THE FIRST HAPLOID LINKAGE MAP IN A COREGONID (*COREGONUS ARTEDI*) IMPROVES KNOWLEDGE OF CHROMOSOMAL EVOLUTION AND REDIPLOIDIZATION ACROSS SALMONIDS

by

Danielle Blumstein

A Thesis

Submitted in partial fulfillment of the requirement of the degree

MASTER OF SCIENCE

IN

NATURAL RESOURCES (FISHERIES)

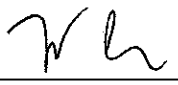
College of Natural Resources

UNIVERSITY OF WISCONSIN

Stevens Point, Wisconsin

May 2019

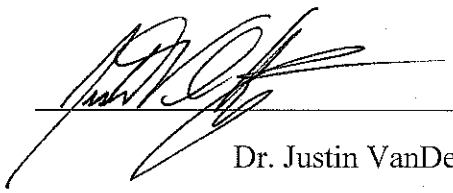
APPROVED BY THE GRADUATE COMMITTEE OF:



Dr. Wesley A. Larson, Committee Chair

Assistant Unit Leader

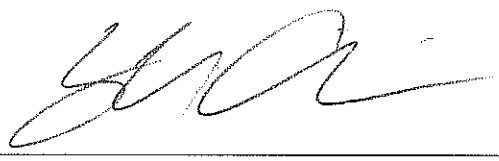
Wisconsin Cooperative Fisheries Research Unit



Dr. Justin VanDeHey

Assistant Professor

University of Wisconsin-Stevens Point



Dr. Shawn Crimmins

Assistant Professor

University of Wisconsin-Stevens Point

Acknowledgements

This project was funded by the Great Lakes Restoration Initiative (GLRI). Special thanks to Dr. Wes Larson, Dr. Wendylee Stott, Dr. Garrett McKinney, Dr. Ben Sutherland, Dr. Justin VanDeHey, Dr. Shawn Crimmins, Dr. Matt Hale, Dr. Matt Campbell, 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.

Abstract

Whole genome duplication (WGD) is an important evolutionary mechanism that can facilitate adaptation and speciation. The salmonid family represents an ideal model to study the effect of WGDs because of the extensive diversity that has evolved following WGD by autotetraploidization in the common ancestor. As salmonids evolved and diversified after the WGD, much of the genome returned to a diploid state. However, ~20% of the genome display residual tetrasomy, and the genomic processes that influence rediploization are still poorly understood. To refine the understanding of the effects of the WGD in salmonids, female (20,450 loci) and male (6,340 loci) linkage maps were constructed for cisco *Coregonus artedi*. These linkage maps identified homologous chromosomes for three coregonines and one representative species for each of *Salmo, Salvelinus,* and *Oncorhynchus* genera and the nonduplicated sister group of salmonids, *Esox*. Using this information, a cross species comparison of homeologous regions was conducted to identify regions that still exhibit residual tetrasomy, that diverged prior to speciation, and intermediate regions that are diverging independently. The further development of genomic resources in less described salmonids will aid in understanding genomic variation post-WGD. Additionally, the linkage map constructed here will facilitate future research with the aim of determining the degree of heritable genetic differences among cisco forms.

Table of Contents

Introduction

Whole genome duplication (WGD) is an important evolutionary mechanism that can facilitate reproductive isolation between populations and eventually lead to speciation (Kassahn et al. 2009; Ohno 1970). A variety of outcomes can occur after a WGD ranging from stasis to large chromosomal polymorphisms and rearrangements that fundamentally change genomic structure (Hufton and Panopoulou 2009). These changes have variable effects on a gene-specific level including loss of function at one copy of a duplicated gene, retention of function in both copies, or neofunctionalization of one or both of the gene copies (Wendel 2000). WGDs are common throughout evolutionary history (Crow et al. 2012; Kenny et al. 2016; Nossa et al. 2014) and in some taxa it has been linked to adaptive radiations (Taylor et al. 2003; Schwager et al. 2017; Robertson et al. 2017), piquing interest in understanding the potential selective advantages and evolutionary consequences of WGDs.

One of the difficulties associated with studying WGDs is the ability to properly genotype and successfully map duplicated loci, as most analysis pipelines are optimized for diploids (Limborg et al. 2016; Kodama et al. 2014; Waples et al. 2016). Unmapped duplicated regions leave large sections of the genome undiscovered, severely hampering studies investigating adaptive diversity and genome architecture in species with duplicated genomes (Limborg et al. 2016; Brown et al. 2010; Fedorova et al. 2008). Fortunately, methodological advances that employee gametic manipulation (Limborg et al. 2016) or high sequencing coverage (McKinney et al. 2018) have begun to facilitate investigation of duplicated loci. For example, haploids, which appear heterozygous at duplicated loci and homozygous at non-duplicated loci, have been used to construct high-

density genetic linkage maps that incorporate both duplicated and non-duplicated regions (Brieuc et al. 2014; Kodama et al. 2014; Larson et al. 2015; McKinney et al. 2016). The ability to identify duplicated regions on linkage maps has made it possible to conduct cross-species analysis to investigate the distribution of duplicated regions.

Salmonids are an ideal model for studying WGDs because they are derived from an ancestral species that underwent a WGD ~100 million years ago and have since diversified into a broad array of both ecologically and genetically distinct taxa (Lien et al. 2016; Macqueen and Johnston 2014). The Salmonidae family is composed of three subfamilies, Salmoninae, the true salmon including the genera *Salmo*, *Oncorhynchus*, and *Salvelinus*, Thymallinae (graylings), and Coregoninae (whitefish and ciscoes) (Crete-Lafreniere et al. 2012). Currently, there is at least one genome assembly per genus except for the Coregoninae subfamily (Savilammi et al. 2019; Varadharajan et al. 2018; Berthelot et al. 2014; Christensen et al. 2018a; Christensen et al. 2018b; Lien et al. 2016; Macqueen et al. 2017). Analysis of these genome sequences has revealed that the majority of the salmonid genome returned to a diploid state prior to the basal split of each of the three subfamilies ~50 million years ago (Robertson et al. 2017). However, the rediploidization process is still incomplete and approximately 20% of salmonid genomes still display signals of tetrasomic inheritance (Robertson et al. 2017). Regions that rediploidized prior to species diversification are referred to as Ancestral Ohnologue Resolution (AORe) regions and maintain only slight sequence similarity with ancestral ohnologues (Robertson et al. 2017). Contrastingly, the regions of the genome that were still tetraploid after speciation and maintain residual tetrasomy are referred to as Lineage-specific Ohnologue Resolution (LORe) regions (Robertson et al. 2017). These LORe

regions still display evidence of homologous recombination at ohnologues, resulting in exchange of alleles across non-homologous chromosomes (Allendorf et al. 2015). LORe regions have the potential to evolve independently among species, leading to different evolutionary trajectories. Therefore, it is important to understand the process of rediploidization across salmonid species to understand how duplications may influence evolutionary processes.

High density linkage maps, some of which include both duplicated and non-duplicated markers, are a valuable tool for understanding the WGD in salmon and have been constructed for many species across the salmonid phylogeny (Leitwein et al. 2017; Waples et al. 2016; McKinney et al. 2016). Most of these single nucleotide polymorphism (SNP)-based maps are constructed using restriction site-associated DNA (RAD) sequence data from the same restriction enzyme (*SbfI*), allowing for cross-species comparisons of markers to identify homologous chromosomes (Everett and Seeb 2014; Kodama et al. 2014; Sutherland et al. 2016; Hale et al. 2017). Although a linkage map has been constructed for two coregonids, Lake whitefish (*Coregonus clupeaformis*) (Gagnaire et al. 2013) and European whitefish (*Coregonus* sp. "Albock") (De-Kayne and Feulner 2018), these maps did not include duplicated markers. Expanding genomic resources for coregonids is important because they represent an outgroup diverged from the Salmoninae subfamily ~50 million years ago (Robertson et al. 2017) that can be used to understand and complete the systematic analysis of genome evolution among the major clades in the salmonid phylogeny.

Linkage maps can also be used to elucidate the genetic architecture of complex traits by facilitating the detection of chromosomal regions linked to loci that control

phenotypes, referred to as quantitative trait loci (QTL; Lynch and Walsh 1998). Phenotypes used in QTL analysis can include morphological, behavioral, or life history traits (Everett and Seeb 2014; Larson et al. 2015; Laporte et al. 2015). The power to identify QTL is determined by the number of individuals (Beavis et al. 1994), the type of cross, the number of genes controlling the trait, and the strength of the effect (Miller et al. 2012a; Otto and Jones 2000). Additionally, using multiple family groups can increase power for detecting QTL as well as identify shared QTL that are not dependent on family-specific genomic architecture (Hecht et al. 2012; Palti et al. 2015). The successful detection of QTL can significantly improve the understanding of adaptively important traits that are highly relevant for conservation.

This study focuses on the North American cisco *C. artedi*. Cisco are an important forage fish for top predators including lake trout (*Salvelinus namaycush*), walleye (*Sander vitreus*), and muskellunge (*Esox masquinongy*), and also provide valuable commercial and recreational fisheries (Stockwell et al. 2011). Additionally, cisco display extremely high phenotypic diversity, which has led to the definition of multiple forms based primarily on morphological evidence (Eshenroder et al. 2016; Yule et al. 2013). Overfishing and rapid environmental changes have led to a large reduction in populations and in some cases the loss of entire forms across large areas (Anneville et al. 2015). However, recent environmental shifts and a renewed focus on conservation of native species has resulted in increased interest in restoring cisco in the Laurentian Great Lakes (Zimmerman and Krueger 2009; Eshenroder et al. 2016). Key to this restoration effort is understanding the relative roles of phenotypic plasticity and adaptive genetic diversity in shaping phenotypic diversity within cisco.

Objectives

The overall goal of this research was to develop the first high density linkage map incorporating duplicated regions for cisco *Coregonus artedi*. This linkage map will serve as an important resource that can be leveraged to provide context for future studies attempting to identify the genetic basis of phenotypic differentiation in cisco. The specific research objectives were to (1) produce detailed sex-specific linkage maps for cisco, (2) identify paralogous loci, (3) determine if the number and location of LORe and AORe regions in cisco differed from other salmonids, (4) conduct QTL analysis for four phenotypic traits in four families from a wild cisco population. Additionally, this study expands on the investigation of post WGD chromosome evolution in the Salmonidae reported by Sutherland et al. (2016) by adding two species of coregonids, *C.* sp. "Albock" (De-Kayne and Feulner 2018) and *C. artedi*, to the previously investigated species. These results provide important context for understanding the evolutionary significance of the WGD in salmonids and reveal that the rediploidization process is variable both within and among genera.

Methods

Creating experimental crosses

Data from diploid and haploid families were used to build sex-specific linkage maps (Table 1, Fig. 1). Diploid crosses were constructed from *C. artedi* collected in Northern Lake Huron (45º 58'51.6" N -84 º19'40.8" W, USA, Fig. 2) during the spawning season (November) in 2015 by United States Fish and Wildlife Service crews using standardized gill net assessment methods. Haploid crosses were collected following

the same methods and from the same location during the November 2017 spawning season. Gametes were extracted from mature fish and eggs were combined directly with sperm for diploid crosses or with sperm that had been irradiated with UV light to break down the DNA (Limborg et al. 2016). UV irradiation leaves the sperm intact so that the egg can be activated but no paternal genetic material is contributed. Haploid embryos derive from maternal genetic material only. Crosses were made in the field and transported to the U. S. Geological Survey-Great Lakes Science Center, Ann Arbor, Michigan (USA) for rearing. Tissue samples (fin clips) from diploid crosses were taken when fish were ~2 years old, and tissue samples from haploids were taken from whole eggs sacrificed approximately 50 days after fertilization. All samples were preserved in a combination of 95% ethanol and 5% EDTA and sent to the University of Wisconsin - Stevens Point Molecular Conservation Genetics Lab for processing.

