FEDERAL ASSISTANCE FINAL PERFORMANCE REPORT	ALASKA DEPARTMENT OF FISH AND GAME SPORT FISH DIVISION PO Box 115526 Juneau, AK 99811-5526

STATE WILDLIFE GRANT (SWG)

STATE: Alaska **GRANT AND SEGMENT NR.:** T-10-6
PROJECT NR.: P-22

WORK LOCATION: Lower Kenai Peninsula

PROJECT DURATION: July 1, 2013–June 30, 2014

PROJECT REPORTING PERIOD: July 1, 2013–June 30, 2014

PROJECT TITLE: Genetic variation of populations of juvenile Coho Salmon and Dolly Varden within and among river basins on the lower Kenai Peninsula

---

PROJECT GOAL:
The purpose of this study is to identify micro-geographic genetic distinctions between populations of the salmonid species coho Salmon (*Oncorhynchus kisutch*) and Dolly Varden Trout (*Salvelinus malma malma*) native in the Lower Kenai Peninsula to inform meaningful conservation practices (Funk et al. 2012). We examined the partitioning of genetic variation to the within and between drainage scale, and resolved whether juveniles of both species in headwater streams are genetically distinct from those found primarily in river margin habitats of the mainstream lower river year round (Walker et al., unpublished data).

PROJECT OBJECTIVES:

1. Identify geographic scale of population subdivision of coho salmon and Dolly Varden.
2. Create hundreds of SNP markers to be released to the salmonid research community.
3. Identify population specific alleles at SNP loci which can be used in future studies of migration in these populations.
4. Identify regions of the genome experiencing drainage specific selection.

SUMMARY OF PROJECT ACCOMPLISHMENTS:

Sampling
Tissue was taken from 17-20 individuals of each species collected from seven or ten geographic locations (7 coho locations, 10 Dolly Varden locations) spanning four adjacent drainages on the lower Kenai Peninsula (King et al. 2012) to permit the partitioning of genetic variation to the within drainage and between drainage scale. Primary experimental samples were taken from one location at each of the Ninilchik River, Deep Creek, Stariski Creek, and North Fork of the Anchor River drainages. Samples were taken at five locations within the South Fork of the

Anchor River basin representing sites along the watershed continuum from headwaters to main river. Locations are presented in Figures 1 & 2 and coordinates are given in Tables 1 & 2.

Sample Processing

Genomic DNA was extracted from each individual tissue sample using Qiagen DNeasy kits. DNA concentrations were obtained using the Life Technologies PicoGreen assay. Any samples that did not yield greater than sufficient concentration for library preparation (>25 ng/μL) were concentrated using a vacuum centrifuge. All samples were of sufficient quality for library preparation. Restriction site-associated DNA (RAD) libraries were prepared, and reduced complexity genome representation to allow for identification of a large number of single nucleotide polymorphism (SNP) markers. Prepared libraries were sequenced using next-generation sequencing (NGS). NGS data were trimmed and Quality Control (QC) techniques were applied to identify a large number of high confidence variant (SNP) loci to be released to the salmonid research community. Allele frequency was measured and population pairwise differentiation (F_{st}) was quantified to identify geographic scale of population subdivision of coho salmon and Dolly Varden. Significantly differentiated populations (F_{st} outliers) were separated as likely identifiers of genome regions under differential and purifying selection between drainages. Private alleles at SNP loci were identified that can be used in future studies of migration in these populations.

Bioinformatics

Sequence data was sorted to separate reads by individual sample. Reads of an individual from separate lanes were concatenated. These sorted data were trimmed of all sequences resulting from restriction digestion, barcoding and amplification. The resulting genotypic data were subjected to basic filtering for quality using the FASTX-Toolkit. Bases with a Phred quality score of less than 20 were discarded and reads with less than 90% of bases above the quality threshold were discarded. Initial processing produced ~ 3.4×10^{10}, accurately sequenced bases spanning ~9.7×10^{8} filtered reads.

We acquired the newly published genome for rainbow trout (*Oncorhynchus mykiss*) (Berthelot et al. 2014) as a reference genome. A genome index was created using Bowtie 2 (Langmead et al. 2012) to reduce memory footprint and access times. A sequence dictionary was created using Picard. Sequences were then provisionally aligned to the trout genome (approximately 25% of reads for each sample aligned uniquely).

