

Development of a genetic assay to differentiate Tucannon River and Asotin Creek steelhead populations using RAD-seq methods

Todd R. Seamons¹, Amelia Whitcomb¹, Glen Mendel², Joseph D. Bumgarner³, and Kenneth I. Warheit¹

¹Washington Department of Fish and Wildlife, Molecular Genetics Laboratory, Olympia, WA

²Washington Department of Fish and Wildlife, Fish Management Office, Dayton, WA - Retired

³Washington Department of Fish and Wildlife, Snake River Laboratory, Dayton, WA

21 September 2015

*Corresponding Author: WDFW Molecular Genetics Laboratory. 600 Capitol Way N., Olympia, WA 98501-1091. Todd.Seamons@dfw.wa.gov. (360) 902-2765

Introduction

Prior evaluation by the Washington Department of Fish and Wildlife (WDFW) of southeast Washington and lower Snake River steelhead (*Oncorhynchus mykiss*) genetic structure using microsatellite DNA suggested that steelhead in Asotin Creek and the Tucannon River, listed under the Endangered Species Act (ESA), were genetically distinct and belonged to different aggregates of populations (Blankenship et al. 2011, Blankenship et al. 2009, Blankenship et al. 2007). A proposal submitted by WDFW to evaluate the genetic relationships of steelhead in small Snake River tributaries to steelhead in nearby major spawning areas in the Tucannon River and Asotin Creek was funded by the WDFW Columbia River Endorsement Board in 2013. That study was intended to evaluate the accuracy of genetic assignments of steelhead to either of these larger steelhead populations or to specific small Snake River tributaries. However, a separate analysis conducted by Idaho Dept. of Fish and Game using 188 single-nucleotide polymorphism (SNP) loci and different tissue collections from the same locations suggested little genetic difference between the steelhead from the Tucannon River and Asotin Creek, or other areas of southeast Washington, and put both the Tucannon and Asotin steelhead in the same reporting group (Ackerman et al. 2012) casting doubt on previous genetic analyses and the ability to distinguish these southeast Washington populations using genetic data. Ackerman (personal communication) stated that individual assignment using genetic stock identification would not be possible with his dataset. Therefore, the 2013 WDFW study was replaced with this current genetic study using RAD-seq reduced representation genomic sequencing methods. Without a high level of power to correctly assign steelhead to Asotin and Tucannon populations, it would also likely not be feasible to assign steelhead from small Snake River tributaries that lie between them. Natural origin steelhead in these small tributaries are currently assigned to either the Asotin or Tucannon populations based on geographic proximity for ESA recovery planning/implementation and fisheries management (Figure 1). Sporadic hatchery releases of unknown stock summer steelhead into Asotin Creek and the Tucannon River occurred from the 1930s-1970s, but both received consistent plants of hatchery steelhead of various stocks beginning in the early 1980's (Table 1). Stocking of hatchery steelhead was discontinued in Asotin Creek following the ESA listings of steelhead in the Snake River in 1997, but Lyons Ferry stock steelhead continued to be released into the Tucannon River until 2010. Genetic differences may have been reduced beyond detectability between these two populations due to a variety of factors: 1) the history of hatchery releases in both basins from a variety of stocks, 2) documented straying of both Lyons Ferry stock and other Snake River hatchery stocks into the Tucannon River from other release locations (e.g., Walla Walla, Touchet, Lyons Ferry; Figure 2), 3) documented hatchery steelhead straying from southeast Washington releases, including steelhead released into the Tucannon River, returning to Asotin Creek, 4) relatively large proportion of stray natural origin steelhead (primarily from the middle Columbia River) entering the Tucannon River (Figure 3), and 5) the proximity of lower Snake River dams and impoundments that may impair steelhead from returning to their natal waters.

Given the apparent uncertainty of the genetic relationship of Asotin and Tucannon steelhead, we undertook a project to scan the genomes of individuals using RAD-seq next generation sequencing (NGS) methods to search for SNP loci that differentiate these populations. Since Touchet River (natural origin) and Lyons Ferry stock hatchery steelhead are known to stray into the Tucannon River and Asotin Creek, we included individuals from these populations as well. If a new SNP panel had high resolution and statistical power to separate steelhead from Tucannon and Asotin Creeks, we may be able to determine if steelhead from Alpowa Creek, Almota Creek, or Penawawa Creek (and possibly other small tributaries) are more closely related to, or the same as, Asotin or Tucannon populations. The ability to distinguish among natural origin steelhead stocks in southeast Washington is key to appropriate ESA recovery planning and implementation and steelhead management.

Methods

Overview – Reduced-representation genomic libraries were constructed using steelhead fin tissue collected from 48 natural origin adults each from the Tucannon River, Touchet River, Asotin Creek, and 48 hatchery origin adults from the Lyons Ferry Hatchery stock. Libraries were developed using restriction-site associated DNA (RAD-seq) methods (e.g., Baird et al. 2008). From the genomic libraries, we identified a large panel of polymorphic single-nucleotide loci (SNPs), and from the panel of polymorphic SNP loci we isolated a subset of loci that were especially effective at differentiating the four ascertained populations.

Tissue Collections Description – Forty-eight natural origin adults each were selected from tissue collections of adult fish returning to spawn in the Tucannon River (WDFW code 05AI), the Touchet River (05AJ), Asotin Creek (05AX), and from Lyons Ferry Hatchery stock (05AL) in 2005. Individuals were chosen from both sexes and the primary age-classes (total-age 4 and 5), which were determined by scale age or fork length.