Laboratory analysis

DNA was extracted using QIAGEN DNeasy 96 Blood and Tissue Kit (Qiagen, Valencia, California) per the manufacturer's instructions. Quality and quantity of the extracted genomic DNA was measured using the Quant-iT PicoGreen double-stranded DNA Assay (Life Technologies) using a Biotek Plate Reader (BioTek). To confirm ploidy of haploid samples, parents and offspring were genotyped using six polymorphic microsatellite loci developed by Angers et al. (1995), Patton et al. (1997), and Rogers et al. (2004). The probability of not detecting a diploid if a diploid was present was ~0.0109 (Unpublished data, Stott). Genomic DNA from diploids and confirmed haploids was prepared for RAD sequencing using the *SbfI* restriction enzyme following the methods outlined in Ali et al. (2016) and Baird et al. (2008). Digested DNA was sheered with

NEBNext® dsDNA Fragmentase® (New England Biolabs, Inc), purified, and indexed using NEBNext® Ultra™ DNA Library Prep Kit for Illumina® per the manufacturer's instructions (New England Biolabs, Inc). Libraries were sequenced on a HiSeq4000 at the Michigan State Genomics Core Facility (East Lansing, MI).

SNP discovery and genotyping

Quality filtering, SNP identification, and genotyping was conducted using *STACKS* v.2.2 (Rochette and Catchen 2017). First, samples were demultiplexed with *process_radtags* (flags = c, -q, -r, -t, --bestrad). Stacks of similar sequences (loci) for each individual were identified with *ustacks* (flags = -m 3, -M 5, -H --max_locus_stacks 4, --model_type bounded, --bound_high 0.05, --disable-gapped), and a catalog of loci was created using a subset of individual (diploid parents = 8, haploid parents = 5, wild fish = 38, total cisco = 51) with *cstacks* (-n of 3, --disable-gapped). The 38 wild fish used to build the catalog were collected from the same geographic area using the same collection methods as listed above. The wild fish were originally included to search for a sex-identification marker using the program RADSex© 2018, but this analysis did not reveal any candidate markers (data not shown) and the genotypes of those wild fish were not included in analyses presented here.

Putative loci within each individual fish were matched against the catalog with *sstacks* (flag = --disable-gapped), *tsv2bam* was used to orient the data by SNP, and *gstacks* was used to call the loci. Only the forward reads from the paired-end data were used in *gstacks* when calling genotypes, but *gstacks* was also run separately with the forward and reverse reads to assemble longer contigs for sequence alignment and annotation. VCFtools (Danecek et al. 2011) was used with an initial VCF file to identify

and remove individuals missing greater than 30% of data. Final genotype calls were output as VCF files with *populations* (flags = -r 0.75) and each family was grouped as a separate population in the popmap file.

Maximum-likelihood based methods developed by Waples et al. (2016) were used to identify loci that could be mapped in haploid crosses and to identify potentially duplicated loci. Custom Python scripts (Python Software Foundation) that filtered the population.hap.vcf file generated by *populations* were used to identify loci that could be mapped in the diploid families. Loci missing more than 25% of data and loci that were genotyped as heterozygous in both parents of diploid families and therefore could not be reliably mapped were removed (Larson et al. 2015). Individual genotypes were exported as Lep-MAP3 input files and Rqtl2 input files. As a final step before linkage mapping, genotypes from all seven families (haploid and diploid) were combined into a single dataset to form the final female Lep-MAP3 input file and the four diploid families were combined into a single dataset to form the final male Lep-MAP3 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 differences in the recombination rates of males and females in the Salmonidae family (Sakamoto et al. 2000), a separate map was constructed for each sex (Fig. 1). 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 using the *SeparateChromosomes2* module. The LOD scores were chosen by slowly increasing the LOD value until the number of LGs stabilized and was similar to that expected for cisco

(1N= 40, Phillips et al. 1996). LOD scores for the study was 15 for the female map and 5 for the male map. Loci were then ordered within the LGs by utilizing paternal and maternal haplotypes as inheritance vectors with the *OrderMarkers2* module with minimum number of markers per LG set to 100 for the female map and 40 for the male map. LGs were reordered and markers were removed until there were no large gaps.

QTL Analysis

The aim of QTL mapping is to identify regions of the genome that contribute to important phenotypic variation. When considering multiple QTLs at once, linked QTL can be separated, interactions between QTL can be investigated, and the unexplained residual phenotypic variation is reduced (Broman 2001). Multiple-QTL analysis was performed for each family separately using the R-based software (R Core Team 2018) package *R/qtl2* (Broman et al. 2019), which utilizes a Markov model for the calculation of genotype probabilities. To quantify differences in morphometric measurements between families, standard length, total length, body depth, and weight were taken for all diploid individuals. A Haley-Knott regression was used with a linear model for genome scans to establish an association between the genotype and the phenotype. Interactions between QTL were identified with the *scan1* function in R (R Core Team 2018). QTL with a minimum LOD score of 3.00 were considered significant (Lander and Kruglyak 1995).

Synteny and Identifying AORes and LORes

$M_{AP}C_{OMP}$ (Sutherland et al. 2016) was used to identify synteny, genomic regions derived from a single ancestral genomic region, across the salmonid phylogeny. Species included in this analysis were northern pike *Esox lucius* (Rondeau et al. 2014), the

nonduplicated sister group for the salmonids, three representatives from the Coregonus genus (cisco *C. artedi,* lake whitefish *C. clupeaformis* (Gagnaire et al. 2013), and European whitefish *C.* sp. "Albock" (De-Kayne and Feulner 2018)), and representatives from the *Salmo, Salvelinus, and Oncorhynchus* genera [Atlantic salmon *Salmo salar* (Lien et al. 2016), brook trout *Salvelinus fontinalis* (Sutherland et al. 2016), and Chinook salmon *Oncorhynchus tshawytscha* (Mckinney *et al.* In Review)]. MAPCOMP compares genetic resources from two different species against a reference genome. Loci are paired if they align to the same contig or scaffold. Due to large phylogenetic distance covered in this analysis, two reference genomes were used; the Atlantic salmon genome (Lien et al. 2016) when comparing between genera and the grayling genome (Varadharajan et al. 2018) when comparing species within the *Coregonus* genus. Because salmonid chromosomal evolution is characterized by Robertsonian fusions (Phillips and Rab 2001), metacentric chromosomes in cisco were assigned by aligning cisco markers to multiple salmonid resources to identify cases where one cisco LG aligned with at least two chromosome arms in other species.

A full comparison of homeologous relationships across all genera in the salmonid family has not been conducted. Homeologous arms exhibiting residual tetraploid can be identified on linkage maps that include duplicate markers in haploid crosses. Homeologs within the cisco, coho salmon *Oncorhynchus kisutch* (Kodama et al. 2014), chum salmon *Oncorhynchus keta* (Waples et al. 2016), sockeye salmon *Oncorhynchus nerka* (Larson et al. 2015), and Chinook salmon (Mckinney *et al.* In Review) linkage maps were identified by analyzing duplicated markers that mapped to more than one LG. A chromosome pair was considered a homeolog if more than one locus pairs successfully mapped to both

LGs (Brieuc et al. 2014). Homeology can also be identified by sequence similarity at 1 Mb windows using LASTZ (Harris 2007). This analysis was conducted for Atlantic salmon (Lien et al. 2016), rainbow trout *Oncorhynchus mykiss* (Pearse et al. 2018), Arctic char *Salvelinus alpinus* (Christensen et al. 2018b), Chinook salmon (Christensen et al. 2018a), and grayling *Thymallus thymallus* (Varadharajan et al. 2018). If homeology was identified, the region was differentiated as either LORe or AORe by using available genomic resources to compare the number of marker pairs and sequence similarity supporting homeology. AORe regions were identified in linkage maps if the region had an order of magnitude less markers supporting the relationship and in genomes if sequences had less than 75% similarity. Homeologs were then visualized for cisco using a Circos plot made with the *circlize* package (Gu et al. 2014) in R (R Core Team 2018) (Fig. 4).

Results

RADseq, SNP discovering, and data filtering

Initial screening of haploid families with microsatellites showed that family P46 was 74% haploid (diploids = 47 and haploids = 133), P47 was 80% haploid (diploids = 25 and haploids=105), and P95 was 86% haploid (diploids=12 and haploids=141). RADseq data were obtained from 746 individuals across seven families, and an average of 4.1 million reads were produced for each individual (range: 1.1 million – 30.8 million reads per individual). Individuals genotyped at > 30% of loci and loci genotyped in > 75% of individuals were retained, resulting in a data set of 676 individuals (333 diploid offspring, 330 haploid offspring, and 13 parents) and 49,998 unique loci (Table 1).

Linkage Mapping

Female and male linkage maps (Fig. 3) contained a total of 22,459 unique loci, including 4,012 loci that were found on both maps (Table 2). The female map included 20,450 loci distributed across 38 LGs. Eight LGs (Cart01 – Cart08) were identified as metacentric based on homologies to two salmonid chromosome arms. On average, metacentric LGs were 85.44 cM (57.72 – 101.35 cM) and contained 739 loci (685 - 867), while acrocentric LGs were 59.12 cM (50.97 - 64.53 cM) and contained 484 loci (296 - 585). The total length of the female map was 2456.98 cM. The male map contained a total of 6,340 loci distributed across 40 LGs, corresponding to the 40 chromosomes observed in karyotype studies of coregonid fishes from the Great Lakes (Phillips et al. 1996). The number of loci per LG varied from 41 to 278, and the lengths of LGs varied from 40.54 cM to 87.14 cM. The total length of the male map was 2,357.97 cM with an average LG length of 58.95 cM. The maps showed an average female:male recombination rate of 1.06:1 (1.12:1 excluding LGs containing only male informative loci) (Table 2). Finally, 3,383 potentially duplicated loci were identified on the female linkage map. Of the potentially duplicated loci, 2,671 loci mapped to one paralog, and 709 loci mapped to two paralogs

QTL Analysis

QTL analysis for 4 phenotypic traits (total length, standard length, weight, and body depth) was conducted separately for each diploid family using 9,761 loci placed on the consensus female linkage map (Fig. 5). Family 18_P7 contained 3,435 makers suitable for QTL mapping, family 24_P7 contained 4,793 loci suitable for QTL mapping, family 24_P2 contained 3080 loci suitable for QTL mapping, and family 24_P9

contained 3,317 loci suitable for QTL mapping. 64 QTL peaks were identified with LOD scores > 3 at 50 unique genomic positions. The number of QTL varied substantially by family, phenotypic trait, and LG. 28 QTL were identified in family 18_P7, 24 QTL in family 24_P7, eight QTL in family 24_P9, and four QTL in family 24_P2. The phenotypic trait with the most QTL was total length (27), followed by standard length (16), body depth (10), and weight (11). Eight LGs contained more than two QTL and the remaining LGs contained 2 or fewer QTL. Four QTL shared three phenotypes within a family and six QTL that shared two phenotypes within a family.

