Canadian Journal of Fisheries and Aquatic Sciences
 Journal canadien des sciences halieutiques et aquatiques

**Rapid discovery of SNPs that differentiate hatchery
 steelhead trout from ESA-listed natural-origin steelhead
 trout using a 57K SNP array**

Journal:	<i>Canadian Journal of Fisheries and Aquatic Sciences</i>
Manuscript ID	cjfas-2017-0116.R1
Manuscript Type:	Article
Date Submitted by the Author:	31-Jul-2017
Complete List of Authors:	Larson, Wesley; University of Wisconsin Stevens Point College of Natural Resources, Palti, Yniv; National Center for Cool and Cold Water Aquaculture, ARS-USDA Gao, Guangtu; National Center for Cool and Cold Water Aquaculture, ARS-USDA Warheit, Kenneth; Washington Department of Fish and Wildlife Seeb, James; University of Washington, School of Aquatic and Fishery Sciences
Is the invited manuscript for consideration in a Special Issue? :	N/A
Keyword:	GENOMICS < General, CONSERVATION < General, steelhead trout, adaptation, SNP array

SCHOLARONE™
 Manuscripts

**Rapid discovery of SNPs that differentiate hatchery steelhead trout from
ESA-listed natural-origin steelhead trout using a 57K SNP array**

Wesley A. Larson^{1,4}, Yniv Palti², Gunagtu Gao², Kenneth I. Warheit^{1,3}, and James E. Seeb^{1,5}

¹School of Aquatic and Fishery Sciences, University of Washington, 1122 NE Boat Street, Box
355020, Seattle WA 98195-5020, USA. (e-mail: jseeb@uw.edu)

²National Center for Cool and Cold Water Aquaculture, ARS-USDA, 11861 Leetown Road,
Kearneysville, WV 25430, USA. (emails: Yniv.Palti@ars.usda.gov,
Guangtu.Gao@ars.usda.gov)

³Washington Department of Fish and Wildlife, 600 Capitol Way N., Olympia, WA, 98501, USA.
(email: Kenneth.Warheit@dfw.wa.gov)

⁴Current address: U. S. Geological Survey, Wisconsin Cooperative Fishery Research Unit,
University of Wisconsin-Stevens Point, 800 Reserve St., Stevens Point, WI 54481, (email:
Wes.Larson@uwsp.edu)

⁵**Corresponding author:** James E. Seeb (e-mail: jseeb@uw.edu, phone: 1-206-685-2097)

Keywords: steelhead trout, conservation, genomics, adaptation, SNP array

Abstract

Natural-origin steelhead trout in the Pacific Northwest USA are threatened by a number of factors including habitat destruction, disease, decline in marine survival and a potential erosion of genetic viability due to introgression from hatchery strains. Our major goal was to use a recently developed SNP array containing ~57,000 SNPs to identify a subset of SNPs that differentiate hatchery- and natural-origin populations. We analyzed 35,765 polymorphic SNPs in nine populations of steelhead trout sampled from Puget Sound, Washington State, USA. We then conducted two outlier tests and found 360 loci that were candidates for divergent selection between hatchery and natural-origin populations (average $F_{CT} = 0.29$, max = 0.65) and 595 SNPs that were candidates for selection among natural-origin populations (average $F_{ST} = 0.25$, max = 0.51). Comparisons with a linkage map revealed that two chromosomes (Omy05, Omy25) contained significantly more outliers than other chromosomes, suggesting that regions on Omy05 and Omy25 may be of adaptive significance. Our results highlight several advantages of the 57K SNP array as a tool for population and conservation genomics studies.

Introduction

Many steelhead trout (*Oncorhynchus mykiss*) populations in the Pacific Northwest USA are at a fraction of their historic numbers even after more than seventy years of conservation efforts (Gayeski et al. 2011). Five of the seven distinct population segments (Waples 1995) that exist in Washington State are currently listed for protection under the US Endangered Species Act (Ford 2011). Recovery strategies originating from several sources are embedded in the Washington Department of Fish and Wildlife Statewide Steelhead trout Management Plan (WDFW 2008) where stakeholders and public and private agencies call for an increased commitment towards protecting the genetic integrity of natural-origin populations. The broad-scale hatchery and management reform proposed therein embraces the conservation genomics strategies of genetic monitoring to detect and amend practices that may encourage harmful gene flow, identification of candidate genes that control or are linked to traits affecting abundance and productivity, establishment of natural-origin population gene banks, and finally, identifying specific genetic strategies to reduce the potentially harmful effects of the segregated hatchery programs, Chambers Creek (winter run) and Skamania River (summer run) propagated throughout Washington States (WDFW 2008).

Accumulating evidence over recent decades suggests that hatchery fish in some salmonid species have lower fitness in natural environments than natural-origin fish (reviewed in Christie et al. 2014). For example, the reproductive performance of hatchery fish spawning naturally is often only a fraction of that observed in natural-origin fish (Fleming et al. 2000; McGinnity et al. 2009; Chilcote et al. 2011). Importantly, evidence suggests that this fitness decline can occur very rapidly. For example, Araki et al. (2008) show that in steelhead trout fitness declines of up to 30% can occur within the first generation or two of hatchery culture. The authors attribute this

decline to domestication selection and suggest that selection is most likely acting on multiple genetic traits throughout the life cycle (Araki et al. 2008); see also Christie et al. (2016). Results in farmed Atlantic salmon (*Salmo salar*) suggest strong selection for convergent evolution in a set of genes involved in domestication among three well-differentiated Norwegian farmed populations (Karlsson et al. 2011). In concert, these studies suggest that domestication may evolve rapidly and across differing populations due to intense selection on genes controlling traits that are advantageous in hatchery environments. Population genetic analyses based on a relatively few neutral loci or demographic data based on presumed origin of individuals (e.g. adipose fin clips) could likely fail to identify individuals carrying these signatures of domestication.

