Novel Method for Comparing RADseq Linkage Maps Reveals Chromosome Evolution in Salmonids

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

***** Corresponding author

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

Running title: Linkage Map Comparison in Salmonids

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

Abstract

Genome duplication can provide material for evolutionary innovation, and much remains unknown about its functional effects. Assembly of large, outbred eukaryotic genomes is 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. However, these methods often result in a low number of homologous markers between species due to phylogenetic distance or technical differences in library preparation. Family Salmonidae is ideal for studying the effects of whole genome duplication. The ancestral salmonid underwent whole genome duplication around 65 million years ago and the tetraploid genome has undergone rediploidization during the salmonid diversification. In the salmonids, synteny occurs between orthologous chromosomes, but each species exhibits conserved and unique chromosome arm fusions and fissions. In this study, we identify orthologous chromosome arms within the salmonids using available RADseq salmonid linkage maps along with a new high-density linkage map (3826 markers) constructed for the *Salvelinus* genera (Brook Charr *S. fontinalis*). 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 (e.g. Rainbow Trout *Oncorhynchus mykiss*). We greatly increased the number of comparable markers between linkage maps relative to that obtained using only identical markers. This enabled a characterization of the most likely history of retained chromosomal rearrangements post-whole genome duplication in five species of *Oncorhynchus*, and one species of each of *Salvelinus*, *Salmo*, and *Coregonus*, representing all of the main salmonid genera. Additionally, integration with the genetic map of the pre-duplicated sister species Northern Pike *Esox lucius* permitted the identification of homeologous chromosomes in all species. Putative conserved inversions within

chromosome arms were also identified among species. Analyses of RADseq-based linkage maps from other taxa are likely to benefit from MAPCOMP, available at: https://github.com/enormandeau/mapcomp/.

Author Summary

Whole genome duplication occurred in the common ancestor of the salmonids, after which the salmonid lineage returned to the original diploid state but with twice the number of chromosomes. Over evolutionary time, the genome was rearranged by fusions and fissions changing chromosome structures. This process has taken different paths in each salmonid lineage, and in living salmonid species some genomic regions still have not completely returned to a diploid state. Here we generate the first high-density genetic map (3826 markers) for a species within an under characterized lineage, Brook Charr *Salvelinus fontinalis.* Then we integrate this map with all other available salmonid maps by identifying the corresponding chromosome arms among the species, including eight species across four genera, as well as a pre-duplicated sister species to the salmonids, Northern Pike *Esox lucius*. This characterization was permitted by the use of a novel map comparison tool, MAPCOMP, which greatly increases the numbers of comparable markers between linkage maps relative to traditional methods. MAPCOMP uses a reference genome of a related species to pair both identical and proximal markers. Integrating the maps allowed for the characterization of all major retained fusion events as well as some smaller rearrangements (i.e. inversions).

Introduction

Whole genome duplication (WGD) can provide the raw material for evolutionary innovation through generating redundant copies of chromosomes. Gene copies can then evolve new functions, sub-functionalize the original functions between the two copies, or most frequently, accumulate mutations that disrupt functionality of one copy [1]. Cross-taxa analyses suggest that during rediploidization, retention of a single gene copy often occurs within the same homeologous chromosome [2]. Generally, the biological effects of WGD and subsequent rediploidization on genome function and speciation remain poorly understood.

Genomes of eukaryotic organisms with retained ancestral WGD (e.g. pseudotetraploid genomes) are challenging to assemble, and therefore linkage maps are useful for comparing chromosomal evolution among lineages [3,4]. High quality, dense linkage maps can also be valuable for validating and orienting genomic scaffolds [5,6], especially for cases of residual polyploidy, large genome size, and high repeat content [7,8]. Recent advances in sequencing, such as through reduced-representation library sequencing (e.g. RADseq) [9-11], 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 [12]. RADseq-based SNP markers are contained in short sequence fragments, which allows for mapping against a genome to identify nearby genes or physical distances between markers [13,14]. RADseq also enables comparative genomics through the use of direct marker-to-marker comparisons to find homologous markers between linkage maps [4]. For this use, however, the number of homologous markers available for comparisons between species decreases with phylogenetic distance [15]. This issue is compounded further when different protocols or restriction enzymes are used. Due to this, it has been suggested to use a common

enzyme and protocol to ensure compatibility of maps [16]. Since this may not always be possible or desirable, 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 over the course of evolutionary time [17,18]. Post-WGD, the salmonid lineage diversified into three subfamilies, 11 genera and more than 60 described species [19], although this diversification was likely due to environmental factors rather than being caused by WGD [20]. This gradual return to diploidy among different lineages led to different evolutionary paths to rediploidization. Although much remains to be understood about this process in salmonids, fundamental work on chromosomal evolution has been conducted using cytogenetics and genetic maps [3,21]. For example, in the study of sex determination in salmonids, a sex-linked region is known to be present in different autosomes [22]. This region has locally conserved synteny [23], probably contained within a cassette moved throughout the genome by transposition events [24]. Comparative genomic analyses have also been conducted for some of the more characterized species, such as Rainbow Trout *Oncorhynchus mykiss* or Atlantic Salmon *Salmo salar*. In particular, these analyses have focused on loss of gene duplicates (~50%) [25] paralog conservation [26], and more. For comparative purposes, Northern Pike *Esox lucius* (2n = 50) is used as a sister species for the salmonid WGD [26,27].

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 (NF = 100) [21]. Fissions also occur, subsequently separating the fused metacentric chromosomes. Cytogenetics in salmonids has identified the presence of two major karyotype groups that differ in the number of retained chromosome fusion events that form larger metacentric chromosomes. Type A species (2n = ~80) have more acrocentric than metacentric chromosomes, whereas Type B species (2n = ~60) have more metacentric than acrocentric chromosomes [21]. Adaptive mechanisms or selective forces driving these rearrangements and correlation with habitat or species biology remain generally unknown [21]. Among species, strong synteny is expected between orthologous chromosome arms [4]. Within a species collinearity is observed between homeologs [25]. 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 [4]. 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.

Rearrangements within a species affect rediploidization efficiency, and thus are valuable to consider for comparative genomics. Rediploidization occurs by gradual sequence divergence between homeologous chromosomes, starting closer to the centromere and spreading towards telomeric regions [28]. The pace of this rediploidization process differs throughout the genome, being impeded in regions with fused metacentric chromosomes. As such, each salmonid lineage with different metacentric fusions will have different regions of impeded rediploidization, and therefore will provide different information for the overall salmonid rediploidization process. Rediploidization is most likely impeded in metacentric chromosomes because of increased tetravalent formation at meiosis allowing homologous recombination between homeologs (i.e.

residual tetrasomy) and thus reducing sequence divergence [4,29] particularly in male salmonid telomeric regions [28-31]. In this regard, it is important to characterize orthologous relationships between chromosome arms and states of chromosomes (metacentric or acrocentric) in order to gain information regarding the expectation for levels of homeologous differentiation within each species.

High-density linkage maps have been constructed for Atlantic Salmon *S. salar* [32,33], members of *Oncorhynchus* including Coho salmon *O. kisutch* [4], Sockeye Salmon *O. nerka* [16,34], Chinook Salmon *O. tshawytscha* [35], Rainbow Trout *O. mykiss* [36,37] and Chum Salmon *O. keta* [30], as well as Lake Whitefish *Coregonus clupeaformis* [38]. 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* [39,40], as well as a low-density (~300 marker) EST-derived SNP map for *S. fontinalis* [41]. High quality genome assemblies exist for Rainbow Trout *O. mykiss* [25,42] and Atlantic Salmon *Salmo salar* [18]. A genomic assembly and genetic map are also available for Northern Pike *Esox lucius*, a sister species to the salmonid WGD [27]. 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 [41] to generate 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 [43]. It identifies on average 4-6 times 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.

Results

Generation of a Brook Charr linkage map

An average of 10M single-end reads were obtained for each parent and 5M for each individual offspring. Using STACKS v1.32 [12], 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 [4] due to low recombination rate and increased residual tetrasomy in males [29].

A total of 42 linkage groups were characterized in the female map (Figure 1), corresponding to the expected haploid chromosome number for Brook Charr [21]. 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 S1. This is in the range of other high-density salmonid maps, such as the Coho Salmon linkage map (6596.7 cM) [4], although is larger than the Chinook Salmon map (4164 cM) [35]. The female map contains 3826 markers with the following marker types, as defined by Wu et al. [44]: 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 S2. The male map contained an additional 2385 fully informative in male parent (*aa* x *ab*) markers, but these markers did not position well (described briefly below), and therefore were not used, as is typical for salmonid maps.