Synteny and Identifying AORes and LORes

Homologous relationships were determined for all but five chromosome arms in cisco (both arms on Cart02, both arms on Cart03, one of two arms on Cart08), and three arms in European whitefish (both arms of Calb02, Calb05) using $M_{AP}C_{OMP}$. Homologous relationships for these arms were likely difficult to determine because they contained a high proportion of duplicated loci, making it difficult to differentiate homologs from homeologs (Cart03-Cart34, Cart07-Cart02, and Cart08-Cart02) (Table 2). Finally, very few alignments were documented for three ancestral chromosomes in cisco and few alignments to two ancestral chromosomes in European whitefish.

Eighteen pairs of homeologous relationships were identified in cisco, with each pair sharing between 1 and 86 duplicated loci (Fig. 4). Of those, six were likely LORes, with a large number of markers supporting homeology (42 - 86 loci, Table 3) and twelve were identified as AORes with low support (1 to 6 loci). LORe regions were compared across currently available genomic resources and haploid linkage maps. Eight LORe regions were identified in haploid maps for sockeye, chum, and coho salmon (2.1 – 2.2,

6.1 – 6.2, 9.1 – 9.2, 11.1 – 11.2, 20.1 – 20.2, 22.1 – 22.2, 23.1 – 23.1, 25.1 – 25.2). Pink salmon shared 7 of the 8 pairs but displayed low number of markers supporting the homeologous pair 25.1 – 25.5, and Chinook salmon shared 7 of 8 pairs as well but displayed low number of markers supporting the homeologous pair 6.1 -6.2 (Table 3). Seven LORe regions were identified (2.1 – 2.2, 9.1 – 9.2, 11.1 – 11.2, 20.1 – 20.2, 22.1 – 22.2, 23.1 – 23.1, 25.1 – 25.2) in the Atlantic salmon, rainbow trout, and Chinook salmon genomes with the exception of grayling, which have low sequence similarity at the homeologous pair 22.1 – 22.1 and Arctic charr, which have low sequence similarity at the homeologous pairs 11.1 – 11.2 and 25.1 – 25.2.

Discussion

The primary aim of this study was to generate the first linkage map for cisco and leverage the map to investigate genomic architecture across the salmonid phylogeny. This linkage map incorporates duplicated regions and is much denser than most RAD-based linkage maps that have previously been constructed for salmonids. Higher marker density helped resolve orthologous relationships between coregonids and the rest of the salmonids as well as identify AORe regions and LORe regions in cisco. Comparisons with other salmonids revealed that patterns of rediploidization are highly variable across genera and do not necessarily correspond with phylogeny.

Synteny

Comparisons among three coregonine species allowed homology and variation in karyotypes to be assessed across the genus. Homology was confirmed for 36 LGs between cisco, lake whitefish, and European whitefish, with both species having few to

no informative alignments for seven cisco arms. This was most likely due to a low number of mapped loci rather than different karyotypes between the three species. The cisco linkage map contained eight metacentric LGs (Cart01-Cart08). Six metacentric homologs were identified in lake whitefish (Cclu04, Cclu01, Cclu06, Cclu24, Cclu05, Cclu10) and seven were identified in European whitefish (Calb02, Calb20, Calb08, Calb09, Calb13, Calb01, Calb03). Metacentric chromosomes in salmonids are usually characterized by having at least one homeologous arm for recombination of homeologs to continue (Sutherland et al. 2016; Kodama et al. 2014). The metacentric LGs identified fit this pattern and exhibited residual tetrasomy in cisco and other species across the salmonid phylogeny (Sutherland et al. 2016).

Identifying AORes and LORes

This study was the first to map duplicated loci in a coregonine and this analysis revealed that coregonines showed similar patterns of homeology as other salmonids. Specifically, high concentrations of duplicated loci were found in six of the previously discovered eight pairs of homeologous chromosomes. Although pairs 22.1 – 22.1 and 25.1-25.2 couldn't be mapped, they may also display residual tetrasomy. It is likely that the reason these LGs couldn't be mapped is because these chromosomes are small and contain many duplicates and are therefore very difficult to map, even when duplicates are included. A similar hypothesis was presented by De-Kayne and Feulner (2018), who failed to map these LGs and also noted that the LGs were also not mapped in lake whitefish (Gagnaire et al. 2013). It is possible that these chromosomes have been lost in coregonids, but a more likely explanation is that limitations in data analysis have prevented them from being mapped.

Recent analysis of salmonid genomes shows that following the salmonid-specific WGD, most of the genome returned to a diploid state prior to the basal split of the family into multiple species. The more ancient homeologs are referred to as AORe regions (Robertson et al. 2017). However, about ~20% of the genome is still displaying residual tetrasomy in extant species and these regions are referred to as LORe regions (Robertson et al. 2017). Because the rediploidization of salmonids has co-occurred with speciation, rediploidization at LORe regions may have followed different paths in individual species. Evidence supporting this hypothesis can be obtained from two primary data sources: (1) sequence similarity in genome sequences between homeologs, and (2) markers that are mapped in each homeologous chromosome (Brieuc et al. 2014; Lien et al. 2011).

Eight homeologous regions are LORes in at least one salmonid (2.1 – 2.2, 6.1 – 6.2, 9.1 – 9.2, 11.1 – 11.2, 20.1 – 20.2, 22.1 – 22.2, 23.1 – 23.1, 25.1 – 25.2, Table 3). Four of these pairs are consistently identified as LORes in all species investigated here and are likely still undergoing frequent homeologous recombination (2.1 – 2.2, 9.1 – 9.2, 20.1 – 20.2, and 23.1 – 23.2). Of the remaining four pairs, three are LORes in all but 1-2 species. However, pair 6.1 – 6.2 appears to be a LORe region in only about half of the species. It is likely homeologous recombination between these chromosomes ceased early after speciation. It is interesting to note that patterns of similarity at these homeologs are not associated with phylogenetic splits. Pair 6.1 – 6.2 is a LORe in 4/6 representatives of the *Oncorhynchus* genus, an LORe in cisco, and an AORe in *Salmo*, *Salvelinus*, and *Thymallus*. This pattern suggests that rediploidization at this pair proceeded completely independently in different species and that, for some reason, this pair appears to cease homeologous recombination more frequently than the other seven LORe pairs.

It is important to note that different methods for identifying AORes and LORes (linkage maps vs genomes) show differences in patterns of residual tetraploidy (see 6.1 – 6.2, Table 3). Linkage maps tend to show an additional LORe region that is not found using genome sequence data. This is likely due to the length of the sequence used during each analysis. Genomes tend to use large sliding windows along chromosomes (often 1 Mb) (Harris 2007) while linkage maps are comparing alleles within ~100-150 bp RADtags. The formation of loci from RADtags is largely a function of analysis parameters, most often those set during the *ustacks* and *cstacks* modules in the program STACKs (Rochette and Catchen 2017). The small sequences analyzed by STACKs may make it possible to construct loci, even when sequence divergence in a given region is relatively large, likely making linkage maps a less conservative characterization method for determining homeology. On the other hand, genome assemblers are not optimized for paralogous and orthologous regions in polyploid genomes. Many times, duplicated regions are detected as single copies as a result of sequence collapse during the assembly process (Varadharajan et al. 2018; Alkan et al. 2011). If sequences don't collapse during assembly contigs might be fragmented and misassembled in the genome, making it difficult to differentiate between homologs and homeologs (Kyriakidou et al. 2018). This could lead to homeologous regions being missed altogether in genome sequences.

More recent linkage maps such as the map presented in this study, Mckinney *et al.* (In Review), and Tarpey et al. (2017), have 3-5 times higher marker density than "first generation" RAD maps. The increased marker density allows for the identification of AORes that would have been ambiguous in less dense linkage maps. For example, three AORe pairs were found with approximately an order or magnitude fewer markers

supporting homeology. These higher density linkage maps can also help to differentiate ambiguous patterns such as at pair 6.1 – 6.2. This is best illustrated in Chinook salmon because the species has several generations of linkage maps available as well as a genome. The first generation linkage map showed substantial support for a LORe region with 11 markers mapped between 6.1 – 6.2 and an average of 12 marker pairs across all homeologous pairs (Brieuc et al. 2014) while the second generation linkage map showed less support for a LORe region with 95 marker pairs mapped between 6.1 – 6.2 and an average of 128 marker pairs across all homeologous pairs (Mckinney *et al.* In Review). This lower sequence similarity is further supported with the Chinook salmon genome, which has less than 75% sequence similarity between 6.1 – 6.2 (Christensen et al. 2018a).

QTL Analysis

In the present study 64 QTL related to the four phenotypes among four diploid families were identified. However, inconsistences were present in the number and location of QTL between families. 10 QTL that shared a peak marker with another QTL were found within the same family. Localized QTL might indicate genomic regions of locally adapted mechanisms (Larson et al. 2015; Hecht et al. 2012; Gagnaire et al. 2013). The phenotypes that shared QTL peaks were not consistent between families, however, all four phenotypes measured in this study, total length, standard length, body depth, and weight, are correlated with one another. These traits are expected to be highly polygenic with complex inheritance patterns, therefore requiring many individuals to maximize statistical power and estimate QTL effect sizes (Ashton et al. 2017; Mackay 2001). Furthermore, the power to identify QTL is determined by the number of individuals (Beavis et al. 1994), the type of cross, the number of genes controlling the trait, and the

strength of the effect (Otto and Jones 2000; Miller et al. 2012b). This study was limited to single-generation families with less than 100 individuals per family and low size variation between individuals within families. This was mostly due constraints of tank rearing and regular culling as individuals aged.

Sexually antagonistic selection can strongly affect genetic architecture and many of the traits measured in this study show sexual dimorphism in other species. This requires sex to be a covariate in the analysis to improve the resolution of QTL effects (Sutherland et al. 2017; Hecht et al. 2012; Laporte et al. 2015). A sex marker in cisco was not successfully identified using RADseq data from wild adults. Diploid individuals were not expressing gametes and could not be sacrificed for sex determination due to ongoing research, so sex was unable to be determined. Therefore, the QTL results should be interpreted with caution because the confounding variable of sex. Without more detailed phenotypic measurements it is difficult to determine whether the QTL peaks found in this study represent true QTL markers or spurious evidence of linkage.

Conclusions

This study presents a high-density linkage map for cisco that is also the first coregonine linkage map to include duplicated markers. This study also provides the most complete description of chromosomal arrangements within the coregonid genus. Furthermore, paralogs were identified as LORe regions and AORe regions. Comparisons of these regions with other salmonid linkage maps and genomes sequences suggest that some regions are returning to a diploid state along independent timelines for each species, resulting in intermediate ohnologue regions in addition to the already described LORe

and AORe regions. The further development of high-density linkage maps and genome assemblies in a diverse group of less described salmonids will aid in understanding the genomic variation at both macro- and micro-evolutionary timescales post-WGD.

References

Ali, O.A., S.M. O'Rourke, S.J. Amish, M.H. Meek, G. Luikart *et al.*, 2016 RAD Capture (Rapture): Flexible and Efficient Sequence-Based Genotyping. *Genetics* 202 (2):389-400.

Alkan, C., S. Sajjadian, and E.E. Eichler, 2011 Limitations of next-generation genome sequence assembly. *Nat Methods* 8 (1):61-65.

Allendorf, F.W., S. Bassham, W.A. Cresko, M.T. Limborg, L.W. Seeb *et al.*, 2015 Effects of crossovers between homeologs on inheritance and population genomics in polyploid-derived salmonid fishes. *J Hered* 106 (3):217-227.

