Jan 24 2022

Swan River Basin Trout Genetics: Assessing Admixture from Rainbow and Yellowstone Cutthroat into Native Westslope Cutthroat Trout

Steve Amish, Beth Gardner[2] & Gordon Luikart, Flathead Lake Biological Station, University of Montana, Polson, MT. [2]US Forest Service, Bigfork, MT.

Executive summary and background

Here we summarize results from the 2021 sampling and SNP genotyping (RAD capture) in 14 sample collections of westslope cutthroat trout (WCT) admixed (e.g., "hybridized") with rainbow trout and Yellowstone cutthroat trout (RBT, YCT) from across the Swan River Basin. We also compare our 2021 admixture proportion estimates to those from two past time points 2011/2012 and 2016 that are each separated by approximately 1-2 trout generations. Thanks to improving technology, we have increased our statistical power by genotyping more loci each time period including 59 species-diagnostic SNPs in 2011/2012, 277 SNPs in 2016, and 1,718 SNPs in 2021. The 1,718 SNPs included 1,070 RBT-diagnostic SNPs, 373 YCT- diagnostics, and 326 WCT-diagnostic loci.

Importantly, in 2016 and 2021, we detected RBT admixture using evidence of contiguous admixture blocks of DNA sequence. These blocks consist of 2-or-more adjacent RBT alleles (SNP alleles) mapped along a chromosome. The identification of blocks of multiple linked RBT-diagnostic alleles (SNPs) was most effective in 2021 because of the larger number of SNPs. In 2021, we also used additional improved data filtering steps that were not available in 2016 to identify (and remove) shared ancestral polymorphisms, which may result in slightly-higher estimates of the percentage of genome-wide WCT ancestry (%WCT alleles) in 2021. Other sources of variation in population-level estimates of %WCT include differences in the time of year and location sampled within a stream, variation in the number of individuals sampled, and variable numbers of admixed immigrant or vagrant fish. Small samples in sites with low admixture and high inter-individual variation in admixture cause variation in %WCT estimates between time points. For example, the presence of a single individual with dozens of nonnative alleles can significantly affect population-level estimates of %WCT when most fish have 100% WCT alleles across their genome.

In 2021, we detected no RBT or YCT alleles at 1718 diagnostic loci in 3 streams: Herrick Run, Dog Creek, and Red Butte Creek. Furthermore, Smith Creek had only 3 nonnative alleles (across the 1718 diagnostic loci), some of which could be shared ancestral alleles (and not from recent admixture). Cold Creek had >5 RBT alleles but only in one individual. Four creeks had 99.5% to 99.9% WCT alleles: Cat, Cedar, Piper, and SixMile Creeks. Bond and Kraft Creeks had a slightly lower percentage of WCT alleles with 96.6%, and 93.7% WCT, respectively. The 2021 sampling included 2 new lakes that were headwaters of two sampled stream populations; Cedar Lake had only ~53% WCT alleles with most of the admixture from YCT (46.5% YCT) and less from RBT (0.5% RBT; Table 1). Piper Lake had 99.7% WCT alleles with most of the non-native alleles from RBT and only 0.07% from YCT.

The typical population admixture pattern consisted of numerous individuals with very low individual admixture and no highly-admixed individuals. However, Bond Creek had two fish with ~50% RBT alleles (suggesting recent admixture) with most fish having only a few small contiguous RBT-admixture blocks,

along with 14 (37%) non-admixed individuals. Kraft Creek also had several individuals with high RBT admixture (%RBT >20-50%).

YCT admixture was highest in Cedar Lake (45.6% YCT as mentioned), 2nd highest in Cedar Creek (0.41% YCT), followed by Bond Creek (0.3% YCT), and Cat Creek (0.17% YCT). Kraft and Piper each only had 0.07% YCT alleles (Table 1).

Temporal changes in the estimated %WCT alleles were small. However, the Kraft Creek %WCT fluctuated from ~94% to 98% and back to 94%, in 2011, 2016, and 2021 (Table 2). This fluctuation could result from occasional immigrants with high admixture or sampling a different proportion of lower versus upper Kraft individuals in different years. A possible decline in %WCT could be occurring in (lower) Cat Creek where the %WCT was 99.5, 99.6, and 99.2, however, more sampling time points are needed to test for a decline. Our results must be interpreted with caution because variation in our %WCT estimates could result from differences in filtering (removal) of loci and variation in sampling between years as mentioned. Continued genetic monitoring is warranted given the risk of admixture spread and climate change. Future sampling downstream of current sites (near mouths of streams), in the main stem of the Swan, and in other streams with higher admixture would improve understanding of admixture and help predict spread of admixture (via modeling). Sampling stream mouths and the mainstem would help identify the origins of highly-admixed immigrants and guide monitoring and management - e.g., ranking the vulnerability of conservation streams and identifying sites for barriers to RBT invasion.

Below, for each stream, we first discuss the 2021 genotyping and admixture results and then discuss temporal changes from two past samplings (2011/2012, and 2016). Finally, we explain reasons to not compute confidence intervals (on admixture point estimates) when quantifying admixture in recently-admixed populations with unclear population boundaries and non-zero migration. As a background note, it's helpful to remember that a WCT-diagnostic locus only detects a "non-WCT" allele as being either from YCT or RBT because the non-WCT allele is in both RBT and YCT (because SNPs have only 2 alleles). This is why we use 3 sets of species-diagnostic loci (WCT-, RBT-, and YCT-diagnostic alleles). Note: we report admixture as the percent (or proportion) of WCT alleles, %WCT (or pWCT). Here, we define an admixed population as anything with %WCT<99.9% (e.g., pRBT > 0.001).

Bond Creek

Of the 38 samples submitted 100% were successfully genotyped. The population-level pWCT was ~ 0.962 with high variation in individual admixture levels among fish; two fish had ~50% RBT alleles suggesting recent admixture, while others had fewer RBT alleles in small contiguous admixture blocks, but 14 (37%) had no RBT alleles. YCT alleles in contiguous admixture blocks were observed in 10 fish (26%). The pattern of RBT admixture across this Bond Creek collection is consistent with either recent hybridization within the population or a mixed sample collection representing 2 or more spawning populations (e.g., the 2-3 high-admixed fish might not be resident spawners). The three individuals with the highest number of RBT alleles were all collected in reach A, the most upstream sample location.

Bond Creek was not sampled as part of this project in 2011 or 2016, making temporal monitoring challenging. However, in 2013 MFWP sampled 29 fish across 2 locations in Bond Creek and reported that Bond Creek may have multiple distinct populations within this single creek. The upper one had slightly more RBT admixture. Both the 2013 and 2021 samples suggest RBT alleles may be coming from

headwater lakes instead of Swan Lake. Future sampling of the headwater lakes would be helpful. Future researchers could also resample Bond, and look for pre-2011 admixture data from MFWP.

Cat Creek

Of the 30 samples submitted, 97% were successfully genotyped and one sample was repeated (with identical results) to ensure data quality. The pWCT ~0.992 reflects an admixed population where small amounts of RBT were clearly detected in most fish (67%) and small amounts of YCT were detected in a few. RBT alleles in contiguous admixture blocks were observed in approximately 20 individuals. Identical YCT alleles in a contiguous admixture block on chromosome omy07 were observed in six fish. The individual Cat_21_076 was sequenced twice as a genotyping control and had a small difference in individual hybrid index scores of 0.004. Individuals with the highest number of RBT alleles were collected in reach C (the downstream of two sampled reaches). Getting samples for Cat Lake may help to identify the source of the admixture.

In 2011, 2016, and 2021, the proportion WCT alleles (pWCT) was ~0.995, 0.996, and 0.992 suggesting variation in admixture estimates in the Cat Creek. This variation may be negligible. However, this potentially-lower pWCT in 2021 suggests future monitoring is warranted. The proportion of non-hybridized individuals in 2011, 2016, and 2021, was 69% (20/29), 44% (23/53), and 33% (10/30), suggesting a decrease in the proportion of non-hybridized individuals - although more loci were sampled each year (potentially) leading to more power to detect low-admixed individuals. The highest RBT- and YCT-admixed individual fish were only 0.019 and 0.001, respectively, showing no individuals with high admixture.

Cedar Creek & Cedar Lake

A total of 76 samples were collected in Cedar Creek and its headwater lake, Cedar Lake; 96% were successfully genotyped. The pWCT ~0.996 for the creek samples reflects the presence of a small amount of YCT and RBT admixture. YCT alleles were observed in only 16 of the Cedar Creek individuals. The maximum individual admixture was less than ~3% (see Table 1, column "HI_YCT" = 0.03). The pWCT of ~0.530 in the lake sample collection reflects the presence of extensive YCT admixture and a small amount of RBT admixture in the majority of individuals. Six fish from the lake had pYCT >0.8 (>80% alleles from YCT, Table 1) and 11 other fish had individual pYCT > 0.5. In addition, approximately 12 fish had small RBT admixture blocks. Ten of the 12 fish with RBT admixture blocks were from the Cedar Lake collection. The high YCT admixture in the lake, with some individuals with >0.99 pYCT suggests the lake is a source of spread of YCT admixture downstream into Cedar Creek and perhaps across the Swan River system.

