



# CRITFC

TECHNICAL REPORT 12-06

**Columbia River Inter-Tribal Fish Commission**  
503.238.0667  
www.critfc.org

729 NE Oregon, Suite 200  
Portland, OR 97232

## **Genetic Evaluation of Steelhead Trout (*Oncorhynchus mykiss*) in the Rock Creek Watershed of the Middle Columbia River Basin**



**Andrew Matala**

**April 13, 2012**

Appendix in Yakama Nation Annual Report to the Bonneville Power Administration, Portland Oregon, contract #43057, project #2007-156-00.

# Genetic evaluation of steelhead trout (*Oncorhynchus mykiss*) in the Rock Creek watershed of the middle Columbia River Basin

Drafted by: Andrew P. Matala, Columbia River Inter-Tribal Fish Commission

Date submitted: February 02, 2012

## ***Objective/Background***

This report summarizes a genetic evaluation of steelhead trout in Rock Creek and Squaw Creek (a tributary of Rock Creek) in the Middle Columbia River Basin (Figure 1). The watershed supports both anadromous steelhead and resident *O. mykiss*, and is part of the Mid-Columbia Evolutionarily Significant Unit (ESU) for steelhead, currently listed as “threatened” under the Endangered Species Act (Busby et al. 1996). However, status of Rock Creek steelhead is unknown (WDF and WDW, 1993). Rock Creek summer steelhead are wild origin and presumably distinct from other mid-Columbia stocks based on geographic isolation of the spawning population (WDF and WDW, 1993). Adult fish begin migrating up the Columbia River from May to November, and spawning occurs in the Rock Creek watershed from February through April. Juvenile life history information specific to this watershed remains largely undocumented (Lautz 2000). Anadromous steelhead trout are known to occur in Rock Creek up to 1/4 mile above the confluence with Quartz Creek, and in Squaw Creek up to the confluence with Harrison Creek (Figure 1).

## ***Methods***

In this evaluation, juvenile fish were characterized using two 96-SNP panels developed by CRITFC. The panels complement those used in development of a basin-wide genotypic baseline for steelhead and will aid in identifying distinctions (or similarities) between the Rock Creek/Squaw Creek (RSC) population and populations throughout the Columbia River Basin (see Ackerman et al. 2011; Hess et al. 2011 for SNP panel details). Genotypes compiled for juvenile *O. mykiss* sampled in 2008 include: n=34 from a location in Rock Creek identified as Highway 8 Bridge, n=51 from a location in Rock Creek identified as “downstream of the Bickleton Rd. Crossing”, and n=72 from Squaw Creek. The 2009-2010 juvenile *O. mykiss* collections include: n=43 from unknown locations in Rock Creek, and n=69 from unknown locations in Squaw Creek. Due to small sample sizes, the temporal samples from 2009 and 2010 were combined; collections for analysis are identified as “Rock09” and “Squaw09”. Populations used to define the baseline for these analyses are described in appendix 1.

Genomic DNA was extracted from digested tissue samples using a standard Qiagen® DNeasy™ protocol. Samples were genotyped at 192 SNP markers with Taqman chemistry and Fluidigm 96.96 dynamic array chips to generate high throughput genotyping. For additional details on locus/primer specifications, locus optimizations, and detailed laboratory methods, see: Hess et al. 2011, and Matala et al. 2011.