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

Anneville, O., E. Lasne, J. Guillard, R. Eckmann, J.D. Stockwell *et al.*, 2015 Impact of Fishing and Stocking Practices on Coregonid Diversity. *Food and Nutrition Sciences* 06 (11):1045-1055.

Ashton, D.T., P.A. Ritchie, and M. Wellenreuther, 2017 Fifteen years of quantitative trait loci studies in fish: challenges and future directions. *Mol Ecol* 26 (6):1465-1476.

Baird, N.A., P.D. Etter, T.S. Atwood, M.C. Currey, A.L. Shiver *et al.*, 2008 Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3 (10):e3376.

Beavis, W., O. Smith, D. Grant, and R. Fincher, 1994 Identification of quantitative trait loci using a small sample of topcrossed and F4 progeny from maize. *Crop science* 34 (4):882-896.

Berthelot, C., F. Brunet, D. Chalopin, A. Juanchich, M. Bernard *et al.*, 2014 The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nature communications* 5.

Brieuc, M.S., C.D. Waters, J.E. Seeb, and K.A. Naish, 2014 A dense linkage map for Chinook salmon (Oncorhynchus tshawytscha) reveals variable chromosomal divergence after an ancestral whole genome duplication event. *G3: Genes, Genomes, Genetics* 4 (3):447-460.

Broman, K.W., 2001 Review of statistical methods for QTL mapping in experimental crosses. *Lab animal* 30 (7).

Broman, K.W., D.M. Gatti, P. Simecek, N.A. Furlotte, P. Prins *et al.*, 2019 R/qtl2: Software for Mapping Quantitative Trait Loci with High-Dimensional Data and Multiparent Populations. *Genetics* 211 (2):495-502.

Brown, C.A., A.W. Murray, and K.J. Verstrepen, 2010 Rapid expansion and functional divergence of subtelomeric gene families in yeasts. *Current Biology* 20 (10):895-903.

Christensen, K.A., J.S. Leong, D. Sakhrani, C.A. Biagi, D.R. Minkley *et al.*, 2018a Chinook salmon (Oncorhynchus tshawytscha) genome and transcriptome. *PLoS One* 13 (4):e0195461.

Christensen, K.A., E.B. Rondeau, D.R. Minkley, J.S. Leong, C.M. Nugent *et al.*, 2018b The Arctic charr (Salvelinus alpinus) genome and transcriptome assembly. *PLoS One* 13 (9):e0204076.

Crete-Lafreniere, A., L.K. Weir, and L. Bernatchez, 2012 Framing the Salmonidae family phylogenetic portrait: a more complete picture from increased taxon sampling. *PLoS One* 7 (10):e46662.

Crow, K.D., C.D. Smith, J.F. Cheng, G.P. Wagner, and C.T. Amemiya, 2012 An independent genome duplication inferred from Hox paralogs in the American paddlefish--a representative basal ray-finned fish and important comparative reference. *Genome Biol Evol* 4 (9):937-953.

Danecek, P., A. Auton, G. Abecasis, C.A. Albers, E. Banks *et al.*, 2011 The variant call format and VCFtools. *Bioinformatics* 27 (15):2156-2158.

De-Kayne, R., and P.G.D. Feulner, 2018 A European Whitefish Linkage Map and Its Implications for Understanding Genome-Wide Synteny Between Salmonids Following Whole Genome Duplication. *G3 (Bethesda)* 8 (12):3745-3755.

Eshenroder, R.L., P. Vecsei, O.T. Gorman, D.L. Yule, T.C. Pratt *et al.*, 2016 Ciscoes of the Laurentian Great Lakes and Lake Nipigon [online].

Everett, M.V., and J.E. Seeb, 2014 Detection and mapping of QTL for temperature tolerance and body size in Chinook salmon (Oncorhynchus tshawytscha) using genotyping by sequencing. *Evol Appl* 7 (4):480-492.

Fedorova, N.D., N. Khaldi, V.S. Joardar, R. Maiti, P. Amedeo *et al.*, 2008 Genomic islands in the pathogenic filamentous fungus Aspergillus fumigatus. *PLoS genetics* 4 (4):e1000046.

Gagnaire, P.A., E. Normandeau, S.A. Pavey, and L. Bernatchez, 2013 Mapping phenotypic, expression and transmission ratio distortion QTL using RAD markers

in the Lake Whitefish (Coregonus clupeaformis). *Molecular ecology* 22 (11):3036-3048.

Gu, Z., L. Gu, R. Eils, M. Schlesner, and B. Brors, 2014 circlize Implements and enhances circular visualization in R. *Bioinformatics* 30 (19):2811-2812.

Hale, M.C., G.J. McKinney, C.L. Bell, and K.M. Nichols, 2017 Using Linkage Maps as a Tool To Determine Patterns of Chromosome Synteny in the Genus Salvelinus. *G3 (Bethesda)* 7 (11):3821-3830.

Harris, R.S., 2007 Improved pairwise alignment of genomic DNA. The Pennsylvania State University.

Hecht, B.C., F.P. Thrower, M.C. Hale, M.R. Miller, and K.M. Nichols, 2012 Genetic architecture of migration-related traits in rainbow and steelhead trout, Oncorhynchus mykiss. *G3: Genes, Genomes, Genetics* 2 (9):1113-1127.

Hufton, A.L., and G. Panopoulou, 2009 Polyploidy and genome restructuring: a variety of outcomes. *Current opinion in genetics & development* 19 (6):600-606.

Kassahn, K.S., V.T. Dang, S.J. Wilkins, A.C. Perkins, and M.A. Ragan, 2009 Evolution of gene function and regulatory control after whole-genome duplication: comparative analyses in vertebrates. *Genome research* 19 (8):1404-1418.

Kenny, N.J., K.W. Chan, W. Nong, Z. Qu, I. Maeso *et al.*, 2016 Ancestral whole-genome duplication in the marine chelicerate horseshoe crabs. *Heredity (Edinb)* 116 (2):190-199.

Kodama, M., M.S. Brieuc, R.H. Devlin, J.J. Hard, and K.A. Naish, 2014 Comparative mapping between Coho Salmon (Oncorhynchus kisutch) and three other salmonids suggests a role for chromosomal rearrangements in the retention of duplicated regions following a whole genome duplication event. *G3 (Bethesda)* 4 (9):1717-1730.

Kyriakidou, M., H.H. Tai, N.L. Anglin, D. Ellis, and M.V. Stromvik, 2018 Current Strategies of Polyploid Plant Genome Sequence Assembly. *Front Plant Sci* 9:1660.

Lander, E., and L. Kruglyak, 1995 Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature genetics* 11 (3):241.

Laporte, M., S.M. Rogers, A.M. Dion-Cote, E. Normandeau, P.A. Gagnaire *et al.*, 2015 RAD-QTL Mapping Reveals Both Genome-Level Parallelism and Different Genetic Architecture Underlying the Evolution of Body Shape in Lake Whitefish (Coregonus clupeaformis) Species Pairs. *G3 (Bethesda)* 5 (7):1481-1491.

Larson, W.A., G.J. McKinney, M.T. Limborg, M.V. Everett, L.W. Seeb *et al.*, 2015 Identification of Multiple QTL Hotspots in Sockeye Salmon (Oncorhynchus nerka) Using Genotyping-by-Sequencing and a Dense Linkage Map. *J Hered* 107 (2):122-133.

Leitwein, M., B. Guinand, J. Pouzadoux, E. Desmarais, P. Berrebi *et al.*, 2017 A dense brown trout (Salmo trutta) linkage map reveals recent chromosomal rearrangements in the Salmo genus and the impact of selection on linked neutral diversity. *G3: Genes, Genomes, Genetics* 7 (4):1365-1376.

Lien, S., L. Gidskehaug, T. Moen, B.J. Hayes, P.R. Berg *et al.*, 2011 A dense SNP-based linkage map for Atlantic salmon (Salmo salar) reveals extended chromosome homeologies and striking differences in sex-specific recombination patterns. *BMC genomics* 12 (1):615.

Lien, S., B.F. Koop, S.R. Sandve, J.R. Miller, M.P. Kent *et al.*, 2016 The Atlantic salmon genome provides insights into rediploidization. *Nature* 533 (7602):200-205.

Limborg, M.T., L.W. Seeb, and J.E. Seeb, 2016 Sorting duplicated loci disentangles complexities of polyploid genome masked by genotyping by sequencing. *Molecular ecology* 25:2117–2129.

Lynch, M., and B. Walsh, 1998 *Genetics and analysis of quantitative traits*: Sinauer Sunderland, MA.

Mackay, T.F., 2001 The genetic architecture of quantitative traits. *Annual review of genetics* 35 (1):303-339.

Macqueen, D.J., and I.A. Johnston, 2014 A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. *Proc Biol Sci* 281 (1778):20132881.

Macqueen, D.J., C.R. Primmer, R.D. Houston, B.F. Nowak, L. Bernatchez *et al.*, 2017 Functional Annotation of All Salmonid Genomes (FAASG): an international initiative supporting future salmonid research, conservation and aquaculture. *BMC genomics* 18 (1):484.

McKinney, G.J., L.W. Seeb, W.A. Larson, D. Gomez-Uchida, M.T. Limborg *et al.*, 2016 An integrated linkage map reveals candidate genes underlying adaptive variation in Chinook salmon (Oncorhynchus tshawytscha). *Mol Ecol Resour* 16 (3):769-783.

McKinney, G.J., R.K. Waples, C.E. Pascal, L.W. Seeb, and J.E. Seeb, 2018 Resolving allele dosage in duplicated loci using genotyping-by-sequencing data: A path forward for population genetic analysis. *Mol Ecol Resour* 18 (3):570-579.

Miller, M.R., J.P. Brunelli, P.A. Wheeler, S. Liu, C.E. Rexroad, 3rd *et al.*, 2012a A conserved haplotype controls parallel adaptation in geographically distant salmonid populations. *Mol Ecol* 21 (2):237-249.

Miller, M.R., J.P. Brunelli, P.A. Wheeler, S. Liu, C.E. REXROAD *et al.*, 2012b A conserved haplotype controls parallel adaptation in geographically distant salmonid populations. *Molecular Ecology* 21 (2):237-249.

Nossa, C.W., P. Havlak, J.-X. Yue, J. Lv, K.Y. Vincent *et al.*, 2014 Joint assembly and genetic mapping of the Atlantic horseshoe crab genome reveals ancient whole genome duplication. *GigaScience* 3 (1):9.

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

Otto, S.P., and C.D. Jones, 2000 Detecting the undetected: estimating the total number of loci underlying a quantitative trait. *Genetics* 156 (4):2093-2107.

Palti, Y., G. Gao, S. Liu, M. Kent, S. Lien *et al.*, 2015 The development and characterization of a 57K single nucleotide polymorphism array for rainbow trout. *Molecular Ecology Resources* 15 (3):662-672.

Patton, J.C., B.J. Gallaway, R.G. Fechhelm, and M.A. Cronin, 1997 Genetic variation of microsatellite and mitochondrial DNA markers in broad whitefish (Coregonus nasus) in the Colville and Sagavanirktok rivers in northern Alaska. *Canadian Journal of Fisheries and Aquatic Sciences* 54 (7):1548-1556.

Pearse, D.E., N.J. Barson, T. Nome, G. Gao, M.A. Campbell *et al.*, 2018 Sex-dependent dominance maintains migration supergene in rainbow trout.

Phillips, R., and P. Rab, 2001 Chromosome evolution in the Salmonidae (Pisces): an update. *Biological Reviews* 76 (1):1-25.