In Cedar Creek in 2011, 2016, and 2021, the proportion of WCT alleles (pWCT) was 0.998, 0.995, and 0.9958, suggesting stability or a possible slight decline in the proportion of WCT alleles (pWCT; Table 2). The Cedar Creek pYCT in 2012 and 2016 was 0.001, and 0.041 suggesting a possible very slight increase in admixture from Yellowstone cutthroat trout. In Cedar Creek, the proportion of non-hybridized individuals in 2011, 2016, and 2021 was 9/10 (90%), 27/35 (77%), and 28/44 (66%). In 2016 and 2017 the individual with the highest admixture (max pRBT) was 0.024 and 0.003, respectively, suggesting rainbow trout admixture is not increasing. In 2016 and 2017 the individual with the highest admixture (max pYCT) was 0.01 and 0.03, respectively. The slight increase in YCT admixture for multiple statistics (in and among individuals) could be caused by possible downstream movement (leakage) of YCT-

admixed individuals from Cedar Lake above. Population genetic assignment (of individuals) could help test for possible downstream movement of admixed in individuals.

Dog Creek

Of the 34 samples submitted, 100% were successfully genotyped. The pWCT ~1.00 reflects a non-admixed WCT population. The 6 RBT alleles detected were not present in contiguous blocks (Fig. 1) and may represent an ancestral polymorphism (APM), sequencing errors, or very low levels of historic admixture.

In 2011, 2016, and 2021, the proportion WCT alleles (pWCT) was 0.999, 0.998, <0.999, suggesting no evidence of increased admixture since at least 2011.

Interestingly, Dog Creek has NO barrier. It is actually connected to Cat Creek, which does have some introgression. Cat Creek has a headwater lake, but Dog does not. Yet the Cat Creek data suggest introgression might be coming from downstream, not the lake. Future additional sampling up and downstream could determine the source or direction of influx of the non-WCT admixture.

Herrick Run

Of the 90 samples submitted, only 42% were successfully genotyped resulting in data for just 38 individuals. This was atypical in that all other collections had >94% genotyping rate. The pWCT was ~1.00 (proportion of WCT alleles observed across all samples) with only 1 RBT allele and 0 YCT alleles detected across all individuals. No evidence of contiguous admixture blocks (2 or more adjacent RBT alleles on a chromosome). Quality assessment of the DNA extractions suggested moderate to high DNA concentrations with no evidence of contaminants that might impede sequencing. DNA from 52 samples was genotyped a second time and the results were similar to the initial run.

There was no change in pRBT from past samples, as expected because Herrick Run is isolated. From 2011 samples, Herrick Run had only one individual with one putative nonnative allele (among 114 diagnostic alleles genotyped) among 34 individuals sampled, which represented 99.97% WCT alleles.

The cause of the unsuccessful genotyping is difficult to assess. It could be poor preservation of field samples (insufficient amount or concentration of ethanol) or contamination of field samples with a PCR inhibitor. Also, contaminants can cause RAD sequencing library prep to fail (e.g., restriction or ligation enzyme reactions are sensitive to contaminants). The problem seems more likely to be with sampling (initial preservation) than DNA extraction or PCR (in the lab) because the samples were extracted and PCR-ed side by side with other samples that yielded genotypes. DNA quality (seemed good) as lab measurements (nanodrop and cybergreen) showed that DNA was present with no protein or polysaccharide contamination. The Herrick Run samples that did not genotype were in the same 96-well plates that yielded genotypes. Re-genotyping with new RAD capture and DNA sequencing library preparations did not yield results (genotypes) for the samples with missing genotypes.

Kraft Creek

Of the 37 samples submitted, 100% were successfully genotyped and one sample was repeat genotyped to ensure data quality. The low proportion of WCT (pWCT ~0.937) compared to other streams was due mostly to RBT admixture. RBT alleles in contiguous admixture blocks were observed in approximately 18 individuals, YCT alleles in contiguous admixture blocks were observed in only 3

individuals, while no RBT alleles were observed in 17 individuals. The pattern of admixture across the sample collection is consistent with either recent admixture within the population or a mixed sample collection representing 2 or more spawning populations. Individuals with the highest number of RBT alleles were collected at Lower Kraft (aka Reach B). The highest-admixed individual had 58% RBT alleles (Table 1). Sample Kraft_21_473 was sequenced twice as a genotyping control and had identical RBT hybrid index scores.

In 2011, 2016, and 2021, the proportion of WCT alleles (pWCT) in Upper Kraft was ~0.94, 0.98, and ~0.94. This is by far the largest temporal change in admixture estimates in this study. The fluctuations could result in part from migration of highly-admixed individuals and from sampling different proportions of Upper and Lower Kraft in different years. In 2016, three individuals sampled from (Upper) Kraft Creek had a high proportion of RBT alleles (pRBT = 0.27, ~0.20, ~0.15), while in 2021 an individual had pRBT = 0.57, suggesting highly-admixed individuals could be arriving at Kraft from somewhere with highly-admixed fish (e.g., downstream). Importantly, Lower Kraft Creek in 2011 showed far higher admixture (pWCT = 0.88) than Upper Kraft (pWCT=0.94, Table 2 footnote), which is consistent with the hypothesis that highly-admixed fish arrive from downstream. Upper Kraft Creek could have lower non-native admixture because it's farther from the mainstem which has higher RBT-admixed fish. Lower Kraft is 2.23 kilometers (2,230 m) downstream of Upper Kraft (aka Reach A). Additional data analyses and the sampling of more time points are needed to understand causes of the admixture fluctuations. To understand the origins of this admixture, and if admixture is increasing, additional research is needed including more sampling near the outlet from Kraft and along the Swan River.

Owl Creek

Of the 18 samples submitted, 94% were successfully genotyped. The pWCT ~0.999 reflects the presence of a very small amount of RBT and YCT admixture in 2 fish. A potential RBT admixture block in individual Owl_21_574 could be re-sequenced to confirm data given its 35% missing genotypes. A contiguous YCT admixture block was detected in individual Owl_21_561 and was confirmed based on a contiguous non-WCT admixture block in the same chromosome region using the WCT-diagnostic SNP genotypes. A larger sample size would help confirm the current extent of admixture in Owl Creek given the low admixture and high variance among individuals in this collection. However, Owl Creek apparently has very low numbers of WCT making larger sample sizes difficult to obtain.

In 2011, 2016, and 2021, the proportion WCT alleles (pWCT) was 0.997, 0.995, and 0.999. In 2011, 24 individuals were sampled and five individuals had 1-3 nonnative alleles (out of 114 alleles from the 57 loci on the SNP chip). However, in 2016 only seven individuals were sampled and 6 individuals had a few RBT alleles. No Yellowstone admixture was detected in 2011 or 2016 and very little in 2021 (pYCT = 0.0005). Overall, the limited results here are consistent with no increase in non-native admixture evident in the Owl Creek sampling area between 2011 and 2021.

Piper Creek & Lake

Of the 56 samples submitted, 100% were successfully genotyped. The pWCT ~0.995 in the creek sample collection reflects a low RBT admixture in all but 3 of the 31 individuals. The pWCT ~0.997 in the lake sample collection reflects RBT admixture in approximately half of the individuals, with 12 of the 25 individuals having no RBT alleles. Contiguous blocks of YCT admixture were observed in 2 individuals from Piper Lake. Overall, less RBT admixture (~0.001 vs ~0.005) was observed in the Piper Lake samples

than in the Piper Creek samples. Since lower admixture and fewer hybridized individuals were observed in the lake collection, this difference could be biologically significant and is consistent with the creek being the source of RBT alleles.

In 2011, 2016, and 2021, in Piper Creek, the proportion of WCT alleles (pWCT) was 0.99, 0.994, and ~0.995. Lower Piper was sampled in 2011 and had slightly more admixture (pWCT = 0.95) consistent with RBT alleles and introgression arriving from downstream. In 2011, 2016, and 2021, 59% (17/29), 32% (13/41), and 35% (11/31) of individuals were non-hybridized however more diagnostic loci were used each time likely giving more power to detect more individuals with very low admixture. Overall, results suggest no increase in population-level non-native admixture in the Piper Creek sampling area since 2011. No clear difference in the pRBT exists between 2021 and 2016 RAD capture data. Additional sampling would help determine if RBT alleles are spreading among more individuals and increasing in frequency (pRBT).

Red Butte Creek

Of the 40 samples submitted, 98% were successfully genotyped. The pWCT ~1.00 reflects a non-admixed WCT population. The 9 RBT alleles detected were not present in contiguous blocks and may represent shared ancestral alleles (polymorphisms), sequencing errors, or very low levels of admixture. One individual, RedB_21_781, was sequenced twice as a genotyping control and had a difference in RBT hybrid index scores of less than 0.0001.

In 2011, 2016, and 2021, the proportion WCT alleles (pWCT) was 1.00, 0.992, and 1.00. It's possible the 2016 putative-RBT alleles were ancestral alleles (transmitted from WCT ancestors) and not from admixture. Results suggest very little or no admixture and no increase in non-native admixture in Red Butte Creek. No increase in admixture is expected since this stream was intentionally isolated in 2014.

Sixmile Creek

Of the 36 samples submitted, 97% were successfully genotyped and one sample was repeated. The pWCT ~0.999 reflects very low non-WCT admixture. No RBT or YCT alleles were detected, but an identical block of 2 non-WCT alleles on omy17 was observed in 9 individuals. There are no YCT or RBT diagnostic SNPs in this region of omy17 so we were unable to confirm the presence of the admixture block or identify species of origin. RBT admixture had been previously detected in Sixmile Creek WCT samples (from 2016). Individual SixM_21_913 was sequenced twice as a genotyping control and had identical RBT hybrid index scores.

In 2011, 2016, and 2021, the proportion WCT alleles (pWCT) was 1.000, ~0.99, and 0.999. In 2016, lower and upper Sixmile (2,400 meters apart) had pWCT of 0.985 and 0.991 suggesting that, if nonnative alleles exist, they are likely downstream near the barrier dam. Future research including genotyping of diagnostic loci from other trout taxa could help resolve the origins of potential nonnative admixture in Sixmile.