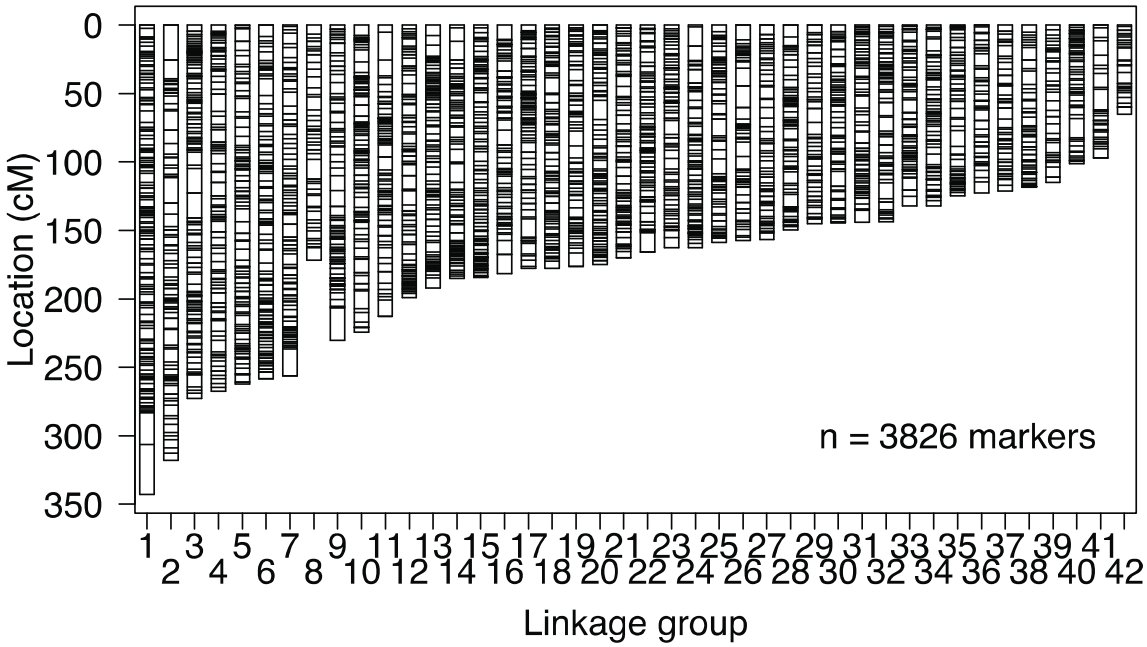

Figure 1. Brook Charr *Salvelinus fontinalis* Linkage Map. Eight metacentric (LG1-8) and 34 acrocentric linkage groups (LG9-42) were identified in the female map. Horizontal lines within each linkage group are markers (n = 3826 markers).

Identification of orthologous chromosome arms among the salmonids

Assignment of linkage groups to chromosome arms has been determined with homologous microsatellite and RADseq markers (using the same library preparation protocols) among

Chinook Salmon, Coho Salmon, Rainbow Trout and Atlantic Salmon [3,4,35,45,46] and recently Sockeye Salmon [16]. 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 [27], but not yet with the rest of the salmonids. Details on the linkage maps and species used in this analysis are provided in Table 1.

Table 1. Overview of species compared. The common name, genus and species name, source of the genetic map, the abbreviation are shown along with the type of map and number markers, the chromosome number for the species, and expected genome size (C-value and Gbp) (obtained from [76]).

Common and Scientific Name and Source	**Abbr.**	**Map Type (Num. markers)**	**Num. Chr. (1n)**	**Exp. Size of Genome (C-value)** [76]	**Exp. Size of Genome (Gbp)**
Northern Pike* *Esox lucius* [27]	*Eluc*	EST-based microsatellite (524)	25	0.85-1.40	0.8-1.3
Lake Whitefish *Coregonus clupeaformis* [38]	*Cclu*	RADseq with *Sbf*I (3438)	40	2.44-3.44	2.3-3.3
Atlantic Salmon *Salmo salar* [32]	*Ssal*	EST-based SNP chip (5650)	29	2.98-3.27	2.9-3.1
Brook Charr *Salvelinus fontinalis*	*Sfon*	RADseq with *Pst*I and *Msp*I (3826)	42	2.86-3.50	2.7-3.4
Sockeye Salmon *Oncorhynchus nerka* [16]	*Oner*	RADseq with *Sbf*I (6262)	28**	2.77-3.04	2.7-2.9
Chum Salmon *O. keta* [30]	*Oket*	RADseq with *Sbf*I (6119)	37	2.49-2.76	2.4-2.6
Chinook Salmon *O. tshawytscha* [35]	*Otsh*	RADseq with *Sbf*I (6352)	34	2.45-3.30	2.3-3.2
Coho Salmon *O. kisutch* [4]	*Okis*	RADseq with *Sbf*I (5377)	30	2.60-3.05	2.5-2.9
Rainbow Trout *O. mykiss* [37]	*Omyk*	RADseq with *Sbf*I (955)	29	1.87-2.92	1.8-2.8

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

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 [25] (see MAPCOMP schematic in Figure 2, 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, 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 *O. mykiss* (*Eluc* 25.1) required using a second linkage map [36] for *O. mykiss* to unambiguously identify orthology. Most arms were also identified in the more distantly related Lake Whitefish, but six arms remained unidentifiable and two remained not definitive.

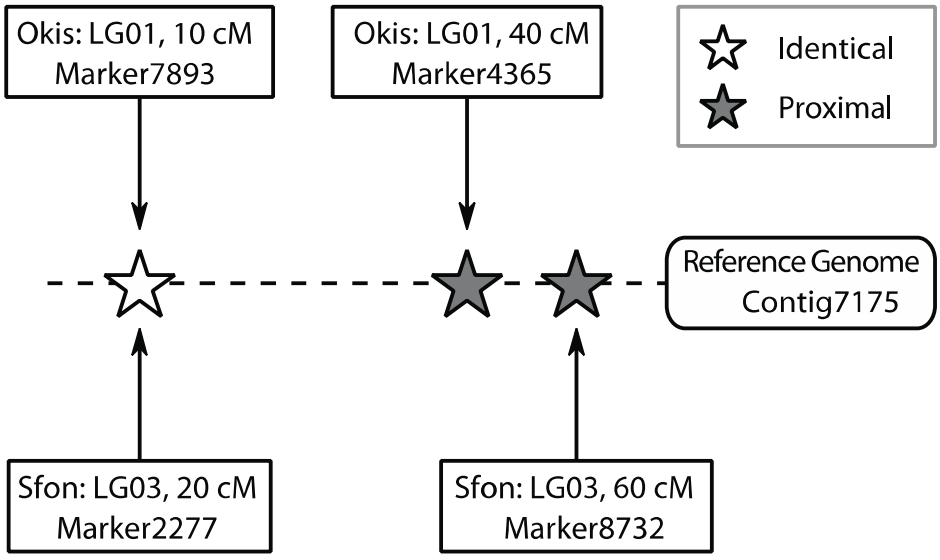

Figure 2. Schematic of MAPCOMP using a reference genome to pair markers. MAPCOMP works by comparing genetic maps from two different species by mapping marker sequences against a reference genome, then retaining high quality mappings that only hit at one place in the genome. Markers from each species are paired if they hit against the same contig/scaffold by taking the closest two markers together as each pair. Each marker can only be paired once, and any other marker that was second closest (or further) to the now-paired marker is discarded. This captures identical markers (open star in image) and non-identical markers (closed stars). Finally, the

linkage group and cM position of each marker is plotted in an Oxford grid (see Figure 3). Note that the marker names and contig ID in the schematic are for demonstration purposes only and do not reflect actual pairings.

