Comparative genomic analyses and a novel linkage map for cisco (*Coregonus artedi)* provides insight into chromosomal evolution and rediploidization across salmonids

Danielle M. Blumstein[1], Matthew A. Campbell[2], Matthew C. Hale[3], Ben J. G. Sutherland[4], Garrett J. McKinney[5], Wendylee Stott[6], Wesley A. Larson[7, 1]

[1] College of Natural Resources, University of Wisconsin-Stevens Point
[2] University of Alaska Museum of the North, University of Alaska Fairbanks, Fairbanks, AK, 99775, USA
[3] Department of Biology, Texas Christian University, Fort Worth, TX, 76122
[4] Pacific Biological Station, Fisheries and Oceans Canada, Nanaimo, V9T 6N7, British Columbia, Canada
[5] National Marine Fisheries Service, Northwest Fisheries Science Center
[6] U.S. Geological Survey, Great Lakes Science Center, Ann Arbor, MI
[7] U.S. Geological Survey, Wisconsin Cooperative Fishery Research Unit, College of Natural Resources, University of Wisconsin-Stevens Point

Data deposition: Raw sequence data has been uploaded to SRA under BioProject PRJNA555579

Corresponding author: Wesley Larson (e-mail: wlarson@uwsp.edu, phone: 715-346-3150, office mailing address: U.S. Geological Survey, Wisconsin Cooperative Fishery Research Unit, College of Natural Resources, University of Wisconsin)

Abstract

Whole-genome duplication (WGD) is hypothesized to be an important evolutionary mechanism that can facilitate adaptation and speciation. Genomes that exist in states of both diploidy and residual tetraploidy simultaneously are of particular interest, as understanding the mechanisms that maintain this mosaic of ploidy after WGDs may provide important insights into evolutionary processes. Salmonids are a useful model to study the effect of WGDs due to the extensive diversity that has evolved following an ancestral autotetraploidization event. Although most of the salmonid genome has reverted to diploidy following WGD, approximately 25% of the chromosomes continue to exist in a state of residual tetrasomy. In this study, we generate a novel linkage map for cisco (*Coregonus artedi*) and conduct comparative genomic analyses to refine our understanding of chromosomal fusion/fission history across salmonids. Additionally, we propose a new naming strategy, protokaryotypes, that facilitates comparisons across Salmonids by standardizing naming to ancestral chromosomes from northern pike (*Esox lucius*). The female linkage map for cisco contains 20,450 loci, 3,383 of which are likely contained within residually tetraploid regions of the genome. Comparative genomic analysis revealed that patterns of residual tetrasomy are generally conserved across species, but there is interspecific variation in the relative extent of the residual tetrasomy. Additionally, with the current reference genomes available, we only find evidence of residual tetrasomy in seven of the eight chromosomes that have been previously hypothesized to show this pattern. This interspecific variation may have important implications for understanding salmonid evolutionary histories and informing future conservation efforts.

Keywords: coregonines; comparative genomics; linkage mapping; residual tetrasomy; Salmonidae; whole genome duplication

Introduction

The evolutionary significance of whole-genome duplications (WGDs) has been intensively debated for decades (e.g; Ohno 1970; Santini et al. 2009; Wood et al. 2009; Zhan et al. 2014; Mayrose et al. 2015; Van de Peer et al. 2017). Multiple studies have hypothesized that WGD is an important evolutionary mechanism that can facilitate adaptation on short- and long-term evolutionary timescales (Ohta 1989, Selmecki et al. 2015; Van de Peer et al. 2017). For example, genes found in polyploid regions are able to gain new function (i.e., neofunctionalization) without the consequences of deleterious mutations affecting the main function of the second gene copy. This may facilitate adaptive molecular divergence and the evolution of new phenotypes (Wittbrodt et al 1998; Wendel 2000; Rastogi and Liberles 2005). However, other studies have hypothesized that WGD presents significant challenges for meiosis and mitosis (Hollister 2015) and may not have as much of an effect on evolution as originally thought (Mayrose et al. 2011; Arrigo and Barker 2012; Vanneste et al. 2014; Clarke et al. 2016). A consensus is therefore yet to be reached on the evolutionary impact of WGD relative to other evolutionary forces.

Conducting genetic studies on organisms having underwent relatively recent WGD has been challenging due to the inability to differentiate alleles and sequences from the same chromosome (homologs) from those on the duplicated chromosome (homeologs) (Limborg et al. 2016b). Fortunately, approaches leveraging gamete manipulation, high sequencing coverage, and long read sequencing have improved our ability to characterize duplicated regions. For example, linkage mapping with haploids and doubled haploids has facilitated analysis of duplicated regions in salmonids (Brieuc et al. 2014; Kodama et al. 2014; Lien et al. 2016; Waples et al. 2016). Further, long-read sequencing technologies, such as PacBio and Oxford Nanopore, have made it possible to assemble complex genomes with convoluted duplication histories (Kyriakidou et al. 2018). These technological advances have revolutionized our ability to understand genomic architecture in that have adaptively radiated to form dozens of species following an ancestral WGD in lineages such as salmonids (Lien et al. 2016; Robertson et al. 2017; Campbell et al. 2019) and many plant species (Alix et al. 2017).

Salmonids are derived from an ancestral species that underwent a WGD ~100 million years ago (Ss4R, Allendorf and Thorgaard 1984; Berthelot et al. 2014; Macqueen and Johnston 2014; Lien et al. 2016) and have since diversified into a broad array of both ecologically and genetically distinct taxa. The Salmonidae family is comprised of three subfamilies: Salmoninae (salmon, trout, and char), Thymallinae (graylings), and Coregoninae (whitefish and ciscoes) (Norden 1961), and diversification of these subfamilies post-dates the Ss4R, occurring 40-50 million years ago (Campbell et al. 2013; Macqueen and Johnston 2014). Phylogenetic analysis has thus far revealed that the majority of the salmonid genome returned to a diploid inheritance state prior to the basal split of each of the three subfamilies (Robertson et al. 2017). The rediploidization process is still incomplete and approximately 20-25% of each salmonid genome still shows signals of tetrasomic inheritance (i.e., residual tetrasomy, or the recombination

between homeologs that results in the exchange of alleles between homeologous chromosomes (Allendorf et al. 2015; Lien et al. 2016; Robertson et al. 2017)).

Evidence for residual tetrasomy in the salmonids was found using allozyme studies in experimental crosses (Allendorf and Danzmann 1997) and, more recently, by linkage maps, sequenced genomes, and sequence capture (Kodama et al. 2014; McKinney et al. 2017; Robertson et al. 2017; Christensen et al. 2018b; Pearse et al. 2018). High-density linkage maps that include both duplicated and non-duplicated markers have revealed that eight pairs of homeologous chromosomes display evidence of conserved residual tetraploidy despite independent fusion and fission events, as well as general evolutionary history across species (Brieuc et al. 2014; Kodama et al. 2014; Sutherland et al. 2016). These chromosomes, referred to as the "magic eight," have been found in linkage mapping studies of coho salmon *Oncorhynchus kisutch* (Kodama et al. 2014), Chinook salmon *O. tshawytscha* (Brieuc et al. 2014; McKinney et al. 2016; McKinney et al. 2019), pink salmon *O. gorbuscha* (Tarpey et al. 2017), chum salmon *O. keta* (Waples et al. 2016), and sockeye salmon *O. nerka* (Larson et al. 2015). Additionally, previous mapping studies have revealed that at least one of the two homeologs exhibiting residual tetraploidy in these pairs is within a fusion, suggesting that chromosomal structure may play a role in maintenance of residual tetraploidy (Brieuc et al. 2014; Kodama et al. 2014; Sutherland et al. 2016). Analysis of sequenced genomes for Atlantic salmon (*Salmo salar*) and rainbow trout (*O. mykiss*) has revealed similar patterns, although in these studies only seven of these pairs were identified as displaying clear signals of conserved residual tetraploidy (Lien et al. 2016; Campbell et al. 2019). Sequence capture analysis also suggests that seven of the pairs display conserved signals of residual tetrasomy across species (Robertson et al. 2017). This led to the definition of two types of homeologous regions: 1) ancestral ohnologue resolution (AORe) regions with relatively low sequence similarity with ancestral homeologs that likely rediploidized prior to species diversification; and 2) lineage ohnologue resolution (LORe) regions with high sequence similarity among homeologs, likely maintained by residual tetraploidy (Robertson et al. 2017). Collectively, these studies have provided important insights into the salmonid WGD, but many unresolved questions related to the rediploidization process and maintenance of residual tetrasomy still exist.

Over the past decade, a proliferation of genomic resources has occurred within salmonids. Currently, linkage maps that include duplicated regions are available for five *Oncorhynchus* species, and genome assemblies are available for grayling *Thymallus thymallus* (Savilammi et al. 2019), Atlantic salmon (Lien et al. 2016), Arctic char *Salvelinus alpinus* (Christensen et al. 2018b), rainbow trout (Pearse et al. 2018), and Chinook salmon (Christensen et al. 2018a). These resources permit investigation into the processes of rediploidization and residual tetrasomy across the salmonid family. However, the genome resources for some of the lineages, such as the Coregoninae subfamily, remain undercharacterized. In the current study, we develop a high-density linkage map for North American cisco (*Coregonus artedi*), the first haploid linkage map for the Coregoninae subfamily, and analyze existing genomic resources for other salmonids with the goal of investigating patterns of residual tetrasomy and chromosomal

129 fusion and fission history across the salmonid family, with particular focus on the coregonines.
130 Our results suggest that (1) interspecific patterns of residual tetrasomy differ more than
131 previously observed; (2) binary definitions of chromosome ploidy status may not adequately
132 capture variation within and among species; (3) linkage maps and sequenced genomes identify
133 slightly different patterns regarding residual tetrasomy; and (4) a large number of fissions and
134 fusions are specific to the base of the Coregoninae subfamily, and species-specific fusions of this
135 lineage are rare. This study uses new and existing resources to conduct the most comprehensive
136 analysis of residual tetrasomy across the salmonid phylogeny to date, revealing important
137 patterns that likely impacted the evolutionary history of this diverse and important taxonomic
138 group.
139
140 **Methods**
141 *Experimental crosses for linkage mapping*
142 Genotypes from diploid and haploid families were used to build sex-specific linkage
143 maps. Diploid crosses were constructed from cisco collected in northern Lake Huron (45º
144 58'51.6" N -84 º19'40.8" W, USA) during spawning season (November 2015) by U. S.Fish and
145 Wildlife Service crews using standardized gill net assessment methods. Gametes for haploid
146 crosses were collected following the same methods, from the same location and month but in
147 2017. Gametes were extracted from mature fish and eggs were combined directly with sperm to
148 produce diploid crosses or with sperm that had been irradiated with 300,000 $\mu J/cm^2$ UV light for
149 two minutes to break down the DNA to produce haploid crosses. UV irradiation leaves the sperm
150 intact so that the egg can be activated but no paternal genetic material is contributed (i.e.,
151 gynogenesis, Chourrout 1982), resulting in haploid embryos developing with maternal genetic
152 material only. Crosses were made in the field and transported to the U. S. Geological Survey-
153 Great Lakes Science Center, Ann Arbor, Michigan (USA) for rearing. Tissue samples (fin clips)
154 were taken from adult parents, from offspring of the diploid crosses at age two, and from
155 haploids approximately 50 days post fertilization. All samples were preserved in a combination
156 of 95% ethanol and 5% EDTA and sent to the University of Wisconsin, Stevens Point Molecular
157 Conservation Genetics Lab for processing. Laboratory and field collections were conducted
158 under the auspices of the U.S. Fish and Wildlife Service and U.S. Geological Survey-Great
159 Lakes Science Center and all necessary animal care and use protocols were filed by these
160 agencies.
161
162 *DNA extraction and RAD-sequencing library preparation*
163 DNA was extracted using DNeasy 96 Blood and Tissue Kits (Qiagen, Valencia,
164 California) per the manufacturer's instructions. Quality and quantity of the extracted genomic
165 DNA was measured using the Quant-iT PicoGreen double-stranded DNA Assay (Life
166 Technologies) with a plate reader (BioTek). To confirm ploidy of haploid samples, parents and
167 offspring were genotyped using six polymorphic microsatellite loci known to occur at diploid
168 sites developed by Angers et al. (1995), Patton et al. (1997), and Rogers et al. (2004), and