Using the Broad Institute's Genome Analysis Toolkit (GATK) probable indels in the individual sequences were identified and sequence data was locally realigned to the reference genome around those transformed indel consensus regions. We used GATK's Unified Genotyper to identify an initial set of SNPs, of which only very high confidence variant loci were used to refine the aligned reads in Base Quality Score Recalibration (BQSR). The refined alignments were used with Unified Genotyper to call a more accurate set of variant loci. We then used GATK's recommended set of variant quality filters to identify SNPs that are likely to be the result of natural variation and not sequencing artifacts. We removed from the SNP set all loci that were out of Hardy-Weinberg Equilibrium within any population. Finally we filtered out loci that are genotyped in less than 80% of individuals using VCFtools. For coho salmon a total of 124,968 SNP loci were identified, of which 5839 passed all filters and were considered for analysis. For Dolly Varden 115,263 SNP loci were identified, and 9277 passed all filters and were analyzed. The resulting DNA data was formatted for various population genetic analysis programs using PGDspider2.

Analysis of Genetic Diversity

An analysis of genetic diversity in the coho (Table 1) indicates that all populations maintain low to moderate genetic diversity. Observed heterozygosity ranged from 0.323 (North Fork Anchor River) to 0.486 (South Fork Anchor River, SA1). Within a drainage, observed heterozygosity were similar suggesting that multiple estimates of heterozygosity per drainage may not be necessary in future studies (but see below). Generally populations found within the Anchor River had a higher genetic diversity relative to those populations found outside this drainage. The North Anchor population, however, contradicts this pattern in having the lowest observed heterozygosity. This may be a function of this population's relatively long distance from the Anchor River inlet. Additional data from Peri8, a site on the South Fork of the Anchor that is also far from the river mouth, may provide support for this hypothesis.

An analysis of genetic diversity in the Dolly Varden (Table 2) is lower than that observed in coho salmon possibly reflecting differences in the life history of these species. Observed heterozygosity was lowest in the Deep Creek population ($H_o = 0.274$) and highest in the Stariski Creek population ($H_o = 0.379$). Very little variation in genetic diversity was observed either within or between drainages for all populations, except the Deep Creek population, possessing observed heterozygosity between 0.309-0.379. This suggests that sampling multiple populations within the same drainage may not provide additional information on the average genetic diversity of populations in the focal drainage.

Coho Genetic Isolation

Pairwise F_{st} averaged across all loci was calculated using Arlequin 3.5 (Table 3). An analysis of pairwise F_{st} values indicates that very little genetic differentiation occurs within a drainage: F_{st} values of the South Fork Anchor River sites ranged from 0 to 0.013. Interestingly, the only within Anchor F_{st} estimates significantly greater than 0 were observed when comparing SA1 to other Anchor River populations. Otherwise, all within Anchor comparisons indicate that no genetic differentiation occurs within a drainage.

Across drainages, F_{st} values were relatively large. At $F_{st} = 0.037$, between drainage Fst was lowest between Anchor SA6 and Stariski. Between drainage F_{st} was maximized when comparing Anchor SA1 and the North Anchor populations ($F_{st} = 0.107$).

To statistically examine the hypothesis that genetic differentiation is greatest between rather than within drainages, the difference between means of pairwise Fst between and within drainages was tested using Welch's t-test. Differentiation between populations at the Lower Anchor and North Anchor are not significant (one tailed test, $p=0.12$, $df = 5$), while populations at the Lower Anchor and Stariski Creek are significantly different (one tailed test, $p=0.004$, $df = 5$). Our findings suggest genetic differentiation is greatest between rather than within drainages.

Of additional interest was the relationship between populations found in headwater streams to those found primarily in river margin habitats. To test for population differentiation of headwater versus river margin habitats, the pairwise F_{st} value between Anchor River SA1, a headwater site, and Anchor River SA6, a lower river site was calculated. At $F_{st} = 0.103$, the genetic differentiation of river and headwater populations is low, particularly when compared to across drainage genetic differentiation.