Genomic Library Preparation – To extract and isolate genomic DNA from tissue, Qiagen DNEasy[®] kits (Qiagen Inc., Valencia, CA) were used, following the recommended protocol for animal tissues. DNA was then quantitated using Quant-It[™] BR assay kit (Life Technologies) and a QuantiFluor[®] ds DNA system (Promega) to normalize or standardize DNA from all individuals at 1µg/40µL. Genomic DNA was then digested using the enzyme *Sbf* I-HF[®] (New England Biolabs) at 50 µL reaction volumes (400 U/ml *Sbf*I-HF[®], 1X Cutsmart[™] buffer). Digests were conducted at 37°C for 3h followed by 65°C for 20m. P1 adapters (Integrated DNA Technologies), which included a DNA barcode specific to each individual fish, were ligated to digested DNA in 60 µL reaction volumes (8.3 nM P1 adapters, 0.17X NEBuffer 2 (New England Biolabs), 1 nM rATP (Promega), 16,666.7 U/ml T4 DNA Ligase (New England Biolabs)). The reaction was incubated at room temperature for 1h followed by 65°C for 20m, after which DNA

from individuals was pooled into a single reaction. Pooled DNA was sheared using a Bioruptor[®] (Diagenode) for 4 to 9 cycles of 30s of shearing and 59s resting, depending on DNA quality. DNA was then purified and size selected using Agencourt[®] AMPure[®] XP PCR purification kits (Beckman Coulter Inc.) following the manufacturers protocol. Genomic libraries were prepared, including the ligation of the P2 adapter, using the KAPA LTP Library Preparation Kit for Illumina[®] platforms (KAPA Biosystems) following the manufacturers protocol with the optional final PCR amplification step with an annealing temperature of 68°C. Following preparation, library DNA concentration was evaluated using qPCR with the KAPA Library Quantification Kit for Illumina[®] platforms and an Applied Biosystems[™] 7900 real-time PCR system (Life Technologies) following the manufacturers protocol. Libraries were normalized to 10 nM and sent to University of Oregon Genomics Core Facility (UOGCF), where they were run on an Illumina[®] HiSeq 2000 sequencer. After the first round of sequencing, the sequence data were processed using the algorithms in the *process_radtags* module of *STACKS* (Catchen et al. 2013, Catchen et al. 2011) to evaluate the average depth of coverage per individual. To increase the total yield per individual, individual libraries were again normalized at the P1 ligation step based on the average depth of coverage such that DNA from individuals with high depth of coverage was reduced and DNA from individuals with low depth of coverage was increased. Library preparation proceeded as described above and the new libraries were submitted to the UOGCF for the second round of sequencing on the HiSeq 2000.

SNP Discovery and Genotyping – SNP discovery and genotyping were performed using the algorithms employed by the software package *STACKS*. Modules were run separately, starting with *process_radtags*, which evaluated read quality and collated sequence reads for each individual based on barcode sequences. Barcodes and RAD-tags were rescued, reads with uncalled bases were removed, default values for sliding window size and score limit were used, and reads with low quality scores were discarded. Sequences from round 1 and round 2 of sequencing were processed separately. Processed sequences from both rounds of sequencing were combined for each individual for further analysis.

After processing raw sequencing data, sequences were assembled into sets or stacks of sequences that represent unique loci (RAD-tags) using the *STACKS* module *ustacks*. Minimum depth of coverage to form a stack was set to 3. Maximum distance allowed between stacks was set to 3. The removal and deleveraging algorithms were enabled. The maximum number of stacks at a *de novo* RAD-tag was set to 4. The bounded SNP model was used with $\alpha = 0.05$ and the upper bound for epsilon, $\epsilon = 0.1$.

After building the RAD-tags for each individual, RAD-tags were compared among all individuals to create a catalog of RAD-tags comprise of all unique RAD-tags from all individuals, each with a unique identifier, using the *STACKS* module *cstacks*. The number of mismatches allowed between sample RAD-tags when generating the catalog was set to 2. RAD-

tags for each individual were then compared to the catalog and renamed according the unique identifier using the *STACKS* module *sstacks*. Polymorphism can occur at any nucleotide base pair along a RAD-tag. These polymorphisms are called single nucleotide polymorphism or SNPs. Files containing genotypes at each SNP locus for each individual were generated using the *STACKS* module *populations*. No data filters were used. Summary statistics were generated for each collection to be used in downstream data filtering.

Filtering and Final SNP Panel Selection – SNP loci to be used in downstream analyses were identified by filtering the list of loci using several criteria. To eliminate uninformative loci, loci with minor allele frequencies (MAF) less than 0.10, with all populations combined, were removed. To remove loci found in the most error prone region of the RAD-tag and to provide enough flanking sequence to create amplicon sequencing panels, loci found within 20 base pairs of either end of the RAD-tag were removed. To remove potentially over-merged RAD-tags (over-merged during the *ustacks* and *cstacks* analyses described above), RAD-tags with more than four SNPs per RAD-tag were removed. The genomes of salmonids duplicated relatively recently in their evolutionary history creating tetraploid genomes. Because re-diploidization of their genomes is incomplete, duplicate, or paralogous, loci are relatively abundant in the genome. Paralogous loci must be identified and removed to prevent erroneous statistical results and interpretation in downstream analyses. Putative paralogous SNPs were identified and removed two ways. First, loci for which all individuals in all populations were heterozygous were assumed to be paralogs and were removed. Second, our *de novo* catalog was matched to a separate catalog wherein paralogous loci had been identified by using doubled-haploid families (S. Young and K. Warheit, WDFW, unpublished data). Any locus identified as paralogous in the separate catalog was removed from our list of loci. Loci that were genotyped in <81% (156 of 192) of individuals were removed. Finally, from RAD-tags with 2 to 4 SNPs per RAD-tag, the SNP locus with the highest heterozygosity was retained and the others discarded.