Smith

Of the 36 samples submitted, 94% were successfully genotyped. The pWCT ~0.999 could reflect very limited non-WCT admixture. The 3 RBT alleles (SNPs) detected were not present in a contiguous block and may represent a shared ancestral allele (allele transmitted to both species from the common

ancestor), sequencing errors, or very low levels of historic admixture. In individual Smith_21_844, a contiguous block of 2 non-WCT alleles was observed. We could not confirm whether this admixture block was from RBT due to missing data in RBT diagnostic genotypes. The presence of non-WCT admixture in only a single fish may reflect a recent immigrant, nonrandom sampling of the population, insufficient sampling of the population, or that the collection represents multiple spawning populations. A rerun of individuals with high missingness could confirm genotypes and the presence of a non-WCT admixture block. Additional sampling of Smith Creek and downstream individuals could also help resolve the frequency and origin of this admixture block.

In 2011, 2016, and 2021, the proportion WCT alleles (pWCT) was 1.00, ~0.99, 0.999. There is no clear evidence of admixture or an increase in admixture in Smith Creek. This is important as a barrier was installed a barrier in 2016.

South Cold Creek

Of the 26 samples submitted, 100% were successfully genotyped. The pWCT ~0.998 reflects the presence of RBT admixture in only 1 fish. The 2 YCT alleles detected were not present in a contiguous block and may represent either an ancestral polymorphism (APM; alleles shared by RBT and WCT), sequencing errors, or very low levels of historic admixture. The RBT admixture block in sample ColdS_21_941 on omy07 was confirmed by the detection of an overlapping admixture block in nonWCT alleles with the WCT diagnostic loci. The presence of RBT admixture in only a single fish may reflect a recent immigrant (or transient fish), insufficient sampling of the population, or inclusion of multiple spawning populations represented in our collection. In 2011 and 2016, we had no samples or data.

Difficulties associated with confidence intervals around admixture estimates

We did not compute confidence intervals (CI's) for sample collection because potential violations of genetic assumptions and geographic origins of individuals (e.g., random distribution of RBT alleles) make interpreting CI's difficult and possibly misleading. It's better to look at multiple summary statistics (e.g., pRBT, individual max pRBT, % of non-admixed ("pure") individuals) and their change over time. Specifically, the use of confidence intervals to compare population estimates of admixture (e.g., pRBT) is problematic due to the non-random distribution of non-native alleles across the genome, and the non-random distribution of admixed individuals within sample collections (Della Croce et al. 2016). At the genomic level, nonWCT alleles in individuals back-crossed to the parental species will result in blocks of non-native alleles in limited regions of the genome being non-randomly distributed. This prohibits bootstrapping across loci to generate confidence intervals. Similarly, at the sample collection level, the presence of non-admixed individuals, or the presence of a few highly admixed individuals, results in a highly skewed distribution of individual hybrid indices, that suggest the sample collection does not represent a single population (e.g., some individuals might be non-breeding transients with high admixture (pRBT). In these cases, confidence intervals will not be very accurate because the sample collection is not representative of a single population. The most obvious example of this would result in the misleading inclusion of zero (pRBT = 0) in a confidence interval for a sample collection that contains a few clearly hybridized individuals.

Acknowledgements: We thank Jared Grummer for helpful suggestions and editing this report. We are especially thankful for sample collection by Swan Valley Connections (Andrea Stephens, Luke Lamar, Sara Lamar), Montana FWP (Leo Rosenthal, Scott Haxhurst), Flathead National Forest (Dustin Allen, Zach Morris, Gershom Bigham, Sarah Caffrey, Emily Miller, Jerick Graves), University

of Montana – Flathead Lake Biological Station (Tyler Tappenbeck, Ben Triano), MPG Ranch (Beau Larkin), and numerous citizen volunteers. Funding provided by the Southwest Crown of the Continent Collaborative."

Table 1. Estimates of non-native admixture for 14 sample collections (streams & 2 lakes) in 2021. Waterbody name, population abbreviation with sample year (Pop_year), drainage, species, and region are listed for each sample collection. Sample collection details include the number of individuals submitted for genotyping (N_seq), the number of samples successfully genotyped (Pop_year_N), the genotyping rate or proportion of submitted samples successfully genotyped, and the number of missing samples (that failed to genotype). Admixture summary metrics are reported as pINV= sum pRBT + pYCT or pNonWCT, whichever is greater; pWCT = 1 – pINV. For RBT, YCT, & WCT diagnostic loci, Pop_n = number of samples, RBT_tot = total number of alleles observed, pRBT = proportion of alleles which are RBT, avgHIS = average hybrid index score for individuals, sdHS = standard deviation hybrid index, max_HI = max hybrid index score, F1 = count of first generation hybrids, RBT_0 = count of individuals with 0 RBT alleles, NAtot = total number of missing genotypes, perc_missing = percent missing data for the sample collection.

Swan & Herrick Run 2021 WCT RAD capture

Summary Table of RBT, YCT, WCT admixture in all samples v3

Summary metrics:

pINV= sum pRBT + pYCT or pNonWCT whichever is greater

pWCT = 1 - pINV

For RBT, YCT, & WCT diagnostic loci the following data is shown:

Pop_n = number of samples, RBT_tot = total number of alleles observed, pRBT = proportion of alleles which are RBT, avgHIS = average hybrid index score for individuals, sdHS = standard deviation hybrid index, max_HI = max hybrid index score, F1 = count of first generation hybrids, RBT_0 = count of indivduals with 0 RBT alleles, NA_tot = total number of missing genotypes, perc_missing = percent missing data for the sample collection.

Total # loci

1719 **Species diagnostic SNPs used to evaluate WCT hybridization**

														WCT Diagnostic SNPs 326 Diag_loci_nonWCT									**RBT Diagnostic SNPs 1070 Diag_loci_RBT**										
Waterbody	**Pop_year**	**Sample_Year**	**Drainage**	**Species**	**Region**	**N_Seq**	**Pop_year_N**	**Genotyping_rate**	**N_Missing**	**pWCT**	**pRBT_pYCT**	**pINV**		**Pop_n_nonWCT**	**nonWCT_tot**	**pNonWCT**	**avgHIS_nonWCT**	**sdHIS_nonWCT**	**max_HI_nonWCT**	**F1_nonWCT**	**nonWCT_0**	**perc_missing_nonWCT**		**Pop_n_RBT**	**RBT_tot**	**pRBT**	**avgHIS_RBT**	**sdHIS_RBT**	**max_HI_RBT**	**F1_RBT**	**RBT_0**	**perc_missing_RBT**		**Pop_n_YCT**	**YCT_tot**
Herrick Run	HR_21	2021	Swan River	WCT	MT	90	38	0.42	52	1.0000	0.0000	0.0000		38	0	0.0000	0.0000	0.0000	0.0000	0	38	1.2		38	1	0.0000	0.0000	0.0001	0.0006	0	37	1.9		38	0
Bond Cr	Bond_21	2021	Swan River	WCT	MT	38	38	1.00	0	0.9612	0.0388	0.0388		38	626	0.0335	0.0336	0.1029	0.4345	0	15	6.0		38	2231	0.0356	0.0357	0.1147	0.4889	0	15	8.2		38	63
Cat Creek	Cat_21	2021	Swan River	WCT	MT	30	30	1.00	0	0.9917	0.0083	0.0083		30	105	0.0074	0.0074	0.0064	0.0239	0	3	9.7		30	313	0.0066	0.0065	0.0053	0.0186	0	1	12.2		30	25
Cedar Creek	Cedar_21	2021	Swan River	WCT	MT	45	44	0.98	1	0.9958	0.0042	0.0042		44	82	0.0038	0.0036	0.0061	0.0285	0	25	6.9		45	12	0.0002	0.0002	0.0005	0.0025	0	38	9.7		45	90
Cedar Lake	CedarLk_21	2021	Swan River	WCT	MT	31	30	0.97	1	0.5296	0.4704	0.4704		28	5596	0.4181	0.4219	0.3147	0.8731	1	4	8.6		29	234	0.0050	0.0048	0.0140	0.0741	0	12	10.0		28	6350
Dog Creek	Dog_21	2021	Swan River	WCT	MT	35	34	0.97	1	0.9999	0.0001	0.0001		35	0	0.0000	0.0000	0.0000	0.0000	0	35	1.9		35	6	0.0001	0.0001	0.0003	0.0012	0	30	1.5		35	0
Kraft Creek	Kraft_21	2021	Swan River	WCT	MT	37	37	1.00	0	0.9438	0.0562	0.0562		37	868	0.0505	0.0477	0.1023	0.5253	0	21	11.1		37	3171	0.0556	0.0514	0.1130	0.5749	0	18	13.9		37	12
Owl Creek	Owl_21	2021	Swan River	WCT	MT	18	17	0.94	1	0.9993	0.0007	0.0007		17	3	0.0004	0.0004	0.0011	0.0039	0	15	12.5		17	5	0.0002	0.0002	0.0005	0.0017	0	13	16.6		17	4
Piper Creek	Piper_21	2021	Swan River	WCT	MT	31	31	1.00	0	0.9945	0.0055	0.0055		31	70	0.0054	0.0053	0.0065	0.0224	0	11	19.7		31	233	0.0055	0.0055	0.0059	0.0210	0	3	22.7		31	1
Piper Lake	PiperLk_21	2021	Swan River	WCT	MT	25	25	1.00	0	0.9973	0.0019	0.0027		25	30	0.0027	0.0027	0.0041	0.0160	0	14	14.6		25	43	0.0012	0.0011	0.0016	0.0047	0	12	18.2		25	8
Red Butte Creek	RedB_21	2021	Swan River	WCT	MT	40	40	1.00	0	0.9999	0.0001	0.0001		40	0	0.0000	0.0000	0.0000	0.0000	0	40	2.6		40	7	0.0001	0.0001	0.0002	0.0006	0	33	3.4		40	0
Sixmile Creek	SixM_21	2021	Swan River	WCT	MT	36	36	1.00	0	0.9990	0.0000	0.0010		36	19	0.0010	0.0010	0.0017	0.0039	0	26	1.9		36	0	0.0000	0.0000	0.0000	0.0000	0	36	2.5		36	0
Smith Creek	Smith_21	2021	Swan River	WCT	MT	36	34	0.94	2	0.9999	0.0000	0.0001		36	2	0.0001	0.0001	0.0009	0.0052	0	35	13.4		36	1	0.0000	0.0000	0.0001	0.0006	0	35	16.8		36	0
South Cold Creek	ColdS_21	2021	Swan River	WCT	MT	26	26	1.00	0	0.9987	0.0013	0.0013		26	14	0.0011	0.0010	0.0053	0.0269	0	25	5.4		26	51	0.0012	0.0011	0.0056	0.0285	0	25	7.0		26	2

