Genomic and Fitness Consequences of Hybridization between Cutthroat and Rainbow Trout

Daniel P. Drinan

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2015

Reading Committee:

Kerry A. Naish, Chair

Steven T. Kalinowski

David A. Beauchamp

Program Authorized to Offer Degree:

School of Aquatic and Fishery Sciences

©Copyright 2015

Daniel P. Drinan

Abstract

Genomic and fitness consequences of hybridization between cutthroat and rainbow trout

Daniel P. Drinan

Chair of the Supervisory Committee: Kerry A. Naish, Associate Professor

School of Aquatic and Fishery Sciences

Hybridization is an important and common evolutionary process that can contribute to diversification, adaptation, and speciation. When species hybridize, divergent genomes are combined through recombination and may result in phenotypic changes. Such phenotypic changes may be the result of differences in chromosomal structure or adaptive divergence between the parental species and may ultimately affect fitness. Understanding how phenotypes change following hybridization, as well as the genetic mechanisms responsible for changes is critical for understanding divergent selection, speciation, and identifying populations that may be at risk from hybridization. Here, the effect of hybridization between cutthroat (*Oncorhynchus clarkii*) and rainbow trout (*O. mykiss*) on fitness was investigated using three techniques. In the first chapter, fitness influencing traits were compared among individuals at various hybridization levels. In the second chapter, genomic changes that could affect fitness were identified in the hybrid relative to rainbow trout. And, in the third chapter, correlations between reproductive success and hybridization were investigated in a wild population, as well as the genomic and ecological mechanisms responsible for those changes.

The first chapter of this dissertation aimed to identify how traits potentially involved in fitness (embryonic survival, ova size, ova energy concentration, sperm motility, burst swimming performance, juvenile survival, and juvenile growth) changed with hybridization between cutthroat and rainbow trout and whether those changes could explain previously observed reductions in reproductive success of individuals with increased rainbow trout ancestry. Using progeny from wild caught fish, differences in phenotypes based on hybridization were observed for embryonic survival, ova energy concentration, juvenile weight, and burst swimming based on ancestry. However, the correlations differed from previously observed patterns of reproductive success and likely do not explain declines in reproductive success associated with hybridization.

The second chapter of this dissertation aimed to identify how hybridization affects the genome by identifying genomic regions with changes in recombination rates in the hybrid relative to rainbow trout as well as genomic areas with excess species-specific ancestry in the hybrid. Previous studies of hybridization have observed recombination suppression in genomic regions where structural differences, such as inversions or karyotype differences, exist between parental species. Such regions may retain groups of adaptive alleles. Additionally, adaptive divergence between the parental species may result in alleles that are preferentially selected in the hybrid progeny. Identification of regions with suppressed recombination or excess species-specific ancestry would provide insight into markers that may be important to fitness and that have differentially evolved in each of the parental species. In total, eight and seven chromosomes were identified to have changes in recombination rates in the hybrid female and male relative to *O. mykiss*. Recombination was suppressed in the hybrids on two chromosomes with known structural differences between the parental species. In addition, changes in recombination rates were observed on five chromosomes with high proportions of duplicated markers and may be

due to increased homeologous chromosome pairing. Recombination patterns were similar between the sexes which suggests that hybridization affects recombination in the same way in females and males. Regions of excess species-specific ancestry covered 11 and 10% of the mapped genome in the female and male and regions of excess were evenly split between cutthroat trout and *O. mykiss*. Genetic drift may be responsible for much of the observed patterns of excess species-specific ancestry, but selection may also play a role.

The aim of the third chapter of this dissertation was to identify the fitness consequences of hybridization, mechanisms responsible for the retention of hybridization, and genomic regions correlated with changes in reproductive success in a wild population of westslope cutthroat trout hybridized with non-native rainbow trout. Adult samples from a previous study, collected over a five year period, were sequenced at 3027 loci. Increased admixture from non-native rainbow trout had a strong, negative effect on reproductive success. A decline of 53% was observed for individuals with an increased genetic contribution of 0.20 from rainbow trout. Despite apparent strong selection against rainbow trout ancestry, hybridization appears to be maintained largely by the invasion of rainbow trout from outside populations as well as the relatively high fitness of few hybrid individuals. Ten loci correlated with reproductive success were identified in females. Seven of the ten loci were linked to chromosomes and three were positioned on chromosomes. Loci linked to reproductive success were identified on chromosomes with excess species-specific ancestry in hybrid progeny (RYHyb14 and RYHyb18) as well as chromosomes with a high proportion of duplicated markers (RYHyb02) and known Robertsonian polymorphism (RYHyb20).

The research presented in this dissertation will elucidate our understanding of the phenotypic and genetic changes correlated with hybridization between rainbow and cutthroat

trout as well as identify genetic and ecological mechanisms that may be responsible for those changes. In addition, results from this study provide insight into differences in adaptive divergence and markers that may be involved in the early stages of speciation in the wild. Results could be used by managers to identify populations that are at risk from hybridization.

Contents

Chapter 1 - Effects of hybridization between nonnative Rainbow Trout and native Westslope Cutthroat Trout on fitness-related traits 1

Abstract 1

Introduction 1

Methods and Materials 3

Sampling 3

Hybridization levels in study system 4

Embryonic survival 4

Egg size 5

Egg energy concentration 6

Sperm motility 6

Juvenile survival & weight at 60 days post hatch 6

Burst swimming endurance 7

Statistical methods - embryonic and juvenile survival 8

Statistical methods - ova size, ova energy concentration, juvenile weight at 60 days, and burst swimming endurance 9

Statistical methods - sperm motility 10

Justification of ancestry estimates 11

Results 12

Embryonic survival 12

Ova size 13

Egg energy 13

Sperm motility 13

Juvenile survival at 60 days post hatch 14

Juvenile weight at 60 days post hatch 14

Burst swimming endurance 15

Discussion 15

Acknowledgements 19

References 19

Tables 27

Figures 45

Supplemental Material 56

Chapter 2 - Changes to recombination and limited segregation distortion revealed in the hybrid trout genome 61
Abstract 61
Introduction 62
Methods 68
Description of experimental crosses 68
DNA extraction, sequencing and quality screening 68
Reference database of O. mykiss RAD loci 69
O. mykiss haploid cross genotyping 70
Hybrid progeny genotyping 71
Linkage mapping 72
Segregation distortion and ancestry bias 73
Differences in relative recombination distances between O.mykiss and F_1 hybrids 74
Results 75
Reference database of O. mykiss RAD loci 75
O. mykiss haploid map 75
Hybrid maps 76
Segregation distortion and ancestry bias 77
Differences in relative recombination distances between O.mykiss and F_1 hybrids 78
Discussion 79
Creating sex-specific maps in a hybrid cross 81
Recombination suppression 83
Excess species-specific ancestry 88
Summary 89
References 90
Tables 103
Figures 106
Supplemental materials 149

Chapter 3 - Identification of genomic regions correlated with a reduction in reproductive success and the ecological mechanisms maintaining admixture in a hybridized population of trout......163
Abstract 163
Introduction 164
Methods 169
Sampling 169
Laboratory methods 169
Bioinformatic methods 170
Admixture estimation 171
Hybrid class estimation 171
Effect of hybridization on reproductive success 172
Identification of loci associated with reproductive success 173
Results 176
Genotyping and admixture estimates 176
Effect of hybridization on reproductive success 178
Maintenance of admixture based on estimates of hybrid classes 178
Identification of genomic regions correlated with reproductive success 179
Discussion 180
Admixture and reproductive success 181
Maintenance of admixture 182
Identification of genomic regions linked to reproductive success 184
Management implications 187
Conclusions 188
References 189
Tables 202
Figures 207

Chapter 1 - Effects of hybridization between nonnative Rainbow Trout and native Westslope Cutthroat Trout on fitness-related traits

Abstract

Hybridization between introduced and native fauna is a risk to native species and may threaten the long-term persistence of numerous taxa. Rainbow Trout *Oncorhynchus mykiss* has been one of the most widely introduced species around the globe and often hybridizes with native Cutthroat Trout *O. clarkii* in the Rocky Mountains. Previous work has shown that hybridization negatively affects reproductive success, but identification of traits contributing to that reduction has been elusive. Here, we used a combination of field and laboratory techniques to assess how Rainbow Trout hybridization affects traits over several stages of Westslope Cutthroat Trout development (embryonic survival, ova size, ova energy concentration, sperm motility, juvenile weight, juvenile survival, and burst swimming endurance). Embryonic survival, ova energy concentration, juvenile weight, and burst swimming endurance were all significantly correlated with ancestry. However, the correlations differed from previously observed patterns of reproductive success and likely do not explain declines in reproductive success associated with hybridization. Future investigation of additional unstudied traits and the use of different environments may shed light on the traits responsible for reproductive success in hybridized Cutthroat Trout.

Introduction

Hybridization between native and non-native species may cause many conservation problems. These include: loss of local adaptations, disruption of co-adapted gene complexes, and

ultimately threaten the persistence of numerous taxa (Rhymer and Simberloff 1996, Allendorf et al. 2001, Templeton 1986, Lynch 1996).

Hybridization with Rainbow Trout *Oncorhynchus mykiss* threatens Cutthroat Trout *O. clarkii* throughout their native range (May 2009, Gresswell 2009, Young 2008, Young 2009). Rainbow Trout has been one of the most widely introduced species on the planet (Halverson 2011), and often hybridizes with native Cutthroat Trout when in sympatry (Shepard et al. 2005, Young 2008, Gresswell 2009, Young 2009). The effect of hybridization upon Westslope Cutthroat Trout has not received much investigation (but see Allendorf et al. 2001), however, Muhfeld et al. (2009a) recently showed that hybridization between Rainbow and Westslope Cutthroat Trout *O. c. lewisi* had a very strong affect upon reproductive success. Muhlfeld et al. showed that hybrid Cutthroat Trout with 20% Rainbow Trout ancestry had half the reproductive success as genetically pure Cutthroat Trout.

The mechanisms responsible for this dramatic decline in reproductive success of hybridized Westslope Cutthroat Trout are unknown. Previous research has shown that hybridization can affects many aspects of the phenotype of Cutthroat Trout (Hawkins 1997, Bear et al. 2007, Seiler and Keeley 2009, Seiler and Keeley 2007, Ostberg et al. 2011, Henderson et al. 2000, Muhlfeld et al. 2009b, DeRito et al. 2010, Rasmussen et al. 2011), but almost nothing is known about the mechanism by which hybridization lowers fitness in the wild.

The goal of this work was to identify traits in Westslope Cutthroat Trout that caused the decline in reproductive success observed by Muhlfeld et al. (2009a). Specifically, we measured how hybridization affected embryonic survival, gamete quality (sperm motility, egg size, and egg energy concentrations), juvenile growth, juvenile survival, and juvenile burst swimming endurance.

Methods and Materials

Adult Rainbow, Westslope Cutthroat, and hybrid trout were captured in the wild. Crosses were made between captured fish to create offspring with varying levels of Cutthroat ancestry. The resulting offspring were raised in a hatchery and studied to identify how genetic ancestry affected the growth and development of hybrid fish in the hatchery.

Sampling

Fish were sampled throughout the upper Flathead River drainage using boat electrofishing in the NF Flathead River and migrant weir trapping in five streams, Abbot, Cyclone, Langford, Ivy, and Rabe creeks (Figure 1.1). Weir traps were checked every other day and electrofishing was performed when water flows allowed. Spawning fish were collected April – June 2010 and 2011. Fork length of each fish was recorded, and captured fish were housed within stream pens.

Families were created by crossing captured females and males. Captured females were checked at least twice weekly for signs of gonadal maturation. A female was deemed mature when gentle ventral pressure resulted in the release of eggs. Ova were manually stripped at maturation, split into two lots, and each lot was fertilized by one male. Pairings were designed to include as many individuals as possible and occurred both within and between populations. However, in some instances, the number of ripe individuals was limited. Therefore opportunistic pairings and multiple uses of some males occurred. Ancestry of each parent was unknown at the time of spawning. The resulting offspring were used to investigate fitness influencing traits.

Genetic ancestry was estimated for each spawned fish. A tissue sample (fin clip) was collected from each individual and stored in 95% ethanol. DNA from each sample was extracted using DNeasy Blood & Tissue Kit (QIAGEN, Venlo, Netherlands) and manufacturer specifications. For each sample, the ancestry proportion Westslope Cutthroat and Rainbow Trout was estimated by genotyping each individual at eight diagnostic single nucleotide polymorphisms (one mitochondrial, seven nuclear; Kalinowski et al. 2011) and calculating the proportion of alleles contributed from each species.

Hybridization levels in study system

Within the upper Flathead River drainage system, hybridization level varies based on an individual's population of origin (Boyer et al. 2008). Levels of hybridization are greatest at Abbot Creek (Proportion Rainbow Trout = 0.92), and decline with increased distance until reaching average proportion Rainbow Trout ancestry below 0.10 at 50 – 60 km from Abbot Creek (Boyer et al. 2008). Although average Rainbow Trout ancestry decreases away from Abbot Creek, populations are not completely homogenized and pure Westslope Cutthroat Trout persist in numerous populations through the drainage. This spatial pattern of hybridization was likely caused by the release of Rainbow Trout from a private hatchery near Abbot Creek in 1997 (Muhlfeld et al. 2009b). Additional admixture from Yellowstone Cutthroat Trout *O. c. bouvieri* has been detected in the Flathead River drainage (MTFWP 2014). However, hybridization with Yellowstone Cutthroat Trout is rare within the system (M. Boyer, Montana Department of Fish, Wildlife and Parks, unpublished data), and was not estimated as part of this study.

Embryonic survival

We estimated embryonic survival rates for all created families in a hatchery. Each half-sib family was divided and incubated at either "warm" (2010: mean = 11.8°C, SD = 0.62°C; 2011: mean = 11.9°C, SD = 0.48°C) or "cold" (2010: mean = 7.6°C, SD = 0.46°C; 2011: mean = 7.9°C, SD = 0.32°C) temperatures in Heath tray incubators (8-Tray Vertical Incubator, MariSource, Milton,Washington). Each of four Heath tray stacks contained eight trays, and each tray was divided into 48 individual compartments (hereafter "wells"). Embryos from each family were placed in two wells per incubation temperature and each well served as the experimental unit. Incubation temperatures were selected as representative of streams throughout the drainage. Water temperatures in the mainstem Flathead River range from about 6°C in April/May up to the high teens by August (Muhlfeld et al. 2009b). Water was supplied to the Heath tray stacks by two natural springs, one warm and one cool, and was mixed to achieve the desired temperatures. Embryos were monitored every two days for mortality and hatching success. Embryos that hatched or died were removed from the system. Embryos that died before the first observation were not included in the experiment as they were deemed to have died as a result of transport.

Egg size

A random subsample of about 12 eggs from each dam was individually collected to measure egg diameter. Eggs were collected while spawning and stored in formalin at room temperature. The diameter of each egg (measured to the nearest hundredth mm) was measured using Spot Software 4.1 and a LeicaDM 2000 microscope (Leica Microsystems, Wetzler, Germany). For each egg, the experimental unit, two measurements were made and the average of the two was used in the data analysis.

Egg energy concentration

A random subsample of about 10 eggs from each dam was collected to estimate the energy concentration of ovum (joules per milligram) using bomb calorimetry (6300 Automatic Isoperibol Calorimeter, Moline, Illinois). Eggs from each dam were separated into two groups, the experimental unit, of approximately equal weight. Each group was dried for eight hours at 80 °C before being homogenized. The volume of each sample was below the minimum required for our bomb calorimetry machine; therefore, we supplemented each sample with a known volume of mineral oil. Each sample was then fired and the temperature rise was measured. In order to improve the accuracy of estimates, 1 mL of phenol red was added to the bomb rinse water and titrated with sodium carbonate to obtain an acid correction factor. An additional correction factor based on mineral oil was used to estimate final gross heat of ova.

Sperm motility

Sperm quality was assessed using measures of motility duration. Sperm from each male was placed on a microscope slide, water was added to activate sperm, and "time of motility" (from sperm activation [addition of water] to 95% immotile) was recorded. Percent of sperm cells motile was estimated by observation (fraction of visible sperm which displayed any movement). To maintain consistency, one person performed all trials. Trials where sperm were still motile after 60 seconds were stopped, as increased fertilization is not expected beyond this point (Liley et al. 2002).

Juvenile survival & weight at 60 days post hatch

Juvenile growth and survival was investigated using embryos that successfully hatched from cold incubation temperatures. Families of embryos that hatched were placed in randomly assigned tanks (one tank per family). Only individuals from cold incubation temperatures were used as they were believed to have been less affected by stress and selection than individuals incubated at warmer temperatures. Tanks measured 120 × 35 × 25 cm and water temperature was maintained at 10°C. Enriched tank environments were used in an attempt to minimize the influence of captivity on survival and behavior of juvenile fish (Smith 2011). Small, fabric, artificial "plants", similar to those used by Smith (2011), were placed in each tank to emulate a more natural environment. Fish were fed to satiation each day and tanks were cleaned daily. Fish in each tank were classified as satiated when excess food was observed on the bottom of the tank. Food volumes were increased as needed. For each family, group weight and survival were measured 60 days following their modal hatching day. Average weight per individual was estimated for each family, the experimental unit, by dividing the bulk weight by the total number of individuals.

Burst swimming endurance

We investigated differences in burst swimming endurance using one year old fish and a swim chamber. Here, we define burst swimming as swimming activity that can be maintained for less than 15 sec. (Hammer 1995). One fish at a time was placed into a swim chamber (Loligo 5L swim tunnel, Tjele, Denmark) and allowed to acclimate to orientation upstream at the lowest water velocity for one hour. Following acclimation, water velocity was increased to experimental levels (previously determined using similarly sized fish, median = 7 body lengths per second). The total time from the start of the trial until a fish was pinned to the back screen for five

consecutive seconds was recorded. The same water source was used for the swim chamber and the tanks housing each test subject.

Statistical methods - embryonic and juvenile survival

Embryonic and juvenile survival were investigated using logistic mixed model regression, with R version 2.15 (R Core Team 2012) and package *lme4* (Bates et al. 2013). The full model for embryonic survival included dam fork length, familial admixture (average of parental admixture), incubation temperature, F_1 status, dam average egg calories, sire sperm motility, dam average egg diameter, and spawning year (Table 1.1). The full model for juvenile survival included familial admixture, F_1 status, starting tank density, and final tank density (Table 1.1). In both analyses, familial admixture was used rather than admixture estimates for individual progeny because the experimental unit was groups of progeny (Heath tray well and tank). F_1 status was also used in addition to familial admixture in both analyses because F_1 progeny have intact parental genomes and the effect of admixture may be different than in other hybrid groups (Muhlfeld et al. 2009a). For both analyses, correlated fixed effects were removed from each model based on variance inflation factors (VIF), using a VIF threshold of three as described by Zuur et al. (2009). Because it is possible for correlated predictors to have VIF scores below the threshold, we also removed one predictor from the full model to see if coefficients were drastically affected (a sign of co-linearity). This was done for all predictors. Changes in coefficients were minimal among models, suggesting that correlated fixed effects had been removed. The inclusion of interaction terms in each model was assessed using conditional plots. For both analyses, three random effects structures were employed to account for relatedness among embryos, dam and sire, dam only, and sire only. All three structures produced

similar results and only those from analyses containing both dam and sire will be discussed here (see supplemental tables for additional analyses). For the analysis of embryonic survival, model selection was performed by removing non-significant fixed effects one at a time until only significant fixed effects remained. This technique was used because an incomplete dataset existed for predictors, which can bias AIC (Akaike information criterion, Akaike 1974) values. In contrast, model selection for juvenile survival was performed by comparing models with all possible combination of covariates using AIC. Models with ΔAIC less than or equal to 2 compared to the best model, but not containing uninformative parameters were retained as best models (Burnham and Anderson 1998, Arnold 2010). For each model, dispersion parameters were assessed. Standard errors and *p*-values were adjusted for models with dispersion parameters greater than 1.5. *P*-values for regression coefficients were estimated using Wald's Z-test.

Statistical methods - ova size, ova energy concentration, juvenile weight at 60 days, and burst swimming endurance

Linear mixed models were used to investigate ova size, ova energy concentration, juvenile weight at 60 days, and burst swimming endurance using the package *lme4* (Bates et al. 2013). Linear mixed models assume constant variance and normality of residuals. Both were assessed graphically. Burst swimming endurance was log transformed to meet these assumptions, while no transformations were needed for the other traits. The experimental units for analyses of ova size and burst swimming endurance was the individual (egg and fish), while the analyses of ova energy concentration and juvenile weight had a group level experimental unit (about 5 eggs and average weight of all fish from a family).

For each trait, full models were created using all known explanatory variables as fixed effects (Table 1.1). For all analyses, correlated fixed effects were removed from each model based on VIF, using the same techniques as for the analyses of embryonic and juvenile survival. The initial full model for burst swimming included hybridization as a continuous variable. However, further analysis found no difference in swimming performance among hybrid individuals (no pure Westslope Cutthroat Trout were available for burst swimming analysis). Average familial admixture for hybrid families ranged from 0.03 to 0.76. Therefore, admixture was changed to a categorical predictor of Rainbow or hybrid trout.

Random effects included in each analysis varied. Models explaining ova size and ova energy concentration included dam as a random effect in order to account for the lack of independence among eggs from the same female. Models for juvenile weight at 60 days and burst swimming endurance had three random effects structures, dam only, sire only, and both dam and sire in order to account for relatedness among half-sib and full-sib progeny. The conclusions drawn between random effects structures did not differ, and for analyses of juvenile weight at 60 days and burst swimming endurance we will only present results for models that included both dam and sire (see supplemental material of other analyses).

Models containing all possible combinations of covariates from the full model were created and AIC was used to identify the best models using the same techniques as for analyses of embryonic and juvenile survival. *P*-values for all fixed effects were estimated using a Satterthwaite approximation implemented by the package *lmerTest* (Kuznetsova et al. 2013).

Statistical methods - sperm motility

Duration of sperm motility was investigated using multinomial logistic regression (MLR) in R version 2.15 (R Core Team 2012) with the package *VGAM* (Yee 2010). MLR was used because trials were stopped after 60 seconds, creating a truncated distribution. To fit a multinomial distribution, motility categories were created (Category A = 60+ sec., Category B = 50-59 sec., Category C = 40-49 sec., Category D = 30-39 sec., Category E = 20-29 sec., Category F = 10-19 sec., Category G = 1-9 sec., Category H = 0). Other categorical frameworks were used with no difference in the conclusions (results not reported). The experimental unit was each male and the full model included sire hybridization, Julian day, and sire length as fixed effects. For all analyses, correlated fixed effects were removed from each model using VIF in the same manner as previously mentioned analyses. AIC was used for model selection using the same techniques as for analyses of ova size, ova energy concentration, juvenile weight at 60 days, and burst swimming endurance. *P*-values for regression coefficients were estimated using Wald's *z*-test with a Bonferroni correction for multiple comparisons, as all pairwise comparisons between motility categories were made.

Justification of ancestry estimates

The use of few diagnostic markers to estimate genome wide hybridization may result in inaccurate estimates and could bias results. To test the effect of hybridization uncertainty on the analysis, a series of simulations and sensitivity analyses were performed. One thousand genomes for each of four known hybrid classes were simulated (pure, F_1, F_2, and backcross). Each genome contained 100 unlinked, diagnostic diploid markers. Genomes were simulated by first creating pure individuals for two species. Pure individuals, one from each species, were paired and alleles

were randomly drawn to create F_1 genomes. F_2 and backcrossed individuals were created using the same methodology, but with F_1 - F_1 and pure - F_1 individuals as their respective parents. For each simulated genome, eight markers were randomly selected and hybridization was estimated by calculating the average number of alleles contributed from each founding species. Estimated hybridization values were compared with known hybridization to identify hybrid classes that were difficult to accurately estimate with few markers. Not surprisingly, hybrid classes differed in their accuracy of estimated hybridization (Figure S1.1). Pure and F_1 individuals were consistently estimated correctly, while F_2 and backcross individuals had a greater variance in estimated admixture (Figure S1.1).

To test the influence of ancestry uncertainty on the results, the final model for each analysis was re-estimated using only individuals from accurately estimated hybrid classes (pure and F_1). No differences in the conclusions drawn from this subset of data and the entire data set existed (Supplemental material). Here we will only present the findings from the full data set.

Results

Embryonic survival

Unlike the findings of Muhlfeld et al. (2009a), we found that increased genetic contribution from Rainbow Trout was correlated with increased embryonic survival. In total, 66 families were created (Table S1.1). The genetic ancestry of created families was skewed toward Rainbow Trout, but families were created across the range of hybridization (Figure 1.2). Embryonic survival was high for all families at both incubation temperatures (median proportion survival 8°C = 0.97, 25th and 75th quantiles = 0.91, 1.00; median proportion survival 12°C = 0.95, 25th and 75th quantiles = 0.89, 1.00). Model selection resulted in a final model that included

average familial genetic proportion Westslope Cutthroat Trout and incubation temperature as fixed effects (Table 1.2). Both genetic proportion Westslope Cutthroat Trout (p-value < 0.001, z = -3.134) and incubation temperature (p-value = 0.03, z = -3.145) had significant, negative correlations with embryonic survival (Figures 1.3-5, Table 1.2).

Ova size

We observed no correlation between any model predictor and ova size. Ova size was assayed in 19 females. Average egg diameter was 5.06 mm (SD = 0.31 mm; Table S1.3). Genetic ancestry of sampled females was skewed toward Rainbow Trout (Figure 1.2). Regression analyses identified three best fit models containing year, hybridization, or both as fixed effects (Table 1.2). In all final models, no fixed effect was significantly correlated with ova size (p-value > 0.25; Table S1.4).

Egg energy

Ova energy concentration was greater in Rainbow Trout than in Westslope Cutthroat Trout. Egg energy concentration was assayed in 37 dams (Table S1.5). Joules per milligram of egg ranged from 6.48 to 10.33 with an average of 8.7 J/mg. In the best model (selected by AIC), genetic proportion Westslope Cutthroat Trout was the only fixed effect (Table 1.2) and was significantly correlated with egg energy concentration (J/mg, Table 1.6, Figures 1.3 and 1.6, p-value = 0.001). On average, Westslope Cutthroat Trout had 0.97 fewer J/mg of ova than Rainbow Trout (95% confidence interval = 0.42, 1.52).

Sperm motility

Duration of sperm motility was not correlated with hybridization, or any predictors. Motility was investigated in 48 males (Table S1.7). Time of motility was high with a median value of 43.5 sec. (25th and 75th quantiles = 33.0 and 60 sec.; Table S1.7). Genetic analyses of tested males showed that Rainbow Trout ancestry contributed more to test subjects than Westslope Cutthroat Trout ancestry (Figure 1.2). Model selection resulted in one best model containing fork length as the only fixed effect (Table 1.2). After a Bonferroni correction for multiple comparisons (pairwise comparisons among motility categories) no significant differences were detected (Table S1.8).

Juvenile survival at 60 days post hatch

No predictors were correlated with juvenile survival at 60 days post hatch. In general, juvenile survival was high for all 51 families investigated (Table S1.9). Median proportion survival was 0.83 (25th and 75th quantiles = 0.77 and 0.89). One model containing tank density was identified as best (Table 1.2). However, tank density was not significantly correlated with juvenile survival (p-value = 0.17; Table S1.10).

Juvenile weight at 60 days post hatch

Juvenile weight at 60 days post hatch was greater for families with increased genetic contribution from Westslope Cutthroat Trout. However, unlike the findings of Muhlfeld et al. (2009a), the progeny of F_1 families had intermediate phenotypes relative to pure Rainbow and Westslope Cutthroat Trout. In total, 51 families were assayed for average juvenile weights 60 days post hatch (Table S1.11). Genetic analyses showed that on average Rainbow Trout ancestry was more prevalent than Westslope Cutthroat Trout in sampled families (Figure 1.2). For all

families, average weight was 0.68 grams (SD = 0.16) 60 days after hatching. Regression analyses found that the final best model describing juvenile weight included average familial genetic proportion Westslope Cutthroat Trout as the only fixed effect (Figures 1.3 and 1.7). Average familial proportion Westslope Cutthroat Trout was positively correlated with average weight (Table 1.4, p-value < 0.001). Westslope Cutthroat Trout weighed about 0.28 g more than Rainbow Trout 60 days after modal hatch date (95% confidence interval = 0.18, 0.39).

Burst swimming endurance

Rainbow Trout swam for less time than hybrid trout. Burst swimming endurance at high water velocity was tested in 17 families (average 4.7 individuals / family, SD = 0.7, Table S1.13). Familial admixture for test subjects ranged from pure Rainbow Trout to nearly pure Westslope Cutthroat Trout (Prop. Westslope Cutthroat Trout = 0.97; Figure 1.2). Time spent swimming varied based on test subject, but had a median value of 8 sec. (25th and 75th quantiles = 8, 21 sec.) The final model describing burst swimming endurance included a categorical variable describing an individual's hybridization category (Rainbow or hybrid trout) as the only fixed effect (Table 1.2). Hybridized trout swam for significantly longer time than Rainbow Trout (Figures 1.3 and 1.8, Table 1.5, p-value = 0.008). On average, Rainbow Trout swam 0.5 times as long as hybrid trout (95% confidence interval = 0.33, 0.77). No difference in burst swimming endurance was observed among hybrid trout based on admixture (p-value = 0.50).

Discussion

Muhlfeld et al (2009a) showed that the reproductive success of hybrid Cutthroat Trout in the wild declined exponentially with increasingly amounts of Rainbow Trout ancestry. The present study sought to identify the specific traits during the life cycle of Cutthroat Trout

responsible for this decline. We examined how hybridization affected embryonic survival, ova size, ova energy concentration, sperm motility, juvenile growth, juvenile survival, and burst swimming endurance. The results of these investigations were somewhat complex, but it was clear that none of these traits are a "smoking gun" that can directly explain the low reproductive success of hybrids observed in the wild by Muhlfeld et al. (2009a). We suggest three possible reasons for this.

First, limitations in our study design may have hampered our ability to detect a link between hybridization and traits investigated as part of this study. Second, the fitness effects observed by Muhlfeld et al. (2009a) may have been caused by traits that we did not measure, or a combination of traits may be responsible for reproductive success in hybridized trout. Lastly, hybridization with Rainbow Trout may only lower the fitness of Cutthroat Trout in some environments, which do not include the hatchery environment used in this study. We discuss each of these potential explanations in turn below.

Like all studies, this investigation had limitations that may have reduced our capacity to detect relationships among the traits we studied. The most significant limitation of this study was the proportion of Rainbow Trout genes in the fish we studied; many of the fish being Rainbow-dominated hybrids. This occurred because we obtained all of the adult fish in our study from the wild. This ensured that the fish had genomes representative of hybrid fish in the wild, but had the disadvantage of not allowing us to control the ancestry of the fish included in the study. The samples we obtained were not evenly distributed across the spectrum of hybridization. Instead, Rainbow Trout were the most commonly sampled fish, and replication at each hybridization level was limited. Although we would have preferred to have more pure Cutthroat Trout and low-level hybrids in our study, we do not believe this aspect of our study design made it

unrealistically difficult to measure how hybridization affected the traits studied. Our samples spanned a wide range of ancestries, from pure Rainbow Trout to pure Westslope Cutthroat Trout for all analyses except that of burst swimming endurance and egg energy concentration. Muhlfeld et al. (2009a) found smooth and continuous responses to hybridization across all levels of hybridization, so if these effects were present in our study, we should have observed them.

A second explanation for why we did not identify the trait causing reduced fitness in hybrid Cutthroat Trout in the wild is that our study did not include the trait (or traits) responsible for this decline. The present study was the most extensive performed to date, but did not include all traits that may influence reproductive success. In particular, this study did not investigate behavioral traits, such as spawn timing and mate selection, which could be important for surviving or reproducing in the wild. Spawn timing differences have been observed between Rainbow, Cutthroat, and hybridized trout (Muhlfeld et al. 2009b, DeRito et al. 2010, Corsi et al. 2013). Rainbow Trout spawn on the rise of the spring hydrograph, often weeks to months earlier than Cutthroat Trout (Muhlfeld et al. 2009b, DeRito et al. 2010, Corsi et al. 2013). As such, high flows at the peak of runoff could scour, and ruin, Rainbow Trout redds. Additionally, Rainbow Trout redds could be disturbed by later spawning Cutthroat Trout who reproduce on the fall of the spring hydrograph (Muhlfeld et al. 2009b, DeRito et al. 2010, Corsi et al. 2013). Another unmeasured trait that may affect reproductive success is mate selection. Salmonids have been observed to have strong mate preferences. For example, studies have found mate size (Foote 1989), ornamentation (Wedekind et al. 2001), and genetic diversity (Landry et al. 2001, Evans et al. 2010) to be important to reproductive success, and therefore may play an important role in mate selection. No known studies of mate selection have been performed in hybridized Cutthroat Trout.

A third explanation for why we did not identify traits responsible for declines in reproductive success is that we conducted our work in a hatchery and hybridization does not have the same effects on traits in a hatchery environment as in the wild. Differences in genotype by environment interactions between captive and wild habitats have been well documented (Olla *et al.* 1998, Waples 1999, Araki *et al.* 2007, 2008, Pulcini *et al.* 2013), and changes in environments could produce vastly different results. For example, Einum and Fleming (1999) found that egg size in Brown Trout (*S. trutta*) influenced juvenile survival and growth in a semi-natural environment, but not in a hatchery environment (Einum and Fleming 1999). We attempted to simulate natural environments in the hatchery by using "enriched" tank environments, but it is not possible to completely replicate the wild environment and differences may have influenced the results observed. Performing such experiments in the wild would be an important next step.

Differences in results between hatchery and wild studies may also highlight how little is known about how traits translate from the hatchery to the wild (Adriaenssens and Johnsson 2010, Saikkonen *et al.* 2011). For example, Saikkonen et al. (2011) observed that growth in the hatchery was negatively correlated with survival and growth in a semi-natural environment for Atlantic salmon (*Salmo salar*). Thus, the extrapolation of findings from studies performed in hatcheries, including the present study, must be done carefully.

Although this study did not identify what caused hybrid Cutthroat Trout in the wild to have low fitness, we have shown that several simple explanations are unlikely. As we have discussed above, several plausible explanations remain untested. This includes the mating ability of hybrid Cutthroat Trout in the wild and the effect of local environments upon the fitness of hybrids. These are obvious topics for future research.

Acknowledgements

The U.S. Geological Survey and U.S. Fish & Wildlife Service provided funding through the Science Supported Partnership program. We thank Aaron Nistler, Cal Fraser, Jason Ilgen, Matt Toner, Robert Muth, Taylor Wilcox, Luke Holmquist, Vincent D'Angelo, Brady Miller, Carter Fredenberg, Durae Belcer-Daniels, Rick Hunt, Steve Glutting, and Ninh Vu for help with field and laboratory work. We thank Kotaro Ono, Jake Ferguson, David Beauchamp, and James Kenagy for help with data analysis and interpretation. We thank two anonymous reviewers and Charles Waters for their reviews of earlier versions of this manuscript. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

Adriaenssens, B., and J. I. Johnsson. 2010. Shy trout grow faster: exploring links between personality and fitness-related traits in the wild. Behavioral Ecology 22:135-143.

Akaike, H. 1974. A new look at the statistical model identification. Automatic Control, IEEE Transactions 19: 716-723.

Allendorf, F. W., R. F. Leary, P. Spruell, and J. K. Wenburg. 2001. The problems with hybrids: setting conservation guidelines. Trends in Ecology & Evolution 16:613–622.

Araki, H., W. R. Ardren, E. Olsen, B. Cooper, and M. S. Blouin. 2007. Reproductive success of captive-bred steelhead trout in the wild: evaluation of three hatchery programs in the Hood River. Conservation Biology 21:181-190.

Araki, H., B. A. Berejikian, M. J. Ford, and M. S. Blouin. 2008. Fitness of hatchery-reared salmonids in the wild. Evolutionary Applications 1:342–355.

Arnold, T. W. 2010. Uninformative parameters and model selection using Akaike's information criterion. Journal of Wildlife Management 74:1175-1178

Bates, D., M. Maechler, and B. Bolker. 2013. lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-2. http://CRAN.R-project.org/package=lme4

Bear, E. A., T. E. McMahon, and A. V. Zale. 2007. Comparative thermal requirements of Westslope Cutthroat Trout and Rainbow Trout: implications for species interactions and development of thermal protection standards. Transactions of the American Fisheries Society 136:1113-1121.

Boyer, M. C., C. C. Muhlfeld, and F. W. Allendorf. 2008. Rainbow Trout *Oncorhynchus mykiss* invasion and the spread of hybridization with native Westslope Cutthroat Trout *Oncorhynchus clarkii lewisi*. Canadian Journal of Fisheries and Aquatic Sciences 65:658-669.

Burnham, K. P., and D. R. Anderson. 1998. Model Selection and Inference: A Practical Information-Theoretical Approach. New York: Springer-Verlag.

Corsi, M. P., L. A. Eby, and C. A. Barfoot. 2013. Hybridization with Rainbow Trout alters life history traits of native Westslope Cutthroat Trout. Canadian Journal of Fisheries and Aquatic Sciences 70:95–904.

DeRito, J. N., A. V. Zale, and B. B. Shepard. 2010. Temporal reproductive separation of fluvial Yellowstone Cutthroat Trout from Rainbow Trout and hybrids in the Yellowstone River. North American Journal of Fisheries Management 30:866-886.

Einum, S., and I. A. Fleming. 1999. Maternal effects of egg size in Brown Trout *Salmo trutta*: norms of reaction to environmental quality. Proceedings of the Royal Society B 266:2095-2100.

Evans, M. L., B. D. Neff, and D. D. Heath, 2013, Behavioural and genetic analyses of mate choice and reproductive success in two Chinook salmon populations: Canadian Journal of Fisheries and Aquatic Sciences, v. 70, p. 263-270.

Foote, C. J. 1989. Female mate preference in Pacific salmon. Animal Behaviour 38:721-723.

Gresswell, R. E. 2009. Yellowstone Cutthroat Trout (*Oncorhynchus clarkii bouvieri*): a technical conservation assessment. USDA Forest Service, Rocky Mountain Region, Golden, C.A.

Halverson, A. 2011. An entirely synthetic fish: how Rainbow Trout beguiled America and overran the world. Yale University Press, New Haven, Connecticut.

Hammer, C. 1995. Fatigue and exercise tests with fish. Comparative Biochemistry and Physiology Part A: Physiology 112:1-20.

Hawkins, D. K. 1997. Hybridization between Coastal Cutthroat Trout (*Oncorhynchus clarki clarki*) and steelhead (*O. mykiss*). Doctoral dissertation. University of Washington, Seattle, Washington.

Henderson, R., J. L. Kershner, and C. A. Toline. 2000. Timing and location of spawning by nonnative wild Rainbow Trout and native Cutthroat Trout in the South Fork Snake River, Idaho, with implications for hybridization. North American Journal of Fisheries Management 20:584-596.

Kalinowski, S. T., B. J. Novak, D. P. Drinan, R. Jennings, and N. V. Vu. 2011. Diagnostic single nucleotide polymorphisms for identifying Westslope Cutthroat Trout *Oncorhynchus clarki lewisi*, Yellowstone Cutthroat Trout *Oncorhynchus clarkii bouvieri* and Rainbow Trout *Oncorhynchus mykiss*. Molecular Ecology Resources 11:389–393.

Kuznetsova, A., P. B. Brockhoff, and R. H. B. Christensen. 2013. lmerTest: Tests for random and fixed effects for linear mixed effect models lmer objects of lme4 package). R-Version:1.1-0. http://cran.r-project.org/web/packages/lmerTest/index.html.

Landry, C., D. Garant, P. Duchesne, and L. Bernatchez. 2001. 'Good genes as heterozygosity': the major histocompatibility complex and mate choice in Atlantic salmon (*Salmo salar*). Proceedings of the Royal Society of London B 268:1279-1285.

Liley , N. R., P. Tamkee, R. Tsai, and D. J. Hoysak. 2002. Fertilization dynamics in Rainbow Trout (*Oncorhynchus mykiss*): effect of male age, social experience, and sperm concentration and motility on in vitro fertilization. Canadian Journal of Fisheries and Aquatic Sciences 59:144-152.

Lynch, M. 1996. A quantitative genetic perspective on conservation issues. Pages 471-501. in J.C. Avise and J.L. Hamrick, editors. Conservation genetics: case histories from nature. Chapman and Hall, New York.

May, B. E. 2009.Westslope Cutthroat Trout status update summary. Wild Trout Enterprises, Bozeman, Montana.

MTFWP (Montana Department of Fish, Wildlife and Parks). 2012. Montana Fisheries Information System. Available: http://fwp.mt.gov/fishing/mFish/.

Muhlfeld, C. C., S. T. Kalinowski, T. E. McMahon, M. L. Taper, S. Painter, R. F. Leary, and F. W. Allendorf. 2009a. Hybridization rapidly reduces fitness of a native trout in the wild. Biology Letters 5:328-331.

Muhlfeld, C. C., T. E. McMahon, D. Belcer, and J. J. Kershner. 2009b. Spatial and temporal spawning dynamics of native Westslope Cutthroat Trout, *Oncorhynchus clarkii lewisi*, introduced Rainbow Trout, *Oncorhynchus mykiss*, and their hybrids. Canadian Journal of Fisheries and Aquatic Sciences 66:1153-1168.

Olla, B. L., M. W. Davis, and C. H. Ryer. 1998. Understanding how the hatchery environment represses or promotes the development of behavioral survival skills. Bulletin of Marine Science 62:531-550.

Ostberg, C. O., J. J. Duda, J. H. Graham, S. Zhang, K. P. Haywood, B. Miller, and T. L. Lerud. 2011. Growth, morphology, and developmental instability of Rainbow Trout, Yellowstone Cutthroat Trout, and four hybrid generations. Transactions of the American Fisheries Society 140:334–344.

Pulcini, D., P. A. Wheeler, S. Cataudella, T. Russo, and G. H. Thorgaard. 2013. Domestication shapes morphology in rainbow trout *Oncorhynchus mykiss*. Journal of Fish Biology 82:390-407.

R Core Team 2012. R: a language and environment for statistical computing. Vienna, Austria.

Rasmussen, J. B., M. D. Robinson, A. Hontela, and D. D. Heath. 2011. Metabolic traits of Westslope Cutthroat Trout, introduced Rainbow Trout and their hybrids in an ecotonal hybrid zone along an elevation gradient. Biological Journal of the Linnean Society 105:56-72.

Rhymer, J. M., and D. Simberloff. 1996. Extinction by hybridization and introgression. Annual Review of Ecology, Evolution, and Systematics 27:83–109.

Saikkonen, A., J. Kekäläinen, and J. Piironen. 2011. Rapid growth of Atlantic salmon juveniles in captivity may indicate poor performance in nature. Biological Conservation 144:2320-2327.

Seiler, S. M., and E. R. Keeley. 2007. Morphological and swimming stamina differences between Yellowstone Cutthroat Trout (*Oncorhynchus clarkii bouvieri*), Rainbow Trout (*Oncorhynchus mykiss*), and their hybrids. Canadian Journal of Fisheries and Aquatic Sciences 64:127–135.

Seiler, S. M., and E. R. Keeley. 2009. Competition between native and introduced salmonid fishes: Cutthroat Trout have lower growth rate in the presence of Cutthroat–Rainbow Trout hybrids. Canadian Journal of Fisheries and Aquatic Sciences 66:133–141.

Shepard, B. B., B. E. May, and W. Urie. 2005. Status and conservation of Westslope Cutthroat Trout within the western United States. North American Journal of Fisheries Management 25:1426-1440.

Smith, C.J. 2011. Assessment of altered rearing environments on survival and performance of hatchery-reared trout: implications for Cutthroat Trout reintroduction programs. M.Sc. thesis, Department of Ecology, Montana State University, Bozeman, M.T.

Templeton, A. R. 1986. Coadaptation and outbreeding depression. Pages 105-116 in M.E. Soule. Sinauer, editor. Conservation biology: the science of scarcity and diversity. Sunderland, Massachusettes.

Waples, R. S. 1999. Disspelling some myths about hatcheries. Fisheries 24:12-21.

Wedekind, C., R. Muller, and H. Spicher. 2001. Potential genetic benefits of mate selection in whitefish. Journal of Evolutionary Biology 14:980-986.

Yee, T. W. 2010 The VGAM package for categorical data analysis. Journal of Statistical Software 32:1–34.

Young, M. K. 2008. Colorado River Cutthroat Trout: a technical conservation assessment. USDA Forest Service, Rocky Mountain Station, Golden, C.O.

Young, M. K. 2009. Greenback Cutthroat Trout (*Oncorhynchus clarkii stomias*): a technical conservation assessment. USDA Forest Service, Rocky Mountain Station, Golden, C.O.

Zuur, A., E. N. Ieno, N. Walker, A. A. Saveliev, and G. M. Smith. 2009 Mixed effects models and extensions in ecology with R. Springer, New York

Tables

	Analysis	Full model	Final model(s)	Random effects
(A)	Embryonic survival	logit(survival) = $\beta_0 + \beta_1$female fork length + $\beta_2$familial admixture + $\beta_3$temperature + β_4F$_1$ status + $\beta_5$average egg calories + $\beta_6$motility category + $\beta_7$average egg diameter + $\beta_8$year	logit(survival) = $\beta_0 + \beta_1$familial admixture + $\beta_2$temperature	dam, sire
(B)	Egg diameter	diameter = $\beta_0 + \beta_1$admixture + $\beta_2$fork length + $\beta_3$year + β_4Julian day	diameter = $\beta_0 + \beta_1$year diameter = $\beta_0 + \beta_1$admixture + $\beta_2$year diameter = $\beta_0 + \beta_1$admixture	dam
(C)	Egg energy	energy = $\beta_0 + \beta_1$admixture + $\beta_2$fork length + $\beta_3$year + β_4Julian day	energy = $\beta_0 + \beta_1$admixture	dam
(D)	Sperm motility	logit(motility category) = $\beta_0 + \beta_1$admixture + β_2Julian day + $\beta_3$length	logit(motility category) = $\beta_0 + \beta_1$length	-
(E)	Juvenile growth	weight = $\beta_0 + \beta_1$familial admixture + β_2F$_1$ status + $\beta_3$starting tank density + $\beta_4$final tank density	weight = $\beta_0 + \beta_1$familial admixture	dam, sire
(F)	Juvenile survival	logit(survival) = $\beta_0 + \beta_1$familial admixture + β_2F$_1$ status + $\beta_3$starting tank density + $\beta_4$final tank density	logit(survival) = $\beta_0 + \beta_1$starting tank density	dam, sire
(G)	Swimming endurance	log(time) = $\beta_0 + \beta_1$fork length + $\beta_2$weight + $\beta_3$water velocity + $\beta_4$admixture category	log(time) = $\beta_0 + \beta_1$admixture category	dam, sire

Table 1.1: Full and final models after model selection. Model selection was performed using AIC (B, C, D, E, F, and G) or by dropping non-significant parameters (A).

Random effect	Variance		
Dam	0.34		
Sire	0.71		

Fixed effect	Estimate	Std. Err.	P-value
Familial admixture	-1.46	0.46	0.002
Temperature	-0.07	0.02	0.002

Table 1.2: Summary of the best model describing embryonic survival. The full model included female fork length, familial admixture, incubation temperature, F_1 status, average egg caloric unit, sire sperm motility category, average egg diameter, and spawning year as fixed effects. All models included dam and sire as random effects. The final model included only familial admixture and incubation temperature as fixed effects.

Random effect	Variance				
Dam	0.2512				

Fixed effect	Estimate	Std. Err.	D.F.	*t*-value	***P*-value**
Dam proportion Westslope Cutthroat Trout	-0.9716	0.2809	34.9900	-3.459	0.00144

Table 1.3: Summary of best model describing ova energy concentration (based on AIC). Dam proportion Westslope Cutthroat Trout was the only fixed effect in the final model.