To examine geographic patterns of F_{st} variation more broadly, Isolation by Distance was examined. Pairwise stream distance between sites was estimated manually (Table 5) and these distances were compared to pairwise F_{st}'s using a Mantel Test for isolation by distance

implemented in Arlequin 3.5. The results indicate that genetic differentiation in coho salmon significantly increases with estimated stream distance ($r = 0.758$, $p = 0.009$).

Dolly Varden Genetic Variation

We examined the genetic differentiation of populations of Dolly Varden in the Lower Kenai Peninsula using the same analytical framework as was used for coho salmon. Pairwise F_{st} ranged were generally an order of magnitude lower than those values observed for the coho populations and ranged from 0 to 0.012 (Table 4). Among the Anchor River populations, genetic differentiation was very low ($0 \geq F_{st} \geq 0.012$) though comparisons within this drainage revealed the largest value of F_{st} observed in this study.

While there were many populations with no significant genetic difference between drainages, some F_{st} values, while low, were greater than 0. Nonetheless, comparing within versus between drainage average F_{st} values indicate that population differentiation is no greater between drainages than within (within Anchor River sites versus Stariski Creek data, one tailed test df = 14.26, p-value = 0.014, within Anchor River sites versus Ninilchik or Deep Creek were not significant). Unlike coho salmon, Dolly Varden populations sampled at the headwater and river margin habitats of the Anchor River show no genetic differentiation: F_{st} between Anchor River, Lower (SA7) and Anchor River headwaters, (SA2) equals 0.

When examining the overall geographic pattern of Fst variation, the Mantel test indicated that stream distance is not effective in predicting genetic differentiation ($r = 0.758$, $p = 0.850$). More complex models of isolation are needed to explain the pattern of genetic differentiation observed in Dolly Vardens on the lower Kenai Peninsula. It is worth noting, however, than much of the overall patterns of genetic differentiation are driven by two populations; Deep Creek and the Anchor River SA3 site. All pairwise comparisons involving Deep Creek were significantly greater than 0. As Deep Creek is the most remote Dolly Varden sampling site included in this study, the observation of low, but significant F_{st}'s between this and all other Dolly Varden population suggests that Isolation by Distance may play an important role in structuring Dolly Varden populations at geographic scales greater than those included in this study. What is not understood are the forces driving the differentiation of SA3 from all other populations included in this study. In all pairwise comparisons, SA3 was significantly differentiated from other populations. In fact, the pairwise comparisons involving SA3 yielded the largest F_{st} values observed in the Dolly Varden comparisons ($F_{st} = 0.012$). Remarkably, one of the pairwise comparisons yielding this F_{st} value involved SA3 and a population less than 7km from the SA3 collection site (SA4). The genetic differentiation of SA3 warrants additional study and may require particular protections to insure the viability of the genetically distinct population.

Identification of loci of interest in coho

We examined the distribution of SNP allele frequencies to identify 1) loci in which Fst were greater than expected based on the distribution of observed pairwise comparisons and 2) loci which contained alleles which were specific to individual populations, i.e. private alleles. The identification of loci with greater than expected F_{st} values suggests that these loci may be under directional selection which prevents the genetic exchange at this position in the genome. To identify outlier loci, the distribution of all pairwise F_{st} values was examined (Figure 2). The distribution of F_{st} values approximates a negative exponential distribution. No outliers from this distribution could be identified, however those loci with F_{st} values between 0.44 and 0.46 are provided with their flanking sequence in Appendix 1. Future population genetic work and molecular evolution studies of these loci are needed to test the hypothesis that these loci are under selection.

To provide resources for genetic identification of samples of unknown geographic origin, we attempted to identify loci possessing private alleles. No loci with private alleles were identified. Though disappointing, the absence of private alleles confirms the conclusion that the genetic differentiation of coho salmon is not great at the sampling scale used in this study.

Identification of loci of interest in Dolly Varden

We approached the search for loci under selection in Dolly Varden populations as described for coho. We found that the distribution of F_{st} in this species also approximates a negative exponential distribution (Figure 5). We distinguished loci with F_{st} values greater than 0.2 as being under putative selection. They are provided with their locations and flanking sequence in Appendix 2. We did not find private alleles in these samples of Dolly Varden populations.