Descriptive statistics including allele frequencies were generated using the analysis program GenAlEx version 6.2 (Peakall and Smouse 2006). The program GENEPOP v. 3.3 (Raymond and Rousset 1995) was used to test for deviations from Hardy-Weinberg equilibrium (HWE) expectations, evaluated across SNP loci and populations. Deviations may be indicative of non-random mating (e.g., population mixtures) or possible marker amplification problems (e.g., null

alleles). The significance level ( $\alpha = 0.05$ ) was corrected for multiple tests using a modified BY-FDR method (Benjamini and Yekutieli 2001) as implemented by Narum (2006). A matrix of pairwise  $F_{ST}$  (among-group variation) for all pairs of collections was generated in GENEPOP. A pairwise matrix of Nei's standard genetic distance (Nei 1972) and an un-rooted neighbor-joining (NJ) tree were generated using PHYLIP version 3.68 (Felsenstein 2008). The NJ tree displays the relationship among all populations as their respective proximities in the tree topology, where the sum of branch lengths represents the genetic distance between any two populations. The SEQBOOT option was implemented to generate 1000 simulated data sets, and a consensus topology with bootstrap support was generated using the CONSENSE option in PHYLIP. The analysis program GenAlEx version 6.2 (Peakall and Smouse 2006) was used to conduct multivariate principal coordinate analysis (PCA) to graphically display patterns or clusters of genetic similarity among 11 analysis populations. The method reduces redundant variables into a smaller subset of the most informative, where each successive PCA axis explains proportionately less of the total variation. Generally the first 2-3 axes will reveal most of the separation among distinct groups. We used both the option to convert the  $F_{ST}$  distance matrix to a covariance matrix, and the option to standardize (dividing the covariance input by  $\sqrt{n - 1}$ ). Eigen values by axis and sample Eigen vectors from PCA analysis (GenAlEx version 6.2) were displayed graphically using a free Microsoft Excel add-in available from W. A. Kamakura, and found at <http://faculty.fuqua.duke.edu/~kamakura/bio/WagnerKamakuraDownloads.htm>

A maximum likelihood method described by Wang and Santure (2009) and implemented in the program COLONY version 2.0 (Jones and Wang 2009) was used to estimate relatedness among all sampled individuals from Rock Creek and Squaw Creek. Parameter settings for the likelihood computation were: model – 1 (inbreeding), mating system - monogamy, species - diploid, number of runs – 1, length of run – medium, analysis method – full likelihood, marker type – codominant, allelic dropout rate – 0.0001, and mutation rate – 0.0001. Only full sibling groups identified with at least 80% probability were considered and no half-sib groups were considered.

Genetic stock identification (GSI) analyses were conducted in order to ascertain the probable origin/s of the sampled Rock Creek and Squaw Creek population. Although the RSC group may comprise a distinct population within the region (e.g., the Columbia River Basin), these tests help differentiate which populations or regions represented in the SNP baseline for steelhead are most similar to the RSC group. Note that the baseline for GSI analysis did not include any lower Columbia River populations (coastal lineage). The program ONCOR version 1.0 (available at: <http://www.montana.edu/kalinowski>) was first used to conduct tests of 100% Simulations on the baseline to determine its resolving power (or accuracy) for differentiating all representative stocks in the baseline. This is achieved by simulating a “fishery mixture sample” for each baseline population, where 100% of the individuals in the sample are from the same population. ONCOR uses the method of Anderson et al. (2007) to simulate mixture genotypes (based the observed population allele frequencies), and estimate their probability of occurrence in the baseline population being evaluated. For the 100% simulations, the mixture sample size parameter was set at 200 and the number of iterations was set at 1000. In addition to 100% simulations, a ‘leave-one-out’ (LOO) individual assignment procedure was performed on the baseline using ONCOR. In this jackknife method each individual is sequentially removed from the baseline, the allele frequencies in the population are recalculated in the absence of that individual, and the removed individual is assigned back to the most likely population of origin

among all baseline populations. Both the 100% simulation and LOO methods provide the assignment accuracy to population of origin, and to reporting group; population groups established on the basis of genetic similarity and geographic proximity. Following the power analysis, the RSC groups were treated as an unknown fishery mixture and assigned origin (using the LOO method) among three reporting groups: the middle Columbia River, upper Columbia River, and Snake River.