Random	Variance				
Dam	0.0104364				
Sire	0.0019584				

Fixed effect	Estimate	Std. Err.	D.F.	t-value	P-value
Familial prop. Westslope Cutthroat Trout	0.28308	0.05203	47.69000	5.441	< 0.001

Table 1.4: Best model describing juvenile growth 60 days post hatch. The full model included familial admixture, F_1 status, starting tank density, and final tank density as fixed effects.

Random effect	Variance				
Dam	0.06988				
Sire	0.00000				

Fixed effect	Estimate	Std. Err.	D.F.	t-value	P-value
Rainbow trout	-0.695	0.2178	11.2910	-3.191	0.008

Table 1.5: Best model describing burst swimming endurance. The full model included as fixed effects fork length, weight, water velocity, and admixture category (hybrid or Rainbow Trout [yes or no] - initial investigation found no difference in swimming performance among hybrids, thus they were grouped).

Study	**Sex**	**Year**	**Avg. fork length (mm)**	**n (length)**	**Median prop. WCT**	**n (prop. WCT)**	**Avg. egg energy (J / mg)**	**n (egg energy)**
Embryonic survival	Female	2010	380 (74)	15	0.47 (0.00, 1.00)	16		13
		2011	381 (40)	20	0.00 (0.00, 0.00)	20		19
	Male	2010	259 (110)	18	0.00 (0.00, 0.67)	20		
		2011	302 (83)	25	0.00 (0.00, 0.40)	30		

Study	**Sex**	**Year**	**Avg. egg diameter (mm)**	**n (egg diameter)**	**Median motility (sec.)**	**n (motility)**
Embryonic survival	Female	2010	5.03 (0.34)	5		
		2011	5.12 (0.29)	13		
	Male	2010			39 (34, 50)	17
		2011			50 (31, 60+)	27

Table S1.1: Summary data of all samples used analysis of embryonic survival.

(A)

Random effect	Variance		
Dam	0.34		
Sire	0.71		

Fixed effect	Estimate	Std. Err.	P-value
Familial admixture	-1.46	0.46	0.002
Temperature	-0.07	0.02	0.002

(B)

Random effect	Variance		
Dam	0.7		

Fixed effect	Estimate	Std. Err.	P-value
Familial admixture	0.59	0.38	< 0.001
Temperature	-0.07	0.02	0.12
Motility			
Cat. B	0.54	0.3	0.07
Cat. C	-0.03	0.22	0.88
Cat. D	0.09	0.19	0.63
Cat. E	-0.57	0.34	0.09
Cat. G	-1.28	0.3	< 0.001
Cat. H	0.17	0.5	0.74
Year	1.09	0.33	0.001

(C)

Random effect	Variance		
Sire			

Fixed effect	Estimate	Std. Err.	P-value
F_1 status	-0.7	0.19	< 0.001
Temperature	-0.07	0.02	0.001
Motility			
Cat. B	0.62	0.75	0.41
Cat. C	-0.73	0.23	0.002
Cat. D	0.95	0.31	0.002
Cat. E	-0.14	0.66	0.83
Cat. G	-1.96	0.76	0.01
Cat. H	0.69	0.85	0.41

Year	1.1	0.34	0.001

Table S1.2: Summary of best models describing embryonic survival. All full models included female fork length, familial admixture, incubation temperature, F_1 status, average egg caloric unit, sire sperm motility category, average egg diameter, and spawning year as fixed effects. However, random effects between the analyses varied with (A) including both dam and sire, (B) dam only, and (C) sire only.

Study	Sex	Year	n	Avg. fork length (mm)	Median prop. WCT	Avg. egg diameter (mm)
Ova size	Female	2010	6	378 (37)	0.00 (0.00, 0.47)	4.92 (0.36)
		2011	13	386 (41)	0.00 (0.00, 0.00)	5.11 (0.30)

Table S1.3: Summary information for samples used in analysis of ova size.

(A)

Random effect	Variance
Dam	0.09920

Fixed effect	Estimate	Std. Err.	D.F.	*t*-value	*P*-value
Year	0.2692	0.2262	16.0140	1.190	0.251
Familial Prop. Westslope Cutthroat Trout	0.2756	0.4847	16.0930	0.569	0.578

(B)

Random effect	Variance
Dam	0.10169

Fixed effect	Estimate	Std. Err.	D.F.	*t*-value	*P*-value
Familial Prop. Westslope Cutthroat Trout	-0.13918	0.34090	17.18000	-0.408	0.688

(C)

Random effect	Variance
Dam	0.09519

Fixed effect	Estimate	Std. Err.	D.F.	*t*-value	*P*-value
Year	0.1767	0.1541	17.1030	1.147	0.267

Table S1.4: Summary of best models describing ova size (based on AIC). Each model contained either year and familial proportion Westslope Cutthroat Trout, or both. In all models, no fixed effects were identified as significant.

Study	Sex	Year	n	Avg. energy conc. (Joules / mg)	Avg. fork length (mm)	Median Julian date	Median prop. (WCT)
Ova energy concentration	Female	2010	18	8.57 (0.81)	380 (67)	May 21 (May 10, May 24)	0.40 (0.00, 0.68)
		2011	19	8.88 (0.44)	376 (55)	May 19 (May 14, May 29)	0.00 (0.00, 0.00)

Table S1.5: Summary information for samples used in analysis of ova energy concentration.

Random effect	Variance				
Dam	0.2512				

Fixed effect	Estimate	Std. Err.	D.F.	*t*-value	*P*-value
Dam proportion Westslope Cutthroat Trout	-0.9716	0.2809	34.9900	-3.459	0.00144

Table S1.6: Summary of best model describing ova energy concentration (based on AIC). Dam proportion Westslope Cutthroat Trout was the only fixed effect and dam was the only random effect.

Study	Sex	Year	n	Median motility (sec.)	Median prop. WCT	Avg. fork length (mm)	Median Julian day
Motility duration	Male	2010	19	40.0 (34.5, 49.3)	0.40 (0.00, 0.63)	263 (109)	May 21 (May 16, May 24)
		2011	29	48.0 (32.0, 60.0)	0.00 (0.00, 0.40)	310 (93)	May 19 (May 16, May 26)

Table S1.7: Summary information for samples used in the analysis of sperm motility duration.

	Cat. A	Cat. B	Cat. C	Cat. D	Cat. E	Cat. G	Cat. H
Cat. A	-						
Cat. B	0.78	-					
Cat. C	0.25	0.13	-				
Cat. D	0.51	0.24	0.75	-			
Cat. E	0.21	0.11	0.34	0.21	-		
Cat. G	0.49	0.29	0.64	0.49	0.72	-	
Cat. H	0.30	0.18	0.44	0.30	0.57	0.35	-

Table S1.8: *P*-values for significant differences in male length between the categories. The final model included male length as the only fixed effect.

Study	Year	n	Avg. prop. survival	Avg. prop. WCT	Avg. initial tank density	Avg. final tank density
Juvenile survival	2010	30	0.83 (0.78, 0.90)	0.39 (0.00, 0.64)	74 (32)	63 (30)
	2011	21	0.83 (0.71, 0.87)	0.00 (0.00, 0.24)	101 (56)	79 (46)

Table S1.9: Summary information for samples used in analysis of juvenile survival.

(A)

Random effects	Variance		
Dam	0.1888		
Sire	0.2268		

Fixed effect	Estimate	Std. Err.	P-value
Starting tank density	0.004	0.003	0.17

(B)

Random effects	Variance		
Dam	0.3974		

Fixed effect	Estimate	Std. Err.	P-value
F_1 status	0.2156	0.2243	0.336

(C)

Random effects	Variance		
Dam	0.3979		

Fixed effect	Estimate	Std. Err.	P-value
Familial proportion westslope cutthroat trout	0.0018	0.3412	0.996

(D)

Random effects	Variance		
Sire	0.514		

Fixed effect	Estimate	Std. Err.	P-value
F_1 status	0.3242	0.1954	0.0971

(E)

Random effects	Variance		
Sire	0.5256		

Fixed effect	Estimate	Std. Err.	P-value
Familial proportion westslope cutthroat trout	0.3947	0.2853	0.167

Table S1.10: Best models describing juvenile survival at 60 days post median hatch date. All analyses included familial admixture, F_1 status, starting tank density, and final tank density as fixed effects in the full model. However, three random effects structures were employed dam and sire (A), dam only (B and C), and sire only (D and E).

Study	Year	n	Avg. weight (g)	Avg. prop. WCT	Avg. initial tank density	Avg. final tank density
Juvenile weight	2010	30	0.72 (0.17)	0.39 (0.00, 0.64)	74 (32)	63 (30)
	2011	21	0.63 (0.11)	0.00 (0.00, 0.24)	101 (56)	79 (46)

Table S1.11: Summary of samples used for analyses of juvenile weight.

(A)

Random	Variance				
Dam	0.0104364				
Sire	0.0019584				

Fixed effect	Estimate	Std. Err.	D.F.	t-value	P-value
Familial prop. Westslope Cutthroat Trout	0.28308	0.05203	47.69000	5.441	< 0.001

(B)

Random	Variance
Dam	0.010281

Fixed effect	Estimate	Std. Err.	D.F.	t-value	P-value
Familial proportion Westslope Cutthroat Trout	0.28980	0.05279	47.82000	5.49	< 0.001

(C)

Random	Variance
Sire	0.007097

Fixed effect	Estimate	Std. Err.	D.F.	t-value	P-value
Familial proportion Westslope Cutthroat Trout	0.34615	0.04932	46.87000	7.018	< 0.001

Table S1.12: Best models describing juvenile growth 60 days post hatch. Full models included familial admixture, F_1 status, starting tank density, and final tank density as fixed effects. Random effects included dam and sire (A), dam only (B), and sire only (C).

Study	**Year**	**N (families)**	**Avg. no. individuals**	**Avg. water velocity (m / sec.)**	**Median time (sec.)**	**Avg. fork length (mm)**	**Median prop. WCT**
Swimming endurance	2011	17	4.7 (0.7)	0.63 (0.06)	15 (8, 21)	89.9 (13.9)	0.44 (0.00, 0.67)

Table S1.13: Summary information for samples used in analysis of swimming endurance.

(A)

Random effect	Variance				
Dam	0.06988				
Sire	0.00000				

Fixed effect	Estimate	Std. Err.	D.F.	t-value	P-value
Rainbow trout	-0.695	0.2178	11.2910	-3.191	0.008

(B)

Random effect	Variance				
Dam	0.06988				

Fixed effect	Estimate	Std. Err.	D.F.	t-value	P-value
Rainbow trout	-0.6950	0.2178	11.2910	-3.191	0.008

(C)

Random effect	Variance				
Sire	0.0435				

Fixed effect	Estimate	Std. Err.	D.F.	t-value	P-value
Rainbow trout	-0.6481	0.1856	23.0280	-3.492	0.002

Table S1.14: Best models describing burst swimming endurance. Full models included as fixed effects fork length, weight, water velocity, and admixture category (hybrid or Rainbow Trout - initial investigation found no difference in swimming performance among hybrids, thus they were grouped). Random effects varied between analysis and included both dam and sire (A), dam only (B), and sire only (C).

Figures

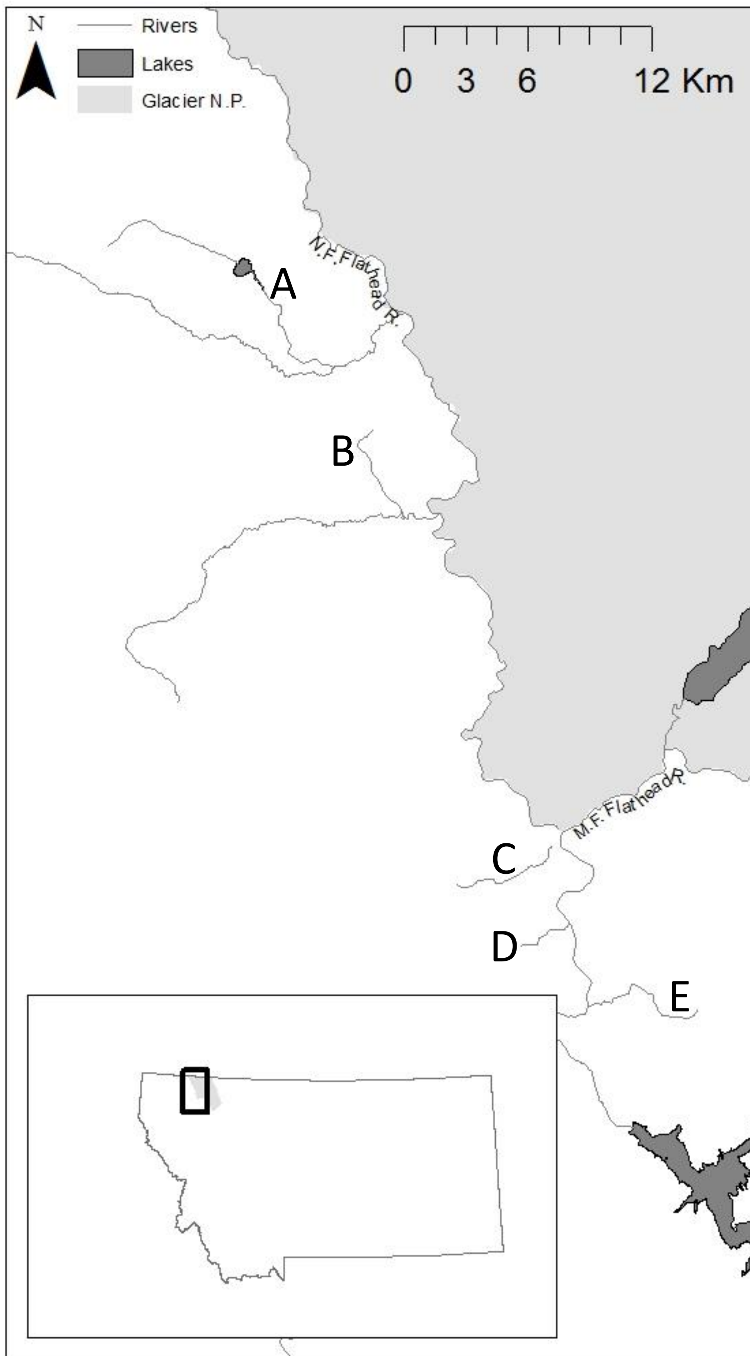


Figure 1.1: Study sampling locations in the upper Flathead River drainage. Fish were sampled in the N.F. Flathead River using boat eletrofishing and using weir traps in Cyclone (A), Langford (B), Rabe (C), Ivy (D), and Abbot (E) creeks.

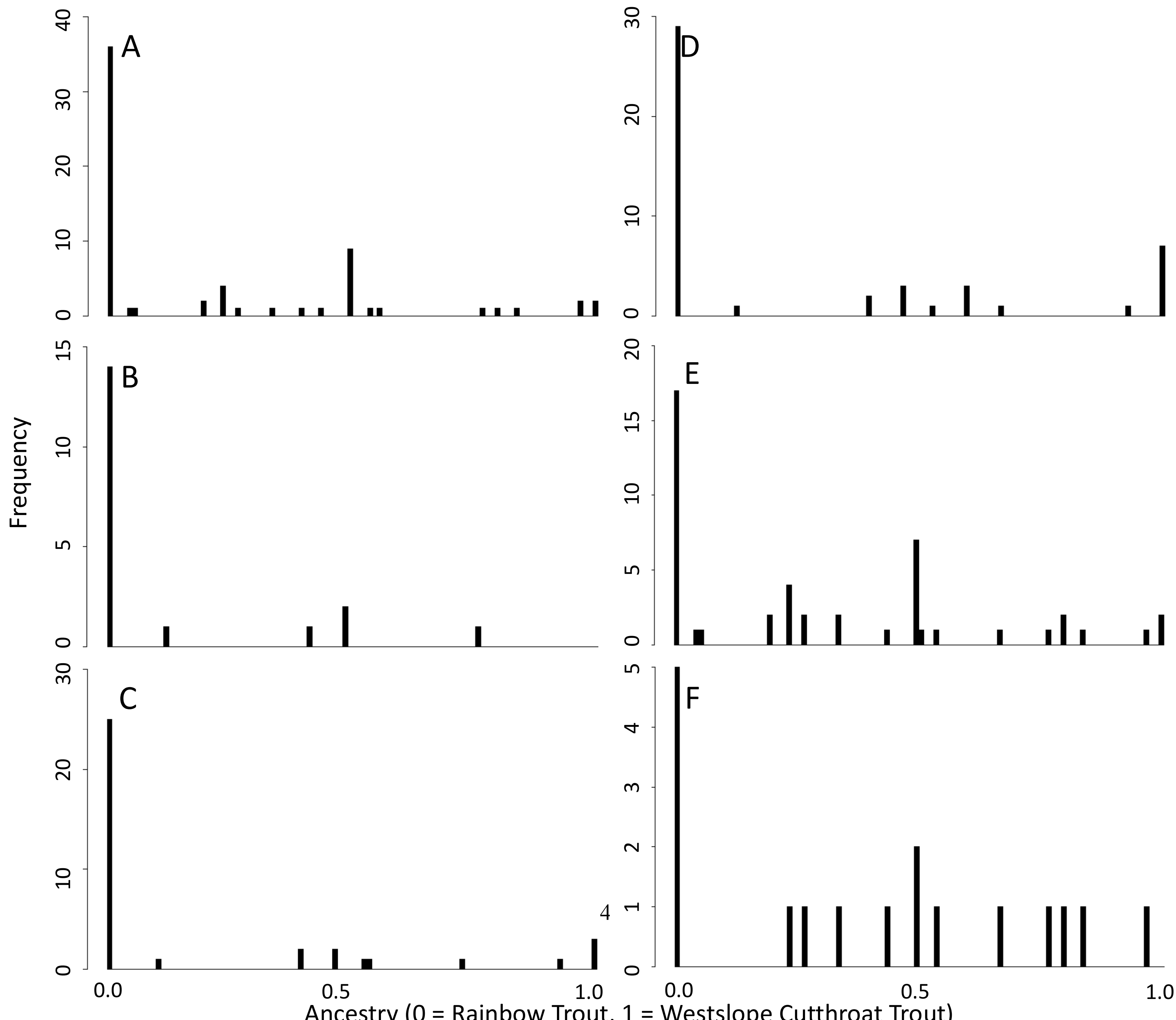
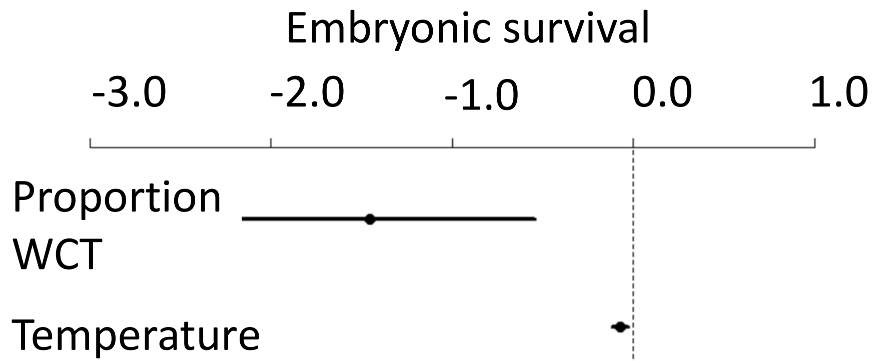
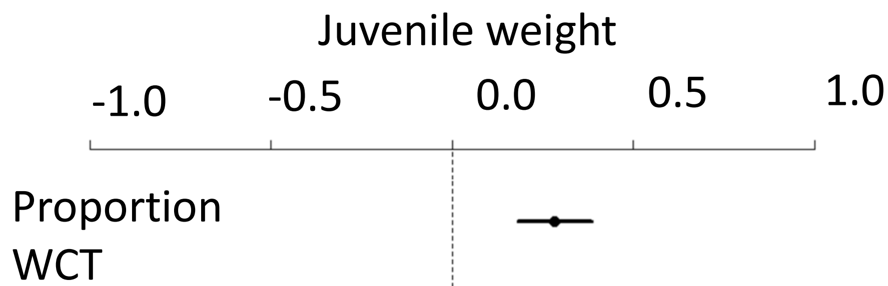
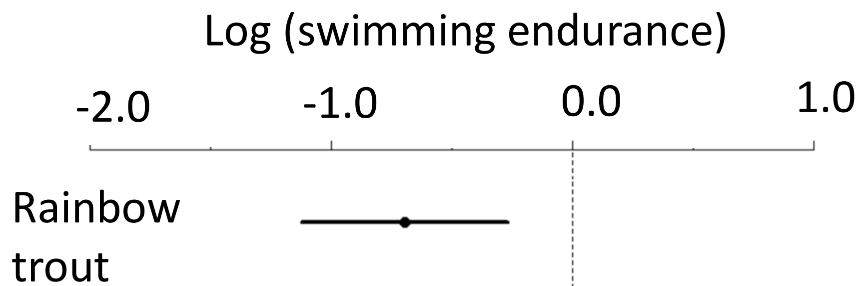

A
B
C
D
E
F
Frequency
0.0
0.5
1.0
Ancestry (0 = Rainbow Trout, 1 = Westslope Cutthroat Trout)
4

Figure 1.2: Ancestry of samples included in each analysis. Average familial ancestry was used in analyses of embryonic survival (A), juvenile weight and survival (E), and burst swimming endurance (F). Dam ancestry was used in analyses of ova size (B) and ova energy concentration (C). Sire ancestry was used in the analysis of sperm motility (D).


Embryonic survival
-3.0
-2.0
-1.0
0.0
1.0
Proportion WCT
Temperature


Juvenile weight
-1.0
-0.5
0.0
0.5
1.0
Proportion WCT


Log (swimming endurance)
-2.0
-1.0
0.0
1.0
Rainbow trout

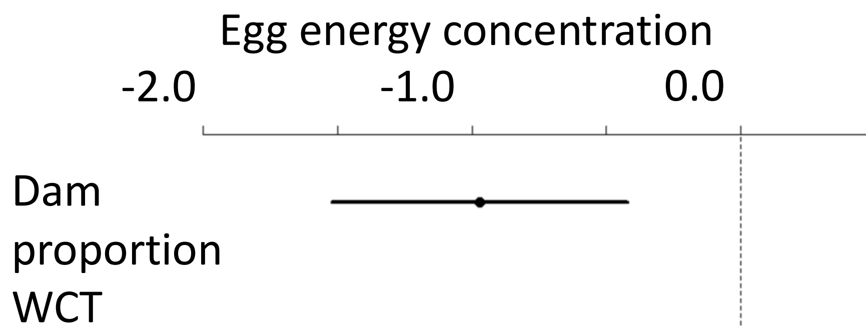

Egg energy concentration
-2.0
-1.0
0.0
Dam proportion WCT

Figure 1.3: Regression coefficients for analyses where hybridization was significantly correlated with the investigated trait. (A) Both familial proportion Westslope Cutthroat Trout and incubation temperature were negatively correlated with embryonic survival. (B) Juvenile weight at 60 days was positively correlated with familial proportion Westslope Cutthroat Trout. (C) Categorical classification of Rainbow Trout (versus non-Rainbow Trout) was negatively correlated with swimming endurance. (D) Dam proportion Westslope Cutthroat Trout was negatively correlated with ova energy concentration.

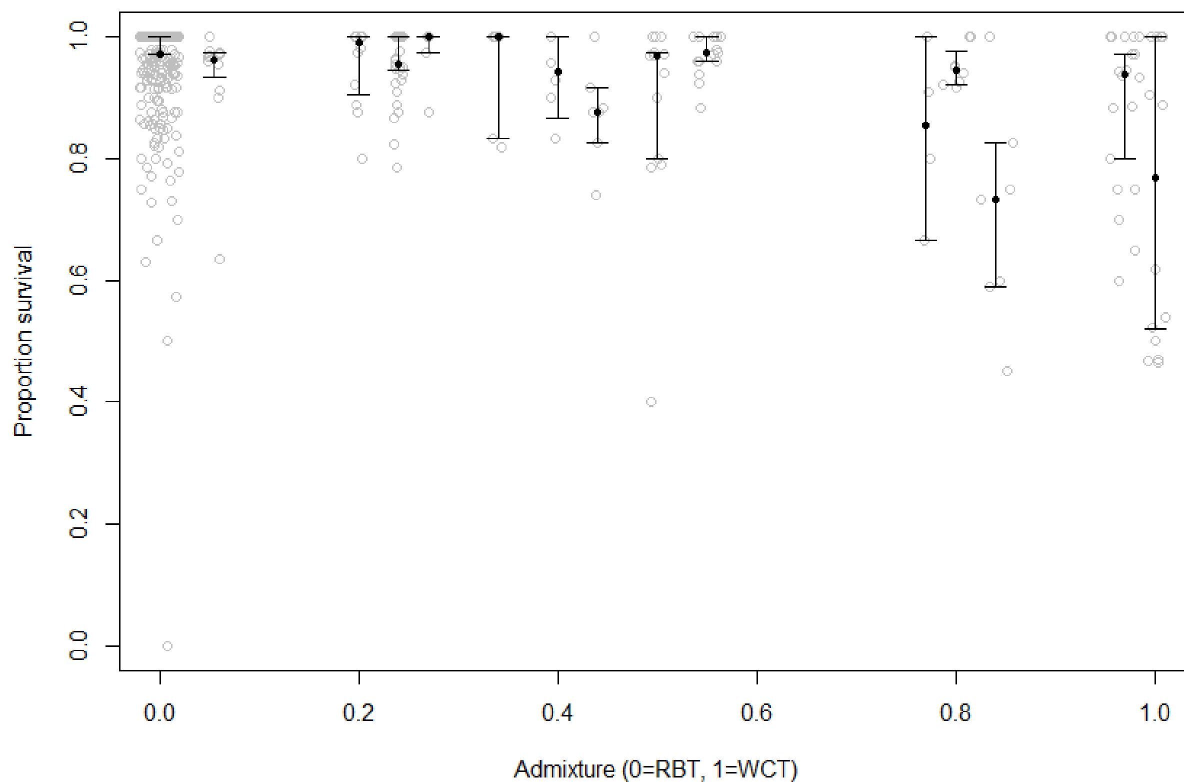


Figure 1.4: Embryonic survival based on admixture. Grey circles represent single observations (proportion survived per well). Grey dots are jittered along the x-axis. Black dots and lines represent the median value with 95% confidence interval of the median at each admixture level.

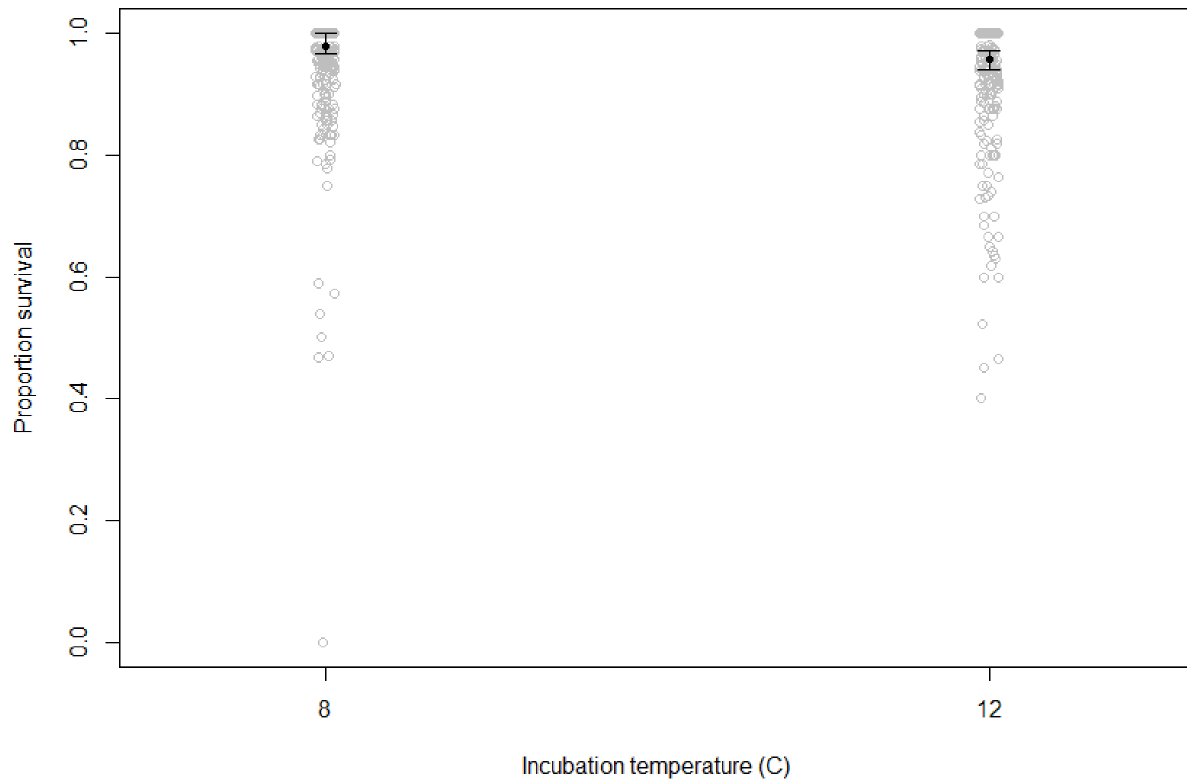


Figure 1.5: Embryonic survival at each incubation temperature. Grey circles represent single observations (proportion survived per well). Grey dots are jittered along the x-axis. Black dots and lines represent the median value with 95% confidence interval of the median at each incubation temperature.

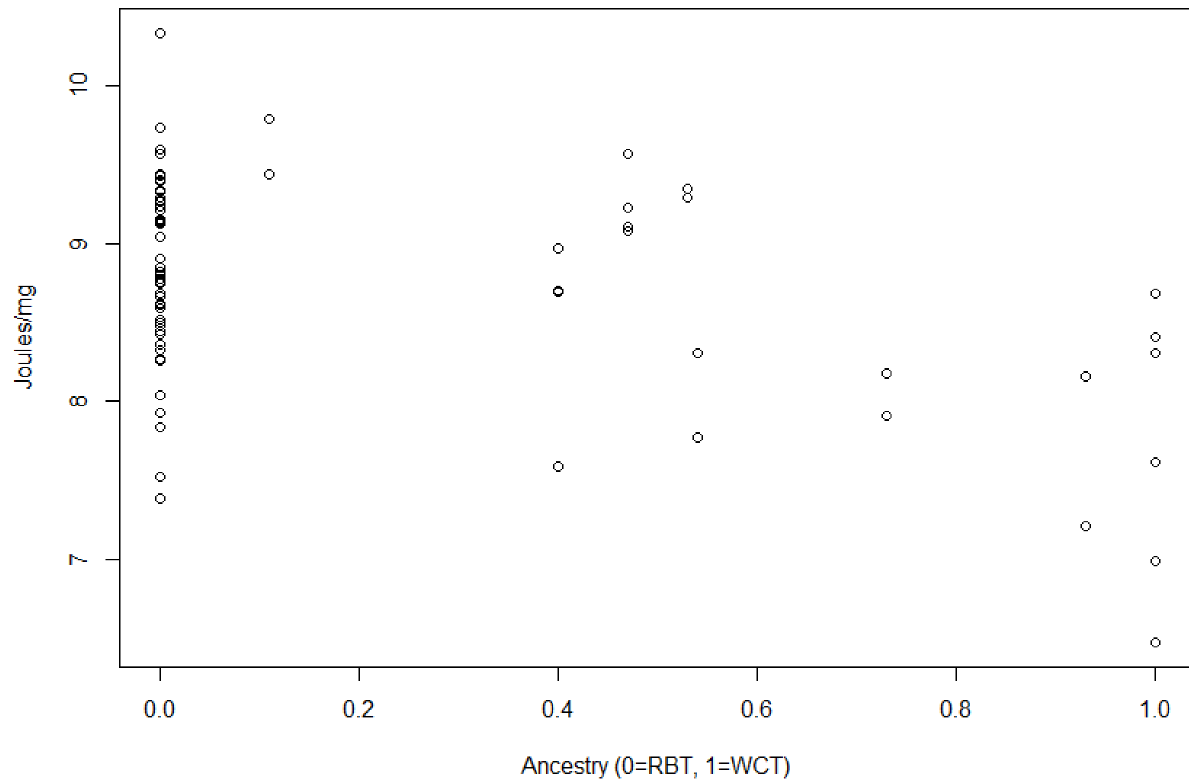


Figure 1.6: Egg energy concentration based on ancestry. Each dot represents measurement of a single female.

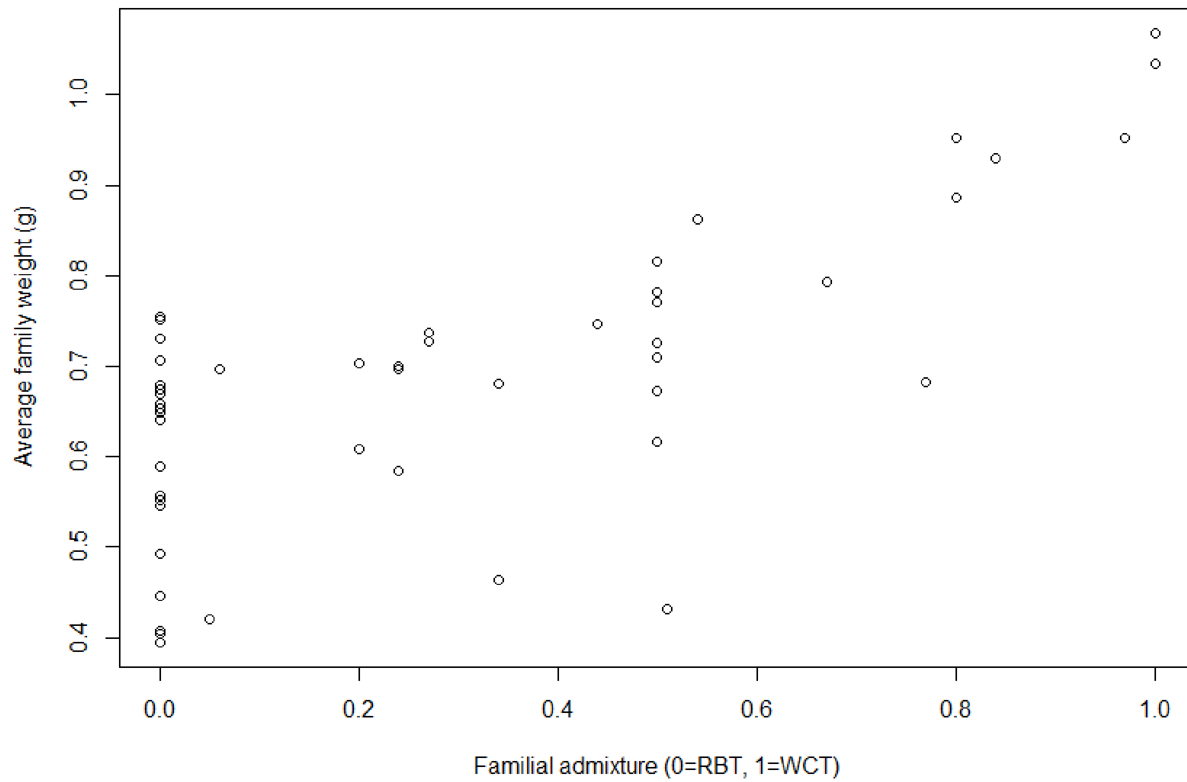


Figure 1.7: Average juvenile weight based on familial admixture. Each dot represents the average weight of progeny from one family. Average weight was estimated by weighing all progeny together and dividing by the total number of progeny. Familial admixture was estimated by averaging the admixture estimates of the spawned female and male. Admixture was estimated using eight species-specific markers.

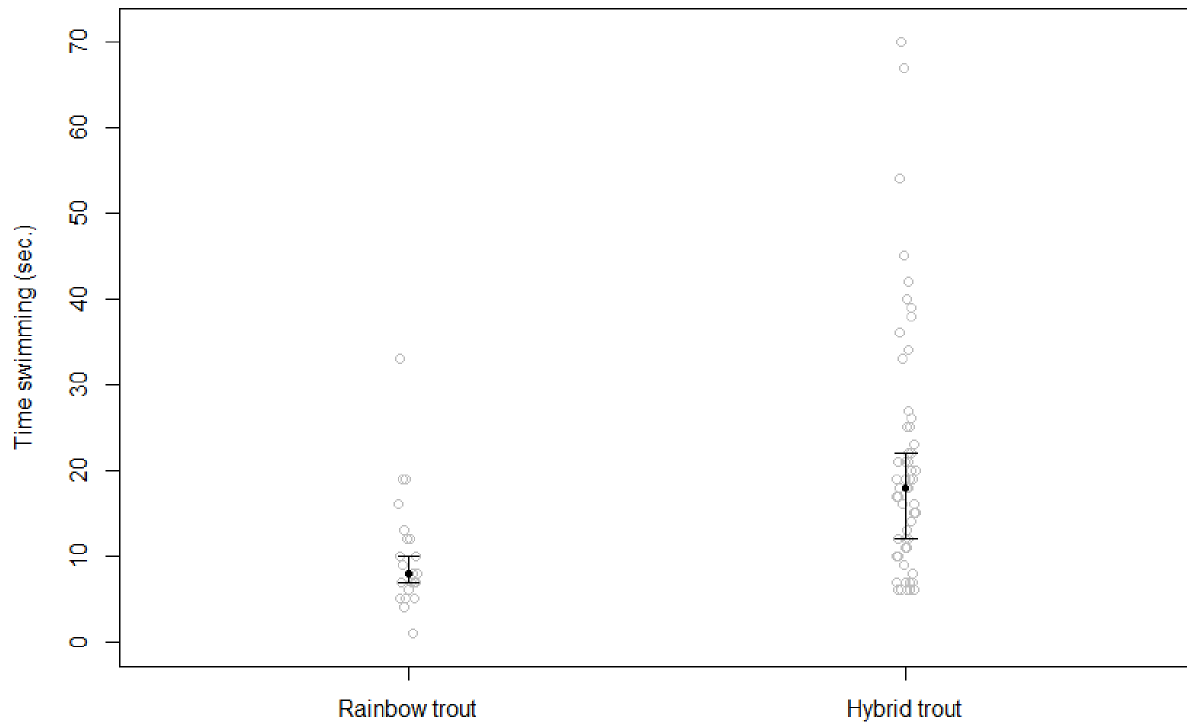


Figure 1.8: Swimming performance of one year old rainbow and hybrid trout. Grey dots represent length of time swam by one individual. Black dots and lines represent the medians and 95% confidence intervals of the medians. Water velocity was approximately seven body lengths per second during testing.

Supplemental Material

Power and sensitivity of admixture estimates

First we created 1000 genomes each for pure, F_1, F_2, and backcrossed individuals. From each genome, genotypes at seven unlinked, diploid markers were randomly sampled to compare known admixture levels to estimates calculated from few markers. Estimates were observed to be accurate for pure and F_1 individuals, but were less accurate for F_2 and backcross individuals (Figure S1.2).

Next, analyses were performed again using only individuals (ova diameter, egg energy concentration, and sperm motility) or families (embryonic survival, juvenile growth, juvenile survival, and high velocity swimming endurance) that were estimated to be genetically pure or F_1. For each analysis, we used the final models identified in the initial analyses. In general, the conclusions drawn for admixture were consistent with those using the full dataset. Below are the details.

Embryonic survival – Conclusions drawn using all data were the same as conclusions drawn using only pure and F_1 families.

Random effects:

Groups	Name	Variance	Std.Dev.
Sire	(Intercept)	1.0866	1.0424
Dam	(Intercept)	0.3321	0.5763

Number of obs: 262, groups: sire, 30; dam, 24

Fixed effects:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	4.16677	0.40275	10.346	<2e-16 ***
Avg. % WCT	-1.90032	0.91105	-2.086	0.0370 *
Temperature	-0.05914	0.03036	-1.948	0.0515 .

Juvenile growth – Conclusions drawn using all data were the same as conclusions drawn using only pure and F_1 families.

Random effects:

Groups	Name	Variance	Std.Dev.
Dam	(Intercept)	0.0096028	0.09799
Sire	(Intercept)	0.0013147	0.03626
Residual		0.0001598	0.01264

Number of obs: 29, groups: damn, 23; sire, 23

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	0.59535	0.02577	22.16500	23.101	< 2e-16 ***
Proportion WCT	0.38047	0.06761	21.04800	5.627	1.37e-05 ***

Juvenile survival – Two models were selected as "best" in the original analysis. Both models are analyzed here. Conclusions drawn using all data were the same as conclusions drawn using only pure and F_1 families.

Random effects:

Groups	Name	Variance	Std.Dev.
Dam	(Intercept)	0.2046	0.4524
Sire	(Intercept)	0.2634	0.5132

Number of obs: 29, groups: dam, 23; sire, 23

Fixed effects:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.227516	0.377588	3.251	0.00115 **
Avg. % WCT	0.128288	0.480181	0.267	0.78934
Tank density	0.006435	0.004184	1.538	0.12406

Random effects:

Groups	Name	Variance	Std.Dev.
Dam	(Intercept)	0.2074	0.4554
Sire	(Intercept)	0.2621	0.5120

Number of obs: 29, groups: dam, 23; sire, 23

Fixed effects:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.265557	0.351463	3.601	0.000317 ***
Tank density	0.006234	0.004128	1.510	0.130968

<u>High velocity swimming endurance</u>

In the initial analysis, rainbow trout were compared to hybrid trout. No differentiation was made among hybrid classes. Therefore, no sensitivity analysis was needed.

<u>Ova diameter</u>

The analysis was dominated by pure rainbow trout and the removal of hybrid individuals that were not F_1s resulted in only three remaining data points that were not rainbow trout. Because the sample size was so small, the analysis was not performed.

<u>Egg energy concentration – conclusions do not change</u>

Random effects:

Groups	Name	Variance	Std.Dev.
Dam	(Intercept)	0.2314	0.4811
Residual		0.1557	0.3946

Number of obs: 53, groups: Dam, 28

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	8.8579	0.1126	26.5570	78.671	< 2e-16 ***
Proportion WCT	-1.1099	0.3403	25.6440	-3.262	0.00312 **

Sperm motility

The final model did not include admixture as a parameter.

Chapter 2 - Changes to recombination and limited segregation distortion revealed in the hybrid trout genome

Abstract

Interspecific hybridization is an important evolutionary process, and results in the combination of genomes that are often structurally and adaptively differentiated. Changes in the composition and function of the hybrid genome as a result of genomic differences in the parental species may influence fitness, and the identification of altered genomic regions is important for understanding adaptation and speciation. Here we used genetic maps to identify how hybridization between Yellowstone cutthroat trout (*Oncorhynchus clarkii bouvieri*) and *O. mykiss* alters recombination and species-specific ancestry compared with pure *O. mykiss*. Recombination changes were identified on seven chromosomes in the hybrid male and eight in the hybrid female. Identified groups were consistent between the sexes. Recombination was suppressed on chromosomes with known architectural polymorphisms between the species. Such regions may retain co-adapted genes and could be important to fitness. In addition, differences in recombination between hybrid and *O. mykiss* were observed on five chromosomes with a high proportion of duplicated markers. Changes in recombination on chromosome arms with a high proportion of duplicated markers may be the result of increased homeologous pairing in hybrid individuals. Regions of excess species-specific ancestry were identified in small proportions throughout the genome in both the male and female. Excess species-specific ancestry appeared to be largely random and may be driven by genetic drift, but we could not exclude the influence of natural selection. Patterns of recombination and species-specific ancestry suggest that few

genetic incompatibilities exist between cutthroat and *O. mykiss*, but observed differences may be important to adaption and speciation.

Introduction

Hybridization is common among both plant and animal taxa (Mallet, 2005) and can serve as an important mechanism for diversification, adaptation, and speciation (Baack and Rieseberg, 2007; Barton and Hewitt, 1989; Burke and Arnold, 2001). During interspecific hybridization, divergent genomes are combined. Changes in the hybrid genome composition and function are common (Baack and Rieseberg, 2007; Li*, et al.*, 2010; Ostberg*, et al.*, 2013) and may be indicative of structural or adaptive divergence in the parental species (Borowsky and Cohen, 2013; Dumas, Catalan and Britton-davidian, 2015). Altered regions may contain genes or gene complexes important to fitness in wild hybrid populations (Chapman and Abbott, 2010), and could be lost or retained as a result of events shortly following hybridization. The identification of both how the hybrid genome is organized, given differences in parental chromosomal structure and divergence, as well as where in the genome these differences occur, is critical for understanding the future evolution of wild hybrid populations. Identifying changes in genomic patterns is particularly important during early generations, as the initial reproductive outcomes will influence what genetic material is available to natural selection in later generations.

Differences in chromosomal structure are common among hybridizing species and may alter recombination patterns in early generation hybrids (Alves*, et al.*, 2014; Dumas*, et al.*, 2015; McGaugh and Noor, 2012; Ostberg*, et al.*, 2013). Interspecific hybridization often results in the

formation of individuals that are heterozygous for polymorphisms in karyotype number or marker orders. As a result, recombination is often suppressed in regions where structural differences occur and genomic areas with large intact species-specific haplotypes are often conserved (Davisson and Akeson, 1993; Kulathinal, Stevison and Noor, 2009; Stevison, Hoehn and Noor, 2011). Recombination suppression results in genomic regions that are more resistant to introgression and may serve to protect species-specific adaptations or gene complexes present in the region. Changes in recombination patterns are most visible in early generations before drift or selection fixes one of the alleles and recombination is restored.

Interspecific hybridization may also alter phenotypes of individuals and the identification of genomic regions responsible for phenotypic changes is critical for understanding divergent adaptation and speciation. When species interbreed, novel combinations of alleles are created. These combinations can have profound effects on phenotypes, including fitness, and may result in genomic regions with excess species-specific ancestry in cases where species alleles are preferentially selected (Borowsky and Cohen, 2013; Martin, Bouck and Arnold, 2006; Ostberg, *et al.*, 2013). Such regions have been widely observed in early generation hybrids (Borowsky and Cohen, 2013; Martin, *et al.*, 2006; Ostberg, *et al.*, 2013), and the identification of genomic regions with excess species-specific ancestry is important for understanding what genomic mechanisms are responsible for adaptation.

The investigation of hybridization is particularly interesting in species derived from a common autotetraploid ancestor, as characterization of the hybrid genome would provide insight into genome evolution subsequent to the duplication event. Further, if the duplicated genomes are not fully restored to the diploid state, ongoing recombination between duplicated

chromosomes or chromosome arms may influence patterns of introgression in the hybrid. Autopolyploids are the result of a doubling of a single genome, resulting in an organism with polysomic inheritance patterns (Wendel, 2000). Autopolyploids may suffer from genome instability due to meiotic and mitotic irregularities, and as a result, frequently revert to a diploid, or partial diploid, state through genomic rearrangements and divergence (Hufton and Panopoulou, 2009; Ohno, 1999; Otto, 2007; Storchova and Pellman, 2004; Wendel, 2000; Wright, *et al.*, 1983). These evolutionary processes are expected to reduce genomic similarity between species diverged from a common autopolyploid ancestor. By investigating the hybrid genome, differences in genome evolution between sister taxa derived from a common autopolyploid ancestor may be observed.

Mesopolyploids are descendants from a whole genome duplication event (Mayfield-Jones, *et al.*, 2013), where diploidization, the process of returning to a diploid state, is ongoing. Therefore, recombination may still occur between duplicated (homeologous) chromosome arms, and some residual tetrasomic inheritance may persist (Allendorf and Danzmann, 1997; Quiros, 1982). Tetrasomic inheritance occurs in parental species when homeologous chromosomes form multivalents and exchange genetic material during meiosis (Allendorf and Thorgaard, 1984; Parisod, Holderegger and Brochmann, 2010; Wolf, Soltis and Soltis, 1989; Wright, *et al.*, 1983). The formation of multivalents may be increased in hybrids between species descended from the same ancestral duplication event (Wright, *et al.*, 1983). Homeologous chromosomes of the same species ancestry are expected to be more similar than homologs from different species. Pairings between homeologs will therefore result in the exclusive exchange of genetic material between chromosomes from the same parental species. In contrast, the pairing of homologous

chromosomes in mesopolyploid hybrids would result in recombination between the chromosomes from both parental species. Therefore, an increase in tetrasomic inheritance in the hybrid genome may result in genomic regions that are less susceptible to introgression.