Of special interest for steelhead management is that introgression of alleles associated with domestication selection into natural-origin populations may erode fitness and stall recovery of depleted populations (Blankenship et al. 2009; Lister 2014). The effectiveness of genetic monitoring efforts to evaluate the impacts of introgression is limited by the resolving power of existing SNP panels which were not explicitly designed to differentiate hatchery- and natural-origin populations (e.g., Abadia-Cardoso et al. 2011; Hansen et al. 2011; Limborg et al. 2012; Jones et al. 2015). In general, genomic resources for Pacific salmon and trout (*Oncorhynchus spp.*) have lagged behind those available for Atlantic salmon (Barson et al. 2015; Lien et al. 2016), and to date no panel of DNA markers has been specifically designed to identify hatchery strays or hatchery hybrids in Pacific salmon, while multiple panels are available for Atlantic salmon (e.g., Glover et al. 2010; Karlsson et al. 2011; Pritchard et al. 2016). However, two recent advances in genomics may rapidly change that situation: Palti et al. (2015) present an Axiom genotyping array for 57,501 SNPs (57K array) that provides substantial opportunity to

yield high-resolution SNPs for genetic monitoring, and Campbell et al. (2015) present a high throughput genotyping platform ideal for panels of 100s of high-resolution SNPs. These advances may also facilitate the genotyping of duplicated loci resulting from a whole genome duplication event in salmon, which are removed from most studies because they are often difficult to genotype (Allendorf et al. 2015; but see Gidskehaug et al. 2011). Duplicated loci are likely involved in adaptation (Limborg et al. 2016) and may improve accuracy of population assignment (Gilbey et al. 2016).

Here, we use the 57K array to identify SNPs that differentiate between three hatchery populations derived in part from the Chambers Creek population and six steelhead trout populations of natural-origin from Puget Sound, Washington State, USA. We also investigate the genomic distribution of outlier loci that are potentially under selection and explore the utility of the 57K array for genotyping duplicated loci.

Materials and Methods

Sample collection and genotyping

Extracted DNA from 71 adult steelhead trout was obtained from Washington Department of Fish and Wildlife archives. We included samples from nine populations: six natural-origin populations of late winter run from North and Central Puget Sound, and three populations of early winter run hatchery-origin, derived in part from a now extirpated population in Chambers Creek, South Puget Sound (Crawford 1979, Table 1, Fig. 1). Genotyping was attempted for 57,501 SNPs on each sample using an Axiom genotyping array (Affymetrix, Santa Clara, California, USA) developed by Palti et al. (2015). Genotyping, genotype calling, and quality control were conducted according to Axiom genotyping procedures described by Affymetrix. Quality control thresholds for each sample were a dish QC ≥ 0.82 and call rate $\geq 97\%$. Loci

were retained if they were classified as PolyHighResolution or NoMinorHom by the Axiom genotyping algorithm (high confidence in genotyping calls) and had a minor allele frequency ≥ 0.05 in at least one population.

Analyses

We calculated allele frequencies and overall and pairwise F_{ST} (Weir and Cockerham 1984) in *Genepop* 4 (Rousset 2008) to investigate population structure. Relationships among populations were visualized with a principal coordinate analysis (PCoA) conducted in GenAIEx 6.5 (Peakall and Smouse 2006).

We conducted outlier tests in *Arlequin* 3.5 (Excoffier and Lischer 2010) in order to identify loci displaying high levels of differentiation between hatchery and natural-origin populations, and among natural-origin populations, consistent with divergent selection. The outlier test comparing hatchery- and natural-origin populations was conducted using a hierarchical island model (Excoffier et al. 2009) and both outlier tests were run using the default parameters. We combined the Sauk River and Skagit River natural-origin populations as well as all three of the hatchery populations in outlier tests to improve estimates of F -statistics because populations within these two groups showed very little genetic differentiation (see results). Loci that fell above the 99% quantile of the F_{CT} distribution (hatchery vs natural-origin outlier test) or F_{ST} distribution (among natural-origin outlier test) were considered candidates for divergent selection.

Outlier loci were located on a linkage map constructed using the 57K array (Gonzalez-Pena et al. 2016) to determine their genomic location. We then conducted chi-squared tests to identify chromosomes with more outlier SNPs than expected given the total number of SNPs

mapped to each chromosome. Relationships between total SNPs per chromosome and outlier SNPs per chromosome were visualized with scatterplots and simple linear regressions.

Fifty-one potentially duplicated loci on the Axiom array were identified by determining whether loci aligned to multiple scaffolds of the rainbow trout reference genome (Berthelot et al. 2014; alignments performed by Palti et al. 2015). We visually examined allelic discrimination plots for these loci to determine if discrete genotype clusters could be identified.