Table 2. Estimated percent WCT alleles (%WCT) in Swan Valley streams at three time points: 2011/12, 2016, and 2021. Time points are separated by approximately 1-2 WCT generations. The number of species-diagnostic SNPs that we genotyped increased across time points from 57, 277, to 1718 SNPs thanks to improving genotyping technology. Two decimal places are reported mainly for the 2021 estimates of %WCT because the larger number of loci (1718) provided more precision. For the seven streams sampled at three time points, the time point with the lowest pWCT estimate is highlighted yellow (below). See the Executive Summary for a discussion of possible sources of variation in %WCT estimates over time. Note: The three population estimates below 96% WCT are in **bold** and are from Cedar Lake and Kraft Creek. Among the 13 collections in 2021, only Cat Creek had a lower estimated %WCT ("purity") compared to 2016 (99.20 down from 99.6). Most changes in %WCT between 2011/12 and 2016 are small and potentially insignificant biologically. For more information on the 2011/12 and 2016 data, see our past reports.

Sample Collection	**2011/2012** %WCT (indiv-min)	**2016** %WCT (indiv-min)	**2021** %WCT (indiv-min)
Herrick Run Creek	99.9		100.00
Bond Creek			96.10 (96.5)
Cat Creek	99.5	99.6 (97.4)	99.20 (97.6)
Cedar Creek – Lower & Upper	^99.5	99.5 (97.6)	99.58 ([b]97.1)
Cedar Lake			[a]**53.0** (22.7)
Cold Creek - S. Fork		99.6 (96.4)	99.87 ([**]97.31)
Cooney Creek		99.4 (96.2)	
Dog Creek	99.9	99.7 (97.2)	99.99 (0)
Owl Creek	99.7 (97.4)	99.3 (97.3)	99.93 (99.6)
Piper Creek	*99 (97.4)	99.4 (96.8)	99.45 (97.8)
Piper Creek -Lower	96 (94.8)		
Piper Lake			99.7 (98.4)
Red Butte Creek	100	99.2 (94.3)	99.99 (99.99)
SixMile Creek		98.5 (95.8)	99.90 (99.6)
SixMile Creek -Upper	100	99.1 (95.1)	
Smith Creek	100	98.8 (60.3)	99.99 (99.5)
Soup		98.8 (95.3)	
Sunset (Beaver Lake Outlet)	99.4	98.9 (96.4)	
S. Fork Cold Creek		99.6 (96.4)	99.87 (97.31)
Woolf Creek	100		
Groom Creek	100	99.7 (96.6)	
Groom Creek – Upper		99.5 (96.7)	
Lion Creek	100	0.0 (0.0)	
Kraft Creek – Upper + Lower	**93.6**^^	^^98.2 (71.7)	**94.38** (47.0)
Upper Swan River	99.8 (95.6)	99.8 (97.9)	
Pony	99.3	99.2 (96.3)	
N. Fork Lost Creek	98	97.8 (93.6)	
S. Fork Lost Creek - Lower		99.1 (96.4)	
S. Fork Lost Creek - Upper		98.8 (97.5)	
Whitetail		99.8 (99.3)	

^The 2011 Cedar Creek sample collection included Upper Cedar Creek (%WCT=0.998) and Lower (%WCT=0.990) with an average of ~0.995. Future research could test if upper and lower sampling sites are genetically differentiated demes (populations). In 2011, the highest-admixed fish had 98% WCT alleles in Lower, and 93% WCT alleles in Upper Cedar Creek.

^^Upper and Lower Kraft was sampled in 2011 and had %WCT or 99.2% and 88%, respectively; we report the mean here (93.6). One individual in 2011 in Lower Kraft had 44% RBT alleles suggesting recent immigration of a highly-admixed individual, similar to 2021 where an individual had 57% RBT alleles.

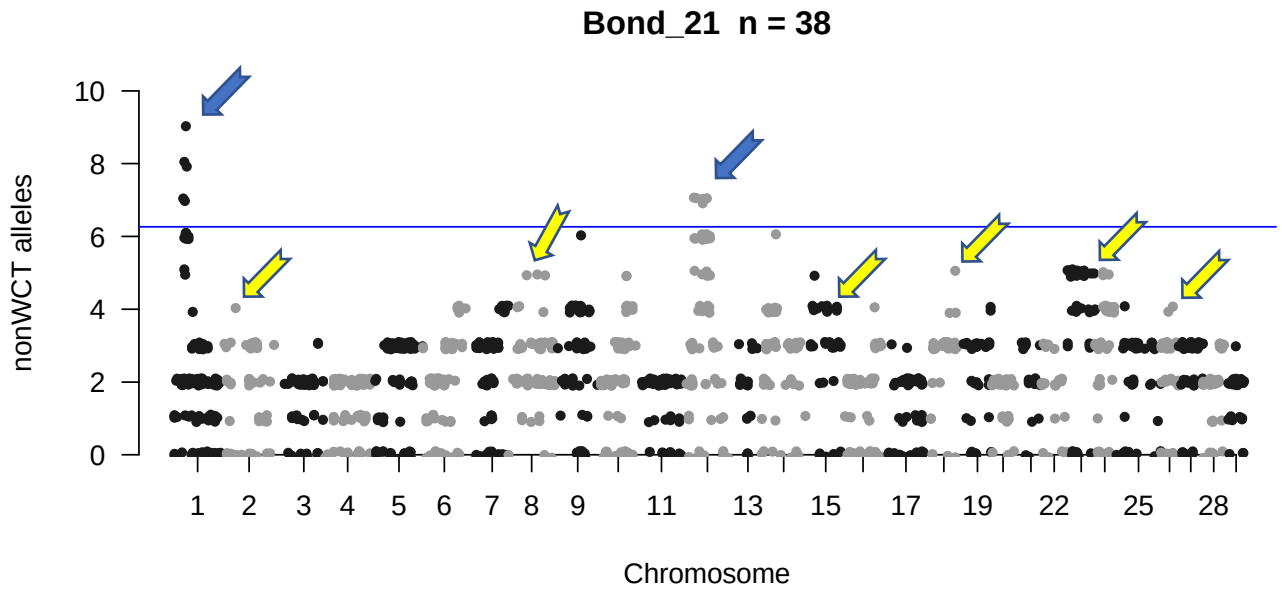
*The 2011 Piper Creek sample was reported as from "Upper Piper" in our 2011/12 report.

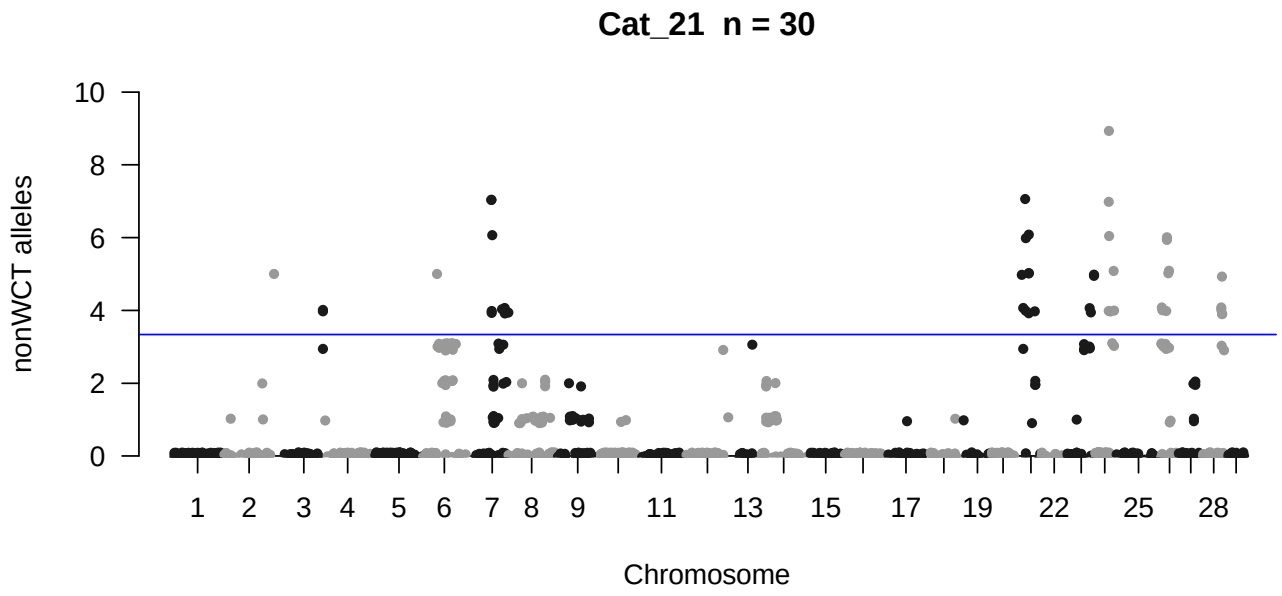
[**]only one individual had non-native alleles: 5 total including 4 from RBT and one from YCT.