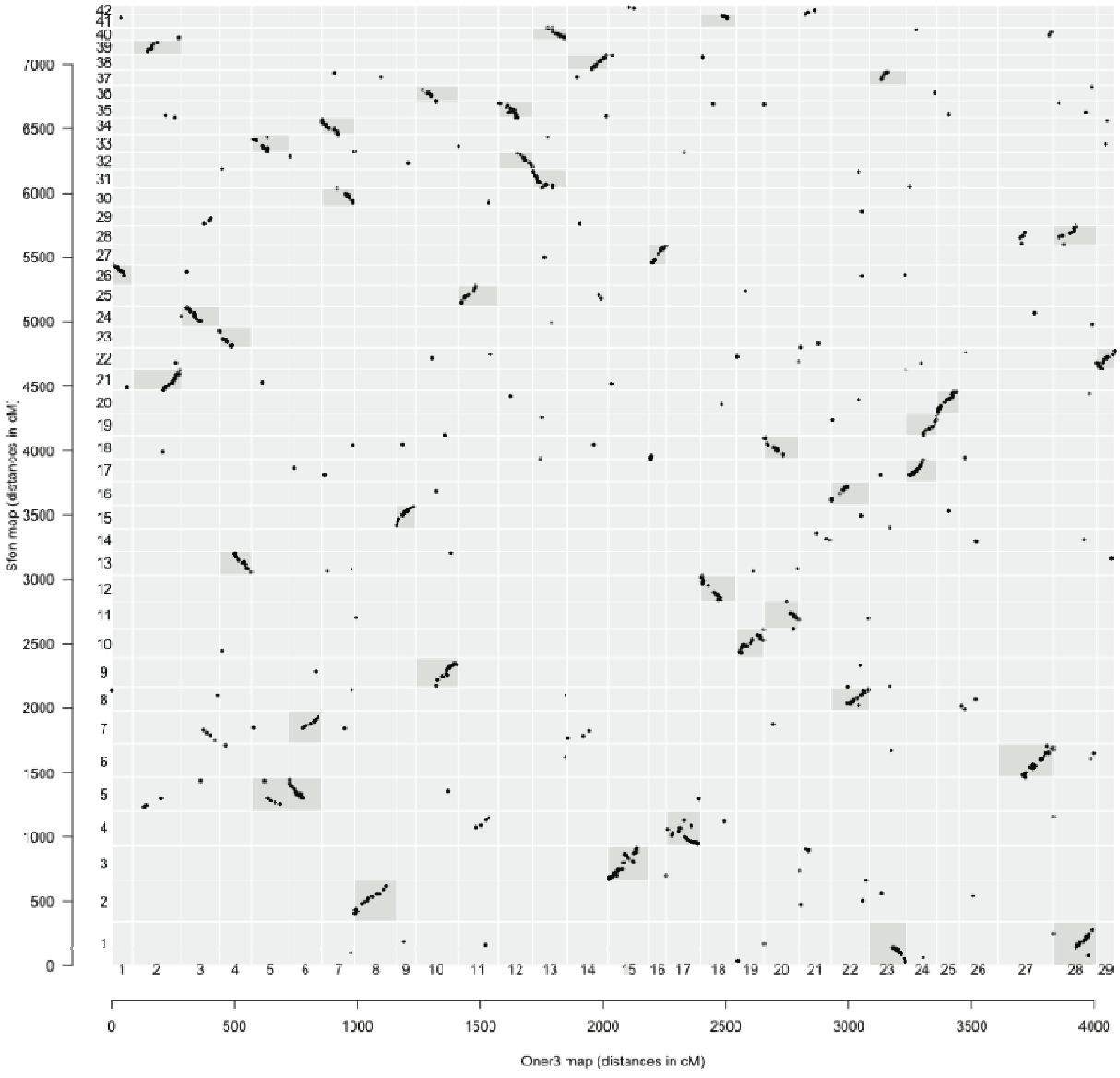

Figure 3. MAPCOMP determination of orthologous chromosome arms. Brook Charr (*Sfon*; y-axis) compared with Sockeye Salmon (*Oner*; y-axis) with markers paired through the Rainbow Trout genome identifies orthology between chromosome arms. A putative inversion can be seen at *Sfon* LG03 and *Oner* LG15. Other comparison Oxford grids are presented in Additional File 1.

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. See Table 1 for common names, abbreviated names, Latin names and source citations for maps.

Eluc	***Cclu***	***Ssal***	***Sfon***	***Oket***	***Oner***	***Omyk***	***Okis***	***Otsh***
1.1	Cclu28	Ssa20b	Sf25	Ok18	On11a	Omy27	Co15b	Ots13q
1.2	Cclu35	Ssa09c	Sf38	Ok01a	On14b	Omy24	Co18a	Ots14q
2.1	Cclu04a	Ssa26	Sf06a	Ok14b	On27a	Omy06b	Co03b	Ots04q
2.2	Cclu04b(mid)	Ssa11a	Sf28	Ok02b	On28a	Omy26	Co08b	Ots12q
3.1	Cclu25	Ssa14a	Sf22	Ok23	On29	Omy08b	Co30	Ots10q
3.2	Cclu26	Ssa03a	Sf11	Ok15	On20b	Omy28	Co27	Ots28
4.1	Cclu16	Ssa09b(mid)	Sf33	Ok30a	On05a	Omy25a	Co15a	Ots08q
4.2	Cclu29	Ssa05a	Sf07b	Ok04b	On06b	Omy14b*	Co19b	Ots21
5.1	Cclu05a	Ssa19b	Sf01a	Ok03	On23b	Omy16a	Co20a	Ots24
5.2	Cclu15	Ssa28	Sf27	Ok24	On16	Omy20a+b?	Co25	Ots25
6.1	Cclu05b(mid)	Ssa01b(mid)	Sf01b	Ok02a	On28b	Omy23	Co11a	Ots01q
6.2	Cclu05c	Ssa18a	Sf36	Ok36	On10a	Omy01b	Co04b	Ots06q
7.1	Cclu13	Ssa13b	Sf08b	Ok12	On22b	Omy12a	Co06a	Ots09p
7.2	Cclu08	Ssa04b	Sf09	Ok11	On10b	Omy10a	Co28	Ots30
8.1	Cclu36	Ssa23	Sf04a	Ok26	On17	Omy04a	Co10a	Ots01p
8.2	Cclu06	Ssa10a	Sf17	Ok01b	On24a	Omy05b	Co13a	Ots05q
9.1	Cclu06b	Ssa02b	Sf42	Ok32a	On21a	Omy13a	Co20b	Ots32
9.2	*missing*	Ssa12a	Sf03b	Ok10b	On15b	Omy17b	Co01b	Ots02q
10.1	Cclu10	Ssa27	Sf23	Ok20	On04a	Omy18b	Co17a	Ots13p
10.2	Cclu24a	Ssa14b	Sf34	Ok19a	On07a	Omy14a	Co14b	Ots31
11.1	Cclu18	Ssa06a	Sf14	Ok32b	On21b	Omy13b	Co10b	Ots27
11.2	*missing*	Ssa03b	Sf08a*	Ok05	On26	Omy12b	Co06b	Ots09q
12.1	Cclu27	Ssa13a	Sf18	Ok27	On20a	Omy16b	Co24	Ots22
12.2	Cclu14	Ssa15b	Sf30	Ok28a	On07b	Omy09b	Co17b	Ots16q
13.1	Cclu34	Ssa24	Sf06b	Ok14a	On27b	Omy06a	Co03a	Ots04p
13.2	Cclu37	Ssa20a	Sf40	Ok25	On13b	Omy11a	Co08a	Ots12p
14.1	Cclu04b	Ssa01c	Sf13	Ok09	On04b	Omy05a	Co23	Ots20
14.2	Cclu33	Ssa11b	Sf10	Ok06	On19	Omy29^*	Co29	Ots33
15.1	Cclu31	Ssa09a	Sf35	Ok35a	On12a	Omy25b	Co14a	Ots08p
15.2	Cclu22	Ssa01a	Sf12	Ok29a	On18a	Omy19b	Co07b	Ots11q
16.1	Cclu02	Ssa21	Sf26	Ok07	On01	Omy22	Co26	Ots26
16.2	Cclu32	Ssa25	Sf24	Ok34a	On03a	Omy03b	Co02b	Ots03q
17.1	Cclu38	Ssa12b	Sf03a	Ok10a	On15a	Omy17a	Co01a	Ots02p
17.2	Cclu21	Ssa22	Sf21	Ok21	On02b	Omy07b	Co05b	Ots07q
18.1	Cclu40	Ssa15a	Sf19	Ok19b	On24b	Omy08a	Co12a	Ots05p
18.2	Cclu17	Ssa06b	Sf31	Ok08	On13a	Omy04b	Co21	Ots18
19.1	Cclu30	Ssa10b	Sf15	Ok17	On09	Omy02b	Co22	Ots19