Salmonids (salmon, trout, char, and whitefish) are derived from a whole genome duplication event that occurred in an autotetraploid ancestor between 88-103 mya (Macqueen and Johnston, 2014). Since duplication, salmonid genomes have been reverting to a diploid state, characterized largely by Robertsonian rearrangements and differentiation of homeologous chromosomes (Phillips and Rab, 2001; Wright*, et al.*, 1983). Robertsonian rearrangements, or the centric fusions and fissions of entire chromosome arms, have resulted in a range of karyotypes among salmonid species (Allendorf and Thorgaard, 1984; Phillips and Rab, 2001). However, the number of chromosome arms (n = approximately 100), and the synteny between arms, has remained largely conserved (Allendorf and Thorgaard, 1984; Phillips and Rab, 2001). There is significant evidence that diploidization is ongoing (Allendorf and Danzmann, 1997; Allendorf and Thorgaard, 1984), with the persistence of residual tetrasomic inheritance, primarily observed in males (Wright*, et al.*, 1983). In salmonids, only a subset of chromosome arms is hypothesized to be involved in ongoing or recent homeologous pairing (Brieuc*, et al.*, 2014; Kodama*, et al.*, 2014). In species of the Pacific salmon genus *Oncorhynchus*, evidence of recent or ongoing tetrasomic inheritance on only a subset of arms (eight pairs) is likely a result of chromosomal rearrangements prior to the divergence of taxa (Kodama*, et al.*, 2014; Naish*, et al.*, 2013), but the role of duplicated regions in adaptation and innovation is not yet understood.

Interspecific hybridization between *Oncorhynchus mykiss* and cutthroat trout (*O. clarkii*) provides an ideal system for studying the effects of hybridization between species descended

from an autotetraploid event. *O. mykiss* and cutthroat trout are sister taxa native to western North America (Behnke, 1992; Crete-Lafreniere, Weir and Bernatchez, 2012; Loxterman and Keeley, 2012). The species share a common ancestry 7.8 – 12.0 million years ago (Crete-Lafreniere*, et al.*, 2012). Throughout the range of cutthroat trout, *O. mykiss* has been widely introduced (Halverson, 2010). When in sympatry, the two species readily hybridize and produce viable offspring. Hybridization has been shown to reduce fitness (Muhlfeld*, et al.*, 2009) and is of major conservation concern throughout the range of cutthroat trout (Allendorf*, et al.*, 2001; Gresswell, 2011; Shepard, May and Urie, 2005; Young, 2008; Young, 2009). The two species *O. mykiss* and cutthroat trout differ in karyotype, but the number of chromosome arms has been conserved. The quantity of chromosomes in both species is polymorphic; *O. mykiss* have 58, 60, or 64 chromosomes (Phillips*, et al.*, 2005; Thorgaard, 1983), while cutthroat trout have 64, 66, or 68 chromosomes (Gold, Avise and Gall, 1977; Loudenslager and Thorgaard, 1979). In spite of these karyotypic differences, chromosome arm number and synteny within arms remains largely conserved between cutthroat and rainbow trout (Ostberg*, et al.*, 2013). Thus, divergence is great enough that evolution has resulted in different chromosome rearrangements and divergence between the species, yet similar enough to allow *O. mykiss* and cutthroat trout to readily hybridize and produce viable offspring.

Comparative genetic mapping provides a powerful basis for studying genome evolution (Brieuc*, et al.*, 2014; Fishman*, et al.*, 2001; Kodama*, et al.*, 2014), and identifying the genomic consequences of hybridization (Ostberg*, et al.*, 2013; Stevison*, et al.*, 2011). By comparing map distances between F_1 hybrid and *O. mykiss* maps, Ostberg *et al.* (2013) identified broad regions of recombination suppression in the F_1 on five linkage groups, including two with known

chromosomal rearrangement differences between cutthroat trout and *O. mykiss*. In addition, few regions of excess species-specific ancestry were observed consistently in the F_2 offspring in both mapping families investigated (Ostberg*, et al.*, 2013). These results suggest that there was little bias in the inheritance of parental chromosomal segments, and selection did not preferentially act on the gametes of the F_1 hybrid parents. However, the study of Ostberg *et al*. (2013) relied on relatively few loci (327 and 312 loci mapped in the merged female and male maps). Therefore, fine-scale patterns in the fate of the early hybrid genome may not have been observed. Recent advances in sequencing technology now make the genotyping of thousands of loci in non-model species, such as cutthroat trout, feasible (Baird*, et al.*, 2008).

The aim of this study was to examine the genomic consequences of hybridization following one generation of recombination between *O. mykiss* and cutthroat trout, and to determine how early generations of hybridization might influence the persistence of species-specific alleles. This aim was achieved identifying genomic regions of recombination suppression, and by performing a whole genome analysis of species-specific ancestry bias in F_2 hybrids to determine whether species-specific alleles are preferentially retained in early generations. Identified regions may be adaptively important and could provide insight into which genomic regions are important to fitness. To achieve these objectives, two genetic linkage maps were created, one from a cross between F_1 hybrids derived from a Yellowstone cutthroat (*O. c. bouvieri*) and *O. mykiss*, and the other from *O. mykiss* haploids individuals, to provide a reference framework for comparison of the hybrid map.

Methods

Description of experimental crosses

We used a comparative analysis between two cross types to characterize the effects of introgression on the hybrid genome. The first was a haploid *O. mykiss* family that served as a reference for the hybrid map, permitting estimation of the effects of recombination on the hybrid genome. The second was a cross between F_1 hybrids, where the F_1 parents were created from an F_0 cross between Yellowstone cutthroat trout and *O. mykiss*.

First, tissue samples from 53 progeny of the F_1 cross were collected. This same family was mapped in a previous study (Ostberg *et al.* 2013). Second, a haploid *O. mykiss* family was created at the U.S. Fish and Wildlife Quinault National Fish Hatchery (QNFH), Washington, USA. QNFH maintains a production line of steelhead for release into the Hoh River, Washington, USA. Haploids were created by fertilizing ova with irradiated milt following the protocols of (Thorgaard *et al.* 1983) . Embryos were euthanized prior to hatching.

DNA extraction, sequencing and quality screening

Genomic DNA was extracted from all tissue samples using DNeasy Blood and Tissue Kits following manufacturer's recommendations (QIAGEN, Valencia, CA, USA). Ploidy of *O. mykiss* embryos was confirmed using eight multiallelic microsatellite markers. Hybrid and *O. mykiss* libraries for restriction-site associated DNA libraries were created using standard laboratory procedures (Baird *et al.* 2008). DNA was digested using the *Sbf-I* restriction enzyme and 6-nucleotide barcodes were annealed to each sample for identification purposes. Samples were pooled into libraries of 36 (diploid samples) or 48 (haploid samples) individuals. Each

library was sequenced on the Illumina HiSeq platform. Single read sequencing was employed with each read having a length of 100 basepairs. Poor quality reads were removed from the analysis using a sliding window approach implemented by STACKS v1.00 (Catchen *et al.* 2011; Catchen *et al.* 2013). Using a window size of 15 bp, reads were removed if the average PHRED quality score within the window dropped below 90% probability of being correct (raw Phred score of 10), or if the sequence contained a missing base. The remaining sequences were separated into individuals based on barcodes.

Reference database of O. mykiss RAD loci

A reference database of all Sbf-I RAD loci for *O. mykiss*, both polymorphic and non-polymorphic, was created. The reference database of *O. mykiss* loci was created using 128 steelhead from QNFH and 18 hatchery fish from Ennis National Fish Hatchery, Ennis, Montana, USA. For each individual, loci were identified *de novo* using STACKS (Catchen *et al.* 2011; Catchen *et al.* 2013) and placed in a temporary database. Up to three sequence mismatches were allowed and a read depth of 5 was required to call a genotype. Loci aligning to repeat regions through the genome and loci with repeat elements within them were filtered from the temporary database using the same alignment based methodologies of (Brieuc *et al.* 2014). First, each locus was aligned to the temporary reference database using BOWTIE 0.12.9 (Langmead *et al.* 2009). A maximum of three mismatches was allowed. A locus that aligned to multiple loci was removed from the temporary database. Second, each locus was also aligned to the temporary reference database using BLAST (Basic Local Alignment Search Tool; Altschul *et al.* 1990). The BLAST algorithm uses a filter to remove segments of low-complexity, such as repeat motifs, when

estimating sequence similarity. As such, sequences that contain repeat motifs are expected to have no single best alignment or alternatively, to align to multiple loci, because the flanking regions that remain after filtering the repeat section are short. Loci that passed both alignments were retained in the final reference database of loci. The final reference database was aligned to the draft *O. mykiss* genome (Berthelot *et al.* 2014) as well as RAD databases produced for *O. mykiss* (Miller *et al.* 2012), coho salmon (Kodama *et al.* 2014), and Chinook salmon (Brieuc *et al.* 2014) using BOWTIE to identify similarities between databases.

O. mykiss haploid cross genotyping

The genotyping of haploid *O. mykiss* was performed by aligning haploid *O. mykiss* sequence reads to the final database of non-repetitive loci using BOWTIE 0.12.9 (Langmead *et al.* 2009). Up to three mismatches between sequences were allowed. Genotypes were called using a maximum likelihood approach implemented by STACKS.

In order to call a genotype at a locus, a read depth of at least five was required. Simulations showed that estimating genotypes with a read depth of only five was sufficient to accurately identify between duplicated and non-duplicated loci when using the maximum likelihood approach implemented by this study (Supplemental S1). The median read depth at a locus for samples included in this analysis was eight.

Because sequencing error is inherent in all genotype-by-sequencing projects and may affect the alleles observed, identification of alleles at each locus cannot be achieved simply by counting unique sequence reads. Instead, we calculated a log-likelihood value for each possible locus type (haploid: AB x 00, duplicated: AA x BB, AA x AB, AB x AB, AA x BC, AB x AC,

and AB x CD) at each marker (Supplemental S2). Log-likelihoods for the two most probable locus types were compared, and the most probable locus genotype was accepted if Δlog-likelihood between the two locus types was 4 or greater. Markers identified as duplicated were separated into their two respective loci by discerning the genotypes at the two duplicated markers (AA x BB, AA x AB, AB x AB, AA x BC, AB x AC, AB x CD) using a maximum likelihood approach (Supplemental S3).

Hybrid progeny genotyping

The identification of duplicated loci can be challenging in a diploid cross such as the F_1 hybrid cross mapped here, and such loci might confound linkage analyses. Therefore, sequence reads for all hybrid progeny were aligned to a reference baseline set of non-duplicated loci identified in Chinook salmon (*O. tshawytscha*) (Brieuc *et al.* 2014). This data set was the most reliable, since the duplicated loci were identified using three haploid families, and centromere locations were identified using three gynogenetic crosses. The set of non-duplicated loci in the Chinook salmon database were 74 bp in length: therefore, each sequence read from the F_1 hybrids was trimmed to the same length. Alignments were performed using the same criteria described earlier for the haploid *O. mykiss*. After alignment, F_2 progeny were genotyped at a locus if 10 or more reads aligned to a single locus in the reference baseline of Chinook salmon loci.

Given that the F_1 parents were derived from outbred F_0, the genotypes and hence range of parental cross types were expected to differ across loci. Given some sequencing error, it was necessary to determine the most likely F_1 genotypes, and hence cross type for each locus (AA x

AB, AB x AB, AA x BB). Therefore, we used the maximum likelihood approach previously described to determine genotypes (Supplemental S2). This information was compared with the parental genotypes derived directly from the sequences. Microsatellite markers from (Ostberg *et al.* 2013) were added to the data set to allow identification of previously named linkage groups (designated RYHyb 1-29).

Linkage mapping

Loci were grouped, ordered, and filtered using the mapping package ONEMAP (Margarido *et al.* 2007) in the statistical software R 3.1.0 (R Core Team 2013). Grouping was performed using a maximum recombination fraction of 0.5 and an LOD score starting at 3.0 and progressively increasing, until all linkage groups identified by (Ostberg *et al.* 2013) could be discriminated for the F_1 hybrid map. Further verification of linkage groups was obtained by aligning the hybrid and haploid maps against homologous linkage groups identified in rainbow trout (Miller *et al.* 2012), Chinook salmon (Brieuc *et al.* 2014) , and coho salmon (Kodama *et al.* 2014). Ordering of markers within linkage groups was performed using rapid chain delineation (Doerge 1996) in ONEMAP. Locus orders were verified using heat maps of recombination fractions within each linkage group.

The creation of sex-specific maps from an F_1 diploid cross can be difficult due to the large proportion of species-specific AB x AB markers. For each AB x AB locus, half of progeny are expected to inherit AB genotypes that are uninformative as to which parent contributed each allele. We therefore used a correction technique to infer ancestry for cross types AB x AB. Sex-specific maps were created for the F_1 hybrid cross using ordered markers, phased parental

chromosomes, and ancestry informative markers (Figure 2.1). F_1 parental chromosomes were phased based on species-specific markers, since chromosomes in these individuals would be inherited intact from the parental species (Figure 2.1A). Haplotype ancestry for ancestry informative markers was identified within each F_2 progeny using the parental and progeny genotypes (Figure 2.1B). Allelic ancestry for markers located between informative loci was then inferred within a progeny chromosome (Figure 2.1C). Double crossovers on a chromosome arm are rare in salmonids, particularly over small regions (Thorgaard *et al.* 1983; Guyomard *et al.* 2006; Brieuc *et al.* 2014). Therefore, if two markers originated from the same parental chromosome and no evidence of recombination was observed between them, all intermediate markers were inferred to have the same chromosomal ancestry (Figure 2.1C). If haplotype ancestry from one parent was known, the contribution from the other parent was inferred (Figure 2.1D). Lastly, distances between markers on the consensus map were estimated by combining sex specific genotypes into multi-allelic, diploid genotypes and estimating marker distance based on the previously identified marker order.

Segregation distortion and ancestry bias

A sliding window approach was used to identify whether recombination in the F_1 parents and subsequent survival of offspring (sampled 10 months post hatch) led to excess species-specific ancestry in the inheritance of chromosomal segments across the genome in the progeny used to create the sex specific maps. A sliding window frame of 180 cM on the consensus map was used to identify chromosomal distribution of species-specific alleles across all progeny for each parent. A consensus map was used to so that direct comparisons between the sexes could be

made. The use of other window sizes was investigated (data not shown), but 180 cM was found to be optimal to overcome bias due to sampling variance while still providing enough detail to detect regional patterns. A window frame of 180 cM is large compared to other analyses (Hohenlohe *et al.* 2010; Perrier *et al.* 2013). However, the size of the consensus linkage map for the hybrid parents was larger than other salmonid RAD maps (Miller *et al.* 2012; Brieuc *et al.* 2014; Kodama *et al.* 2014). Thus the window size was proportional to the map size and was comparable to previous studies in salmon (Perrier *et al.* 2013). For each 180cm window frame, loci were bootstrapped 100,000 times. A chromosomal segment was considered to have biased species-specific ancestry if segregation ratios deviated in the same direction over at least 10 cM. If less than 80% of F_2 individuals were genotyped at a locus, the locus was not included in the analysis.

Differences in relative recombination distances between O.mykiss and F_1 hybrids

We compared the relative map distances of the sex-specific hybrid maps to the haploid *O. mykiss* map in order to identify how recombination across the hybrid genome may be affected by introgression, relative to the parental species. We aligned loci from the *O. mykiss* haploid map to the female hybrid map. Each chromosome marker position was plotted, and both a linear and logistic model was fit to the data. Akaike information criterion (AIC) was used to determine which model was a better fit (ΔAIC >= 4). Chromosomes where the logistic model fit better suggested a relative change in recombination across the chromosome following hybridization.

Results

Reference database of O. mykiss RAD loci

A reference database of RAD loci present in *O. mykiss* was created as a baseline for sequence alignments. A total of 58,468 loci were identified in at least 64 individuals with a minimum depth of 5 sequence reads per locus. Of these loci, 10,457 (~18%) were identified as duplicated based on the maximum likelihood analysis.

Agreement between the reference database and other salmonid databases was high. Most loci (50,413, 86%) from the present *O. mykiss* database were uniquely aligned with the recently published *O. mykiss* reference genome (Berthelot *et al.* 2014). Overlap between the RAD database generated here and other RAD databases produced with salmonids was also high, 59% for *O. mykiss* (34,487 loci; Miller *et al.* 2012), 51% for coho (29,721 loci; Kodama *et al.* 2014), and 49% for Chinook salmon (28,607 loci; Brieuc *et al.* 2014).

O. mykiss haploid map

A total of 4728 non-duplicated and 465 duplicated markers in the *O. mykiss* database were mapped to 29 linkage groups (Figure S2.1) using an LOD score of 5 to 7. Of the mapped, duplicated markers, 102 loci were polymorphic across both duplicated loci, while 363 were polymorphic in only one of the duplicated loci. The total map distance was 6798 cM.

Duplicated markers were not evenly distributed across linkage groups (X^2 test, p-value = 0.005). On most chromosome arms, few markers were identified as duplicated. However, 10% or more of the markers on Omy01q, Omy02p, Omy10c, Omy12q, Omy13p, Omy13q, Omy17q, and Omy19p, were duplicated markers (Figure S2.2). The duplicated markers were also more

common at distal regions (Figure 2.2). Arms with a high proportion of duplicated markers in this study were homologs to those previously observed in two other *Oncorhynchus* species, and may be involved in recent or ongoing homeologous pairing (Brieuc *et al.* 2014; Kodama *et al.* 2014).

Hybrid maps

The order of loci in the hybrid map was identified by creating a temporary consensus map using progeny genotypes. In total, 3844 loci were ordered on all 29 linkage groups that had been previously identified (Ostberg *et al.* 2013). Sex specific ancestry was inferred using the ordered loci (Figure 2.1), and sex specific map distances were estimated after inferred ancestries (Figure 2.1). In total, 3834 and 3832 markers were mapped for the female and the male F_1 parents respectively. The female map contained 28 linkage groups and had a total distance of 14,199 cM, while the male map contained 29 linkage groups and had a distance of 13,549 cM. Both the female and male linkage maps included microsatellite markers on all chromosomes that had been previously identified by (Ostberg *et al.* 2013) and were used to identify linkage groups. Previous mapping of this family (Ostberg *et al.* 2013) observed that female contained a metacentric fusion between RYHyb25 and RYHyb29 (hereafter the fused metacentric chromosome is referred to as RYHyb25_29), while the male contained two acrocentric chromosomes. These differences are reflected in sex-specific hybrid maps in this study.

The female map was larger than the male map (1.05 times greater; Figures S2.3-5). All but five groups (RYHyb05, RYHyb08, RYHyb14, RYHyb16, and RYHyb19) had a distance that was greater in the female map compared to the male map. Differences in map distances between the sexes have been well documented in salmonids (Sakamoto *et al.* 2000; Lien *et al.* 2011) and

may be related to reduced recombination in males or differences in the location of recombination between the sexes.

Segregation distortion and ancestry bias

Identification of genomic regions of excess species-specific ancestry contributed by either the dam or the sire was identified by comparing expected ancestries estimated by bootstrapping across all loci, with observed ancestry estimates, where values were smoothed across the chromosome using a sliding window approach. The consensus hybrid map was used for both male and female analyses so that results could be compared. Chromosomal regions with species-specific alleles spanning 10 or more continuous cM outside the expected bootstrapped range were considered evidence of bias.

Excess species-specific ancestry was identified on numerous chromosomes, but limited in the proportion of the genome that contained excess species-specific alleles for both the female and male. In total, 59 and 47 chromosomal segments with species bias over greater than 10 cM were observed in the female and male respectively (Table 2.1; Figures S2.6-33). These regions fell on 33 chromosome arms for the female, and 23 arms for the male (Table 2.1; Figure S2.6–33).

Patterns of excess species-specific ancestry were similar between the sexes. Sixteen chromosome arms contained segments that were biased towards the same species in both sexes. Only three chromosome arms (RYHyb02q, RYHyb06q, and RYHyb09q) showed species-specific biases that differed between the sexes. The lengths of species-biased segments were also similar between the sexes, with the median length being 17.28 cM (25th and 75th quantiles =

12.33, 34.16 cM) and 15.53 cM (25th and 75th quantiles = 12.22, 24.86 cM) in the female and male respectively. In total, 11% of the genomic distance contributed from the female was biased compared to 10% in the male. Approximately half of the segments were biased toward cutthroat trout ancestry and half toward *O. mykiss* ancestry in both the female and male.

The distribution of segments with excess species-specific ancestry was not correlated with chromosome type (metacentric vs acrocentric). Due to the karyotype of cutthroat and *O. mykiss*, approximately 90% of the genomic distance investigated in this analysis was on metacentric chromosomes, while only 10% was on acrocentric chromosomes. Species biased segments were similarly distributed with 88 and 81% of excess segments observed on metacentric chromosomes in the female and male, while only 12 and 19% were on acrocentric chromosomes.

The distribution of biased segments was not related to whether a chromosome arm had an excess of duplicated markers. Sixteen chromosome arms have been previously identified to contain a high proportion of duplicated markers in *Oncorhynchus* (Omy01q, Omy02p, Omy03p, Omy06q, Omy07p, Omy10q, Omy12q, Omy13p, Omy13q, Omy15q, Omy17q, Omy18p, Omy19p, Omy21p, Omy23, and Omy26) (Brieuc *et al.* 2014; Kodama *et al.* 2014), a subset of which were identified in the *O. mykiss* haploid map in this study. These chromosome arms were just as likely to have regions of excess species-specific ancestry as arms without duplicated markers in both the female (*p-value* = 0.40, Wilcoxon rank sum test) and male (*p-value* = 0.47, Wilcoxon rank sum test).

Differences in relative recombination distances between O.mykiss and F_1 hybrids

Changes in recombination rates across the chromosome in a hybrid organism may serve to protect important genomic regions from introgression and conserve co-adapted gene complexes. In this study, changes in recombination rates in the hybrid genome were assessed by comparing map distances between markers mapped in both the sex specific RYHyb maps and the haploid *O. mykiss* map using linear and logistic models. Differences in recombination distances between markers in *O. mykiss* and the female hybrid map were observed in eight linkage groups (RYHyb02, RYHyb07, RYHyb10, RYHyb11, RYHyb12, RYHyb17, RYHyb20, and RYHyb25; Figure 2.3) and in seven groups when comparing *O. mykiss* and the male hybrid map (RYHyb02, RYHyb07, RYHyb10, RYHyb11, RYHyb12, RYHyb17, and RYHyb20; Figure 2.4).

Recombination differences between *O. mykiss* and the hybrid maps appeared to be largely consistent between the male and female hybrid parents. The seven linkage groups identified in the hybrid male parent were observed in the female parent. However, differences in RYHyb25_29 were identified in the female comparison parent only. Qualitatively, the shape of the lines fit to each data set were similar between the male and female hybrid parents, suggesting that hybridization may affect recombination in similar ways in both sexes.

Discussion

Here we aimed to examine the consequences of hybridization on the genome, following one generation of recombination between *O. mykiss* and cutthroat trout, to determine how hybridization might influence the persistence of species-specific chromosomal ancestry and recombination rate in hybrid individuals. Two hybrid sex-specific and one haploid *O. mykiss* linkage maps were created with 28 and 29 linkage groups respectively. Marker order between the

hybrid and *O. mykiss* maps was largely in agreement with previously produced maps (Miller *et al.* 2012; Brieuc *et al.* 2014; Kodama *et al.* 2014). Development of the *O. mykiss* haploid map provided additional genetic resources to the recently published draft genome (Berthelot *et al.* 2014) and previous genetic maps in the species (Sakamoto *et al.* 2000; Rexroad *et al.* 2008; Guyomard *et al.* 2012; Miller *et al.* 2012). In addition, placement of duplicated markers on the *O. mykiss* map confirms previous findings that duplicated genomic regions are not randomly distributed in *Oncorhynchus* and are conserved among species (Brieuc *et al.* 2014; Kodama *et al.* 2014). Genomic regions of excess species-specific ancestry were identified in the hybrid maps and covered 11 and 10% of the genome (based on distance) in the female and male. These regions existed throughout the genome and bias was evenly distributed between the parental species in both sexes. Differences in recombination between the hybrid and *O. mykiss* maps were identified in eight linkage groups in the female and seven in the male. Patterns of recombination were consistent between the sexes, with all seven linkage groups identified in the male also being identified in the female. Of the linkage groups identified, five contain a high proportion of duplicated markers in salmonids and two have known chromosomal polymorphisms between cutthroat and *O. mykiss*.

In this study, we investigated ancestry bias and recombination differences in one F_1 family. Differences in excess species-specific ancestry have previously been observed among families (Ostberg *et al.* 2013), likely driven by a combination of selection on family level polymorphisms as well as genetic drift. In this study, we were interested in species level differences that were driving patterns of excess species-specific ancestry. Ancestry bias in early generation hybrids that are caused by species level differences in genome structure are

anticipated to be present in all mapped families. Therefore, the results identified here are expected to describe deviations largely caused by processes such as selection and genetic drift. Because only one mapping family was investigated, differentiating between these causes is not possible, but instead, this study serves as a starting point for further investigation of how species specific differences affect the hybrid genome. Future studies using additional families could add insight into mechanisms responsible for excess species-specific ancestry patterns observed in this study.

The second limitation was that the results of the present study may have been influenced by our technique of inferring genotypes. The ancestry of genotypes uninformative to recombination was inferred by comparing the ancestries of neighboring loci. This technique is dependent on marker order and could introduce error if loci are not in the correct position. Marker order on chromosome arms is largely conserved among salmonids (Phillips *et al.* 2009; Naish *et al.* 2013). We compared marker order of the hybrid map to the haploid rainbow trout, Chinook (Brieuc *et al.* 2014), and coho salmon (Kodama *et al.* 2014) maps and observed similar marker orders suggesting that the inference technique used in this study likely did not introduce considerable error.

Creating sex-specific maps in a hybrid cross

Creating sex-specific genetic maps from interspecific hybrid crosses and bi-allelic markers can be difficult due to the high proportion of species-specific markers. Traditionally, when creating genetic maps of both sexes from diploid crosses, two individuals are crossed and sex-specific patterns of recombination are examined in the progeny (Rexroad *et al.* 2008; Lien *et*

al. 2011; Kodama *et al.* 2014). However, interspecific F_1 hybrids contain intact genomes contributed from each founding species, and for all bi-allelic loci with species-specific alleles, F_1s will be heterozygous resulting in an AB x AB marker when females and males are crossed. Crosses at such markers are expected to result in 50% of progeny with AB genotypes which are uninformative as to the allele contributed from each parent. AB x AB markers are often removed from sex-specific maps, but removing markers limits the power of a map, and if the distribution of species-specific loci is uneven throughout the genome, the resulting map could have low coverage in some areas. Analyses using maps in which species-specific markers have been removed may produce biased results if some genomic regions are missing from the map due to the non-random distribution of species-specific markers. In this study, we found that species-specific markers were non-randomly distributed and certain genomic regions contained a higher density of species-specific loci than others (Data not provided).

In order to retain species-specific markers in the female and male maps, we used a series of steps to infer the ancestry of each allele at genotypes that were uninformative on recombination. Our method used phased parental data to estimate ancestry at markers where genotypes were homozygous, multi-allelic (microsatellites contributed from Ostberg et al. 2013), and non-species-specific to build a framework of known ancestries. Because salmonids have almost complete recombination interference (Thorgaard *et al.* 1983; Brieuc *et al.* 2014), double crossover events are not anticipated over short chromosomal segments. Therefore, we used the framework of known ancestries to infer parental contribution if the surrounding markers both originated from the same parental chromosome and no evidence of recombination was observed between them. In addition, we used known ancestries identified as contributed from one parent to

inform ancestry contributed from the other. Such a technique could be useful for future studies of mapping families with a high proportion of AB x AB loci, such as F_1 interspecific hybrids.

Recombination suppression

In this study, recombination patterns were altered in the hybrid map compared to the *O. mykiss* map in eight chromosomes in the female and seven in the male. At least two of these linkage groups, RYHyb20 and RYHyb25, likely contain structural polymorphisms between cutthroat and *O. mykiss* (Naish *et al.* 2013; Ostberg *et al.* 2013). First, RYHyb20 contains two chromosomal polymorphisms. Evidence from Chinook salmon suggests that the entire q-arm of Omy20 (*O. mykiss* homolog to RYHyb20) was involved in a pericentric inversion, resulting in the conversion of a metacentric chromosome to acrocentric in *O. mykiss* (Naish *et al.* 2013). Additionally, Omy20 and Omy28 are acrocentric chromosomes in *O. mykiss*, but appear to have been involved in a Robertsonian fusion in cutthroat trout, resulting in the formation of a metacentric chromosome (Ostberg *et al.* 2013). During meiosis, an inversion loop may form on RYHyb20 to account for the pericentric inversion found in *O. mykiss*. Recombination in the hybrid maps appears to be least on the p-arm of RYHyb20 near the centromere. This area likely contains both the inversion in Omy20 and the fusion between homologs to Omy20 and Omy28. Second, RYHyb25 contains an additional difference in chromosome architecture between the species. *O. mykiss* has two common chromosome races, one with two acrocentric chromosomes Omy25 and Omy29 and one with a fusion between the two chromosomes (Phillips *et al.* 2005). For this study, *O. mykiss* from the Kamloops stock were used, which likely is polymorphic for these two chromosome races (Ostberg *et al.* 2013). In contrast, cutthroat trout have two

acrocentric chromosomes at their homologs for Omy25 and Omy29 (Ostberg *et al.* 2013). The mapping family used in this study likely contains a female heterozygous for the Omy25 and Omy29 fusion, while the male is likely homozygous for the acrocentric morphs (Ostberg *et al.* 2013). In RYHyb25_29 (fusion between RYHyb25 and RYHyb29), recombination appears suppressed in the female near the centromere or the presumed fusion site. Thus, the results from this study add additional evidence to the well-established phenomena that recombination is suppressed near Robertsonian polymorphisms (Colombo 2013; Dumas *et al.* 2015) and inversion sites (Jaarola *et al.* 1998; Andolfatto *et al.* 2001; Kirkpatrick 2010). Reduced recombination in heterokaryotypes may be caused by meiotic irregularities, such as pairing problems and segregation irregularities between heterokaryotypes, or selection against recombined genomes. Our data could not discern between these causes, but regardless of the mechanism, reduced recombination may protect alleles from disruption which could be important if fitness related genes are found within heterokaryotypic regions (Rane *et al.* 2015).

We hypothesize that RYHyb11 may also contain a chromosomal polymorphism between cutthroat trout and *O. mykiss*. Differences in recombination were observed between both the female and male hybrid and *O. mykiss*, with relatively less recombination observed at centromeric regions compared to the telomeric regions on the q-arm in the hybrid individuals. Similar patterns were observed by Ostberg *et al.* (2013). Reduced recombination at centric regions is expected if species differ in their Robertsonian arrangements. However, some marker order differences were also observed between the hybrid and *O. mykiss* maps, which suggests an inversion in one of the species may be responsible for the changes in recombination. Further

analysis is necessary to identify the mechanisms that may be driving changes in recombination on RYHyb11.

Recombination was altered in the hybrid relative to *O. mykiss* on chromosomes likely to contain a high proportion of duplicated markers and may be a consequence of increased pairing and recombination between homeologous chromosomes in hybrids. The results of this study confirm the previous findings that re-diploidization after whole genome duplication has not been uniform throughout the salmonid genome (Brieuc *et al.* 2014; Kodama *et al.* 2014). Sixteen chromosome arms, eight of which were identified in *O. mykiss* as part of this study, contain elevated proportions of duplicated markers and are hypothesized to be involved in tetrasomic inheritance (Brieuc *et al.* 2014; Kodama *et al.* 2014). In this study, we observed changes in recombination patterns between the hybrid maps and *O. mykiss* on five chromosome arms (RYHyb02p, RYHyb07p, RYHyb10q, RYHyb12q, and RYHyb17q) that a high proportion of duplicated markers have been identified on in other species (Brieuc *et al.* 2014; Kodama *et al.* 2014). One explanation for this pattern could be increased tetrasomic inheritance caused by the preferential pairing of homeologous chromosomes in the hybrids. F_1 hybrids contain intact genomes contributed from each founding species, and homeologous chromosomes with the same species-specific ancestry may be more similar than homologous chromosomes from different species ancestries. As such, homeologs may preferentially pair during meiosis resulting in increased tetrasomic inheritance (Hickok 1978). Increased recombination between homeologous chromosomes is expected to reduce the genetic distance between loci on homologous chromosomes. This pattern is clearly visible on RYHyb02p, RYHyb12q, and RYHyb17q, but not on RYHyb07p and RYHyb10q, where distances between markers appear to increase in the

hybrid maps. In addition, if increased homeologous pairing was responsible for changes in recombination patterns on chromosomes with a high proportion of duplicated markers, one would expect to observe changes in recombination in both homeologs (see Kodama et al. 2014 for homeologous pairings). In this study, changes in hybrid recombination rates relative to *O. mykiss* was not observed in all pairs of homeologous chromosomes. It is unclear if this is due to inadequate marker coverage on some arms reducing our ability to detect differences in recombination between the hybrids and *O. mykiss*, or whether some other mechanism may be at least partially driving patterns of recombination in the hybrids. The lack of congruence between homeologous chromosomes and the unexpected reduction of map distances on RYHyb07p, and RYHyb10q, suggest that the influence of other mechanisms, such as existing differences in recombination between the founding species, cannot be overlooked as a factor in some of the observed patterns of recombination.

For genomic regions where increased homeologous pairing may be occurring in the hybrid, the result could be chromosomes that are less susceptible to introgression. Homeologs of the same species are expected to pair and exchange genetic material because of their reduced divergence relative to homologs of different species ancestries (Hickok 1978). As a result, recombination would only occur between chromosomes from the same species and introgression would be limited in such genomic regions. Homeologous pairing may serve to protect species-specific adaptations from disruption, but also would hinder the creation of novel gene combinations that recombination between homologous chromosomes of different species ancestries would produce. It is unclear why duplicated genomic regions have been retained in

only some chromosome arms of the salmonid genome. Therefore, the full implication of increased homeologous pairing to evolution is unclear.

Differences in recombination between the hybrid and *O. mykiss* were largely the same in both the female and male. All seven chromosomes identified to have changes in recombination in the hybrid male compared to *O. mykiss* were also identified in the hybrid female. In addition, the relationship of recombination (shape of the lines) was similar in the two sexes, suggesting that hybridization has similar effects on recombination in both the female and male. These findings are unsurprising for chromosomes with known physical polymorphisms. However, for chromosomes in which increased tetrasomic inheritance may be related to changes in recombination, this result is surprising. Previous research of tetrasomy in salmonids has observed strong sex bias, with tetrasomy occurring exclusively in males (Wright *et al.* 1983; Allendorf & Danzmann 1997). Given the patterns of recombination suppression observed in this study, we hypothesize that tetrasomic inheritance may be occurring in hybrid females. The discrepancy between this study and previous work may be partially due to differences in marker type, genomic resources, and test statistics. Earlier studies used allozymes at a limited number of loci and often focused on identifying pseudolinkage, a statistical excess of non-parental gametes (Davisson *et al.* 1973; Wright *et al.* 1983; Allendorf & Danzmann 1997). In contrast, the present study used thousands of loci to explore patterns of recombination on chromosomes with known homeologous relationships. This information was then used to hypothesize causes of the observed changes in recombination. Low levels of homeologous recombination could result in changes to map distances that could be detected using the techniques outlined in this study, but may not be enough to identify statistical linkage. Further analysis is needed to identify if a link

exists between homeologous pairing and recombination patterns in the female and male, or if the patterns observed in this study are the result of other biological processes.

Excess species-specific ancestry

A relatively small proportion of the genome was identified as having excess species-specific ancestry as part of this study, but regions identified are likely the result of genetic drift. Here, regions of excess species-specific ancestry were distributed across 11 and 10% of the genome in the female and male respectively. This is a considerably smaller portion of the genome than identified in other taxa (Gadau *et al.* 1999; Fishman *et al.* 2001; Lopez-Fernandez & Bolnick 2007) and is suggestive that the genomes of cutthroat trout and *O. mykiss* are highly similar and few genomic incompatibilities exist between the species. Although few regions of excess species-specific ancestry were identified, those that were observed are likely largely the result of genetic drift. Genetic drift is expected to result in the random distribution of excess species-specific ancestry throughout the genome, with similar proportions of cutthroat trout and *O. mykiss* bias. In this study, regions of species-specific ancestry were widespread, identified on 33 and 23 chromosome arms in the female and male, and both female and male maps had nearly equal proportions of excesses for cutthroat trout and *O. mykiss*. In addition, 16 chromosome arms had excess ancestry of the same species in the female and male maps, which is no different than what would be expected by chance alone (*p-value* = 0.06, permutation test). Thus the distribution of excess species-specific ancestry suggests that genetic drift may have a strong influence in species-specific alleles observed in this study. Ostberg *et al.* (2013), using two mapping families of hybridized cutthroat trout and *O. mykiss*, observed similar proportions of excess species-

specific ancestry as observed in this study, and with few regions in common between mapping families. Taken together, the results from this study and Ostberg et al. (2013) suggest genetic drift may play an important role in the formation of the hybrid genome. As a result, predictions of wide scale ancestry patterns may be difficult in hybridized populations.

Summary

Our results indicate that changes to recombination following hybridization are largely driven by chromosomal polymorphisms between founding species and may be influenced by increased homeologous pairing between chromosomes with the same species ancestry. Both processes could have profound implications by limiting recombination between interspecific chromosomes, thus resulting in genomic regions with reduced introgression. If adaptively important loci are located in these regions, changes to recombination patterns following hybridization may serve to protect these areas from disruption, potentially conserving local adaptations.

Only a small proportion of the hybrid genome contained species-specific bias in this study. This finding has important implications as it may suggest that few genetic incompatibilities exist between cutthroat trout and *O. mykiss* and may explain the high proportion of hybrid populations that exist (Shepard *et al.* 2005; Metcalf *et al.* 2008; Kovach *et al.* 2011). Additionally, the outcomes of early generation hybridization influences the genetic material available for natural selection to operate on in the future, thus the influence of genetic drift may play an important role in hybrid evolution. The low levels of species-specific bias observed in this study suggest that rapid evolution following hybridization is likely not

ubiquitous across the genome and that a high proportion of species-specific genetic material may be conserved within a population. For genomic regions that contained excess species-specific ancestry, the pattern appeared to be largely random and not chromosome specific. As such, predicting the outcomes of introgression throughout the genome is unlikely.

Although this study focused on hybridization between Yellowstone cutthroat trout and *O. mykiss*, the findings are likely generalizable to hybridization between *O. mykiss* and all cutthroat trout subspecies. In addition, the genetic resources created as part of this study provide a novel framework for future genomic studies of cutthroat trout and *O. mykiss*. Lastly, this study provides important guidance for mapping hybrid families with genetic data dominated by bi-allelic markers.

References

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

Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001) The problems with hybrids: setting conservation guidelines. *Trends in Ecology & Evolution* **16**, 613-622.

Allendorf FW, Thorgaard GH (1984) Tetraploidy and the evolution of salmonid fishes. In: *Evolutionary Genetics of Fishes* (ed. Turner BJ), pp. 1-27. Plenum, New York.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403-410.

Alves JM, Chikhi L, Amorim A, Lopes AM (2014) The 8p23 Inversion polymorphism determines local recombination heterogeneity across human populations. *Genome Biology and Evolution* **6**, 921-930.

Andolfatto P, Depaulis F, Navarro A (2001) Inversion polymorphisms and nucleotide variability in drosophila. *Genetical Research* **77**, 1-8.

Baack EJ, Rieseberg LH (2007) A genomic view of introgression and hybrid speciation. *Current Opinion in Genetics & Development* **17**, 513-518.

Baird NA, Etter PD, Atwood TS, *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *Plos One* **3**.

Barton NH, Hewitt GM (1989) Adaptation, speciation, and hybrid zones. *Nature* **341**, 497-503.

Behnke R (1992) *Native Trout of Western North America* American Fisheries Society, Bethesda Md.

Berthelot C, Brunet F, Chalopin D*, et al.* (2014) The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nature Communications* **5**.

Borowsky R, Cohen D (2013) Genomic consequences of ecological speciation in *Astyanax* cavefish. *Plos One* **8**.

Brieuc MSO, Waters CD, Seeb JE, Naish KA (2014) A dense linkage map for Chinook salmon (*Oncorhynchus tshawytscha*) reveals variable chromosomal divergence after an ancestral whole genome duplication event. *G3-Genes Genomes Genetics* **4**, 447-460.

Burke JM, Arnold ML (2001) Genetics and the fitness of hybrids. *Annual Review of Genetics* **35**, 31-52.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology* **22**, 3124-3140.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping loci de novo from short-read sequences. *G3-Genes Genomes Genetics* **1**, 171-182.

Chapman MA, Abbott RJ (2010) Introgression of fitness genes across a ploidy barrier. *New Phytologist* **186**, 63-71.

Colombo PC (2013) Micro-evolution in grasshoppers mediated by polymorphic Robertsonian translocations. *Journal of Insect Science* **13**.

Crete-Lafreniere A, Weir LK, Bernatchez L (2012) Framing the Salmonidae family phylogenetic portrait: A more complete picture from increased taxon sampling. *Plos One* **7**.

Davisson MT, Akeson EC (1993) Recombination suppression by heterozygous Robertsonian chromosomes in the mouse. *Genetics* **133**, 648-667.

Davisson MT, Wright JE, Atherton LM (1973) Cytogenetic analysis of pseudolinkage of LDH loci in teleost genus *Salvelinus*. *Genetics* **73**, 645-658.

Doerge RW (1996) Constructing genetic maps by rapid chain delineation. Journal of Quantitative Trait Loci.

Dumas D, Catalan J, Britton-davidian J (2015) Reduced recombination patterns in Robertsonian hybrids between chromosomal races of the house mouse: chiasma analyses. *Heredity* **114**, 56-64.

Ellstrand NC, Schierenbeck KA (2006) Hybridization as a stimulus for the evolution of invasiveness in plants? *Euphytica* **148**, 35-46.

Fishman L, Kelly AJ, Morgan E, Willis JH (2001) A genetic map in the *Mimulus guttatus* species complex reveals transmission ratio distortion due to heterospecific interactions. *Genetics* **159**, 1701-1716.

Gadau J, Page RE, Werren JH (1999) Mapping of hybrid incompatibility loci in *Nasonia*. *Genetics* **153**, 1731-1741.

Gold JR, Avise JC, Gall GAE (1977) Chromosome cytology in the cutthroat trout series *Salmo-clarki Salmonidae*. *Cytologia (Tokyo)* **42**, 373-382.

Gresswell RE (2011) Biology, Status, and Management of the Yellowstone Cutthroat Trout. *North American Journal of Fisheries Management* **31**, 782-812.

Guyomard R, Boussaha M, Krieg F, Hervet C, Quillet E (2012) A synthetic rainbow trout linkage map provides new insights into the salmonid whole genome duplication and the conservation of synteny among teleosts. *Bmc Genetics* **13**.

Guyomard R, Mauger S, Tabet-Canale K*, et al.* (2006) A Type I and Type II microsatellite linkage map of rainbow trout (*Oncorhynchus mykiss*) with presumptive coverage of all chromosome arms. *Bmc Genomics* **7**.

Halverson A (2010) *An entirely synthetic fish: how rainbow trout beguiled America and overran the world* Yale University Press, New Haven.

Hickok LG (1978) Homoeologous chromosone-pairing - frequency differences in inbred and intraspecific hybrid polyploid ferns. *Science* **202**, 982-984.

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *Plos Genetics* **6**.

Hufton AL, Panopoulou G (2009) Polyploidy and genome restructuring: a variety of outcomes. *Current Opinion in Genetics & Development* **19**, 600-606.

Jaarola M, Martin RH, Ashley T (1998) Direct evidence for suppression of recombination within two pericentric inversions in humans: A new sperm-FISH technique. *American Journal of Human Genetics* **63**, 218-224.

Kirkpatrick M (2010) How and why chromosome inversions evolve. *Plos Biology* **8**.

Kodama M, Brieuc MSO, Devlin RH, Hard JJ, Naish KA (2014) Comparative mapping between coho salmon (*Oncorhynchus kisutch*) and three other dalmonids duggests a role for chromosomal rearrangements in the retention of duplicated regions following a whole genome duplication event. *G3-Genes Genomes Genetics* **4**, 1717-1730.

Kovach RP, Eby LA, Corsi MP (2011) Hybridization between Yellowstone cutthroat trout and rainbow trout in the upper Snake River basin, Wyoming. *North American Journal of Fisheries Management* **31**, 1077-1087.

Kulathinal RJ, Stevison LS, Noor MAF (2009) The genomics of speciation in drosophila: diversity, divergence, and introgression estimated using low-coverage genome sequencing. *Plos Genetics* **5**.

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**.

Li S, Chen Y, Gao H, Yin T (2010) Potential chromosomal introgression barriers revealed by linkage analysis in a hybrid of *Pinus massoniana* and *P. hwangshanensis*. *Bmc Plant Biology* **10**.

Lien S, Gidskehaug L, Moen T*, 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**.

Lopez-Fernandez H, Bolnick DI (2007) What causes partial F_1 hybrid viability? Incomplete penetrance versus genetic variation. *Plos One* **2**.

Loudenslager EJ, Thorgaard GH (1979) Karyotypic and evolutionary relationship of the Yellowstone (*Salmo clarki bouvieri*) and west-slope (*Salmo clarki lewisi*) cutthroat trout. *Journal of the Fisheries Research Board of Canada* **36**, 630-635.

Loxterman JL, Keeley ER (2012) Watershed boundaries and geographic isolation: patterns of diversification in cutthroat trout from western North America. *Bmc Evolutionary Biology* **12**.

Macqueen DJ, Johnston IA (2014) A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. *Proceedings of the Royal Society B-Biological Sciences* **281**.

Mallet J (2005) Hybridization as an invasion of the genome. *Trends in Ecology & Evolution* **20**, 229-237.

Margarido GRA, Souza AP, Garcia AAF (2007) OneMap: software for genetic mapping in outcrossing species. *Hereditas* **144**, 78-79.

Martin NH, Bouck AC, Arnold ML (2006) Detecting adaptive trait introgression between *Iris fulva* and *I. brevicaulis* in highly selective field conditions. *Genetics* **172**, 2481-2489.

McGaugh SE, Noor MAF (2012) Genomic impacts of chromosomal inversions in parapatric drosophila species. *Philosophical Transactions of the Royal Society B-Biological Sciences* **367**, 422-429.

Metcalf JL, Siegle MR, Martin AP (2008) Hybridization dynamics between Colorado's native cutthroat trout and introduced rainbow trout. *Journal of Heredity* **99**, 149-156.

Miller MR, Brunelli JP, Wheeler PA*, et al.* (2012) A conserved haplotype controls parallel adaptation in geographically distant salmonid populations. *Molecular Ecology* **21**, 237-249.

Muhlfeld CC, Kalinowski ST, McMahon TE*, et al.* (2009) Hybridization rapidly reduces fitness of a native trout in the wild. *Biology Letters* **5**, 328-331.

Naish KA, Phillips RB, Brieuc MSO, Newton LR, Elz AE, Park LK (2013) Comparative genome mapping between Chinook salmon (*Oncorhynchus tshawytscha*) and rainbow

trout (*O. mykiss*) based on homologous microsatellite loci. *G3-Genes Genomes Genetics* **3**, 2281-2288.

Ostberg CO, Hauser L, Pritchard VL, Garza JC, Naish KA (2013) Chromosome rearrangements, recombination suppression, and limited segregation distortion in hybrids between Yellowstone cutthroat trout (*Oncorhynchus clarkii bouvieri*) and rainbow trout (*O. mykiss*). *Bmc Genomics* **14**.