169 individuals were classified as haploids if only a single allele was present at all loci. The
170 probability of not detecting a diploid if a diploid was present is ~1.09% based on microsatellite
171 heterozygosity in the parental population (unpublished data, Wendylee Stott).
172 Genomic DNA from diploids and confirmed haploids was prepared for RAD sequencing
173 using the SbfI restriction enzyme following the methods outlined in Ali et al. (2016) except
174 shearing restriction digested DNA was done with NEBNext® dsDNA Fragmentase® (New
175 England Biolabs, Inc) instead of sonication. DNA was then purified and indexed using
176 NEBNext® Ultra™ DNA Library Prep Kit for Illumina® per the manufacturer's instructions
177 (New England Biolabs, Inc). Libraries were sequenced on a HiSeq4000 with paired end 150bp
178 chemistry at the Michigan State Genomics Core Facility (East Lansing, MI).
179
180 *SNP discovery and genotyping*
181 Quality filtering, SNP identification, and genotyping was conducted using *STACKS* v.2.2
182 (Rochette and Catchen 2017). First, samples were demultiplexed with *process_radtags* with
183 flags -c, -q, -r, -t 140, --*bestrad*. Markers were discovered *de novo* and genotyped within
184 individuals with *ustacks* (flags = -m 3, -M 5, -H --*max_locus_stacks* 4, --*model_type* bounded, --
185 *bound_high* 0.05, --*disable-gapped*). A catalog of loci was created using a subset of the
186 individuals (diploid parents = 8, haploid parents = 5, wild fish = 38, total cisco = 51) with *cstacks*
187 (-n of 3, --*disable-gapped*). The 38 wild fish used in the catalog were collected from the same
188 geographic area using the same collection methods as listed above and were included to search
189 for a sex identification marker, which was unsuccessful (*data not shown*).
190 Putative loci within each individual fish were matched against the catalog with *sstacks*
191 (flag = --*disable-gapped*), *tsv2bam* was used with only the forward reads to orient the data by
192 SNP, and *gstacks* was used to combine genotypes across individuals. Only the forward reads
193 from the paired-end data were used in *gstacks* due to variable read depth in reverse reads and
194 thus less reliable genotyping. *gstacks* was also run separately with the forward and reverse reads
195 using *tsv2bam* to assemble longer contigs for sequence alignment and annotation. Final genotype
196 calls were output as VCF files with *populations* (flags = -r 0.75), with each family grouped as a
197 separate population in the *popmap* sample interpretation file. VCFtools (Danecek et al. 2011)
198 was used to identify and remove individuals from the study that were missing more than 30% of
199 data.
200 Maximum likelihood-based methods developed by Waples et al. (2016) were used to
201 identify loci that could be mapped in haploid crosses and to identify potentially duplicated loci.
202 Custom Python scripts available on GitHub (Python Software Foundation version 2.7) (see *Code*
203 *Availability*), were used to filter the haplotype VCF file output from the *populations* module to
204 identify loci that could be mapped in the diploid families. Loci missing more than 25% of data
205 and loci that were genotyped as heterozygous in both parents of diploid families (and therefore
206 could not be reliably mapped) were removed (as in Larson et al. 2015). Individual genotypes
207 were exported with the custom Python scripts as *LepMap3* input files. As a final step before
208 linkage mapping, genotypes from all seven families (haploid and diploid) were combined into a

single dataset to form the final female *LepMap3* input file and the four diploid families were combined into a single dataset to form the final male *LepMap3* input file.

Linkage mapping

The program *LepMap3* (Rastas 2017) was used to construct linkage maps following the methods of McKinney et al. (2016). Due to heterochiasmy (i.e., recombination rate differences between males and females) that occurs in the Salmonidae family (Sakamoto et al. 2000), a separate map was constructed for each sex. Loci were filtered and clustered into linkage groups (LGs) based on recombination rates by calculating logarithm of the odds (LOD) scores between all pairs of loci with the *SeparateChromosomes2* module. The LOD scores were chosen by increasing the LOD value by one until the number of LGs stabilized and was similar to that expected based on the haploid karyotype of cisco (N=40, Phillips et al. 1996). The final LOD scores used to generate the map were LOD = 15 and 5 for the female and the male maps, respectively. Loci were then ordered within LGs by utilizing paternal and maternal haplotypes as inheritance vectors with the *OrderMarkers2* module. We used a minimum marker number per LG of 100 for the female map and 40 for the male map. LGs were reordered and markers removed until no large gaps remained (Rastas 2017).

Comparative analysis of syntenic regions of linkage maps via MAPCOMP

MAPCOMP can be used to compare syntenic relationships among markers between linkage maps of any related species using a genome intermediate from another related species (Sutherland et al. 2016). Here, MAPCOMP was used to compare the cisco map with other *Coregonus* spp.; lake whitefish *C. clupeaformis* (Gagnaire et al. 2013), European whitefish *C. lavaretus* "Albock" (De-Kayne and Feulner 2018), as well as other representative species from other salmonid genera (i.e., Atlantic salmon (Lien et al. 2016), brook trout *S. fontinalis* (Sutherland et al. 2016), and Chinook salmon (Brieuc et al. 2014) and a representative of the outgroup to the salmonid WGD, northern pike *Esox lucius* (Rondeau et al. 2014). All code to collect and prepare maps, and run the analysis are available on GitHub (see *Code Availability*).

MAPCOMP pairs loci between the two compared linkage maps if they align at the same locus or close to each other on the same contig or scaffold (Sutherland et al. 2016). Due to the large phylogenetic distance covered in this analysis, two reference genomes; grayling (Savilammi et al. 2019) and Atlantic salmon (Lien et al. 2016) were used for comparisons. As salmonid chromosomal evolution is typified by Robertsonian fusions (Phillips and Rab 2001), fused chromosome arms in cisco, lake whitefish, and European whitefish were identified by aligning cisco markers to multiple salmonid resources to identify cases where one cisco LG corresponded with at least two chromosome arms in another species. The fusion and fission phylogenetic history was plotted based on the most parsimonious explanation of common fusions among *Coregonus* spp. basing the approximate timings of fusions on shared fusions among species. Fusion history shared at the base of the Salmoninae lineage was taken from earlier work (Sutherland et al. 2016).

Homeolog identification, similarity and inheritance mode

Homeologous chromosome arms can be identified in haploid crosses by mapping multiple alleles of duplicated markers as described in Brieuc et al. (2014). Duplicated markers in cisco were mapped using this method, and data obtained on duplicated markers from previously constructed linkage maps for coho salmon (Kodama et al. 2014), Chinook salmon (McKinney et al. 2019), pink salmon (Tarpey et al. 2017), chum salmon (Waples et al. 2016), and sockeye salmon (Larson et al. 2015). Homeologs were then ranked based on the number of markers supporting each known homeologous relationship.

Homeology was assessed by comparing DNA sequence similarity between homeologous arms from chromosome-level genome assemblies. Genomes included in this analysis were grayling (GCA_004348285.1, Savilammi et al. 2019) Atlantic salmon (GCF_000233375.1, Lien et al. 2016), Arctic char (GCF_002910315.2, Christensen et al. 2018b), rainbow trout (GCA_002163495.1, Pearse et al. 2018), and Chinook salmon (GCF_002872995.1, Christensen et al. 2018a). Homeologous pairs of chromosomes originating from the Ss4R duplication were known from original studies or through MAPCOMP comparisons and whole chromosomes containing the target chromosome pairs to be compared in each contrast were aligned to each other with LASTZ v1.02 (Harris 2007) following methods outlined in Lien et al. (2016). Options specified with LASTZ included --chain --gapped --gfextend --identity=75.0..100. --ambiguous=iupac --exact=20. The analysis was restricted to alignments with minimum percent match values of 75%, and a minimum length of 1,000 base pairs to minimize the likelihood of spurious alignments. Overall similarity of a homeologous pair was represented by the median percent similarity of all alignments, weighted by alignment length, and summarized with boxplots for each homeologous pair in each species ordered based on descending median percentage sequence similarity.

Each homeologous pair was classified into one of two categories, tetrasomic or disomic, using a machine learning approach. From the 25 homeologous pairs originating at the Ss4R for each species, a training set was constructed containing the four highest and lowest sequence similarity homeolog pairs. A *k* – nearest neighbor classification (knn) approach was then applied to the dataset using this training set. The *k* nearest-neighbors was determined by using a repeated 10-fold cross validation of 100 iterations with the trainControl function in the R package caret (v6.0-84, Kuhn 2019). The *k* for each species with the largest number of neighbors exhibiting the highest accuracy with the training set was selected. This *k* was then used to classify homeologous pairs as disomic or tetrasomic, along with the predefined training set (knn function, Ripley and Venables 2019). Overall, similarity between the two predicted categories from the *knn* classification was tested with a Wilcoxon test (R Core Team 2018) incorporating percent similarity from all alignments to determine whether the categories displayed significantly different sequence similarity between homeologs (alpha = 0.01). All scripts used in this analysis are available on GitHub (see *Code Availability*).

Results

RADseq, SNP discovery, and data filtering

RADseq data were obtained from 746 cisco across seven families, with an average of 4.1 M reads per individual (range: 1.1 – 30.8 M reads per individual). Individuals that were genotyped at more than 30% of loci and loci that were genotyped in more than 75% of the total individuals were retained, resulting in a dataset of 676 individuals (n = 333 diploid offspring; 330 haploid offspring; and 13 parents) and 49,998 unique loci (Supplementary file S1).

Linkage Mapping

A total of 22,459 unique loci were identified in the female (Figure 1) and male linkage maps (Supplementary file S2). The female map included 20,450 loci distributed across 38 LGs (Table 1), the male map included 6,340 loci distributed across 40 LGs, and 4,012 loci were present on both maps (Supplementary file S3). A total of 40 chromosomes was expected from karyotyping of coregonine fishes from the Great Lakes (Phillips et al. 1996), which matches the number of LGs mapped in males. However, male LGs 39 and 40 in males contained relatively few markers and may be fragments of other linkage groups rather than the two linkage groups that were not mapped in females. Eight LGs (i.e., Cart01 – Cart08) were identified as metacentric based on homology to two chromosome arms in other salmonids using MapComp (*see below*). In the female map, metacentric LGs were on average 85.44 cM (57.72 – 101.35 cM) and contained 739 loci (range: 685 - 867). Putative acrocentric LGs in the female map were on average 59.12 cM (50.97 - 64.53 cM) and contained 484 loci (range: 296 - 585). The total length of the female map was 2,456.98 cM. The average lengths of metacentric LGs on the male map were 66.76 cM (51.62 – 87.14 cM) and they contained 212 loci on average (range: 159 – 278). Putative acrocentric LGs in the male map were on average 57.00 cM (40.54 – 83.66 cM) and contained 145 loci (range: 41 – 224). The total length of the male map was 2,357.97 cM. We identified 3,383 putatively duplicated loci on the female linkage map, and of these, 2,671 loci mapped to one paralog and 709 loci mapped to both paralogs.