Once polymorphic SNP loci were identified, loci were ranked using per-locus global (across all populations) F_{ST} (Storer et al. 2012). Per-locus global F_{ST} s were calculated using the AMOVA analysis employed by the software ARLEQUIN v3.0 (Excoffier et al. 2005). Unlike the scenario presented by Storer et al. (2012), we currently do not have a limit on the number of SNPs we can use. Therefore, rather than take the 48 or 96 top ranked SNP loci, we retained all loci for which the F_{ST} was statistically significant at $\alpha = 0.05$ after correction for multiple testing. From this subset of loci, the minimum F_{ST} was 0.029. As we develop further the SNP panel for high throughput SNP genotyping, we may need to re-evaluate the final list to narrow it down to a specific number of the top ranked loci. That number will depend on the genotyping method employed; therefore we did not explore various scenarios at this time.

SNP panel evaluation – The loci defining the final panel were evaluated for conformation to Hardy-Weinberg expectations by examining observed and expected heterozygosity, F_{IS} , and

linkage disequilibrium (LD). Observed and expected heterozygosity and F_{IS} were calculated using GDA (Lewis and Zaykin 2001). Probability tests for LD for each pair of loci in each population were conducted using GENEPOP v4.0 (Raymond and Rousset 1995, Rousset 2008). Significance tests of F_{IS} and LD were evaluated with and without correction for multiple tests using false discovery rate (FDR, Verhoeven et al. 2005). Including many relatives in population collections may bias some analyses. In order to identify relatives, genetic data were analyzed using COLONY2 (Wang 2004, 2013, Wang and Santure 2009), which genetically identifies full-sibling relationships. All but one randomly chosen member of each full sibling family of more than three members (if any) were removed from individual assignment test analysis.

The Tucannon population had many more locus pairs in LD than other populations (see Results). Factors related to the effective population size (N_e) could cause such a pattern (Waples 2015). Thus the N_e of each population was calculated using the methods employed in the software LDNE (Waples and Do 2008) using the genetic data sets with COLONY-identified relatives included. LDNE estimates may be affected by allele frequencies, therefore P_{crit} values, the threshold allele frequency for an allele to be included in analysis, can be adjusted for analysis. Only loci with MAF greater than 0.10 were included in the panel at this point, but MAF within collections could be smaller than 0.01. Default P_{crit} values (0.10, 0.05, 0.02, and 0.01) were used.

The final panel of loci was evaluated for its power to distinguish the four populations in the ascertainment panel, with a particular focus on distinguishing Tucannon steelhead from Asotin steelhead. We did not have the sample sizes needed to properly evaluate the SNP panels by using holding samples and detection samples (Anderson 2010). Thus, all individuals were used to discover and evaluate the SNP panel. This may create an upward bias in the power to resolve population structure. Population structure was evaluated by calculating pair-wise F_{ST} among collections using GENEPOP v4.0. The power of the final SNP panel to assign individuals to source populations was evaluated using the fishery simulation method employed in the software ONCOR (Anderson et al. 2008).

Results

We compiled 1,036,866 unique RAD-tags from the raw sequencing data. Many RAD-tags were invariant at all nucleotides in all individuals in which it was found, thus a smaller number of SNP loci, 557,820, were identified. After filtering using the criteria defined above, 3,795 loci remained for evaluating population differentiation. Global F_{ST} values estimated for these loci ranged from a low of -0.01 to a high of 0.11. Of these, 827 F_{ST} values were significant with P -values < 0.05 before correcting for multiple tests and 199 remained significant with P -values < 0.05 after correcting for multiple tests using false discovery rate. These 199 loci comprised the final test panel.

No systematic issues were found in the final panel of loci. The average observed and expected heterozygosity for all loci within collections ranged from 0.25 to 0.28 and the average

F_{IS} ranged from 0.04 to 0.05. No locus of the final panel was estimated to be in violation of HWE in any of the four populations after correction for multiple tests at $\alpha = 0.05$ (Table 2). Each population had four, five, or six loci estimated to be in violation of HWE before multiple testing, but no locus was found in violation of HWE in more than one population. Approximately the same percent of pairwise tests among loci for LD per population had P-values < 0.01 before correction for multiple testing (between 0.75 and 1.1% of pairwise tests, out of a total of 19,701 tests). However, the Asotin collection had no significant pairwise tests, the Touchet collection had one significant test, the Lyons Ferry collection had four significant tests, and the Tucannon collection had 17 significant tests (0.09%) after correction for multiple tests. No significant pairwise tests of the same pair of loci were found in more than one population. Thus, the final panel of loci was acceptable for further analyses.

Minor allele frequencies among all populations varied from zero (one allele fixed within a population) to 0.5 (the maximum possible value for MAF). All populations had some (three to seven) loci with $MAF < 0.01$; those minor alleles were eliminated from analysis using LDNE. As expected, LDNE produced N_e estimates that increased with a decreasing P_{crit} value showing the bias that occurs when alleles with frequencies near zero or one are included. For three populations, Tucannon, Lyons Ferry, and Asotin, the N_e estimates when using $P_{crit} = 0.10$ were not dramatically different from those estimated when $P_{crit} = 0.01$ (Table 3). For the Touchet population, a lack of LD when using $P_{crit} = 0.10$ or 0.05 precluded accurate estimation of N_e (-6195.0 and 1183.1, respectively, both with infinite upper bounds). LDNE estimates using $P_{crit} = 0.02$ or 0.01 produced nearly equal estimates of N_e (506.3 and 505.0, respectively) that were likely biased slightly high.

