes (see 18qa-1qa in Figure 3b in Lien et al. 2016) as well as suggested by the lack of identifiable isoloci in this pair (Lien et al. 2011). Conversely, in *Oncorhynchus*, this homeologous pair is residually tetrasomic as demonstrated by isoloci in Coho, Chinook , Chum and Sockeye Salmon (see Table 1 for references). This difference in rediploidization between Atlantic Salmon and the Pacific salmonids is shown from the orthologous relationships of chromosome arms characterized in Table 2.

It is interesting to note that one of the homeologous chromosome arms, 6.1, is fused in the center of one of the triple chromosome arm fusions specifically in Atlantic Salmon (Figure 4). Although it is known that residual tetrasomy requires at least one of the two homeologous chromosomes to be in a metacentric fusion, the effect of being in the middle of a triple chromosome arm fusion on rediploidization is not known. It is possible that this position could hinder homeologous pairing at meiosis. Regardless of this mechanism, this result indicates that the path to rediploidization differs for this chromosome pair between Atlantic Salmon and the evaluated Pacific salmonids. Information regarding residual tetrasomy (e.g. through haploid crosses) in additional maps from members of *Coregonus*, *Salvelinus*, *Thymallus* or other salmonid genera will be valuable to understand this process further in a broader range of genera.

Fusions and inversions

Chromosomal rearrangements include chromosome fusions or fissions, region amplifications or deletions, segment inversions or non-homologous chromosome segment translocations (Rieseberg 2001). The characterization of the fusion events across all published salmonid maps (Figure 4) provides a new resolution of the exact identities of chromosome arms in the pre-duplicated genome that have fused together at different moments during the salmonid

diversification. This demonstrates the stepwise process of generating the extant salmonid karyotypes, with fusions occurring at each step along the diversification process. Notably, for most salmonid species, most fusions are not ancestrally conserved, but rather occur individually within each species (Figure 4). It remains unclear why some species retain their high number of acrocentric chromosomes (e.g. Brook Charr, three species-specific fusions), whereas others do not (e.g. Pink Salmon, 17 species-specific fusions). Furthermore, this variation in numbers of species-specific fusions can occur between closely related species (e.g. Chum and Sockeye Salmon).

Inversions can occur when a segment of a chromosome is cut out by two breakpoints and then reinserted in the opposite orientation (Kirkpatrick 2010). Effects of inversions on fitness are highly unpredictable and vary across taxa. In general, they tend to reduce recombination rates at the site of the inversion, potentially playing an important role in speciation and local adaptation (Rieseberg 2001; Kirkpatrick 2010; Noor et al. 2001a; 2001b). For example, two inversions reduce recombination and maintain genetic differentiation between migratory and stationary ecotypes of Atlantic Cod (*Gadus morhua*), preserving the co-occurrence of adaptive alleles within the migratory form (Kirubakaran et al. 2016) Additionally, lower recombination rates were observed in heterokaryotypic regions of Yellowstone Cutthroat Trout (*O. clarkii*) and Rainbow Trout hybrids compared to collinear regions (Ostberg et al. 2013). Recombination suppression may allow for conservation of fitness-related gene complexes that are locally adapted, or involved in reproductive isolation (Ostberg et al. 2013). Robertsonian rearrangements (e.g. fusions and fissions) have less of an effect on recombination rates than do rearrangements affecting synteny (e.g. inversions) (Rieseberg 2001; Ostberg et al. 2013).

MapComp: *Potential and limitations*

By using the information from both identical and proximal marker pairing, MapComp solves the issue of low marker homology between reduced representation sequencing (RADseq)-based linkage maps generated with different protocols or restriction enzymes, or from relatively more distantly related species. Synteny is still required in order to pair proximal markers through the intermediate reference genome. Previously, polymorphic microsatellite markers highly conserved among salmonids have enabled exploration of salmonid chromosomal evolution by integrating across species and genera (Naish et al. 2013). Although RADseq-based linkage maps routinely provide an order of magnitude more markers than microsatellite maps with less effort, identical markers are not always abundant between species. Low marker homology among species has also hindered cross-species comparisons when using microsatellite-based genetic maps, for example when Coho Salmon was compared with Sockeye Salmon and Pink Salmon (Naish et al. 2013). As such, with the generation of additional high-density maps for the salmonids, the use of MapComp will continue to be highly useful in characterizing these relationships.