Phillips, R.B., K.M. Reed, and P. Rab, 1996 Revised karyotypes and chromosome banding of coregonid fishes from the Laurentian Great Lakes. *Canadian Journal of Zoology* 74:323 -329.

Python Software Foundation, Python Language Refrence, version 2.7. Available at http://www.python.org.

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

Rastas, P., 2017 Lep-MAP3: robust linkage mapping even for low-coverage whole genome sequencing data. *Bioinformatics* 33 (23):3726-3732.

Robertson, F.M., M.K. Gundappa, F. Grammes, T.R. Hvidsten, A.K. Redmond *et al.*, 2017 Lineage-specific rediploidization is a mechanism to explain time-lags between genome duplication and evolutionary diversification. *Genome Biol* 18 (1):111.

Rochette, N.C., and J.M. Catchen, 2017 Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols* 12 (12):2640.

Rogers, S.M., M.H. Marchand, and L. Bernatchez, 2004 Isolation, characterization and cross-salmonid amplification of 31 microsatellite loci in the lake whitefish (Coregonus clupeaformis, Mitchill). *Molecular Ecology Resources* 4 (1):89-92.

Rondeau, E.B., D.R. Minkley, J.S. Leong, A.M. Messmer, J.R. Jantzen *et al.*, 2014 The genome and linkage map of the northern pike (Esox lucius): conserved synteny revealed between the salmonid sister group and the Neoteleostei. *PLoS One* 9 (7):e102089.

Sakamoto, T., R.G. Danzmann, K. Gharbi, P. Howard, A. Ozaki *et al.*, 2000 A Microsatellite Linkage Map of Rainbow Trout (Oncorhynchus mykiss) Characterized by Large Sex-Specific Differences in Recombination Rates. *Genetics* 155:1331–1345.

Savilammi, T., C.R. Primmer, S. Varadharajan, R. Guyomard, Y. Guiguen *et al.*, 2019 The Chromosome-Level Genome Assembly of European Grayling Reveals Aspects of a Unique Genome Evolution Process Within Salmonids. *G3 (Bethesda).*

Schwager, E.E., P.P. Sharma, T. Clarke, D.J. Leite, T. Wierschin *et al.*, 2017 The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. *BMC Biol* 15 (1):62.

Stockwell, J.D., M.P. Ebener, J.A. Black, O.T. Gorman, T.R. Hrabik *et al.*, 2011 A Synthesis of Cisco Recovery in Lake Superior: Implications for Native Fish Rehabilitation in the Laurentian Great Lakes. *North American Journal of Fisheries Management* 29 (3):626-652.

Sutherland, B.J.G., T. Gosselin, E. Normandeau, M. Lamothe, N. Isabel *et al.*, 2016 Salmonid Chromosome Evolution as Revealed by a Novel Method for Comparing RADseq Linkage Maps. *Genome Biol Evol* 8 (12):3600-3617.

Sutherland, B.J.G., C. Rico, C. Audet, and L. Bernatchez, 2017 Sex Chromosome Evolution, Heterochiasmy, and Physiological QTL in the Salmonid Brook Charr Salvelinus fontinalis. *G3 (Bethesda)* 7 (8):2749-2762.

Tarpey, C.M., J.E. Seeb, G.J. McKinney, and L.W. Seeb, 2017 A dense linkage map for odd-year lineage pink salmon incorporating duplicated loci, pp. 1-50. School of

Aquatic and Fishery Sciences, University of Washington, Alaska Department of Fish and Game.

Taylor, J.S., I. Braasch, T. Frickey, A. Meyer, and Y. Van de Peer, 2003 Genome duplication, a trait shared by 22000 species of ray-finned fish. *Genome Res* 13 (3):382-390.

Varadharajan, S., S.R. Sandve, G.B. Gillard, O.K. Torresen, T.D. Mulugeta *et al.*, 2018 The Grayling Genome Reveals Selection on Gene Expression Regulation after Whole-Genome Duplication. *Genome Biol Evol* 10 (10):2785-2800.

Waples, R.K., L. Seeb, and J.E. Seeb, 2016 Linkage mapping with paralogs exposes regions of residual tetrasomic inheritance in chum salmon (Oncorhynchus keta). *Molecular Ecology Resources* 16 (1):17-28.

Wendel, J.F., 2000 Genome evolution in polyploids, pp. 225-249 in *Plant molecular evolution*. Springer.

Yule, D.L., S.A. Moore, M.P. Ebener, R.M. Claramunt, T.C. Pratt *et al.*, 2013 Morphometric variation among spawning cisco aggregations in the Laurentian Great Lakes: are historic forms still present? *Advances in Limnology* 64:119-132.

Zimmerman, M.S., and C.C. Krueger, 2009 An Ecosystem Perspective on Re-establishing Native Deepwater Fishes in the Laurentian Great Lakes. *North American Journal of Fisheries Management* 29:1352–1371.

Table 1: Sampling information for cisco (*Coregonus artedi*) families collected from Lake Huron and the number of individuals used in each analysis by family. Number of offspring only include those used for mapping and QTL analysis after the removal of individuals with low sequencing coverage. Number of SNPs include SNPs retained following quality control and filtering but before linkage mapping.

		Number of Individuals				
Family	Ploidy	Mapping	QTL Analysis	Sequencing method	Average reads/ individual	Number of SNPs
P46	Haploid	111	0	HiSeq4000	3.14E+06	23109
P47	Haploid	80	0	HiSeq4000	3.05E+06	22601
P95	Haploid	139	0	HiSeq4000	4.30E+06	25033
18-P7	Diploid	81	81	HiSeq4000	4.78E+06	28817
24-P2	Diploid	84	84	HiSeq4000	4.43E+06	22815
24-P7	Diploid	95	95	HiSeq4000	5.51E+06	29698
24-P9	Diploid	73	73	HiSeq4000	4.98E+06	30291
Total		663	333			49998

Table 2: Results for the consensus female and male linkage maps. Duplicated and singleton loci are from the female linkage map. LG type is denoted with acrocentric (A) and metacentric (M).

			Length		# SNPs		SNP/cM			
Cisco LG	Duplicated Loci	Singleton Loci	Female (cM)	Male (cM)	Female	Male	Female	Male	LG Type	Female:Male Recombination
Cart1	212	510	101.353	87.14	722	224	7.12	2.57	M	1.16
Cart2	394	384	97.766	61.11	778	159	7.96	2.6	M	1.60
Cart3	226	477	93.564	78.98	703	248	7.51	3.14	M	1.18
Cart4	48	547	92.451	67.8	595	194	6.44	2.86	M	1.36
Cart5	283	584	91.727	58.07	867	219	9.45	3.77	M	1.58
Cart6	108	729	91.364	68.95	837	278	9.16	4.03	M	1.33
Cart7	189	496	57.72	60.4	685	204	11.87	3.38	M	0.96
Cart8	189	532	57.559	51.62	721	170	12.53	3.29	M	1.12
Cart9	238	271	64.526	44.47	509	92	7.89	2.07	A	1.45
Cart10	21	488	63.98	49.68	509	152	7.96	3.06	A	1.29
Cart11	38	510	63.721	49.41	548	165	8.60	3.34	A	1.29
Cart12	393	177	63.179	74.69	570	80	9.02	1.07	A	0.85
Cart13	27	558	62.904	53.67	585	181	9.30	3.37	A	1.17
Cart14	73	454	62.66	59.74	527	140	8.41	2.34	A	1.05
Cart15	96	318	62.45	48.26	414	121	6.63	2.51	A	1.29
Cart16	24	456	62.253	45.8	480	153	7.71	3.34	A	1.36
Cart17	27	478	62.247	53.19	505	160	8.11	3.01	A	1.17
Cart18	25	446	62.155	42.08	471	136	7.58	3.23	A	1.48
Cart19	29	541	61.852	75.08	570	213	9.22	2.84	A	0.82
Cart20	31	536	60.188	48.36	567	224	9.42	4.63	A	1.24
Cart21	31	388	59.042	61.23	419	142	7.10	2.32	A	0.96
Cart22	20	544	58.958	49.49	564	204	9.57	4.12	A	1.19
Cart23	52	473	58.571	55.39	525	162	8.96	2.92	A	1.06
Cart24	13	556	58.377	57.23	569	195	9.75	3.41	A	1.02

Cart25	26	458	58.282	51.08	484	161	8.30	3.15	A	1.14
Cart26	30	445	58.092	50.91	475	147	8.18	2.89	A	1.14
Cart27	32	499	57.996	60.69	531	213	9.16	3.51	A	0.96
Cart28	27	444	57.793	54.44	471	141	8.15	2.59	A	1.06
Cart29	15	404	57.777	62.64	419	149	7.25	2.38	A	0.92
Cart30	29	384	57.761	71.5	413	141	7.15	1.97	A	0.81
Cart31	25	541	56.925	77.82	566	184	9.94	2.36	A	0.73
Cart32	19	478	56.557	75.17	497	177	8.79	2.35	A	0.75
Cart33	20	460	56.371	49.57	480	146	8.52	2.95	A	1.14
Cart34	262	82	55.542	83.66	344	62	6.19	0.74	A	0.66
Cart35	42	254	54.738	40.54	296	81	5.41	2	A	1.35
Cart36	19	332	54.449	45.5	351	117	6.45	2.57	A	1.20
Cart37	25	402	53.159	55.82	427	122	8.03	2.19	A	0.95
Cart38	25	431	50.97	73.21	456	171	8.95	2.34	A	0.70
Cart39	0	0	0	45.13	0	41	0.00	0.91	A	0.00
Cart40	0	0	0	58.46	0	71	0.00	1.21	A	0.00
Averages	89.03	449.13	64.66	58.95	538.16	158.50	8.36			1.06
Total	3383	17067	2456.979	2357.98	20450	6340	317.71			

Table 3. Homeologous chromosome pairs across salmonids. Regions were characterized as 'Lineage-specific Ohnologue Resolution' (L) or 'Ancestral Ohnologue Resolution' (A). Depending on the resources available, characterizations were determined by either the number of marker pairs supporting each relationship on a linkage map (listed in parentheses) or by high sequence similarity at 1 Mb windows. Northern pike 22.1-22.2 and 25.1-25.2 were not mapped in cisco (see results and discussion). Homeolog 6.1 – 6.2 showed lower support than the other pairs on the Chinook linkage map, but support was still relatively high. This pair is denoted with a "?".

		Corresponding northern pike chromosome							
Source	Resource	2.1 - 2.2	6.1 - 6.2	9.1 - 9.2	11.1 - 11.2	20.1 - 20.2	22.1 - 22.2	23.1 - 23.2	25.1 - 25.2
Cisco (current study)	Map (20450)	L (69)	L (42)	L (86)	L (46)	L (52)	-	L (58)	-
Grayling (Savilammi et al. 2019)	Genome	L	A	L	L	L	A	L	L
Atlantic salmon (Lien et al. 2016)	Genome	L	A	L	L	L	L	L	L
Arctic char (Christensen et al. 2018b)	Genome	L	A	L	A	L	L	L	A
Rainbow (Pearse et al. 2018)	Genome	L	A	L	L	L	L	L	L
Chinook (Christensen et al. 2018a)	Genome	L	A	L	L	L	L	L	L
Chinook (Mckinney *et al.* In Review)	Map (23715)	L (165)	A? (95)	L (126)	L (161)	L (123)	L (116)	L (131)	L (108)
Coho (Kodama et al. 2014)	Map (5377)	L (15)	L (17)	L (15)	L (7)	L (11)	L (10)	L (16)	L (9)
Chum (Waples et al. 2016)	Map (6119)	L (10)	L (6)	L (12)	L (6)	L (10)	L (13)	L (11)	L (8)
Sockeye (Larson et al. 2015)	Map (6262)	L (13)	L (6)	L (17)	L (10)	L (19)	L (9)	L (9)	L (11)
Pink (Tarpey et al. 2017)	Map (12275)	L (68)	L (34)	L (45)	L (65)	L (56)	L (39)	L (54)	A (9)

Table 4: MapComp results documenting homologous chromosomes for cisco (*Coregonus artedi*), lake whitefish (Gagnaire et al. 2013), European whitefish (De-Kayne and Feulner 2018), Atlantic salmon (Lien et al. 2016), brook trout (Sutherland et al. 2016), and Chinook salmon (Mckinney et al. In Review) using the northern pike chromosome names as a reference (Rondeau et al. 2014). Italics indicate problems resolving orthology, 'missing' indicates no informative alignments, and 'not.clear' indicate confounding ortholgy across multiple species.