Otto SP (2007) The evolutionary consequences of polyploidy. *Cell* **131**, 452-462.

Parisod C, Holderegger R, Brochmann C (2010) Evolutionary consequences of autopolyploidy. *New Phytologist* **186**, 5-17.

Perrier C, Bourret V, Kent MP, Bernatchez L (2013) Parallel and nonparallel genome-wide divergence among replicate population pairs of freshwater and anadromous Atlantic salmon. *Molecular Ecology* **22**, 5577-5593.

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

Phillips RB, Keatley KA, Morasch MR*, et al.* (2009) Assignment of Atlantic salmon (*Salmo salar*) linkage groups to specific chromosomes: conservation of large syntenic blocks

corresponding to whole chromosome arms in rainbow trout (*Oncorhynchus mykiss*). *BMC Genetics* **10**.

Phillips RB, Morasch MR, Wheeler PA, Thorgaard GH (2005) Rainbow trout (*Oncorhynchus mykiss*) of Idaho and Alaskan origin (2n=58) share a chromosome fusion relative to trout of California origin (2n=60). *Copeia*, 661-664.

Quiros CF (1982) Tetrasomic segregation for multiple alleles in alfalfa. *Genetics* **101**, 117-127.

R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rane RV, Rako L, Kapun M, Lee SF, Hoffmann AA (2015) Genomic evidence for role of inversion 3RP of drosophila melanogaster in facilitating climate change adaptation. *Molecular Ecology* **24**, 2423-2432.

Rexroad CE, III, Palti Y, Gahr SA, Vallejo RL (2008) A second generation genetic map for rainbow trout (*Oncorhynchus mykiss*). *Bmc Genetics* **9**.

Sakamoto T, Danzmann RG, Gharbi K, *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.

Shepard BB, May BE, Urie W (2005) Status and conservation of westslope cutthroat trout within the western United States. *North American Journal of Fisheries Management* **25**, 1426-1440.

Stevison LS, Hoehn KB, Noor MAF (2011) Effects of inversions on within- and between-species recombination and divergence. *Genome Biology and Evolution* **3**, 830-841.

Storchova Z, Pellman D (2004) From polyploidy to aneuploidy, genome instability and cancer. *Nature Reviews Molecular Cell Biology* **5**, 45-54.

Thorgaard GH (1983) Chromosomal differences among rainbow trout populations. *Copeia*, 650-662.

Thorgaard GH, Allendorf FW, Knudsen KL (1983) Gene-centromere mapping in rainbow trout - high interference over long map distances. *Genetics* **103**, 771-783.

Thornton DH, Murray DL (2014) Influence of hybridization on niche shifts in expanding coyote populations. *Diversity and Distributions* **20**, 1355-1364.

Wendel JF (2000) Genome evolution in polyploids. *Plant Molecular Biology* **42**, 225-249.

Wolf PG, Soltis PS, Soltis DE (1989) Tetrasomic inheritance and chromosome-pairing behavior in the naturally occurring autotetraploid *Heuchera grossulariifolia* (Saxifragaceae). *Genome* **32**, 655-659.

Wright JE, Johnson K, Hollister A, May B (1983) Meiotic models to explain classical linkage, pseudolinkage, and chromosome-pairing in tetraploid derivative salmonid genomes. *Isozymes-Current Topics in Biological and Medical Research* **10**, 239-260.

Young MK (2008) Colorado River Cutthroat Trout (*Oncorhynchus clarkii pleuriticus*): A Technical Conservation Assessment. USDA Forest Service, Rocky Mountain Region.

Young MK (2009) Greenback Cutthroat Trout (*Oncorhynchus clarkii stomias*): a technical conservation assessment. USDA Forest Service, Rocky Mountain Region.

Tables

	Excess species specific ancestry	
Chromosome arm	Female	Male
RYHyb01p	*O. mykiss*	-
RYHyb01q	Cutthroat trout	-
RYHyb02p	*O. mykiss*	-
RYHyb02q	*O. mykiss*	Cutthroat trout
RYHyb03p	-	-
RYHyb03q	*O. mykiss*	*O. mykiss*
RYHyb04p	*O. mykiss*	-
RYHyb04q	*O. mykiss*	-
RYHyb05p	*O. mykiss*	*O. mykiss*
RYHyb05q	*O. mykiss*	-
RYHyb06p	Cutthroat trout	-
RYHyb06q	Cutthroat trout	*O. mykiss*
RYHyb07p	*O. mykiss*	*O. mykiss*
RYHyb07q	*O. mykiss*	*O. mykiss*
RYHyb08p	-	Cutthroat trout
RYHyb08q	-	Cutthroat trout
RYHyb09p	-	-
RYHyb09q	Cutthroat trout	*O. mykiss*
RYHyb10p	*O. mykiss*	*O. mykiss*

RYHyb10q	-	-
RYHyb11p	*O. mykiss*	-
RYHyb11q	-	-
RYHyb12p	Cutthroat trout	-
RYHyb12q	-	-
RYHyb13p	-	-
RYHyb13q	*O. mykiss*	-
RYHyb14p	-	Cutthroat trout
RYHyb14q	Cutthroat trout	-
RYHyb15p	-	-
RYHyb15q	*O. mykiss*	-
RYHyb16p	-	-
RYHyb16q	*O. mykiss*	-
RYHyb17p	Cutthroat trout	Cutthroat trout
RYHyb17q	-	-
RYHyb18p	Cutthroat trout	Cutthroat trout
RYHyb18q	Cutthroat trout	Cutthroat trout
RYHyb19p	-	Cutthroat trout
RYHyb19q	Cutthroat trout	Cutthroat trout
RYHyb20p	-	-
RYHyb20q	-	-
RYHyb21p	-	-

RYHyb21q	-	-
RYHyb22p	-	-
RYHyb22q	Cutthroat trout	Cutthroat trout
RYHyb23	Cutthroat trout	Cutthroat trout
RYHyb24	*O. mykiss*	*O. mykiss*
RYHyb25p	Cutthroat trout	Cutthroat trout
RYHyb25q	Cutthroat trout	-
RYHyb26	*O. mykiss*	*O. mykiss*
RYHyb27	*O. mykiss*	*O. mykiss*
RYHyb28	Cutthroat trout	Cutthroat trout
Total *O. mykiss*	18	10
Total cutthroat trout	15	13

Table 2.1: Chromosome arms with observed excess species-specific ancestry in the male and female. p- and q-arms are denoted. Chromosomes without p or q designations are acrocentric. Numbers in parentheses represent number of regions within the chromosome arm that were identified as bias.

Figures

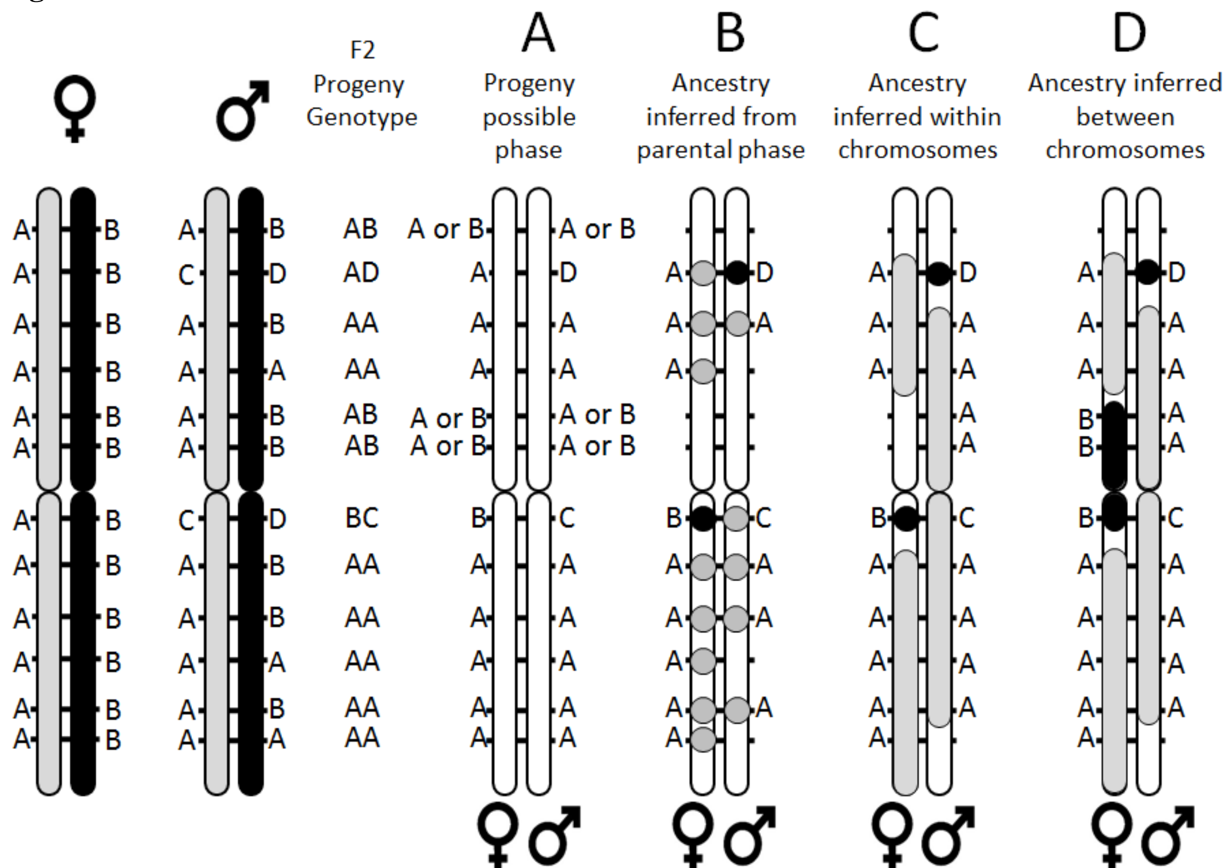


Figure 2.1: Techniques used to create sex specific maps. On the left, species specific chromosomes in F_1 hybrids are designated grey and black. Markers are either biallelic (AB) or multiallelic (ABCD). The genotype of a single F_2 progeny is given. A, B, C,

and D illustrate the sequential computation of likely genotypes in one F_2 offspring, following recombination in F_1 parents. (A) Multiple phase possibilities exist given the observed genotypes at each marker. (B) Ancestry informative markers (multiallelic markers, AB x AA, or F_2 homozygotes from AB x AB crosses) are used to phase known species-specific (grandparental) haplotypes at markers where possible. (C) Double recombination events are rare in salmonids, particularly across short chromosomal segments. Therefore, the species ancestry of F_2 heterozygotes resulting from an AB x AB cross can be inferred by extending previously identified haplotypes when the neighboring ancestries are in agreement. (D) Finally, if the haplotype of one F_1 parental chromosome is known, that information may be useful to infer the chromosomal ancestry of the other F_1 parent.

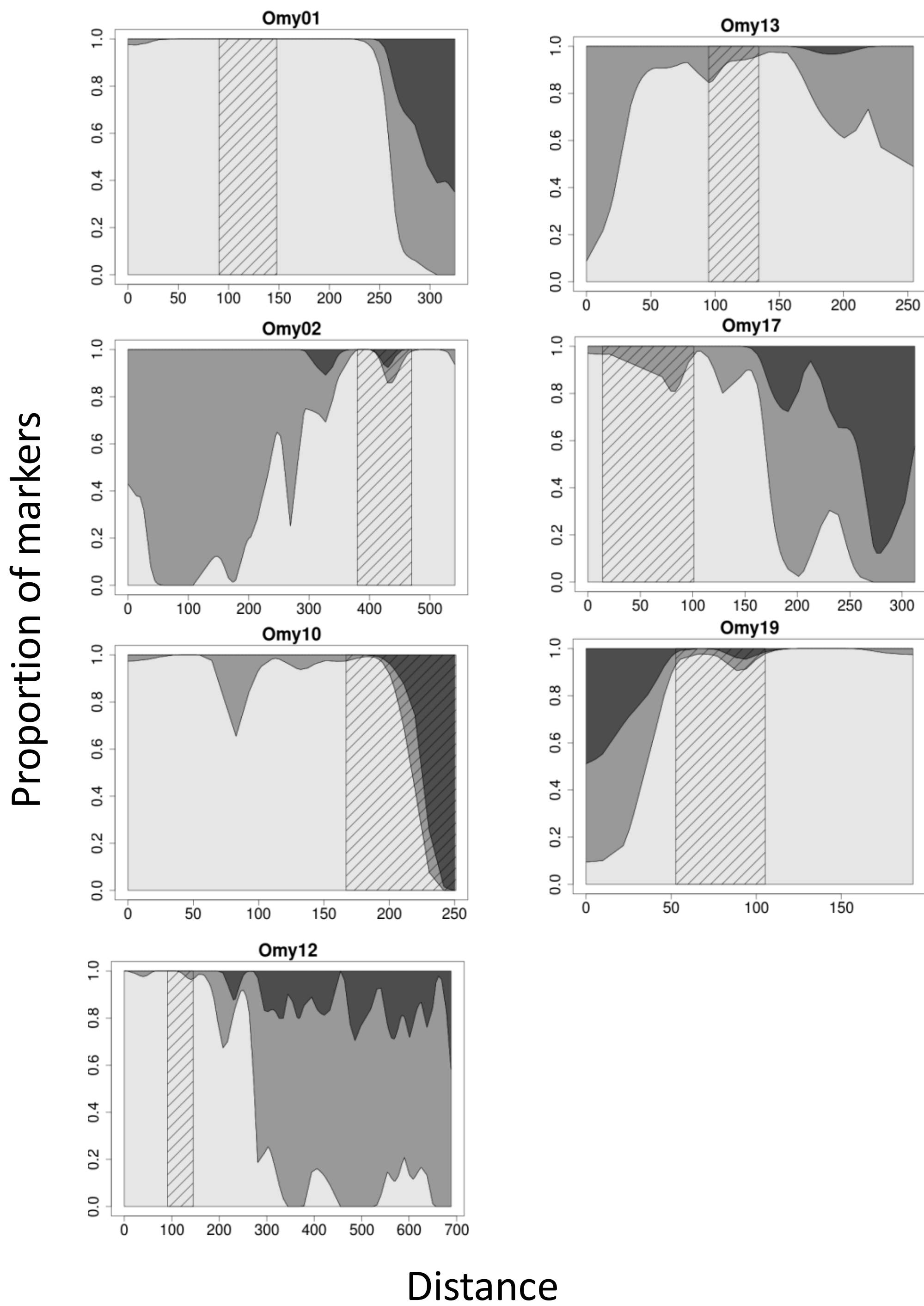


Figure 2.2: Distribution of duplicated markers on each of the eight chromosome arms in the rainbow trout haploid map that had greater than 0.1 proportion of duplicated markers per arm.

Light grey represents non-duplicated markers. Grey represents duplicated markers in which only one locus was polymorphic. Dark grey represents duplicated markers in which both loci were polymorphic. Dashed rectangles represent centromeric regions. Linkage groups are oriented p-arm to q-arm.

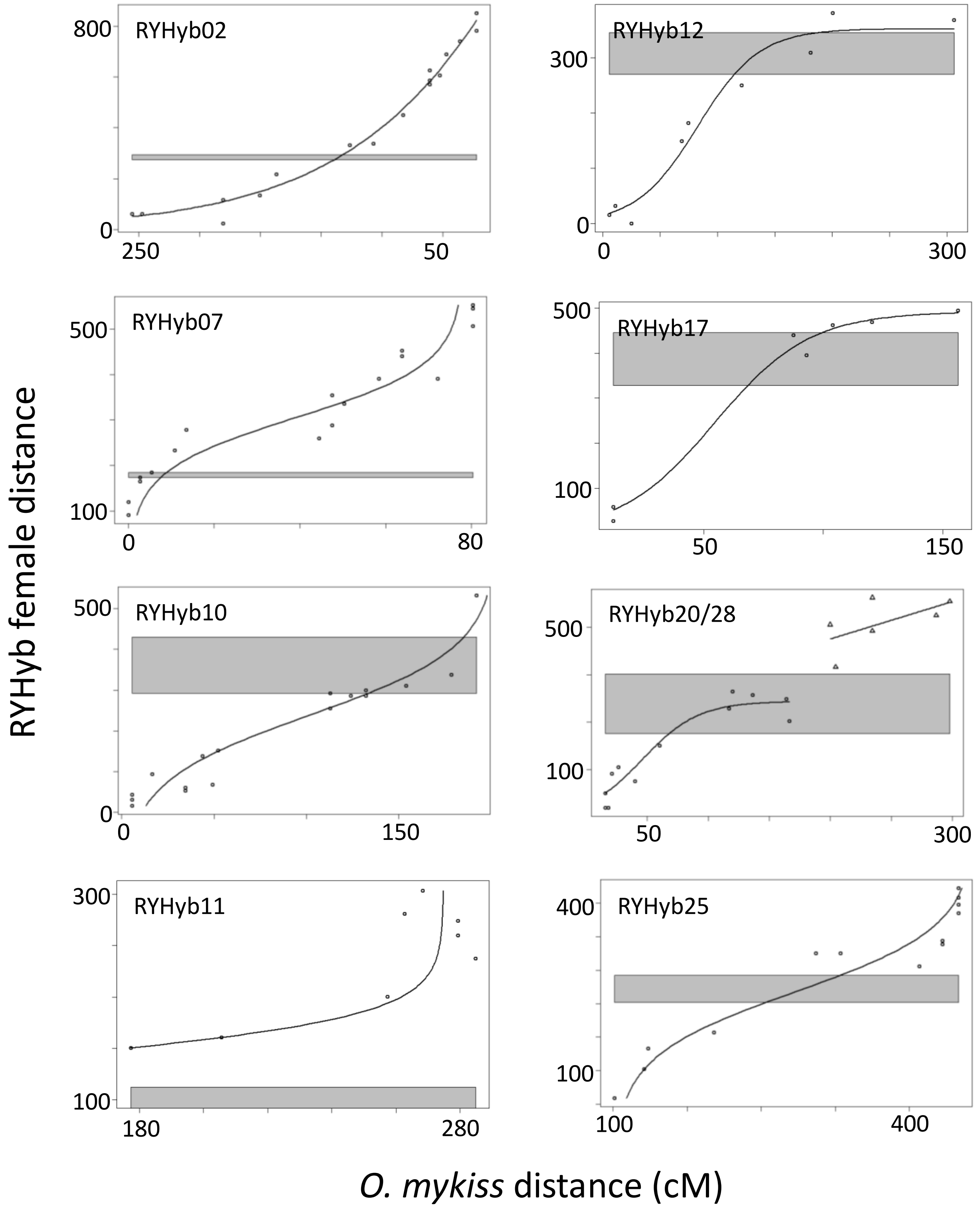

RYHyb02
800
0
250
50
RYHyb12
300
0
0
300
RYHyb07
500
100
0
80
RYHyb17
500
100
50
150
RYHyb10
500
0
0
150
RYHyb20/28
500
100
50
300
RYHyb11
300
100
180
280
RYHyb25
400
100
100
400
RYHyb female distance
O. mykiss distance (cM)

Figure 2.3: Comparison of genetic distances between the F_1 female hybrid and haploid *O. mykiss* maps. Black circles represent positions of markers mapped in both studies. Grey areas represent centromeric regions in the F_1 female hybrid map. RYHyb20 is a fusion of two chromosomes in the *O. mykiss* map. Therefore, black circles represent loci that mapped to Omy20 and triangles represent loci that mapped to Omy28. Markers are oriented p-arm to q-arm. The accompanying linkage groups are represented above (*O. mykiss*) and to the right (hybrid) of each plot. Red areas represent purported centromeres, while yellow regions show the areas where the markers from each plot are mapped in each linkage group.

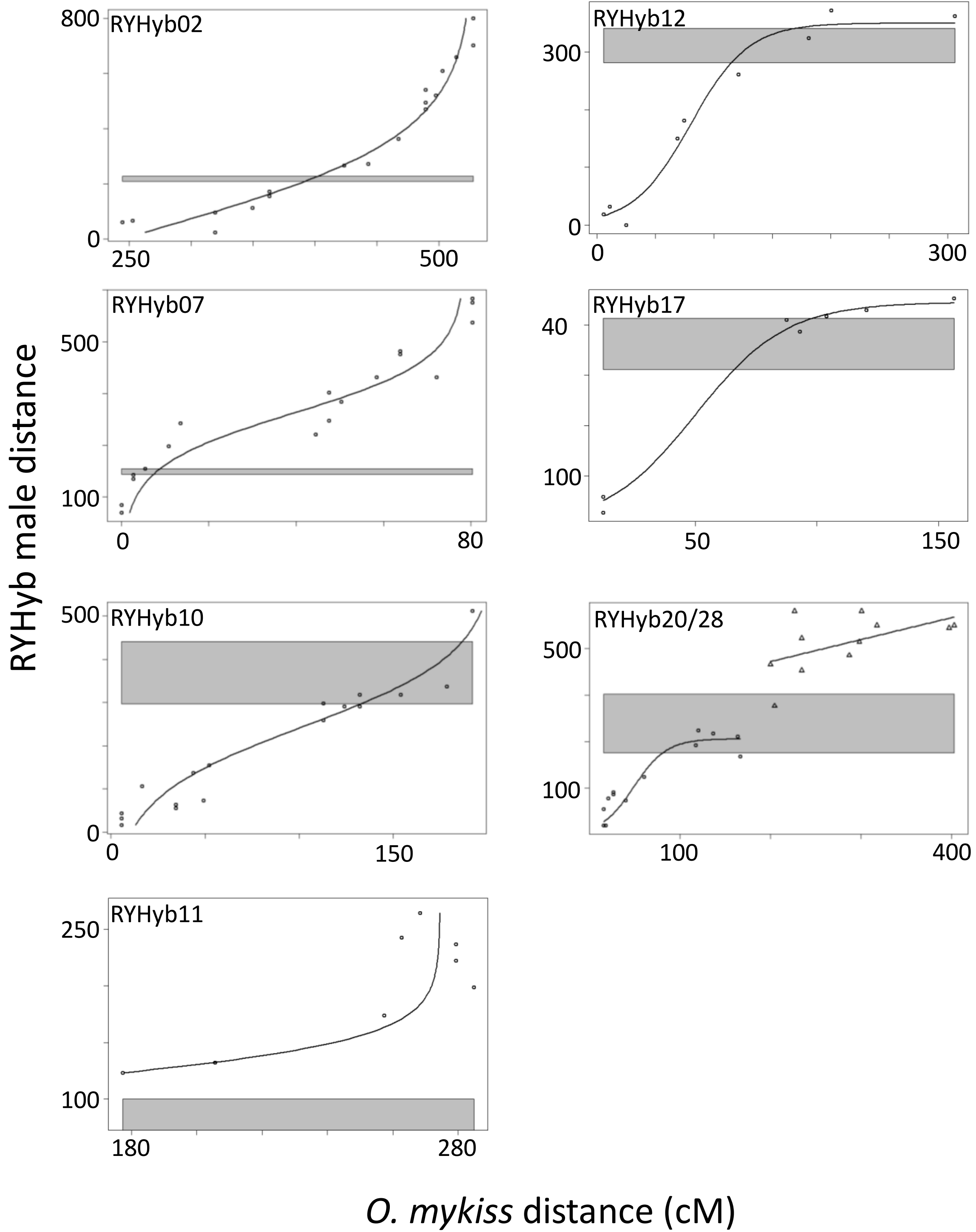

RYHyb male distance
O. mykiss distance (cM)
RYHyb02
800
0
250
500
RYHyb12
300
0
0
300
RYHyb07
500
100
0
80
RYHyb17
40
100
50
150
RYHyb10
500
0
0
150
RYHyb20/28
500
100
100
400
RYHyb11
250
100
180
280

Figure 2.4: Comparison of genetic distances between the F_1 male hybrid and haploid *O. mykiss* maps. Black circles represent positions of markers mapped in both studies. Grey areas represent centromeric regions in the F_1 male hybrid map. RYHyb20 is a fusion of two chromosomes in the *O. mykiss* map. Therefore, black circles represent loci that mapped to Omy20 and triangles represent loci that mapped to Omy28. Markers are oriented p-arm to q-arm. The accompanying linkage groups are represented above (*O. mykiss*) and to the right (hybrid) of each plot. Red areas represent purported centromeres, while yellow regions show the areas where the markers from each plot are mapped in each linkage group.

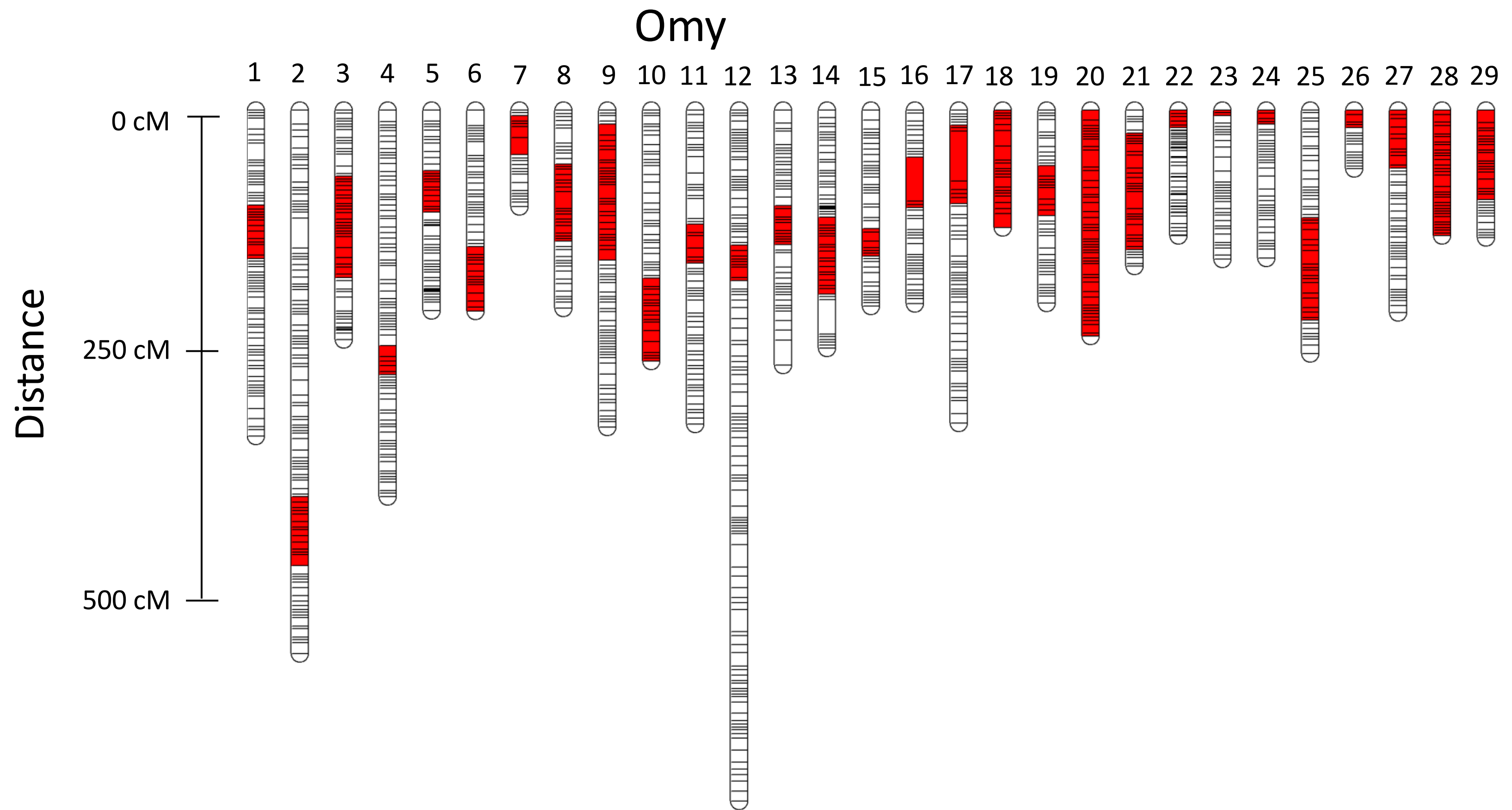


Figure S2.1: Graphical representation of 29 *Oncorhynchus mykiss* haploid female linkage groups. Red sections represent areas where centromeres are inferred based on alignment with Chinook salmon linkage map. Groups 1 – 21 and 25 are metacentric, while 22 – 24 and 26 – 29 are acrocentric. Linkage groups are aligned p-arm to q-arm.

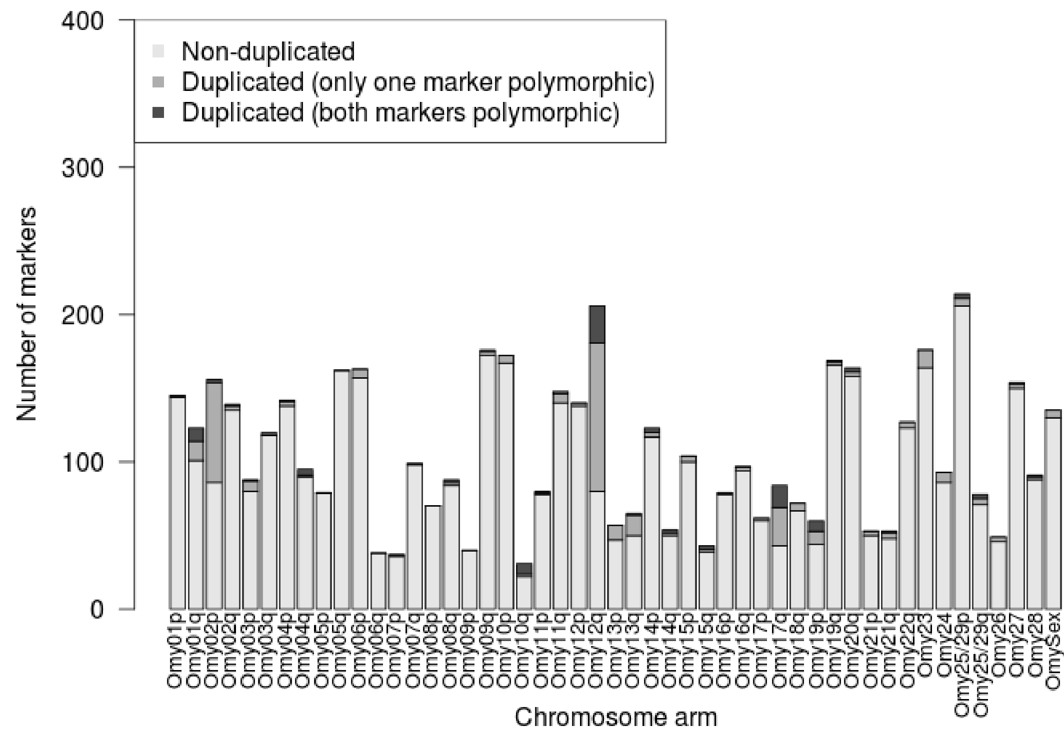
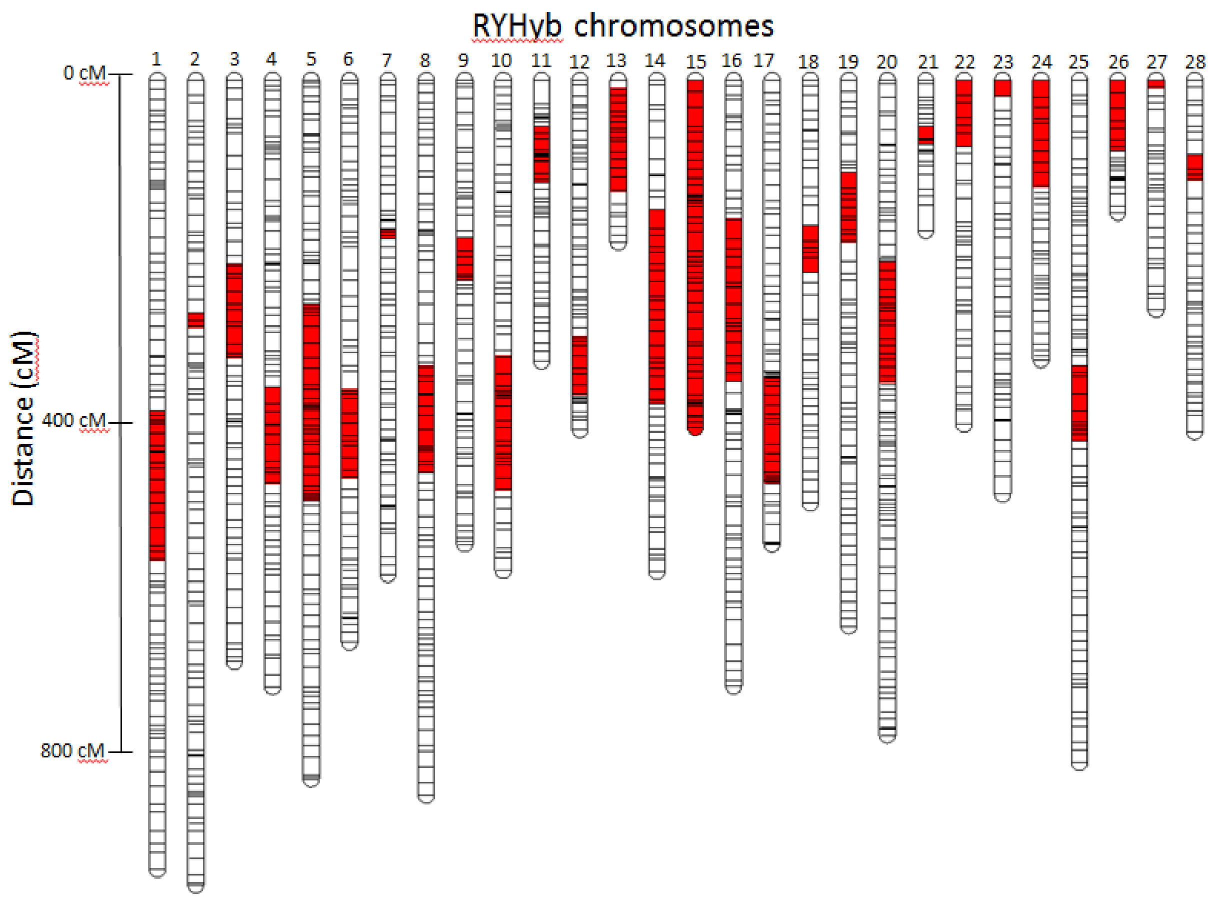


Figure S2.2: Distribution of duplicated markers across the *Oncorhynchus mykiss* linkage map.



RYHyb chromosomes
Distance (cM)
0 cM
400 cM
800 cM
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

Figure S2.3: Graphical representation of 28 linkage groups identified in the F_1 hybrid cross between Yellowstone cutthroat trout (*Oncorhynchus clarki bouvieri*) and *O. mykiss*. The map was constructed as a consensus map of both sexes. Red sections represent areas where centromeres are inferred based on alignment with Chinook salmon linkage map. Linkage groups are aligned p-arm to q-arm.

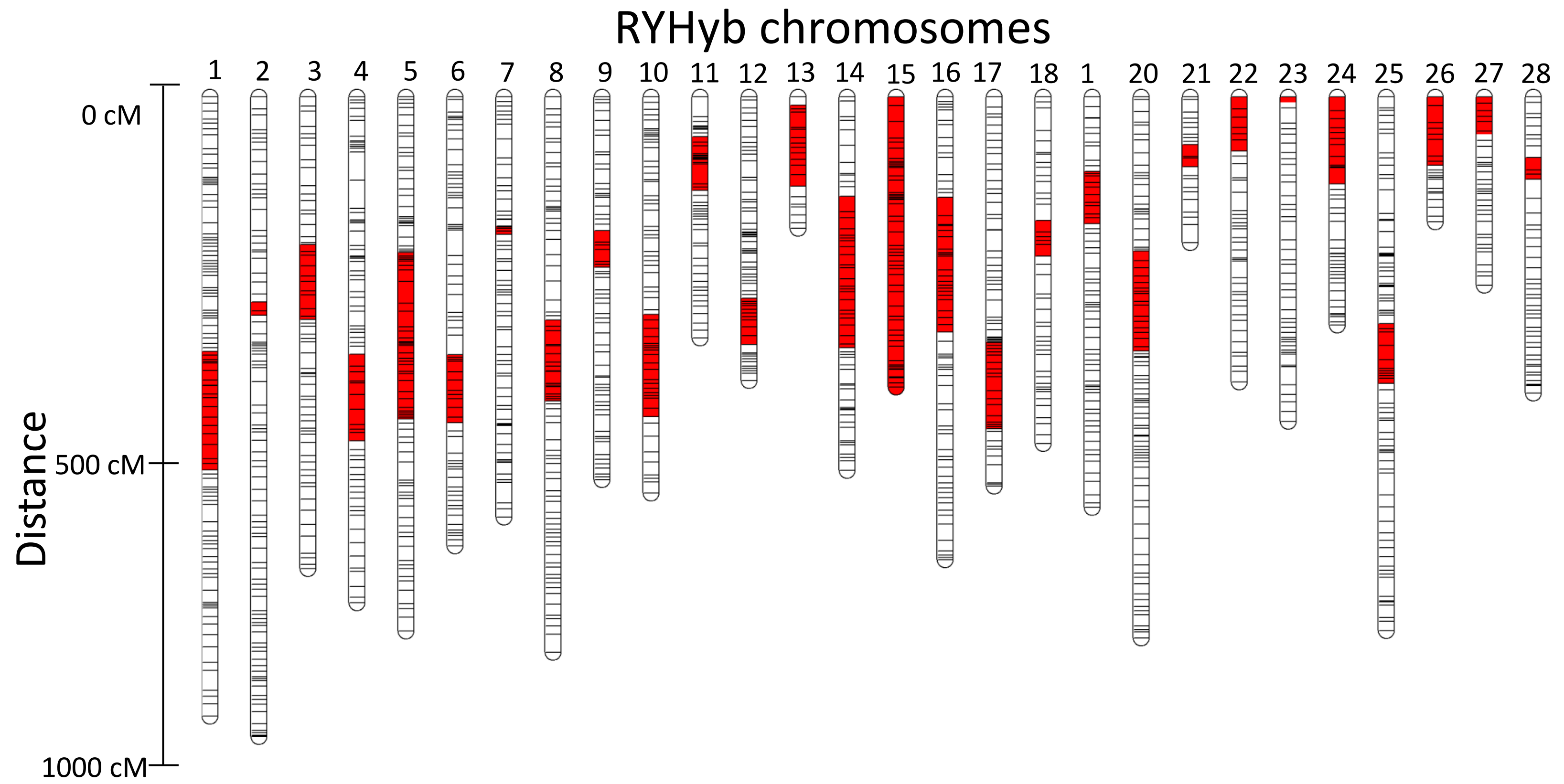


Figure S2.4: Graphical representation of 28 linkage groups identified in a female F_1 hybrid between Yellowstone cutthroat trout (*Oncorhynchus clarki bouvieri*) and *O. mykiss*. Red sections represent areas where centromeres are inferred based on alignment with Chinook salmon linkage map. Linkage groups are aligned p-arm to q-arm.

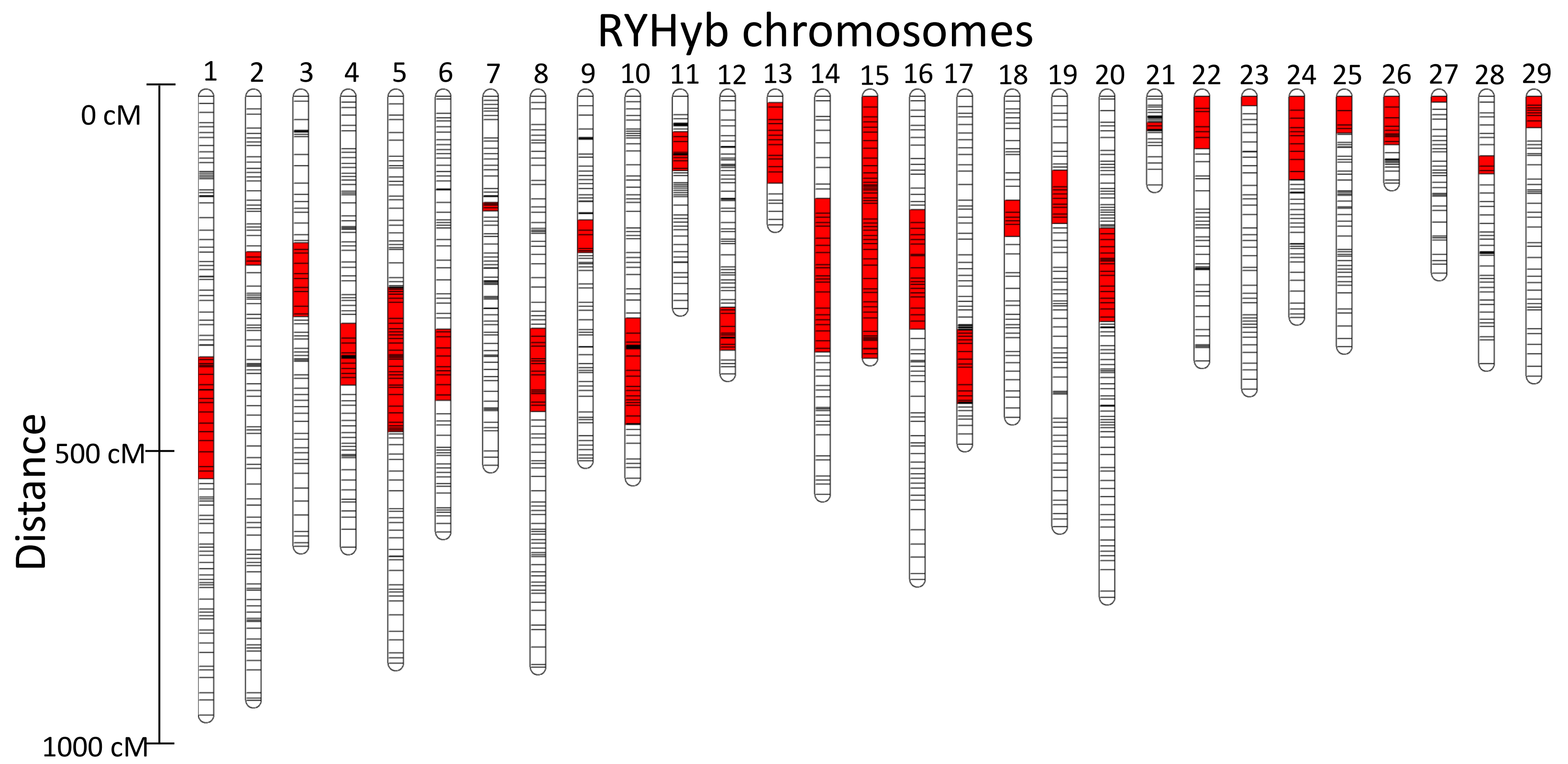

RYHyb chromosomes
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
0 cM
500 cM
1000 cM
Distance

Figure S2.5: Graphical representation of 29 linkage groups identified in a male F_1 hybrid between Yellowstone cutthroat trout (*Oncorhynchus clarki bouvieri*) and *O. mykiss*. Red sections represent areas where centromeres are inferred based on alignment with Chinook salmon linkage map. Linkage groups are aligned p-arm to q-arm.

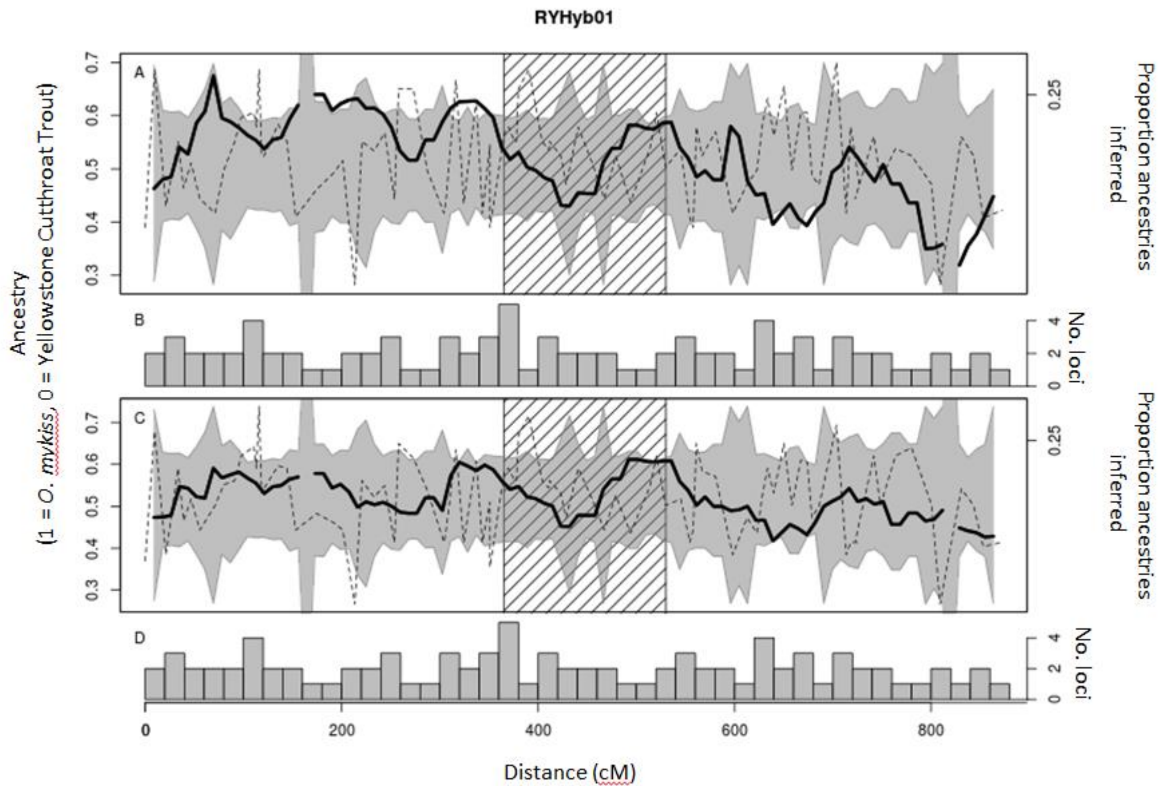


Figure S2.6: Distribution of mean species specific marker ancestries across linkage group RYHyb01, contributed from the dam (A and B) and sire (C and D), as observed in F_2 progeny. In (A) and (C), black lines represent mean ancestry (1 = *O. mykiss*, 0 = Yellowstone cutthroat trout) across the linkage group (left Y axis). Confidence intervals at each region are designated in grey. Confidence intervals were estimated by bootstrapping loci 100,000 times at each position. Ancestries are based on sex specific maps. Some ancestries were inferred (Figure 2.1); the dotted line represents the proportion of ancestries inferred at each locus (right y axis). The dashed box is the centromeric region identified by comparative mapping with other salmonids. Plots B and D are the distribution of markers used to produce A and C respectively.