### **Results**

A total of ten significant HWE deviations were observed across 940 tests; 5 collections and 188 SNPs (excluded sex marker and hybrid markers). Both the “Rock08-Hwy 8 Bridge” and “Rock09” collections had zero deviations. There were 2 deviations in the “Squaw09” collection, 3 deviations in the “Rock08-Bickleton” collection, and 5 deviations in the “Squaw08” collection. Results indicated no inherent population or marker related problems, although kinship may have contributed to some issues of genotypic disequilibrium (Table 1). Kinship analyses revealed 61 total groups of full siblings, of which 20 were comprised of three or more individuals. Some observed sibships spanned multiple years (i.e., individuals from the same family sampled at different ages), and a substantial number spanned multiple sampling locations, including those between Rock Creek and Squaw Creek (Table 1). This latter result provides anecdotal evidence of movement throughout the watershed and a likely verification of gene flow among resident fish.

Pairwise  $F_{st}$ , indicating the amount of variation attributable to differences between populations, did not show significant differentiation among RSC collections exclusively (mean  $F_{st}$ =0.006). The mean pairwise  $F_{st}$  across all collections evaluated in these analyses, including 36 baseline and 5 RSC, was 0.073 (Figure 2). Results of pairwise testing revealed a high degree of similarity between RSC and many collections from the middle Columbia River, upper Columbia River, and Snake River regions. Those populations within the lower Columbia River, representing the coastal lineage of *O. mykiss*, were genetically discrete and highly differentiated from all remaining populations. Therefore the RSC population can be confidently characterized as an inland lineage group, with no discernible influence from lower Columbia stocks (Figures 2, 3 and 4).

Neighbor joining tree topology and clusters of collections in PCA plots provide corroborating evidence of genetic similarity between RSC and Snake River stocks. Those showing greatest similarity include Asotin Creek, the Clearwater River, and the Grande Ronde River. Collections that are more highly differentiated from RSC include the south and middle forks of both The Salmon River and Clearwater River subbasins with the Snake River Basin. With the exception of the Klickitat River and Yakima River, middle Columbia River collections also appear to share a relatively high degree of genetic similarity with RSC population (Figures 3 and 4).

The GSI results show a high level of accuracy for baseline assignment to reporting group using the 100% simulation method. Results are less definitive based on the LOO method, where Snake River collections are relatively discrete, but middle Columbia River collections exhibit a large degree of ambiguity. Using LOO, the percent correct assignment to reporting group of origin is marginal, with the exception of the Yakima and Klickitat rivers (Table 2). Evaluation of RSC using the “unknown” mixed stock fishery approach revealed a significant Snake River influence

Nearly 80% of assignments across the 5 RSC collections were to the Snake River reporting group, and the mean assignment probability was ~87% (Figure 5). Ten percent of total assignments were to the middle Columbia River reporting group.

### ***Discussion***

These results indicate genetic similarity between collections of juvenile *O. mykiss* within the Rock Creek watershed (including the Squaw Creek Tributary) and Snake River stocks. The different analyses employed in this evaluation were generally complementary. Other studies suggest a relatively high stray rate within middle Columbia River tributaries including the Klickitat River and Deschutes River, where as many as 1/3 of a hatchery sample in the latter was recently identified (using a parentage based tagging approach; PBT) as progeny of Snake River broodstocks (pers. comm. Maureen Hess, CRITFC; pers. comm. Matt Smith, USFWS). Given that there may be a significant Snake River stray influence throughout the middle Columbia region, it stands to reason that there is likely to be difficulty in differentiating populations between the two regions. In fact, high rates of mis-assignment between the two regions in GSI tests specifically points to such a conclusion. If and when resources allow, characterization of the Rock Creek/Squaw Creek population is likely to benefit greatly from continued genetic monitoring and evaluations that include archival samples (if available) and adult sampling. Archival samples will allow a glimpse through time that may indicate significant temporal variation, where historically the population may have been well differentiated from, and/or less impacted by exogenous stocks. If it is feasible to collect adult steelhead samples, our current genetic tools will allow highly confident identification of any individuals that are the progeny of Snake River hatcheries. It is important to note that the samples analyzed to date are from juvenile fish. This indicates that if the observed genetic similarity in Rock Creek to stocks from the Snake River accurately reflects out of basin straying, those individuals represent effective migrants and a high level of gene flow (e.g., reproductive success in Rock Creek).