19.2	Cclu11	Ssa16a	Sf20	Ok22	On25	Omy01a	Co04a	Ots06p
20.1	Cclu10b^	Ssa05b	Sf07a	Ok31*	On14a	Omy02a	Co13b	Ots23
20.2	Cclu01a^	Ssa02a	Sf29	Ok34b	On03b	Omy03a	Co02a	Ots03p
21.1	Cclu12	Ssa29	Sf05b	Ok04a	On06a	Omy15a	Co11b	Ots29
21.2	Cclu39	Ssa19a	Sf16	Ok28b	On22a	Omy11b	Co18b	Ots16p
22.1	Cclu19?^	Ssa17a	Sf39	Ok37	On02a	Omy07a	Co05a	Ots07p
22.2	Cclu19?^	Ssa16b	Sf05a	Ok30b	On05b	Omy18a*	Co16b	Ots14p
23.1	*missing*	Ssa07b	Sf02b	Ok13b	On08b	Omy21a	Co09a	Ots15p
23.2	*missing*	Ssa17b	Sf37	Ok33	On23a	Omy15b*	Co19a	Ots17
24.1	Cclu24b	Ssa07a	Sf02a	Ok13a	On08a	Omy21b	Co09b	Ots15q
24.2	Cclu23	Ssa18b	Sf32	Ok35b	On12b	Omy09a	Co16a	Ots10p
25.1	*missing*	Ssa04a	Sf04b	Ok16	On11b	Omy10b**	Co12b	Ots34
25.2	*missing*	Ssa08a	Sf41	Ok29b^	On18b	Omy19a	Co07a	Ots11p

Notation: ^ = evidence for orthology is weak; * = needed to use second species (*Oner*) to identify; ** = needed to use a second Omyk map to identify [36]

In Brook Charr, a total of eight metacentric and 36 acrocentric chromosomes were expected from salmonid cytogenetics [21,47] and all were identified here (Table 2), increasing the resolution of the Brook Charr linkage maps from existing microsatellite-based linkage maps [40,41]. Since Brook Charr has the fewest number of metacentric chromosomes, often two acrocentric chromosomes in Brook Charr correspond to two fused chromosome arms in another species. In some cases, due to tandem chromosome fusions observed in Atlantic Salmon [46] three linkage groups in Brook Charr correspond to one linkage group in Atlantic Salmon. For example, *Sfon* Sf33, Sf35, Sf38 are in tandem fusions in *Ssal* Ssa09. We compared orthology identified with MAPCOMP between Coho Salmon and Chinook Salmon with an analyses that used homologous markers [4], 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 Chum Salmon, Sockeye Salmon and Lake Whitefish.

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 against the maps of all species using MAPCOMP. Parameters were adjusted to allow pairing of markers even when they hit two positions in the reference genome, as was expected for a sister species that did not undergo the salmonid WGD. As the number of markers was lower for the Northern Pike genetic map than the salmonid maps, a consensus approach was used to gain evidence from more than one species for unambiguous homeolog identification. These results were compared to that of Rondeau et al. [27], in which BLAST was used with Atlantic Salmon linkage groups against the Northern Pike genome to identify salmonid WGD homeologs. All homeologous pairs in Atlantic Salmon identified by MAPCOMP using the Rainbow Trout intermediate reference genome were concurrent with those originally identified [27], but here were also extended to all other species (Table 2).

Homeologous chromosomes can also be identified by mapping isoloci in haploid crosses [4,16,35]. Without a haploid cross for our Brook Charr map, here we cannot identify which homeologs still undergo residual tetrasomy at meiosis. The homeologs identified with MAPCOMP were compared with those identified using isoloci. All eight pairs of homeologs identified in Chinook Salmon [35] were confirmed. Another previously identified homeology that was not confirmed in Chinook Salmon (Ots24-Ots29) [35], was also not identified here. Eight pairs of homeologs that have been identified in Coho Salmon [4] and Sockeye Salmon [16] were also confirmed. 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).

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 eight metacentric chromosomes in Coho Salmon and Chinook Salmon [4]. Here, orthology characterization of chromosome arms allowed the inclusion of genera *Salvelinus* and *Coregonus* as well as the Sockeye Salmon/Chum Salmon clade within *Oncorhynchus.* We identified 16 different fusion events conserved in at least two species, five fission events conserved in at least two species, 69 species-specific fusion events, and three species-specific fission events (Figure 4). For simplicity, we use chromosome names from the Northern Pike (*Eluc*) chromosomes to refer to chromosome arms, including the duplicate designation, as shown in Table 2.

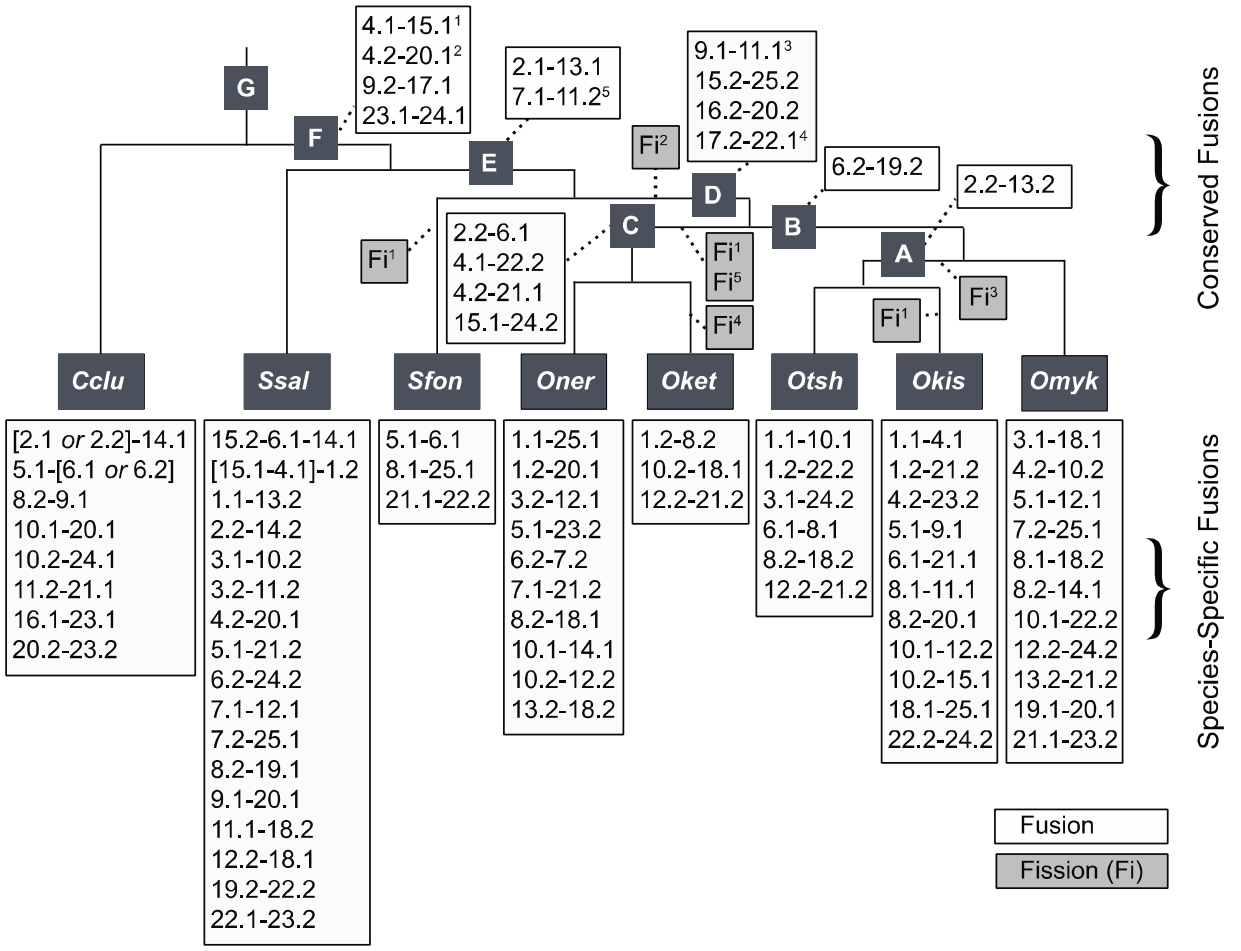

Figure 4. Fusions and fissions across the salmonid lineage. Different fusions and fissions have occurred during the evolution of the salmonids. White boxes display the fusion events, where the orthologous chromosomes for all species are named according to their Northern Pike linkage group ID, with .1 or .2 assigned for duplicate 1 or 2, as this species did not undergo the salmonid whole genome duplication. 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), where a previously fused metacentric divides into two acrocentric chromosomes. For example, the fusion between 4.2-20.1 at point (F) in the phylogeny is divided at Fi^2 prior to point (D). Fissions are detected as the original fused metacentric is divided in all species after that point in the phylogeny, and the newly separated arms are often paired in new fusions. The superscript number corresponds to the superscript in the fusion event. See Table 1 for full species names and Table 2 for all orthologous relationships. The phylogeny is adapted from [19].