Results

Sample collection and genotyping

Sixty-one of 71 samples passed quality control thresholds for the Axiom SNP array (Table 1). Samples that failed to meet quality control standards were generally distributed evenly across populations. However, a large proportion of individuals from the Nooksack River population failed to meet QC standards, possibly due to poor sample handling. A total of 35,765 SNPs produced high quality genotypes and had a minor allele frequency > 0.05 in at least one population. On average 98.8% of these SNPs were genotyped in each individual. Genotype data from this study will be archived in DRYAD upon acceptance.

Analyses

Hatchery and natural-origin populations formed discrete clusters according to a PCoA based on pairwise F_{ST} (Fig. 2, Table S1). Additionally, differentiation was much lower among hatchery populations compared to natural-origin populations, an unsurprising result given that all three hatchery populations in this study primarily were of Chambers Creek-origin. Values of pairwise F_{ST} between natural-origin and hatchery steelhead trout ranged from 0.029 to 0.083 (Table S1). The overall F_{ST} in this study was 0.034, similar to other previous genetic analyses

which included many of the same populations (Washington Department of Fish and Wildlife, unpublished data).

We identified 360 loci that were candidates for divergent selection between hatchery and natural-origin populations; the average F_{CT} for these loci was 0.29 (Fig. 3, Table S2). Allele frequency differences between pooled hatchery-origin and pooled natural-origin populations at these loci ranged from 0.13 to 0.64 (average = 0.34, standard deviation = 0.1), and the four loci with the highest F_{CT} values displayed consistent allele frequency differences of more than 0.6 among hatchery and natural-origin populations (Fig. 4). We also identified 595 SNPs that were candidates for divergent selection among natural-origin populations (average F_{ST} = 0.25, max = 0.51).

Examination of the number of outlier SNPs found on each chromosome revealed that Omy25 contained approximately seven times more outliers than would be expected between hatchery and natural-origin populations and Omy05 contained approximately nine times more outliers than would be expected among natural-origin populations (Fig. 5, Table S3).

Comparisons of outlier loci to the most recent draft of the rainbow trout reference genome (GenBank assembly Accession GCA_002163495) indicated that the highly differentiated region on Omy05 spans approximately 40 megabases (Mb) between positions 40 and 80 Mb and the highly differentiated region on Omy25 spans approximate 35 Mb between position 35 and 70 Mb.

Most (63%) of the duplicated loci were assigned to the low quality (Other) category by the Axiom genotyping software rather than the high quality categories (PolyHighResolution or NoMinorHom). This was expected given that the Axiom genotyping algorithm is designed for genotyping duplicated loci. Nevertheless, visual examination of allelic discrimination plots for

putatively duplicated loci showed that genotyping of duplicated loci is possible with the Axiom genotyping chemistry (for example see Figure 6). It is important to note that separation of genotype clusters is more difficult as the number of clusters increases from a maximum of three in a non-duplicated SNP to a maximum of five in a duplicated SNP. It is therefore likely that a higher proportion of duplicated SNPs will have to be discarded because they cannot be reliably genotyped.

Discussion

Our results illustrate the utility of the 57K array for rapidly and effectively identifying high-resolution SNPs that can be used for a variety of applications including genetic monitoring. We identified over 300 SNPs that may display differentiation between hatchery and natural-origin steelhead trout in Puget Sound, despite the fact that pairwise- F_{ST} values between these groups are not especially large (as low as 0.029). However, since all hatchery populations were founded in part from Chambers Creek, which is located in South Puget Sound, the loci we found to differentiate between hatchery and natural-origin populations may at least in part be differentiating between South Puget Sound steelhead trout populations versus these other Puget Sound populations. Regardless of the origin of the differentiation that we observed (geography and history versus hatchery-induced selection), these loci will likely be useful for differentiating steelhead trout from hatchery- and natural-origin populations included in this study, which is a management requirement associated with the production of the hatchery fish.

The loci identified here represent excellent candidates for inclusion in a high-throughput SNP panel that can be used to differentiate hatchery- and natural-origin steelhead trout in Washington State. However, we were not able to cross-validate the patterns of differentiation observed at each locus using an independent dataset as suggested by Anderson (2010) due to sample size

limitations. It is therefore possible that some of the loci that we identified may be false positives that displayed substantial differentiation in our dataset but may not display this pattern when screened in other individuals or populations (Anderson 2010; Waples 2010). Patterns of differentiation at these loci should be validated using a training and holdout approach before they are used to differentiate hatchery- and natural-origin steelhead trout. Once appropriate loci are identified, a panel including these loci can be constructed using the genotyping-in-thousands by sequencing (GT-seq) technique recently developed by (Campbell et al. 2015). This technique utilizes highly-multiplexed PCR followed by amplicon sequencing to facilitate genotyping of thousands of individuals at hundreds of SNPs in a single sequencing lane and represents a significant cost savings over previous methods used to genotype SNPs and microsatellites (Campbell et al. 2015).