Estimated pairwise F_{STs} using the full suite of 3,795 SNP loci were between 0.007 and 0.015 (Table 4). Estimated pairwise F_{STs} using the suite of 199 SNP loci high-graded by global F_{ST} were, as expected, much larger, from 0.015 to 0.118 (Table 4). The patterns among pairwise population estimated F_{STs} were the same using either suite of loci; the estimated pairwise F_{ST} comparing Tucannon to Asotin was the largest and the F_{ST} comparing Lyons Ferry to Touchet was the smallest with the others falling between. Estimated F_{STs} using the full suite of loci were at or below empirically established thresholds of the limit of population assignment algorithms (~ 0.02 , e.g., Hauser et al. 2006), whereas F_{STs} using the high-graded suite of loci were well above those thresholds, suggesting assignment success would be improved.

Individual assignment success, using ONCOR fishery simulations was nearly perfect ($>98\%$ success for Touchet and Lyons Ferry and $>99.99\%$ for Asotin and Tucannon) using the high-graded loci (Table 5). In comparison, assignment success using 199 loci chosen randomly from the full suite of loci was much lower, especially for the Touchet and Lyons Ferry populations (66% and 69% respectively) (Table 5).

Discussion

Using RAD-seq methods, we discovered a panel of 199 SNP loci that distinguished steelhead from among four southeast Washington populations: Tucannon River, Touchet River, Lyons Ferry Hatchery, and Asotin Creek. This panel of SNP loci had sufficient statistical power to distinguish steelhead from the Tucannon River from steelhead from Asotin Creek, the primary objective, as well as distinguishing those populations from the Touchet River and Lyons Ferry Hatchery populations.

No systematic scoring or inheritance issues were found with the high-graded loci in these populations. However, the Tucannon River population showed higher levels of pairwise locus linkage disequilibrium than the other populations. Genetic drift associated with small N_e may be one cause of LD; however the estimated N_e of the Tucannon River population was larger than that of the Lyons Ferry Hatchery population and similar in size to Asotin Creek suggesting the LD in the 17 locus-pairs in the Tucannon River is not related to N_e . Significant LD may also be present when a collection is actually a combination of more than one population (Wahlund effect, e.g., Waples 2015). Wahlund effects typically also manifest as significant tests of HWE, but statistical tests of loci in the Tucannon collection were not statistically significant (at $\alpha = 0.05$). However, the combining of populations may have occurred in previous generations; one generation of random mating eliminates HWE, whereas LD erodes more slowly (Waples 2015). Washington Department of Fish and Wildlife (WDFW), through the use of PIT tag detections in the lower Tucannon River since 2005, have documented a relatively large number of out-of-basin natural and hatchery origin steelhead entering and likely spawning (based on entry time) in the Tucannon River (Figure 3). Assuming these fish are successful at spawning, and their progeny survive to return as adults, could explain the higher levels of pairwise locus linkage disequilibrium.

A contradiction was evident in the estimated genetic relationships between the Tucannon River and Asotin Creek steelhead (Ackerman et al. 2012 vs., Blankenship et al. 2011). Our results support those of Blankenship et al. (2011) that Asotin Creek and Tucannon River steelhead are genetically differentiated. Ackerman et al. (2012) placed Asotin Creek and Tucannon River steelhead in the same Lower Snake River aggregate (aka reporting group), and suggested (personal communication) that with their panel of 188 SNPs they could not differentiate steelhead from those populations. Perhaps the main reason Ackerman et al. failed to find similar patterns as Blankenship et al. was because they did not have the same collections or populations in their evaluations. Blankenship et al. had many more collections from the aggregate wherein the Tucannon was placed than did Ackerman et al. making it difficult for Ackerman et al. to identify or accurately characterize the same aggregate. Ackerman et al. appear to have aggregated Tucannon River and Asotin Creek steelhead mainly because of a lack of a strong signal in the dendrogram produced with their genetic data (Ackerman et al. Figure 8) but also using unidentified “other information”, which was likely their geographic location. Ackerman et al. reported that their Lower Snake River aggregate had the lowest individual

assignment with ≥ 80 posterior probability to any aggregate and the lowest successful individual assignment back to the Lower Snake River aggregate (Ackerman et al. Figure 15). The low overall assignment success to any of their aggregates suggests that the aggregate is poorly defined, or rather, that not all source populations were included in the baseline. Using ONCOR assignment methods, it is possible to assign individuals with very high posterior probability to an incorrect population or aggregate, especially when the true source population is not in the baseline. This could be evaluated with estimated likelihood values; however, Ackerman et al. did not report likelihood values. It is possible to statistically test the hypothesis that individuals originated from un-sampled source populations (e.g., Paetkau et al. 2004); however, Ackerman et al. apparently did not conduct such tests. Further genotyping and testing using the IDFG SNP panel might resolve the issue.