The oldest and most conserved of the identified rearrangements were two fusion events (*Eluc* 9.2-17.1 and 23.1-24.1) occurring prior to the divergence of *Salmo* and *Salvelinus* and found in all species within this lineage (see F in Figure 4). Another metacentric fusion event at this same point in the phylogeny was also identified (*Eluc* 4.2-20.1) that is still present in both *Salmo* and *Salvelinus*, but not in *Oncorhynchus*, suggesting the occurrence of a fission event prior to the radiation of *Oncorhynchus* (Fi^2 in Figure 4). Occurring prior to the divergence of *Salvelinus* and *Oncorhynchus* (E in Figure 4), one fusion (*Eluc* 2.1-13.1) is found in all descendants of this lineage, and a second fusion is found in all descendants except for the *Oner/Oket* clade (fission at Fi^5 in Figure 4).

More recent rearrangements include four fusions prior to the radiation within the *Oncorhynchus* lineage (D in Figure 4), of which two are present in all *Oncorhynchus* species (*Eluc* 15.2-25.2 and 16.2-20.2), one that was lost by fission in the lineage leading to *Otsh/Okis* (*Eluc* 9.1-11.1; see Fi^3), and one that was lost by fission specifically in Chum Salmon (*Eluc* 17.2-22.1; see Fi^4). Within the clade containing *Otsh*/*Okis*/*Omyk*, one fusion occurred prior to the divergence of *Otsh*/*Okis* (*Eluc* 2.2-13.2; A in Figure 4) and four fusions occurred prior to the divergence of *Oner*/*Oket* (C 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 18 in Atlantic Salmon and 11 in Coho Salmon (Figure 4).

Some rearrangements are more complex and thus it is more difficult to unambiguously describe their history. For example, a three chromosome fusion in Atlantic Salmon occurred through a single fusion (*Eluc* 4.1-15.1) that either a) fused once prior to the divergence of Atlantic Salmon and divided by fission three different times (at C in Figure 4, once specifically in each of Brook Charr and Coho Salmon); or b) fused two times independently (once in Atlantic

Salmon only, and once prior to the divergence of *Otsh*/*Okis*/*Omyk* (at B in Figure 4) then lost by fission in Coho Salmon. It is not clear which of these possibilities is correct, but in Figure 4 we display the first scenario. After this metacentric fusion, an additional fusion occurred specifically in Atlantic Salmon, adding a third chromosome arm (*Eluc* 1.2 with the 15.1-4.1 metacentric). Additionally, three different fusions appeared to have occurred two independent times: *Eluc* 8.2-18.1 in *Otsh* and *Oner*; *Eluc* 12.2-21.2 in *Otsh* and *Oket*; and *Eluc* 7.2-25.1 in *Ssal* and *Omyk*. For each of these multiple independent origins, the alternate explanation only requiring a single fusion would require three independent fission events each. Although this is not entirely clear, we display the rearrangements that would require the fewest independent fusions/fissions in Figure 4.

Putative lineage-specific inversions

With the higher number of comparable markers, several regions that were inverted and 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.

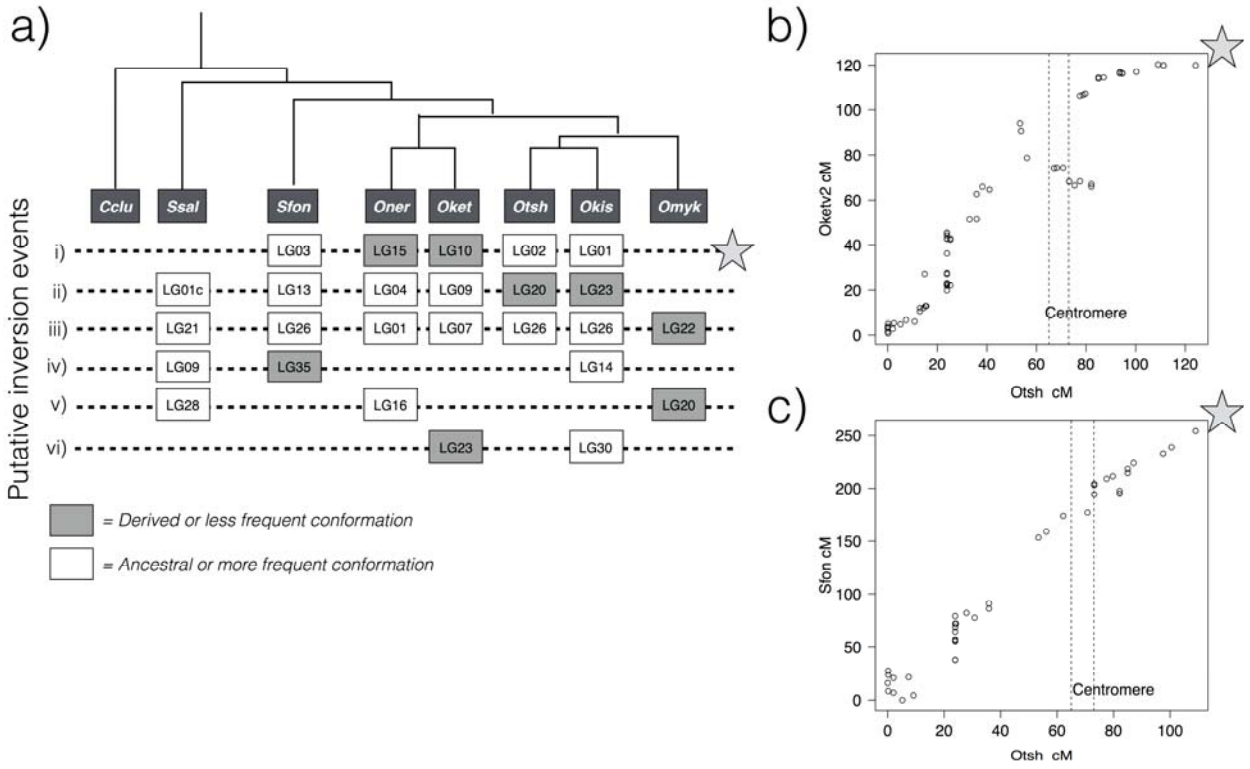

Figure 5. Putative conserved and species-specific inversions. a) The salmonid phylogeny is shown highlighting six different inversion events (labeled *i*-*vi*), each represented on a dotted line below the phylogeny. Per line (inversion event), white boxes containing a linkage group are displayed when the inversion can be identified in pair-wise comparisons against the less frequent (derived; grey) conformation. For the inversion (*vi*) with only one species showing each conformation, the derived (grey) conformation was randomly assigned. The light grey star denotes the *Oket*/*Oner*-specific inversion event (*i*) displayed in (b) between one ancestral conformation (*Otsh*) and derived (*Oket*). For comparison, two species with ancestral comparisons for this linkage group (*Sfon* and *Otsh*) are shown at (c). Predicted centromere positions previously identified in Chinook (*Otsh*) [35] are also shown in (b-c). Full names for species are defined in Table 1. The phylogeny is adapted from [19].

A striking putative inversion was identified in one of the metacentric chromosomes that is conserved across all evaluated salmonids (except Lake Whitefish; *Eluc* 9.2-17.1; Figure 4). An

inversion near the center of the linkage group is present in only the Sockeye Salmon/Chum Salmon clade, visible in comparisons with the maps of Coho Salmon, Chinook Salmon and Brook Charr (Figure 5a-b). As a result, the conformation observed in Chum Salmon 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 Sockeye Salmon/Chum Salmon, as there is a gap with no marker pairs available at the inverted locus in the Rainbow Trout linkage group.

To further characterize this inversion, centromere locations obtained from Chinook [35] were compared to the location of the inversion. This clearly indicates that the inverted region, containing 12 marker pairs mapping through seven different Rainbow Trout scaffolds, occurs between 49-82 cM of Chinook Ots02 (Figure 5b). The centromere for Ots02 is at 65-73 cM. Therefore, the inverted region probably contains the centromere and is considered a putative pericentric inversion in Chum and Sockeye Salmon. Near 78 cM of Ots02, the inversion is less clear as another Rainbow Trout scaffold is mapped against, identifying the end of the inverted region. A preliminary analysis of the presence of genes within this region by mapping Chinook markers against Rainbow Trout mRNA sequences [25] (only retaining high scoring, uniquely mapping markers) indicates four potential mRNA transcripts with top BLAST hits in NCBI annotated as: *cysteine-rich EGF-like protein*; *solute carrier family 2 facilitated glucose transporter member M*; and *tricohyalin-like*. Also, BLAST analysis of the 55 markers within this region in the Chinook linkage map identified additional evidence of gene content for nine markers. Regardless of the exact identity of genes within the region, this inverted region most likely contains genes and therefore is a disruption of synteny for the derived species.