At its core, MapComp is similar to the approach used by Sarropoulou et al. (2008), in which EST-based markers from two species were aligned to a reference genome of a third species, to identify orthologous linkage groups. However, this earlier approach did not retain marker positions from original maps for plotting within an Oxford grid, and only provided the total number of markers found to correspond for each linkage group pair. Other cross-species map comparison approaches exist, for example cMAP (Fang et al. 2003), although these often require shared markers between maps. Another similar approach was used by Amores et al. (2011) for Spotted Gar *Lepisosteus oculatus*, where paired-end sequencing was performed on a single-digest then random shear library. The authors therefore obtained a larger amount of

sequence near their marker allowing them to identify genes near the marker. Then the order of the identified genes was used to compare synteny of orthologs in assembled genomes such as humans *Homo sapiens* or Zebrafish *Danio rerio*. In contrast, MapComp works without prior knowledge of specific gene orthology, providing map comparisons at a much higher marker density without being restricted to coding regions. Another recent approach compared a linkage map for the European tree frog *Hyla arborea* with the genome of the western clawed frog *Xenopus tropicalis* and identified many syntenic regions (Brelsford et al. 2015). A recent approach in salmonids used a RADseq high-density linkage map for Chinook Salmon with the Atlantic Salmon reference genome to anchor Atlantic Salmon scaffolds to the Chinook Salmon linkage map when enough markers were present and the order was as expected (McKinney et al. 2015). Orthology relationships between Chinook Salmon and Atlantic Salmon have been characterized previously (Brieuc et al. 2014), and so the aligned scaffolds could then be classified as homologous, homeologous, or unsupported to further improve the anchoring of scaffolds onto the linkage map, and to identify potential genes around loci of interest (McKinney et al. 2015). MapComp is not meant to be used for RADseq based phylogenetic analysis, which requires identical markers for comparisons; this is rather performed using the direct marker approach with reciprocal best hit BLAST (Cariou et al. 2013; Pante et al. 2014).

MapComp is thus an easy solution to compare genetic maps in a way that is more tolerant of different library preparation protocols and phylogenetic distances. As shown here, MapComp is effective at finding orthology between chromosomes (Table 2), permitting the characterization of chromosomal rearrangements since whole genome duplication (Figure 4) and identifying putative structural rearrangements (Figure 5). This method will allow for the exploration of corresponding regions between species, such as regions harboring QTLs (Sarropoulou et al.

2008). Advances in genomics have resulted in many taxonomic groups having at least one species with a reference genome at some stage of assembly, providing the intermediate genome needed for this approach, and opening up this approach for a number of other taxonomic groups. MapComp is freely available at: https://github.com/enormandeau/mapcomp/

Conclusions

We provide the most complete analysis to date of the chromosomal rearrangements that lead to the current chromosome conformations in salmonids using the newly developed MapComp method. This analysis permitted the integration of all high-density salmonid maps across the lineage, identifying the timing of fusions of all chromosomes, including those still undergoing residual tetrasomy in the characterized species. Large inversions were also identified using this method, including a pericentric inversion that has occurred after the ancestral fusion of two chromosome arms and potentially rearranged the position of 11 genes across a centromere. These analyses will be further refined through the continued availability of other high-density salmonid maps, and can provide insights into the chromosomal evolution in both salmonids and other taxa.

Acknowledgements

This work was funded by a Fonds de Recherche Nature et Technologies (FRQNT) research grant awarded to Céline Audet, Louis Bernatchez and Nadia Aubin-Horth and a grant from the Société de Recherche et de Développement en Aquaculture Continentale (SORDAC) awarded to Céline Audet and Louis Bernatchez. We would like to thank the staff of the Institut des Sciences de la Mer de Rimouski (ISMER), Pierre East from Pisciculture de la Jacques-Cartier, the Réseau Aquaculture Québec (RAQ), and Pierre Dubé from SORDAC. We also express our thanks to G. Côté, G. Légaré, C. Rougeux and B. Boyle from the Institut de Biologie Intégrative et des

Systèmes (IBIS) for help with DNA extraction and library preparation. We also thank A-M Dion-Côté, E Rondeau, M. Limborg, two anonymous Reviewers and D. Huchon for valuable comments on the manuscript. BJGS was supported by an NSERC postdoctoral fellowship. Thanks to the members of the Bernatchez Laboratory for discussion and support.