An additional contradiction was evident in the estimated genetic relationships of Touchet and Tucannon steelhead and the Lyons Ferry Hatchery steelhead stock. Blankenship et al. (2007) reported that there was “not strong evidence for hatchery introgression in the Touchet, ... from LFH [Lyons Ferry Hatchery] based on the individual assignment results” and “[t]here is evidence for hatchery introgression in the Tucannon from LFH based on the individual assignment results”. Though we did not directly address introgression from LFH stocks to the other populations, we reported here that the 2005 Touchet River collection was more genetically similar to the 2005 Lyons Ferry Hatchery stock than to the 2005 Tucannon River or 2005 Asotin Creek steelhead collections. We also reported that, based on individual assignment tests using the high-graded panel of SNP loci or a randomly drawn panel of the same size, simulated Touchet River individuals were more likely than simulated Tucannon or Asotin individuals to be misassigned to the LFH stock. It is not possible to reconcile the contradiction at this time. Blankenship et al. (2009, 2007) reported only estimated pairwise F_{ST} values for pairwise comparisons of collections within rivers or the Lyons Ferry Hatchery stock precluding even a comparison of the relative differences among populations (a direct comparison is also difficult because different marker types were used in both studies). Blankenship et al. (2009, 2007) reported assignment test results with all individuals combined from all collection years, i.e., it is possible, for example, that the misassigned Tucannon samples in Blankenship et al. (2009, 2007) were all sampled in years other than 2005. However, that does not explain the misassignment rate of Touchet steelhead to LFH stock in our study. The difference in misassignment rates reported here and in Blankenship et al. (2009, 2007) is too large to be a sampling effect, especially given the assumption that the same samples were used in both studies. The misassignment rates were calculated using different methods – a leave-one-out procedure in Blankenship et al. (2009, 2007) and a simulation procedure in this report – however, this also seems unlikely to explain the difference in misassignment rates.

The high-graded panel of SNPs is not yet ready for use in management. Several things need to occur before the high-graded panel of SNPs should be used in management specifically of Tucannon River and Asotin Creek steelhead. First, the panel needs more laboratory development for high-throughput genotyping. This would require primer development for

genotyping by sequencing or development of TaqMan assays and testing of the primers or assays. Second, once high-throughput genotyping is possible, testing of the baseline marker panel would need to be done. Temporal stability of the baseline needs to be determined. One, or preferably two or more, years of samples from the Tucannon River and Asotin Creek need to be genotyped and compared to the current baseline. If some slight intra-population variability is present while maintaining the ability to distinguish Tucannon River and Asotin Creek steelhead, those collections can be incorporated into the baseline. A good target is 50 samples per collection, so two years of collections from two source populations equals 200 samples. Once the baseline is set, the population assignment success using samples that are not included in the baseline (Anderson 2010) needs to be determined. For this test, the more samples we can test the more precise our estimate of assignment success will be. A good target would be 50 samples from a couple of different years from each population, so around 200 samples. Third, if the baseline is stable, the baseline markers' ability to distinguish steelhead from the populations lying between the Tucannon River and Asotin Creek – Alpowa Creek, Almota Creek, or Penawawa Creek (and possibly other small tributaries) could be evaluated. Similar to previous analyses, a good target number of samples to test is 50 per population per year. It would also be prudent to test samples from populations known to produce fish that stray into the Tucannon River, Asotin Creek or the small tributaries in order to determine the power of the dataset to identify steelhead from outside populations as non-Tucannon, non-Asotin fish. The number of fish to process for this test may depend on how many populations need to be tested outside of the Lyons Ferry and Touchet River populations, which are included in this report.

A final note: the use of the panel of high-graded loci is limited. The ascertainment panel of populations was limited, which means that the high-graded loci may be of no use to distinguish populations outside those four used in this analysis. These loci are also not useful in the unbiased estimation of phylogenetic relationships. That is, parameters like genetic distances will be biased due to the high-grading of the loci.

Acknowledgements

Funding for this project was provided by the Columbia River Salmon and Steelhead Endorsement Advisory Board.

References cited

- Ackerman, M.W., McCane, J., Steele, C.A., Campbell, M.R., Matala, A.P., Hess, J.E., and Narum, S.R. 2012. Chinook and steelhead genotyping for genetic stock identification at Lower Granite Dam IDFG Report 12-15. Idaho Dept. of Fish and Game, Boise, ID.
- Anderson, E.C. 2010. Assessing the power of informative subsets of loci for population assignment: standard methods are upwardly biased. *Mol. Ecol. Res.* **10**(4): 701-710.
- Anderson, E.C., Waples, R.S., and Kalinowski, S.T. 2008. An improved method for predicting the accuracy of genetic stock identification. *Can. J. Fish. Aquat. Sci.* **65**: 1475-1486.
- Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., and Johnson, E.A. 2008. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS ONE* **3**(10): e3376.
- Blankenship, S.M., Campbell, M.R., Hess, J.E., Hess, M.A., Kassler, T.W., Kozfkay, C.C., Matala, A.P., Narum, S.R., Paquin, M.M., Small, M.P., Stephenson, J.J., Warheit, K.I., and Moran, P. 2011. Major lineages and metapopulations in Columbia River *Oncorhynchus mykiss* are structured by dynamic landscape features and environments. *Trans. Am. Fish. Soc.* **140**(3): 665-684.
- Blankenship, S.M., Small, M.P., and Bumgarner, J.D. 2009. Temporal Stability of Genetic Variation within Natural Populations of Summer Steelhead Receiving Mitigation Hatchery Fish. *Trans. Am. Fish. Soc.* **138**(5): 1052-1064.
- Blankenship, S.M., Small, M.P., Bumgarner, J.D., Schuck, M., and Mendel, G. 2007. Genetic relationships among Tucannon, Touchet, and Walla Walla River summer steelhead (*Oncorhynchus mykiss*) receiving mitigation hatchery fish from Lyons Ferry Hatchery. Washington Dept. of Fish and Wildlife, Olympia, WA.
- Bumgarner, J.D., and Dedloff, J.T. 2015. Lyons Ferry Hatchery Complex summer steelhead evaluations: 2012 run year annual report to the U.S. Fish and Wildlife Service Lower Snake River Compensation Plan Office FPA 15-06. Wash. Dept. of Fish and Wildlife.
- Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., and Cresko, W.A. 2013. Stacks: an analysis tool set for population genomics. *Mol. Ecol.* **22**(11): 3124-3140.
- Catchen, J.M., Amores, A., Hohenlohe, P., Cresko, W., and Postlethwait, J.H. 2011. Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. *G3: Genes, Genomes, Genetics* **1**(3): 171-182.
- Excoffier, L., Laval, G., and Schneider, S. 2005. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics* **1**: 47-50.
- Hauser, L., Seamons, T.R., Dauer, M., Naish, K.A., and Quinn, T.P. 2006. An empirical verification of population assignment methods by marking and parentage data: hatchery and wild

steelhead (*Oncorhynchus mykiss*) in Forks Creek, Washington, USA. *Mol. Ecol.* **15**(11): 3157-3173.