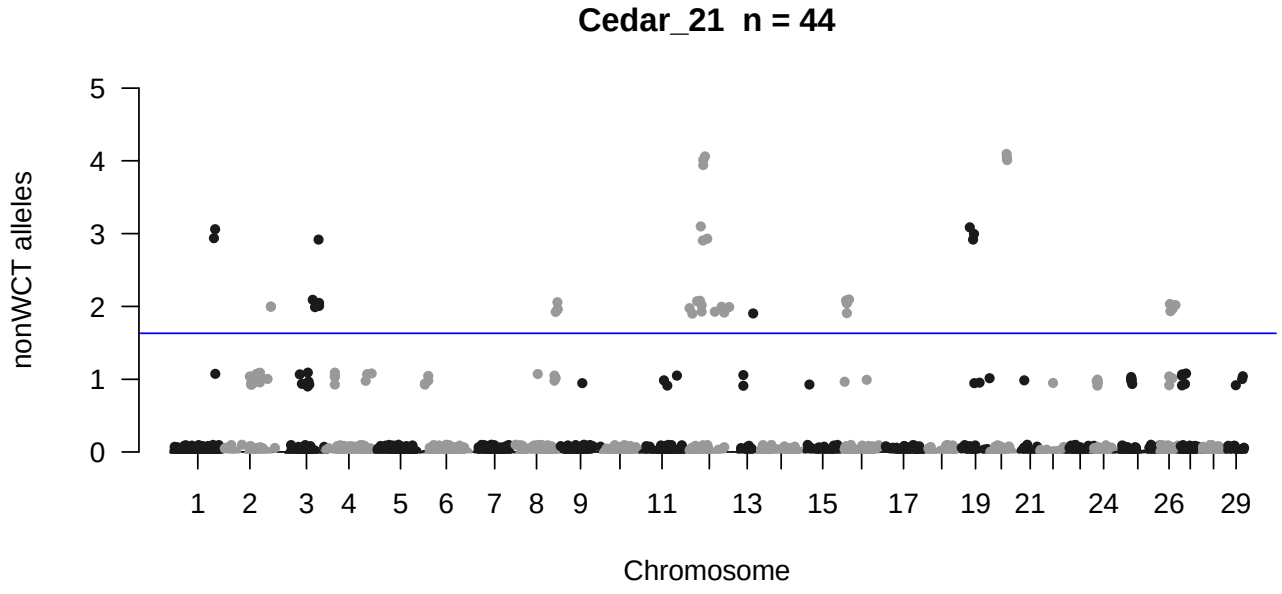
[a] Yellowstone admixture is high (pYCT=46.5) in Cedar Lake; admixture from RBT is only 0.5% (pRBT=0.005).

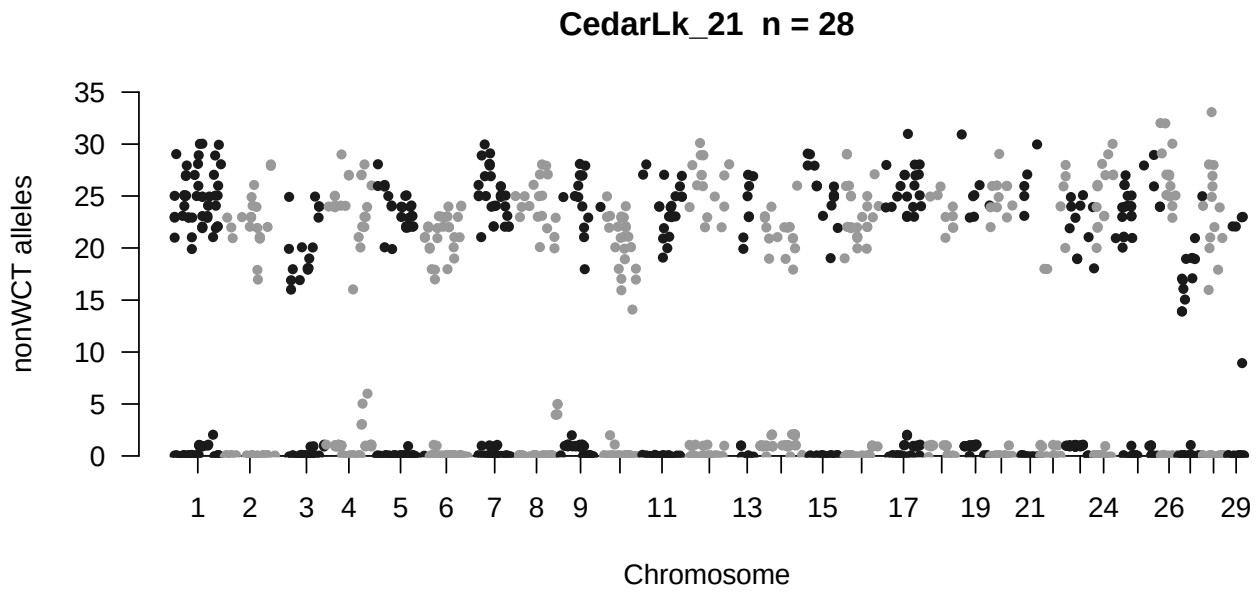
[b] This highest-admixed fish in Cedar Creek had 2.85% non-WCT alleles (max_HI_nonWCT = 0.0285, Table 1).

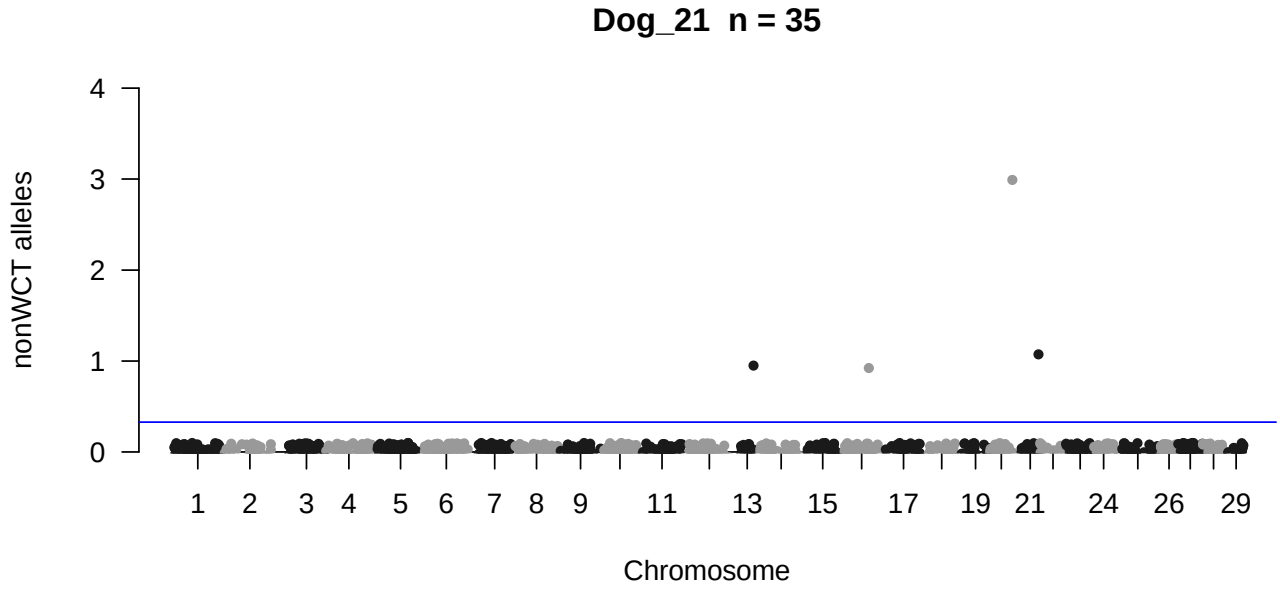
Figure 1. Manhattan plots showing the chromosomal location (x-axis) of all diagnostic SNPs (dots) mapped across the rainbow trout genome for sample collections from 2021. The x-axis shows the 29 chromosomes alternating in color grey and then black. Dots are SNP alleles from non-WCT taxa. Black dots without a red or yellow outline reflect alleles detected at WCT diagnostic loci. Red or yellow outlines around the dots indicate alleles from RBT or YCT diagnostic loci respectively. Y-axis shows the count of non-WCT alleles among individuals at each genomic location (dot) in the sample collection (population); "n" is the number of individuals genotyped. The maximum possible number of alleles at a locus is 2x the number of individuals because individuals are diploid. The blue horizontal line equals the mean + 3SD of the non-WCT allele count among all diagnostic loci. Thus, a dot above the blue line is an outlier region with a non-WCT allele count far higher than the mean count genome-wide. Bond Creek (first panel) shows two blocks of non-WCT alleles in outlier regions (blue arrows) along with many other non-WCT SNPs at relatively high allele counts (frequency) across the genome (yellow arrows). At the first blue arrow on the Bond Creek plot, the mix of black dots and black dots with a yellow outline in close proximity on chromosome 1 indicate alleles detected at both WCT and YCT diagnostic loci. A broader distribution of dots across the x-axis represents a broader distribution of outlier (high count) non-WCT alleles across the genome appears in Cat Creek, 2nd panel below. Panel 4, Cedar Lake, has the highest counts of non-WCT alleles (e.g., 20-30) at many loci genome-wide. In Cedar Lake, the blue line is not shown because the value exceeds the y-axis. The 6th panel "HR_21" (Herrick Run), has a blue arrow showing a single non-WCT allele ("RBT diagnostic allele") on Chromosome 11; such a pattern with only a single non-WCT SNP (not a block of SNPs) could be a shared ancestral allele and not a true RBT-diagnostic allele (thus a 'false-positive' admixture signal). See Table 1 for population abbreviations ("Pop_year" column). Note: figures like those below are available (upon request) with black dots (loci) colored yellow for YCT alleles and red for RBT alleles. Mainly Cedar Lake and Creek have YCT alleles.