In addition to identifying candidate loci for high-throughput SNP panels, our study also provides preliminary evidence that chromosomes Omy05 and Omy25 contain regions that may be involved in adaptive differentiation among populations of steelhead trout in Puget Sound. SNPs across a large portion of chromosome Omy05 have previously been shown to be strongly associated with differences in migratory behavior within *O. mykiss* (Nichols et al. 2008; Hecht et al. 2012; Pearse et al. 2014) and have also been associated with environmental differences among populations (Pearse and Garza 2015; Abadia-Cardoso et al. 2016; Apgar et al. 2017; Leitwein et al. 2017) as well as differences in development rates (Miller et al. 2012). It is hypothesized that a chromosomal inversion that limits recombination between adaptively important alleles may be at least partially responsible for the adaptive significance of chromosome Omy05 (Pearse et al. 2014). To our knowledge, no previous studies have suggested that similar processes are occurring on chromosome Omy25. However, variation in

chromosome number has been observed across populations of steelhead trout in Washington (Thorgaard 1983; Ostberg and Thorgaard 1999), and this variation may be at least partially due to variation in chromosome Omy25 (Y. Palti, unpublished data). Specifically, Omy25, as defined in the linkage map of Gonzalez-Pena et al. (2016), has been shown to be split into two acrocentric chromosomes or maintained as a metacentric chromosome across different populations (Y. Palti, unpublished data). Although we were not able to achieve the resolution necessary to examine chromosomes Omy05 and Omy25 in detail, the preliminary results documented here suggest that structural polymorphisms such as inversions and chromosome fissions/fusions influence large regions of these two chromosomes and may be involved in adaptive differentiation of steelhead trout populations across Washington State and beyond. Structural polymorphisms can create large blocks of loci that show similar patterns of differentiation and are in strong linkage disequilibrium (Pearse et al. 2014); inclusion of multiple loci from these linkage blocks in SNP panels can result in pseudoreplication. Therefore, future panels constructed to differentiate hatchery- and natural-origin steelhead should investigate patterns of linkage disequilibrium and limit the number of SNPs included from large blocks of linked loci such as those that are likely to be found on chromosomes Omy05 and Omy25.

One potential drawback of static SNP arrays such as the 57K array used here is a reduction in the number of polymorphic loci when the panel is genotyped in populations not included in the ascertainment set (ascertainment bias, Smith et al. 2007). Ascertainment populations for the 57K SNP panel were largely derived from aquaculture strains (Palti et al. 2015). Consequently, Palti et al. (2015) found that over 40,000 loci were polymorphic in aquaculture strains. Palti et al. (2015) also investigated the number of polymorphic markers within seven natural-origin populations and found that 10,000 – 25,000 loci were polymorphic within each of these

populations (Palti et al. 2015 did not examine the number of polymorphic loci across natural-origin populations). Our data demonstrated that over 35,000 SNPs from the 57K array are polymorphic across six populations of natural-origin and three populations of hatchery-origin steelhead trout in Puget Sound. Taken together, these results suggest that studies investigating multiple rainbow trout populations across medium to large geographic scales should be able to genotype at least 30,000 polymorphic loci using the 57K array.

Although SNP arrays developed for domesticated species have been used for population and conservation genomics applications (e.g., Johnston et al. 2011; vonHoldt et al. 2011; Bourret et al. 2013), studies based on restriction site-associated DNA sequencing (RADseq) and other reduced representation methods have largely dominated this field over the last few years (but see Hubert et al. 2010; Andrews et al. 2016). RADseq is an extremely useful and cost effective tool for genotyping thousands of loci, but is not without its drawbacks (McKinney et al. 2017). For example, RADseq data can contain errors due to null alleles, PCR duplicates, and variation in coverage across loci (Andrews et al. 2016). Additionally, datasets based on RADseq are difficult to share across laboratories because few laboratories share identical analysis parameters. SNP arrays, on the other hand, represent a static and curated resource, facilitating simple exchange of genotypes among labs and comparisons to previously developed genomic resources (Palti et al. 2015; Gonzalez-Pena et al. 2016; Vallejo et al. 2016). Furthermore, the software used to analyze SNP arrays is usually developed by the company that developed the chemistry, leading to less variation in analysis methods among laboratories. Finally, the data from SNP arrays are much more manageable, lessening the need for high-powered computers, storage space for terabytes of data, and sophisticated bioinformatics. For example, the genotype data from this project was < 500 megabytes and genotype calling took < 10 min; data from a RAD project with a similar

number of individuals would likely be over 100 gigabytes and analysis would take at least 24 hours on a typical desktop or cluster. Methods that combine RAD sequencing with sequence capture such as RAD capture (Ali et al. 2015) represent an improvement over traditional RAD sequencing that should simplify data analysis and improve consistency across laboratories. However, these methods will still be affected by limitations associated with traditional RAD including PCR duplicates and variation in coverage across loci.

Another potential advantage of SNP arrays compared to RADseq data is the ability to genotype duplicated loci that may contain up to four alleles in a single individual (Gidskehaug et al. 2011; Limborg et al. 2016). Duplicated loci are likely involved in adaptive differentiation within many species, but are filtered out of most RADseq studies because they are difficult to genotype using this technique (Limborg et al. 2016). We investigated genotype clusters at a handful of duplicated loci found on the 57k array and found that clusters at many of these loci can be easily differentiated by eye. These results demonstrate that large-scale genotyping of duplicated loci will likely be possible with Axiom chemistry paired with automated genotyping approaches for tetraploid loci such as the one implemented in the R package fitTetra (Voorrips et al. 2011). Unfortunately, attempts were made to filter as many duplicated loci as possible from the 57K array developed by Palti et al. (2015); any future SNP arrays developed for salmonids should include a representative number of duplicated loci to facilitate interrogation of the important genomic regions where these loci are found.