References

Aljanabi SM, Martinez I. 1997. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. Nucleic Acids Research. 25:4692–4693.

Allendorf FW et al. 2015. Effects of crossovers between homeologs on inheritance and population genomics in polyploid-derived salmonid fishes. J. Hered. 106:217–227. doi: 10.1093/jhered/esv015.

Allendorf FW, Danzmann RG. 1997. Secondary tetrasomic segregation of MDH-B and preferential pairing of homeologues in rainbow trout. Genetics. 145:1083–1092.

Allendorf FW, Thorgaard GH. 1984. Tetraploidy and the evolution of salmonid fishes. Evolutionary Genetics of Fishes. 1–53. doi: 10.1007/978-1-4684-4652-4_1.

Amores A et al. 2014. A RAD-tag genetic map for the Platyfish (*Xiphophorus maculatus*) reveals mechanisms of karyotype evolution among teleost fish. Genetics. 197:625–641. doi: 10.1534/genetics.114.164293.

Amores A, Catchen J, Ferrara A, Fontenot Q, Postlethwait JH. 2011. Genome evolution and meiotic maps by massively parallel DNA sequencing: Spotted Gar, an outgroup for the teleost genome duplication. Genetics. 188:799–808. doi: 10.1534/genetics.111.127324.

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet. 17:81–92. doi: 10.1038/nrg.2015.28.

Baird NA et al. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE. 3:e3376. doi: 10.1371/journal.pone.0003376.

Berthelot C et al. 2014. The Rainbow Trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. Nat Comms. 5:3657. doi: 10.1038/ncomms4657.

Brelsford A, Dufresnes C, Perrin N. 2015. High-density sex-specific linkage maps of a European tree frog (*Hyla arborea*) identify the sex chromosome without information on offspring sex. 116:177–181. doi: 10.1038/hdy.2015.83.

Brieuc MSO et al. 2014. A dense linkage map for Chinook Salmon (*Oncorhynchus tshawytscha*)

reveals variable chromosomal divergence after an ancestral whole genome duplication event. G3 - Genes|Genomes|Genetics. 4:447–460. doi: 10.1534/g3.113.009316.

Broman KW, Wu H, Sen S, Churchill GA. 2003. R/qtl: QTL mapping in experimental crosses. Bioinformatics. 19:889–890. doi: 10.1093/bioinformatics/btg112.

Brunet FG et al. 2006. Gene loss and evolutionary rates following whole-genome duplication in teleost fishes. Molecular Biology and Evolution. 23:1808–1816. doi: 10.1093/molbev/msl049.

Cariou M, Duret L, Charlat S. 2013. Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. Ecology and Evolution. 3:846–852. doi: 10.1002/ece3.512.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013. Stacks: an analysis tool set for population genomics. Mol. Ecol. 22:3124–3140. doi: 10.1111/mec.12354.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. 2011. Stacks: building and genotyping loci *de novo* from short-read sequences. G3 - Genes|Genomes|Genetics. 1:171–182. doi: 10.1534/g3.111.000240.

Crête-Lafrenière A, Weir LK, Bernatchez L. 2012. Framing the Salmonidae Family phylogenetic portrait: a more complete picture from increased taxon sampling Fontaneto, D, editor. PLoS ONE. 7:e46662. doi: 10.1371/journal.pone.0046662.

Danzmann RG et al. 2008. Distribution of ancestral proto-Actinopterygian chromosome arms within the genomes of 4R-derivative salmonid fishes (Rainbow Trout and Atlantic Salmon). BMC Genomics. 9:557. doi: 10.1186/1471-2164-9-557.

Davidson WS et al. 2010. Sequencing the genome of the Atlantic Salmon (*Salmo salar*). Genome Biol. 11:403. doi: 10.1186/gb-2010-11-9-403.

Elshire RJ et al. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE. 6:e19379. doi: 10.1371/journal.pone.0019379.

Everett MV, Miller MR, Seeb JE. 2012. Meiotic maps of Sockeye Salmon derived from massively parallel DNA sequencing. BMC Genomics. 13:521. doi: 10.1186/1471-2164-13-521.