Northern Pike	Lake Whitefish	Cisco	European Whitefish	Atlantic Salmon	Brook Trout	Chinook Salmon
1.1	Cclu28	Cart23	Calb16	Ssa20b	Sf25	Ots13q
1.2	Cclu35	Cart14	Calb33	Ssa09c	Sf38	Ots14q
2.1	Cclu04a	Cart01a	*Calb02*	Ssa26	Sf06a	Ots04q
2.2	Cclu04b	Cart12	*Calb02*	Ssa11a	Sf28	Ots12q
3.1	Cclu25	Cart25	Calb19	Ssa14a	Sf22	Ots10q
3.2	Cclu26	Cart26	Calb22	Ssa03a	Sf11	Ots28
4.1	Cclu16	Cart30	Calb29	Ssa09b	Sf33	Ots08q
4.2	Cclu29	Cart21	Calb30	Ssa05a	Sf07b	Ots21
5.1	Cclu05a	Cart06b	Calb01a	Ssa19b	Sf01a	Ots24
5.2	Cclu15	Cart18	Calb40	Ssa28	Sf27	Ots25
6.1	Cclu05b	Cart06a	Calb01b	Ssa01b	Sf01b	Ots01q
6.2	Cclu05c	Cart15	Calb27	Ssa18a	Sf36	Ots06q
7.1	Cclu13	Cart20	Calb06	Ssa13b	Sf08b	Ots09p
7.2	Cclu08	Cart19	Calb07	Ssa04b	Sf09	Ots30
8.1	Cclu36	Cart27	Calb17	Ssa23	Sf04a	Ots01p
8.2	Cclu06	Cart03a	Calb08	Ssa10a	Sf17	Ots05q
9.1	Cclu06b	*Cart03b or Cart34*	missing	Ssa02b	Sf42	Ots32
9.2	missing	*Cart03b or Car34*	*Calb05*	Ssa12a	Sf03b	Ots02q
10.1	Cclu10	Cart08	Calb20b	Ssa27	Sf23	Ots13p
10.2	Cclu24a	Cart04b	Calb09	Ssa14b	Sf34	Ots31
11.1	Cclu18	Cart05a	Calb13	Ssa06a	Sf14	Ots27
11.2	missing	Cart09	Calb34	Ssa03b	Sf08a	Ots09q
12.1	Cclu27	Cart33	Calb14	Ssa13a	Sf18	Ots22
12.2	Cclu14	Cart16	Calb28	Ssa15b	Sf30	Ots16q
13.1	Cclu34	Cart17	Calb25	Ssa24	Sf06b	Ots04p
13.2	Cclu37	Cart32	Calb31	Ssa20a	Sf40	Ots12p
14.1	Cclu04b	Cart01b	Calb02a	Ssa01c	Sf13	Ots20
14.2	Cclu33	Cart38	Calb11	Ssa11b	Sf10	Ots33
15.1	Cclu31	Cart10	Calb18	Ssa09a	Sf35	Ots08p
15.2	Cclu22	Cart31	Calb10	Ssa01a	Sf12	Ots11q
16.1	*Cclu02 or Clu03*	Cart07	Calb03b	Ssa21	Sf26	Ots26
16.2	Cclu32	Cart28	Calb21	Ssa25	Sf24	Ots03q

17.1	Cclu38	Cart24	Calb05	Ssa12b	Sf03a	Ots02p
17.2	Cclu21	Cart22	Calb12	Ssa22	Sf21	Ots07q
18.1	Cclu40	Cart29	Calb24	Ssa15a	Sf19	Ots05p
18.2	Cclu17	Cart37	Calb23	Ssa06b	Sf31	Ots18
19.1	Cclu30	Cart13	Calb04	Ssa10b	Sf15	Ots19
19.2	Cclu11	Cart11	Calb15	Ssa16a	Sf20	Ots06p
20.1	Cclu10b	*Cart08 or Cart02*	Calb36	Ssa05b	Sf07a	Ots23
20.2	Cclu01a	Cart02b	Calb20a	Ssa02a	Sf29	Ots03p
21.1	Cclu12	Cart05b	Calb13	Ssa29	Sf05b	Ots29
21.2	Cclu39	Cart36	Calb26	Ssa19a	Sf16	Ots16p
22.1	*Cclu19*	not.clear	missing	Ssa17a	Sf39	Ots07p
22.2	*Cclu19*	Cart21	missing	Ssa16b	Sf05a	Ots14p
23.1	missing	*Cart07 or Cart02*	Calb03b	Ssa07b	Sf02b	Ots15p
23.2	missing	*Cart07 or Cart02*	missing	Ssa17b	Sf37	Ots17
24.1	Cclu24b	Cart04a	Calb09b	Ssa07a	Sf02a	Ots15q
24.2	Cclu23	Cart35	Calb32	Ssa18b	Sf32	Ots10p
25.1	missing	not.clear	Calb39	Ssa04a	Sf04b	Ots34
25.2	missing	missing	missing	Ssa08a	Sf41	Ots11p

Fig. 1: Work flow to develop the linkage maps and conduct QTL analysis. Sample numbers indicate number of offspring that were successfully sequenced and passed quality control filters. Further information on each family can be found in Table 1.

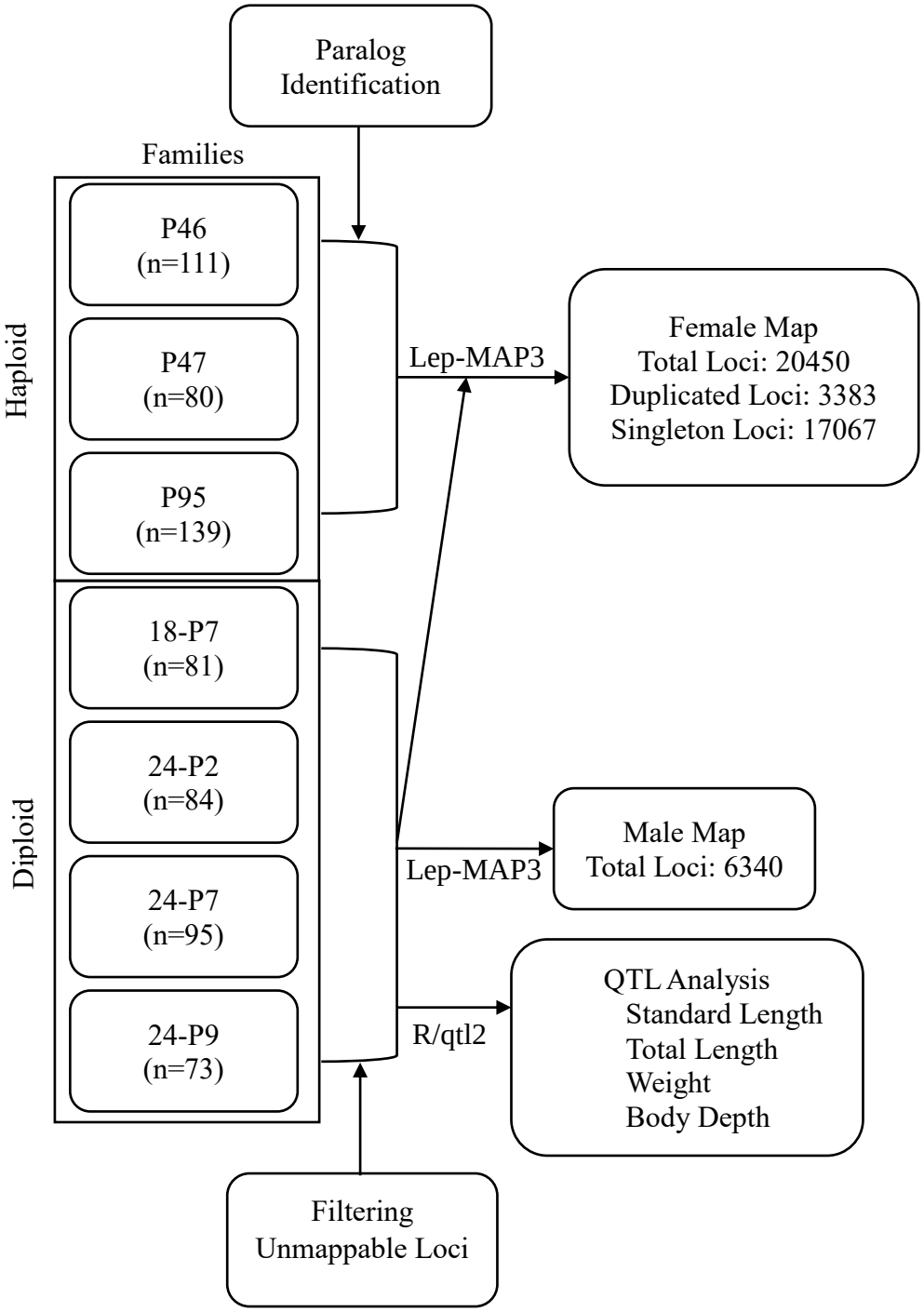


Fig. 2 Study area showing the 2015 and 2017 cisco (*Coregonus artedi*) sampling location in northern Lake Huron.