Future Directions

This study yielded a huge amount of data that provides opportunity for future work. The results of this study indicate that genetic differentiation of coho populations within drainages is limited while differentiation is greater across drainages. This suggests that future sampling of coho should focus on between drainages, however the inclusion of several additional within drainage populations would be useful to conduct a formal Analysis of Molecular Variation. Such an analysis would estimate the pattern of genetic differentiation at the within and between scales in a more formal manner.

Genetic differentiation was much lower in the Dolly Varden populations and our results suggest that population differentiation likely is of greater importance across geographic scales larger than those included in the sampled sites. The exception to this observation is the population at SA3. This population is genetically differentiated from even its nearest neighbor. This finding suggests that SA3 warrants additional genetic and ecological study and possibly special protection.

Because this study was focused on populations on the Kenai Peninsula, our data cannot speak to the degree of genetic differentiation at larger geographic scales. The populations identified in this study are all less than ~ 90km apart. How patterns of population genetic differentiation are shaped across the distribution of these species is unknown.

A significant limitation of this study is the absence of a fully assembled salmonid species more closely related to the focal species. For this study, we utilized the recently released rainbow trout genome. While this genome is useful in assembling short read sequences into loci, the potential lack of synteny between the rainbow trout and coho salmon and Dolly Varden prevents the examination of population differentiation at the genomic scale.

FINAL REPORT STATUS: Genetic data from this project are available from the Kachemak Bay Research Reserve.

PREPARED BY: William J. Matthaeus (graduate student, Baylor University), Dr. Pat Danley (Professor, Baylor University, Dr. Ryan King (Professor, Baylor University, Coowe Walker (Project Lead, Kachemak Bay Research Reserve)

DATE: SEPTEMBER 9, 2014

Table 1. Coho salmon general population information. Sampling site names and abbreviations. Number of individuals sequenced and variant loci identified.

	Basin	Catchment ID	Reach ID	Lat.dd	Lon.dd	N (sequenced)	N (analyzed)	SNPs (analyzed)	Observed Heterozygosity
SA1	Anchor River	1203	HWS2_1203M	59.779675	-151.5551056	16	10	5839	0.485
SA2	Anchor River	8	Peri_8	59.7405453	-151.3027552	0	0	0	
SA3	Anchor River	V03	Val_V03	59.7095941	-151.6987938	19	10	5839	0.457
SA4	Anchor River	V05	Val_V05	59.7510472	-151.7131527	17	9	5839	0.465
SA6	Anchor River, lower	GW03	NAD83	59.7711751	-151.8435987	20	10	5839	0.396
S	Stariski Creek	171	HWS2_171M	59.8406447	-151.7829479	19	10	5839	0.381
NA	Anchor River, north branch	44	HWS2_44L	59.8613647	-151.6583606	18	10	5839	0.323

Table 2. Dolly Varden general population information. Sampling site names and abbreviations. Number of individuals sequenced and variant loci identified.

	Basin	Catchment ID	Reach ID	Lat.dd	Lon.dd	N (sequenced)	N (analyzed)	SNPs (analyzed)	Observed Heterozygosity
SA1	Anchor River	1203	HWS2_1203M	59.779675	-151.5551056	20	10	9277	0.309
SA2	Anchor River	8	Peri_8	59.7405453	-151.3027552	20	10	9277	0.308
SA3	Anchor River	V03	Val_V03	59.7095941	-151.6987938	20	10	9277	0.308
SA4	Anchor River	V05	Val_V05	59.7510472	-151.7131527	19	10	9277	0.374
SA5	Anchor River	V24	Val_V24	59.7245924	-151.5285954	20	10	9277	0.334
SA7	Anchor River lower	SANC Main	NA	59.756693	-151.783863	19	10	9277	0.363
S	Stariski Creek	171	HWS2_171M	59.8406447	-151.7829479	19	10	9277	0.379
NA	Anchor River, north branch	44	HWS2_44L	59.8613647	-151.6583606	19	10	9277	0.337
N	Ninilchik River	545	HWS2_545L	60.049777	-151.6320347	12	0	0	-
D	Deep Creek	V12	Val_V12	59.9900454	-151.4919892	20	10	9277	0.274

Table 3. Estimates of population differentiation (Fst) among six populations of coho salmon. Below the diagonal is the average pairwise F_{ST} between the two populations. Population abbreviations are as indicated in Table 1. Significant values of Fst ($p<0.05$) values are bolded. Negative F_{ST} values are presented as 0. .