Fang Z et al. 2003. cMap: the comparative genetic map viewer. Bioinformatics. 19:416–417. doi: 10.1093/bioinformatics/btg012.

Fierst JL. 2015. Using linkage maps to correct and scaffold de novo genome assemblies: methods, challenges, and computational tools. Front. Genet. 6:220. doi: 10.3389/fgene.2015.00220.

Force A et al. 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics. 151:1531–1545. doi: 10.1007/s10709-008-9311-5.

Gagnaire P-A, Normandeau E, Pavey SA, Bernatchez L. 2013. Mapping phenotypic, expression and transmission ratio distortion QTL using RAD markers in the Lake Whitefish (*Coregonus*

clupeaformis). Mol. Ecol. 22:3036–3048. doi: 10.1111/mec.12127.

Gonen S et al. 2014. Linkage maps of the Atlantic Salmon (*Salmo salar*) genome derived from RAD sequencing. BMC Genomics. 15:166. doi: 10.1186/1471-2164-15-166.

Gonen S, Bishop SC, Houston RD. 2015. Exploring the utility of cross-laboratory RAD-sequencing datasets for phylogenetic analysis. BMC Research Notes. 8:299. doi: 10.1186/s13104-015-1261-2.

Gosselin T, Bernatchez L. stackr: GBS/RAD Data Exploration, Manipulation and Visualization using R. doi: 10.5281/zenodo.19647.

Grattapaglia D, Sederoff R. 1994. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. Genetics. 137:1121–1137.

Gregory TR. 2016. Animal Genome Size Database. http://www.genomesize.com.

Henning F, Lee HJ, Franchini P, Meyer A. 2014. Genetic mapping of horizontal stripes in Lake Victoria cichlid fishes: benefits and pitfalls of using RAD markers for dense linkage mapping. Mol. Ecol. 23:5224–5240. doi: 10.1111/mec.12860.

Ilut DC, Nydam ML, Hare MP. 2014. Defining loci in restriction-based reduced representation genomic data from nonmodel species: sources of bias and diagnostics for optimal clustering. Biomed Res Int. 2014:675158. doi: 10.1155/2014/675158.

Kinnison MT, Hendry AP. 2004. From Macro- to Micro-Evolution. In: Evolution Illuminated: Salmon and their Relatives. Hendry, AP & Stearns, SC, editors. New York pp. 208–231.

Kirkpatrick M. 2010. How and why chromosome inversions evolve. Plos Biol. 8:e1000501. doi: 10.1371/journal.pbio.1000501.

Kirubakaran TG et al. 2016. Two adjacent inversions maintain genomic differentiation between migratory and stationary ecotypes of Atlantic cod. Mol. Ecol. n/a–n/a. doi: 10.1111/mec.13592.

Kodama M, Brieuc 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. 4:1717–1730. doi: 10.1534/g3.114.012294.

Larson WA et al. 2015. Identification of Multiple QTL Hotspots in Sockeye Salmon (Oncorhynchus nerka) Using Genotyping-by-Sequencing and a Dense Linkage Map. J. Hered. doi: 10.1093/jhered/esv099.

Lee GM, Wright JE. 1981. Mitotic and meiotic analyses of Brook Trout, *Salvelinus fontinalis*. J. Hered. 321–327.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform.

Bioinformatics. 25:1754–1760. doi: 10.1093/bioinformatics/btp324.

Lien S et al. 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. 12:615. doi: 10.1186/1471-2164-12-615.

Lien S et al. 2016. The Atlantic salmon genome provides insights into rediploidization. Nature. 1–13. doi: 10.1038/nature17164.

Limborg MT, Waples RK, Allendorf FW, Seeb JE. 2015. Linkage mapping reveals strong chiasma interference in Sockeye Salmon: Implications for interpreting genomic data. G3 - Genes|Genomes|Genetics. 5:2463–2473. doi: 10.1534/g3.115.020222.

Limborg MT, Waples RK, Seeb JE, Seeb LW. 2014. Temporally isolated lineages of pink salmon reveal unique signatures of selection on distinct pools of standing genetic variation. J. Hered. 105:741–751. doi: 10.1093/jhered/esu063.