Comparative analysis of syntenic regions of linkage maps via MapComp

The main focus of our comparative analysis was to define the homologous and homeologous relationships among the linkage maps available for the coregonines, specifically in cisco (current study), lake whitefish (Gagnaire et al. 2013), and European whitefish (De-Kayne and Feulner 2018), and bring these species into the context of the broader chromosomal correspondence by identifying the homologous chromosome arms in brook trout, Atlantic salmon, and Chinook salmon, as well as the non-duplicated northern pike (Table 2, Supplementary file S4). To facilitate these comparisons, we have defined the term protokaryotype (PK). PKs correspond to hypothetical ancestral salmonid chromosomes pre-duplication inferred from observing a representative genome of the non-duplicated sister lineage of salmonids from the Esociformes and are ordered 01-25. Protokaryotypes correspond 1:1 with the northern pike genome but have two descendant homeologous regions within salmonid

329 genomes. For example, PK 01 corresponds to northern pike chromosome 01 and was an ancestral
330 pre-duplication salmonid chromosome which gave rise to homeologous Atlantic salmon
331 chromosomes Ssa09c (PK 01.2) and Ssa20b (PK 01.1) and to homeologous rainbow trout
332 Omy27 (PK 01.1) and Omy24 (PK 01.2) (Supplemental file S7). PKs in the previously
333 hypothesized "magic eight" from linkage mapping studies are PKs 02, 06, 09, 11, 20, 22, 23, 25.
334 PKs defined as LORes by Robertson et al. (2017) and those that displayed residual tetraploidy in
335 previous genome-based studies (Lien et al. 2016; Campbell et al. 2019) are the same as these
336 with the exception of PK 06. The use of this naming system corresponds to the identifiers used in
337 Sutherland et al. (2016), maintaining the correspondence of PK .1 or .2.
338 Most PKs were identifiable in the MapComp analysis, with some notable exceptions for
339 each coregonine species. In cisco, chromosome arms PK 22.2, 25.1 and 25.2 were unidentified;
340 two of these arms (PK 25.1 25.2) were also unidentified in the European whitefish linkage map
341 (De-Kayne and Feulner 2018). Additionally, it was difficult to determine correspondences for
342 PK 09 and 23. In European whitefish, five chromosome arms were unidentifiable (i.e., PK 09.1,
343 09.2, 23.2, 25.1 and 25.2), and there were homeology ambiguities for PK 02, as well as
344 homology ambiguities for PK arm 05.2 (Table 2). In lake whitefish, five arms were
345 unidentifiable (i.e., PK 09.2, 11.2, 20.2, 22.1, and 25.2), and there were homeology ambiguities
346 for PK 02.2 and 06.2. In multiple species, arms where it was difficult to determine homologous
347 relationships often had a high proportion of duplicated loci, presumably making distinguishing
348 homologs and homeologs challenging. Nonetheless, most homologs and homeologs (42/50) were
349 identified in all three coregonine species. This information was then leveraged to characterize the
350 fusion/fission history within the Coregoninae lineage using the methods outlined in Sutherland et
351 al. (2016).
352 The fusion/fission analysis indicated far fewer species-specific fusions than identified for
353 salmonids in Sutherland et al. (2016), with most fusions occurring prior to the divergence of the
354 coregonines (Figure 2). This difference in species-specific fusions may also be related to the
355 general lower number of fusions in coregonines relative to *Salmo* and *Oncorhynchus*
356 (Supplementary file S5); although, in the coregonines the majority of fusions were observed in
357 more than one species, which was not observed in most other species previously characterized.
358 Two strongly supported fusions were observed in all three coregonine species: fusions PK 05.1-
359 06.1 and 10.2-24.1. PK 11.1-21.1 was fused in both cisco and European whitefish, which
360 presumably underwent a fission in lake whitefish (Figure 2). The full characterization of fissions
361 will require the resolution of those ambiguous arms that are considered as probable in the current
362 analysis, and this may be further clarified in future work.
363 In summary, five fusions were likely shared among all three species, and one was shared
364 between cisco and European whitefish (PK 11.1-21.1). Cisco had two species-specific fusions
365 (PK 10.2-20.1 and 22.1-20.2), bringing the total count of observed fusions to eight. European
366 whitefish had one species-specific fusion (PK 20.2-10.2), bringing the total count of observed
367 fusions to seven. Lake whitefish also had one species-specific fusion (PK 20.1-23.2), bringing
368 the total count to six. Interestingly, the PK 09.2-17.1 fusion that was originally proposed to be

369 shared among all known salmonids (Sutherland et al. 2016), was found not to be fused in any of
370 the species here, suggesting either that this fusion occurred after the divergence of Coregonus
371 from the ancestor of the rest of the salmonids, or that a fission occurred at the base of the
372 coregonines (Figure 2).
373
374 *Homeolog identification, similarity, and inheritance mode*
375 A second major goal of this study was to compare homeologous relationships and modes
376 of inheritance within and among species. We identified 17 of the 25 homeologous chromosome
377 pairs (PK) in cisco using the markers that could be mapped to both homeologs in the linkage
378 map, and each homeologous pair shared between one and 86 duplicated loci (Fig. 3,
379 Supplementary file S6). Of the 17 homeolog pairs, six (PK 02, 06, 09, 11, 20, 23) had many loci
380 (42-86) supporting homeology; these are six of the "magic eight" discussed above. The other 11
381 had few markers supporting homeology (i.e., 1-6) and are not members of the "magic eight". The
382 other two arms found in the magic eight were not identifiable in cisco, potentially because they
383 contained a large number of duplicated loci. All of the previously constructed linkage maps for
384 salmonids that included duplicated regions had a large number of markers supporting homeology
385 for the "magic eight" with the exception of pink salmon, where seven of the eight PKs had high
386 support (34 – 68 loci) but one pair (PK25) displayed substantially lower support (nine loci)
387 (Tarpey et al. 2017) (Figure 3, Supplementary file S6).
388 To better understand the genetic similarity between homeologs and infer inheritance
389 mechanisms (i.e., residual tetrasomy or disomy), all 25 known homeologous relationships were
390 compared in reference genomes for grayling, Atlantic salmon, Arctic char, rainbow trout, and
391 Chinook salmon (Figures 3 and 4, Supplementary file S7). Using the machine learning algorithm
392 (see Methods), the optimal *k* – nearest neighbor for each species was identified as five. Those
393 five nearest neighbors from the training sets voted on the assignment of a particular PK to either
394 putatively tetrasomic or disomic classes (Figure 4), and the proportion of votes supporting each
395 assignment are reported in Supplemental Document S8. The highest observed vote proportion for
396 assignment to a class is 4 of 5 as a result of the limit on training set size to four of each class and
397 the five optimal *k* – nearest neighbors indicated for accuracy.
398 For Atlantic salmon and all *Oncorhynchus* spp. (i.e., rainbow trout and Chinook salmon),
399 the same eight PKs (i.e., PK 01, 02, 09, 11, 20, 22, 23, 25) were classified as tetrasomic using the
400 machine learning approach. This list of PKs includes all of those defined as LORes by Robertson
401 et al. (2017), and one additional (i.e., PK 01), but does not include PK 06, which is considered to
402 be part of the "magic eight" using linkage map evidence. Arctic char showed evidence for
403 residual tetrasomy in seven of these eight PKs, with the exception of PK 11 (see below for
404 details regarding this discrepancy due to other chromosome arms in this fusion). Grayling also
405 shared seven of the eight residually tetraploid homeolog pairs, with the exception of PK 01. Most
406 PKs received the highest possible vote proportions for their classifications (0.8), however, PK01
407 in Atlantic salmon and rainbow trout demonstrated a lower vote proportion (0.6) (Figure 4,
408 Supplemental Document S8), suggesting reduced support for this homeologous pair being

409 tetrasomic. Additionally, PK 19 in grayling, didn't have the highest vote proportion and was
410 assigned as diploid but had the highest sequence similarity in that class (Figure 4, Supplemental
411 Document S8). Sequence similarity was significantly higher for the tetrasomic PKs across all
412 species ($P < 0.0001$).
413 Although the group of tetrasomic PKs was largely conserved across species, there was
414 substantial variation in the relative sequence similarity between these homeolog pairs (i.e., order
415 of highest to lowest similarity) among species. PK 01 consistently displayed the lowest sequence
416 similarity of all the PKs in all five species where it was classified as tetrasomic and did not
417 always receive the highest observed vote proportion (see above). However, there were a number
418 of other homeolog pairs that displayed highly variable sequence similarity rankings across
419 species (Figure 4). For example, PK 09 had the highest sequence similarity in the grayling
420 genome, the sixth highest in the rainbow trout genome, and the fourth or fifth highest in the other
421 genomes. This variation suggests that the frequency of tetravalent meiosis for each PK may
422 differ across species and that the process of diploidization has occurred in a species-specific
423 manner post WGD as suggested in the mechanisms proposed by Robertson et al (2017).
424
425 **Discussion**
426 The amount of genomic resources available for salmonids has increased drastically over
427 the last decade. However, many previous studies investigating genome evolution in salmonids
428 focus on one or a few species, with fewer studies considering broader subsets of available taxa to
429 understand patterns of genome evolution across the Salmonidae family (but see Sutherland et al.
430 2016; Robertson et al. 2017). Here, we utilize genomic resources along with a newly generated
431 high-density linkage map for cisco to compare patterns of homology, fusion/fission events,
432 homeology, and residual tetrasomy across species. The cisco linkage map incorporates
433 duplicated regions and contains 20,450 loci, making it denser than most RAD-based haploid
434 linkage maps that have previously been constructed for salmonids (typically built from 3,000 to
435 5,000 loci). Higher marker density helped identify orthologous relationships between
436 coregonines and other salmonids as well as to identify homeologous chromosomes in cisco. We
437 also demonstrate the use of the protokaryotype ID, defined here but first used in Sutherland et al.
438 (2016), for comparative analyses in salmonids in order to unify and facilitate comparative
439 approaches in salmonid linkage maps and chromosome-level assemblies. Comparisons across
440 Salmonidae revealed that patterns of rediploidization are relatively similar across genera and
441 loosely correspond with phylogeny. However, we did identify substantial variation in sequence
442 similarity between homeologs both within species across homeolog pairs, and among species,
443 suggesting that frequently used binary classifications such as AORe/LORe and "magic eight"
444 may be oversimplified.
445
446 *Protokaryotype identifiers to facilitate comparative genomics in salmonids*
447 Comparative genomics within Salmonidae is important for the interpretation of the
448 effects of rediploidization after WGD on genome evolution (e.g. Berthelot et al. 2014; Kodama