	SA1	SA3	SA4	SA6	S	NA
SA1						
SA3	**0.011**					
SA4	**0.005**	0				
SA6	**0.013**	0	0			
S	**0.06**	**0.048**	**0.044**	**0.037**		
NA	**0.107**	**0.098**	**0.094**	**0.087**	**0.095**	

Table 4. Estimates of population differentiation (Fst) among six populations of Dolly Varden. Below the diagonal is the average pairwise F_{ST} between the two populations. Population abbreviations are as indicated in Table 2. Significant values of F_{st} ($p<0.05$) values are bolded. Negative F_{ST} values are presented as 0.

	SA1	SA2	SA3	SA4	SA7	S	NA	N	D
SA1									
SA2	0								
SA3	**0.009**	0							
SA4	0	0	**0.012**						
SA7	**0.002**	0	**0.003**	**0.002**					
S	0	0	**0.009**	0	0				
NA	0	0	**0.001**	0	0	0			
N	0	0	**0.012**	0	**0.001**	0	0		
D	**0.002**	**0.001**	**0.007**	**0.008**	**0.005**	**0.003**	**0.004**	**0.004**	

Table 5. Estimates of pairwise stream distances (km) between coho sampling sites. Abbreviations as in Table 1.

	SA1	SA2	SA3	SA4	SA6	S	NA
SA1	-	21.12	16.59	23.45	31.96	54.91	54.03
SA2		-	22.64	29.50	38.02	60.97	60.08
SA3			-	6.86	15.37	38.48	37.43
SA4				-	12.05	35.00	34.12
SA6					-	22.95	23.29
S						-	46.24
NA							-

Table 6. Estimates of pairwise stream distances (km) between Dolly Varden sampling sites. Abbreviations as in Table 1.

	SA1	SA2	SA3	SA4	SA7	S	NA	N	D
SA1	0								
SA2	21.12	0							
SA3	16.59	22.64	0						
SA4	23.45	29.50	6.86	0					
SA7	11.27	17.12	12.79	19.65	0				
S	26.26	32.32	9.67	6.36	22.47	0			
NA	54.91	60.97	38.48	35.00	51.12	28.65	0		
N	54.03	60.08	37.43	34.12	50.23	27.76	46.24	0	
D	83.30	89.35	66.87	63.39	79.50	57.04	44.03	74.63	0

Figure 1. Map of coho sampling sites in lower Kenai Peninsula noting river name, catchment ID and Reach ID. Coho sampling within the Lower Anchor Rive are labeled: Anchor River, Lower GW, Anchor River Vo3 Val, Anchor River VO5 Val, Anchor River 1203 HWS2. In North Anchor: Anchor River, north br. And In Stariski Creek: Stariski Creek 171 HWS.

Figure 2. Map of Lower and North Anchor river Dolly Varden sampling sites in lower Kenai Peninsula noting river name, catchment ID and Reach ID.

Figure 3. Map of Deep Creek and Ninilchik River Dolly Varden sampling sites in lower Kenai Peninsula noting river name, catchment ID and Reach