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. 281:20132881–20132881. doi: 10.1098/rspb.2013.2881.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 17:10. doi: 10.14806/ej.17.1.200.

Mascher M, Stein N. 2014. Genetic anchoring of whole-genome shotgun assemblies. Front. Genet. 5:208. doi: 10.3389/fgene.2014.00208.

May B, Delany ME. 2015. Meiotic models to explain classical linkage, pseudolinkage, and chromosomal pairing in tetraploid derivative salmonid genomes: II. Wright is still right. J. Hered. 106:762–766. doi: 10.1093/jhered/esv056.

McKinney GJ et al. 2015. An integrated linkage map reveals candidate genes underlying adaptive variation in Chinook Salmon (Oncorhynchus tshawytscha). n–a–n–a. doi: 10.1111/1755-0998.12479.

Miller MR et al. 2012. A conserved haplotype controls parallel adaptation in geographically distant salmonid populations. Mol. Ecol. 21:237–249. doi: 10.1111/j.1365-294X.2011.05305.x.

Ming R, Man Wai C. 2015. Assembling allopolyploid genomes: no longer formidable. Genome Biol. 16:27. doi: 10.1186/s13059-015-0585-5.

Naish KA et al. 2013. Comparative genome mapping between Chinook Salmon (*Oncorhynchus tshawytscha*) and Rainbow Trout (*O. mykiss*) based on homologous microsatellite loci. G3 - Genes|Genomes|Genetics. 3:2281–2288. doi: 10.1534/g3.113.008003.

Noor MA et al. 2001a. The genetics of reproductive isolation and the potential for gene exchange between *Drosophila pseudoobscura* and *D. persimilis* via backcross hybrid males. Evolution.

55:512–521.

Noor MA, Grams KL, Bertucci LA, Reiland J. 2001b. Chromosomal inversions and the reproductive isolation of species. Proceedings of the National Academy of Sciences. 98:12084–12088. doi: 10.1073/pnas.221274498.

Ohno S. 1970. *Evolution by Gene Duplication*. Springer Berlin Heidelberg: Berlin, Heidelberg doi: 10.1007/978-3-642-86659-3.

Ostberg CO, Hauser L, Pritchard VL, Garza JC, Naish KA. 2013. Chromosome rearrangements, recombination suppression, and limited segregation distortion in hybrids between Yellowstone Cutthroat Trout (*Oncorhynchus clarkii bouvieri*) and Rainbow Trout (*O. mykiss*). BMC Genomics. 14:570. doi: 10.1186/1471-2164-14-570.

Palti Y et al. 2015. Detection and validation of QTL affecting bacterial cold water disease resistance in Rainbow Trout using restriction-site associated DNA sequencing Wang, H, editor. PLoS ONE. 10:e0138435. doi: 10.1371/journal.pone.0138435.

Pante E et al. 2014. Use of RAD sequencing for delimiting species. Heredity. 114:450–459.

Phillips R, Ráb P. 2001. Chromosome evolution in the Salmonidae (Pisces): an update. Biol Rev Camb Philos Soc. 76:1–25.

Phillips RB et al. 2009. Assignment of Atlantic Salmon (*Salmo salar*) linkage groups to specific chromosomes: Conservation of large syntenic blocks corresponding to whole chromosome arms in Rainbow Trout (*Oncorhynchus mykiss*). BMC Genet. 10:46. doi: 10.1186/1471-2156-10-46.

Poland JA, Brown PJ, Sorrells ME, Jannink J-L. 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme Genotyping-by-Sequencing approach Yin, T, editor. PLoS ONE. 7:e32253. doi: 10.1371/journal.pone.0032253.

Rieseberg LH. 2001. Chromosomal rearrangements and speciation. Trends in Ecology & Evolution. 16:351–358.

Rondeau EB et al. 2014. The genome and linkage map of the Northern Pike (*Esox lucius*): conserved synteny revealed between the salmonid sister group and the Neoteleostei. PLoS ONE. 9:e102089. doi: 10.1371/journal.pone.0102089.

Sankoff D, Zheng C, Zhu Q. 2010. The collapse of gene complement following whole genome duplication. BMC Genomics. 11:1. doi: 10.1186/1471-2164-11-313.