### ***Acknowledgments***

Contributing CRITFC staff: Lori Maxwell, Laboratory Technician; Shawn Narum, Lead Geneticist

### ***References***

- Ackerman, M., and M. Campbell. 2011. Chinook and steelhead genotyping for genetic stock identification at lower granite dam. ANNUAL PROGRESS REPORT July 1, 2010 — June 30, 2011 Submitted to Bonneville Power Administration. Contract #48347; Project # 2010-026-00
- Anderson EC, RS Waples, ST KALINOWSKI (2007). An improved method for estimating the accuracy of genetic stock identification. Canadian Journal of Fisheries and Aquatic Sciences (Accepted pending revision July 24, 2007). Benjamini Y., and D. Yekutieli. 2001. The control of false discovery rate under dependency. Annals of Statistics 29:1165-1188.
- Busby, P. J., T. C. Wainwright, G. J. Bryant, L. J. Lierheimer, R. S. Waples, F. W. Waknitz, and I. V. Lagomarsino. 1996. Status review of west coast steelhead from Washington, Idaho,

- Oregon, and California. Technical memorandum NMFW-NWFSC-27. National Oceanic and Atmospheric Administration, Seattle, Washington. (Also available from <http://www.nwr.noaa.gov/1salmon/salmesa/pubs.htm#> Status Reports).
- Felsenstein, J. 1992. PHYLIP (Phylogeny Inference Package) Version 3.5C. Department of Genetics. SK-50, University of Washington, Seattle, 98195, USA.
- Hess, J. E., N. Campbell, A. P. Matala and S. R. Narum. 2011. Genetic assessment of Columbia River Stocks: 2010 annual report. Submitted to Bonneville Power Administration. Contract #41224; Project # 2008-907-00  
<https://pisces.bpa.gov/release/documents/documentviewer.aspx?doc=P120015>
- Jones, O. and J. Wang. 2009. COLONY: a program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources* 10: 551–555.
- Lautz, K. 2000. Salmon and Steelhead Habitat Limiting Factors. Washington State Conservation Commission Report.
- Matala, A. P., J. Hess and S. R. Narum. 2011. Resolving adaptive and demographic divergence among Chinook salmon populations in the Columbia River Basin. *Transactions of the American Fisheries Society* 140:783–807.
- Narum, S. R. 2006. Beyond Bonferroni: Less conservative analyses for conservation genetics. *Conservation Genetics* 7:783-787.
- Nei, M. 1972. Genetic distance between populations. *The American Naturalist* 106:283-292.
- Peakall, R., P.E. Smouse. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6, 288-295.
- Raymond, M. and F. Rousset. 1995. Genepop: population genetics software for exact tests and ecumenicism. *Journal of Heredity* 83: 248-249.
- Wang, J., and A. W. Santure. 2009. Parentage and sibship inference from multi-locus genotype data under polygamy. *Genetics* 181: 1579-1594.
- Washington Dept. of Fisheries and Washington Department of Wildlife [WDF and WDW]. 1993. Washington State Salmon and Steelhead Stock Inventory. Appendix Three: Columbia River Stocks. Washington Department of Fisheries, Washington Department of Wildlife. 580 pp.
- Washington Department of Fish and Wildlife [WDFW]. 1998

**Table 1.** A summary of estimated relatedness. Groups are numbers of individual juvenile O. mykiss identified as full siblings using the program COLONY. Results are: a.) 20 groups of three or more full-siblings, b.) 41 sibling pairs. There were 93 individuals with no full siblings identified.