In the comparisons, other inversions were also visible (Figure 5a; Additional File S3). Future exploration is warranted for the region described above, and the other inversions in Figure 5. More information will be obtainable as more genomes become available with assembled contigs in these regions for comparative genomics, for example an exploration of the exact location of the breakpoints around the inversion and whether they occur within genes or in possible regulatory regions around the genes.

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 (e.g. [4]). 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 4-6-fold; Table 3). For example, between Brook Charr and Chinook Salmon, 907 marker pairs were identified using MAPCOMP, whereas direct mapping identified 190 pairs.

In addition to the published Rainbow Trout genome, MAPCOMP was tested using the available Atlantic Salmon genome as the intermediate reference for pairing markers between maps (Table 3). Results using either genome were similar, 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 pairings of markers (on average 1.4-fold). This could be due to slight differences in contiguity of the two genomes. However, as the general results remain the same, the published Rainbow Trout genome [25] was used for the present analysis.

Table 3. MAPCOMP tested using two different intermediate genomes and compared to results from reciprocal BLAST. The number and percent 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 (see Figure 2 for details on pairing). Also shown in 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 *Omyk* or *Ssal* genome assemblies. *N/A* values are present as Brook Charr is not paired against itself.

Species	Total Markers	Map to *Omyk* genome: No. (%)	No. Marker Pairs	Map to *Ssal* genome: No. (%)	No. Marker Pairs	Recip. best-hit BLAST
Okis	5377	3873 (72%)	813	2856 (53%)	1068	182
Otsh	6352	4663 (73%)	907	3472 (54%)	1162	190
Omyk	955	837 (87%)	300	626 (65%)	411	30
Oket	6119	4150 (67%)	795	3139 (51%)	1049	205
Oner	6262	4034 (64%)	771	3138 (50%)	1061	209
Sfon	3826	2321 (60%)	N/A	2454 (64%)	N/A	N/A
Ssal	5650*	2776 (49%)	619	3434 (60%)	1041	208
Cclu	3438	1156 (33%)	346	1185 (34%)	609	111

*from EST sequences

Discussion

Linkage maps have many applications, including QTL analysis, assisting genome assembly and comparative genomics. With advances in sequencing technology and techniques [9], high quality and dense linkage maps are increasingly available for many species, including non-model species. QTL analysis is essential for identifying regions of the genome linked to important agricultural [48], medical [49], or ecological and evolutionary traits [50]. Dense linkage maps can also be used for assisting genome assembly [7] or for comparative genomics, allowing for information transfer from model to related non-model organisms [3]. They are also useful for

cross-species QTL comparisons [16] to understand genome function and evolution, such as that after a whole genome duplication [4].

Salmonids are a valuable taxon for studying genome duplication. Recently, Kodama et al. [4] characterized some of the rearrangements and positioned them in the phylogeny of salmonids based on conservation. This has indicated that structural rearrangements have occurred throughout the evolution of the salmonids, retained from different points in evolutionary history. Here, we further demonstrate the diversity of these rearrangements, identifying all orthologous arms and the most likely timings of the rearrangement events throughout salmonid evolutionary history. As rediploidization efficiency is reduced in metacentric chromosomes, it is valuable to characterize the history of such rearrangements. This occurs by enhanced homologous recombination between distal regions of homeologs in metacentric chromosomes (residual tetrasomy) [29,30]. Fewer tetravalents form at meiosis when fewer metacentrics exist, as has been demonstrated in Rainbow Trout (1n = 20 metacentrics) relative to Brown Trout *Salmo trutta* (1n = 10 metacentrics) [31]. This is important to consider for genome assembly since there will be less divergence between homeologs when the homeolog is fused into a metacentric chromosome. Among the salmonids, species often have different metacentric and acrocentric chromosomes, with the exception of a few conserved fusion events (Figure 4), and therefore as more genomes are assembled for the different salmonids, we will achieve different regions of completeness among the different species. Once the Brook Charr genome (1n = 8 metacentrics) is assembled, additional and different information will be gained on rediploidization to complement that provided from Rainbow Trout and Atlantic Salmon. This highlights the importance of understanding the exact orthology between chromosome arms and identity and conservation of metacentrics across the salmonid lineage.

Fusions, Inversions, and Evolution

Chromosomal rearrangements include chromosome fusion or fission, region amplification or deletion, segment inversion, or segment translocation between non-homologous chromosomes [51]. 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 gradual process of generating the present day salmonid karyotypes, with fusions occurring at each step along the diversification process, and then many fusions occurring specifically within each species. 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, such as Brook Charr with its two species-specific fusions, whereas others have more species-specific fusions. Furthermore, even within a clade, large differences in the number of species-specific fusions exist. For example, in the Sockeye Salmon/Chum Salmon clade, there are ten and three species-specific fusions each, respectively. Similarly, within the Coho Salmon/Chinook Salmon clade, there are six and 11 species-specific fusions each, respectively.

Inversions can occur when a segment of a chromosome is cut out by two breakpoints and then reinserted in the opposite orientation [52]. 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 [51-54]. For example, introgression rates in sunflower hybrids were 50% lower across chromosomes with rearrangements than they were in collinear chromosomes [55]. Additionally, lower recombination rates were observed in heterokaryotypic regions of Yellowstone Cutthroat Trout (*O. clarkii*) and Rainbow Trout hybrids compared to collinear regions [56]. Recombination

suppression may allow for conservation of fitness-related gene complexes that are locally adapted, or involved in reproductive isolation (discussed in [56]). Robertsonian rearrangements, for example fusions and fissions, although having effects on rediploidization efficiency, as discussed above, have less of an effect on recombination rates than do rearrangements affecting synteny (e.g. inversions) [51,56].

In salmonids, one of the metacentric chromosomes conserved in all of the evaluated high-density maps, except possibly *Coregonus* (*Eluc* 9.2-17.1; Figure 4), has a putative pericentric inversion in the Chum Salmon/Sockeye Salmon clade (Figure 5). A high-density genetic map for Pink Salmon, also within this clade [19], will indicate the broader conservation of this putative inversion. Future genome sequencing of either Chum Salmon or Sockeye Salmon will be useful for validating this inversion [57]. Assembly of this region for validation may however be more challenging, depending on physical proximity of this region to the centromere, given the difficulties in assembling centromeric regions [57]. Preliminary results indicate that several genes are present in this Chum Salmon/Sockeye Salmon inversion. More sequence information will allow for an identification of the location of breakpoints of the inversion, to identify whether they occur in or near genes. Many inversions are only visible in a subset of the species. A complete understanding of these inversions will either depend on denser maps, or future genome assemblies. Further exploration of structural variation and conservation of these variants across lineages will contribute to our understanding of the mechanisms underlying speciation in salmonids.

MAPCOMP: *Potential and Limitations*

By using the information of 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 [3]. Although RADseq-based linkage maps often provide two orders 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 [3]. 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. [43], 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 [58], although these often require shared markers between maps. Another similar approach was used by Amores et al. [13] 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 [59]. 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 [60,61].

In two orthology designation cases, it was not clear which homeolog was fused in *Coregonus* (see Figure 4; *Eluc* 2.1 or 2.2 and 6.1 or 6.2). It is unclear why this occurred, but this was rare and orthology is typically unambiguous (Additional File S3). A potential limitation of MAPCOMP is the sequence similarity and synteny required between the maps to be compared as well as the sequence similarity to the reference genome; further use of MAPCOMP on a more diverse set of taxa will provide more insight on this.

When combined with high quality maps from other species, MAPCOMP is also useful for evaluating map construction in new species, or even identifying problems with specific marker types, such as the removed male-specific markers in the consensus Brook Charr map (*data not shown*). This was expected for salmonids [4] and was clear when using MAPCOMP on this subset of markers. Poor positioning of male-specific markers across all linkage groups may occur due to almost complete crossover interference within male salmonids during meiosis [3].

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 [43] 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/

Material and Methods

Brook Charr genetic map

Animals

Full details regarding the experimental mapping family were reported previously [41,62]. Grandparents (F_0) 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 selected F_2 offspring and F_1 parents by high salt extraction [63] with an additional RNase A digestion step (QIAGEN), as previously reported [41]. 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).