449 et al. 2014; Lien et al. 2016). However, most studies use independent naming schemes for each
450 species, making it difficult to directly compare studies without complicated lookup tables or
451 alignments to confirm homology (e.g. Brieuc et al. 2014; Kodama et al. 2014). Recently
452 developed methods for connecting linkage maps through reference genomes (Sutherland et al.
453 2016) facilitated description of homologous relationships for most or all linkage group arms
454 across salmonids (with a few exceptions in coregonines). Additionally, Sutherland et al. (2016)
455 and Savilammi et al. (2019) have explored the utility of naming chromosomes based on
456 homology to northern pike, which as a representative of the non-duplicated sister lineage of
457 salmonids can be used to represent the ancestral (pre-duplication) state of salmonid genomes.
458 Each northern pike chromosome is represented by two homeologous chromosome arms in
459 salmonids, which are designated with a .1 or .2 following the ancestral chromosome number.
460 This naming system, if applied broadly, will facilitate comparative genomics in salmonids by
461 creating a "Mueller element"-like system (reviewed in Schaeffer 2018), where each chromosome
462 arm has a universal identifier. However, there also remains value in species-specific identifiers;
463 for example, Cart03 is the third named linkage group in the Cisco linkage map (Table 2). By
464 comparison, Cart03 named via the PK system could be Cart03 (PK 08.2-09.1) or Cart03$_{PK08.2\text{-}09.1}$
465 as Cart03 represents the fusion of two ancestral salmonid chromosome arms 08.2 and 09.1
466 (Figure 1, Table 2).
467 The PK system allows for simple comparison across species and genera. A case study of
468 the utility of this system is the comparison of the chromosome containing the sex determining
469 gene in salmonids (sdY), which is known to move between chromosomes (see Table 1 in
470 Sutherland et al. 2017). This comparison demonstrated that some chromosome arms more
471 frequently contain the sex determining gene than would be expected by chance or explainable by
472 phylogenetic conservation (i.e., PK 01.2 (AC04q), PK 03.1 (Cclu25, Co30, So09), PK 19.1
473 (So09.5, AC04q.1), PK 15.1 (AC04q.2, BC35), Sutherland et al. 2017). Even more intriguing is
474 that the northern pike naming was based on the threespine stickleback *Gasterosteus aculeatus*
475 (Rondeau et al. 2014), and PK 19 is the sex determining chromosome in three spine stickleback
476 (Peichel et al. 2004). As observed above, this chromosome is often fused with sex chromosomes
477 in salmon (Sutherland et al. 2017). By comparison, using the naming system, it is easy to observe
478 that LG24 in northern pike (i.e., PK 24 in salmon), recently identified to hold the sex
479 determining gene in Northern pike (Pan et al. 2019), does not appear to contain the sex-
480 determining locus in any tested salmonids. Deriving this information would be more difficult
481 without the PK system and would require extensive cross-referencing.
482 Other applications for the PK naming system include facilitating comparisons of the
483 correspondence of chromosomes containing genes for adaptive potential in sockeye salmon
484 (So13$_{PK18.2}$, TULP4, Larson et al. 2017), for run timing in Chinook (Ots28$_{PK03.2}$, GREB1L,
485 Prince et al. 2017), and for age-at-maturity in Atlantic salmon (Ssa25$_{PK16.2}$, VGLL3, Barson et al.
486 2015). While there may be some sections of the PK that are not always retained (e.g., some
487 transposition of sections), as long as the majority of the chromosome is preserved, then the PK
488 system enables general comparisons. The PK system proposed here will facilitate quick and

accurate comparisons across taxa, adding significant value to the myriad studies searching for adaptively important genes and regions in salmonids by leveraging comparative approaches.

Homology and fission/fusion history in coregonines

Comparisons using linkage maps for three coregonine species, cisco, lake whitefish (Gagnaire et al. 2013), and European whitefish (De-Kayne and Feulner 2018), allowed us to assess homology and variation in karyotypes across the genus. Ambiguity in homologous relationships remained for at least five chromosome arms in all three coregonines; this degree of uncertainty was much higher than documented across the *Salmo*, *Oncorhynchus*, and *Salvelinus* genera by Sutherland et al. (2016), where there were only two ambiguities across these groups. Coregonines appear to have a number of relatively small acrocentric chromosomes (Phillips and Rab 2001), some of which contain a high degree of duplicated loci, making constructing linkage maps more difficult than for other salmonids (Gagnaire et al. 2013; De-Kayne and Feulner 2018). For example, PK 25 has never been successfully mapped in coregonines, likely because it is small, submetacentric or acrocentric, and contains many duplicates. In other salmonids, where PK 25 is part of larger and/or metacentric chromosome, mapping becomes easier as there are many disomically inherited markers on the chromosome. Interestingly, the fact that PK 25 displayed evidence of residual tetrasomy in cisco, even though it is likely an acrocentric or submetacentric chromosome, suggests that metacentric chromosomes are not required for homeologous recombination. This contradicts previous theory which suggests that homeologous recombination requires at least one chromosome arm to be metacentric (Kodama et al. 2014). Additionally, PK 25.2 in grayling is a submetacentric chromosome and also displays signals of residual tetrasomy (Savilammi et al. 2019), providing further evidence that a small secondary arm may be sufficient to facilitate tetrasomic meiosis.

The fusion history in coregonines differs substantially from many other members of the salmonid family. Members of the *Coregonus*, *Salvelinus*, and *Thymallus* genera possess the "A karyotype," with a diploid chromosome number (2N) ~80 and many acrocentric chromosomes, whereas *Oncorhynchus* and *Salmo*, possess the "B karyotype," with 2N ~60 and many metacentric chromosomes (Phillips and Rab 2001). Given that these both come from an ancestral type of n = 50 chromosome arms, species with the "A karyotype" have undergone fewer fusions than other lineages. Interestingly, it appears that "A karyotype" species also generally contain a lower proportion of species-specific fusions compared to "B karyotype" species, suggesting that the "B karyotype" status comes from species-specific fusions. Sutherland et al. (2016) investigated fusion history within many species from the *Oncorhynchus* genus and found that most species had many species-specific fusions (e.g., 17 species-specific fusions in pink salmon). However, Sutherland et al. (2016) only investigated one species from the *Coregonus* and *Salvelinus* genera as this was all that was available at the time of publication, and no species from *Thymallus*. Our current study is the first to investigate fusion history across multiple coregonines and illustrates that most fusions are shared among species in the *Coregonus* genus, contrasting the pattern observed in *Oncorhynchus* (Sutherland et al. 2016). The functional effect

529 of differing fusion histories is yet to be determined, and remains an important question
530 differentiating species within the *Coregonus*, *Salvelinus*, and *Thymallus* genera from other
531 salmonids. Further information from genome sequencing projects, for example the European
532 whitefish genome (De Kayne et al. 2019) should facilitate important future studies contrasting
533 genomic processes and structure in species with differing fusion histories.
534
535 *Patterns of homeology and residual tetrasomy across salmonids*
536 Although patterns of residual tetrasomy were generally conserved, variation within and
537 among species was observed when examining results from linkage maps versus reference
538 genomes. Sequence similarity analyses using reference genomes suggested that all species
539 showed evidence for residual tetrasomic inheritance in seven homeologous pairs (PK 02, 09, 11,
540 20, 22, 23, and 25) with the exception of PK11 in Arctic char (see below). Using linkage maps,
541 these same seven homeologous pairs have been found to be tetrasomic in *Oncorhynchus*
542 (Kodama et al. 2014; McKinney et al. 2019), *Salvelinus* (Sutherland et al. 2016; Nugent et al.
543 2017), *Salmo* (Robertson et al. 2017), and likely *Coregonus* (results reported herein), strongly
544 suggesting that tetravalent meioses can and do form between these homeologs in all investigated
545 species to date. However, evidence for residual tetrasomy differed between linkage maps and
546 genomes for multiple PKs, most notably PK 06, which was classified as tetrasomic in linkage
547 mapping studies but not in genome analyses, and PK 01 which was classified as tetrasomic in
548 genome analyses but not linkage maps. It is likely that some of these differences are the result of
549 methodological limitations of the current approach and point to future analysis approaches that
550 may be able to improve upon the framework presented here. This is further described below.
551 The observation that PK11 did not display high sequence similarity in Arctic char might
552 suggest a difference in diploidization rates in char compared to other salmonids for this
553 homeologous pair, but it is more likely that methodological limitations prevented us from
554 detecting residual tetrasomy, as a linkage map study in Arctic char found a high number of
555 duplicated markers on this PK (Nugent et al. 2017). The percentage similarity analysis applied in
556 the present study uses complete chromosome alignments and requires post-filtering to remove
557 non-homeologous alignments. This method appears to be robust when chromosome arms are
558 well defined but, PK 11 in Arctic char appears to be composed of four chromosome arms that
559 have come together in a series of species-specific fusions (inferred from Christensen et al.
560 2018b). Since arm boundaries were not well defined, alignments in this PK produced a wide
561 interquartile range, suggesting that, while some regions of the PK are likely undergoing residual
562 tetrasomy, the alignments may have masked these regions by integrating over multiple
563 chromosome arms. This would be particularly problematic if the chromosomes being compared
564 both contained non-target chromosomes that were homeologous. To improve upon the method
565 applied here, better definition of the breaks between chromosome fusions could be applied and
566 this could prevent such ambiguities or noise in the sequence similarity calculated. We therefore
567 conclude that PK 11 is likely tetrasomic in Arctic char, but that we were unable to classify it as
568 such due to methodological limitations. The sequence similarity method applied here is generally

569 robust, but the fusion history of the species being analyzed needs to be considered to avoid
570 unexpected and erroneous similarity values. Ideally, only the section containing the ancestral
571 chromosome of interest would be being compared between the homeologs. This is an avenue of
572 method development that will be valuable for future work.
573 Contrastingly, the finding that PK 06 is not tetrasomic does not appear to be due to
574 methodological limitations of our genome analysis but may be due to differences in estimating
575 extent of residual tetraploidy between linkage mapping and genome assembly approaches.
576 Linkage mapping in *Oncorhynchus* and *Salvelinus* consistently finds support for tetrasomic
577 inheritance at PK 06 (Larson et al. 2017; Nugent et al. 2017), but the genome analysis conducted
578 here and that conducted for rainbow trout (Campbell et al. 2019) found that this PK displayed
579 intermediate sequence similarity consistent with disomic homeologs. One of the ways the two
580 approaches differ is the length of the sequence used during each analysis. The genome analysis
581 conducted here calculated similarity by using alignments of at least 1,000 bp, whereas linkage
582 maps compare alleles within ~100-150 bp RADtags. The short sequences analyzed by software
583 such as STACKs (Rochette and Catchen 2017; Rochette et al. 2019) make it possible to collapse
584 sequences into a single locus that can be mapped at both paralogs, even when sequence
585 divergence in a given region is relatively large. This therefore makes linkage maps a less
586 conservative characterization method for determining residual tetrasomy. In addition, many
587 genome assemblers applied to salmonid genomes (e.g. Chin et al. 2016; Koren et al. 2017; Ruan
588 and Li 2019) are not optimized for paralogous regions in polyploid genomes. This could be
589 especially problematic for genomes that combine both disomic and tetrasomic regions, such as in
590 salmonids. The end results are that duplicated regions may be detected as single copies as a
591 result of sequence collapse during the assembly process (Alkan et al. 2011; Varadharajan et al.
592 2018). If sequences do not collapse during assembly, contigs might be fragmented and
593 misassembled in the genome, making it difficult to differentiate between homologs and
594 homeologs (Kyriakidou et al. 2018). This could lead to homeologous regions being missed
595 altogether in genome sequences, particularly in comparisons that require chromosome-level
596 assemblies. However, the fact that support for tetrasomic inheritance in other PKs identified as
597 tetrasomic through linkage mapping was consistent with that observed in genome analysis
598 strongly suggests that there is something unique with PK 06 rather than a fault with the genome
599 analyses conducted here. Perhaps, as suggested by Campbell et al (2019), the PK 06
600 chromosome arms are returning to a diploid state faster than the other seven tetrasomic
601 homeolog pairs or the tetrasomically inherited portion of PK 06 is smaller than other tetrasomic
602 PKs.
603 Another notable difference between linkage mapping and genome analysis was the
604 consistent classification of PK 01 as tetrasomic in the genome analysis (five of six species) but
605 not in any linkage map. PK 01 uniformly exhibited the least similarity between tetrasomic
606 homeologous pairs and was assigned to the putatively tetrasomic class of PKs with less certainty
607 by the machine learning algorithm. This suggests that PK 01 may have low levels of tetrasomy.
608 We also observed some consistent patterns of variation in sequence similarity within disomic

markers. For example, homeolog pairs for PK 24 and 21 generally displayed the lowest sequence similarity, and homeolog pairs for PK 07 and 19 displayed higher similarity. Our study therefore presents additional nuances into the rediploidization process by identifying a core group of conserved tetrasomic homeologs, potentially intermediate homeologs (PK 01, 06) and consistently diverged homeologs (PK 21, 24). Future investigations can be refined to examine four well-defined categories across PKs: tetrasomic, intermediate, disomic, and most diverged. This enhanced refinement should reduce noise from the incorrect pooling of homeologs and aid in understanding the rediploidization process in salmonids.