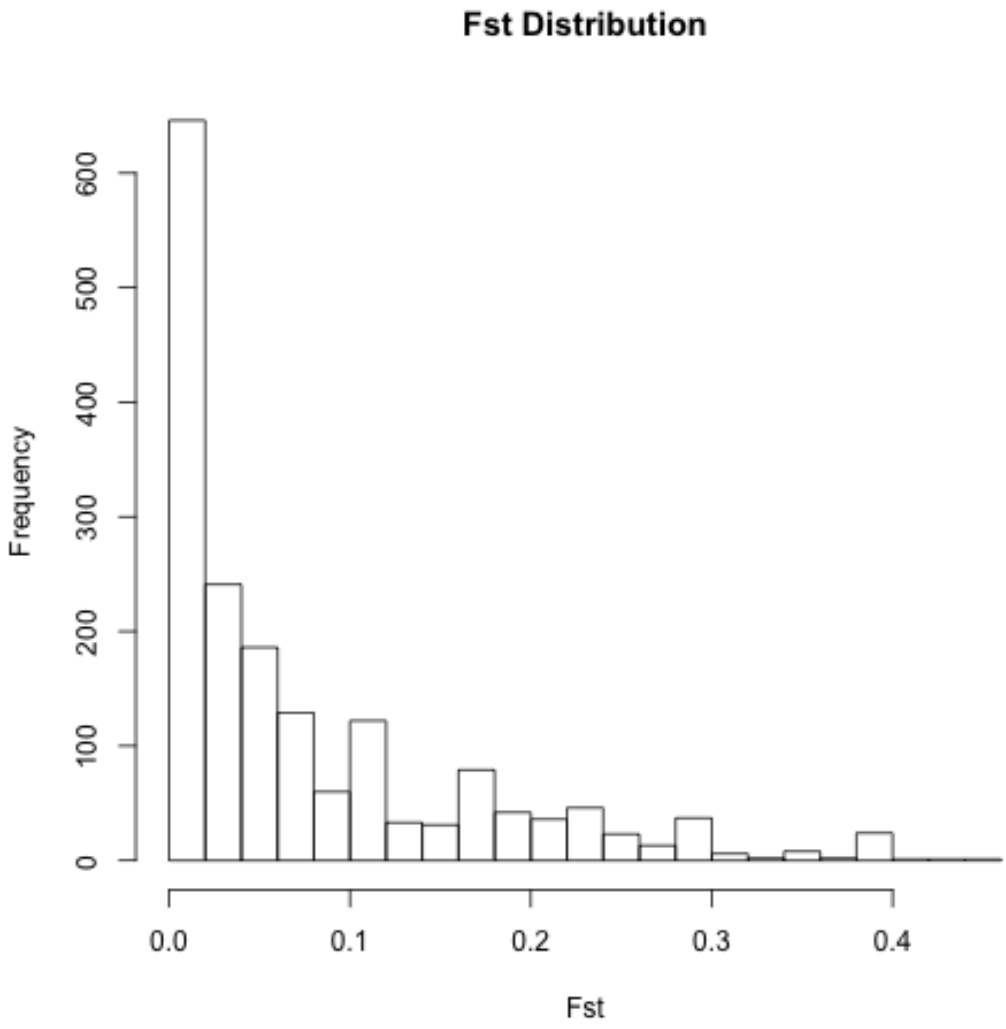

Figure 4: Distribution of Weir and Cockerham's F_{st} within loci, averaged across all coho salmon population pairwise comparisons.