RADseq [9] was performed as per methods previously outlined [10], and described in full in [64]. 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 sequenced on a single lane of HiSeq2000 at Génome Québec Innovation Centre (McGill University, Montréal). In order to obtain deeper sequencing of the parents, each parent individual was sequenced on Ion Torrent at the sequencing platform at IBIS (the Institut de Biologie Intégrative et des Systèmes, Université Laval, Québec City). Due to the use of different sequencing platforms for F_1 and F_2 individuals, 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 [65]. Reads were de-multiplexed, by barcodes, and quality trimmed to 80bp using STACKS v.1.32 [12,66] module's *process_radtags*. We used the ploidy-informed empirical procedure developed by Ilut et al. [67] to optimize *de novo* assembly. Sequence similarity was explored to find the optimum clustering threshold, which is highly important for pseudotetraploid salmonids *de novo* assembly (see Additional File S4 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. We used the *genotypes* module of STACKS to output markers along their allelic state and raw genotypes. The markers were translated by *genotypes_summary.R* into fully- or semi-informative markers types, specifically the four types of markers that our outbreeding design permitted: *ab* x *ac*, *ab* x *ab*, *ab* x *aa* and *aa* x *ab* [44].

Premapping quality control

Several steps of quality control were performed based on the recommendations in van Ooijen and Jansen [68]. Premapping 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 in stackr v.0.2.1 [69]. 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 the markers contribute equally to the construction of the map, because linkage phases change across loci [68]. 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 genetic linkage map was first built in JOINMAP (v4.1; [70]) using the pseudo-testcross approach strategy [68,71] that only uses the markers segregating in a uni-parental configuration (i.e. *ab* x *ac* and *ab* x *ab* markers are excluded). Second, the consensus map 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 [68,72]. The pseudo-testcross maps were used as confirmations. Separate maximum likelihood maps were generated for each parent, and only the female map was retained, as is typical for salmonid mapping studies [4]. Markers were grouped with the independence 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 grouping tree branches with stable marker numbers over increasing consecutive LOD values. This of LGs number 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.

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 in Ooijen and Jansen [68], errors in ordering and genotyping along marker exclusion followed these criteria: (i) oversized LG, which can occur with high marker numbers, (ii) incidence of improbable genotypes (e.g. double recombinants [14], also inspected using the *countXO* function of R/qtl [73]), (iii) drastic changes of orders and (iv) low levels of fit

or high levels of stress. Maps were inspected for distorsion before and after manual exclusion of markers. Mapping distances (cM) were calculated using the Haldane mapping function.

MapComp

Map comparison through intermediate Rainbow Trout reference genome

In order to compare the *S. fontinalis* map to other salmonid maps, marker name, sequence, linkage group and cM position were obtained from published datasets (see Data Accessibility for instructions for obtaining the published data and formatting). Comparisons of linkage group composition and order were investigated for Chinook Salmon *O. tshawytscha* [35], Chum Salmon *O. keta* [30], and Rainbow Trout *O. mykiss* [36,37] Sockeye Salmon *O. nerka* [34,74], Coho Salmon *O. kisutch* [4], Atlantic Salmon *Salmo salar* [32], Lake Whitefish *C. clupeaformis* [38], and the salmonid WGD sister outgroup Northern Pike *Esox lucius* [27], as shown in Table 1.

The basic workflow of MapComp is in Figure 2. First, all marker sequences were combined into a single fasta file and mapped to a reference genome (either Rainbow Trout published scaffolds (http://www.genoscope.cns.fr/trout/data/ added 28-Apr-2014) [25] using *BWA mem* [75]. We also tested MapComp with the current version of the Atlantic Salmon genome AGKD00000000.4 (NCBI) to evaluate the effect of a different reference genome on marker pairing, but we restricted all other analyses to the results obtained using the Rainbow Trout genome [25]. 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. Linkage group pairs (one per species) containing more than five marker pairs were shaded darker on the surrounding grid for ease of visualization. The pipeline developed for MAPCOMP is available at https://github.com/enormandeau/mapcomp/.

To identify homeologs, the stringency of the MAPCOMP parameters were relaxed to allow for multiple hits from the non-duplicated *E. lucius* map against the Rainbow Trout reference genome intermediate, as each marker could be present in at least duplicate in the Rainbow Trout genome (MAPQ $\geq$ 2 and 1 or more match allowed).

Identification of orthology and homeology between chromosome arms

Orthology of chromosome arms between Chinook Salmon and Coho Salmon maps [4] was confirmed using MAPCOMP. Chinook Salmon and Coho Salmon were then individually compared with the Brook Charr map to identify corresponding chromosome arms in Brook Charr. The orthology designation was started with Chinook Salmon and Coho Salmon as these two closely-related species were integrated previously using homologous markers [4,35]. Once these orthology relationships were obtained, the Brook Charr map was compared with 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 identified in the same way, and the original Northern Pike linkage groups were randomly given a .1 or .2 designation to represent the duplicated chromosome number.

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 LG [14]. 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 also considered (i.e. when an inversion was identified within multiple species within a lineage). For centromere identification, the region of interest was obtained from MAPCOMP results to compare with centromere positions in previous datasets to define 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 [19]. The analysis of metacentric conservation was based on the analysis of conservation in Coho, Chinook, Rainbow Trout and Atlantic Salmon [4], but re-analyzed using MAPCOMP and additional maps in the present study (i.e. Chum, Sockeye, Brook Charr and Lake Whitefish). Identification of whether a chromosome was completely present, we required evidence that both arms of the chromosome were matched between species to ensure that missing data did not result in more chromosomes being falsely identified as metacentric only because the second arm was not represented in the compared species.

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 Guillaume Côté, Gaétan Légaré, Clément Rougeux and Brian Boyle from the Institut de Biologie Intégrative et des Systèmes (IBIS) for help with DNA extraction and library preparation. We also thank Anne-Marie Dion-Côté and Eric Rondeau for comments on an earlier version of the manuscript. BJGS was supported by an NSERC postdoctoral fellowship. Thanks to the members of the Bernatchez Laboratory for discussion and support.

Data Availability

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

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

References

1. Brunet FG, Roest Crollius H, Paris M, Aury J-M, Gibert P, Jaillon O, et al. Gene loss and evolutionary rates following whole-genome duplication in teleost fishes. Molecular Biology and Evolution. 2006;23: 1808–1816. doi:10.1093/molbev/msl049

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

3. Naish KA, Phillips RB, Brieuc MSO, Newton LR, Elz AE, Park LK. Comparative genome mapping between Chinook Salmon (*Oncorhynchus tshawytscha*) and Rainbow Trout (*O. mykiss*) based on homologous microsatellite loci. G3 - Genes|Genomes|Genetics. Genetics Society of America; 2013;3: 2281–2288. doi:10.1534/g3.113.008003

4. Kodama M, Brieuc MSO, Devlin RH, Hard JJ, Naish KA. 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. Genetics Society of America; 2014;4: 1717–1730. doi:10.1534/g3.114.012294

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

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

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

8. Amores A, Catchen J, Nanda I, Warren W, Walter R, Schartl M, et al. A RAD-tag genetic map for the Platyfish (*Xiphophorus maculatus*) reveals mechanisms of karyotype evolution among teleost fish. Genetics. 2014;197: 625–641. doi:10.1534/genetics.114.164293

9. Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, et al. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE. 2008;3: e3376. doi:10.1371/journal.pone.0003376

10. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE. 2011;6: e19379. doi:10.1371/journal.pone.0019379

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

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

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

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

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

16. Larson WA, McKinney GJ, Limborg MT, Everett MV, Seeb LW, Seeb JE. Identification of Multiple QTL Hotspots in Sockeye Salmon (Oncorhynchus nerka) Using Genotyping-by-Sequencing and a Dense Linkage Map. J Hered. 2015. doi:10.1093/jhered/esv099

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

18. Davidson WS, Koop BF, Jones SJM, Iturra P, Vidal R, Maass A, et al. Sequencing the genome of the Atlantic Salmon (*Salmo salar*). Genome Biol. 2010;11: 403. doi:10.1186/gb-2010-11-9-403

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

20. Macqueen DJ, Johnston IA. 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. 2014;281: 20132881–20132881. doi:10.1098/rspb.2013.2881

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

22. Phillips RB, Konkol NR, Reed KM, Stein JD. Chromosome painting supports lack of homology among sex chromosomes in *Oncorhynchus*, *Salmo*, and *Salvelinus* (Salmonidae). Genetica. Kluwer Academic Publishers; 2001;111: 119–123. doi:10.1023/A:1013743431738