Interestingly, more variation in sequence similarity was observed within tetrasomic homeologs than was observed in disomic homeologs. For example, PK 23 has the second highest sequence similarity in Arctic char, the fourth highest in rainbow trout, the sixth highest in Atlantic salmon, and the seventh highest in grayling. While this may be in part due to differences in genome assembly method and quality, the fact that variation exists even among the highest quality genomes (Atlantic salmon and rainbow trout) suggests that rediploidization rates at tetrasomic PKs may vary among species, even though the same seven PKs are consistently classified as tetrasomic. In other words, although there appears to be a large amount of conservation of tetrasomic inheritance between species, our genome analyses also suggest some independence in the return to disomy since the three subfamilies of salmonid split ~ 50MYA.

Conclusions

Here we provide the most compete analysis of chromosomal rearrangements in Coregonine using the currently available genomic resources as well and a newly developed haploid linkage map for cisco. We also suggest a naming system, PK, to facilitate a comparative analysis of all currently available salmonid genomes and linkage maps. Our study revealed that patterns of tetrasomic inheritance are largely conserved across the salmonids, but that there is substantial variation in these patterns both within and among species. For example, while the same seven PKs appear to be tetrasomically inherited across all species examined, their relative rates of sequence similarity differ within species, suggesting the potential of independent evolutionary trajectories following speciation. Additionally, we documented that analyses based on linkage maps do not identify the same tetrasomically inherited PKs as genome analyses and postulate that this may be due to inconsistencies with genome assemblies or due to differences the length of sequence used in comparisons. When inferring genomic patterns, more robust results can be obtained by using comparative approaches as opposed to a single resource. We therefore urge researchers to be cautious when inferring patterns based on a single resource and to integrate across resources when possible. In conclusion, our study presents important insights about the WGD in salmon and also provides a framework that can be built upon to improve our understanding of WGDs both within and beyond salmonids.

Acknowledgements

This project was funded by the Great Lakes Restoration Initiative (GLRI). Special thanks to the USFWS crew members Chris Olds, Paul Haver, Kaley Genther, Steve Nimcheski, and Matt McLean for assistance in field sampling, USGS Great Lakes Science Center Aquatic Research Wet Lab for egg rearing, the University of Wisconsin- Stevens Point Molecular Conservation Genetics Lab for assistance in lab work, and the support of the Turing High Performance Computing cluster at Old Dominion University. Thanks to Kris Christensen and Eric Rondeau for valuable and constructive comments on the manuscript. Any use of trade, product, or company name is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Code Availability

Linkage mapping: https://github.com/DaniBlumstein/Cisco-Linkage-Map

MapComp: https://github.com/bensutherland/coregonus_mapcomp

Genome homeology relationships: https://github.com/MacCampbell/residual-tetrasomy

664 **References**

665 Ali OA, et al. 2016. RAD Capture (Rapture): Flexible and Efficient Sequence-Based
666 Genotyping. Genetics 202: 389-400. doi: 10.1534/genetics.115.183665

667 Alkan C, Sajjadian S, Eichler EE 2011. Limitations of next-generation genome sequence
668 assembly. Nat Methods 8: 61-65. doi: 10.1038/nmeth.1527

669 Allendorf FW, et al. 2015. Effects of crossovers between homeologs on inheritance and
670 population genomics in polyploid-derived salmonid fishes. J Hered 106: 217-227. doi:
671 10.1093/jhered/esv015

672 Allendorf FW, Danzmann RG 1997. Secondary Tetrasomic Segregation of MDH-B and
673 Preferential Pairing of Homeologues in Rainbow Trout. Genetics 145: 1083-1092.

674 Allendorf, FW, Thorgaard GH 1984. Tetraploidy and the evolution of salmonid
675 fishes. Evolutionary genetics of fishes. Springer, Boston, MA. 1-53.

676 Angers B, Bernatchez L, Angers A, Desgroseillers L 1995. Specific microsatellite loci for brook
677 charr reveal strong population subdivision on a microgeographic scale. Journal of Fish Biology
678 47: 177-185.

679 Arrigo N, Barker MS 2012. Rarely successful polyploids and their legacy in plant genomes. Curr
680 Opin Plant Biol 15: 140-146. doi: 10.1016/j.pbi.2012.03.010

681 Barson NJ, et al. 2015. Sex-dependent dominance at a single locus maintains variation in age at
682 maturity in salmon. Nature 528: 405-408. doi: 10.1038/nature16062

683 Berthelot C, et al. 2014. The rainbow trout genome provides novel insights into evolution after
684 whole-genome duplication in vertebrates. Nature communications 5: 3657. doi:
685 10.1038/ncomms4657

686 Brieuc MS, Waters CD, Seeb JE, Naish KA 2014. A dense linkage map for Chinook salmon
687 (Oncorhynchus tshawytscha) reveals variable chromosomal divergence after an ancestral whole
688 genome duplication event. G3 (Bethesda) 4: 447-460. doi: 10.1534/g3.113.009316

689 Campbell MA, Hale MC, McKinney GJ, Nichols KM, Pearse DE 2019. Long-Term
690 Conservation of Ohnologs Through Partial Tetrasomy Following Whole-Genome Duplication in
691 Salmonidae. G3 (Bethesda) 9: 2017-2028. doi: 10.1534/g3.119.400070

692 Campbell MA, Lopez JA, Sado T, Miya M 2013. Pike and salmon as sister taxa: detailed
693 intraclade resolution and divergence time estimation of Esociformes + Salmoniformes based on
694 whole mitochondrial genome sequences. Gene 530: 57-65. doi: 10.1016/j.gene.2013.07.068

695 Chin CS, et al. 2016. Phased diploid genome assembly with single-molecule real-time
696 sequencing. Nat Methods 13: 1050-1054. doi: 10.1038/nmeth.4035

697 Chourrout D 1982. Gynogenesis caused by ultraviolet irradiation of salmonid sperm. Journal of
698 Experimental Zoology 223: 175-181.

699 Christensen KA, et al. 2018a. Chinook salmon (Oncorhynchus tshawytscha) genome and
700 transcriptome. PLoS One 13: e0195461. doi: 10.1371/journal.pone.0195461

701 Christensen KA, et al. 2018b. The Arctic charr (Salvelinus alpinus) genome and transcriptome
702 assembly. PLoS One 13: e0204076. doi: 10.1371/journal.pone.0204076

703 Clarke JT, Lloyd GT, Friedman M 2016. Little evidence for enhanced phenotypic evolution in
704 early teleosts relative to their living fossil sister group. Proc Natl Acad Sci U S A 113: 11531-
705 11536. doi: 10.1073/pnas.1607237113

706 Crete-Lafreniere A, Weir LK, Bernatchez L 2012. Framing the Salmonidae family phylogenetic
707 portrait: a more complete picture from increased taxon sampling. PLoS One 7: e46662. doi:
708 10.1371/journal.pone.0046662

709 Danecek P, et al. 2011. The variant call format and VCFtools. Bioinformatics 27: 2156-2158.
710 doi: 10.1093/bioinformatics/btr330

711 De-Kayne R, Feulner PGD 2018. A European Whitefish Linkage Map and Its Implications for
712 Understanding Genome-Wide Synteny Between Salmonids Following Whole Genome
713 Duplication. G3 (Bethesda) 8: 3745-3755. doi: 10.1534/g3.118.200552

714 De-Kayne R, Zoller S, Feulner, PG 2019 A de novo chromosome-level genome assembly
715 of Coregonus sp. “Balchen”: one representative of the Swiss Alpine whitefish radiation. bioRxiv
716 doi: 10.1101/771600

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

720 Gerstein AC, Otto SP 2009. Ploidy and the causes of genomic evolution. Journal of Heredity
721 100: 571-581. doi:10.1093/jhered/esp057
722 Harris RS 2007. Improved pairwise alignment of genomic DNA. The Pennsylvania State
723 University.

724 Hollister JD 2015. Polyploidy: adaptation to the genomic environment. New Phytologist 205:
725 1034-1039. doi: 10.1111/nph.12939

726 Karine Alix et al. 2017. Polyploidy and interspecific hybridization: partners for adaptation,
727 speciation and evolution in plants, Annals of Botany, 120: 183-194,
728 https://doi.org/101093/aob/mcx079

729 Kodama M et al. 2014. Comparative mapping between Coho Salmon (Oncorhynchus kisutch)
730 and three other salmonids suggests a role for chromosomal rearrangements in the retention of
731 duplicated regions following a whole genome duplication event. G3 (Bethesda) 4: 1717-1730.
732 doi: 10.1534/g3.114.012294

733 Koren S et al. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer
734 weighting and repeat separation. Genome Res 27: 722-736. doi: 10.1101/gr.215087.116

735 Kuhn M 2019. caret: Classification and Regression Training. R package version 6.0-84.

736 Kyriakidou M, Tai HH, Anglin NL, Ellis D, Stromvik MV 2018. Current Strategies of Polyploid
737 Plant Genome Sequence Assembly. Front Plant Sci 9: 1660. doi: 10.3389/fpls.2018.01660

738 Larson WA, et al. 2017. Genomic islands of divergence linked to ecotypic variation in sockeye
739 salmon. Mol Ecol 26: 554-570. doi: 10.1111/mec.13933

740 Larson WA, et al. 2015. Identification of Multiple QTL Hotspots in Sockeye Salmon
741 (Oncorhynchus nerka) Using Genotyping-by-Sequencing and a Dense Linkage Map. J Hered
742 107: 122-133. doi: 10.1093/jhered/esv099

743 Lien S, et al. 2016. The Atlantic salmon genome provides insights into rediploidization. Nature
744 533: 200-205. doi: 10.1038/nature17164

745 Limborg MT, McKinney GJ, Seeb LW, Seeb JE 2016a. Recombination patterns reveal
746 information about centromere location on linkage maps. Mol Ecol Resour 16: 655-661. doi:
747 10.1111/1755-0998.12484

748 Limborg MT, Seeb LW, Seeb JE 2016b. Sorting duplicated loci disentangles complexities of
749 polyploid genome masked by genotyping by sequencing. Molecular ecology 25: 2117–2129.

750 Macqueen DJ, Johnston IA 2014. A well-constrained estimate for the timing of the salmonid
751 whole genome duplication reveals major decoupling from species diversification. Proc Biol Sci
752 281: 20132881. doi: 10.1098/rspb.2013.2881

753 Mayrose I, et al. 2015. Methods for studying polyploid diversification and the dead end
754 hypothesis: a reply to Soltis et al. (2014). New Phytologist 206: 27-35. doi: 10.1111/nph.13192
755 Mayrose I, et al. 2011. Recently formed polyploid plants diversify at lower rates. Science 333:
756 1257. doi: 10.1126/science.1207205

757 McKinney GJ, et al. 2016. An integrated linkage map reveals candidate genes underlying
758 adaptive variation in Chinook salmon (Oncorhynchus tshawytscha). Mol Ecol Resour 16: 769-
759 783. doi: 10.1111/1755-0998.12479

760 McKinney G et al. 2017. Paralogs are revealed by proportion of heterozygotes and deviations in
761 read ratios in genotyping-by-sequencing data from natural populations. Mol Ecol Resour, 17:
762 656-669. doi:10.1111/1755-0998.12613

763 McKinney GJ, et al. 2019. Dense SNP panels resolve closely related Chinook salmon
764 populations. Canadian Journal of Fisheries and Aquatic Sciences. doi: 10.1139/cjfas-2019-0067

765

766 Norden CR 1961. Comparative osteology of representative salmonid fishes, with particular
767 reference to the grayling (Thymallus arcticus) and its phylogeny. Journal of the Fisheries Board
768 of Canada 18: 679-791.