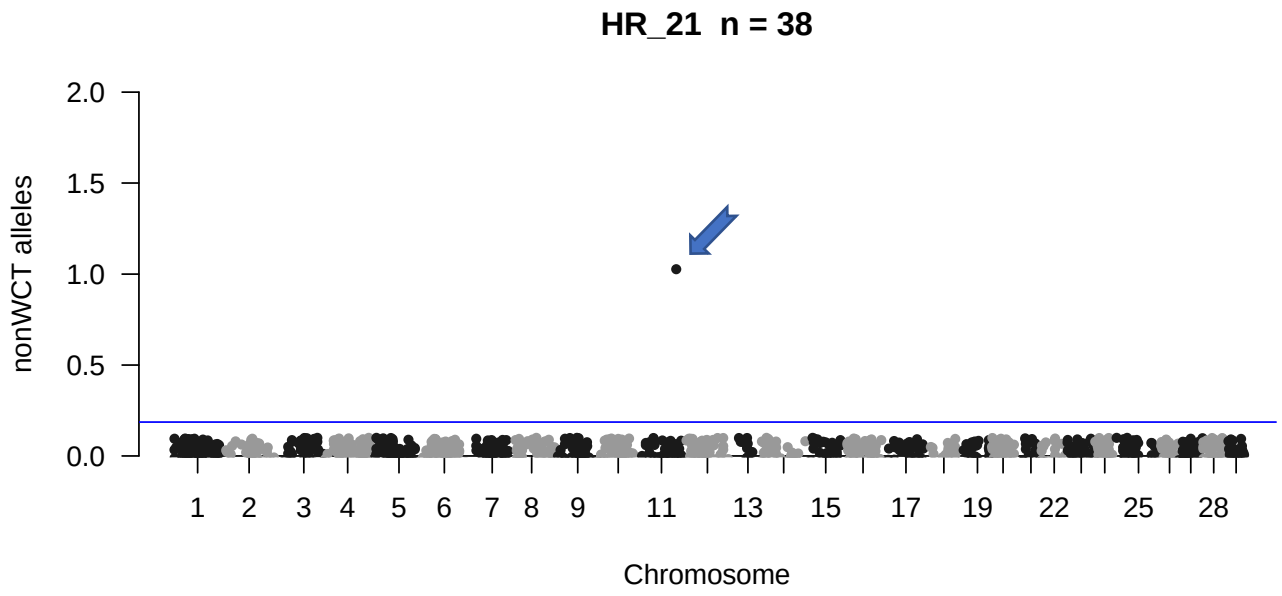


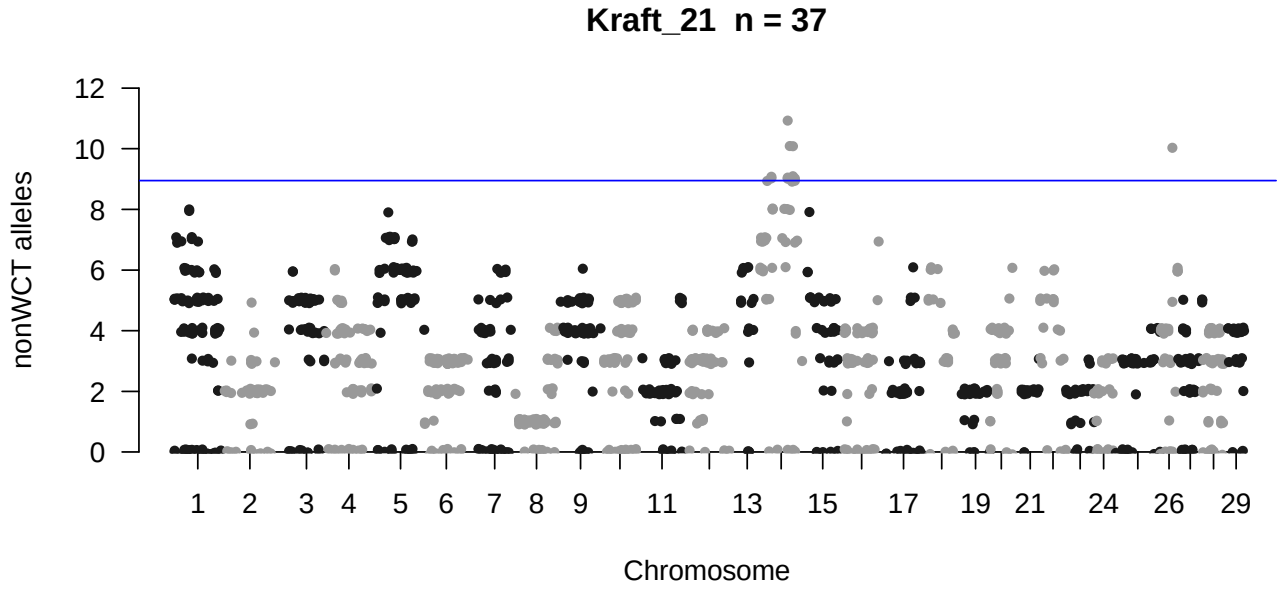

Cat_21 n = 30
nonWCT alleles
0
2
4
6
8
10
1 2 3 4 5 6 7 8 9 11 13 15 17 19 22 25 28
Chromosome

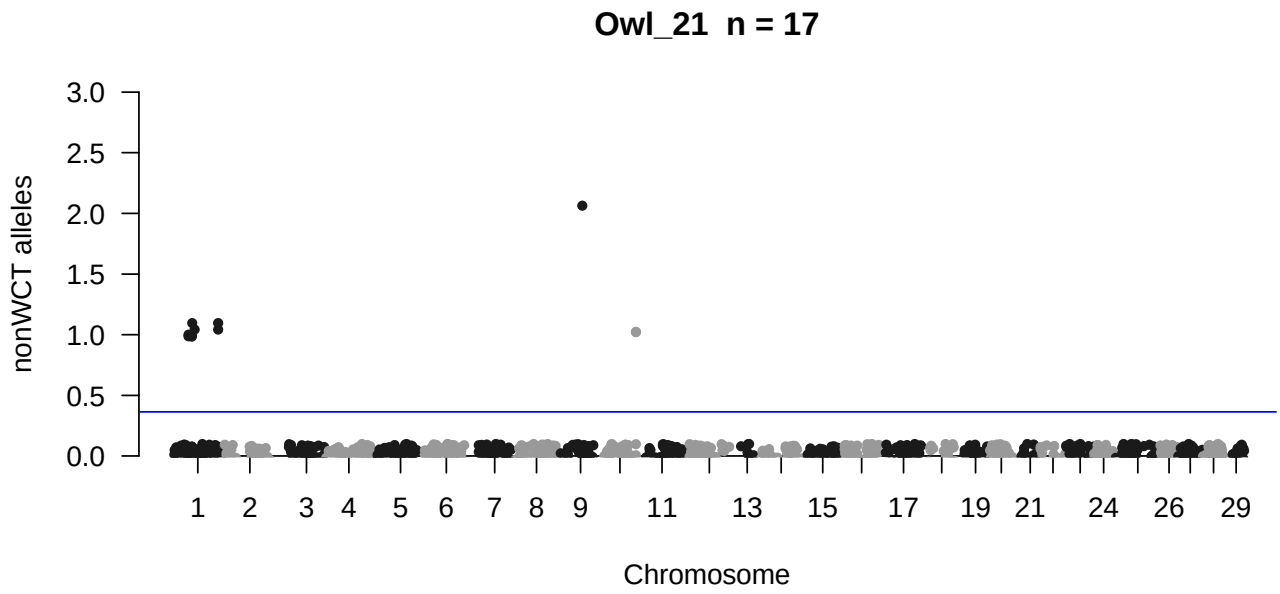

Cedar_21 n = 44
nonWCT alleles
0
1
2
3
4
5
1 2 3 4 5 6 7 8 9 11 13 15 17 19 21 24 26 29
Chromosome