**a.)**

Sib-Group	Rock 08 (Bickleton)	Rock 08 (Hwy. 8 Bridge)	Rock 09	Squaw 08	Squaw 09	<b>Total</b>
1	---	---	---	9	2	11
2	6	3	---	---	---	9
3	2	5	---	---	---	7
4	2	2	1	1	---	6
5	2	2	---	---	---	4
6	2	1	---	1	---	4
7	2	1	---	---	---	3
8	---	---	1	2	---	3
9	---	---	---	3	---	3
10	---	---	---	3	---	3
11	---	---	---	2	1	3
12	---	---	---	---	3	3
13	---	---	---	---	3	3
14	---	3	---	---	---	3
15	4	---	---	---	---	4
16	3	---	---	---	---	3
17	4	1	---	---	---	5
18	---	---	---	2	3	5
19	---	---	---	3	2	5
20	---	---	---	4	---	4
Total	27	18	2	30	14	91

**b.)**

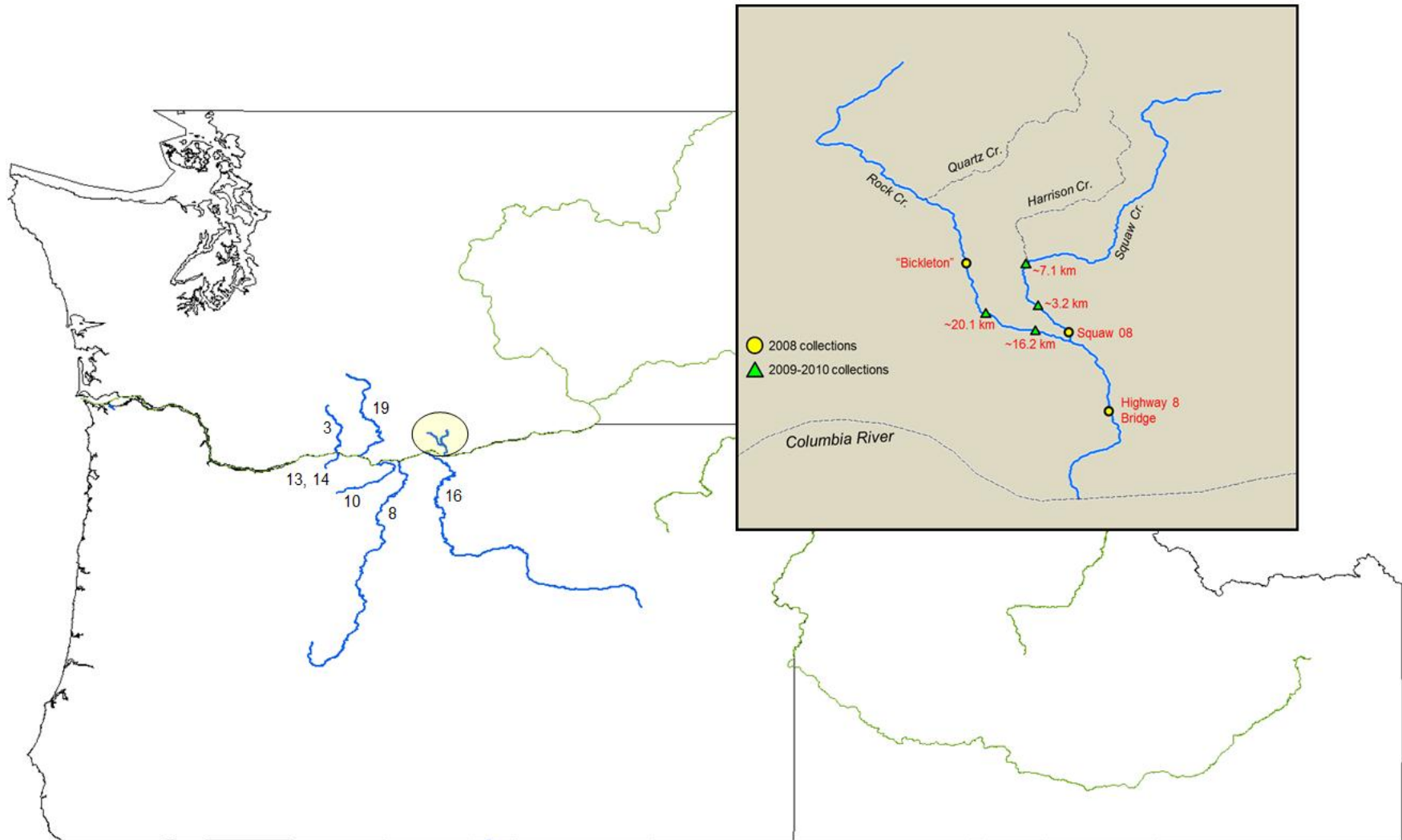
sibling pairs (n=2)	Rock 08 (Bickleton)	Rock 08 (Hwy. 8 Bridge)	Rock 09	Squaw 08	Squaw 09
Rock 08 (Bickleton)	<b>6</b>				
Rock 08 (Hwy. 8)	2	---			
Rock 09	1	---	<b>6</b>		
Squaw 08	---	4	3	<b>8</b>	
Squaw 09	---	---	---	2	<b>9</b>