769 Nugent CM, Easton AA, Norman JD, Ferguson MM, Danzmann RG 2017. A SNP Based
770 Linkage Map of the Arctic Charr (Salvelinus alpinus) Genome Provides Insights into the
771 Diploidization Process After Whole Genome Duplication. G3 (Bethesda) 7: 543-556. doi:
772 10.1534/g3.116.038026

773 Ohno S 1970. Evolution by gene duplication: Springer Science & Business Media.

774 Ohta T 1989. Role of gene duplication in evolution. Genome, 31(1), 304-310. doi:10.1139/g89-
775 048

776 Pan Q et al. 2019. Identification of the master sex determining gene in Northern pike (Esox
777 lucius) reveals restricted sex chromosome differentiation. PLoS Genet 15: e1008013. doi:
778 10.1371/journal.pgen.1008013

779 Patton JC, Gallaway BJ, Fechhelm RG, Cronin MA 1997. Genetic variation of microsatellite and
780 mitochondrial DNA markers in broad whitefish (Coregonus nasus) in the Colville and
781 Sagavanirktok rivers in northern Alaska. Canadian Journal of Fisheries and Aquatic Sciences 54:
782 1548-1556. doi: 10.1139/f97-062
783 Pearse DE, et al. 2018. Sex-dependent dominance maintains migration supergene in rainbow
784 trout. bioRxiv (2018): 504621. doi: 10.1101/504621
785 Peichel CL, et al. 2004. The master sex-determination locus in threespine sticklebacks is on a
786 nascent Y chromosome. Current biology 16: 1416-1424. doi: 10.1016/j.cub.2004.08.030
787 Phillips R, Rab P 2001. Chromosome evolution in the Salmonidae (Pisces): an update. Biological
788 Reviews 76: 1-25. doi: 10.1017/S1464793100005613

789 Phillips RB, Reed KM, Ráb P 1996. Revised karyotypes and chromosome banding of coregonid
790 fishes from the Laurentian Great Lakes. Canadian Journal of Zoology 74: 323-329. doi:
791 10.1139/z96-040
792 Prince DJ, et al. 2017. The evolutionary basis of premature migration in Pacific salmon
793 highlights the utility of genomics for informing conservation. Science advances 3: e1603198.
794 doi: 10.1126/sciadv.1603198
795 Python Software Foundation version 2.7. Python Language Reference, Available at
796 http://www.python.org.

797 R Core Team 2018. R: A language and environment for statistical computing. Vienna, Austria: R
798 Foundation for Statistical Computing.

799 Rastas P 2017. Lep-MAP3: robust linkage mapping even for low-coverage whole genome
800 sequencing data. Bioinformatics 33: 3726-3732. doi: 10.1093/bioinformatics/btx494

801 Rastogi S, Liberles DA 2005. Subfunctionalization of duplicated genes as a transition state to
802 neofunctionalization. BMC Evolutionary Biology 5: 28. doi:10.1186/1471-2148-5-28
803 Ripley B, Venables W 2019. class: Functions for Classification. R package version 7.3-15.

804 Robertson FM, et al. 2017. Lineage-specific rediploidization is a mechanism to explain time-lags
805 between genome duplication and evolutionary diversification. Genome Biol 18: 111. doi:
806 10.1186/s13059-017-1241-z

807 Rochette NC, Catchen JM 2017. Deriving genotypes from RAD-seq short-read data using
808 Stacks. Nat Protoc 12: 2640-2659. doi: 10.1038/nprot.2017.123

809 Rochette NC, Rivera-Colón AG, Catchen JM 2019. Stacks 2: Analytical Methods for Paired-end
810 Sequencing Improve RADseq-based Population Genomics. bioRxiv: 615385. doi:
811 10.1101/615385
812 Rogers SM, Marchand M-H, Bernatchez L 2004. Isolation, characterization and cross-salmonid
813 amplification of 31 microsatellite loci in the lake whitefish (Coregonus clupeaformis, Mitchill).
814 Molecular Ecology Notes 4: 89-92. doi: 10.1046/j.1471-8286.2003.00578.x

815 Rondeau EB, et al. 2014. The genome and linkage map of the northern pike (Esox lucius):
816 conserved synteny revealed between the salmonid sister group and the Neoteleostei. PLoS One
817 9: e102089. doi: 10.1371/journal.pone.0102089
818 Ruan J, Li H 2019. Fast and accurate long-read assembly with wtdbg2. bioRxiv: 530972. doi:
819 10.1101/530972
820 Sakamoto T, et al. 2000. A Microsatellite Linkage Map of Rainbow Trout (Oncorhynchus
821 mykiss) Characterized by Large Sex-Specific Differences in Recombination Rates. Genetics 155:
822 1331–1345.

823 Santini F, Harmon LJ, Carnevale G, Alfaro ME 2009. Did genome duplication drive the origin of
824 teleosts? A comparative study of diversification in ray-finned fishes. BMC Evol Biol 9: 194. doi:
825 10.1186/1471-2148-9-194

826 Savilammi T, et al. 2019. The Chromosome-Level Genome Assembly of European Grayling
827 Reveals Aspects of a Unique Genome Evolution Process Within Salmonids. G3 (Bethesda) 9:
828 1283-1294. doi: 10.1534/g3.118.200919

829 Schaeffer SW 2018. Muller "Elements" in Drosophila: How the Search for the Genetic Basis for
830 Speciation Led to the Birth of Comparative Genomics. Genetics 210: 3-13. doi:
831 10.1534/genetics.118.301084

832 Schwager EE, et al. 2017. The house spider genome reveals an ancient whole-genome
833 duplication during arachnid evolution. BMC Biol 15: 62. doi: 10.1186/s12915-017-0399-x

834 Selmecki AM, et al. 2015. Polyploidy can drive rapid adaptation in yeast. Nature 519: 349-352.
835 doi: 10.1038/nature14187

836 Sutherland BJG, et al. 2016. Salmonid Chromosome Evolution as Revealed by a Novel Method
837 for Comparing RADseq Linkage Maps. Genome Biol Evol 8: 3600-3617. doi:
838 10.1093/gbe/evw262

839 Sutherland BJG, Rico C, Audet C, Bernatchez L 2017. Sex Chromosome Evolution,
840 Heterochiasmy, and Physiological QTL in the Salmonid Brook Charr Salvelinus fontinalis. G3
841 (Bethesda) 7: 2749-2762. doi: 10.1534/g3.117.040915

842 Tarpey CM, Seeb JE, McKinney GJ, Seeb LW 2017. A dense linkage map for odd-year lineage
843 pink salmon incorporating duplicated loci. In. Alaska Department of Fish and Game: School of
844 Aquatic and Fishery Sciences, University of Washington. p. 1-50.

845 Taylor JS, Braasch I, Frickey T, Meyer A, Van de Peer Y 2003. Genome duplication, a trait
846 shared by 22000 species of ray-finned fish. Genome Res 13: 382-390. doi: 10.1101/gr.640303

847 Van de Peer Y, Mizrachi E, Marchal K 2017. The evolutionary significance of polyploidy. Nat
848 Rev Genet 18: 411-424. doi: 10.1038/nrg.2017.26

849 Vanneste K, Maere S, Van de Peer Y 2014. Tangled up in two: a burst of genome duplications at
850 the end of the Cretaceous and the consequences for plant evolution. Philos Trans R Soc Lond B
851 Biol Sci 369. doi: 10.1098/rstb.2013.0353

852 Varadharajan S, et al. 2018. The Grayling Genome Reveals Selection on Gene Expression
853 Regulation after Whole-Genome Duplication. Genome Biol Evol 10: 2785-2800. doi:
854 10.1093/gbe/evy201

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

858 Wendel JF. 2000. Genome evolution in polyploids. In. Plant molecular evolution: Springer. p.
859 225-249.

860 Wittbrodt K, Meyer A, Schartl M 1998. More genes in fish? BioEssays 20: 511-515. doi:
861 10.1002/(sici)1521-1878(199806)20:6<511::Aid-bies10>3.0.Co;2-3

862 Wolfe KH 2001. Yesterday's polyploids and the mystery of diploidization. Nature reviews
863 genetics 2: 333.

864 Wood TE, et al. 2009. The frequency of polyploid speciation in vascular plants. Proc Natl Acad
865 Sci U S A 106: 13875-13879. doi: 10.1073/pnas.0811575106

866 Zhan SH, Glick L, Tsigenopoulos CS, Otto SP, Mayrose I 2014. Comparative analysis reveals
867 that polyploidy does not decelerate diversification in fish. J Evol Biol 27: 391-403. doi:
868 10.1111/jeb.12308

Table 1. Results for the female and male cisco linkage maps. Duplicated and singleton loci are from the female linkage map, and % duplicated is the percentage of duplicated loci on each female linkage group (LG). LG type is denoted with acrocentric (A) and metacentric (M).

				Length		Total Loci		Loci/cM		
Cisco LG	Duplicated Loci	Singleton Loci	% Duplicated	Female (cM)	Male (cM)	Female	Male	Female	Male	LG Type
Cart1	212	510	0.42	101.35	87.14	722	224	7.12	2.57	M
Cart2	394	384	1.03	97.77	61.11	778	159	7.96	2.6	M
Cart3	226	477	0.47	93.56	78.98	703	248	7.51	3.14	M
Cart4	48	547	0.09	92.45	67.80	595	194	6.44	2.86	M
Cart5	283	584	0.48	91.73	58.07	867	219	9.45	3.77	M
Cart6	108	729	0.15	91.36	68.95	837	278	9.16	4.03	M
Cart7	189	496	0.38	57.72	60.40	685	204	11.87	3.38	M
Cart8	189	532	0.36	57.56	51.62	721	170	12.53	3.29	M
Cart9	238	271	0.88	64.53	44.47	509	92	7.89	2.07	A
Cart10	21	488	0.04	63.98	49.68	509	152	7.96	3.06	A
Cart11	38	510	0.07	63.72	49.41	548	165	8.6	3.34	A
Cart12	393	177	2.22	63.18	74.69	570	80	9.02	1.07	A
Cart13	27	558	0.05	62.90	53.67	585	181	9.3	3.37	A
Cart14	73	454	0.16	62.66	59.74	527	140	8.41	2.34	A
Cart15	96	318	0.30	62.45	48.26	414	121	6.63	2.51	A
Cart16	24	456	0.05	62.25	45.80	480	153	7.71	3.34	A
Cart17	27	478	0.06	62.25	53.19	505	160	8.11	3.01	A
Cart18	25	446	0.06	62.16	42.08	471	136	7.58	3.23	A
Cart19	29	541	0.05	61.85	75.08	570	213	9.22	2.84	A
Cart20	31	536	0.06	60.19	48.36	567	224	9.42	4.63	A
Cart21	31	388	0.08	59.04	61.23	419	142	7.1	2.32	A