This study demonstrates that SNP arrays such as the 57K array described by Palti et al. (2015) represent useful resources for population and conservation genomics applications. Development and genotyping costs for SNP arrays are certainly non-trivial, likely isolating their utility to more frequently studied organisms. However, these arrays provide a significant

improvement in genotyping accuracy, data transferability, and ease of analysis compared to methods such as RADseq, making them an attractive option for future studies.

Acknowledgements

Genotyping for this project was funded by a grant from the US Department of Agriculture National Animal Genome Research Program to JES and KIW. WAL was supported by the H. Mason Keeler Endowment for Excellence and a US National Science Foundation Graduate Research Fellowship (grant # DGE-0718124). Any use of trade, product, or company name is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

- Abadia-Cardoso, A., Clemento, A.J., and Garza, J.C. 2011. Discovery and characterization of single-nucleotide polymorphisms in steelhead/rainbow trout, *Oncorhynchus mykiss*. Mol. Ecol. Resour. **11**: 31-49.
- Abadia-Cardoso, A., Pearse, D.E., Jacobson, S., Marshall, J., Dalrymple, D., Kawasaki, F., Ruiz-Campos, G., and Garza, J.C. 2016. Population genetic structure and ancestry of steelhead/rainbow trout (*Oncorhynchus mykiss*) at the extreme southern edge of their range in North America. Conserv. Genet. **17**(3): 675-689.
- Ali, O.A., O'Rourke, S.M., Amish, S.J., Meek, M.H., Luikart, G., Jeffres, C., and Miller, M.R. 2015. RAD capture (Rapture): flexible and efficient sequence-based genotyping. Genetics.
- Allendorf, F.W., Bassham, S., Cresko, W., Limborg, M.T., Seeb, L.W., and Seeb, J.E. 2015. Effects of crossovers between homeologs on inheritance and population genomics in polyploid-derived salmonid fishes. J. Hered. **106**(3): 217-227.

- 317 Anderson, E.C. 2010. Assessing the power of informative subsets of loci for population
318 assignment: standard methods are upwardly biased. *Mol. Ecol. Resour.* **10**(4): 701-710.
- 319 Andrews, K.R., Good, J.M., Miller, M.R., Luikart, G., and Hohenlohe, P.A. 2016. Harnessing
320 the power of RADseq for ecological and evolutionary genomics. *Nat. Rev. Genet.* **17**(2):
321 81-92.
- 322 Apgar, T.M., Pearse, D.E., and Palkovacs, E.P. 2017. Evolutionary restoration potential
323 evaluated through the use of a trait-linked genetic marker. *Evol. Appl.* **10**(5): 485-497.
- 324 Araki, H., Berejikian, B.A., Ford, M.J., and Blouin, M.S. 2008. Fitness of hatchery-reared
325 salmonids in the wild. *Evol. Appl.* **1**(2): 342-355.
- 326 Barson, N.J., Aykanat, T., Hindar, K., Baranski, M., Bolstad, G.H., Fiske, P., Jacq, C., Jensen,
327 A.J., Johnston, S.E., Karlsson, S., Kent, M., Moen, T., Niemelä, E., Nome, T., Næsje,
328 T.F., Orell, P., Romakkaniemi, A., Sægvog, H., Urdal, K., Erkinaro, J., Lien, S., and
329 Primmer, C.R. 2015. Sex-dependent dominance at a single locus maintains variation in
330 age at maturity in salmon. *Nature* **528**(7582): 405-408.
- 331 Berthelot, C., Brunet, F., Chalopin, D., Juanchich, A., Bernard, M., Noel, B., Bento, P., Da Silva,
332 C., Labadie, K., Alberti, A., Aury, J.-M., Louis, A., Dehais, P., Bardou, P., Montfort, J.,
333 Klopp, C., Cabau, C., Gaspin, C., Thorgaard, G.H., Boussaha, M., Quillet, E., Guyomard,
334 R., Galiana, D., Bobe, J., Volff, J.-N., Genet, C., Wincker, P., Jaillon, O., Crollius, H.R.,
335 and Guiguen, Y. 2014. The rainbow trout genome provides novel insights into evolution
336 after whole-genome duplication in vertebrates. *Nat. Commun.* **5**: 3657.
- 337 Blankenship, S.M., Small, M.P., and Bumgarner, J.D. 2009. Temporal stability of genetic
338 variation within natural populations of summer steelhead receiving mitigation hatchery
339 fish. *Trans. Am. Fish. Soc.* **138**(5): 1052-1064.

- 340 Bourret, V., Kent, M.P., Primmer, C.R., Vasemägi, A., Karlsson, S., Hindar, K., McGinnity, P.,
341 Verspoor, E., Bernatchez, L., and Lien, S. 2013. SNP-array reveals genome-wide patterns
342 of geographical and potential adaptive divergence across the natural range of Atlantic
343 salmon (*Salmo salar*). *Mol. Ecol.* **22**(3): 532-551.
- 344 Campbell, N.R., Harmon, S.A., and Narum, S.R. 2015. Genotyping-in-Thousands by sequencing
345 (GT-seq): A cost effective SNP genotyping method based on custom amplicon
346 sequencing. *Mol. Ecol. Resour.* **15**(4): 855-867.
- 347 Chilcote, M.W., Goodson, K.W., and Falcy, M.R. 2011. Reduced recruitment performance in
348 natural populations of anadromous salmonids associated with hatchery-reared fish. *Can.*
349 *J. Fish. Aquat. Sci.* **68**(3): 511-522.
- 350 Christie, M.R., Ford, M.J., and Blouin, M.S. 2014. On the reproductive success of early-
351 generation hatchery fish in the wild. *Evol. Appl.* **7**(8): 883-896.
- 352 Christie, M.R., Marine, M.L., Fox, S.E., French, R.A., and Blouin, M.S. 2016. A single
353 generation of domestication heritably alters the expression of hundreds of genes. *Nat.*
354 *Commun.* **7**: 10676.
- 355 Crawford, B.A. 1979. The origin and history of the trout brood stocks of the Washington
356 Department of Game. Olympia, WA.
- 357 Excoffier, L., Hofer, T., and Foll, M. 2009. Detecting loci under selection in a hierarchically
358 structured population. *Heredity* **103**(4): 285-298.
- 359 Excoffier, L., and Lischer, H.E.L. 2010. Arlequin suite ver 3.5: a new series of programs to
360 perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.*
361 **10**(3): 564-567.