Lewis, P.O., and Zaykin, D. 2001. Genetic Data Analysis: Computer program for the analysis of allelic data. p. Free program distributed by the authors over the internet from <http://lewis.eeb.uconn.edu/lewishome/software.html>.

Paetkau, D., Slade, R., Burden, M., and Estoup, A. 2004. Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Mol. Ecol.* **13**(1): 55-65.

Raymond, M., and Rousset, F. 1995. An exact test for population differentiation. *Evolution* **49**(6): 1280-1283.

Rousset, F. 2008. GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Mol. Ecol. Res.* **8**(1): 103-106.

Storer, C.G., Pascal, C.E., Roberts, S.B., Templin, W.D., Seeb, L.W., and Seeb, J.E. 2012. Rank and Order: Evaluating the Performance of SNPs for Individual Assignment in a Non-Model Organism. *PLoS ONE* **7**(11): e49018.

Verhoeven, K.J.F., Simonsen, K.L., and McIntyre, L.M. 2005. Implementing false discovery rate control: increasing your power. *Oikos* **108**(3): 643-647.

Wang, J. 2004. Sibship Reconstruction From Genetic Data With Typing Errors. *Genetics* **166**(4): 1963-1979.

Wang, J. 2013. A simulation module in the computer program colony for sibship and parentage analysis. *Mol. Ecol. Res.*: n/a-n/a.

Wang, J., and Santure, A.W. 2009. Parentage and sibship inference from multilocus genotype data under polygamy. *Genetics* **181**(4): 1579-1594.

Waples, R.S. 2015. Testing for Hardy–Weinberg Proportions: Have We Lost the Plot? *J. Hered.* **106**(1): 1-19.

Waples, R.S., and Do, C. 2008. LDNE: a program for estimating effective population size from data on linkage disequilibrium. *Mol. Ecol. Res.* **8**(4): 753-756.