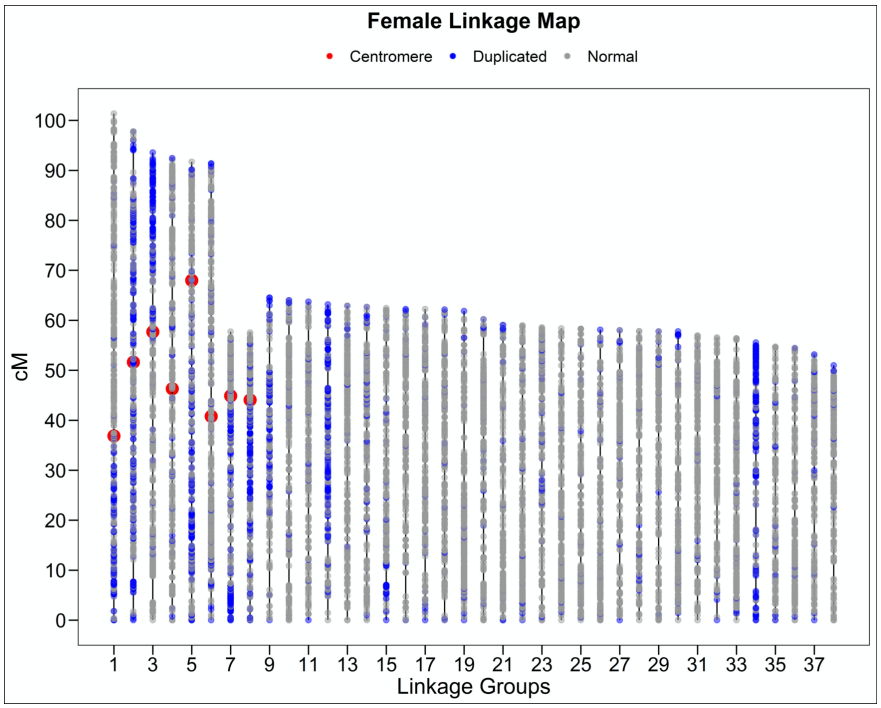
Cart22	20	544	0.04	58.96	49.49	564	204	9.57	4.12	A
Cart23	52	473	0.11	58.57	55.39	525	162	8.96	2.92	A
Cart24	13	556	0.02	58.38	57.23	569	195	9.75	3.41	A
Cart25	26	458	0.06	58.28	51.08	484	161	8.3	3.15	A
Cart26	30	445	0.07	58.09	50.91	475	147	8.18	2.89	A
Cart27	32	499	0.06	58.00	60.69	531	213	9.16	3.51	A
Cart28	27	444	0.06	57.79	54.44	471	141	8.15	2.59	A
Cart29	15	404	0.04	57.78	62.64	419	149	7.25	2.38	A
Cart30	29	384	0.08	57.76	71.50	413	141	7.15	1.97	A
Cart31	25	541	0.05	56.93	77.82	566	184	9.94	2.36	A
Cart32	19	478	0.04	56.56	75.17	497	177	8.79	2.35	A
Cart33	20	460	0.04	56.37	49.57	480	146	8.52	2.95	A
Cart34	262	82	3.20	55.54	83.66	344	62	6.19	0.74	A
Cart35	42	254	0.17	54.74	40.54	296	81	5.41	2	A
Cart36	19	332	0.06	54.45	45.50	351	117	6.45	2.57	A
Cart37	25	402	0.06	53.16	55.82	427	122	8.03	2.19	A
Cart38	25	431	0.06	50.97	73.21	456	171	8.95	2.34	A
Cart39	0	0	0.00	0.00	45.13	0	41	0	0.91	A
Cart40	0	0	0.00	0.00	58.46	0	71	0	1.21	A
Averages	89.03	449.13	0.20	64.66	58.95	538.16	158.5	8.36		
Total	3383	17067	0.20	2456.98	2357.98	20450	6340	317.71		

Table 2. MAPCOMP results documenting homologous chromosomes for the three coregonines; cisco, lake whitefish (Gagnaire et al. 2013), and European whitefish (De-Kayne and Feulner 2018), integrated with Atlantic salmon (Lien et al. 2016), brook trout (Sutherland et al. 2016), and Chinook salmon (Brieuc et al. 2014). Homologous chromosomes for all species are named according to the corresponding Northern Pike linkage group as a reference (Rondeau et al. 2014), as per Sutherland et al. (2016), here termed protokaryotype ID (PK). ^ indicates weak evidence and * indicates uncertainty between homeologs.

Northern Pike (PK)	Cisco	European Whitefish	Lake Whitefish	Brook Trout	Atlantic Salmon	Chinook Salmon
1.1	Cart23	Calb16	Cclu28	Sf25	Ssa20b	Ots13q
1.2	Cart14	Calb33	Cclu35	Sf38	Ssa09c	Ots14q
2.1	Cart01a	Calb02b^*	Cclu04a	Sf06a	Ssa26	Ots04q
2.2	Cart12	Calb02b^*	Cclu04a^*	Sf28	Ssa11a	Ots12q
3.1	Cart25	Calb19	Cclu25	Sf22	Ssa14a	Ots10q
3.2	Cart26	Calb22	Cclu26	Sf11	Ssa03a	Ots28
4.1	Cart30	Calb29	Cclu16	Sf33	Ssa09b	Ots08q
4.2	Cart21	Calb30	Cclu29	Sf07b	Ssa05a	Ots21
5.1	Cart06b	Calb01a	Cclu05a	Sf01a	Ssa19b	Ots24
5.2	Cart18	Calb35 or Calb40	Cclu15	Sf27	Ssa28	Ots25
6.1	Cart06a	Calb01b	Cclu05b	Sf01b	Ssa01b	Ots01q
6.2	Cart15	Calb27	Cclu05b^*	Sf36	Ssa18a	Ots06q
7.1	Cart20	Calb06	Cclu13	Sf08b	Ssa13b	Ots09p
7.2	Cart19	Calb07	Cclu08	Sf09	Ssa04b	Ots30
8.1	Cart27	Calb17	Cclu36	Sf04a	Ssa23	Ots01p
8.2	Cart03a	Calb08	Cclu06a	Sf17	Ssa10a	Ots05q
9.1	Cart03b*	missing	Cclu06b	Sf42	Ssa02b	Ots32
9.2	Cart34*	missing	missing	Sf03b	Ssa12a	Ots02q
10.1	Cart08a	Calb20b	Cclu10	Sf23	Ssa27	Ots13p
10.2	Cart04b	Calb09a	Cclu24a	Sf34	Ssa14b	Ots31
11.1	Cart05a	Calb13b	Cclu18	Sf14	Ssa06a	Ots27
11.2	Cart09	Calb34	missing	Sf08a	Ssa03b	Ots09q
12.1	Cart33	Calb14	Cclu27	Sf18	Ssa13a	Ots22
12.2	Cart16	Calb28	Cclu14	Sf30	Ssa15b	Ots16q
13.1	Cart17	Calb25	Cclu34	Sf06b	Ssa24	Ots04p
13.2	Cart32	Calb31	Cclu37	Sf40	Ssa20a	Ots12p
14.1	Cart01b	Calb02a	Cclu04b	Sf13	Ssa01c	Ots20
14.2	Cart38	Calb11	Cclu33	Sf10	Ssa11b	Ots33

15.1	Cart10	Calb18	Cclu31	Sf35	Ssa09a	Ots08p
15.2	Cart31	Calb10	Cclu22	Sf12	Ssa01a	Ots11q
16.1	Cart07a	Calb03b	Cclu02b or Cclu03	Sf26	Ssa21	Ots26
16.2	Cart28	Calb21	Cclu32	Sf24	Ssa25	Ots03q
17.1	Cart24	Calb05	Cclu38	Sf03a	Ssa12b	Ots02p
17.2	Cart22	Calb12	Cclu21	Sf21	Ssa22	Ots07q
18.1	Cart29	Calb24	Cclu40	Sf19	Ssa15a	Ots05p
18.2	Cart37	Calb23	Cclu17	Sf31	Ssa06b	Ots18
19.1	Cart13	Calb04	Cclu30	Sf15	Ssa10b	Ots19
19.2	Cart11	Calb15b	Cclu11	Sf20	Ssa16a	Ots06p
20.1	Cart08b^	Calb36	Cclu01a	Sf07a	Ssa05b	Ots23
20.2	Cart02b	Calb20a	missing	Sf29	Ssa02a	Ots03p
21.1	Cart05b	Calb13a	Cclu12	Sf05b	Ssa29	Ots29
21.2	Cart36	Calb26	Cclu39	Sf16	Ssa19a	Ots16p
22.1	Cart02a	Calb39^	missing	Sf39	Ssa17a	Ots07p
22.2	missing	Calb15a	Cclu19^	Sf05a	Ssa16b	Ots14p
23.1	Cart07b^*	Calb03a	Cclu02a	Sf02b	Ssa07b	Ots15p
23.2	Cart07b^*	missing	Cclu01b^	Sf37	Ssa17b	Ots17
24.1	Cart04a	Calb09b	Cclu24b	Sf02a	Ssa07a	Ots15q
24.2	Cart35	Calb32	Cclu23	Sf32	Ssa18b	Ots10p
25.1	missing	missing	Cclu09^	Sf04b	Ssa04a	Ots34
25.2	missing	missing	missing	Sf41	Ssa08a	Ots11p

A



B

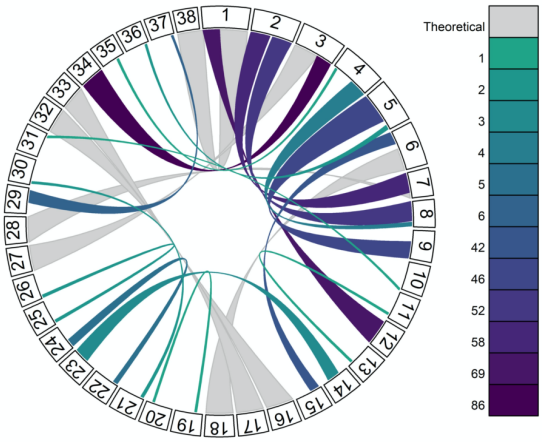


Figure 1. A) Female linkage map for cisco containing 20,450 loci. Each dot represents a locus, duplicated loci are blue and singleton loci are gray. Lengths are in centimorgans (cM). Approximate location of centromeres for metacentric LGs are denoted in red. Metacentric LGs were identified through homologous relationships of chromosome arms with other Salmonids via MapComp. B) Circos plot of cisco LGs highlighting 17 supported homeologous regions within the linkage map. Included in the 17 homeologus regions are six of the eight regions that are likely still residually tetrasomic across the Salmonids. Colors represent the number of markers supporting relationship, with darker colors representing higher marker numbers (maximum support = 86 markers).

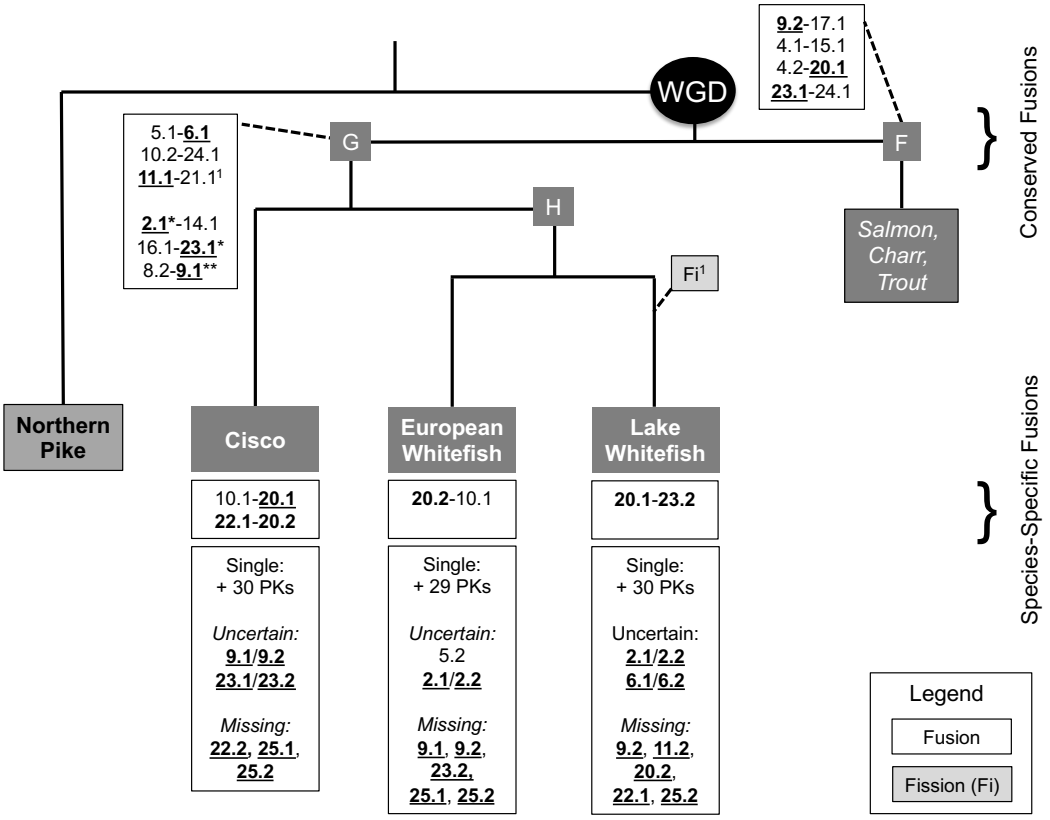


Figure 2. Fusions and fissions in the Coregoninae and Salmoninae lineages. This is an extension of Figure 4 from Sutherland et al. (2016). White boxes display the fusion events, where the homologous chromosomes for all species are named according to the protokaryotype ID. Bold and underlined chromosome numbers are the homeologous pairs that exhibit residual tetraploidy (i.e "magic eight"), * indicate uncertainty in one species, and ** indicates uncertain in two species (i.e., *C. artedi* is ambiguous for homeolog 9.1 or 9.2 while *C. lavaretus* is missing 9.1 and 9.2). Above the species names are conserved fusions, whereas below are the species-specific fusions. The phylogeny is adapted from (Crete-Lafreniere et al. 2012). Branch lengths do not represent phylogenetic distance, only relative phylogenetic position.
[1]Arms 11.1-21.1 were fused in both cisco and European whitefish, but likely underwent fission in lake whitefish.