23. Woram RA, Gharbi K, Sakamoto T, Høyheim B, Holm L-E, Naish K, et al. Comparative

genome analysis of the primary sex-determining locus in salmonid fishes. Genome Research. Cold Spring Harbor Lab; 2003;13: 272–280. doi:10.1101/gr.578503

24. Lubieniecki KP, Lin S, Cabana EI, Li J, Lai YYY, Davidson WS. Genomic instability of the sex-determining locus in Atlantic Salmon (*Salmo salar*). G3 - Genes|Genomes|Genetics. 2015. doi:10.1534/g3.115.020115

25. Berthelot C, Brunet F, Chalopin D, Juanchich A, Bernard M, Noël B, et al. The Rainbow Trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. Nat Comms. 2014;5: 3657. doi:10.1038/ncomms4657

26. Leong JS, Jantzen SG, Schalburg von KR, Cooper GA, Messmer AM, Liao NY, et al. *Salmo salar* and *Esox lucius* full-length cDNA sequences reveal changes in evolutionary pressures on a post-tetraploidization genome. BMC Genomics. 2010;11: 279. doi:10.1186/1471-2164-11-279

27. Rondeau EB, Minkley DR, Leong JS, Messmer AM, Jantzen JR, Schalburg von KR, et al. The genome and linkage map of the Northern Pike (*Esox lucius*): conserved synteny revealed between the salmonid sister group and the Neoteleostei. PLoS ONE. 2014;9: e102089. doi:10.1371/journal.pone.0102089

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

29. Allendorf FW, Bassham S, Cresko WA, Limborg MT, Seeb LW, Seeb JE. Effects of crossovers between homeologs on inheritance and population genomics in polyploid-derived salmonid fishes. J Hered. 2015;106: 217–227. doi:10.1093/jhered/esv015

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

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

32. Lien S, Gidskehaug L, Moen T, Ben J Hayes, Berg PR, Davidson WS, et al. 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. BioMed Central Ltd; 2011;12: 615. doi:10.1186/1471-2164-12-615

33. Gonen S, Lowe NR, Cezard T, Gharbi K, Bishop SC, Houston RD. Linkage maps of the Atlantic Salmon (*Salmo salar*) genome derived from RAD sequencing. BMC Genomics. 2014;15: 166. doi:10.1186/1471-2164-15-166

34. Everett MV, Miller MR, Seeb JE. Meiotic maps of Sockeye Salmon derived from massively parallel DNA sequencing. BMC Genomics. BioMed Central Ltd; 2012;13: 521.

doi:10.1186/1471-2164-13-521

35. Brieuc MSO, Brieuc MSO, Waters CD, Waters CD, Seeb JE, Seeb JE, et al. A dense linkage map for Chinook Salmon (*Oncorhynchus tshawytscha*) reveals variable chromosomal divergence after an ancestral whole genome duplication event. G3 - Genes|Genomes|Genetics. 2014;4: 447–460. doi:10.1534/g3.113.009316

36. Miller MR, Brunelli JP, Wheeler PA, Liu S, Rexroad CE, Palti Y, et al. A conserved haplotype controls parallel adaptation in geographically distant salmonid populations. Mol Ecol. 2012;21: 237–249. doi:10.1111/j.1365-294X.2011.05305.x

37. Palti Y, Vallejo RL, Gao G, Liu S, Hernandez AG, Rexroad CE III, et al. 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. Public Library of Science; 2015;10: e0138435. doi:10.1371/journal.pone.0138435

38. Gagnaire P-A, Normandeau E, Pavey SA, Bernatchez L. Mapping phenotypic, expression and transmission ratio distortion QTL using RAD markers in the Lake Whitefish (*Coregonus clupeaformis*). Mol Ecol. 2013;22: 3036–3048. doi:10.1111/mec.12127

39. Woram RA, McGowan C, Stout JA, Gharbi K, Ferguson MM, Hoyheim B, et al. 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. 2004;47: 304–315. doi:10.1139/g03-127

40. Timusk ER, Ferguson MM, Moghadam HK, Norman JD, Wilson CC, Danzmann RG. 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. BioMed Central Ltd; 2011;12: 68. doi:10.1186/1471-2156-12-68

41. Sauvage C, Vagner M, Derôme N, Audet C, Bernatchez L. Coding gene single nucleotide polymorphism mapping and quantitative trait loci detection for physiological reproductive traits in Brook Charr, *Salvelinus fontinalis*. G3 - Genes|Genomes|Genetics. Genetics Society of America; 2012;2: 379–392. doi:10.1534/g3.111.001867

42. Palti Y, Genêt C, Gao G, Hu Y, You FM, Boussaha M, et al. A second generation integrated map of the Rainbow Trout (*Oncorhynchus mykiss*) genome: Analysis of conserved synteny with model fish genomes. Mar Biotechnol. Springer-Verlag; 2011;14: 343–357. doi:10.1007/s10126-011-9418-z

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

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

45. Danzmann RG, Davidson EA, Ferguson MM, Gharbi K, Koop BF, Høyheim B, et al. Distribution of ancestral proto-Actinopterygian chromosome arms within the genomes of 4R-derivative salmonid fishes (Rainbow Trout and Atlantic Salmon). BMC Genomics. 2008;9: 557. doi:10.1186/1471-2164-9-557

46. Phillips RB, Keatley KA, Morasch MR, Ventura AB, Lubieniecki KP, Ben F Koop, et al. 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. BioMed Central Ltd; 2009;10: 46. doi:10.1186/1471-2156-10-46

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

48. Yáñez JM, Houston RD, Newman S. Genetics and genomics of disease resistance in salmonid species. Front Genet. 2014;5: 415. doi:10.3389/fgene.2014.00415

49. Mackay TFC, Stone EA, Ayroles JF. The genetics of quantitative traits: challenges and prospects. Nat Rev Genet. 2009;10: 565–577. doi:10.1038/nrg2612

50. Naish KA, Hard JJ. Bridging the gap between the genotype and the phenotype: linking genetic variation, selection and adaptation in fishes. Fish and Fisheries. Blackwell Publishing Ltd; 2008;9: 396–422. doi:10.1111/j.1467-2979.2008.00302.x

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

52. Kirkpatrick M. How and why chromosome inversions evolve. Plos Biol. Public Library of Science; 2010;8: e1000501. doi:10.1371/journal.pbio.1000501

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

54. Noor MA, Grams KL, Bertucci LA, Almendarez Y, Reiland J, Smith KR. The genetics of reproductive isolation and the potential for gene exchange between *Drosophila pseudoobscura* and *D. persimilis* via backcross hybrid males. Evolution. 2001;55: 512–521.

55. Rieseberg LH, Whitton J, Gardner K. Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. Genetics. 1999;152: 713–727.

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

57. Suzuki T, Tsurusaki Y, Nakashima M, Miyake N, Saitsu H, Takeda S, et al. Precise

detection of chromosomal translocation or inversion breakpoints by whole-genome sequencing. Journal of Human Genetics. Nature Publishing Group; 2014;59: 649–654. doi:10.1038/jhg.2014.88

58. Fang Z, Polacco M, Chen S, Schroeder S, Hancock D, Sanchez H, et al. cMap: the comparative genetic map viewer. Bioinformatics. 2003;19: 416–417. doi:10.1093/bioinformatics/btg012

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

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

61. Pante E, Abdelkrim J, Viricel A, Gey D, France SC, Boisselier MC, et al. Use of RAD sequencing for delimiting species. Heredity. 2014;114: 450–459.

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

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

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

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

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

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

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

69. Gosselin T, Bernatchez L. stackr [Internet]. Available: https://github.com/thierrygosselin/stackr

70. van Ooijen JW. JoinMap ® 4. 2006;: Software for the calculation of genetic linkage ma.

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

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

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

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

75. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25: 1754–1760. doi:10.1093/bioinformatics/btp324

76. Gregory TR. Animal Genome Size Database [Internet]. Available: http://www.genomesize.com

Supporting Information Captions

Additional File S1. Descriptive statistics for *S. fontinalis* sex-specific maps.

Additional File S2. Female Brook Charr *S. fontinalis* linkage map. Includes species name, linkage group, cM position, marker name and sequence.

Additional File S3. 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) Coho Salmon (*Okis*), 2) Chinook Salmon (*Otsh*), 3) Rainbow Trout (*Omyk*), 4) Chum Salmon (*Oket*), 5) Sockeye Salmon (*Oner*), 6) Atlantic Salmon (*Ssal*), and 7) Lake Whitefish (*Cclu*).

Additional File S4. Mapping software parameters and bioinformatics pipeline overview. STACKS parameters and an outline of bioinformatics steps used to generate the Brook Charr linkage map.