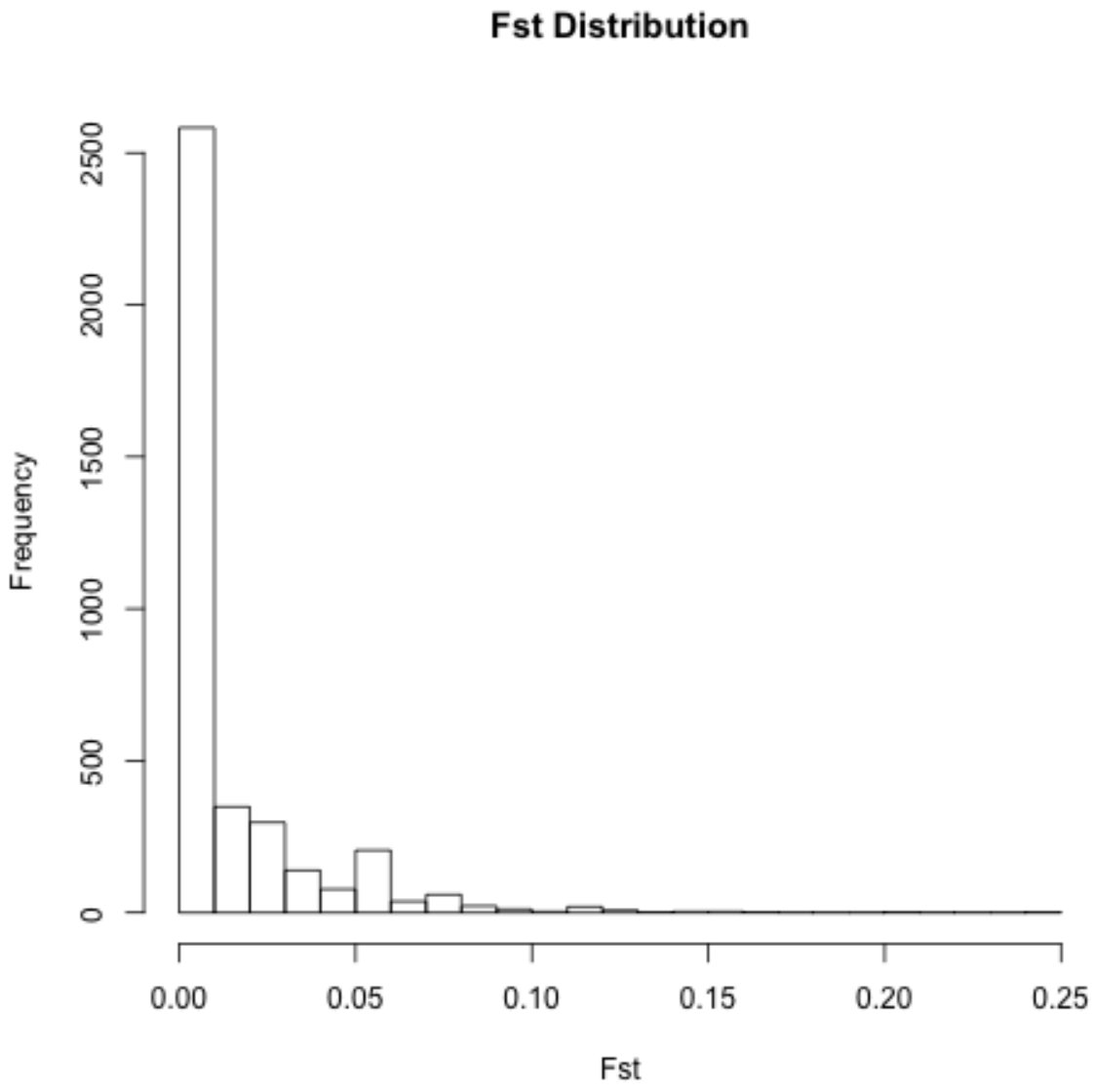

Figure 5: Distribution of Weir and Cockerham's F_{st} within loci, averaged across all Dolly Varden population pairwise comparisons.

References:

Berthelot, C., Brunet, F., Chalopin, D., Juanchich, A., Bernard, M., Noël, B., Bento, P., Da Silva, C., Labadie, K., Alberti, A., et al. (2014). The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. Nature Communications *5*

King, R. S., Walker, C. M., Whigham, D. F., Baird, S. J., & Back, J. A. (2012). Catchment topography and wetland geomorphology drive macroinvertebrate community structure and juvenile salmonid distributions in south-central Alaska headwater streams. Freshwater Science, 31(2), 341-364.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods *9*, 357–359.

Appendix 1: Loci with extreme outlier average Fst values (>0.4) with flanking sequence of the SNP identified by a lower case base.

Chromosome	Position	Fst	Flanking Sequence
ENA|FR904378|FR904378.1	1467944	0.418	TGCAATTGGGAGCAGACACACACGCATGCACgTCGTTGAAGTCCAGGAGGAGTAGCAAGGTTT
ENA|FR955678|FR955678.1	1325	0.432	CCAGGAAACAGTAAGGATACGAGCCTAATGCaGTCATTACCAGTTCAACACTCTCTTTCAATG
ENA|FR907303|FR907303.1	72611	0.444	ATTTAAATTGTCCTTACTTTATGGGCTCGTCtAACGATTGAAGTGCTCAGAAAATAGCTTATT

Appendix 2: Loci with extreme outlier average Fst values (>0.2) with flanking sequence of the SNP identified by a lower case base.

Chromosome	Position	Fst	Flanking Sequence
ENA|FR904554|FR904554.1	980931	0.20736	TGATGCATACATGTCAGAATGTTTACGCAACaGTTTCGATCAATTTCTAAATTCAATGTCAAA
ENA|FR906708|FR906708.1	20210	0.24457	GGGTGGCCGTGTCAATCATCAAACACTGGCAtCGATAGCTCGACTATGTTGCTGTTTCATCAG