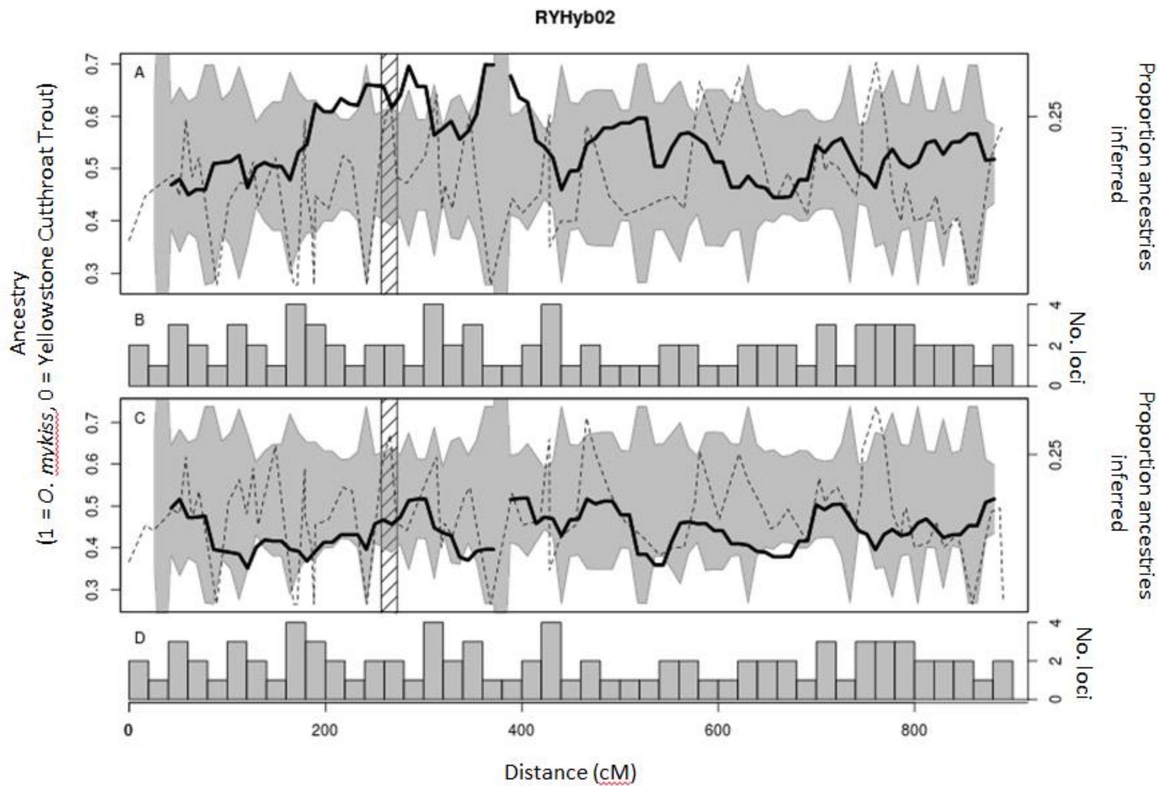


Figure S2.7: Distribution of mean species specific marker ancestries across linkage group RYHyb02, contributed from the dam (A and B) and sire (C and D), as observed in F_2 progeny. In (A) and (C), black lines represent mean ancestry (1 = *O. mykiss*, 0 = Yellowstone cutthroat trout) across the linkage group (left Y axis). Confidence intervals at each region are designated in grey. Confidence intervals were estimated by bootstrapping loci 100,000 times at each position. Ancestries are based on sex specific maps. Some ancestries were inferred (Figure 2.1); the dotted line represents the proportion of ancestries inferred at each locus (right y axis). The dashed box is the centromeric region identified by comparative mapping with other salmonids. Plots B and D are the distribution of markers used to produce A and C respectively.

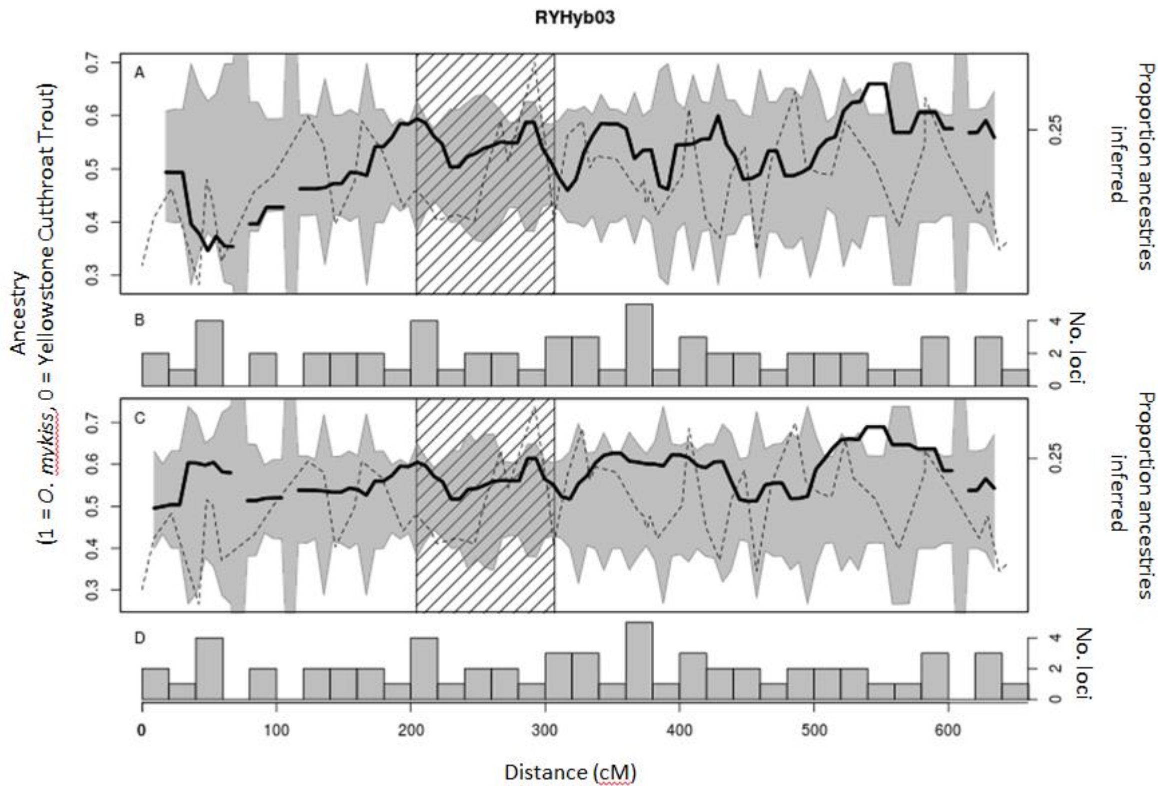


Figure S2.8: Distribution of mean species specific marker ancestries across linkage group RYHyb03, contributed from the dam (A and B) and sire (C and D), as observed in F_2 progeny. In (A) and (C), black lines represent mean ancestry (1 = *O. mykiss*, 0 = Yellowstone cutthroat trout) across the linkage group (left Y axis). Confidence intervals at each region are designated in grey. Confidence intervals were estimated by bootstrapping loci 100,000 times at each position. Ancestries are based on sex specific maps. Some ancestries were inferred (Figure 2.1); the dotted line represents the proportion of ancestries inferred at each locus (right y axis). The dashed box is the centromeric region identified by comparative mapping with other salmonids. Plots B and D are the distribution of markers used to produce A and C respectively.

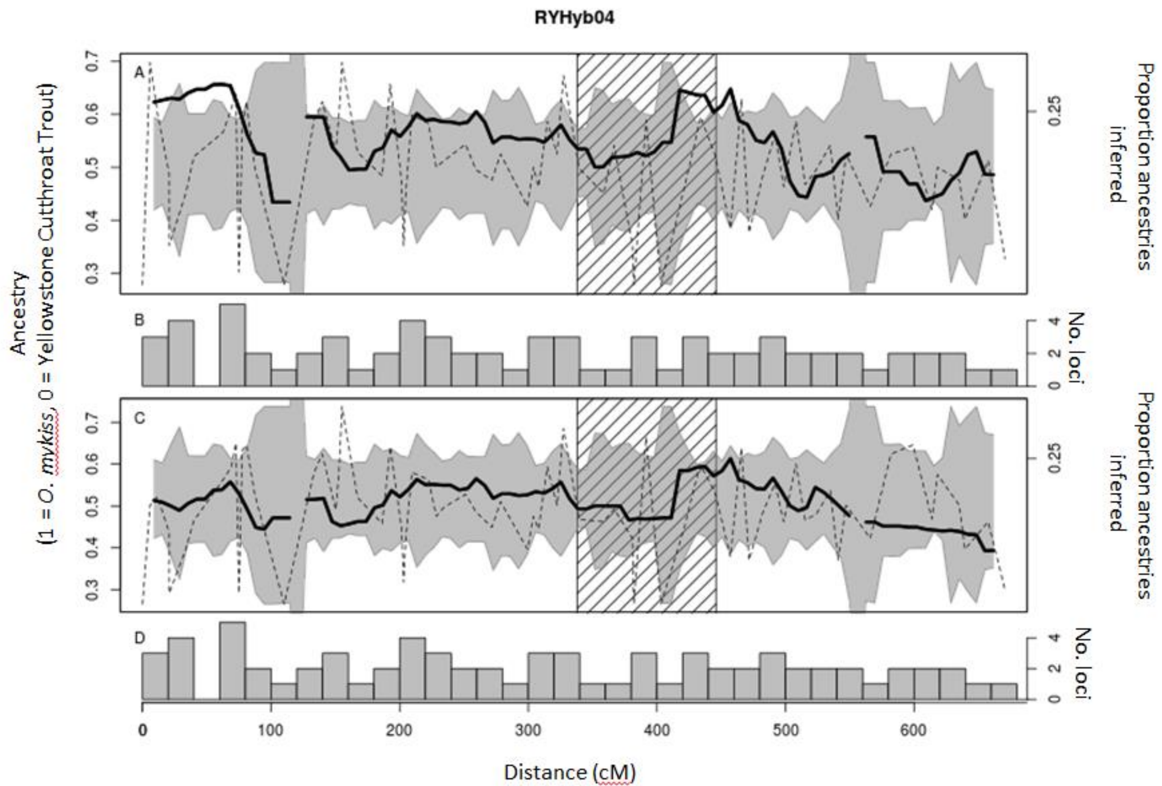


Figure S2.9: Distribution of mean species specific marker ancestries across linkage group RYHyb04, contributed from the dam (A and B) and sire (C and D), as observed in F_2 progeny. In (A) and (C), black lines represent mean ancestry (1 = *O. mykiss*, 0 = Yellowstone cutthroat trout) across the linkage group (left Y axis). Confidence intervals at each region are designated in grey. Confidence intervals were estimated by bootstrapping loci 100,000 times at each position. Ancestries are based on sex specific maps. Some ancestries were inferred (Figure 2.1); the dotted line represents the proportion of ancestries inferred at each locus (right y axis). The dashed box is the centromeric region identified by comparative mapping with other salmonids. Plots B and D are the distribution of markers used to produce A and C respectively.

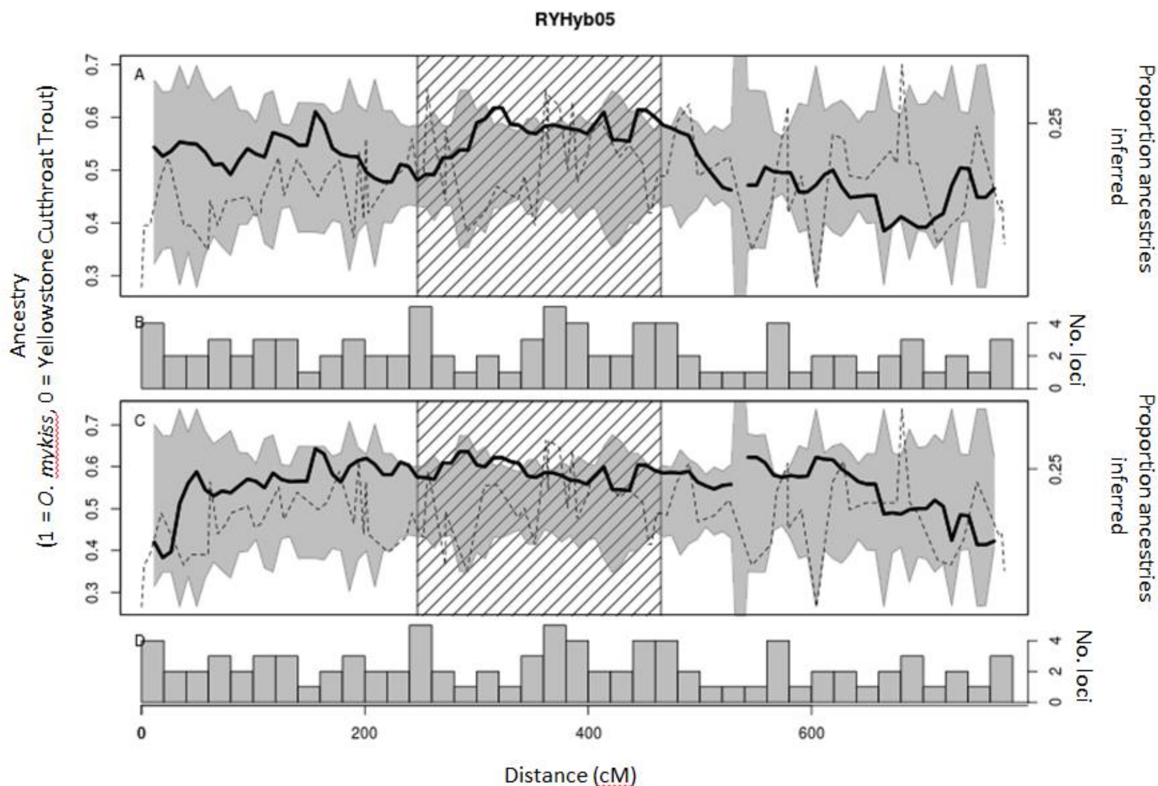


Figure S2.10: Distribution of mean species specific marker ancestries across linkage group RYHyb05, contributed from the dam (A and B) and sire (C and D), as observed in F_2 progeny. In (A) and (C), black lines represent mean ancestry (1 = *O. mykiss*, 0 = Yellowstone cutthroat trout) across the linkage group (left Y axis). Confidence intervals at each region are designated in grey. Confidence intervals were estimated by bootstrapping loci 100,000 times at each position. Ancestries are based on sex specific maps. Some ancestries were inferred (Figure 2.1); the dotted line represents the proportion of ancestries inferred at each locus (right y axis). The dashed box is the centromeric region identified by comparative mapping with other salmonids. Plots B and D are the distribution of markers used to produce A and C respectively.

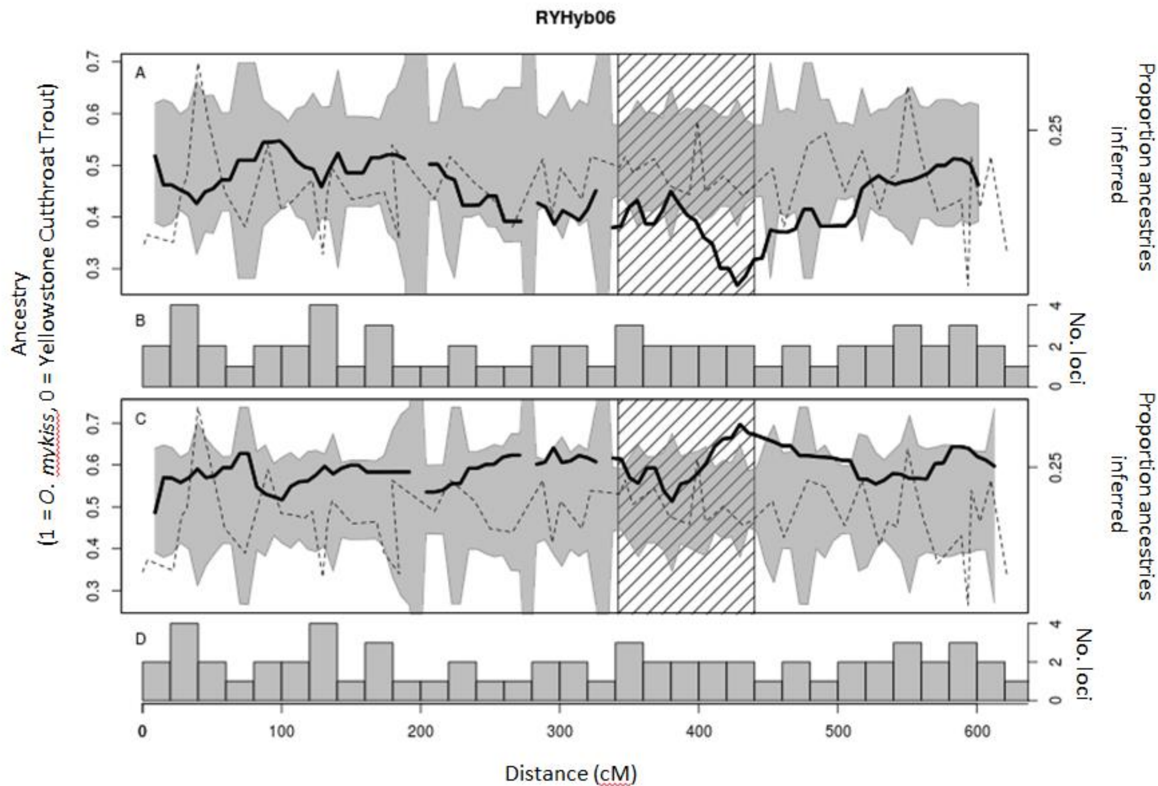


Figure S2.11: Distribution of mean species specific marker ancestries across linkage group RYHyb06, contributed from the dam (A and B) and sire (C and D), as observed in F_2 progeny. In (A) and (C), black lines represent mean ancestry (1 = *O. mykiss*, 0 = Yellowstone cutthroat trout) across the linkage group (left Y axis). Confidence intervals at each region are designated in grey. Confidence intervals were estimated by bootstrapping loci 100,000 times at each position. Ancestries are based on sex specific maps. Some ancestries were inferred (Figure 2.1); the dotted line represents the proportion of ancestries inferred at each locus (right y axis). The dashed box is the centromeric region identified by comparative mapping with other salmonids. Plots B and D are the distribution of markers used to produce A and C respectively.

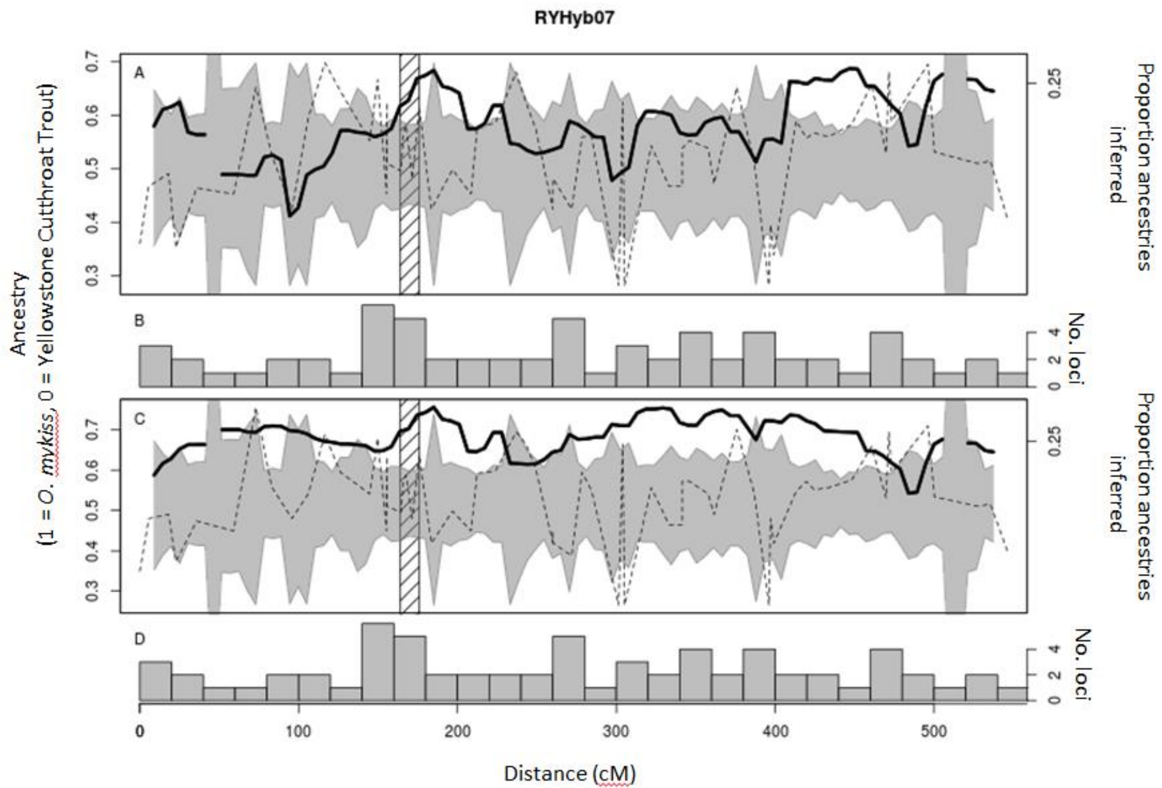


Figure S2.12: Distribution of mean species specific marker ancestries across linkage group RYHyb07, contributed from the dam (A and B) and sire (C and D), as observed in F_2 progeny. In (A) and (C), black lines represent mean ancestry (1 = *O. mykiss*, 0 = Yellowstone cutthroat trout) across the linkage group (left Y axis). Confidence intervals at each region are designated in grey. Confidence intervals were estimated by bootstrapping loci 100,000 times at each position. Ancestries are based on sex specific maps. Some ancestries were inferred (Figure 2.1); the dotted line represents the proportion of ancestries inferred at each locus (right y axis). The dashed box is the centromeric region identified by comparative mapping with other salmonids. Plots B and D are the distribution of markers used to produce A and C respectively.

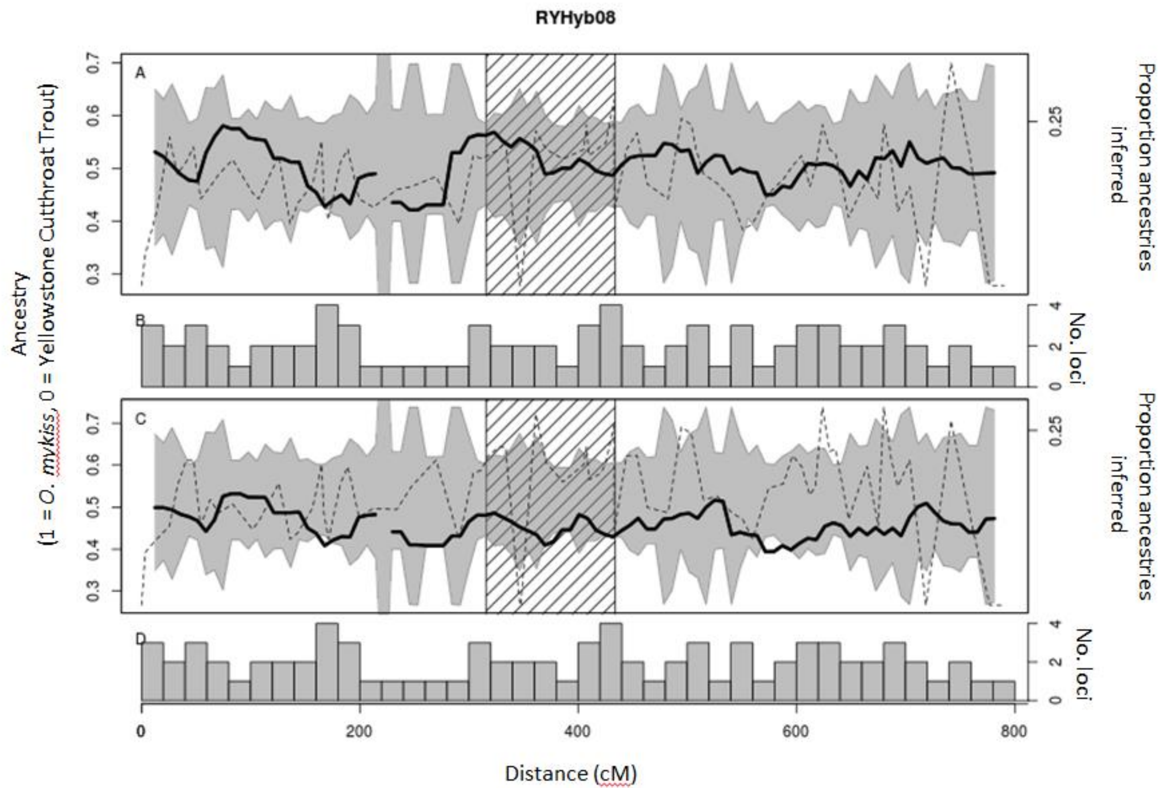


Figure S2.13: Distribution of mean species specific marker ancestries across linkage group RYHyb08, contributed from the dam (A and B) and sire (C and D), as observed in F_2 progeny. In (A) and (C), black lines represent mean ancestry (1 = *O. mykiss*, 0 = Yellowstone cutthroat trout) across the linkage group (left Y axis). Confidence intervals at each region are designated in grey. Confidence intervals were estimated by bootstrapping loci 100,000 times at each position. Ancestries are based on sex specific maps. Some ancestries were inferred (Figure 2.1); the dotted line represents the proportion of ancestries inferred at each locus (right y axis). The dashed box is the centromeric region identified by comparative mapping with other salmonids. Plots B and D are the distribution of markers used to produce A and C respectively.

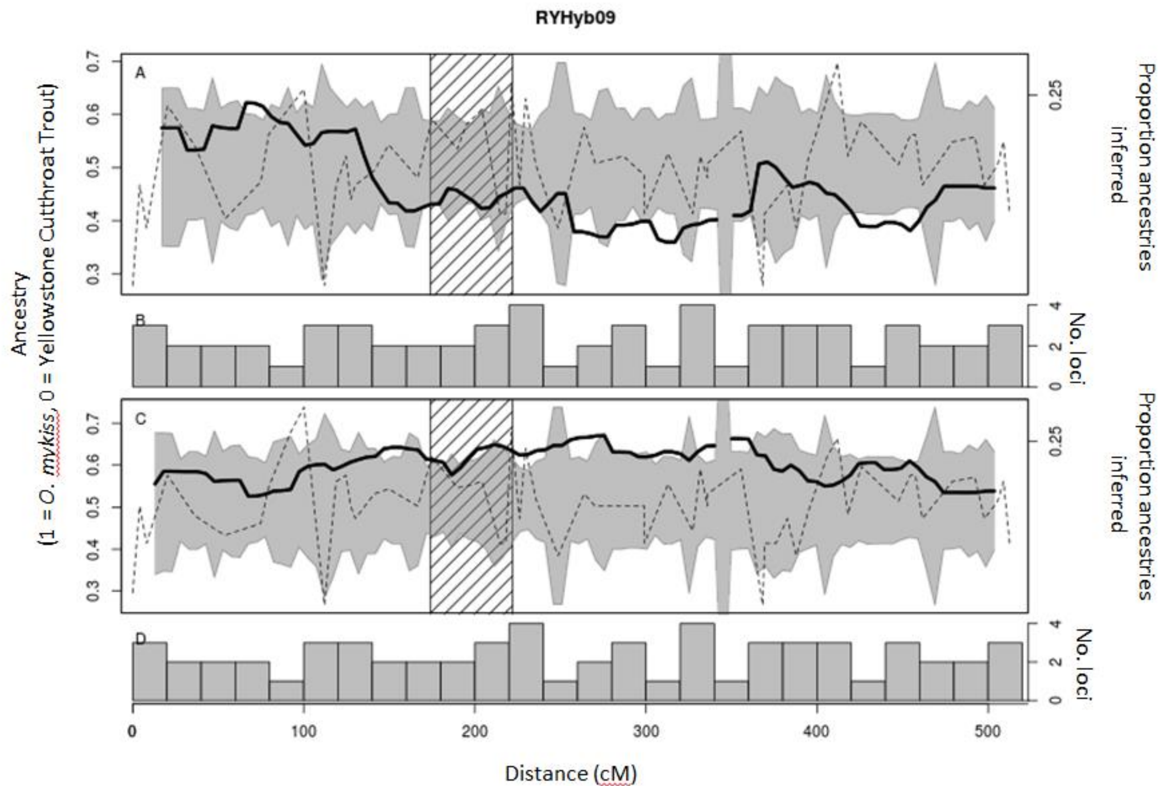


Figure S2.14: Distribution of mean species specific marker ancestries across linkage group RYHyb09, contributed from the dam (A and B) and sire (C and D), as observed in F_2 progeny. In (A) and (C), black lines represent mean ancestry (1 = *O. mykiss*, 0 = Yellowstone cutthroat trout) across the linkage group (left Y axis). Confidence intervals at each region are designated in grey. Confidence intervals were estimated by bootstrapping loci 100,000 times at each position. Ancestries are based on sex specific maps. Some ancestries were inferred (Figure 2.1); the dotted line represents the proportion of ancestries inferred at each locus (right y axis). The dashed box is the centromeric region identified by comparative mapping with other salmonids. Plots B and D are the distribution of markers used to produce A and C respectively.

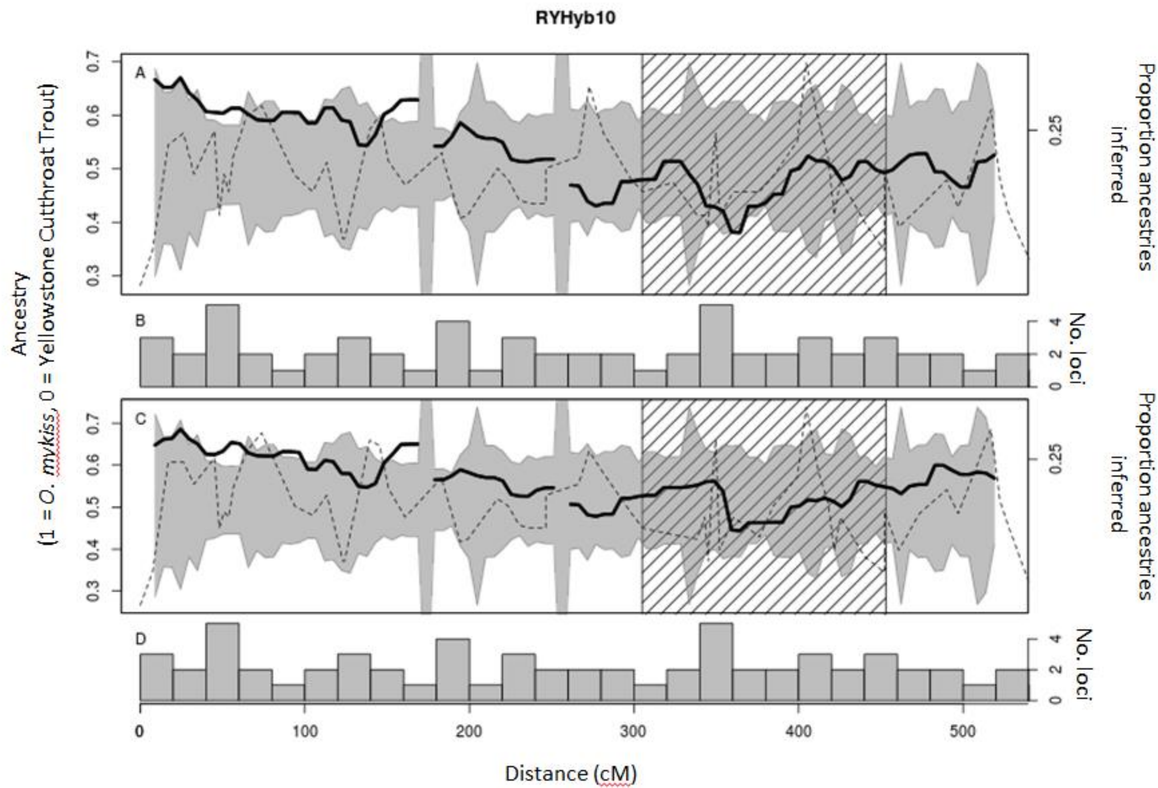


Figure S2.15: Distribution of mean species specific marker ancestries across linkage group RYHyb10, contributed from the dam (A and B) and sire (C and D), as observed in F_2 progeny. In (A) and (C), black lines represent mean ancestry (1 = *O. mykiss*, 0 = Yellowstone cutthroat trout) across the linkage group (left Y axis). Confidence intervals at each region are designated in grey. Confidence intervals were estimated by bootstrapping loci 100,000 times at each position. Ancestries are based on sex specific maps. Some ancestries were inferred (Figure 2.1); the dotted line represents the proportion of ancestries inferred at each locus (right y axis). The dashed box is the centromeric region identified by comparative mapping with other salmonids. Plots B and D are the distribution of markers used to produce A and C respectively.

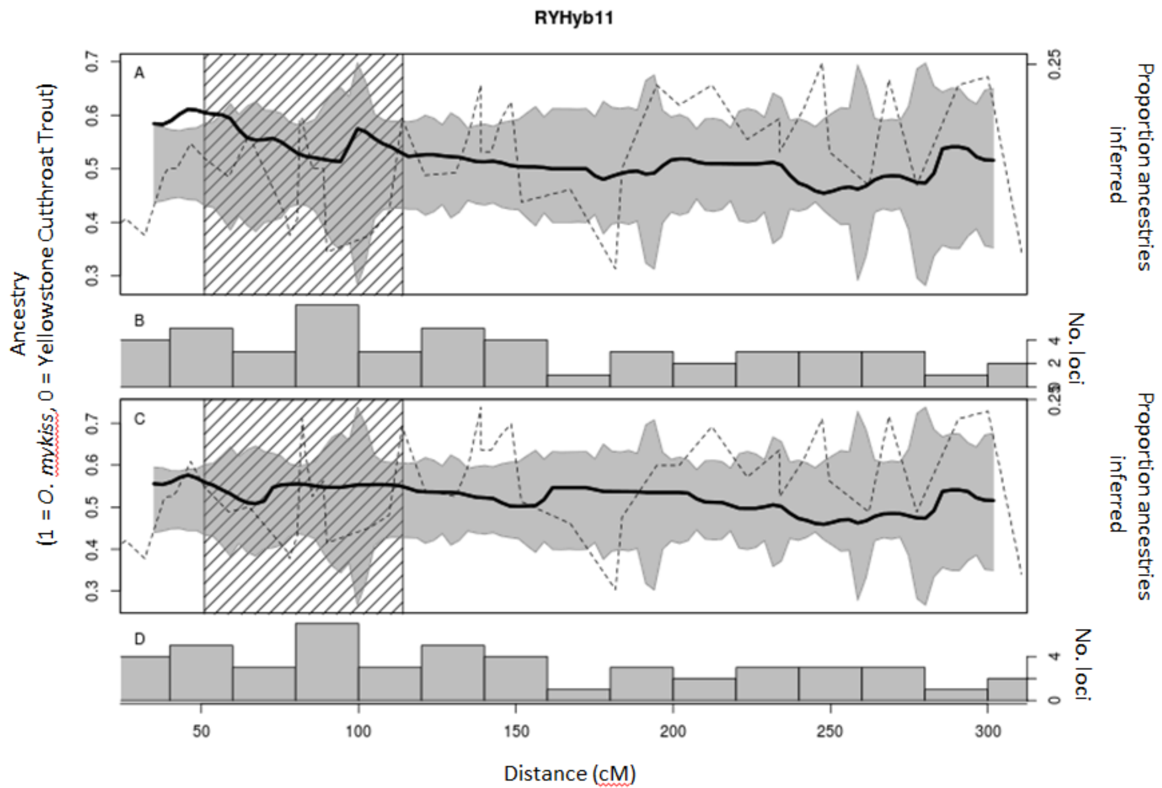


Figure S2.16: Distribution of mean species specific marker ancestries across linkage group RYHyb11, contributed from the dam (A and B) and sire (C and D), as observed in F_2 progeny. In (A) and (C), black lines represent mean ancestry (1 = *O. mykiss*, 0 = Yellowstone cutthroat trout) across the linkage group (left Y axis). Confidence intervals at each region are designated in grey. Confidence intervals were estimated by bootstrapping loci 100,000 times at each position. Ancestries are based on sex specific maps. Some ancestries were inferred (Figure 2.1); the dotted line represents the proportion of ancestries inferred at each locus (right y axis). The dashed box is the centromeric region identified by comparative mapping with other salmonids. Plots B and D are the distribution of markers used to produce A and C respectively.

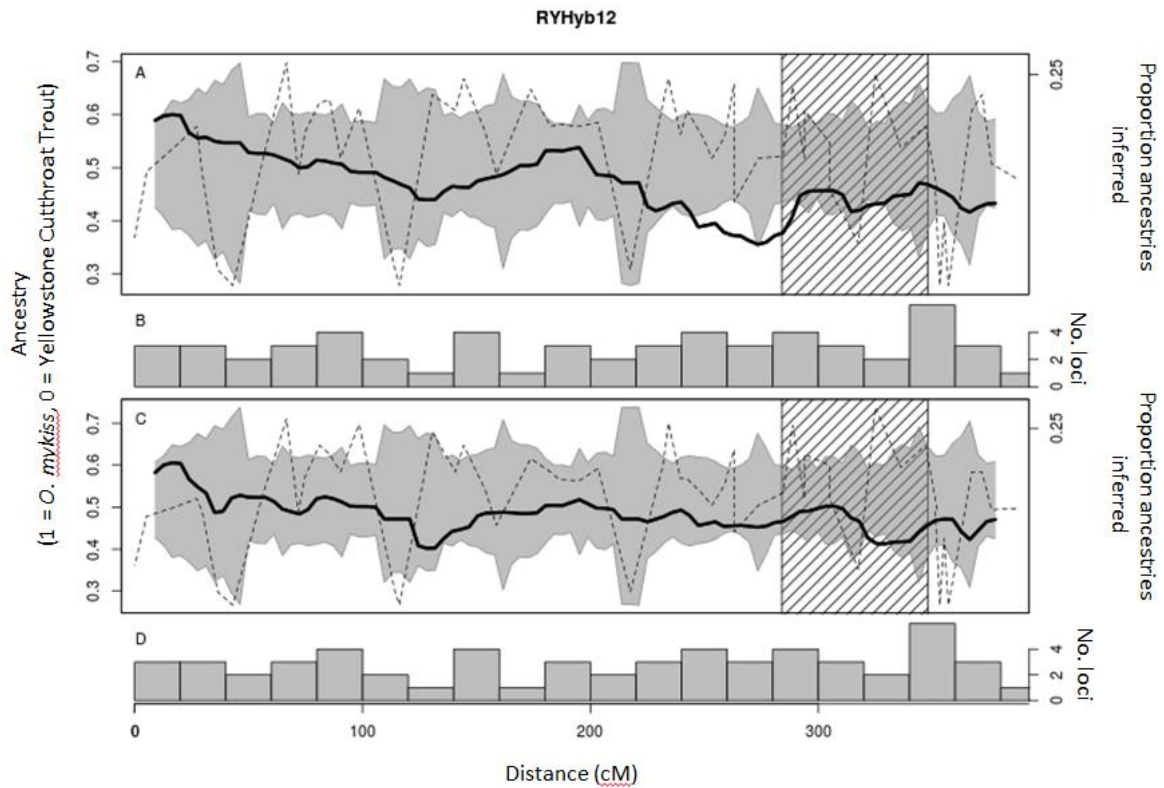


Figure S2.17: Distribution of mean species specific marker ancestries across linkage group RYHyb12, contributed from the dam (A and B) and sire (C and D), as observed in F_2 progeny. In (A) and (C), black lines represent mean ancestry (1 = *O. mykiss*, 0 = Yellowstone cutthroat trout) across the linkage group (left Y axis). Confidence intervals at each region are designated in grey. Confidence intervals were estimated by bootstrapping loci 100,000 times at each position. Ancestries are based on sex specific maps. Some ancestries were inferred (Figure 2.1); the dotted line represents the proportion of ancestries inferred at each locus (right y axis). The dashed box is the centromeric region identified by comparative mapping with other salmonids. Plots B and D are the distribution of markers used to produce A and C respectively.

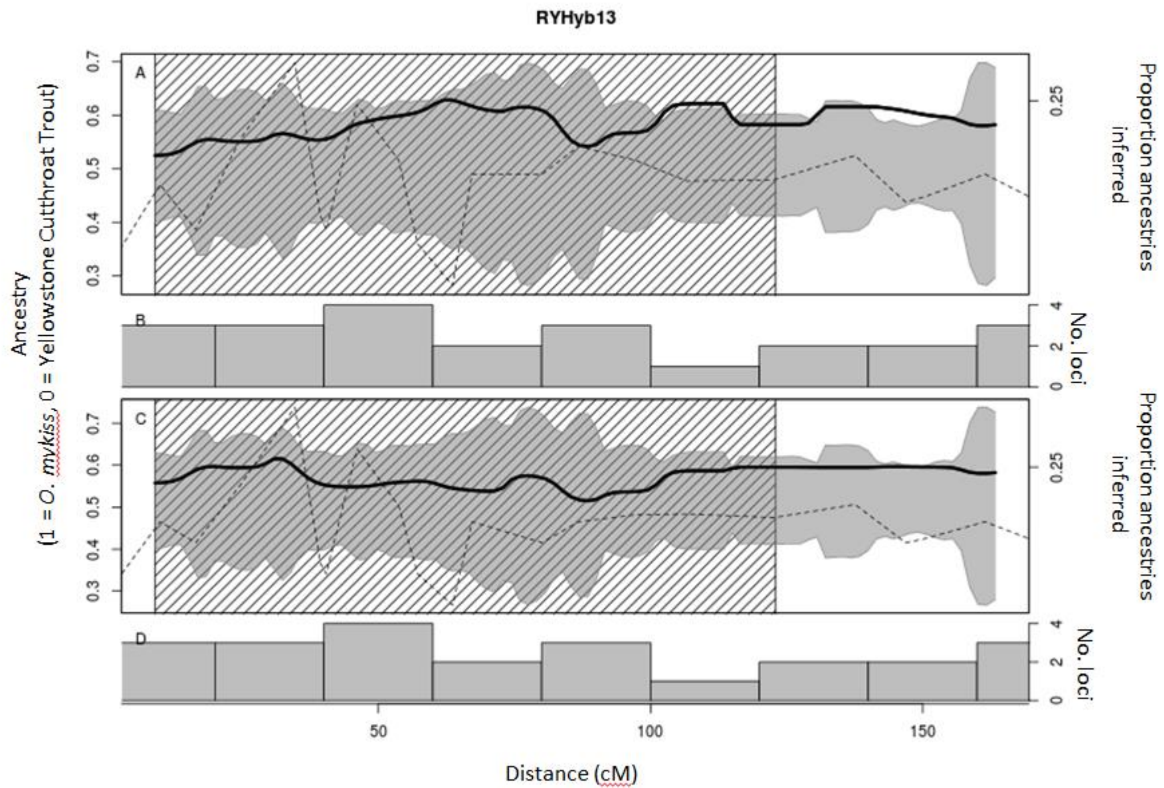


Figure S2.18: Distribution of mean species specific marker ancestries across linkage group RYHyb13, contributed from the dam (A and B) and sire (C and D), as observed in F_2 progeny. In (A) and (C), black lines represent mean ancestry (1 = *O. mykiss*, 0 = Yellowstone cutthroat trout) across the linkage group (left Y axis). Confidence intervals at each region are designated in grey. Confidence intervals were estimated by bootstrapping loci 100,000 times at each position. Ancestries are based on sex specific maps. Some ancestries were inferred (Figure 2.1); the dotted line represents the proportion of ancestries inferred at each locus (right y axis). The dashed box is the centromeric region identified by comparative mapping with other salmonids. Plots B and D are the distribution of markers used to produce A and C respectively.

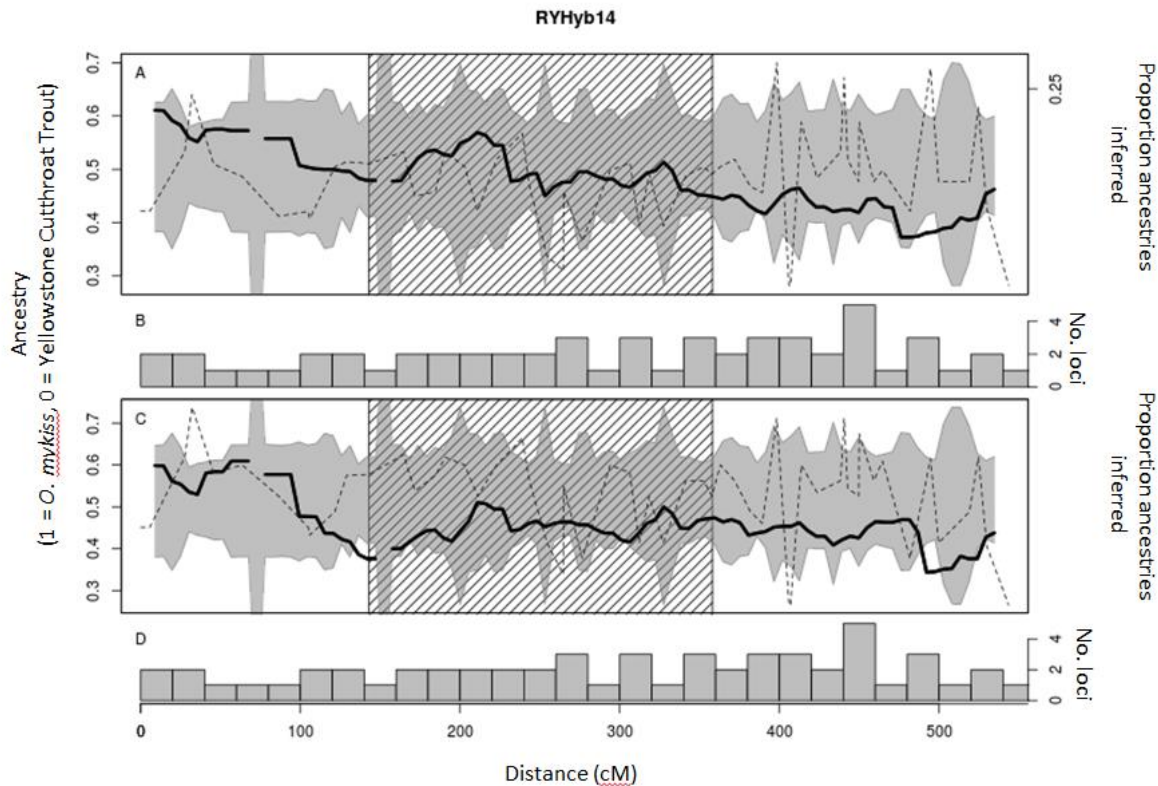


Figure S2.19: Distribution of mean species specific marker ancestries across linkage group RYHyb14, contributed from the dam (A and B) and sire (C and D), as observed in F_2 progeny. In (A) and (C), black lines represent mean ancestry (1 = *O. mykiss*, 0 = Yellowstone cutthroat trout) across the linkage group (left Y axis). Confidence intervals at each region are designated in grey. Confidence intervals were estimated by bootstrapping loci 100,000 times at each position. Ancestries are based on sex specific maps. Some ancestries were inferred (Figure 2.1); the dotted line represents the proportion of ancestries inferred at each locus (right y axis). The dashed box is the centromeric region identified by comparative mapping with other salmonids. Plots B and D are the distribution of markers used to produce A and C respectively.