- 362 Fleming, I.A., Hindar, K., Mjolnerod, I.B., Jonsson, B., Balstad, T., and Lamberg, A. 2000.
363 Lifetime success and interactions of farm salmon invading a native population. Proc.
364 Biol. Sci. **267**(1452): 1517-1523.
- 365 Ford, M.J. 2011. Status review update for Pacific salmon and steelhead listed under the
366 Endangered Species Act: Pacific Northwest. U.S. Dept. Commer., NOAA Tech. Memo
367 **NMFS-NWFSC-113**: 281.
- 368 Gayeski, N., McMillan, B., and Trotter, P. 2011. Historical abundance of Puget Sound steelhead,
369 *Oncorhynchus mykiss*, estimated from catch record data. Can. J. Fish. Aquat. Sci. **68**(3):
370 498-510.
- 371 Gidskehaug, L., Kent, M., Hayes, B.J., and Lien, S. 2011. Genotype calling and mapping of
372 multisite variants using an Atlantic salmon iSelect SNP array. Bioinformatics **27**(3): 303-
373 310.
- 374 Gilbey, J., Cauwelier, E., Coulson, M.W., Stradmeyer, L., Sampayo, J.N., Armstrong, A.,
375 Verspoor, E., Corrigan, L., Shelley, J., and Middlemas, S. 2016. Accuracy of assignment
376 of Atlantic salmon (*Salmo salar* L.) to rivers and regions in Scotland and Northeast
377 England based on single nucleotide polymorphism (SNP) markers. PLOS ONE **11**(10):
378 e0164327.
- 379 Glover, K.A., Hansen, M.M., Lien, S., Als, T.D., Hoyheim, B., and Skaala, O. 2010. A
380 comparison of SNP and STR loci for delineating population structure and performing
381 individual genetic assignment. BMC Genet. **11**.
- 382 Gonzalez-Pena, D., Gao, G., Baranski, M., Moen, T., Cleveland, B.M., Kenney, P.B., Vallejo,
383 R.L., Palti, Y., and Leeds, T.D. 2016. Genome-wide association study for identifying loci

384 that affect fillet yield, carcass, and body weight traits in rainbow trout (*Oncorhynchus*
385 *mykiss*). Front. Genet. **7**(203).

386 Hansen, M.H.H., Young, S., JØRgensen, H.B.H., Pascal, C., Henryon, M., and Seeb, J. 2011.
387 Assembling a dual purpose TaqMan-based panel of single-nucleotide polymorphism
388 markers in rainbow trout and steelhead (*Oncorhynchus mykiss*) for association mapping
389 and population genetics analysis. Mol. Ecol. Resour. **11**: 67-70.

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

393 Hubert, S., Higgins, B., Borza, T., and Bowman, S. 2010. Development of a SNP resource and a
394 genetic linkage map for Atlantic cod (*Gadus morhua*). BMC Genomics **11**: 191.

395 Johnston, S.E., McEwan, J.C., Pickering, N.K., Kijas, J.W., Beraldi, D., Pilkington, J.G.,
396 Pemberton, J.M., and Slate, J. 2011. Genome-wide association mapping identifies the
397 genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep
398 population. Mol. Ecol. **20**(12): 2555-2566.

399 Jones, M.H., Seeb, J.E., Warheit, K.I., Seamons, T.R., Quinn, T.P., and Seeb, L.W. 2015.
400 Consequences of emergence timing for the growth and relative survival of steelhead fry
401 from naturally spawning wild and hatchery parents. Trans. Am. Fish. Soc. **144**(5): 977-
402 989.

403 Karlsson, S., Moen, T., Lien, S., Glover, K.A., and Hindar, K. 2011. Generic genetic differences
404 between farmed and wild Atlantic salmon identified from a 7K SNP-chip. Mol. Ecol.
405 Resour. **11 Suppl 1**: 247-253.