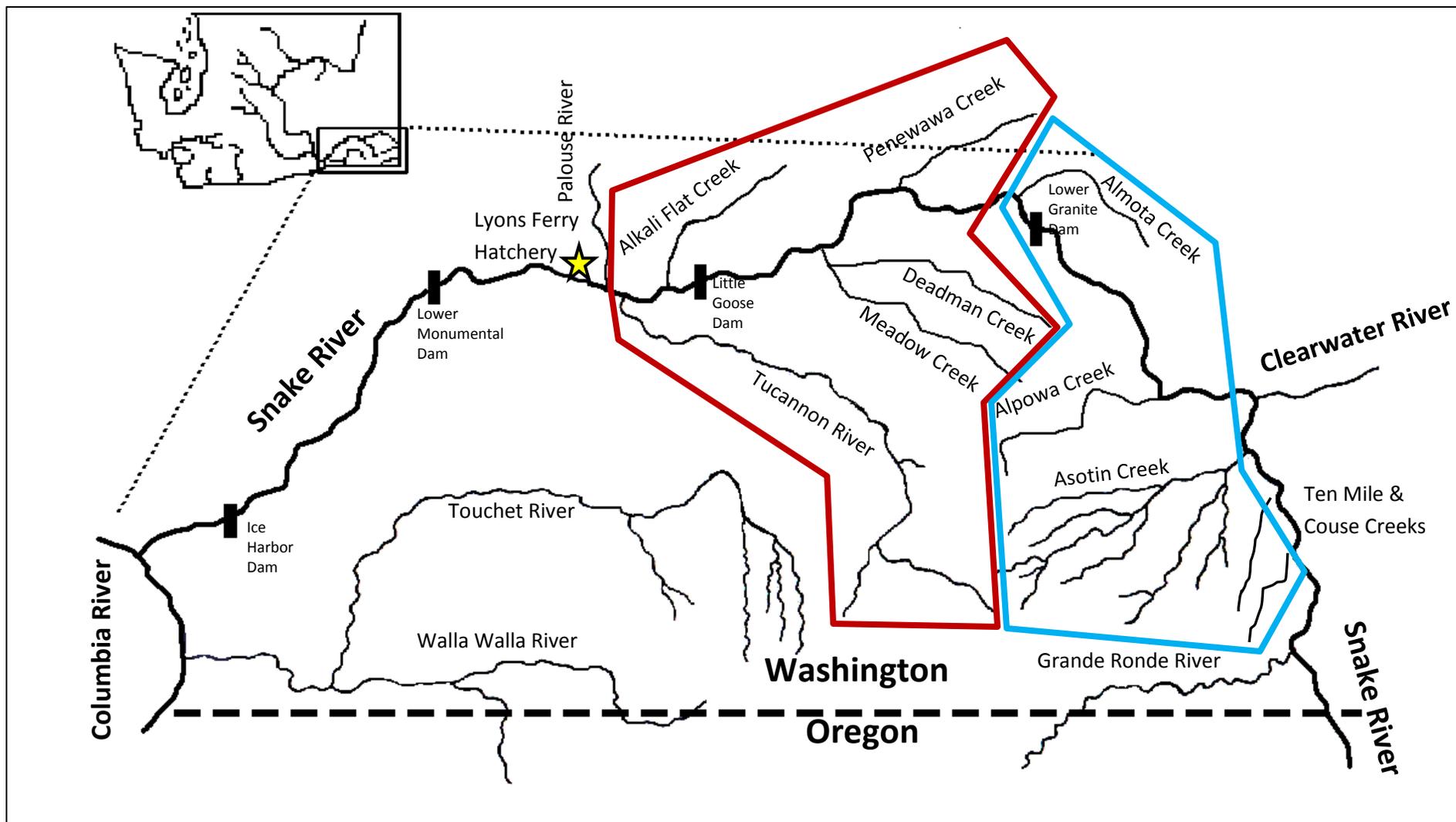


Figure 1. Map of southeast Washington showing the rivers containing managed summer steelhead populations. The red line surrounds the currently defined Tucannon Summer Steelhead population. The blue line surrounds the currently defined Asotin Creek summer steelhead population (adapted from Bumgarner and Dedloff 2015).

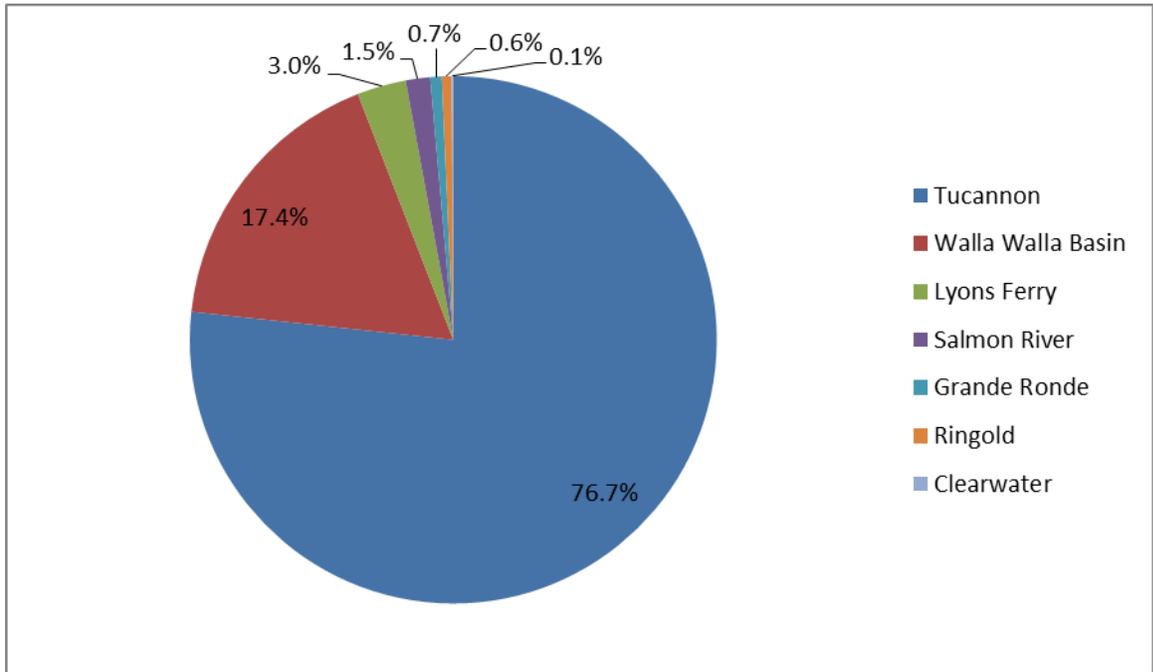


Figure 2. Source population composition of hatchery origin adult steelhead detected in the Tucannon River from PIT tag detections 2005-2015.

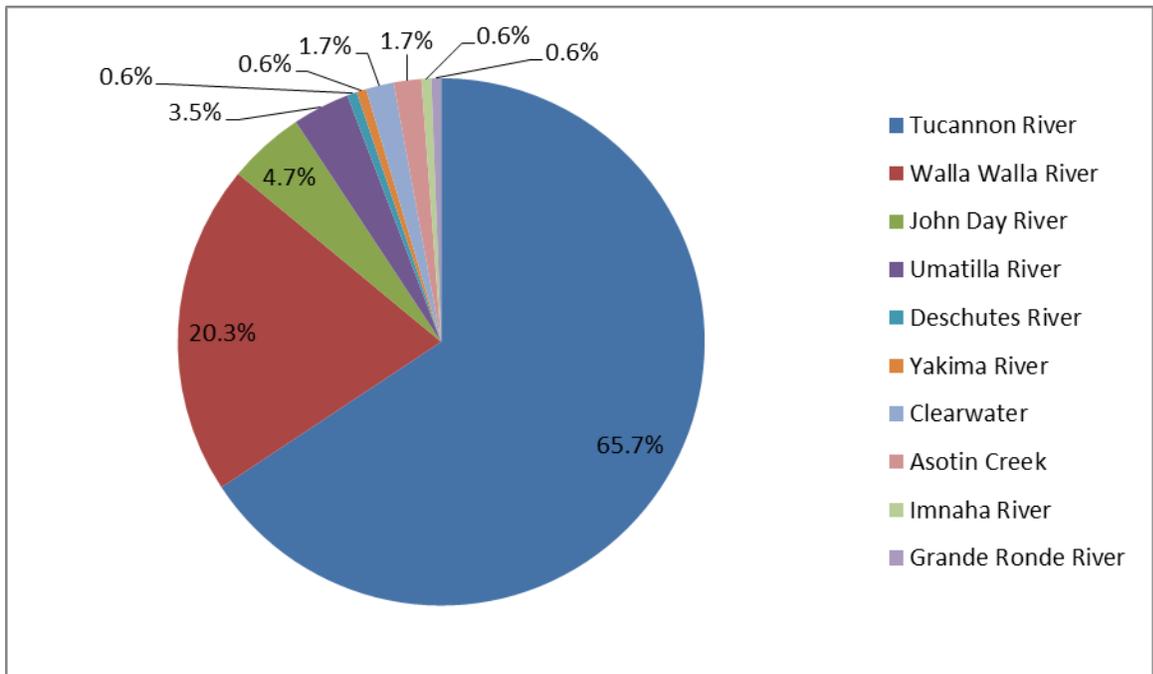


Figure 3. Source population composition of natural origin adult steelhead detected in the Tucannon River from PIT tag detections 2005-2015.