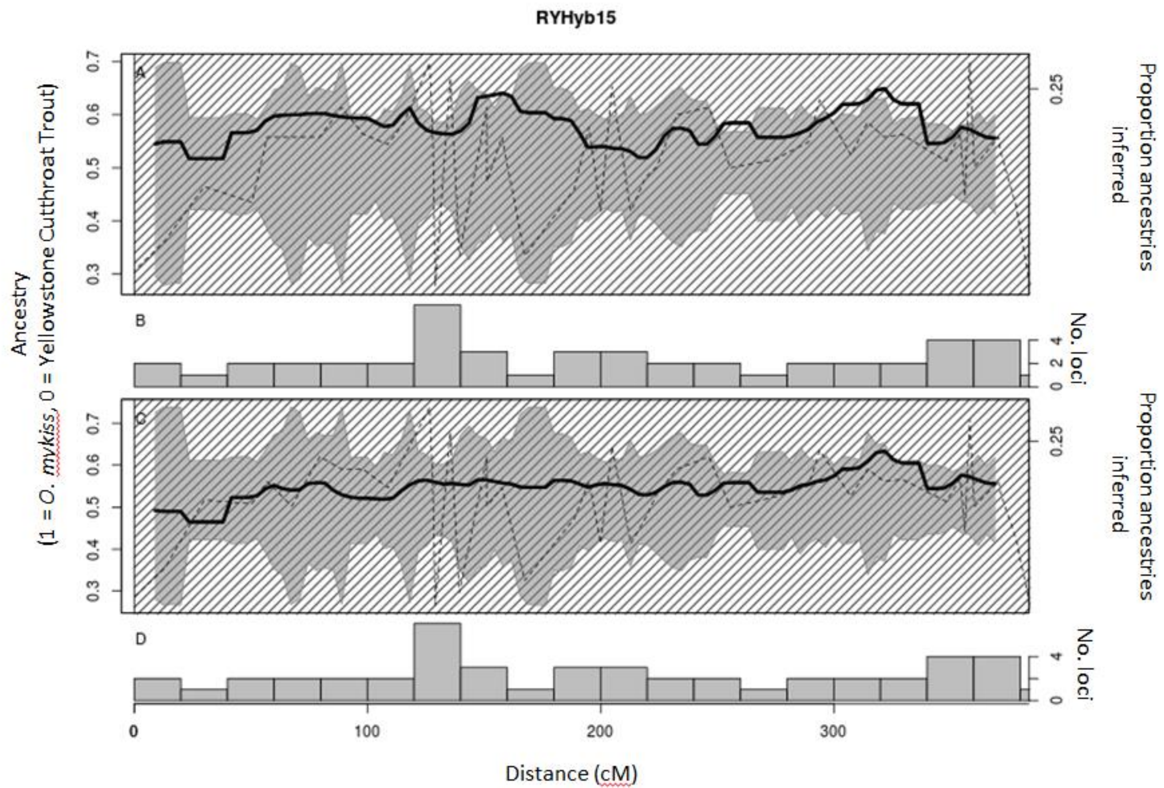


Figure S2.20: Distribution of mean species specific marker ancestries across linkage group RYHyb15, contributed from the dam (A and B) and sire (C and D), as observed in F_2 progeny. In (A) and (C), black lines represent mean ancestry (1 = *O. mykiss*, 0 = Yellowstone cutthroat trout) across the linkage group (left Y axis). Confidence intervals at each region are designated in grey. Confidence intervals were estimated by bootstrapping loci 100,000 times at each position. Ancestries are based on sex specific maps. Some ancestries were inferred (Figure 2.1); the dotted line represents the proportion of ancestries inferred at each locus (right y axis). The dashed box is the centromeric region identified by comparative mapping with other salmonids. Plots B and D are the distribution of markers used to produce A and C respectively.

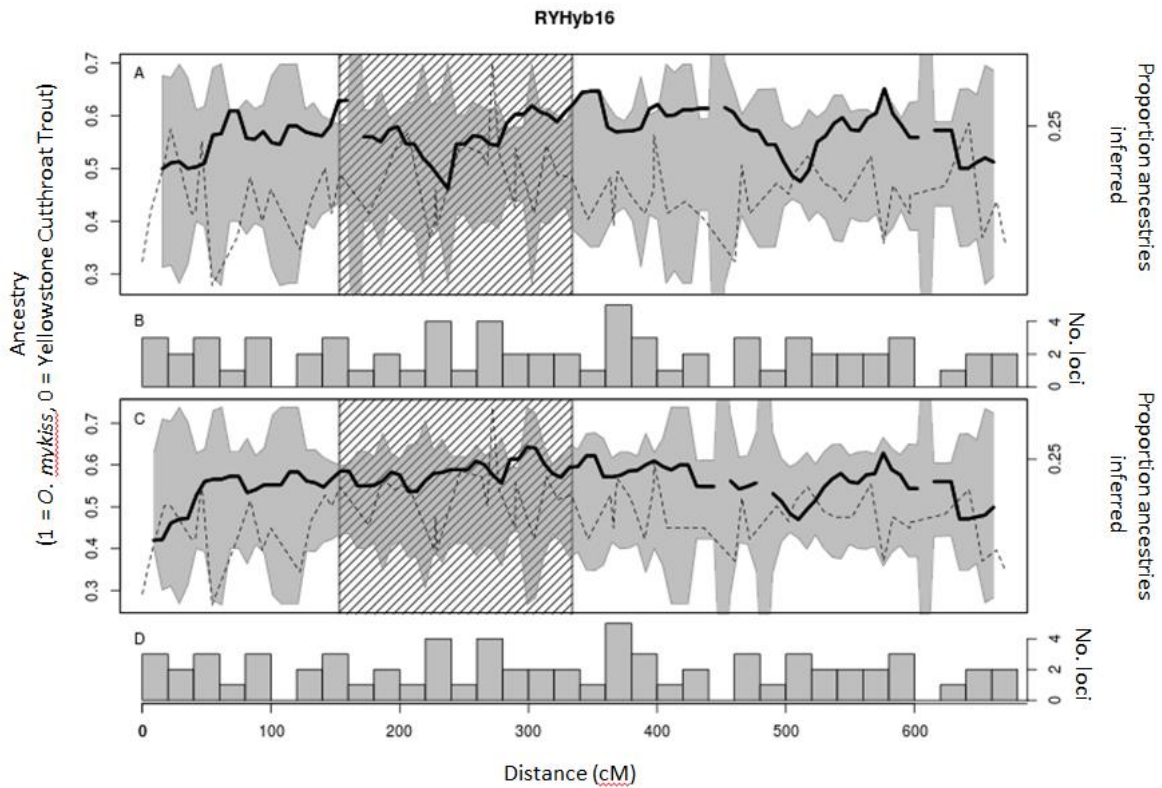


Figure S2.21: Distribution of mean species specific marker ancestries across linkage group RYHyb16, contributed from the dam (A and B) and sire (C and D), as observed in F_2 progeny. In (A) and (C), black lines represent mean ancestry (1 = *O. mykiss*, 0 = Yellowstone cutthroat trout) across the linkage group (left Y axis). Confidence intervals at each region are designated in grey. Confidence intervals were estimated by bootstrapping loci 100,000 times at each position. Ancestries are based on sex specific maps. Some ancestries were inferred (Figure 2.1); the dotted line represents the proportion of ancestries inferred at each locus (right y axis). The dashed box is the centromeric region identified by comparative mapping with other salmonids. Plots B and D are the distribution of markers used to produce A and C respectively.

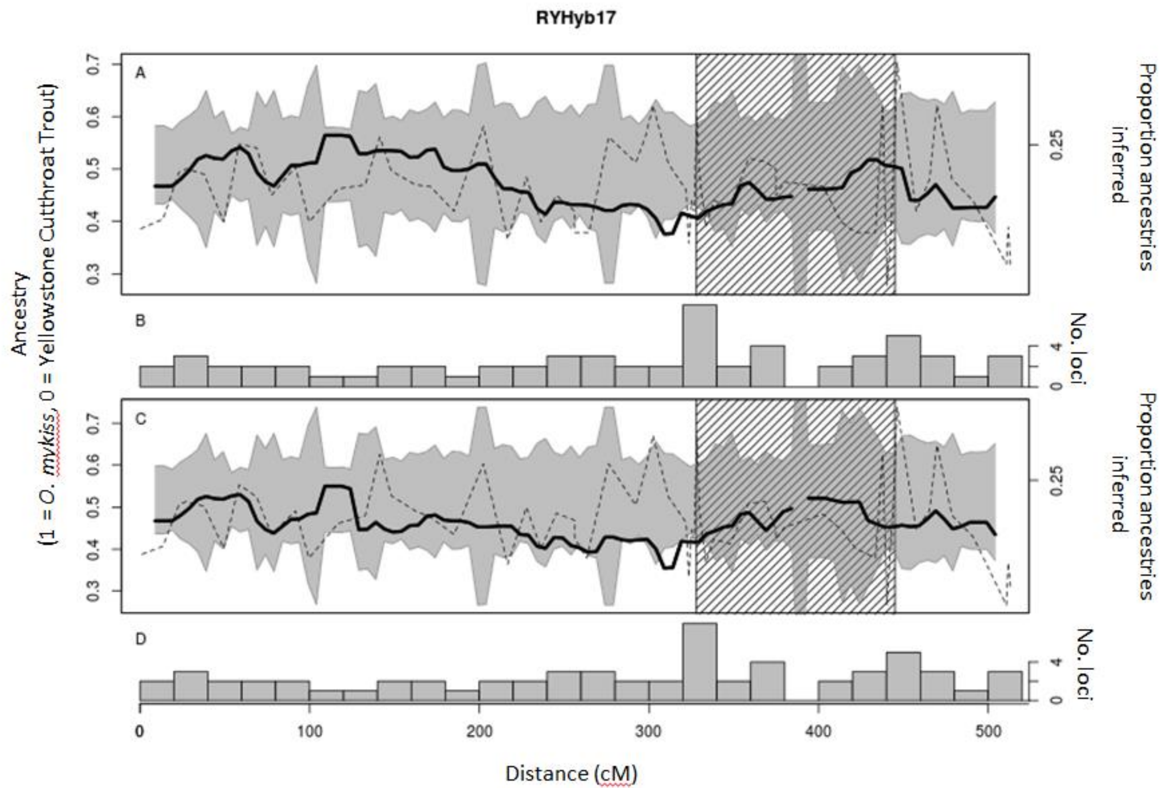


Figure S2.22: Distribution of mean species specific marker ancestries across linkage group RYHyb17, contributed from the dam (A and B) and sire (C and D), as observed in F_2 progeny. In (A) and (C), black lines represent mean ancestry (1 = *O. mykiss*, 0 = Yellowstone cutthroat trout) across the linkage group (left Y axis). Confidence intervals at each region are designated in grey. Confidence intervals were estimated by bootstrapping loci 100,000 times at each position. Ancestries are based on sex specific maps. Some ancestries were inferred (Figure 2.1); the dotted line represents the proportion of ancestries inferred at each locus (right y axis). The dashed box is the centromeric region identified by comparative mapping with other salmonids. Plots B and D are the distribution of markers used to produce A and C respectively.

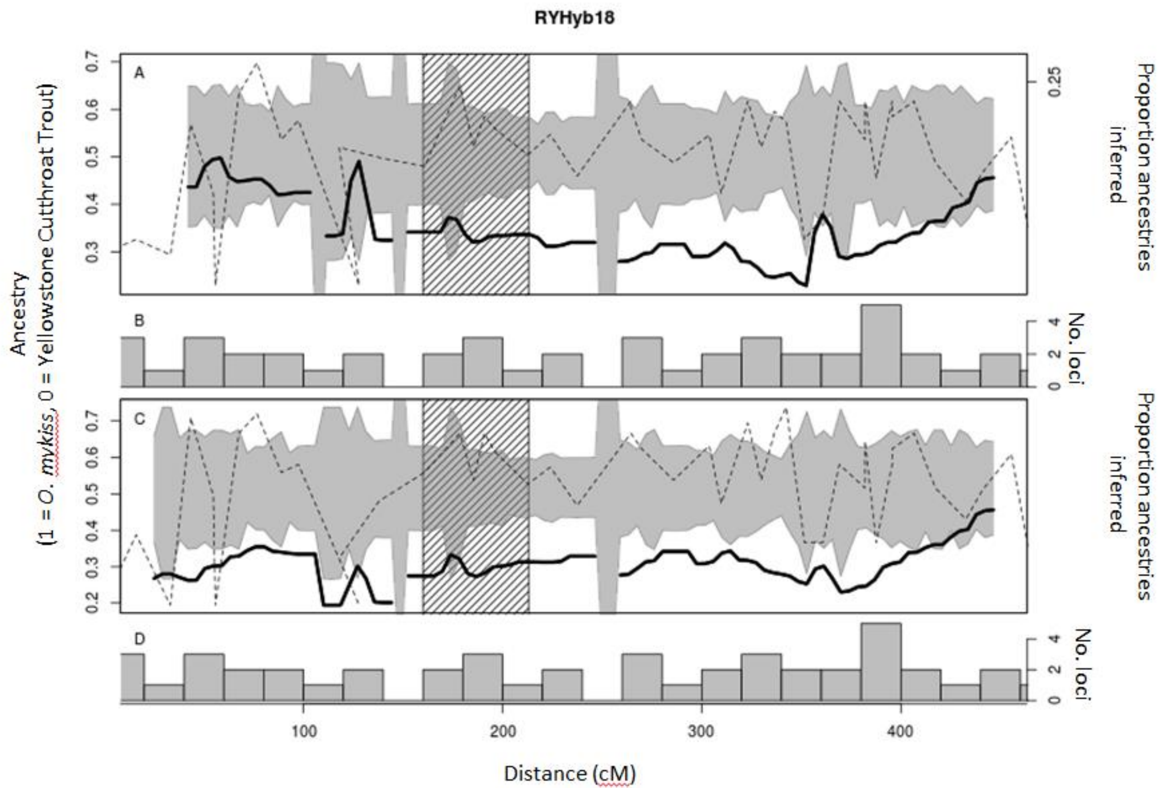


Figure S2.23: Distribution of mean species specific marker ancestries across linkage group RYHyb18, contributed from the dam (A and B) and sire (C and D), as observed in F_2 progeny. In (A) and (C), black lines represent mean ancestry (1 = *O. mykiss*, 0 = Yellowstone cutthroat trout) across the linkage group (left Y axis). Confidence intervals at each region are designated in grey. Confidence intervals were estimated by bootstrapping loci 100,000 times at each position. Ancestries are based on sex specific maps. Some ancestries were inferred (Figure 2.1); the dotted line represents the proportion of ancestries inferred at each locus (right y axis). The dashed box is the centromeric region identified by comparative mapping with other salmonids. Plots B and D are the distribution of markers used to produce A and C respectively.

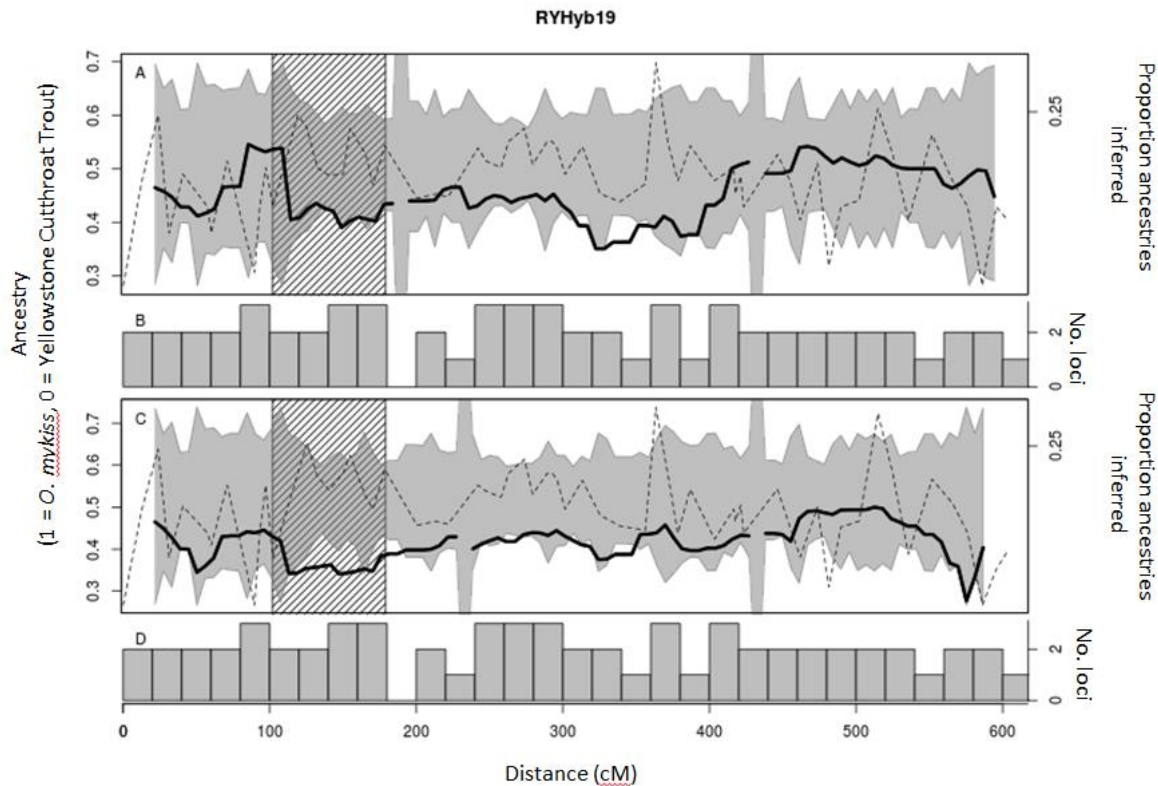


Figure S2.24: Distribution of mean species specific marker ancestries across linkage group RYHyb19, contributed from the dam (A and B) and sire (C and D), as observed in F_2 progeny. In (A) and (C), black lines represent mean ancestry (1 = *O. mykiss*, 0 = Yellowstone cutthroat trout) across the linkage group (left Y axis). Confidence intervals at each region are designated in grey. Confidence intervals were estimated by bootstrapping loci 100,000 times at each position. Ancestries are based on sex specific maps. Some ancestries were inferred (Figure 2.1); the dotted line represents the proportion of ancestries inferred at each locus (right y axis). The dashed box is the centromeric region identified by comparative mapping with other salmonids. Plots B and D are the distribution of markers used to produce A and C respectively.

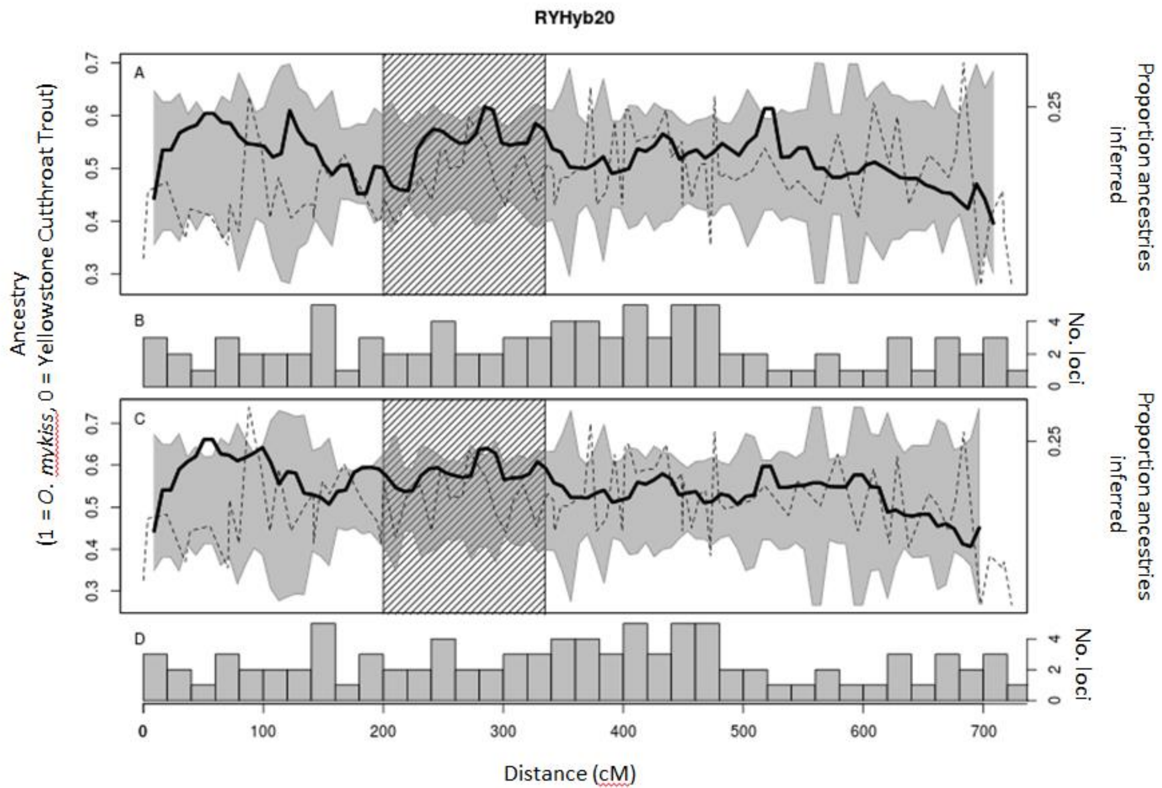


Figure S2.25: Distribution of mean species specific marker ancestries across linkage group RYHyb20, contributed from the dam (A and B) and sire (C and D), as observed in F_2 progeny. In (A) and (C), black lines represent mean ancestry (1 = *O. mykiss*, 0 = Yellowstone cutthroat trout) across the linkage group (left Y axis). Confidence intervals at each region are designated in grey. Confidence intervals were estimated by bootstrapping loci 100,000 times at each position. Ancestries are based on sex specific maps. Some ancestries were inferred (Figure 2.1); the dotted line represents the proportion of ancestries inferred at each locus (right y axis). The dashed box is the centromeric region identified by comparative mapping with other salmonids. Plots B and D are the distribution of markers used to produce A and C respectively.

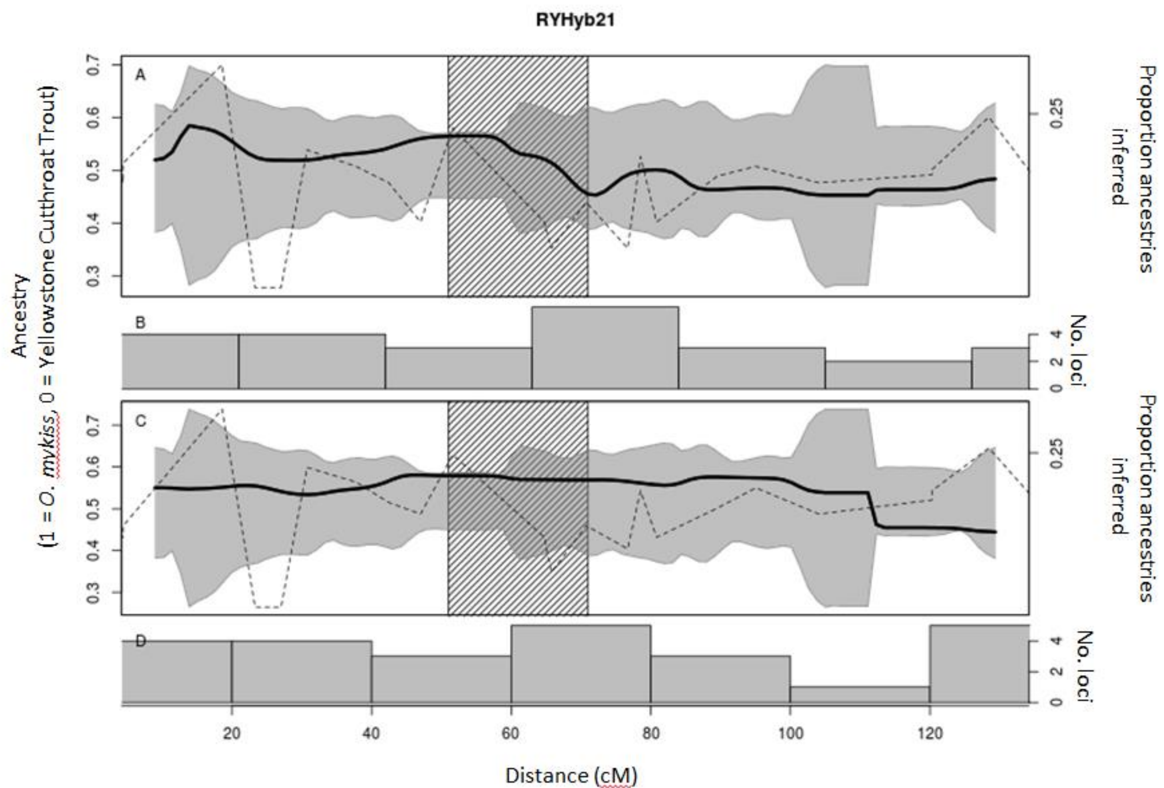


Figure S2.26: Distribution of mean species specific marker ancestries across linkage group RYHyb21, contributed from the dam (A and B) and sire (C and D), as observed in F_2 progeny. In (A) and (C), black lines represent mean ancestry (1 = *O. mykiss*, 0 = Yellowstone cutthroat trout) across the linkage group (left Y axis). Confidence intervals at each region are designated in grey. Confidence intervals were estimated by bootstrapping loci 100,000 times at each position. Ancestries are based on sex specific maps. Some ancestries were inferred (Figure 2.1); the dotted line represents the proportion of ancestries inferred at each locus (right y axis). The dashed box is the centromeric region identified by comparative mapping with other salmonids. Plots B and D are the distribution of markers used to produce A and C respectively.

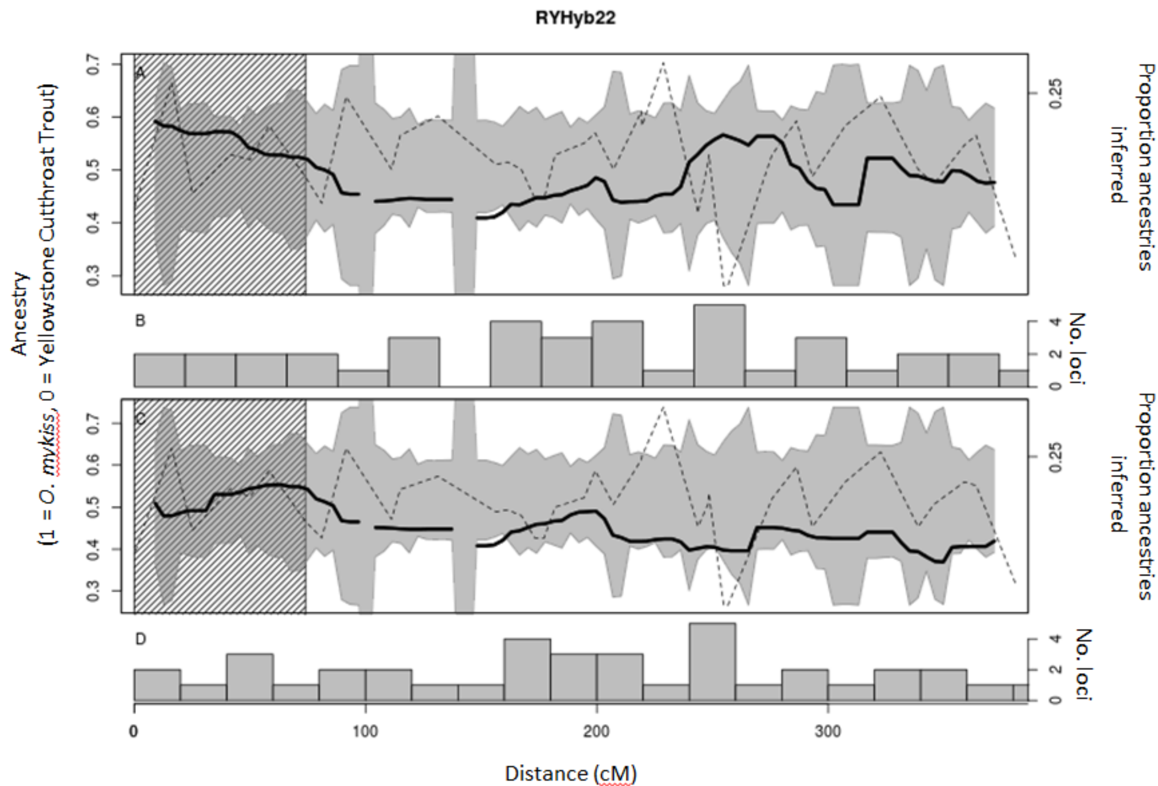


Figure S2.27: Distribution of mean species specific marker ancestries across linkage group RYHyb22, contributed from the dam (A and B) and sire (C and D), as observed in F_2 progeny. In (A) and (C), black lines represent mean ancestry (1 = *O. mykiss*, 0 = Yellowstone cutthroat trout) across the linkage group (left Y axis). Confidence intervals at each region are designated in grey. Confidence intervals were estimated by bootstrapping loci 100,000 times at each position. Ancestries are based on sex specific maps. Some ancestries were inferred (Figure 2.1); the dotted line represents the proportion of ancestries inferred at each locus (right y axis). The dashed box is the centromeric region identified by comparative mapping with other salmonids. Plots B and D are the distribution of markers used to produce A and C respectively.

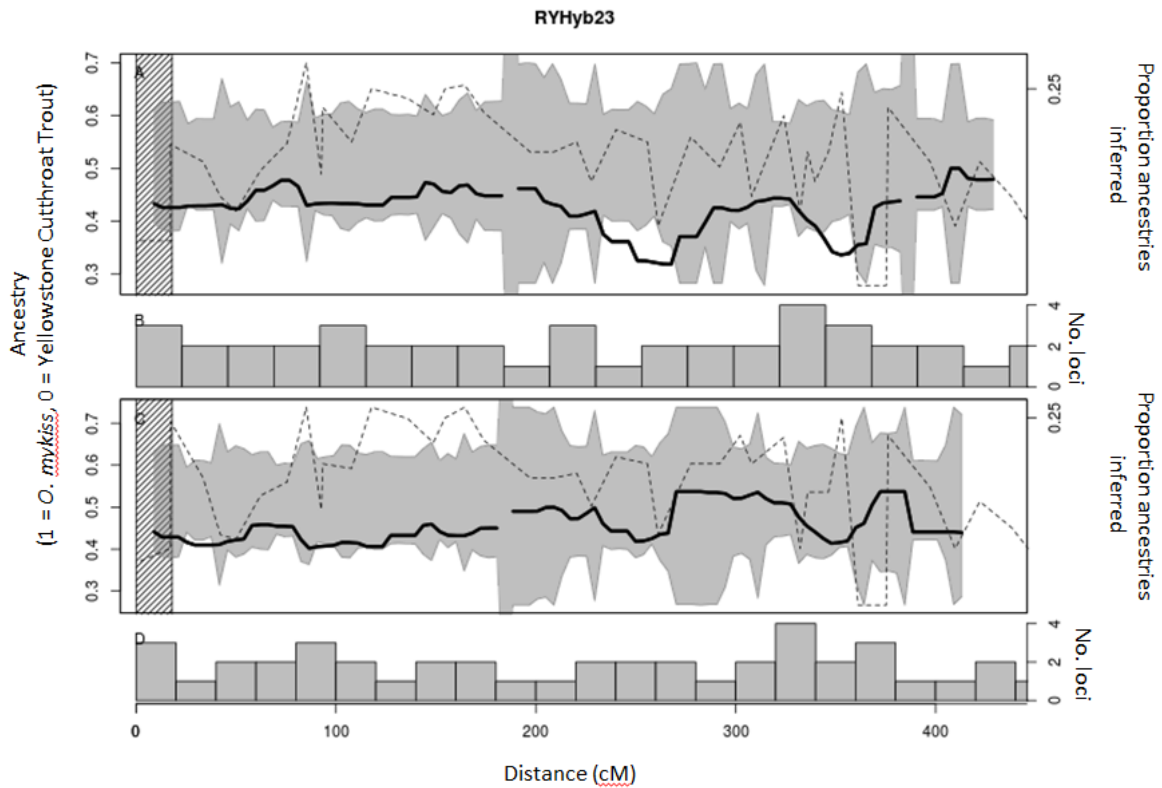


Figure S2.28: Distribution of mean species specific marker ancestries across linkage group RYHyb23, contributed from the dam (A and B) and sire (C and D), as observed in F_2 progeny. In (A) and (C), black lines represent mean ancestry (1 = *O. mykiss*, 0 = Yellowstone cutthroat trout) across the linkage group (left Y axis). Confidence intervals at each region are designated in grey. Confidence intervals were estimated by bootstrapping loci 100,000 times at each position. Ancestries are based on sex specific maps. Some ancestries were inferred (Figure 2.1); the dotted line represents the proportion of ancestries inferred at each locus (right y axis). The dashed box is the centromeric region identified by comparative mapping with other salmonids. Plots B and D are the distribution of markers used to produce A and C respectively.

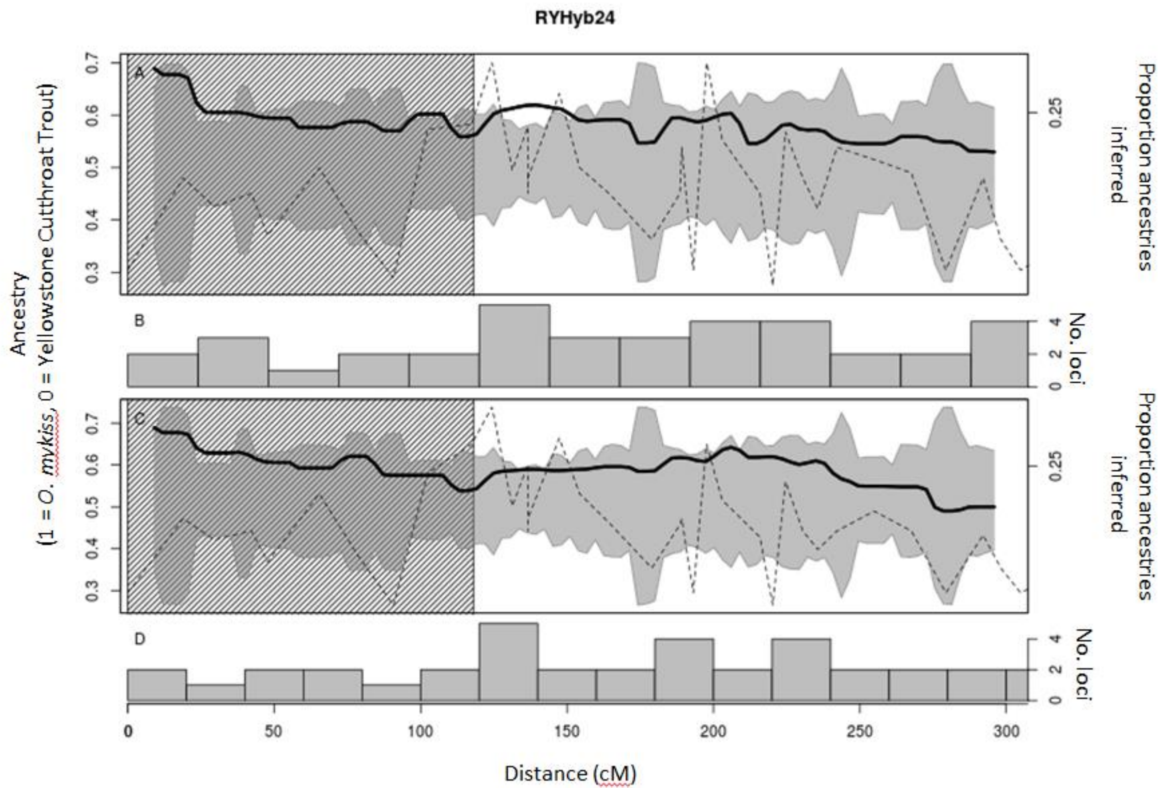


Figure S2.29: Distribution of mean species specific marker ancestries across linkage group RYHyb24, contributed from the dam (A and B) and sire (C and D), as observed in F_2 progeny. In (A) and (C), black lines represent mean ancestry (1 = *O. mykiss*, 0 = Yellowstone cutthroat trout) across the linkage group (left Y axis). Confidence intervals at each region are designated in grey. Confidence intervals were estimated by bootstrapping loci 100,000 times at each position. Ancestries are based on sex specific maps. Some ancestries were inferred (Figure 2.1); the dotted line represents the proportion of ancestries inferred at each locus (right y axis). The dashed box is the centromeric region identified by comparative mapping with other salmonids. Plots B and D are the distribution of markers used to produce A and C respectively.

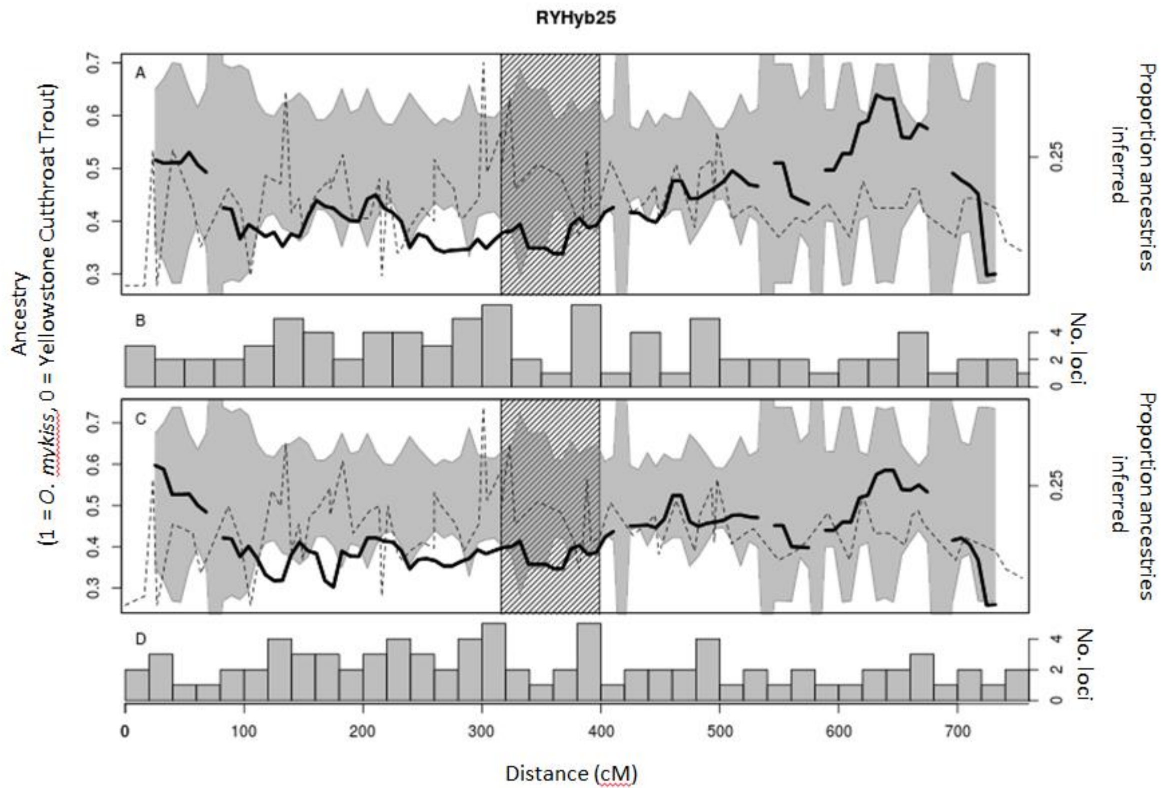


Figure S2.30: Distribution of mean species specific marker ancestries across linkage group RYHyb25, contributed from the dam (A and B) and sire (C and D), as observed in F_2 progeny. In (A) and (C), black lines represent mean ancestry (1 = *O. mykiss*, 0 = Yellowstone cutthroat trout) across the linkage group (left Y axis). Confidence intervals at each region are designated in grey. Confidence intervals were estimated by bootstrapping loci 100,000 times at each position. Ancestries are based on sex specific maps. Some ancestries were inferred (Figure 2.1); the dotted line represents the proportion of ancestries inferred at each locus (right y axis). The dashed box is the centromeric region identified by comparative mapping with other salmonids. Plots B and D are the distribution of markers used to produce A and C respectively.

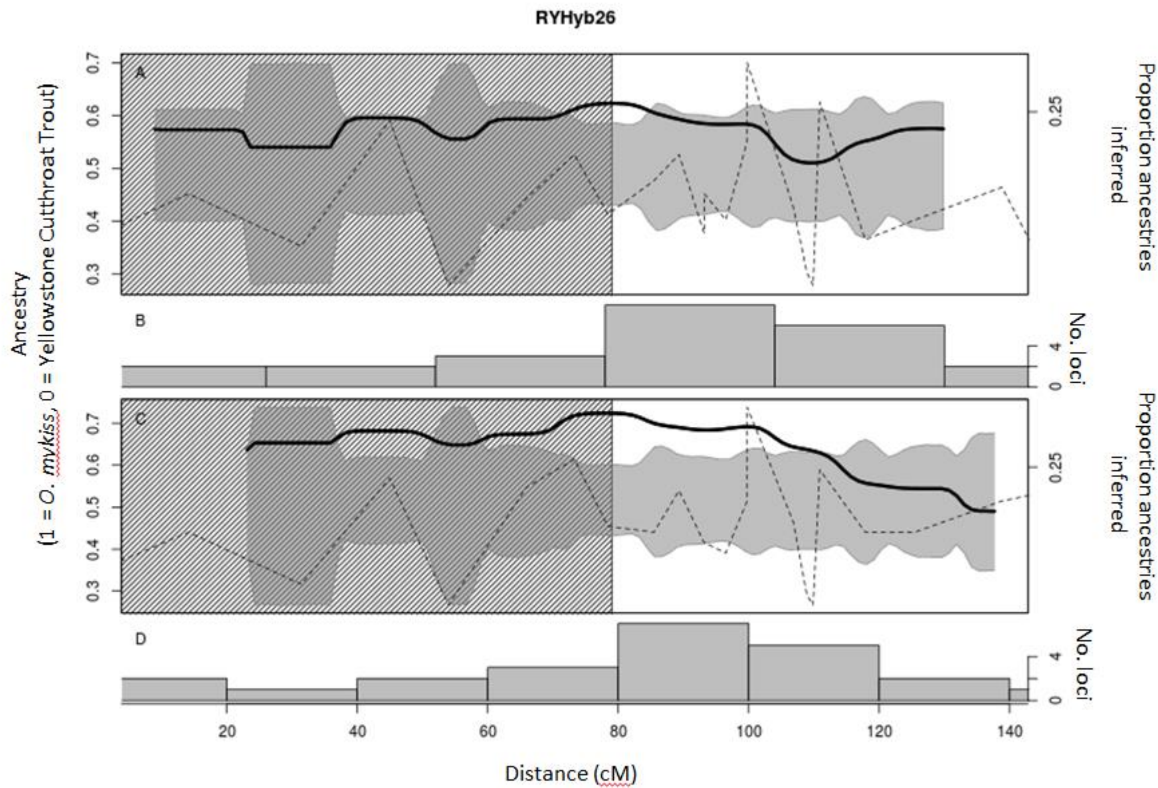


Figure S2.31: Distribution of mean species specific marker ancestries across linkage group RYHyb26, contributed from the dam (A and B) and sire (C and D), as observed in F_2 progeny. In (A) and (C), black lines represent mean ancestry (1 = *O. mykiss*, 0 = Yellowstone cutthroat trout) across the linkage group (left Y axis). Confidence intervals at each region are designated in grey. Confidence intervals were estimated by bootstrapping loci 100,000 times at each position. Ancestries are based on sex specific maps. Some ancestries were inferred (Figure 2.1); the dotted line represents the proportion of ancestries inferred at each locus (right y axis). The dashed box is the centromeric region identified by comparative mapping with other salmonids. Plots B and D are the distribution of markers used to produce A and C respectively.

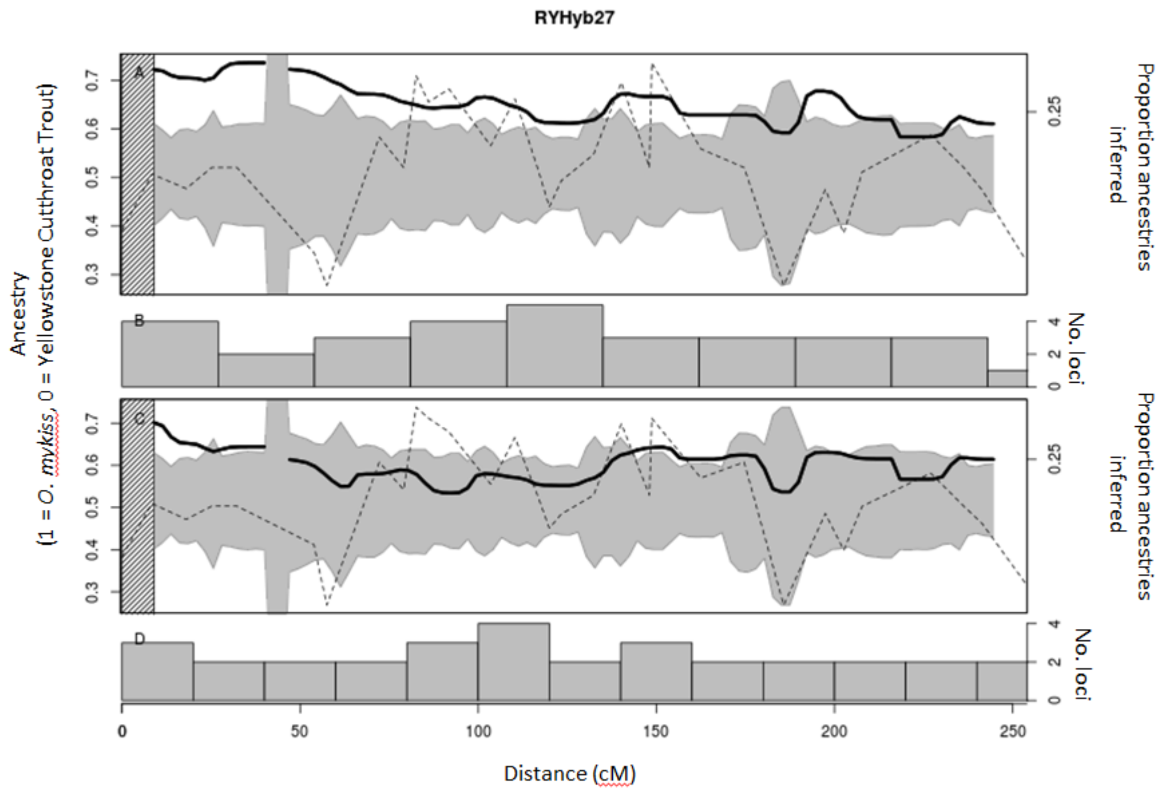


Figure S2.32: Distribution of mean species specific marker ancestries across linkage group RYHyb27, contributed from the dam (A and B) and sire (C and D), as observed in F_2 progeny. In (A) and (C), black lines represent mean ancestry (1 = *O. mykiss*, 0 = Yellowstone cutthroat trout) across the linkage group (left Y axis). Confidence intervals at each region are designated in grey. Confidence intervals were estimated by bootstrapping loci 100,000 times at each position. Ancestries are based on sex specific maps. Some ancestries were inferred (Figure 2.1); the dotted line represents the proportion of ancestries inferred at each locus (right y axis). The dashed box is the centromeric region identified by comparative mapping with other salmonids. Plots B and D are the distribution of markers used to produce A and C respectively.