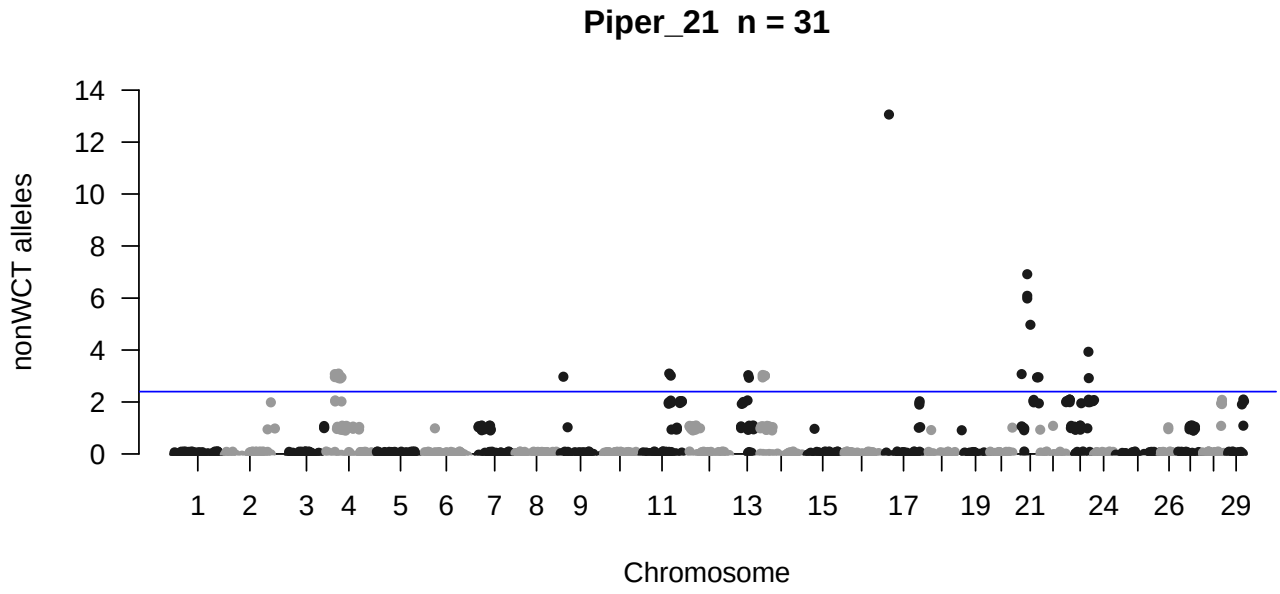

CedarLk_21 n = 28
nonWCT alleles
0
5
10
15
20
25
30
35
1 2 3 4 5 6 7 8 9 11 13 15 17 19 21 24 26 29
Chromosome

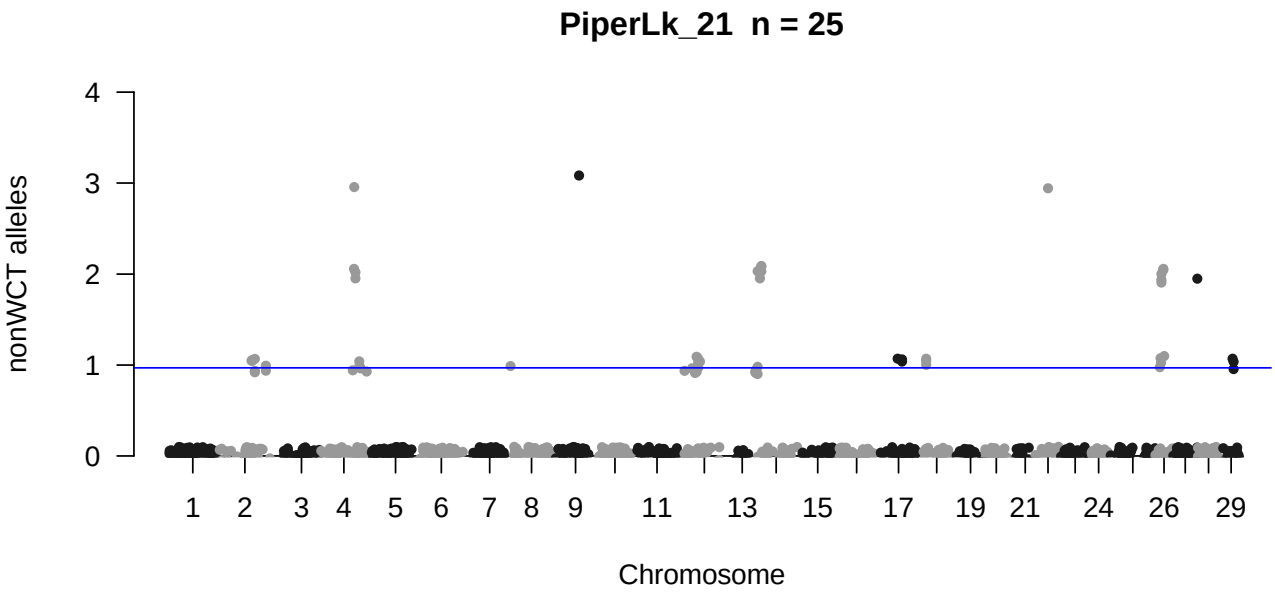

Dog_21 n = 35
nonWCT alleles
0
1
2
3
4
1 2 3 4 5 6 7 8 9 11 13 15 17 19 21 24 26 29
Chromosome

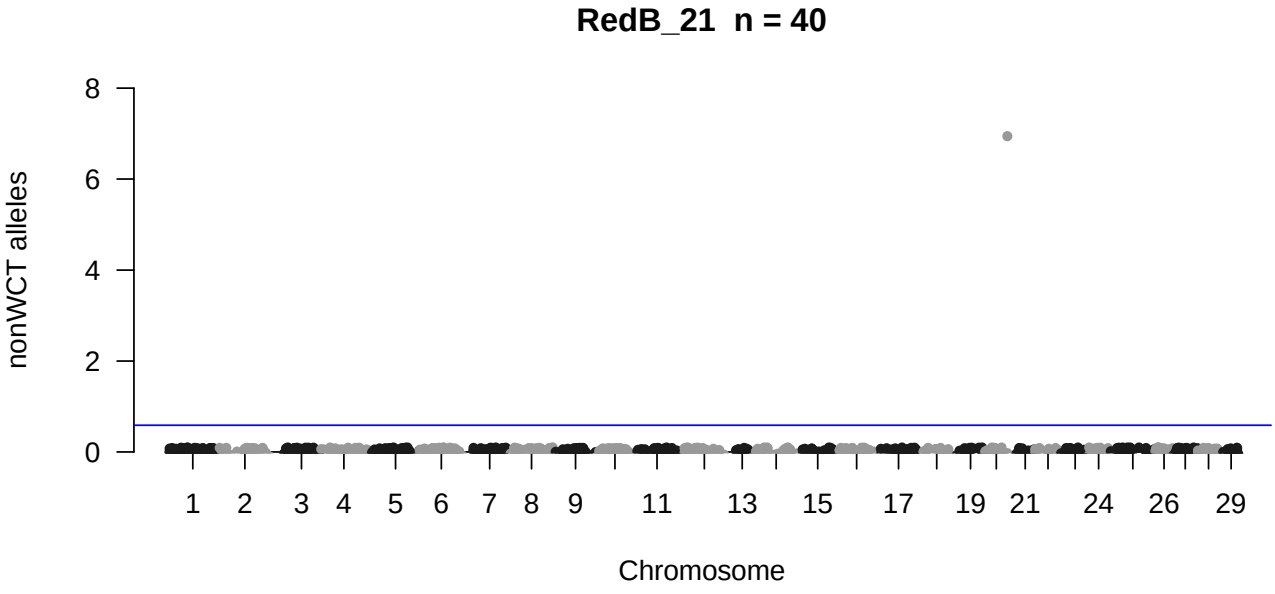

HR_21 n = 38
nonWCT alleles
2.0
1.5
1.0
0.5
0.0
1 2 3 4 5 6 7 8 9 11 13 15 17 19 22 25 28
Chromosome

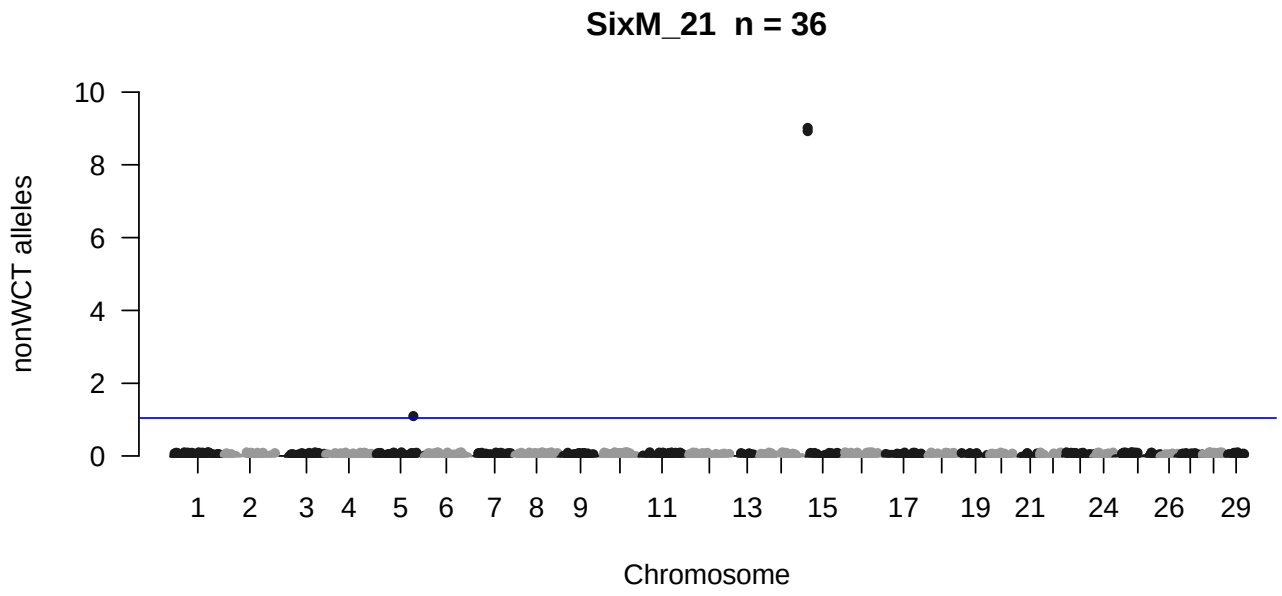

Kraft_21 n = 37
nonWCT alleles
12
10
8
6
4
2
0
1 2 3 4 5 6 7 8 9 11 13 15 17 19 21 24 26 29
Chromosome