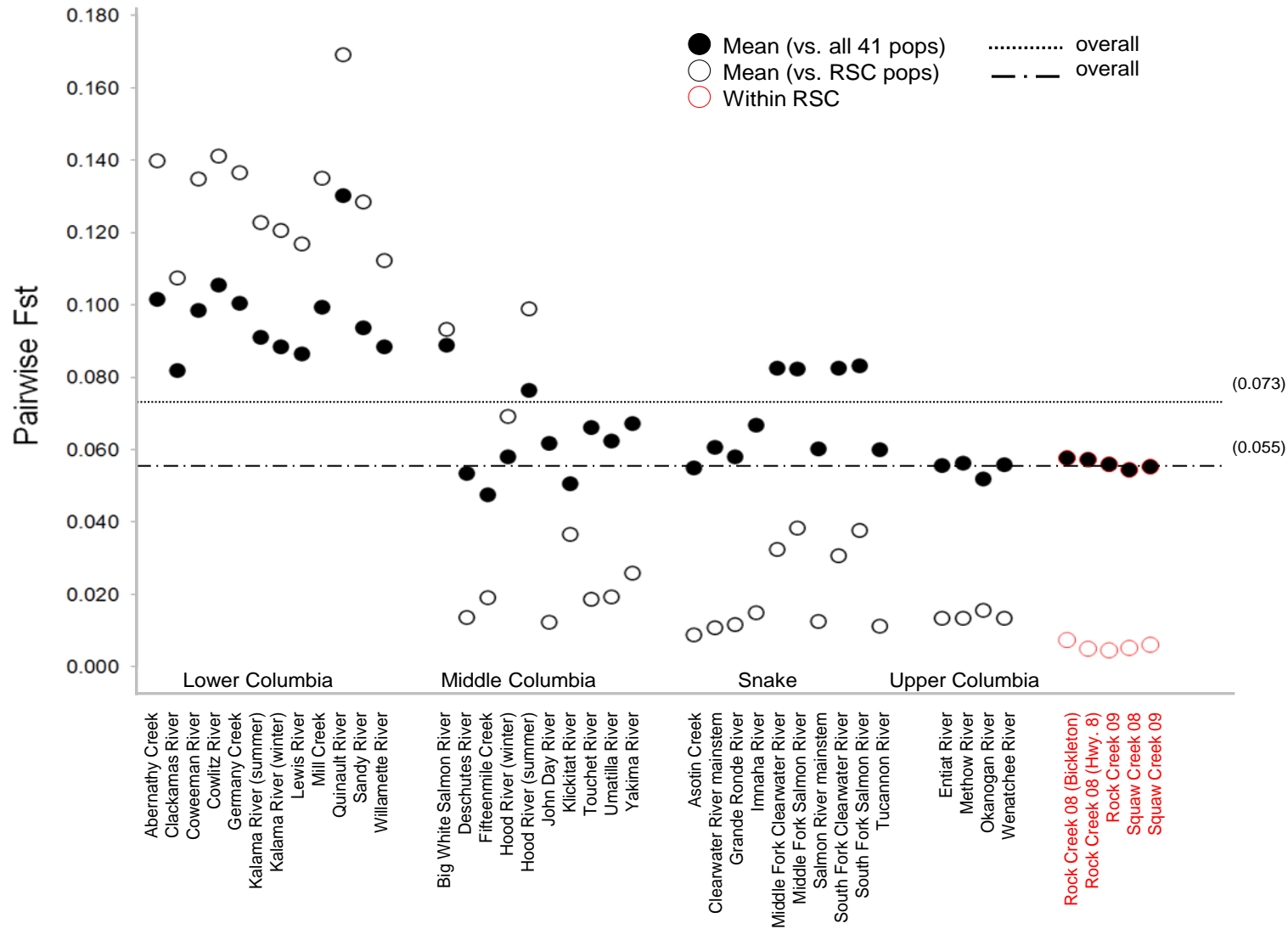
**Table 2.** Assignment tests and genetic stock ID. Values less than 0.75 appear in bold italics. Populations of the inland lineage were not included in GSI analyses. Results were generated using two methods for assignment to reporting group implemented in ONCOR. In the jackknife or leave-one-out (LOO) method results depict number of total individuals per collection that assigned to their reporting group (region) of origin. The simulation results are the average rate of correct assignment to reporting group across tests.