Table 1. Releases of hatchery origin summer steelhead (varied stocks) into Asotin Creek or the Tucannon River since 1983. Wells, Wallowa, Lyons Ferry, Pahsimeroi, and Oxbow stocks are all segregated programs, derived in locations other than Asotin Creek or Tucannon River. Tucannon stock is an integrated program, primarily comprised of natural origin returns to the Tucannon River (WDFW – Snake River Lab unpublished data).

Release Year	Asotin Creek		Tucannon River			
	Number Released	Stock	Number Released	Stock	Number Released	Stock
1983	36,774	Wells	148,275	Wells		
1984	33,005	Wallowa	195,315	Wells/Wallowa		
1985	31,500	Wallowa	151,609	Wallowa		
1986	44,650	Wallowa	201,293	Wells/Wallowa		
1987	22,950	Lyons Ferry	263,639	Lyons Ferry		
1988	28,975	Wallowa	161,293	Lyons Ferry		
1989	29,975	Wallowa	160,131	Lyons Ferry		
1990	137,847	Pahsimeroi	119,264	Pahsimeroi		
1991			200,761	Lyons Ferry		
1992			120,082	Lyons Ferry	9,958	Tucannon
1993	136,050	Oxbow	108,937	Lyons Ferry	4,602	Tucannon
1994	30,460	Lyons Ferry	135,359	Lyons Ferry	10,179	Tucannon
1995	35,800	Lyons Ferry	146,070	Lyons Ferry		
1996	38,500	Lyons Ferry	169,706	Lyons Ferry		
1997	39,997	Lyons Ferry	139,971	Lyons Ferry		
1998			160,068	Lyons Ferry		
1999			179,089	Lyons Ferry		
2000			145,768	Lyons Ferry		
2001			121,390	Lyons Ferry	84,968	Tucannon
2002			135,203	Lyons Ferry	58,616	Tucannon
2003			115,496	Lyons Ferry	43,688	Tucannon
2004			83,726	Lyons Ferry	42,967	Tucannon
2005			102,029	Lyons Ferry	61,238	Tucannon
2006			101,724	Lyons Ferry	65,245	Tucannon
2007			96,690	Lyons Ferry	62,940	Tucannon
2008			102,103	Lyons Ferry	59,630	Tucannon
2009			105,995	Lyons Ferry		
2010			104,646	Lyons Ferry	57,562	Tucannon
2011					122,919	Tucannon
2012					51,124	Tucannon
2013					58,357	Tucannon
2014					90,483	Tucannon

Table 2. Genetic parameters of steelhead populations of the lower Snake River. See attached MS Excel spreadsheet (SEWA_steelhead_RAD-seq_genotyping_Table_1.xlsx).

Table 3. Estimated effective population size (N_e).

Population	P_{crit}	\hat{N}_e	95% CI (Jackknife)
Tucannon	0.1	105.7	82.1-145.0
	0.05	126.9	99.8-171.3
	0.02	158.3	121.8-221.9
	0.01	169.0	128.4-242.3
Touchet	0.1	-6195.0	588.2-Infinite
	0.05	1183.1	404.7-Infinite
	0.02	506.3	283.5-2081.7
	0.01	505.0	282.8-2071.8
Lyons Ferry	0.1	73.0	61.4-88.7
	0.05	81.6	70.2-96.6
	0.02	88.1	75.6-104.7
	0.01	92.2	78.9-110.0
Asotin	0.1	95.7	77.9-122.1
	0.05	107.6	88.4-135.7
	0.02	122.6	100.6-155.1
	0.01	124.3	101.7-158.1

Table 4. Estimated F_{ST} s. Values above diagonal are F_{ST} s estimated using the full suite of loci. Values below diagonal are F_{ST} s estimated using 199 loci high-graded by global F_{ST} for increased differentiation.

Population	Tucannon	Touchet	Lyons Ferry	Asotin
Tucannon	-	0.009	0.010	0.015
Touchet	0.050	-	0.007	0.010
Lyons Ferry	0.057	0.015	-	0.009
Asotin	0.118	0.046	0.046	-

Table 5. Individual assignment success (%) from fishery simulations conducted using the software ONCOR with 199 loci high-graded for differentiation and with 199 randomly chosen loci.

High-graded loci	Actual population	Assigned population			
		Tucannon	Touchet	Lyons Ferry	Asotin
High-graded loci	Tucannon	99.995	0	0.005	0
	Touchet	0.015	98.815	1.16	0.01
	Lyons Ferry	0.005	1.655	98.34	0
	Asotin	0	0.015	0.005	99.98
Randomly chosen loci	Tucannon	94.69	2.405	2.785	0.12
	Touchet	1.485	66.385	26.08	6.05
	Lyons Ferry	1.81	25.48	69.66	3.05
	Asotin	0.06	6.675	4.505	88.76