- 406 Leitwein, M., Garza, J.C., and Pearse, D.E. 2017. Ancestry and adaptive evolution of
407 anadromous, resident, and adfluvial rainbow trout (*Oncorhynchus mykiss*) in the San
408 Francisco bay area: application of adaptive genomic variation to conservation in a highly
409 impacted landscape. *Evol. Appl.* **10**(1): 56-67.
- 410 Lien, S., Koop, B.F., Sandve, S.R., Miller, J.R., Kent, M.P., Nome, T., Hvidsten, T.R., Leong,
411 J.S., Minkley, D.R., Zimin, A., Grammes, F., Grove, H., Gjuvsland, A., Walenz, B.,
412 Hermansen, R.A., von Schalburg, K., Rondeau, E.B., Di Genova, A., Samy, J.K.A., Olav
413 Vik, J., Vigeland, M.D., Caler, L., Grimholt, U., Jentoft, S., Inge Våge, D., de Jong, P.,
414 Moen, T., Baranski, M., Palti, Y., Smith, D.R., Yorke, J.A., Nederbragt, A.J., Tooming-
415 Klunderud, A., Jakobsen, K.S., Jiang, X., Fan, D., Hu, Y., Liberles, D.A., Vidal, R.,
416 Iturra, P., Jones, S.J.M., Jonassen, I., Maass, A., Omholt, S.W., and Davidson, W.S.
417 2016. The Atlantic salmon genome provides insights into rediploidization. *Nature* **533**:
418 200-205.
- 419 Limborg, M.T., Blankenship, S.M., Young, S.F., Utter, F.M., Seeb, L.W., Hansen, M.H.H., and
420 Seeb, J.E. 2012. Signatures of natural selection among lineages and habitats in
421 *Oncorhynchus mykiss*. *Ecol. Evol.* **2**(1): 1-18.
- 422 Limborg, M.T., Seeb, L.W., and Seeb, J.E. 2016. Sorting duplicated loci disentangles
423 complexities of polyploid genomes masked by genotyping by sequencing. *Mol. Ecol.*
424 **25**(10): 2117-2129.
- 425 Lister, D.B. 2014. Natural productivity in steelhead populations of natural and hatchery origin:
426 assessing hatchery spawner influence. *Trans. Am. Fish. Soc.* **143**(1): 1-16.
- 427 McGinnity, P., Jennings, E., DeEyto, E., Allott, N., Samuelsson, P., Rogan, G., Whelan, K., and
428 Cross, T. 2009. Impact of naturally spawning captive-bred Atlantic salmon on wild

- 429 populations: depressed recruitment and increased risk of climate-mediated extinction.
430 Proc. Biol. Sci. **276**(1673): 3601-3610.
- 431 McKinney, G.J., Larson, W.A., Seeb, L.W., and Seeb, J.E. 2017. RADseq provides
432 unprecedented insights into molecular ecology and evolutionary genetics: comment on
433 Breaking RAD by Lowry et al. (2016). Mol. Ecol. Resour. **17**(3): 356-361.
- 434 Miller, M.R., Brunelli, J.P., Wheeler, P.A., Liu, S., Rexroad, C.E., III, Palti, Y., Doe, C.Q., and
435 Thorgaard, G.H. 2012. A conserved haplotype controls parallel adaptation in
436 geographically distant salmonid populations. Mol. Ecol. **21**(2): 237-249.
- 437 Nichols, K.M., Edo, A.F., Wheeler, P.A., and Thorgaard, G.H. 2008. The genetic basis of
438 smoltification-related traits in *Oncorhynchus mykiss*. Genetics **179**(3): 1559-1575.
- 439 Ostberg, C.O., and Thorgaard, G.H. 1999. Geographic distribution of chromosome and
440 microsatellite DNA polymorphisms in *Oncorhynchus mykiss* native to western
441 Washington. Copeia(2): 287-298.
- 442 Palti, Y., Gao, G., Liu, S., Kent, M.P., Lien, S., Miller, M.R., Rexroad, C.E., III, and Moen, T.
443 2015. The development and characterization of a 57K single nucleotide polymorphism
444 array for rainbow trout. Mol. Ecol. Resour. **15**(3): 662-672.
- 445 Peakall, R., and Smouse, P.E. 2006. GENALEX 6: genetic analysis in Excel. Population genetic
446 software for teaching and research. Mol. Ecol. Notes. **6**(1): 288-295.
- 447 Pearse, D.E., and Garza, J.C. 2015. You can't unscramble an egg: population genetic structure of
448 *Oncorhynchus mykiss* in the California Central Valley inferred from combined
449 microsatellite and single nucleotide polymorphism data. San Francisco Estuary and
450 Watershed Science **13**(4).

- 451 Pearse, D.E., Miller, M.R., Abadía-Cardoso, A., and Garza, J.C. 2014. Rapid parallel evolution
452 of standing variation in a single, complex, genomic region is associated with life history
453 in steelhead/rainbow trout. *Proc. Biol. Sci.* **281**(1783).
- 454 Pritchard, V.L., Erkinaro, J., Kent, M.P., Niemelä, E., Orell, P., Lien, S., and Primmer, C.R.
455 2016. Single nucleotide polymorphisms to discriminate different classes of hybrid
456 between wild Atlantic salmon and aquaculture escapees. *Evol. Appl.* **9**(8): 1017-1031.
- 457 Rousset, F. 2008. GENEPOP '007: a complete re-implementation of the GENEPOP software for
458 Windows and Linux. *Mol. Ecol. Resour.* **8**(1): 103-106.
- 459 Smith, C.T., Antonovich, A., Templin, W.D., Elfstrom, C.M., Narum, S.R., and Seeb, L.W.
460 2007. Impacts of marker class bias relative to locus-specific variability on population
461 inferences in Chinook salmon: A comparison of single-nucleotide polymorphisms with
462 short tandem repeats and allozymes. *Trans. Am. Fish. Soc.* **136**(6): 1674-1687.
- 463 Thorgaard, G.H. 1983. Chromosomal differences among rainbow trout populations. *Copeia*
464 **1983**(3): 650-662.
- 465 Vallejo, R.L., Leeds, T.D., Fragomeni, B.O., Gao, G., Hernandez, A.G., Misztal, I., Welch, T.J.,
466 Wiens, G.D., and Palti, Y. 2016. Evaluation of genome-enabled selection for bacterial
467 cold water disease resistance using progeny performance data in rainbow trout: insights
468 on genotyping methods and genomic prediction models. *Front. Genet.* **7**(96).
- 469 vonHoldt, B.M., Pollinger, J.P., Earl, D.A., Knowles, J.C., Boyko, A.R., Parker, H., Geffen, E.,
470 Pilot, M., Jedrzejewski, W., Jedrzejewska, B., Sidorovich, V., Greco, C., Randi, E.,
471 Musiani, M., Kays, R., Bustamante, C.D., Ostrander, E.A., Novembre, J., and Wayne,
472 R.K. 2011. A genome-wide perspective on the evolutionary history of enigmatic wolf-
473 like canids. *Genome. Res.* **21**(8): 1294-1305.