Collection	Reporting Group	<u>LOO</u>		<u>100% simulation test</u>		
		(N)	% correct	Average	SD	(95% C.I.)
Deschutes	Middle Columbia	76	<b><i>0.553</i></b>	0.982	0.014	(0.949 , 1.000)
Fifteenmile	Middle Columbia	35	<b><i>0.657</i></b>	0.993	0.007	(0.974 , 1.000)
John-Day	Middle Columbia	123	<b><i>0.585</i></b>	0.968	0.022	(0.911 , 0.999)
Klickitat	Middle Columbia	156	0.929	1.000	0.001	(0.995 , 1.000)
Umatilla	Middle Columbia	12	<b><i>0.583</i></b>	0.930	0.025	(0.883 , 0.971)
Yakima	Middle Columbia	108	0.852	0.996	0.006	(0.980 , 1.000)
Entiat	Upper Columbia	123	<b><i>0.626</i></b>	0.971	0.016	(0.935 , 0.996)
Methow	Upper Columbia	44	0.795	0.937	0.029	(0.871 , 0.982)
Wenatchee	Upper Columbia	58	<b><i>0.655</i></b>	0.970	0.017	(0.938 , 0.996)
Asotin	Snake	28	0.750	0.802	0.050	(0.700 , 0.893)
Clearwater	Snake	75	0.907	0.978	0.016	(0.945 , 1.000)
Grande Ronde	Snake	177	0.814	0.978	0.017	(0.940 , 1.000)
Imnaha	Snake	69	0.899	0.986	0.013	(0.955 , 1.000)
Salmon	Snake	293	0.874	0.989	0.009	(0.966 , 1.000)
Tucannon	Snake	43	<b><i>0.651</i></b>	0.781	0.045	(0.694 , 0.886)