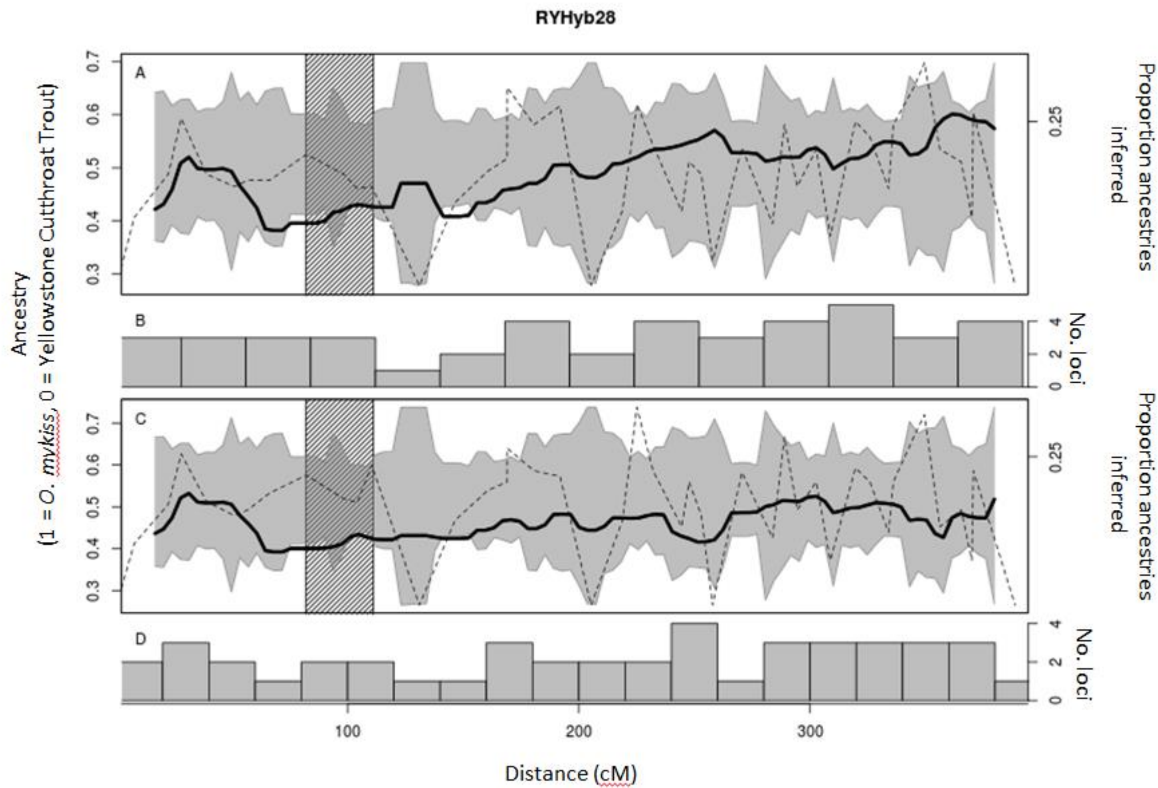


Figure S2.33: Distribution of mean species specific marker ancestries across linkage group RYHyb28, contributed from the dam (A and B) and sire (C and D), as observed in F_2 progeny. In (A) and (C), black lines represent mean ancestry (1 = *O. mykiss*, 0 = Yellowstone cutthroat trout) across the linkage group (left Y axis). Confidence intervals at each region are designated in grey. Confidence intervals were estimated by bootstrapping loci 100,000 times at each position. Ancestries are based on sex specific maps. Some ancestries were inferred (Figure 2.1); the dotted line represents the proportion of ancestries inferred at each locus (right y axis). The dashed box is the centromeric region identified by comparative mapping with other salmonids. Plots B and D are the distribution of markers used to produce A and C respectively.

Supplemental materials

Identification of duplicated loci

We used an R script and simulated genotypes to estimate how well we could 1) differentiate duplicated and non-duplicated markers and 2) differentiation duplicated marker types using a read depth of 5 sequences per locus and 48 progeny from a haploid family.

To achieve these objectives, we first simulated genotypes for 48 progeny at 1000 loci of each locus type. Locus types were:

Haploid: AB x 00,

Duplicated1: AA x BB,

Duplicated2: AA x AB,

Duplicated3: AB x AB,

Duplicated4: AA x BC,

Duplicated5: AB x BC, and

Duplicated6: AB x CD.

For haploid loci, one allele (A or B) was randomly drawn for each progeny. For duplicated markers, one allele was randomly drawn from each of the loci contributing to the duplicated locus. For example, for a Duplicated6 locus, A or B and C or D were randomly drawn, resulting in a genotype of AC, AD, BC, or BD. All alleles at a locus were equally likely to be drawn.

After simulating genotypes, genotyping error was introduced by replacing 5% of alleles randomly with an A, B, C, or D allele.

We then calculated log-likelihood scores for each locus type given the observed genetic data at each locus. Calculations were made as follows.

$$LL(hap|genotypes) = \log\left(C\ \left(0.5\ (1-err) + \frac{err}{10}\right)^{n_{aa}+n_{bb}}\ \frac{err^{n_{total}-n_{aa}-n_{bb}}}{10}\right)$$

$$LL(dup1|genotypes) = \log\left(C\ \left(1\ (1-err) + \frac{err}{10}\right)^{n_{ab}}\ \frac{err^{n_{total}-n_{ab}}}{10}\right)$$

$$LL(dup2|genotypes) = \log\left(C\ \left(0.5\ (1-err) + \frac{err}{10}\right)^{n_{ab}+n_{aa}}\ \frac{err^{n_{total}-n_{ab}-n_{aa}}}{10}\right)$$

$$LL(dup3|genotypes) = \log\left(C\ \left(0.25\ (1-err) + \frac{err}{10}\right)^{n_{aa}+n_{bb}} \left(0.5\ (1-err) + \frac{err}{10}\right)^{n_{ab}}\ \frac{err^{n_{total}-n_{ab}-n_{aa}-n_{bb}}}{10}\right)$$

$$LL(dup4|genotypes) = \log\left(C\ \left(0.5\ (1-err) + \frac{err}{10}\right)^{n_{ab}+n_{ac}}\ \frac{err^{n_{total}-n_{ab}-n_{ac}}}{10}\right)$$

$$LL(dup5|genotypes) = \log\left(C\ \left(0.25\ (1-err) + \frac{err}{10}\right)^{n_{aa}+n_{ab}+n_{ac}+n_{bc}}\ \frac{err^{n_{total}-n_{aa}-n_{ab}-n_{ac}-n_{bc}}}{10}\right)$$

$$LL(dup6|genotypes) = \log\left(C\ \left(0.25\,(1-err)+\frac{err}{10}\right)^{n_{ab}+n_{ad}+n_{bc}+n_{cd}}\ \frac{err}{10}^{n_{total}-n_{ab}-n_{ad}-n_{bc}-n_{cd}}\right)$$

$$C = \frac{n_{total}!}{n_{aa}!\,n_{bb}!\,n_{cc}!\,n_{dd}!\,n_{ab}!\,n_{ac}!\,n_{ad}!\,n_{bc}!\,n_{bd}!\,n_{cd}!}$$

$$err = 0.05$$

Likelihoods were compared and a locus type was assigned if the most likely locus type had a log-likelihood of at least 4 greater than the second most likely locus type.

Results

True locus type: haploid (AB x 00)	
Haploid (AB x 00) 911 (98.4% of called loci) Dup1 (AAxBB) 0 (0% of called loci) Dup2 (AAxAB) 0 (0% of called loci) Dup3 (ABxAB) 15 (1.6% of called loci)	Number of loci identified as haploid: 911 (100% of called loci) Number of loci identified as duplicated: 15 (0% of called loci)

Dup4 (AAxBC) 0 (0% of called loci) Dup5 (ABxAC) 0 (0% of called loci) Dup6 (ABxCD) 0 (0% of called loci) Loci not called 74	

True locus type: dup1 (AA x BB) Haploid (AB x 00) 0 (0% of called loci) Dup1 (AAxBB) 943 (99.5% of called loci) Dup2 (AAxAB) 0 (0% of called loci) Dup3 (ABxAB) 5 (0.5% of called loci) Dup4 (AAxBC) 0 (0% of called loci) Dup5 (ABxAC) 0 (0% of called loci) Dup6 (ABxCD) 0 (0% of called loci) Loci not called 52	Number of loci identified as haploid: 0 (0% of called loci) Number of loci identified as duplicated: 948 (100% of called loci)

True locus type: dup2 (AA x AB)		
Haploid (AB x 00)	0 (0% of called loci)	Number of loci identified as haploid: 0 (0% of called loci)
Dup1 (AAxBB)	0 (0% of called loci)	Number of loci identified as duplicated: 932 (100% of called loci)
Dup2 (AAxAB)	930 (99.8% of called loci)	
Dup3 (ABxAB)	1(0.1% of called loci)	
Dup4 (AAxBC)	0 (0% of called loci)	
Dup5 (ABxAC)	1(0.1% of called loci)	
Dup6 (ABxCD)	0 (0% of called loci)	
Loci not called	68	

True locus type: dup3 (AB x AB)		
Haploid (AB x 00)	0 (0% of called loci)	Number of loci identified as haploid: 0 (0% of called loci)
Dup1 (AAxBB)	0 (0% of called loci)	Number of loci identified as duplicated: 999 (100% of called loci)
Dup2 (AAxAB)	0 (0% of	

called loci) Dup3 (ABxAB) 999 (100% of called loci) Dup4 (AAxBC) 0 (0% of called loci) Dup5 (ABxAC) 0 (0% of called loci) Dup6 (ABxCD) 0 (0% of called loci) Loci not called 1	

<u>True locus type: dup4 (AA x BC)</u> Haploid (AB x 00) 0 (0% of called loci) Dup1 (AAxBB) 0 (0% of called loci) Dup2 (AAxAB) 0 (0% of called loci) Dup3 (ABxAB) 0 (0% of called loci) Dup4 (AAxBC) 995 (100% of called loci) Dup5 (ABxAC) 0 (0% of called loci)	Number of loci identified as haploid: 0 (0% of called loci) Number of loci identified as duplicated: 995 (100% of called loci)

Dup6 (ABxCD) 0 (0% of called loci) Loci not called 5	

True locus type: dup5 (AB x AC)	
Haploid (AB x 00) 0 (0% of called loci) Dup1 (AAxBB) 0 (0% of called loci) Dup2 (AAxAB) 0 (0% of called loci) Dup3 (ABxAB) 0 (0% of called loci) Dup4 (AAxBC) 0 (0% of called loci) Dup5 (ABxAC) 1000 (100% of called loci) Dup6 (ABxCD) 0 (0% of called loci) Loci not called 0	Number of loci identified as haploid: 0 (0% of called loci) Number of loci identified as duplicated: 1000 (100% of called loci)

True locus type: dup6 (AB x CD)	
Haploid (AB x 00) 0 (0% of called loci) Dup1 (AAxBB) 0 (0% of called	Number of loci identified as haploid: 0 (0% of called loci)

loci) Dup2 (AAxAB) 0 (0% of called loci) Dup3 (ABxAB) 0 (0% of called loci) Dup4 (AAxBC) 0 (0% of called loci) Dup5 (ABxAC) 447 (60.2% of called loci) Dup6 (ABxCD) 296 (39.8% of called loci) Loci not called 257	Number of loci identified as duplicated: 743 (100% of called loci)

R script

```
# this script will look to see how well we do at assigning the correct duplication type with
# only a depth of 5 reads and 48 individuals

set.seed(100)

sim <- function(locus1, locus2, err, depth, haploid){
  hap <- 0; dup1 <- 0; dup2 <- 0; dup3 <- 0; dup4 <- 0; dup5 <- 0; dup6 <- 0; unk <- 0
```

```
# run this simulation 1000 times
for(k in seq(1,1000)){
  aa <- 0; ab <- 0; ac <- 0; ad <- 0; bb <- 0; bc <- 0; bd <- 0; cc <- 0; cd <- 0; dd <- 0

  # run 48 individuals
  for(i in seq(1,48)){
    gam1 <- locus1[round(runif(1, min=1, max=2))]
    gam2 <- locus2[round(runif(1, min=1, max=2))]
    tmp_reads <- c()

    for(j in seq(1,depth)){
      tmp_reads[j] <- gam1
      if (runif(1)>0.5 & haploid > 1){tmp_reads[j] <- gam2} # when running the haploid version,
convert half the genotypes to "a/b"

      # introduce error
      if (runif(1) < err){tmp_reads[j] <- c('a', 'b', 'c', 'd')[round(runif(1,min=1, max=4))]}
    }

    if (length(unique(tmp_reads))<3){
      uniq_reads <- unique(tmp_reads)
      if (length(uniq_reads) == 1){uniq_reads[2] <- uniq_reads[1]}
```

```
  if (uniq_reads[1] == 'a' & uniq_reads[2] == 'a'){aa <- aa + 1}
  if (uniq_reads[1] == 'b' & uniq_reads[2] == 'b'){bb <- bb + 1}
  if (uniq_reads[1] == 'c' & uniq_reads[2] == 'c'){cc <- cc + 1}
  if (uniq_reads[1] == 'd' & uniq_reads[2] == 'd'){dd <- dd + 1}

  if (uniq_reads[1] == 'a' & uniq_reads[2] == 'b'){ab <- ab + 1}
  if (uniq_reads[1] == 'b' & uniq_reads[2] == 'a'){ab <- ab + 1}

  if (uniq_reads[1] == 'a' & uniq_reads[2] == 'c'){ac <- ac + 1}
  if (uniq_reads[1] == 'c' & uniq_reads[2] == 'a'){ac <- ac + 1}

  if (uniq_reads[1] == 'a' & uniq_reads[2] == 'd'){ad <- ad + 1}
  if (uniq_reads[1] == 'd' & uniq_reads[2] == 'a'){ad <- ad + 1}

  if (uniq_reads[1] == 'b' & uniq_reads[2] == 'c'){bc <- bc + 1}
  if (uniq_reads[1] == 'c' & uniq_reads[2] == 'b'){bc <- bc + 1}

  if (uniq_reads[1] == 'b' & uniq_reads[2] == 'd'){bd <- bd + 1}
  if (uniq_reads[1] == 'd' & uniq_reads[2] == 'b'){bd <- bd + 1}

  if (uniq_reads[1] == 'c' & uniq_reads[2] == 'd'){cd <- cd + 1}
  if (uniq_reads[1] == 'd' & uniq_reads[2] == 'c'){cd <- cd + 1}
}
```

```
    }

    # calculate likelihoods
    total <- aa + bb + cc + dd + ab + ac + ad + bc + bd + cd
    #constant <-
factorial(total)/(factorial(aa)*factorial(bb)*factorial(cc)*factorial(dd)*factorial(ab)*factorial(ac)*
factorial(ad)*factorial(bc)*factorial(bd)*factorial(cd))
    constant <- 10
    lik_hap  <- log(constant * (0.50 * (1-err)+err/10)^(aa+bb)       * (err/10)^(total-aa-bb))
    lik_dup1 <- log(constant * (1.00 * (1-err)+err/10)^(ab)          * (err/10)^(total-ab))
    lik_dup2 <- log(constant * (0.50 * (1-err)+err/10)^(ab+aa)       * (err/10)^(total-ab-aa))
    lik_dup3 <- log(constant * (0.25 * (1-err)+err/10)^(aa+bb) * (0.5 * (1-err)+err/10)^(ab) *
(err/10)^(total-aa-bb-ab))
    lik_dup4 <- log(constant * (0.50 * (1-err)+err/10)^(ab+ac)       * (err/10)^(total-ab-ac))
    lik_dup5 <- log(constant * (0.25 * (1-err)+err/10)^(aa+ab+ac+bc) * (err/10)^(total-aa-ab-ac-
bc))
    lik_dup6 <- log(constant * (0.25 * (1-err)+err/10)^(ab+ad+bc+cd) * (err/10)^(total-ab-ad-bc-
cd))

    scores <- sort(c(lik_hap, lik_dup1, lik_dup2, lik_dup3, lik_dup4, lik_dup5, lik_dup6),
decreasing=T)

    if ((scores[1] - scores[2]) > 4){
```

```
      if (lik_hap == scores[1]){hap <- hap + 1}
      if (lik_dup1 == scores[1]){dup1 <- dup1 + 1}
      if (lik_dup2 == scores[1]){dup2 <- dup2 + 1}
      if (lik_dup3 == scores[1]){dup3 <- dup3 + 1}
      if (lik_dup4 == scores[1]){dup4 <- dup4 + 1}
      if (lik_dup5 == scores[1]){dup5 <- dup5 + 1}
      if (lik_dup6 == scores[1]){dup6 <- dup6 + 1}

    }
  }

  output <- cbind(c('haploid', 'AAxBB', 'AAxAB', 'ABxAB', 'AAxBC', 'ABxAC',
'ABxCD'),c(hap, dup1, dup2, dup3, dup4, dup5, dup6))
  print(output)
}

# RESULTS
# hap
sim(c('a','b'), c('a','a'), 0.05, 5, 1) # can't run with more than 5
# dup1 (AAxBB)                          reads because the upper bound error would fix some of this
sim(c('a','a'), c('b','b'), 0.05, 5, 2)
# dup2 (AAxAB)
sim(c('a','a'), c('a', 'b'), 0.05, 5, 2)
```

dup3 (ABxAB)

sim(c('a','b'), c('a', 'b'), 0.05, 5, 2)

dup4 (AAxBC)

sim(c('a','a'), c('b', 'c'), 0.05, 5, 2)

dup5 (ABxAC)

sim(c('a','b'), c('a', 'c'), 0.05, 5, 2)

dup6 (ABxCD)

sim(c('a','b'), c('c', 'd'), 0.05, 5, 2)

Log-likelihood equations used to calculate locus types.

$$C = \frac{n_{total}!}{n_{aa}!\,n_{bb}!\,n_{cc}!\,n_{dd}!\,n_{ab}!\,n_{ac}!\,n_{ad}!\,n_{bc}!\,n_{bd}!\,n_{cd}!}$$

$$err = 0.05$$

Haploid locus (AB x 00):

$$LL(hap|genotypes) = \log\left(C\ \left(0.5\ (1-err) + \frac{err}{10}\right)^{n_{aa}+n_{bb}}\ \frac{err}{10}^{n_{total}-n_{aa}-n_{bb}}\right)$$

Duplicated1 locus (AA x BB):

$$LL(dup1|genotypes) = \log\left(C\ \left(1\ (1-err) + \frac{err}{10}\right)^{n_{ab}}\ \frac{err}{10}^{n_{total}-n_{ab}}\right)$$

Duplicated2 locus (AA x AB):

$$LL(dup2|genotypes) = \log\left(C\ \left(0.5\ (1 - err) + \frac{err}{10}\right)^{n_{ab}+n_{aa}}\ \frac{err^{n_{total}-n_{ab}-n_{aa}}}{10}\right)$$

Duplicated3 locus (AB x AB):

$$LL(dup3|genotypes)$$
$$= \log\left(C\ \left(0.25\ (1 - err) + \frac{err}{10}\right)^{n_{aa}+n_{bb}}\ \left(0.5\ (1 - err) + \frac{err}{10}\right)^{n_{ab}}\ \frac{err^{n_{total}-n_{ab}-n_{aa}-n_{bb}}}{10}\right)$$

Duplicated4 locus (AA x BC):

$$LL(dup4|genotypes) = \log\left(C\ \left(0.5\ (1 - err) + \frac{err}{10}\right)^{n_{ab}+n_{ac}}\ \frac{err^{n_{total}-n_{ab}-n_{ac}}}{10}\right)$$

Duplicated5 locus (AB x AC):

$$LL(dup5|genotypes)$$
$$= \log\left(C\ \left(0.25\ (1 - err) + \frac{err}{10}\right)^{n_{aa}+n_{ab}+n_{ac}+n_{bc}}\ \frac{err^{n_{total}-n_{aa}-n_{ab}-n_{ac}-n_{bc}}}{10}\right)$$

Duplicated6 locus (AB x CD):

$$LL(dup6|genotypes)$$
$$= \log\left(C\ \left(0.25\ (1 - err) + \frac{err}{10}\right)^{n_{ab}+n_{ad}+n_{bc}+n_{cd}}\ \frac{err^{n_{total}-n_{ab}-n_{ad}-n_{bc}-n_{cd}}}{10}\right)$$

Chapter 3 - Identification of genomic regions correlated with a reduction in reproductive success and the ecological mechanisms maintaining admixture in a hybridized population of trout.

Abstract

Hybridization is an important and common evolutionary process aiding in the diversification, adaptation, chromosomal evolution, and extinction of species. Divergent genomes are combined when species hybridize, resulting in novel gene combinations that could have profound fitness effects. Understanding how population fitness changes following hybridization and identifying what genetic and ecological mechanisms are responsible for those changes is important for understanding how natural selection operates and providing management guidance for populations of conservation concern. Here, we sought to identify the fitness consequences of hybridization, mechanisms responsible for the retention of hybridization, and genomic regions correlated with changes in reproductive success in a wild population of westslope cutthroat trout hybridized with non-native rainbow trout. We genotyped adults samples from Muhlfeld et al. (2009a), collected over a four year period, at 3027 loci. Increased admixture from non-native rainbow trout had a strong, negative effect on reproductive success. A decline of 53% was observed for individuals with an increased genetic contribution of 0.20 from rainbow trout. Despite apparent strong selection against rainbow trout ancestry, hybridization appears to be maintained largely by the invasion of rainbow trout from outside populations as well as the relatively high fitness of few hybrid individuals. Ten loci correlated with reproductive success were identified in females. Seven of the ten loci were linked to chromosomes and three were positioned on chromosomes. Loci linked to reproductive success were identified in genomic

regions where previous studies have identified excess species-specific ancestry in hybrid populations (RYHyb14 and RYHyb18) as well as chromosomes with a high proportion of duplicated markers (RYHyb02) and known Robertsonian polymorphism (RYHyb20).

Introduction

Hybridization is the interbreeding of genetically divergent individuals, and is common in both plant and animal taxa, occurring in as much as 25 and 10% of plants and animals respectively (Mallet 2005). When species hybridize, new genetic material is introduced to populations which may affect phenotypes on which natural selection can act (Burke & Arnold 2001). As such, hybridization could play an important role in evolution, and has already been shown to contribute to speciation (Mallet 2007), adaptation (Burke & Arnold 2001), chromosome evolution (Baack & Rieseberg 2007), and extinction (Rhymer & Simberloff 1996).

There is substantial interest in understanding the impact of hybridization on the fitness of populations (Arnold & Hodges 1995; Burke & Arnold 2001). Previous research has observed both positive (Benson *et al.* 2011; Fitzpatrick & Shaffer 2007; Grant & Grant 1992) and negative (Lancaster *et al.* 2007; Larcombe *et al.* 2014; Muhlfeld *et al.* 2009a) changes in reproductive success of hybrid individuals. During hybridization divergent genomes recombine, resulting in new genetic combinations (Winkler *et al.* 2010). Some novel alleles, or combination of alleles, may be adaptive, and could result in an increase in hybrid fitness (Burke & Arnold 2001). On the other hand, the introduction of new genetic material could swamp locally adaptive alleles, disrupt co-adapted gene complexes, or result in negative epistasis between loci, ultimately reducing fitness (Burke & Arnold 2001). Understanding how hybridization affects fitness is critical for predicting long-term population persistence in wild hybrid populations.

Understanding how hybridization is maintained when fitness is higher in one or both of the parental species remains an important question in hybrid biology research. Greater fitness in pure species relative to hybrids has been observed in the wild (Lancaster *et al.* 2007; Muhlfeld *et al.* 2009a; Roe *et al.* 2014), and understanding the mechanisms responsible for the maintenance of hybridization in such examples is important for proper management and predicting the long-term outcomes of hybridization. Hybridization in the face of selection against hybrid genotypes could be maintained by a number of mechanisms including ongoing migration from outside populations, heterosis of early generation hybrids, or introgression of genomes where adaptive alleles are maintained. Previous examples of hybridization being mediated by migration from outside populations and heterosis of F_1 hybrids have been observed (Allendorf *et al.* 2001; Kovach *et al.* 2015; Leary *et al.* 1993), and may be the result of conserved epistatic interactions in pure and F_1 hybrids who possess intact genomes from one or both parental species (Burke & Arnold 2001). If hybrid breakdown and selection against non-native genotypes is strong in the population, a hybrid swarm will not form and important adaptations that have evolved over generations will likely be maintained (Allendorf *et al.* 2001). However, if emigration rates are high, the loss of reproductive potential in the native population could affect the long-term population persistence (Allendorf *et al.* 2001).

Hybridization may also be maintained through introgression. Interspecific hybridization produces genomes with a combination of ancestry contributed from each of the founding species. If fitness in a population is controlled by a limited number of loci, some proportion of hybrid individuals will by chance possess the adaptive alleles necessary for high relative fitness. In such cases, hybridization will be maintained in the population through introgression at non-adaptive loci. Populations in which hybridization is maintained through introgression are expected to

produce hybrid swarms, and may require different management techniques compared to hybrid populations maintained by ongoing migration and heterosis (Allendorf *et al.* 2001).

Identification of the genetic basis of fitness is a fundamental goal of biology, and identifying loci correlated with fitness would provide insight into how natural selection operates in wild populations. Despite the importance of understanding fitness, few studies have identified loci correlated directly with reproductive success in the wild (Agren *et al.* 2013; Gratten *et al.* 2008). Instead, the genetic basis of "fitness-related" traits, such as growth (Ballesta *et al.* 2015; Reid *et al.* 2005) and coloration (Gratten *et al.* 2007; Malek *et al.* 2012), is identified. Numerous traits may be important to fitness, but without direct comparison to reproductive success, it is often unclear what controls fitness.

Hybrid populations provide a unique opportunity to identify the genetic basis of fitness (Darvasi & Shifman 2005; Winkler *et al.* 2010). In hybrid populations, individuals have different combinations of alleles contributed from the parental species, and the expression of traits that differ between the founding species is expected to vary in the hybrid individuals depending on their hybrid class, locus specific ancestry, and the genetic effect involved in the trait (heterosis, additive, dominant, epistasis) (Darvasi & Shifman 2005). This variation in both genomic contribution and resulting phenotype can be used to identify the genetic mechanisms responsible for the trait of interest, by genotyping individuals throughout their genome and identifying genotype-phenotype correlations (Darvasi & Shifman 2005; Goldstein *et al.* 2010; Winkler *et al.* 2010).

Random Forest (Breiman 2001), a machine learning algorithm, is a particularly powerful tool for identifying correlations between genotypes and phenotypes (Goldstein *et al.* 2010). Random Forest constructs a "forest" of decision trees by progressively identifying loci that best

group samples based on reduced variation of trait values within groups at each branch node (Goldstein *et al.* 2010). The method makes no assumptions about the genetic basis underlying the trait of interest and is robust to additive, dominant, and epistatic interactions (Goldstein *et al.* 2010; Holliday *et al.* 2012). Random Forest has recently been used in a number of genomic studies to identify loci explaining traits of interest (Brieuc *et al.* 2015; Goldstein *et al.* 2010; Holliday *et al.* 2012).

The westslope cutthroat trout (*Oncorhynchus clarkii lewisi*) provides a unique opportunity to explore how hybridization affects reproductive success, the maintenance of hybridization in a population with low reproductive success for hybrid individuals, and identification of genomic regions responsible for fitness in a hybrid population. The westslope cutthroat trout is one of 14 cutthroat trout subspecies native to western North America (Behnke 1992). Throughout their range populations are in decline largely due to hybridization with introduced rainbow trout (*O. mykiss*; (Liknes & Graham 1988; Shepard *et al.* 2005). Declines have been so dramatic that pure populations of westslope cutthroat trout are estimated to inhabit only 25% of their historical range (May 2009). Thus hybridization is of major conservation concern in the region (Boyer *et al.* 2008; Muhlfeld *et al.* 2009a; Shepard *et al.* 2005).

Increased admixture from rainbow trout has a strong, negative effect on reproductive success in westslope cutthroat trout (Muhlfeld *et al.* 2009a). Reproductive success measures of all spawning adults sampled from a tributary to the North Fork Flathead River, Montana, over a four year period revealed a rapid, decline in reproductive success. Increased rainbow trout ancestry was estimated to reduce reproductive success by nearly 50% for every 0.2 increment in proportion rainbow trout ancestry (Muhlfeld *et al.* 2009a). Although revealing, the study was limited by the number of markers used to estimate hybridization, eight. Inaccuracies in

hybridization estimates could affect the predicted relationship between admixture and fitness and warrants further investigation to confirm the findings of Muhlfeld *et al.* (2009a). Using recently developed genomic techniques (Baird *et al.* 2008; Hohenlohe *et al.* 2010) to genotype individuals at thousands of markers would provide a more robust estimate of hybridization (Allendorf *et al.* 2010), and would allow for investigation into the mechanisms that explain ongoing hybridization. In addition, pairing genotypes produced at thousands of loci with recently developed genomic maps (Berthelot *et al.* 2014; Brieuc *et al.* 2014; Drinan *et al.* in prep; Kodama *et al.* 2014) would allow for the identification of genomic regions correlated with fitness, a trait which is rarely mapped in wild populations.

Here we aim to understand the genomic basis and outcomes of interspecific hybridization on population composition and reproductive success. To understand the effects of hybridization on natural populations of cutthroat trout, we attempted to answer three important questions of hybrid populations: 1) does hybridization affect fitness, 2) what mechanisms maintain hybridization in populations, and 3) what genomic regions are correlated with reproductive success in a hybrid population. Our specific objectives are to first genotype individuals from Muhlfeld et al. (2009a) at thousands of loci to more accurately estimate the effect of hybridization on reproductive success in a wild population of hybrid trout. Second, we will characterize the ancestry of spawners in the hybrid population in order to identify the biological reasons for the maintenance of hybridization. Third, we will use Random Forest and comparative mapping to identify loci and genomic regions correlated with reproductive success in hybridized trout.

Methods

Sampling

Tissue samples were collected as part of a previous study in Langford Creek, Montana, USA (Muhlfeld *et al.* 2009a). Langford Creek is a small tributary to the North Fork Flathead River on the western border of Glacier National Park in Montana. Historically, a migratory population of westslope cutthroat trout was the only *Oncorhynchus* species in Langford Creek. However, the introduction of non-native rainbow trout (*O. mykiss*) into the North Fork Flathead River drainage, purportedly in the 1990s (Muhlfeld *et al.* 2009b), has resulted in hybridization throughout the drainage (Boyer *et al.* 2008). Individuals ranging from pure westslope cutthroat to pure rainbow trout can now be observed in Langford Creek. From 2003 to 2007, a weir trap was used to capture adults as they migrated into the creek to spawn. For each captured adult, a tissue sample (fin clip) was taken before the individual was released above the trap and allowed to spawn naturally. The resulting progeny were captured in a weir trap as they out migrated from Langford Creek. Tissue samples were collected from all progeny, and a parentage analysis was performed using microsatellite markers to identify reproductive success of parents (Muhlfeld *et al.* 2009a). All adult samples collected by Muhlfeld et al. (2009a) were used in the present study.

Laboratory methods

Adult samples were sequenced using restriction-site associated DNA (RAD) sequencing techniques (Baird *et al.* 2008). DNA samples from 57 adult females and 111 adult males, as well as 18 hatchery westslope cutthroat trout from Washoe Park Trout Hatchery, Anaconda, Montana, and 18 rainbow trout from Ennis National Fish Hatchery, Ennis, Montana, were RAD sequenced using standard laboratory techniques (Baird *et al.* 2008). Briefly, DNA was digested using the

Sbf-I restriction enzyme. Unique, six barcode identifiers were ligated to each DNA sequence to distinguish individual samples. Samples were then pooled and sequenced using the Illumina HiSeq platform. Each read was sequenced to 100 basepairs (bp).

Bioinformatic methods

Individuals were genotyped from raw sequence reads in three steps: removal of poor quality reads, alignment of sequence reads to the draft rainbow trout genome, and estimation of genotypes at identified loci using a maximum likelihood approach. First, poor quality reads were removed based on sequencing quality scores using previously published techniques and criteria (Brieuc *et al.* 2014; Drinan *et al.* in prep; Kodama *et al.* 2014). Reads that passed quality filtering were sorted by individual and aligned to the draft rainbow trout genome (Berthelot *et al.* 2014). Alignments were made using BOWTIE 1.1.1 (Langmead *et al.* 2009). Up to three mismatches were allowed between the reference genome and sequence reads. Because the rainbow trout genome contains duplicated loci which cannot be accurately genotyped in SNP population data, reads that aligned to multiple loci were removed from downstream analyses, thus removing duplicated loci. A total of 8 reads were required to call a genotype at a locus for an individual. Genotypes were estimated using a maximum likelihood methodology implemented by STACKS 1.0.8 (Catchen *et al.* 2013; Catchen *et al.* 2011). STACKS has been shown to mis-score heterozygote genotypes due to an overestimation of sequence error rates when no error guidance is provided (Catchen *et al.* 2013). To account for the overestimation of error rates, we used a maximum allowable error rate of 0.05. At least 70% of individuals must have been genotyped at a locus for it to be retained, and all loci with a minor allele frequency less than 0.05 were removed.

Admixture estimation

Individual proportion admixture was assessed using STRUCTURE v 2.3 (Pritchard *et al.* 2000). STRUCTURE uses a model-based clustering algorithm to determine the genetic makeup of individuals. We performed three STRUCTURE analyses and compared results to confirm that each analysis converged to the same findings. For each run, 50,000 burn-ins and 100,000 iterations with a K of 2 was used. We used a K of 2 rather than iterating through multiple possible values of K because westslope cutthroat and rainbow trout are the dominant species in the system and hybridization with other introduced species (Yellowstone cutthroat trout *O. c. bouvieri*) is rare (Boyer personal communication). STRUCTURE input files included all adults from Langford Creek as well as pure hatchery westslope cutthroat and rainbow trout. Hatchery fish were included to confirm that STRUCTURE clustered pure individuals of the same species together, while not grouping individuals of different species. In addition, hatchery fish served as learning samples to improve accuracy of admixture estimates in hybrid fish (Hauser *et al.* 2006).

Hybrid class estimation

The hybrid class of adult individuals from Langford Creek was estimated by plotting genome-wide admixture against average heterozygosity at diagnostic loci. By identifying hybrid classes present in Langford Creek, we can identify if migrants are entering the stream each year and whether hybrid breakdown is occurring. Salmonids (salmon, trout, char, and whitefish) experience almost complete recombination interference, with only one recombination per chromosome arm in approximately 90% of meioses (Brieuc *et al.* 2014; Thorgaard *et al.* 1983). As such, average heterozygosity at diagnostic loci and admixture are expected to vary

systematically between hybrid classes in early generations (Figure 1). For example, an F_1 hybrid receives haploid genomes from each of the founding species and is expected to be 0.5 admixed and have an average heterozygosity of 1.0 at diagnostic loci. An F_2 hybrid is expected to have the same admixture level, but an average heterozygosity of only 0.5 at diagnostic loci because the F_2 hybrid's parents experience interspecific recombination. In this study, we sought to identify F_1 (admixture = 0.5, heterozygosity = 1), F_2 (admixture = 0.5, heterozygosity = 0.5), backcross (admixture = 0.25 or 0.75, heterozygosity = 0. 5), and pure (admixture = 0 or 1, heterozygosity = 0) individuals.

Differences in average heterozygosity and admixture compared to expectations may exist due to recombination occurring at different locations on chromosome arms. In hybrids, this can result in gametes with differing genomic contributions from each of the founding species, even among siblings. To account for deviations from expected values, we classified individuals to a hybrid class if their measures of average heterozygosity and genome-wide admixture were within 0.05 of the expected values for a given hybrid class.

Diagnostic loci were identified using pure hatchery fish (18 westslope cutthroat trout and 18 rainbow trout). The identification of diagnostic loci was necessary as the predicted patterns of average heterozygosity are only true at markers with species-specific alleles, and using non-diagnostic marker to estimate average heterozygosity could bias the assignment of individuals to hybrid classes.

Effect of hybridization on reproductive success

The effect of genome-wide admixture on reproductive success was estimated using generalized linear mixed models. Reproductive success was measured as the number of offspring

produced by each parent, as reported by Muhlfeld *et al.* (2009a). We used a single model to investigate reproductive success. The model included sex and genome-wide admixture calculated by STRUCTURE as fixed effects. No interaction terms were included due to lack of visual evidence of possible interactions (Zuur *et al.* 2012). Spawning year was included as a random effect in all models to account for unmeasured interannual variation that may affect reproductive success. A zero-inflated negative binomial error distribution was used based on visual inspection of the data (Zuur *et al.* 2012). F_1 individuals had reproductive success values similar to parental species and were excluded from the analysis for biological reasons. Analyses were performed using R (R Core Team 2012) with the package glmmADMB (Skaug *et al.* 2010)

Identification of loci associated with reproductive success

A Random Forest (RF) analysis was used to identify candidate genomic regions correlated with reproductive success in hybrid and pure trout. RF is a machine learning algorithm that creates a forest of regression trees by categorizing observations based on predictors (Breiman 2001; Goldstein *et al.* 2010; Strobl *et al.* 2009). In RF, a bootstrapped sample is generated from the original data, and a random subset of predictor variables is chosen. The data is then searched to find the predictor that best splits the response variables into meaningful groups. The samples are split into subtrees and the process is repeated for each subtree until the nodes can no longer be divided. Data not included in the bootstrapped sample are then run down the tree to estimate prediction error rates. The process is repeated to create a forest of trees and the error rate for different combinations of variables can be compared to identify predictors that best explain the response. RF does not make assumptions about the distribution of the response variable, relationships between explanatory variables, or the genetic interactions at each locus

(additive, dominant, over-dominant, under-dominant) thus making it an ideal analysis for studies in which little is known about the mechanisms driving a trait.

In this study, we were interested in using RF to identify genomic regions that explain the variation in reproductive success. To identify such regions, we ran separate RF analyses for the females and males as evidence suggests that the traits influencing reproductive success may differ between the sexes in salmonids (Kodama *et al.* 2012; Muhlfeld *et al.* 2009a; Quinn 2005). In addition, F_1 individuals were removed as reproductive success appeared to follow a different distribution in F_1 than in pure and other hybrid trout. Reproductive success in the F_1 may be the result of heterosis. For each sex, we first performed three RF analyses to identify a general order of loci based on importance, as measured by the change in classification accuracy of a tree after permutation of genotypes at the locus of interest (Strobl *et al.* 2009). For each analysis, estimates of reproductive success from Muhlfeld *et al.* (2009a) were used as the phenotype, all loci were included as possible explanatory variables, and 20,000 trees were created. Loci were sorted based on their importance value, and the amount of variation explained by the top 0.01% - 100% of loci was identified by progressively adding less important loci to the sample and calculating variation explained by the subset of markers. The proportion of variation explained for each subset of top loci was averaged across all three analyses. Because the entire dataset contained thousands of loci and each analysis does not contain all possible combinations of markers, we expected that the importance of loci may vary between the three analyses. Therefore, loci that were identified in any of the three analyses as best at explaining phenotypic variation were retained and a backward purging approach was used on this subset of markers to identify a more precise order as outlined by (Holliday *et al.* 2012). Briefly, to perform the backward purging technique, RF was run on the subset of loci that explained the most phenotypic variation three

times with 10,000 trees per run. The locus with the combined lowest importance value was removed from the analysis. The amount of variation explained by the remaining loci was averaged across analyses. The process was re-run until the minimum number of allowable loci remained (two). The set of loci that explained the most variation in reproductive success using the backward purging approach were retained as candidate loci correlated with reproductive success.

Simulations were performed to identify if the loci selected by RF had significantly greater importance values than would be expected by chance. A challenge of RF analyses is that no explicit hypotheses are tested, but instead correlations are identified between dependent and independent variables without known levels of significance (Goldstein *et al.* 2010; Strobl *et al.* 2009). Importance values, calculated by permuting at a single locus, are often used as a proxy for significance (Strobl *et al.* 2009), but their exact interpretation can be difficult without understanding the distribution of importance values for loci that are known to be completely unlinked to the trait of interest. Therefore, we ran a simulation to identify if importance values overlapped between an unlinked locus, and loci empirically identified by RF to be important, based on methods outlined by Ono *et al.* (in review). One thousand RF analyses were performed. For each analysis, a single locus was simulated by using the genotypes of a randomly selected real locus. For the simulated locus, genotypes were randomly assigned to individuals and an RF analysis was performed using all loci plus the single simulated locus as explanatory variables and reproductive success as the dependent variable. Twenty thousand trees were created per analysis. Importance values calculated for the simulated loci were retained and compared with the importance values calculated for all loci using the true dataset. Overlap in importance values between the unlinked simulated locus and loci identified by RF to be linked to reproductive

success was interpreted as no significant difference existing between the real locus and a random locus.

All RF analyses were performed using the package randomForest (Liaw & Wiener 2002) within the statistical software R (R Core Team 2012). The package randomForest does not allow missing data, therefore we imputed missing genotypes using fastPHASE 1.4.0 (Scheet & Stephens 2006) prior to performing RF analyses.

Genomic regions correlated with reproductive success were identified by aligning loci correlated with reproductive success, as identified by RF, to previously produced genetic maps (Brieuc *et al.* 2014; Drinan *et al.* in prep; Kodama *et al.* 2014). A large proportion of the draft rainbow trout genome is not mapped to position (approximately 95%) or chromosome (approximately 53%), therefore the identification of genomic regions potentially important to reproductive success is difficult, if not impossible, using the rainbow trout genome alone. To overcome this challenge, we aligned the identified loci to previously produced linkage maps of sister taxa (Brieuc *et al.* 2014; Drinan *et al.* in prep; Kodama *et al.* 2014) using BOWTIE 1.1.1 (Langmead *et al.* 2009). Chromosome arms and marker order have remained largely conserved among salmonids, and homologous relationships between chromosome arms of different species have been identified (Kodama *et al.* 2014; Naish *et al.* 2013; Phillips *et al.* 2009). As a result, comparative mapping is a useful tool for positioning unmapped markers.

Results

Genotyping and admixture estimates

A total of 204 individuals were sequenced, 168 adults from Langford Creek, 18 hatchery westslope cutthroat trout, and 18 hatchery rainbow trout. Alignment of RAD sequence reads to

the draft rainbow trout genome resulted in the identification of 3027 loci sequenced in at least 70% of individuals. Of the loci identified, 1516 contained alleles that were diagnostic between westslope cutthroat and rainbow trout based on genotypes of pure hatchery rainbow and cutthroat trout, 1534 loci were mapped to a chromosome, and 140 were mapped to a position on a chromosome.

Admixture estimates were highly consistent among STRUCTURE runs (mean standard deviation of admixture between STRUCTURE runs < 0.01), which suggests that the MCMC algorithm for each run converged to similar results. Estimated hybridization ranged from pure westslope cutthroat to rainbow trout; however, samples of pure westslope cutthroat trout and low level hybrids as well as individuals with admixture near 0.5 were the most common (Figures 1and 2). Admixture estimates calculated from RAD data in this study were largely consistent with previous calculations using eight microsatellite markers (Muhlfeld et al. 2009; $R^2 = 0.93$; Figure S6).

Hybrid classes for adults from Langford Creek were identified by examining patterns of genome-wide admixture and heterozygosity at diagnostic loci in each individual. In total, 43% of adults from Langford Creek were assigned to pure or early generation hybrid classes (F_1, F_2, or BC; Table S1; Figure 1). Of those identified, pure westslope cutthroat trout were the most common, with 17% of all individuals assigned to that class. Nine percent of individuals were each identified as rainbow trout or F_1, 3% as F_2 or backcrossed between F_1 and rainbow trout, and 2% as backcrossed between F_1 and westslope cutthroat trout. Fifty-seven percent of individuals could not be assigned to early generation hybrid classes as they did not fit predicted estimates of genome-wide admixture and average heterozygosity at diagnostic markers for pure, F_1, F_2, or BC individuals. Our inability to assign these individuals to early generation hybrid

classes may be the result of a more complex breeding history than the classes explored here (F_1 x F_2, Pure x F_2, etc.).

Effect of hybridization on reproductive success

Reproductive success varied depending on hybridization level (Figure 2). Of non-F_1 adults, pure westslope cutthroat trout produced the most offspring (median = 12; 25 and 75% quantiles = 4, 20), rainbow trout produced the fewest (median = 0; 25 and 75% confidence interval = 0, 1), and hybrid individuals were intermediate (median = 1; 25 and 75% quantiles = 0, 3.75; Figure 2). In the GLMM analysis, both proportion admixture and sex were significantly correlated with reproductive success (Table 1). Increased admixture from rainbow trout significantly reduced reproductive success (p-value < 0.01, $z = -5.00$). An increase of 0.2 in rainbow trout ancestry was correlated with a decline of 53% in reproductive success (95% confidence limit = 41-68%). In addition, males produced significantly fewer offspring than females (p-value < 0.01, $z = -2.84$). On average females produce 6.4 offspring (variance = 84.7), while males produced only 3.3 (variance = 75.6).

Maintenance of admixture based on estimates of hybrid classes

The maintenance of admixture appears to be largely driven by emigration of rainbow trout from outside sources as well as high fitness of a small number of F_1 and non-F_1 hybrids (Figure 2). In all four years of the study, adult, pure rainbow trout were observed entering Langford Creek as spawners (Figure 1). Given the low reproductive success of rainbow trout in Langford Creek, it is unlikely that these fish were the result of chance successful reproduction between two pure rainbow trout in Langford Creek, but instead were migrants from a

neighboring population in the system. Populations of trout at lower elevations are often dominated by rainbow trout (Boyer *et al.* 2008; Muhlfeld *et al.* 2014).

The invasion of Langford Creek appears to be recent and ongoing. The proportion of F_1 hybrids in the adult spawners peaked in 2004 (25% of total adult population in that year; Figure S7), and was followed by an increase in backcrossed rainbow trout in 2006 and F_2s in 2007. The generation time of westslope cutthroat and rainbow trout in the system is 3-5 years (Kovach *et al.* 2015), suggesting that the largest wave of invasion occurred in 2000. It is hypothesized that the major invasion of rainbow trout into the NF Flathead River drainage were from illegally stocked rainbow trout in 1997 (Muhlfeld *et al.* 2009b), thus it took approximately one generation for rainbow trout alleles to reach Langford Creek, 40 km upstream of the likely invasion site.

Identification of genomic regions correlated with reproductive success

In the females, 52% of the variation in reproductive success was explained by the top 0.07% of loci in the initial Random Forest analysis (22 loci), using all loci. Using a backward purging approach that was based on the best 22 predictor loci, 11 loci explained the maximum phenotypic variation (61%). Ten of the 11 loci identified by RF to be linked to reproductive success had importance values that were significantly different than simulated unlinked loci (Figure S8). A principle components analysis (PCA) of the top 11 loci found that females largely grouped into reproducers and non-reproducers based on the first principle component (PC1; Figure 3). PC1 explained 71% of the phenotypic variation observed, while the second principle component explained only 8%.

In the males, only 6% of the variation in reproductive success was explained by the top 1% of loci identified by the initial Random Forest analysis, which included all loci. Using a

backward purging approach on the top-ranking 37 loci for predicting reproductive success, 8 loci were identified to explain the maximum phenotypic variation, 21%. The importance values of our simulated, unlinked loci overlapped with the importance values of all real loci (Figure S9), suggesting that the loci identified by RF are not actually linked to reproductive success in the males. A PCA of the top 8 loci explained a large amount of the genetic variation in PC1 (57%) and PC2 (12%), but failed to separate reproducers and non-reproducers into discernible groups (Figure 4).

Of the loci linked to reproductive success by RF, we were able to place 13 loci (eight female and five male) to chromosomes and six (four female and two male) to positions in linkage groups by comparative mapping with Chinook salmon (Brieuc *et al.* 2014), coho salmon (Kodama *et al.* 2014), and hybrid cutthroat trout (Drinan *et al.* in prep). In the female, two markers mapped each to RYHyb18 and RYHyb20, while one marker was mapped each to RYHyb02, RYHyb12, RYHyb14, RYHyb18, and RYHyb21 (Figure 5). In rainbow trout, Omy20 and Omy28 are two separate acrocentric chromosomes (Phillips *et al.* 2006). However, in cutthroat trout, Omy20 and Omy28 are hypothesized to be fused (Ostberg *et al.* 2013). In the male, two markers mapped to RYHyb03, and one marker mapped to each of RYHyb11, RYHyb18, and RYHyb21 (Figure 5, Table S1). RYHyb21 was the only linkage group for which markers in both the female and male mapped, but because the marker in the female could not be mapped to position, it is unknown if they are from the same genomic region.

Discussion

Here we aimed to understand the genomic basis of admixture between westslope cutthroat and rainbow trout and the outcomes of hybridization on population composition and

reproductive success. To understand the effects of hybridization on natural populations of cutthroat trout, we attempted to answer three important questions of hybrid populations: 1) does hybridization affect fitness, 2) what ecological mechanisms maintain hybridization in populations, and 3) what genomic regions are correlated with reproductive success in a hybrid population. Results from regression analyses confirm that hybridization dramatically reduces reproductive success. Despite evidence of selection against increased admixture from rainbow trout, hybrid genes appear to be maintained largely through migrants from other populations and a small number of fit hybrid individuals. Loci correlated with reproductive success were identified, and may be important for predicting fitness outcomes in wild populations.