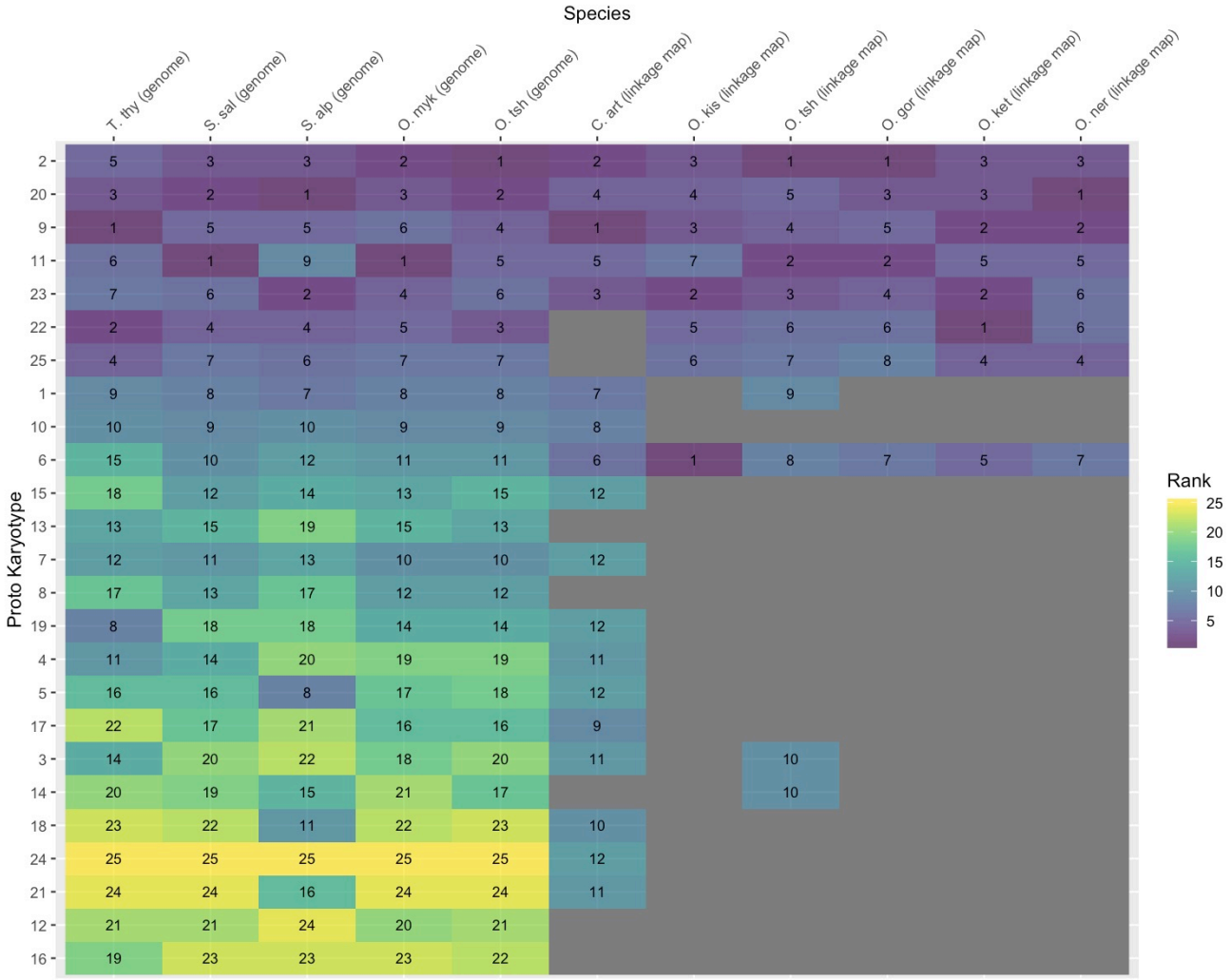


Figure 3. Ranking of homeologous chromosome pairs based on putative residual tetrasomic inheritance as measured by the number of markers shared among homeologs for linkage maps or percent sequence similarity for genomes. A lower rank represents more marker pairs supporting a homeolog and or a higher sequence similarity. Chromosomes for all species are named according to the protokaryotype ID (PK). PKs are ordered in the figure by averaging the ranks across all species and then sorting the averages from smallest to largest (i.e., ordered from highest support for residual tetrasomy to lowest). Grey indicates that no duplicated loci could be mapped to both homeologs. Species abbreviations are grayling (T. thy), Atlantic salmon (S. sal), Arctic char (S. alp), rainbow trout (O. myk), Chinook salmon (O. tsh), cisco (C. art), and coho salmon (O. kis).

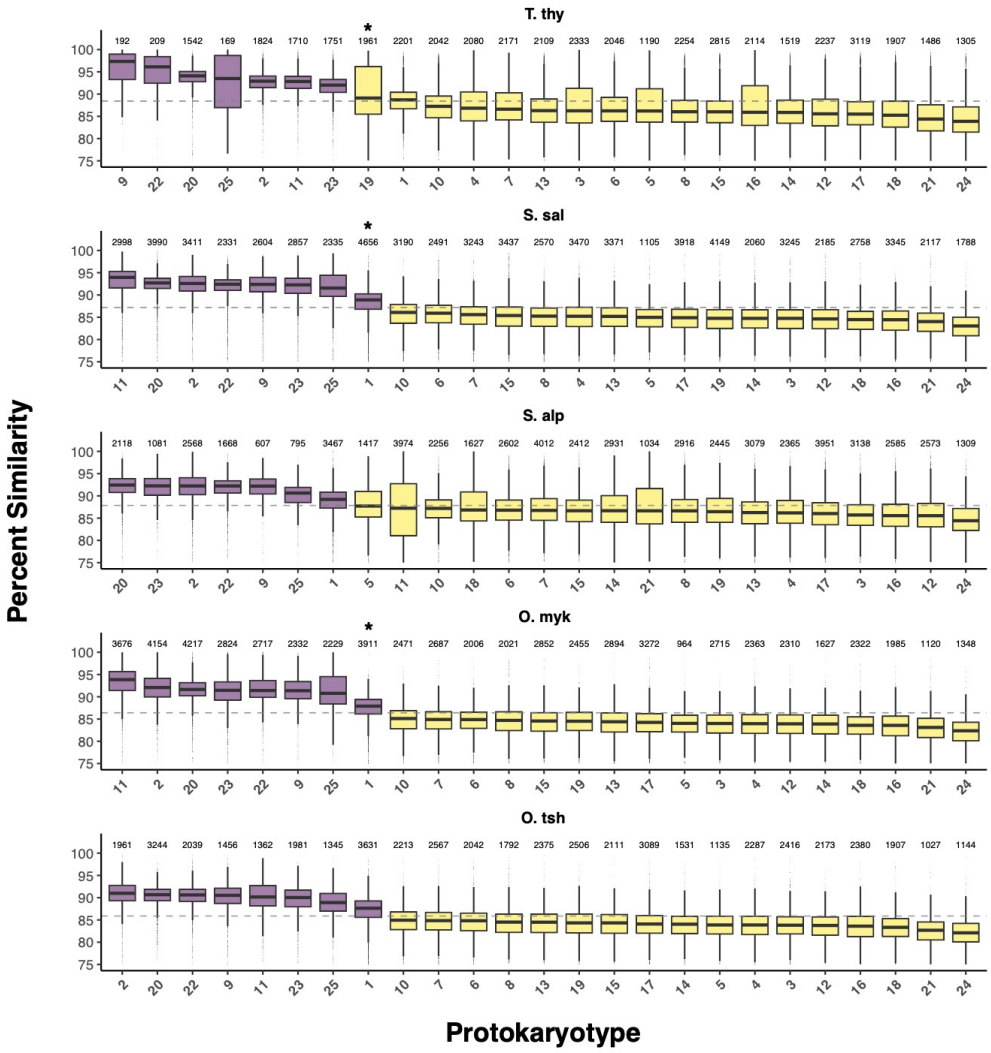


Figure 4. Distribution of protokaryotypes (PK) similarity across salmonids based on genome assemblies. For each species with a genome sequence, the percent similarity (*y* – axis) of the 25 PK pairs as shown as box plots. PK pairs are ranked from highest to lowest median similarity for each species (*x* – axis), with the average similarity of protokaryotypes presented as a dashed line. The classification of protokaryotypes by the machine learning approach described in the main text into putatively tetrasomic and disomic pairs is shown through coloring of the boxplots into purple (putatively tetrasomic) and yellow (putatively disomic). The number of alignments used in computing similarity is presented at the top of each bar. Those protokaryotypes that did not receive the highest observed voting proportion for the assigned class are indicated with an asterisk (*). PKs with high variance (e.g. PK 11 in S. alp) may be due to methodological limitations that have caused additional non-homeologous chromosome arms to be included in the comparisons (see discussion). Species abbreviations are grayling (T. thy), Atlantic salmon (S. sal), Arctic char (S. alp), rainbow trout (O. myk), and Chinook salmon (O. tsh)

Supplementary tables and figures:

S1. Sampling information for cisco (*Coregonus artedi*) families collected from Lake Huron the number of individuals used. The number of offspring shown only includes those used for mapping after the removal of individuals with low sequencing coverage. The numbers of SNPs shown include those retained following quality control filtering but before inclusion in linkage mapping.

S2. Male linkage map for cisco (*Coregonus artedi*) containing 6340 loci. Each dot represents a locus. Lengths are in centimorgans (cM).

S3. Information for each marker on the female and male cisco (*Coregonus artedi*) linkage maps. Tag is the RAD tag, and marker name is the tag followed by a designation used to differentiate duplicated loci. The "Sequence P1 column" is the sequence from the single end read for each RAD tag, and the "Sequence PE" is the sequence obtained from paired-end assemblies.

S4. MAPCOMP determination of homologous chromosome arms. The cisco male and female maps were compared with markers paired through the Atlantic salmon genome to identify homology between chromosome arms. This shows the inconsistencies and poor resolution of the male map relative to the female map, which is expected given low male recombination

S5: Probable metacentric chromosomes from the MAPCOMP analysis for coregonines.

S6. Homeologous chromosome pairs for currently available haploid linkage maps and the number of maker pairs supporting homeology (Kodama et al. 2014; Larson et al. 2015; Waples et al. 2016; Tarpey et al. 2017). Homologous chromosomes are named according to the corresponding Northern Pike linkage group Protokaryotype ID (PK).

S7. Homeologous chromosome pairs for all available salmonid genomic resources.

S8. Support for classifications from k – nearest neighbor machine learning algorithm. Cross validation indicated five nearest neighbors should be used for classification across species. The vote proportion for assignment to either putatively tetrasomic or putatively disomic categories are presented as the proportion of votes out of five at the top of each bar. Median percent similarity for each protokaryotype pair is presented as the *y* – axis and protokaryotype pairs are ranked from most similar to least (*x* – axis).