- Voorrips, R.E., Gort, G., and Vosman, B. 2011. Genotype calling in tetraploid species from bi-allelic marker data using mixture models. *BMC Bioinformatics* **12**(1): 172.
- Waples, R.S. 1995. Evolutionarily significant units and the conservation of biological diversity under the Endangered Species Act. *In* Evolution and the aquatic ecosystem: defining unique units in population conservation. *Edited by* J.L. Nielsen and D.A. Powers. American Fisheries Society Symposium 17, Bethesda, Maryland. pp. 8-27.
- Waples, R.S. 2010. High-grading bias: subtle problems with assessing power of selected subsets of loci for population assignment. *Mol. Ecol.* **19**(13): 2599-2601.
- WDFW. 2008. Washington Department of Fish and Wildlife statewide steelhead management plan: statewide policies, strategies, and actions, Olympia, WA
<http://wdfw.wa.gov/publications/pub.php?id=00149>.
- Weir, B.S., and Cockerham, C.C. 1984. Estimating *F*-statistics for the analysis of population-structure. *Evolution* **38**(6): 1358-1370.

Tables

Table 1: Collection location, origin, sampling year, and sample size for each population included in this study. All populations listed as Hatchery were derived in part from Chambers Creek population. Samples that failed to genotype or didn't pass quality control thresholds are not included in sample sizes. Kendall Creek Hatchery is in the North Fork Nooksack River, Marblemount Hatchery is in the Skagit River, and Tokul Creek Hatchery is in the Snoqualmie River. See Fig. 1 for geographic locations of each population.

#	Population	Origin	Sampling year(s)	Sample size
1	North Fork Nooksack River	Natural-origin	2012	4
2	Kendall Creek Hatchery	Hatchery	2001	9
3	Upper Skagit River	Natural-origin	2008, 2010, 2011	8
4	Sauk River	Natural-origin	2010	6
5	Marblemount Hatchery	Hatchery	2009, 2010	9
6	Skykomish River	Natural-origin	2013	4
7	Snoqualmie River	Natural-origin	2013	8
8	Tokul Creek Hatchery	Hatchery	2001	5
9	Green River	Natural-origin	2013	8
Total				61

Figure legends

Fig.1: Map of sampling locations. See Table 1 for additional details about each sampling site. Hatchery populations are denoted with a red hue and were derived in part from Chambers Creek (white circle).

Fig. 2: Principal coordinate analysis of population structure based on pairwise- F_{ST} calculated from all 35,765 SNPs including outliers. Population numbers are in parentheses and correspond

to Table 1. Points are also labeled with population numbers. Abbreviations are natural-origin (N), and hatchery-origin (H).

Fig. 3: Result from outlier tests conducted in *Arlequin* to find loci that were significantly differentiated (a) between hatchery vs natural-origin populations and (b) among natural-origin populations. Loci colored in red represent candidates for selection at a significance level of 0.01. The upper 99% quantile is plotted in blue.

Fig. 4: Frequency of one allele for the four SNPs that displayed the highest F_{CT} values in a comparison of natural-origin and hatchery origin steelhead trout. Populations with frequencies of 0 are fixed for the alternate allele. Population numbers are in parentheses and correspond to Table 1. Populations are ordered from left to right in the same order as in the legend. The Sauk River and Skagit populations as well as three hatchery populations were pooled for this analysis because they showed very little genetic differentiation (see Fig. 2, Table S1). All four SNPs are located on Chromosome 25.

Fig. 5: Relationship between the number of SNPs found on each chromosome of the linkage map published by Gonzalez-Pena et al. (2016) and the number of outlier SNPs found on each chromosome. Outliers in panel (a) were differentiated between hatchery and natural-origin populations, and outliers in panel (b) were differentiated among natural-origin populations (Table S3). A line of best fit is drawn through each distribution. Chromosomes containing significantly more outliers than expected are denoted in red.

536 Fig. 6: Allelic discrimination plot for a putatively duplicated locus, AX-89940714. Five discrete
537 genotype clouds can be delimited for this locus. Clouds are ordered from left to right in the same
538 order as in the legend.

Draft

Fig.1: Map of sampling locations. See Table 1 for additional details about each sampling site. All hatchery populations in the study were derived in part from Chambers Creek (denoted with a white circle).

215x279mm (300 x 300 DPI)