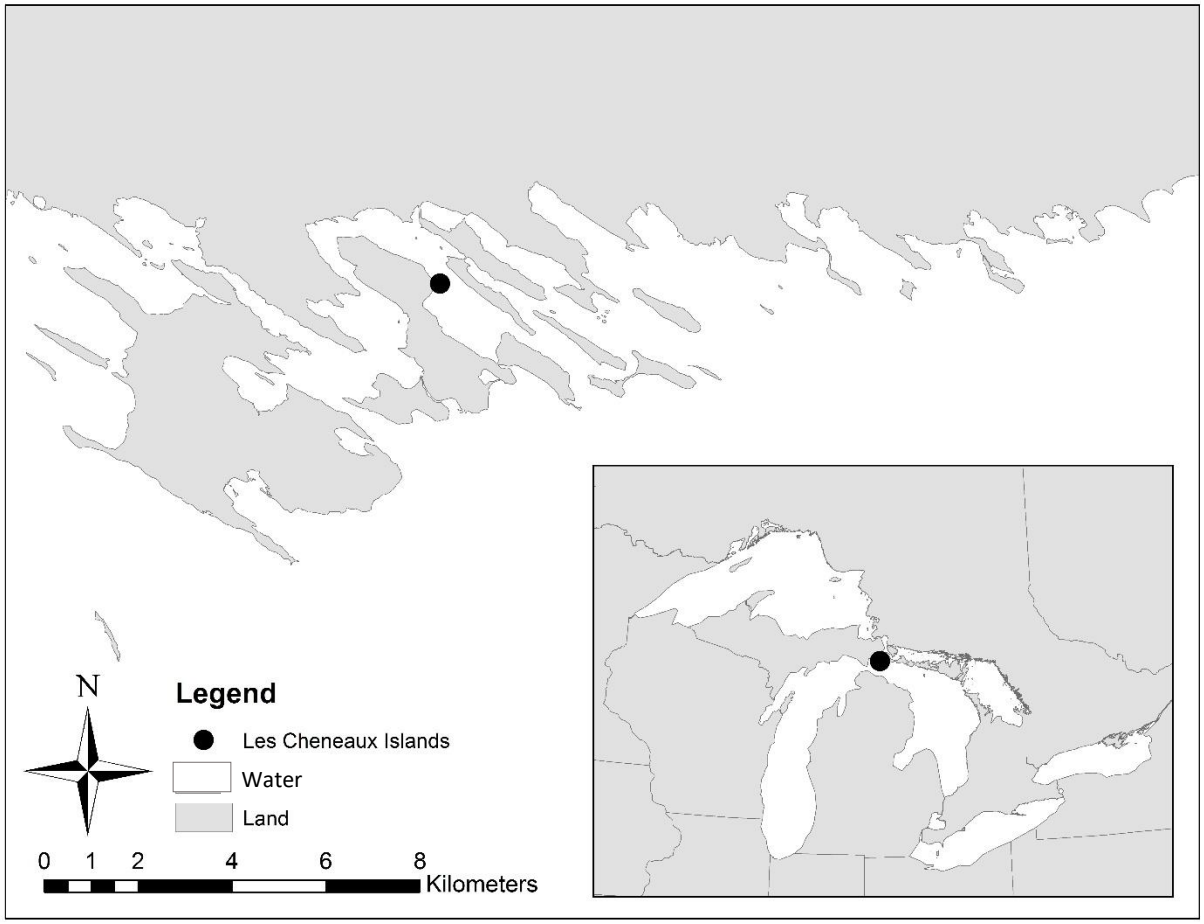
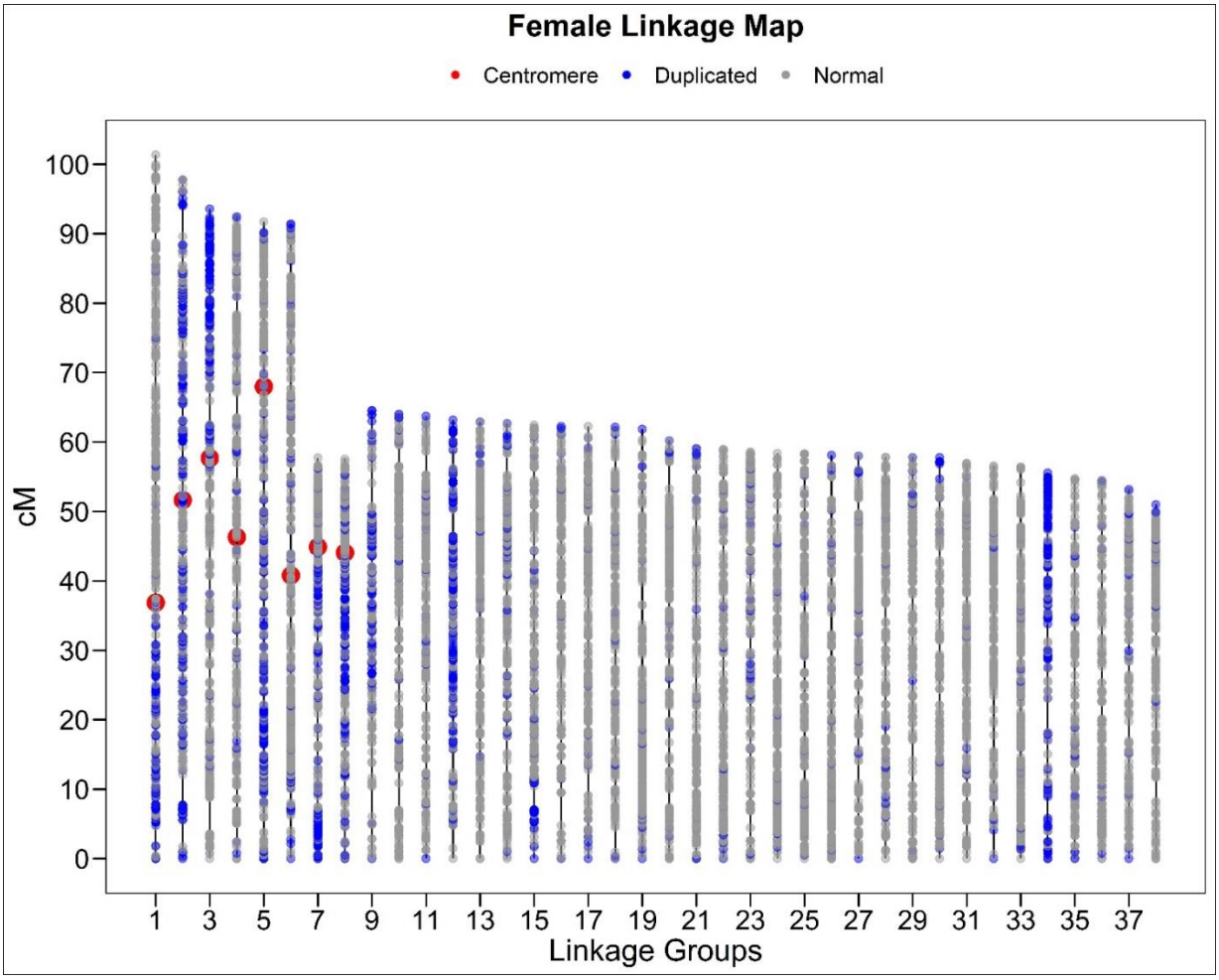


Fig. 3: Female and male linkage maps for cisco (*Coregonus artedi*) containing 20450 and 6340 loci, respectively. Each dot represents a locus, duplicated loci are blue and singleton loci are gray. Lengths are in centimorgans (cM). Approximate location of centromeres for metacentric LGs are denoted in red. Metacentric LGs were identified through homologous relationships of chromosome arms with other salmonids.



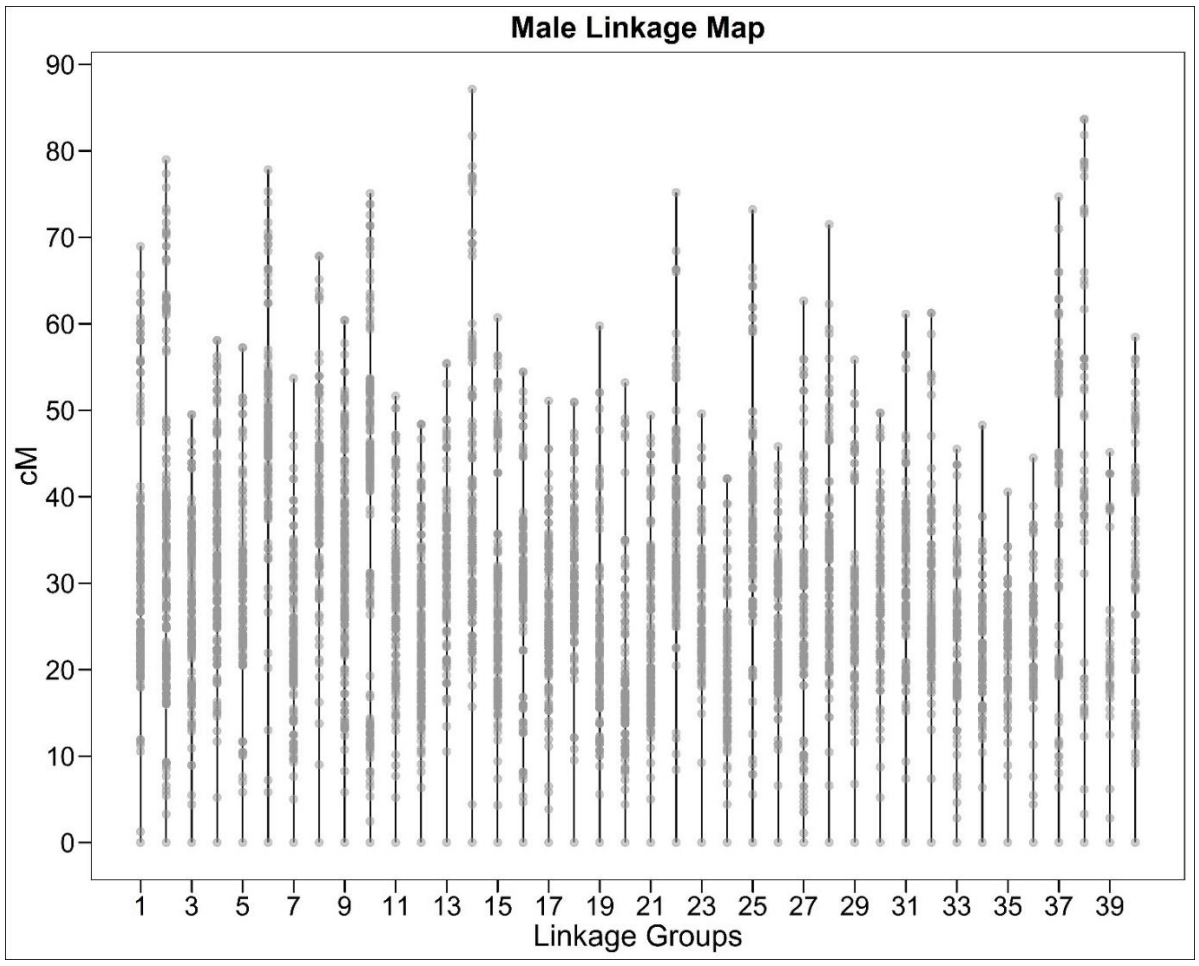

Male Linkage Map
cM
0
10
20
30
40
50
60
70
80
90
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39
Linkage Groups

Fig. 4: Circos plot of cisco (*Coregonus artedi*) LGs highlighting 'Lineage-specific Ohnologue Resolution' (LORe) and 'Ancestral-specific Ohnologue Resolution' (AORe) regions within the linkage map.