Sarropoulou E, Nousdili D, Magoulas A, Kotoulas G. 2008. Linking the genomes of nonmodel teleosts through comparative genomics. Mar Biotechnol. 10:227–233. doi: 10.1007/s10126-007-9066-5.

Sauvage C, Vagner M, Derôme N, Audet C, Bernatchez L. 2012a. Coding gene single nucleotide polymorphism mapping and quantitative trait loci detection for physiological reproductive traits in Brook Charr, *Salvelinus fontinalis*. G3 - Genes|Genomes|Genetics. 2:379–392. doi:

10.1534/g3.111.001867.

Sauvage C, Vagner M, Derôme N, Audet C, Bernatchez L. 2012b. Coding gene SNP mapping reveals QTL linked to growth and stress response in Brook Charr (*Salvelinus fontinalis*). G3 - Genes|Genomes|Genetics. 2:707–720. doi: 10.1534/g3.112.001990.

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.

Timusk ER et al. 2011. Genome evolution in the fish family salmonidae: generation of a Brook Charr genetic map and comparisons among charrs (Arctic Charr and Brook Charr) with Rainbow Trout. BMC Genet. 12:68. doi: 10.1186/1471-2156-12-68.

van Ooijen JW. 2006. JoinMap ® 4. Software for the calculation of genetic linkage map.

van Ooijen JW. 2011. Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species. Genet Res (Camb). 93:343–349. doi: 10.1017/S0016672311000279.

van Ooijen JW, Jansen J. 2013. *Genetic mapping in experimental populations*. Cambridge University Press: Cambridge doi: 10.1017/cbo9781139003889.

Waples RK, Seeb LW, Seeb JE. 2016. Linkage mapping with paralogs exposes regions of residual tetrasomic inheritance in Chum Salmon (*Oncorhynchus keta*). Mol. Ecol. Resour. 16:17–28. doi: 10.1111/1755-0998.12394.

Woram RA et al. 2004. A genetic linkage map for Arctic Char (*Salvelinus alpinus*): evidence for higher recombination rates and segregation distortion in hybrid versus pure strain mapping parents. Genome. 47:304–315. doi: 10.1139/g03-127.

Wright JE, Johnson K, Hollister A, May B. 1983. Meiotic models to explain classical linkage, pseudolinkage, and chromosome pairing in tetraploid derivative salmonid genomes. Isozymes Curr Top Biol Med Res. 10:239–260.

Wu R, Ma C-X, Painter I, Zeng Z-B. 2002. Simultaneous maximum likelihood estimation of linkage and linkage phases in outcrossing species. Theor Popul Biol. 61:349–363. doi: 10.1006/tpbi.2002.1577.

Supporting Information

Additional File S1. Mapping software parameters and bioinformatics pipeline overview. STACKS and JOINMAP parameters and an outline of bioinformatics steps used to generate the Brook Charr linkage map.

Additional File S2. Descriptive statistics for Brook Charr female map.

Additional File S3. Brook Charr female linkage map. Includes species name, linkage group, cM position, marker name and sequence.

Additional File S4. Putative genes within the 9.2-17.1 pericentric inversion. Genes annotated in the Atlantic Salmon genome within the putative inverted region in the Pink, Chum and Sockeye Salmon clade.

Additional File S5. Brook Charr and all other species in Oxford grids. Oxford grids comparing Brook Charr to other species using the Rainbow Trout genome as the reference intermediate, including: 1) Lake Whitefish, 2) Atlantic Salmon, 3) Rainbow Trout, 4) Coho Salmon, 5) Chinook Salmon, 6) Pink Salmon, 7) Chum Salmon, 8) Sockeye Salmon.

Additional File S6. Distribution of paired marker distances. Frequencies of distances between marker pairs in various comparisons via MAPCOMP. As can be seen, more closely-related species have more homologous markers (0 distance between marker pairs), whereas the more distantly related comparisons have fewer homologous markers.

Code and Pipeline Availability

MAPCOMP: https://github.com/enormandeau/mapcomp/

Collecting and formatting available salmonid maps:
https://github.com/bensutherland/2016_ms_sfonmap

RADseq workflow: http://gbs-cloud-tutorial.readthedocs.org

STACKS workflow: https://github.com/enormandeau/stacks_workflow