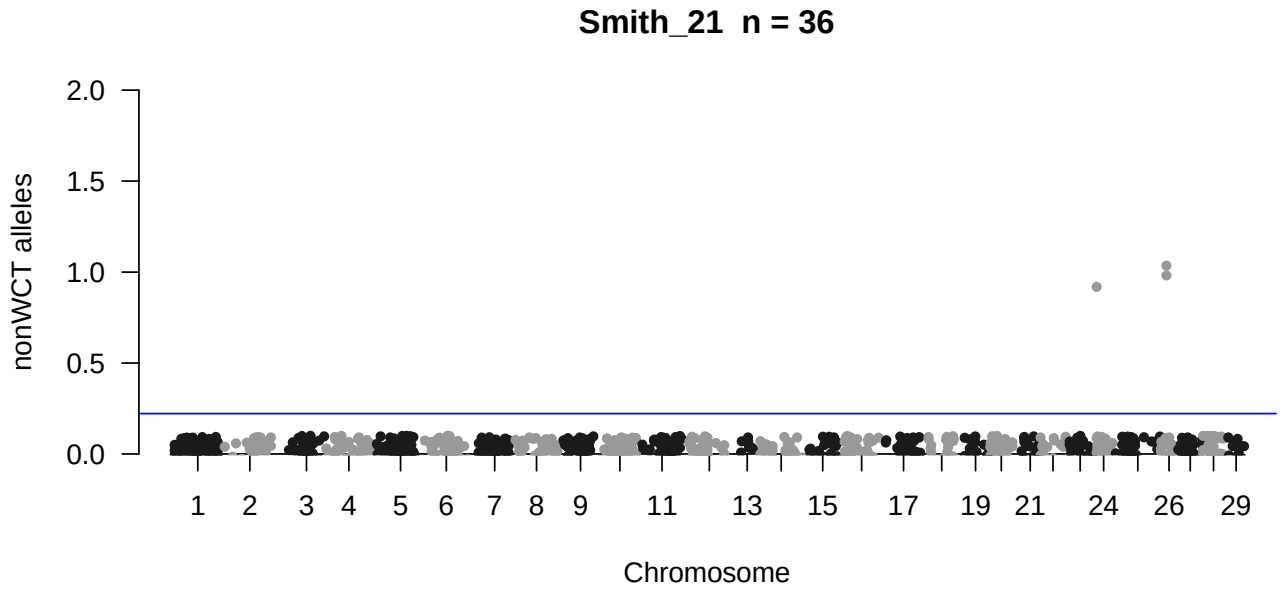

Owl_21 n = 17
nonWCT alleles
3.0
2.5
2.0
1.5
1.0
0.5
0.0
1 2 3 4 5 6 7 8 9 11 13 15 17 19 21 24 26 29
Chromosome

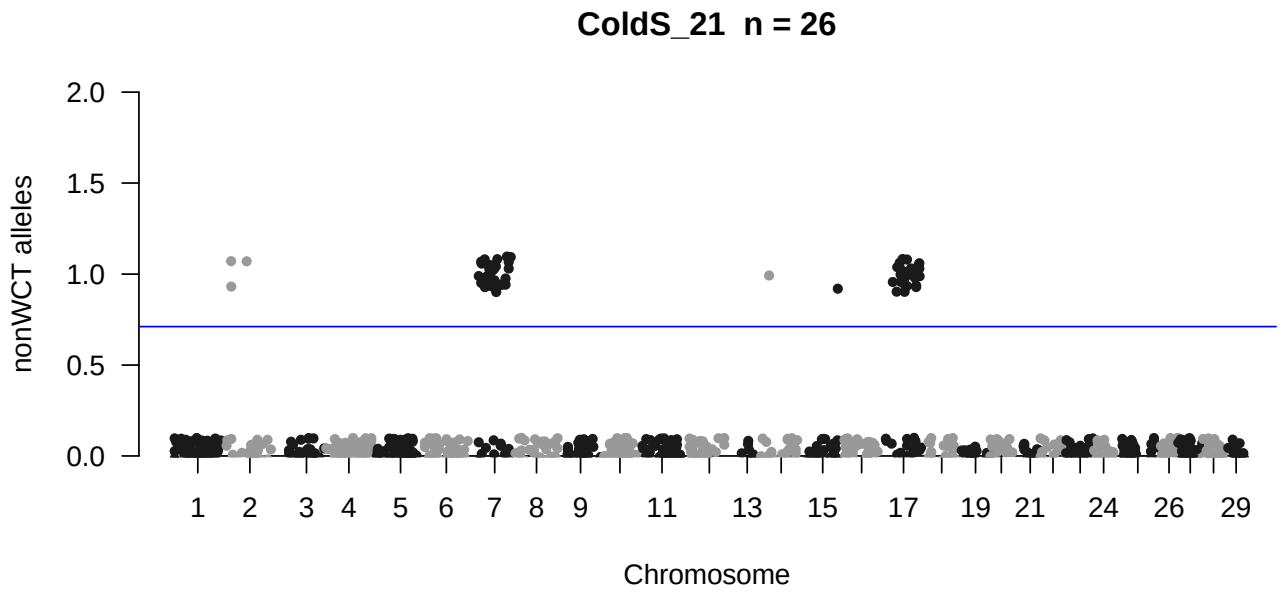

Piper_21 n = 31
nonWCT alleles
14
12
10
8
6
4
2
0
1 2 3 4 5 6 7 8 9 11 13 15 17 19 21 24 26 29
Chromosome


PiperLk_21 n = 25
nonWCT alleles
0
1
2
3
4
1 2 3 4 5 6 7 8 9 11 13 15 17 19 21 24 26 29
Chromosome


RedB_21 n = 40
nonWCT alleles
0
2
4
6
8
1 2 3 4 5 6 7 8 9 11 13 15 17 19 21 24 26 29
Chromosome


SixM_21 n = 36
nonWCT alleles
0
2
4
6
8
10
1 2 3 4 5 6 7 8 9 11 13 15 17 19 21 24 26 29
Chromosome


Smith_21 n = 36
nonWCT alleles
0.0
0.5
1.0
1.5
2.0
1 2 3 4 5 6 7 8 9 11 13 15 17 19 21 24 26 29
Chromosome



Methods

DNA was extracted from fin clip tissue samples and RAD-capture sequence libraries were genotyped targeting approximately 3,000 regions of the WCT genome. Laboratory methods generally followed the extraction, "BestRAD", and capture protocols outlined in Ali et al. (2016). Sequence read data was generated for RAD capture libraries on an Illumina MiSeq sequencer at the University of Montana Genomics Core and produced a median read depth of 24X. High-quality genotypes were selected by mapping raw sequence reads to the rainbow trout genome (Omyk_1.0; Pearse 2019) and filtering data in accordance with Broad Institute's "Best practices for non-model organisms". Genotypes were then filtered based on a minimum mapping quality of 30, minimum read depth of 7, heterozygote allele balance and minimum read depth (.25>AB<.75, 10), missing data per locus < 25%, and missing data per individual less than 30%. Additional filtering of loci with allele frequencies > 2SD above the mean, or missingness > 2SD above the mean, was done to account for potential shared ancestral polymorphisms and to account for potential null alleles respectively.

A set of 1070 RBT, 323 YCT, & 326 WCT diagnostic loci were used to estimate non-native admixture. Reference samples of known hybridization status for rainbow trout, Yellowstone cutthroat trout, and westslope cutthroat trout were included in the data set and analyzed at diagnostic loci to validate the genotyping process. For a locus to be species diagnostic, only one allele of a bi-allelic locus is associated with the species of interest, while the other allele is shared between the two remaining species. For example, in rainbow trout diagnostic loci one allele is associated with RBT, and the other allele is shared with WCT and YCT. For each set of diagnostic loci, the proportion of diagnostic alleles observed out of the total number of alleles is calculated at the individual (hybrid index or HI) and the population level (pRBT = RBT alleles / total number of possible alleles). The presence of first-generation hybrids (F1s) may be indicative of recent admixture. When the sample collection represents a randomly mating

population, non-native alleles are expected to be evenly distributed across individuals and loci. A HI histogram showing a disjointed distribution or a high number of individuals with no diagnostic alleles may be indicative of a recent admixture event, or a sample collection containing individuals from more than one population. A Manhattan plot showing blocks of diagnostic alleles across loci in the same region of a chromosome are indicative of blocks of admixture and can be used to confirm non-native hybridization.

References

Ali, O. A., O'Rourke, S. M., Amish, S. J., Meek, M. H., Luikart, G., Jeffres, C., & Miller, M. R. (2016). RAD capture (Rapture): flexible and efficient sequence-based genotyping. *Genetics*, *202*(2), 389-400.

Pearse, D. E., Barson, N. J., Nome, T., Gao, G., Campbell, M. A., Abadía-Cardoso, A., ... & Lien, S. (2019). Sex-dependent dominance maintains migration supergene in rainbow trout. *Nature Ecology & Evolution*, *3*(12), 1731-1742.