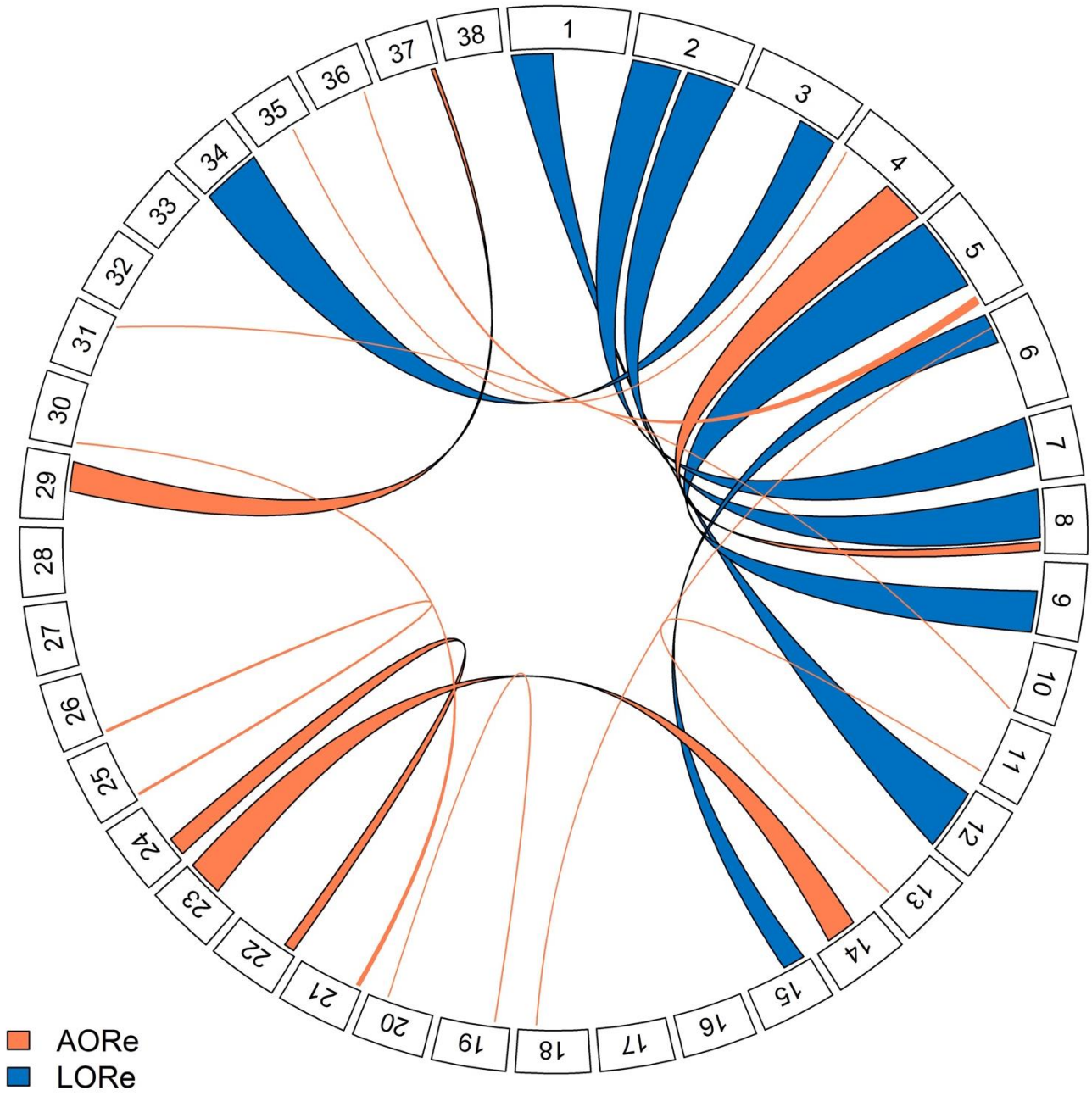
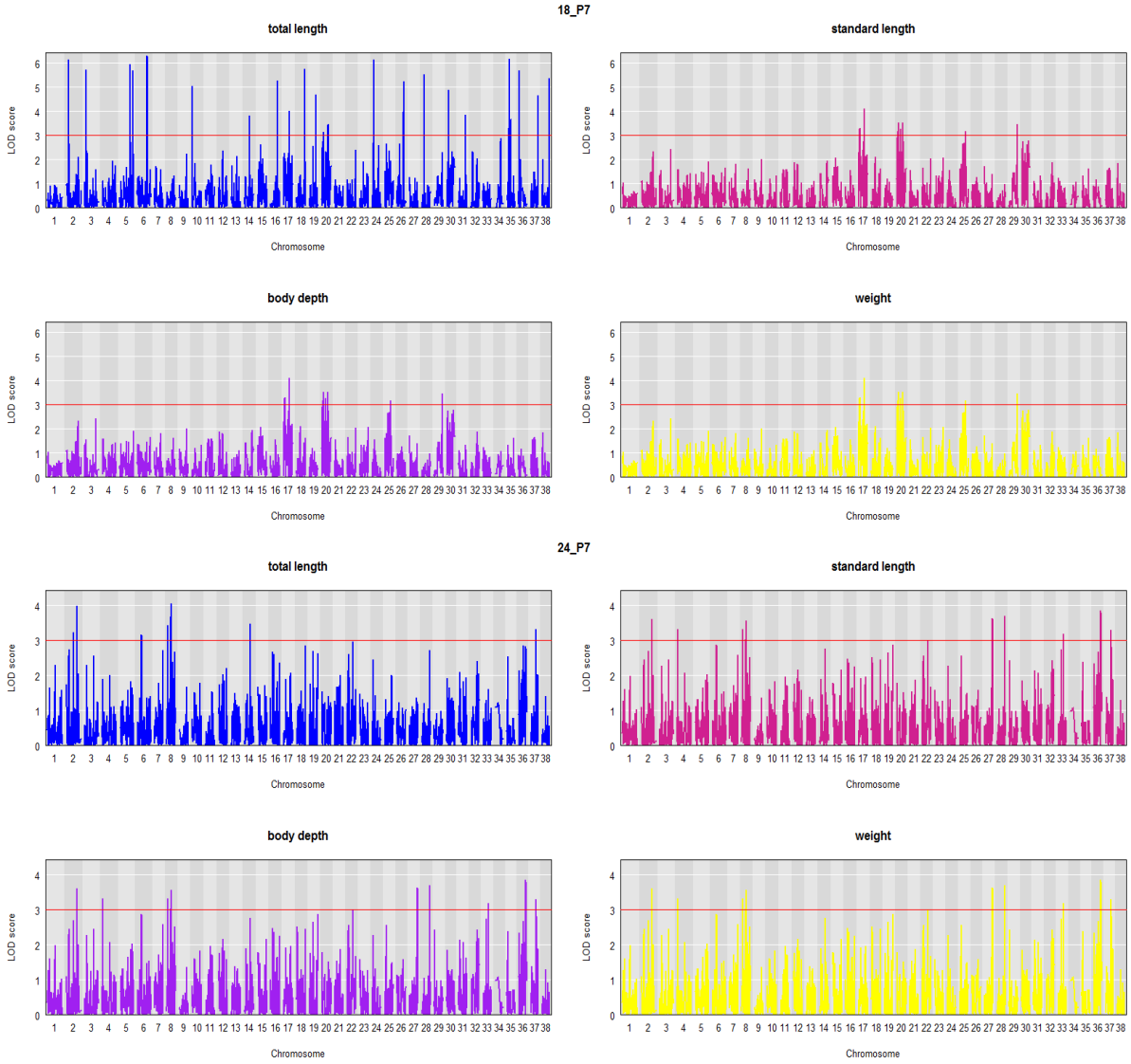
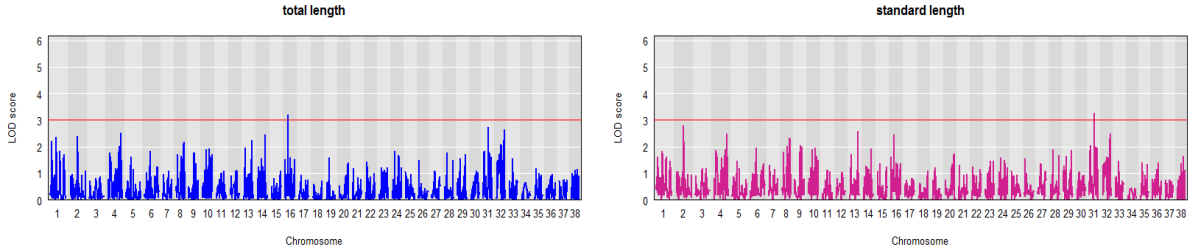


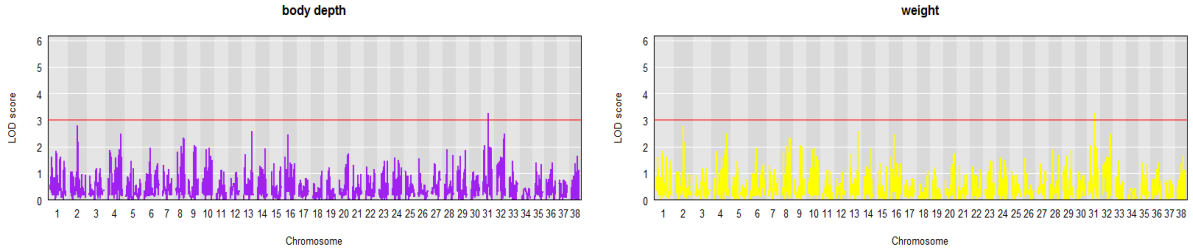
Fig. 5: The results from the QTL analysis by Haley-Knott regression for each diploid cisco (*Coregonus artedi*) family and phenotype. QTL are only considered significant if the LOD peak exceeds the threshold of three (denoted by the red horizontal line).



24_P2

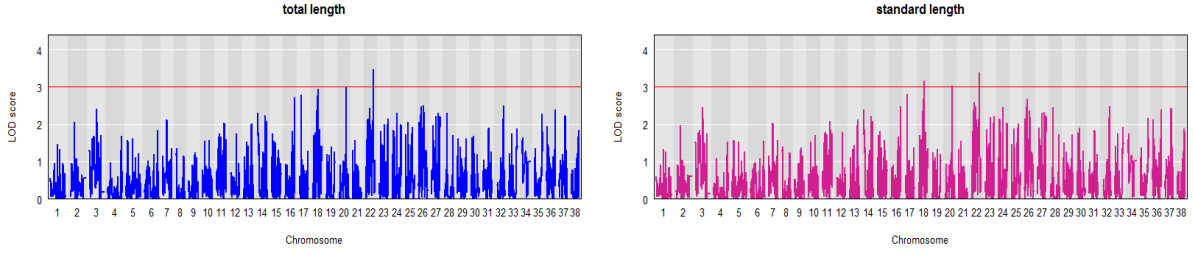

total length
LOD score
Chromosome

standard length
LOD score
Chromosome

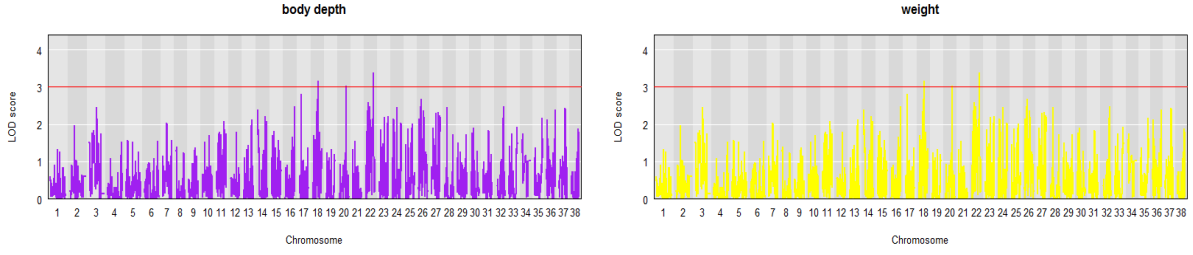

body depth
LOD score
Chromosome

weight
LOD score
Chromosome

24_P9


total length
LOD score
Chromosome

standard length
LOD score
Chromosome


body depth
LOD score
Chromosome

weight
LOD score
Chromosome

Supplementary tables:

Files available for download here: https://www.researchgate.net/project/The-First-Haploid-Linkage-Map-in-a-Coregonid-Coregonus-artedi-Improves-Knowledge-of-Chromosomal-Evolution-and-Rediploidization-Across-Salmonids

Table S1. 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.

Table S2. Identified QTL in cisco (*Coregonus artedi*) with positions for the 64 significant QTL for four phenotypes in four families.