Admixture and reproductive success

Increased admixture from rainbow trout resulted in a sharp decline in reproductive success, confirming the previous results of Muhlfeld *et al.* (2009a). However, limited replication of spawners with low genetic contribution from rainbow trout may have hindered our ability to precisely detect the slope of the decline. In total, we sampled only two individuals in which genomic proportion rainbow trout ranged from 0.05 to 0.15 (Figure 2). These two individuals had low reproductive success, and if they are representative of all hybrids at this level, the decay in reproductive success due to increased genomic contribution from rainbow trout may be stronger than estimated here. However, we do not feel that this limitation affected our major finding that increased admixture from rainbow trout reduces reproductive success.

Although the major finding that increased admixture from rainbow trout results in a decline in reproductive success observed by Muhlfeld *et al.* (2009a) were confirmed by the current study, differences in admixture and hybrid class estimates for some individuals did exist.

First, admixture estimates between the studies were generally similar (Figure S6). However, admixture estimates for hybrid individuals with admixture values closer to the parental values varied between the studies. Most individuals falling into these categories in the earlier study were identified as pure westslope cutthroat or rainbow trout in the present study (Figure S6). A second difference is that some individuals classified as F_1 in the microsatellite based analyses were identified as non-F_1 hybrids in the present study. Both differences are likely due to the number of markers used to estimate admixture and hybrid classes in each study. In the present study, over 3,000 loci were used by STRUCTURE to estimate admixture and over 1500 diagnostic loci were used to estimate hybrid classes. In contrast, the earlier study was limited to eight diagnostic loci to estimate both. More loci are expected to result in a more accurate estimate of admixture and hybrid classes. As such, we believe the estimates of admixture and hybrid class produced from this study to be more accurate than previous estimates, but the results are largely unchanged.

Maintenance of admixture

Admixture appears to be largely mediated by two sources, emigration of rainbow trout from outside populations and high fitness of a relatively few F_1 and non-F_1 hybrid individuals based on reproductive success and estimated hybrid classes. These results are largely in agreement with previous research in Langford Creek and a neighboring stream that observed a higher number of admixed spawning adults entering the stream than juveniles out-migrating (Kovach *et al.* 2015). In the present study, pure adult rainbow trout were observed entering Langford Creek as spawners in all four years. Given the low reproductive success of rainbow trout in the system, it is unlikely that the pure rainbow trout observed entering Langford Creek

each year were the result of chance successful reproduction between two pure rainbow trout in the previous generation. Instead, we hypothesize that the rainbow trout entering Langford Creek were mainly emigrants from other populations persisting at lower elevations in the drainage (Boyer *et al.* 2008; Muhlfeld *et al.* 2014). Throughout the NF and MF Flathead River drainages, introduced rainbow trout have come to dominate lower elevation populations (Boyer *et al.* 2008; Muhlfeld *et al.* 2014), potentially due to differences in thermal tolerance (Bear *et al.* 2007). In other species and anadromous *O.mykiss*, the progeny of hatchery derived fish and hybrid individuals, have been observed to stray at a higher rate than natural fish (Ford *et al.* 2015; Gilk *et al.* 2004; Keefer & Caudill 2012). As such, these highly admixed populations in lower elevations may serve as sources for the dispersal of rainbow trout genes throughout the drainage.

A second mechanism maintaining hybridization in Langford Creek is the high fitness of a small number of F_1 hybrids, and individuals of other hybrid classes. When species interbreed and F_1s are created, the resulting offspring possess an intact genome contributed from each of the parental species. The retention of entire genomes may preserve fitness influencing epistatic interactions among loci that have evolved within species (Lippman & Zamir 2007). In addition, F_1 individuals may benefit from the masking of deleterious alleles that have become fixed in each of the parental species (Lippman & Zamir 2007). These beneficial interactions within the F_1 genome could influence patterns of reproductive success observed in this study. In contrast, it is unclear why two male and two female non-F_1 hybrid individuals have relatively high reproductive success, there are two possible explanations. First, over 20 million rainbow trout have been introduced into the Flathead River drainage starting in the 1800s (Hitt *et al.* 2003). These introductions were largely in the lower sections of the drainage such as Flathead River and Flathead Lake. Rainbow trout derived from these introductions may have gone through more

than 200 years (~50 generations) of natural selection, and many poorly adapted alleles may have been purged from these populations. The more recent, purported rainbow trout introduction in 1997 means that at least two populations of rainbow trout that have been under selection for vastly different time frames may persist in the Flathead River drainage. Both could be contributing to admixture in Langford Creek. The outcomes of hybridization with historical and recent introductions could explain why some chance hybrids have high reproductive success. A second explanation could be the random segregation of fitness-influencing alleles. Fitness is likely controlled by multiple loci, and some proportion of hybrids will likely possess beneficial alleles that could result in relatively high reproductive success by chance. Depending on the genetic mechanisms responsible for fitness, the combination of alleles important to fitness could be either rare or common, and may explain the high reproductive success observed in some hybrid individuals in this study.

Identification of genomic regions linked to reproductive success

The present study is one of the first to identify genomic regions correlated with reproductive success in a wild population salmonid species. Differences in reproductive success of individuals may be due to natural selection on species-specific ancestry at genomic regions identified by RF. In the RF analysis of female reproductive success, 11 loci were linked to the trait of interest. Of these 11 loci, ten were significantly different than a simulated unlinked locus, and three of those ten were positioned within chromosomes. Although we were unable to position all loci linked to reproductive success due to the infancy of the rainbow trout genome, the loci that were positioned provide important insight into genomic regions that may harbor adaptively important loci. Two of the four loci that were positioned were placed in genomic

regions known to contain excess cutthroat ancestry in a hybrid cutthroat and rainbow trout mapping family (RYHyb14 and RYHyb18: Drinan *et al.* in prep.). Regions of excess species-specific ancestry were identified by comparing expected versus observed species ancestries in juvenile F_2 progeny between Yellowstone cutthroat and rainbow trout (Drinan *et al*. in prep). Observed regions of excess species-specific ancestry in the mapping families were hypothesized to be driven by drift or natural selection in the earlier study. However, two regions overlapped in both studies and both genomic regions showed excess of cutthroat trout ancestry in the mapping families. Therefore, we hypothesize that natural selection may be driving the selection of cutthroat alleles on chromosomes RYHyb14 and RYHyb18.

The remaining locus linked to reproductive success in females and positioned in the genome persists on a chromosome arms with a high proportion of duplicated markers retained following an autotetraploid event in a common ancestry. An ancestor to all *Oncorhynchus* species experienced an autotetraploid whole genome duplication event 88-103 mya (Macqueen & Johnston 2014). Since that time, species have diverged largely through Robertsonian rearrangements (fusions and fissions of entire chromosome arms) and sequence divergence, resulting in species with a similar number of chromosome arms (approximately 100), but differing karyotypes (Allendorf & Thorgaard 1984). Re-diploidization following whole genome duplication is ongoing, but the process of diploidization is not occurring at the same rate throughout the genome (Drinan *et al*. in prep., Kodama *et al.* 2014, Brieuc *et al.* 2014). Instead, sixteen chromosome arms appear to have largely retained duplicated regions (Brieuc *et al.* 2014; Kodama *et al.* 2014). One locus (chrUn [861182876]) linked to reproductive success in the female was positioned on RYHyb02p, a chromosome arm that has retained a high proportion of duplicated markers (Brieuc *et al.* 2014; Kodama *et al.* 2014). The evolutionary importance of

duplicated genomic regions is unknown, but it is hypothesized that they may provide additional redundancy of important genes or increase the flexibility of species to adapt to new environments (Comai 2005).

An additional locus (chrUn_28 [16461881]) that could not be positioned to an exact location on a chromosome, but mapped to RYHyb20 persists on a chromosome with a known Robertsonian polymorphism between cutthroat and rainbow trout (Ostberg *et al.* 2013). In cutthroat trout, the homolog of RYHyb20 is a metacentric chromosome formed from the fusion of two homologous acrocentric arms, Omy20 and Omy28 in rainbow trout (Ostberg *et al.* 2013). Karyotype differences have been observed to reduce recombination across long stretches of chromosome arms (Ostberg *et al.* 2013). Previous research has reported that structural polymorphisms can contain adaptively important loci (Rane *et al.* 2015), and the reduction in recombination due to architectural differences could serve to protect adaptations that persist on this chromosomal region in hybrid trout.

We found little evidence in male fish for loci associated with reproductive success. Importance values for loci identified by RF to be linked with reproductive success in the males overlapped with the distribution of importance values of random loci. In addition, a PCA of loci identified by RF in males failed to separate reproducers and non-reproducers. Together, these findings suggest that the loci identified by RF are likely not linked to reproductive success in the males. Our inability to identify loci correlated with reproductive success in the males may be because only a few males were highly reproductively successful. In total, only four males produced more than ten progeny. Reproductive success is usually dominated by only a few males across salmon species (Anderson *et al.* 2011; Kodama *et al.* 2012; Seamons *et al.* 2004), and the

few reproductively successful males observed in this study may have limited our ability to identify genetic mechanisms contributing to reproductive success.

Management implications

The results of this study have important management implications as hybridization becomes an increasingly common occurrence. First, the finding that increased rainbow trout ancestry has a negative effect on reproductive success highlights the need for development of policy on the management of hybrid populations. Currently, no universal guidelines exist in the United States. Instead, each hybrid system is investigated independently. Previous studies have observed no morphological differences between pure westslope cutthroat trout and those with up to 20% admixture contributed from rainbow trout (reviewed by USFWS 2003). Because no morphological differences were observed, populations with up to 20% admixture have the same conservation status as pure populations of cutthroat trout. However, given the updated findings from this study that an increase in rainbow trout ancestry of 0.20 results in a greater than 50% reduction in reproductive success, in addition to other recently published studies (Corsi *et al.* 2013), there is evidence that important differences likely do exist between pure and low level hybrid individuals. Further review is warranted to determine what level of admixture is acceptable.

Second, understanding that migration has a role in mediating admixture has important conservation implications, as even strong selection against migrants can be overcome if rates of emigration are great enough. In the case of hatchery derived invaders, such as rainbow trout in the Flathead River drainage, rates of straying may be elevated even in the progeny of hatchery derived individuals as has been observed in Chinook salmon (Ford *et al.* 2015). Because of the

elevated stray rates of these non-native fish, removal of non-native rainbow trout as has been initiated in the NF Flathead river (Montana FWP 2012) may be necessary to prevent the spread of rainbow trout genes in drainages where they have been introduced.

Lastly, the results of this study may be useful for predicting which wild populations are susceptible to hybridization. By examining allele frequencies in wild populations at loci identified as important to reproductive success in this study, managers may be able to predict the fitness outcomes of admixed populations. Such genetic tools could be valuable for prioritizing limited resources.

Conclusions

In summary, hybridization of westslope cutthroat trout with introduced rainbow trout has a strong, negative effect on reproductive success, but appears to be maintained largely through emigration and by a small number of highly productive hybrids. Genomic regions contributing to reproductive success were identified and previous identification of the same regions in mapping families suggests selection may be driving allele frequencies at these loci. In addition, structural polymorphisms between the species and past genome evolution may contribute to reproductive success. This is one of the first studies to directly link reproductive success, rather than surrogate traits, to genetic loci in a wild population, and the identified loci may serve as important management tools to predict future population susceptibility to hybridization.

Although a major step forward, this research also provides the groundwork for further studies. It is important to identify the phenotypic traits that drive changes in reproductive success observed in Langford Creek. Numerous hatchery and field studies have investigated trait variation between cutthroat and rainbow trout (Corsi *et al.* 2013; Muhlfeld *et al.* 2009b; Ostberg

et al. 2011; Rasmussen *et al.* 2012; Seiler & Keeley 2007a, b), but no known studies have directly linked traits to reproductive success. In addition, it is unclear if the negative effect of hybridization observed in Langford Creek is generalizable across the range of westslope cutthroat trout, or whether the effect of hybridization is dependent on habitat (Rasmussen *et al.* 2012). Understanding such patterns would be important for identifying where limited management resources should be directed.

References

Agren J, Oakley CG, McKay JK, Lovell JT, Schemske DW (2013) Genetic mapping of adaptation reveals fitness tradeoffs in arabidopsis thaliana. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 21077-21082.

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. *Nature Reviews Genetics* **11**, 697-709.

Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001) The problems with hybrids: Setting conservation guidelines. *Trends in Ecology & Evolution* **16**, 613-622.

Allendorf FW, Thorgaard GH (1984) Tetraploidy and the evolution of salmonid fishes. In: *Evolutionary genetics of fishes* (ed. Turner BJ), pp. 1-27. Plenum, New York.

Anderson JH, Faulds PL, Atlas WI, Pess GR, Quinn TP (2011) Selection on breeding date and body size in colonizing coho salmon, *Oncorhynchus kisutch*. *Molecular Ecology* **19**, 2562-2573.

Arnold ML, Hodges SA (1995) Are natural hybrids fit or unfit relative to their parents. *Trends in Ecology & Evolution* **10**, 67-71.

Baack EJ, Rieseberg LH (2007) A genomic view of introgression and hybrid speciation. *Current Opinion in Genetics & Development* **17**, 513-518.

Baird NA, Etter PD, Atwood TS*, et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced rad markers. *Plos One* **3**.

Ballesta P, Mora F, Ruiz E, Contreras-Soto R (2015) Marker-trait associations for survival, growth, and flowering components in eucalyptus cladocalyx under arid conditions. *Biologia Plantarum* **59**, 389-393.

Bear EA, McMahon TE, Zale AV (2007) Comparative thermal requirements of westslope cutthroat trout and rainbow 11-out: Implications for species interactions and development of thermal protection standards. *Transactions of the American Fisheries Society* **136**, 1113-1121.

Behnke RJ (1992) *Native trout of western north america* American Fisheries Society, Bethesda, Md.

Benson JF, Hostetler JA, Onorato DP*, et al.* (2011) Intentional genetic introgression influences survival of adults and subadults in a small, inbred felid population. *Journal of Animal Ecology* **80**, 958-967.

Berthelot C, Brunet F, Chalopin D*, et al.* (2014) The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nature Communications* **5**.

Boyer MC, Muhlfeld CC, Allendorf FW (2008) Rainbow trout (*Oncorhynchus mykiss*) invasion and the spread of hybridization with native westslope cutthroat trout (*Oncorhynchus clarkii lewisi*). *Canadian Journal of Fisheries and Aquatic Sciences* **65**, 658-669.

Breiman L (2001) Random forests. *Machine Learning* **45**, 5-32.

Brieuc MSO, Ono K, Drinan D, Naish KA (2015) Integration of random forest with population-based outlier analyses provides insight on the genomic basis and evolution of run timing in chinook salmon (*Oncorhynchus tshawytscha*). *Molecular Ecology*.

Brieuc MSO, Waters CD, Seeb JE, Naish KA (2014) A dense linkage map for chinook salmon (*Oncorhynchus tshawytscha*) reveals variable chromosomal divergence after an ancestral whole genome duplication event. *G3-Genes Genomes Genetics* **4**, 447-460.

Burke JM, Arnold ML (2001) Genetics and the fitness of hybrids. *Annual Review of Genetics* **35**, 31-52.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: An analysis tool set for population genomics. *Molecular Ecology* **22**, 3124-3140.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: Building and genotyping loci de novo from short-read sequences. *G3-Genes Genomes Genetics* **1**, 171-182.

Comai L (2005) The advantages and disadvantages of being polyploid. *Nature Reviews Genetics* **6**, 836-846.

Corsi MP, Eby LA, Barfoot CA (2013) Hybridization with rainbow trout alters life history traits of native westslope cutthroat trout. *Canadian Journal of Fisheries and Aquatic Sciences* **70**, 895-904.

Darvasi A, Shifman S (2005) The beauty of admixture. *Nature Genetics* **37**, 118-119.

Drinan DP, Ostberg CO, Brieuc MSO, Naish KA (in prep) Changes to recombination and limited segregation distortion revealed in the hybrid trout genome.

Fitzpatrick BM, Shaffer HB (2007) Hybrid vigor between native and introduced salamanders raises new challenges for conservation. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 15793-15798.

Ford MJ, Murdoch A, Hughes M (2015) Using parentage analysis to estimate rates of straying and homing in chinook salmon (*Oncorhynchus tshawytscha*). *Molecular Ecology* **24**, 1109-1121.

Gilk SE, Wang IA, Hoover CL*, et al.* (2004) Outbreeding depression in hybrids between spatially separated pink salmon, *Oncorhynchus gorbuscha*, populations: Marine survival, homing ability, and variability in family size. *Environmental Biology of Fishes* **69**, 287-297.

Goldstein BA, Hubbard AE, Cutler A, Barcellos LF (2010) An application of random forests to a genome-wide association dataset: Methodological considerations & new findings. *Bmc Genetics* **11**.

Grant PR, Grant BR (1992) Hybridization of bird species. *Science* **256**, 193-197.

Gratten J, Beraldi D, Lowder BV*, et al.* (2007) Compelling evidence that a single nucleotide substitution in tyrp1 is responsible for coat-colour polymorphism in a free-living population of soay sheep. *Proceedings of the Royal Society B-Biological Sciences* **274**, 619-626.

Gratten J, Wilson AJ, McRae AF*, et al.* (2008) A localized negative genetic correlation constrains microevolution of coat color in wild sheep. *Science* **319**, 318-320.

Hauser L, Seamons TR, Dauer M, Naish KA, Quinn TP (2006) An empirical verification of population assignment methods by marking and parentage data: Hatchery and wild steelhead (*Oncorhynchus mykiss*) in forks creek, washington, USA. *Molecular Ecology* **15**, 3157-3173.

Hitt NP, Frissell CA, Muhlfeld CC, Allendorf FW (2003) Spread of hybridization between native westslope cutthroat trout, *Oncorhynchus clarki lewisi*, and nonnative rainbow trout, *Oncorhynchus mykiss*. *Canadian Journal of Fisheries and Aquatic Sciences* **60**, 1440-1451.

Hohenlohe PA, Bassham S, Etter PD*, et al.* (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced rad tags. *Plos Genetics* **6**.

Holliday JA, Wang T, Aitken S (2012) Predicting adaptive phenotypes from multilocus genotypes in sitka spruce (*Picea sitchensis*) using random forest. *G3-Genes Genomes Genetics* **2**, 1085-1093.

Keefer ML, Caudill CC (2012) A review of adult salmon and steelhead straying with an emphasis on columbia river populations. Department of Fish and Wildlife Resources, College of Natural Resources, University of Idaho.

Kodama M, Brieuc MSO, Devlin RH, Hard JJ, Naish KA (2014) Comparative mapping between coho salmon (*Oncorhynchus kisutch*) and three other salmonids suggests a role for chromosomal rearrangements in the retention of duplicated regions following a whole genome duplication event. *G3-Genes Genomes Genetics* **4**, 1717-1730.

Kodama M, Hard JJ, Naish KA (2012) Temporal variation and selection on body length and date of return in a wild population of coho salmon, *Oncorhynchus kisutch*. *BMC Evolutionary Biology* **12**, 116.

Kovach RP, Muhlfeld CC, Boyer MC*, et al.* (2015) Dispersal and selection mediate hybridization between a native and invasive species. *Proceedings of the Royal Society B-Biological Sciences* **282**.

Lancaster ML, Bradshaw CJA, Goldsworthy SD, Sunnucks P (2007) Lower reproductive success in hybrid fur seal males indicates fitness costs to hybridization. *Molecular Ecology* **16**, 3187-3197.

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**.

Larcombe MJ, Barbour RC, Vaillancourt RE, Potts BM (2014) Assessing the risk of exotic gene flow from eucalyptus globulus plantations to native e-ovata forests. *Forest Ecology and Management* **312**, 193-202.

Leary RF, Allendorf FW, Forbes SH (1993) Conservation genetics of bull trout in the Columbia and Klamath River drainages. *Conservation Biology* **7**, 856-865.

Liaw A, Wiener M (2002) Classification and regression by randomforest. *R News* **2**, 18-22

Liknes GA, Graham PJ (1988) Westslope cutthroat trout in Montana: Life history, status, and management. *American Fisheries Society Symposium* **4**, 53-60.

Lippman ZB, Zamir D (2007) Heterosis: Revisiting the magic. *Trends in Genetics* **23**, 60-66.

Macqueen DJ, Johnston IA (2014) A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. *Proceedings of the Royal Society B-Biological Sciences* **281**.

Malek TB, Boughman JW, Dworkin I, Peichel CL (2012) Admixture mapping of male nuptial colour and body shape in a recently formed hybrid population of threespine stickleback. *Molecular Ecology* **21**, 5265-5279.

Mallet J (2005) Hybridization as an invasion of the genome. *Trends in Ecology & Evolution* **20**, 229-237.

Mallet J (2007) Hybrid speciation. *Nature* **446**, 279-283.

Montana FWP (2012) Flathead river hybrid trout suppression project draft environmental assessment.

Muhlfeld CC, Kalinowski ST, McMahon TE*, et al.* (2009a) Hybridization rapidly reduces fitness of a native trout in the wild. *Biology Letters* **5**, 328-331.

Muhlfeld CC, Kovach RP, Jones LA*, et al.* (2014) Invasive hybridization in a threatened species is accelerated by climate change. *Nature Climate Change* **4**, 620-624.

Muhlfeld CC, McMahon TE, Belcer D, Kershner JL (2009b) Spatial and temporal spawning dynamics of native westslope cutthroat trout, *Oncorhynchus clarkii lewisi*, introduced rainbow trout, *Oncorhynchus mykiss*, and their hybrids. *Canadian Journal of Fisheries and Aquatic Sciences* **66**, 1153-1168.

Naish KA, Phillips RB, Brieuc MSO*, et al.* (2013) Comparative genome mapping between chinook salmon (*Oncorhynchus tshawytscha*) and rainbow trout (*O. mykiss*) based on homologous microsatellite loci. *G3-Genes Genomes Genetics* **3**, 2281-2288.

Ostberg CO, Duda JJ, Graham JH*, et al.* (2011) Growth, morphology, and developmental instability of rainbow trout, yellowstone cutthroat trout, and four hybrid generations. *Transactions of the American Fisheries Society* **140**, 334-344.

Ostberg CO, Hauser L, Pritchard VL, Garza JC, Naish KA (2013) Chromosome rearrangements, recombination suppression, and limited segregation distortion in hybrids between yellowstone cutthroat trout (*Oncorhynchus clarkii bouvieri*) and rainbow trout (*O. mykiss*). *Bmc Genomics* **14**.

Phillips RB, Keatley KA, Morasch MR*, et al.* (2009) Assignment of atlantic salmon (*Salmo salar*) linkage groups to specific chromosomes: Conservation of large syntenic blocks corresponding to whole chromosome arms in rainbow trout (*Oncorhynchus mykiss*). *Bmc Genetics* **10**.

Phillips RB, Nichols KM, DeKoning JJ, *et al.* (2006) Assignment of rainbow trout linkage groups to specific chromosomes. *Genetics* **174**, 1661-1670.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.

Quinn TP (2005) *The behavior and ecology of pacific salmon and trout* American Fisheries Society, Bethesda, Md.

Rane RV, Rako L, Kapun M, Lee SF, Hoffmann AA (2015) Genomic evidence for role of inversion 3rp of drosophila melanogaster in facilitating climate change adaptation. *Molecular Ecology* **24**, 2423-2432.

Rasmussen JB, Robinson MD, Hontela A, Heath DD (2012) Metabolic traits of westslope cutthroat trout, introduced rainbow trout and their hybrids in an ecotonal hybrid zone along an elevation gradient. *Biological Journal of the Linnean Society* **105**, 56-72.

Reid DP, Szanto A, Glebe B, Danzmann RG, Ferguson MM (2005) Qtl for body weight and condition factor in atlantic salmon (*Salmo salar*): Comparative analysis with rainbow trout (*Oncorhynchus mykiss*) and arctic charr (*Salvelinus alpinus*). *Heredity* **94**, 166-172.

Rhymer JM, Simberloff D (1996) Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics* **27**, 83-109.

Roe AD, MacQuarrie CJK, Gros-Louis M-C*, et al.* (2014) Fitness dynamics within a poplar hybrid zone: I. Prezygotic and postzygotic barriers impacting a native poplar hybrid stand. *Ecology and Evolution* **4**, 1629-1647.

Scheet P, Stephens M (2006) A fast and flexible statistical model for large-scale population genotype data: Applications to inferring missing genotypes and haplotypic phase. *American Journal of Human Genetics* **78**, 629-644.

Seamons TR, Bentzen P, Quinn TP (2004) The mating system of steelhead, *Oncorhynchus mykiss*, inferred by molecular analysis of parents and progeny. *Environmental Biology of Fishes* **69**, 333-344.

Seiler SM, Keeley ER (2007a) A comparison of aggressive and foraging behaviour between juvenile cutthroat trout, rainbow trout and f1 hybrids. *Animal Behaviour* **74**, 1805-1812.

Seiler SM, Keeley ER (2007b) Morphological and swimming stamina differences between yellowstone cutthroat trout (*Oncorhynchus clarkii bouvieri*), rainbow trout (*Oncorhynchus mykiss*), and their hybrids. *Canadian Journal of Fisheries and Aquatic Sciences* **64**, 127-135.

Shepard BB, May BE, Urie W (2005) Status and conservation of westslope cutthroat trout within the western United States. *North American Journal of Fisheries Management* **25**, 1426-1440.

Strobl C, Malley J, Tutz G (2009) An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods* **14**, 323-348.

Thorgaard GH, Allendorf FW, Knudsen KL (1983) Gene-centromere mapping in rainbow-trout - high interference over long map distances. *Genetics* **103**, 771-783.

Winkler CA, Nelson GW, Smith MW (2010) Admixture mapping comes of age. In: *Annual review of genomics and human genetics, vol 11* (eds. Chakravarti A, Green E), pp. 65-89.

Zuur AF, Saveliev AA, Ieno EN (2012) *Zero inflated models and generalized linear mixed models with r* Highland Statistics Ltd., Newburgh, U.K.

Tables

Parameter	Coefficient	Standard error	z-score	p-value
Intercept	2.44	0.36	6.78	<0.01
Proportion rainbow trout	-2.65	0.41	-6.49	<0.01
Sex (male)	-0.80	0.25	-3.20	<0.01

Table 3.1: Regression coefficients for generalized linear mixed model used to investigate factors influencing reproductive success in a hybrid population of trout. A zero-inflated negative binomial error distribution was used. Proportion rainbow trout was estimated using 3027 loci.

Female

Rainbow trout genome locus	Chromosome	Position (cM)
chrUn_14 (11108873)	RYHyb14	382 – 530
chrUn_18 (916565)	RYHyb18	303 – 342
chrUn_18 (9511642)	RYHyb18	-
chrUn_28 (16461881)	RYHyb20	-
chrUn_21 (7080146)	RYHyb21	-
chrUn (372445148)	-	-
chrUn (384507463)	-	-
chrUn (472292170)*	RYHyb20	654
chrUn (599363286)	-	-
chrUn (680823997)	RYHyb12	-
chrUn (861182876)	RYHyb02	0 – 24

Male

Rainbow trout genome locus	Chromosome	Position (cM)
chrUn_3 (10025894)*	RYHyb03	-
chrUn_3 (14562667)*	RYHyb03	144
chrUn_11 (6607603)*	RYHyb11	-
chrUn_18 (4330134)*	RYHyb18	-
chrUn (300198084)*	-	-

chrUn (50282385)*	-	-
chrUn (824759953)*	RYHyb21	38 - 52
chrUn (935517313)*	-	-

Table S3.1: Candidate loci linked to reproductive success by Random Forest analysis. Locus names represent the contig and position in parentheses to which RAD loci aligned. For example, “chrUn_14 (11108873)” aligned to contig chrUn_14 at base 11,108,873 in the genomic sequence. Numbers in the contig name (e.g. chrUn_14) represent the *O. mykiss* chromosome to which the locus belongs. Contigs without numbers in the name are not mapped to chromosome. Asterisks designate loci that did not have importance values significantly different than a simulated unlinked locus. Chromosome and position are based on the RYHyb RAD linkage map (Drinan et al. in prep).

Parameter	**Coefficient**	**Standard error**	**z-score**	**p-value**
Intercept	2.44	0.36	6.78	<0.01
Proportion rainbow trout	-2.65	0.41	-6.49	<0.01
Sex (male)	-0.80	0.25	-3.20	<0.01

Table 1: Regression coefficients for generalized linear mixed model used to investigate factors influencing reproductive success in a hybrid population of trout. A zero-inflated negative binomial error distribution was used. Proportion rainbow trout was estimated using 3027 loci.

Female		
Rainbow trout genome locus	Chromosome	Position (cM)
chrUn_14 (11108873)	RYHyb14	382 – 530
chrUn_18 (916565)	RYHyb18	303 – 342
chrUn_18 (9511642)	RYHyb18	-
chrUn_28 (16461881)	RYHyb20	-
chrUn_21 (7080146)	RYHyb21	-
chrUn (372445148)	-	-
chrUn (384507463)	-	-
chrUn (472292170)*	RYHyb20	654
chrUn (599363286)	-	-
chrUn (680823997)	RYHyb12	-
chrUn (861182876)	RYHyb02	0 – 24

Male		
Rainbow trout genome locus	Chromosome	Position (cM)
chrUn_3 (10025894)*	RYHyb03	-
chrUn_3 (14562667)*	RYHyb03	144
chrUn_11 (6607603)*	RYHyb11	-
chrUn_18 (4330134)*	RYHyb18	-
chrUn (300198084)*	-	-
chrUn (50282385)*	-	-
chrUn (824759953)*	RYHyb21	38 - 52
chrUn (935517313)*	-	-

Table S1: Candidate loci linked to reproductive success by Random Forest analysis. Locus names represent the contig and position in parentheses to which RAD loci aligned. For example, "chrUn_14 (11108873)" aligned to contig chrUn_14 at base 11,108,873 in the genomic sequence. Numbers in the contig name (e.g. chrUn_14) represent the *O. mykiss* chromosome to which the locus belongs. Contigs without numbers in the name are not mapped to chromosome. Asterisks designate loci that did not have importance values significantly different than a simulated unlinked locus. Chromosome and position are based on the RYHyb RAD linkage map (Drinan et al. in prep).

Figures

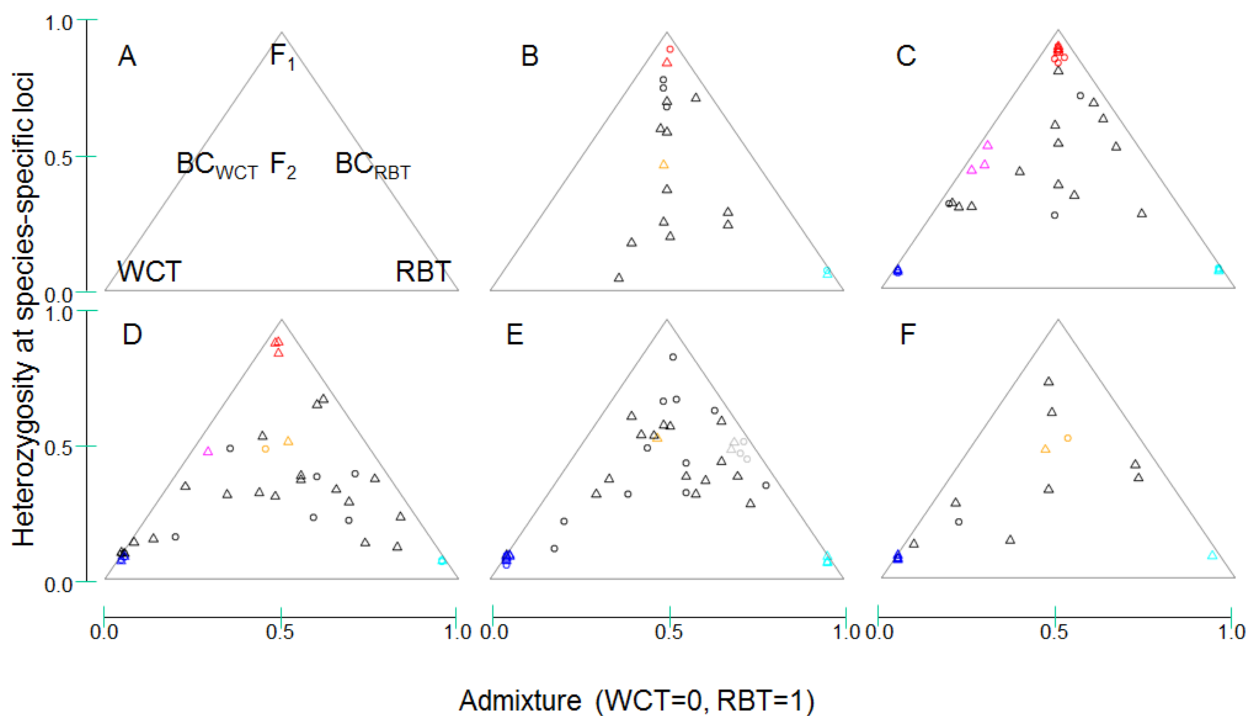


Figure 3.1: Plots used to identify hybrid classes of trout in Langford Creek. Members of *Oncorhynchus* have nearly complete recombination interference, with 90% of meioses resulting in one recombination event per chromosome arm. Early generation hybrid classes are expected to have predictable measures of genome-wide admixture and heterozygosity at species-specific loci depending on their hybrid class. Excepted hybrid classes can be observed in (A). WCT = westslope cutthroat trout, RBT = rainbow trout, BC_{WCT} =

backcross with westslope cutthroat trout, and BC_{RBT} = backcross with rainbow trout. Panels (B), (C), (D), (E), and (F) show spawning adults from 2003, 2004, 2005, 2006, and 2007 respectively. Triangles represent males, while circles represent females. Red = F_1, orange = F_2, pink = BC_{WCT}, grey = BC_{RBT}, blue = WCT, and teal = RBT.

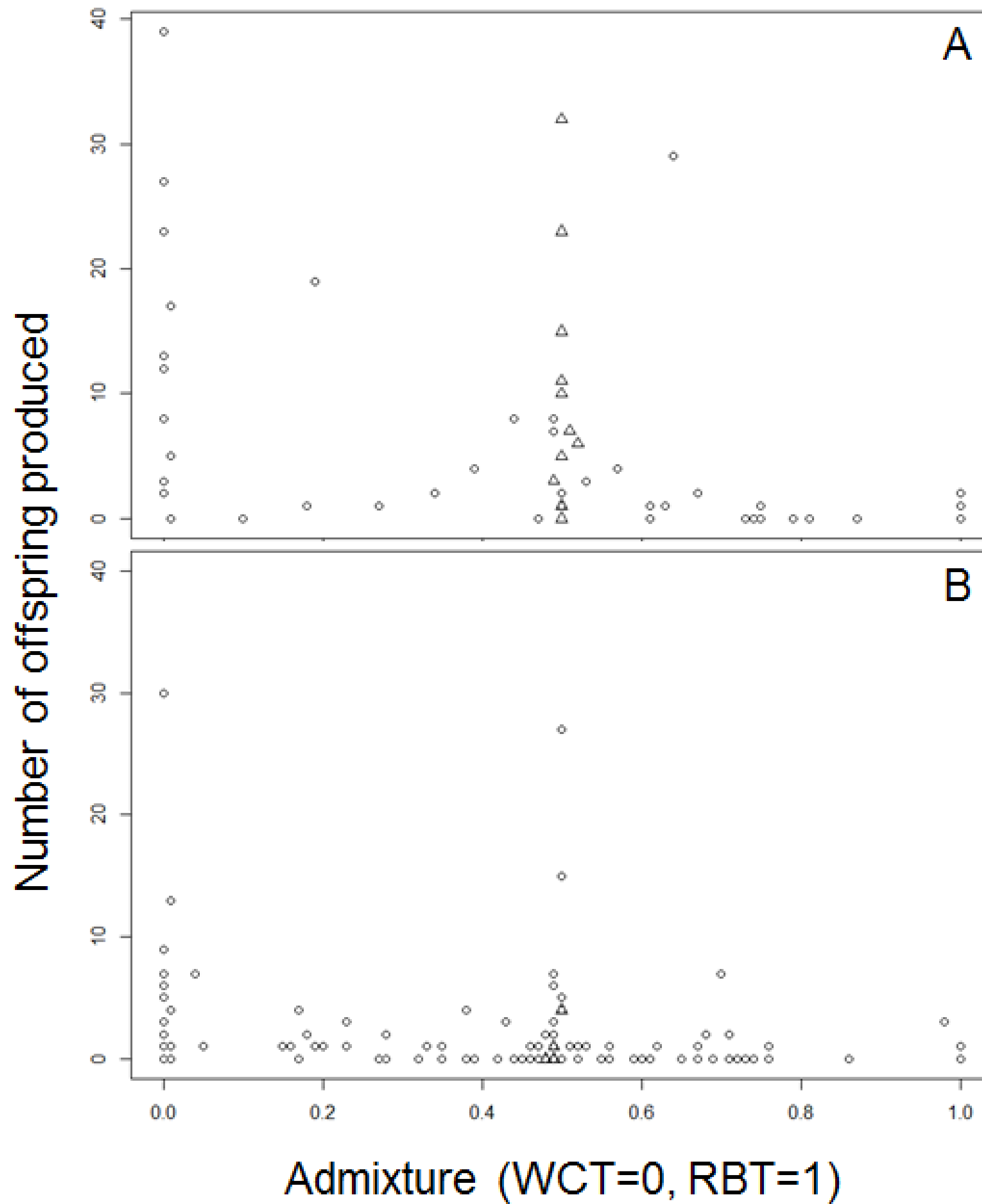


Figure 3.2: Plot of reproductive success for female (A) and male (B) trout in Langford Creek, Montana from 2003 – 2007. Number of offspring produced was estimated by Muhlfeld et al. (2009a). Admixture was estimated using STRUCTURE at 3027 loci. Triangles represent individuals identified as F_1 hybrids. WCT = westslope cutthroat trout and RBT = rainbow trout.

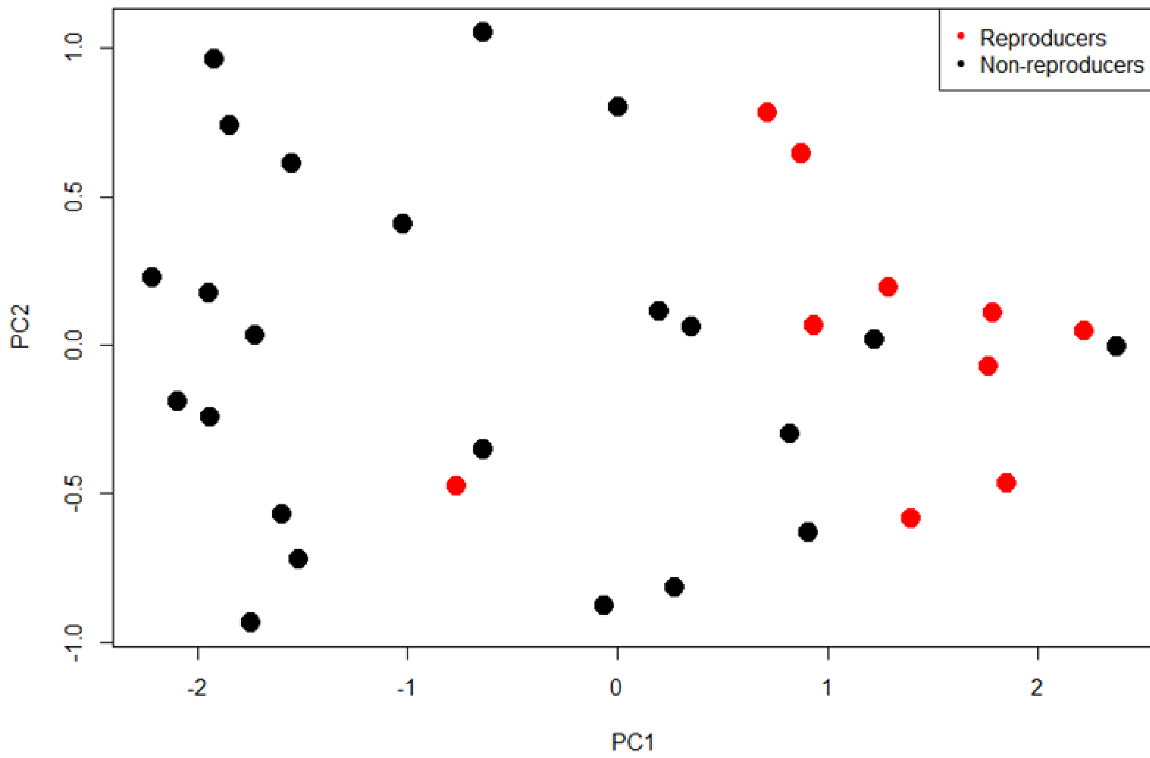


Figure 3.3: Principle components analysis of the 10 loci predictive of reproductive success by Random Forest for female trout. PC1 explained 71% of the phenotypic variation observed, while the second principle component explained only 8%.

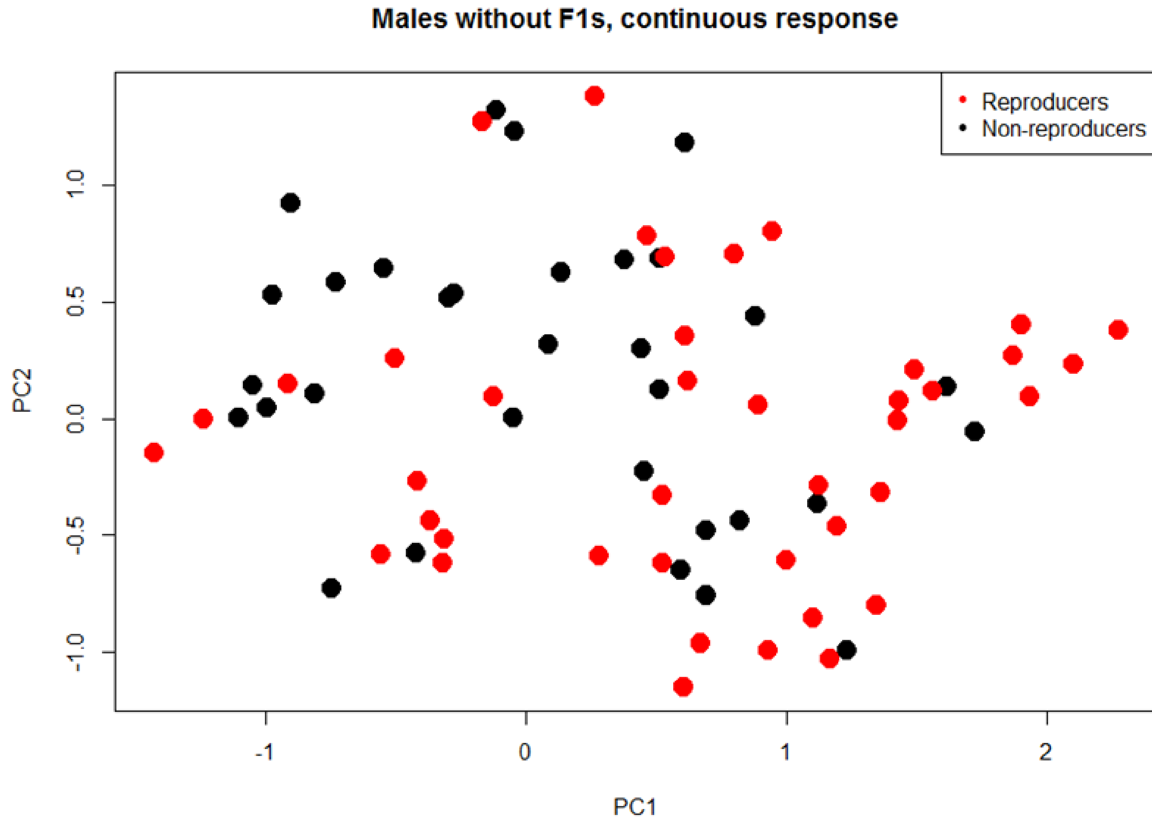


Figure 3.4: Principle components analysis of the 8 loci predictive of reproductive success by Random Forest for male trout. PC1 explained 57% of the phenotypic variation observed, while the second principle component explained only 12%.

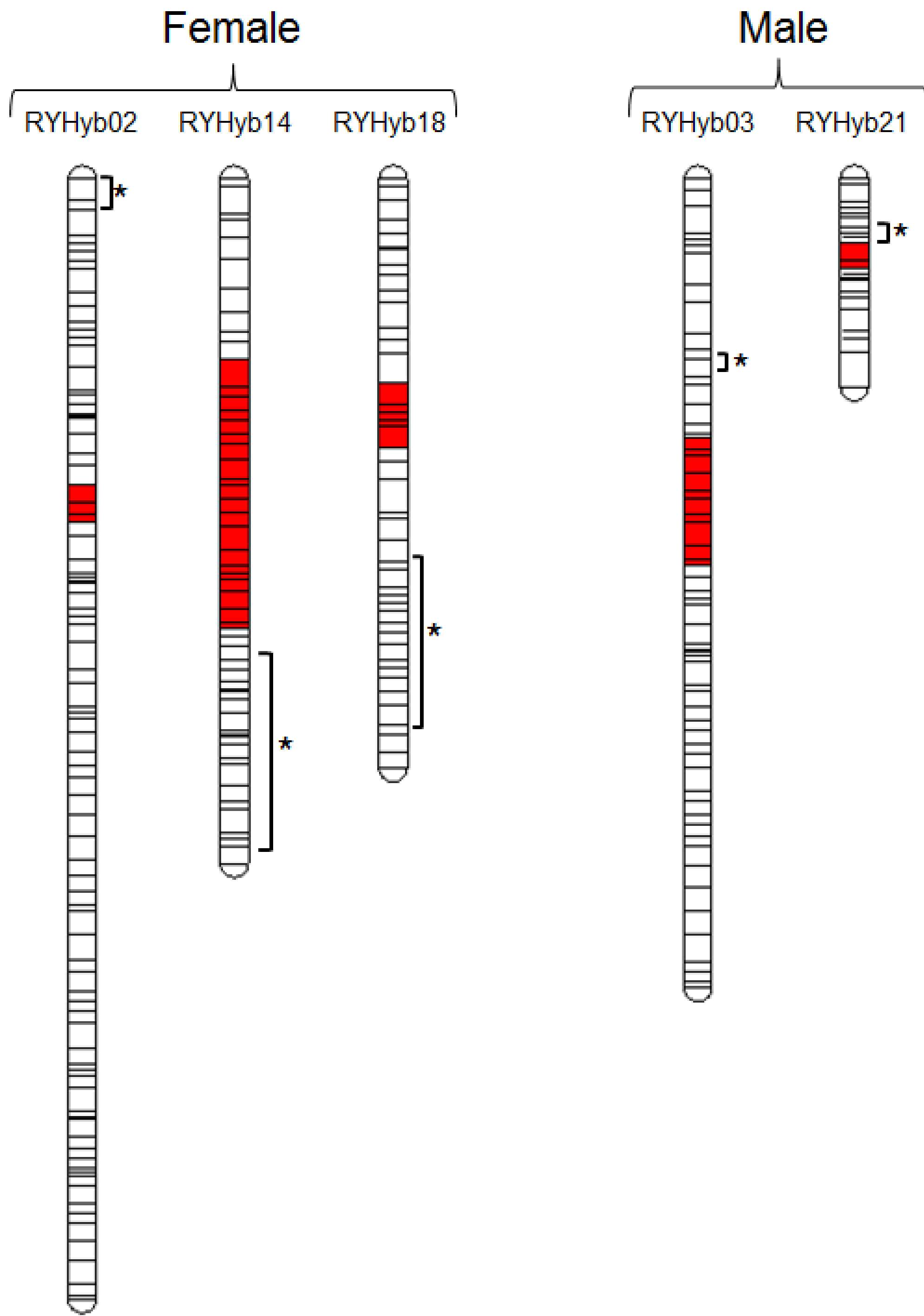


Figure 3.5: Plot of genomic regions linked to reproductive success by Random Forest. Each chromosome contains a series of black lines representing mapped loci from Drinan et al. (in prep). Red regions represent the estimated centromeric region. Brackets show genomic regions

where loci linked to reproductive success by Random Forest are estimated to be based on comparative mapping with other *Oncorhynchus* species.