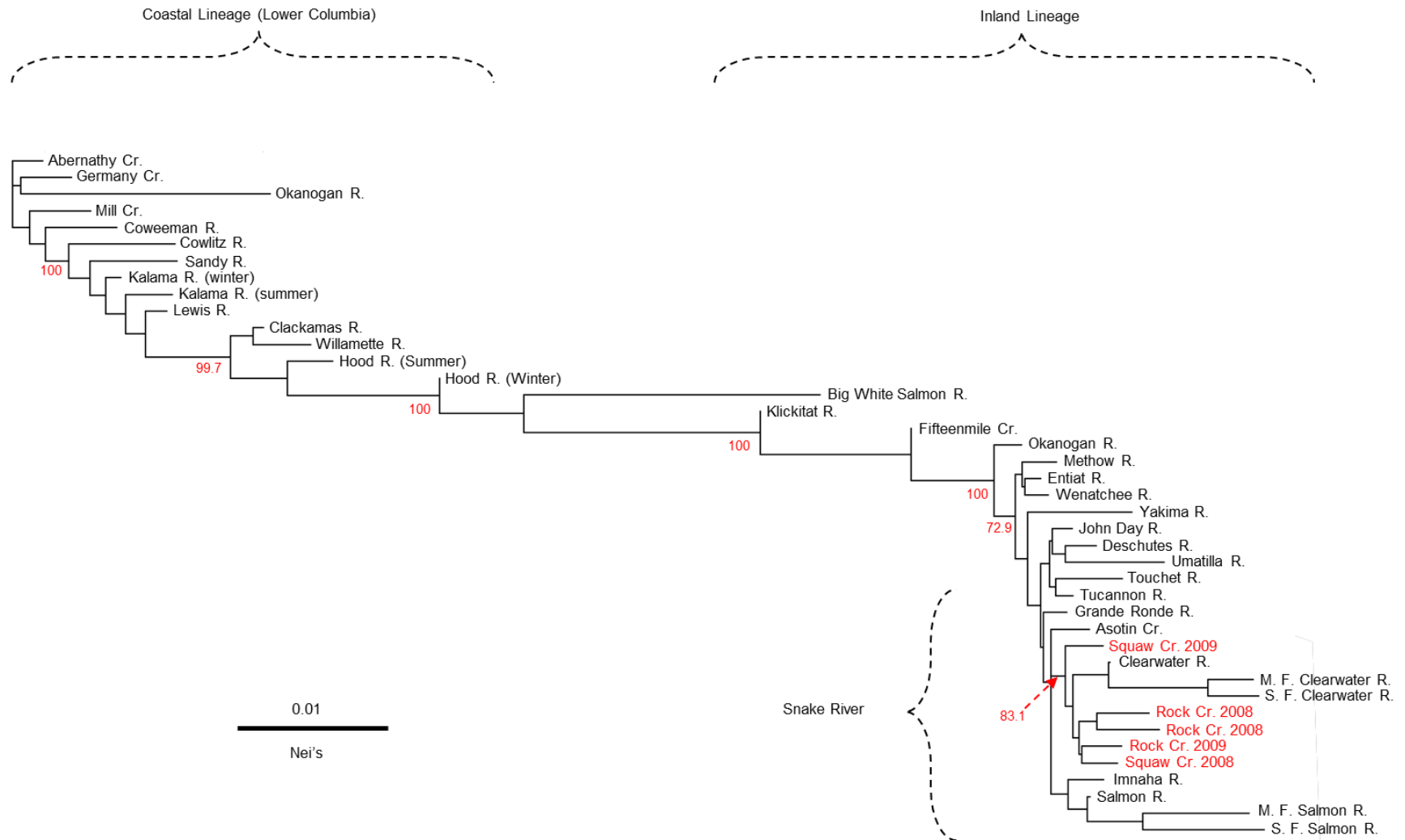
**Figure 1.** Map of the Columbia River Basin. The Rock Creek watershed is highlighted by the yellow oval (inset). Adjacent Columbia River tributaries are as follows: 3 – White Salmon River, 19 – Klickitat River, 13 & 14 – Hood River summer-run and winter-run, 10 – Fifteenmile Creek, 8 – Deschutes River, and 16 – John Day River.



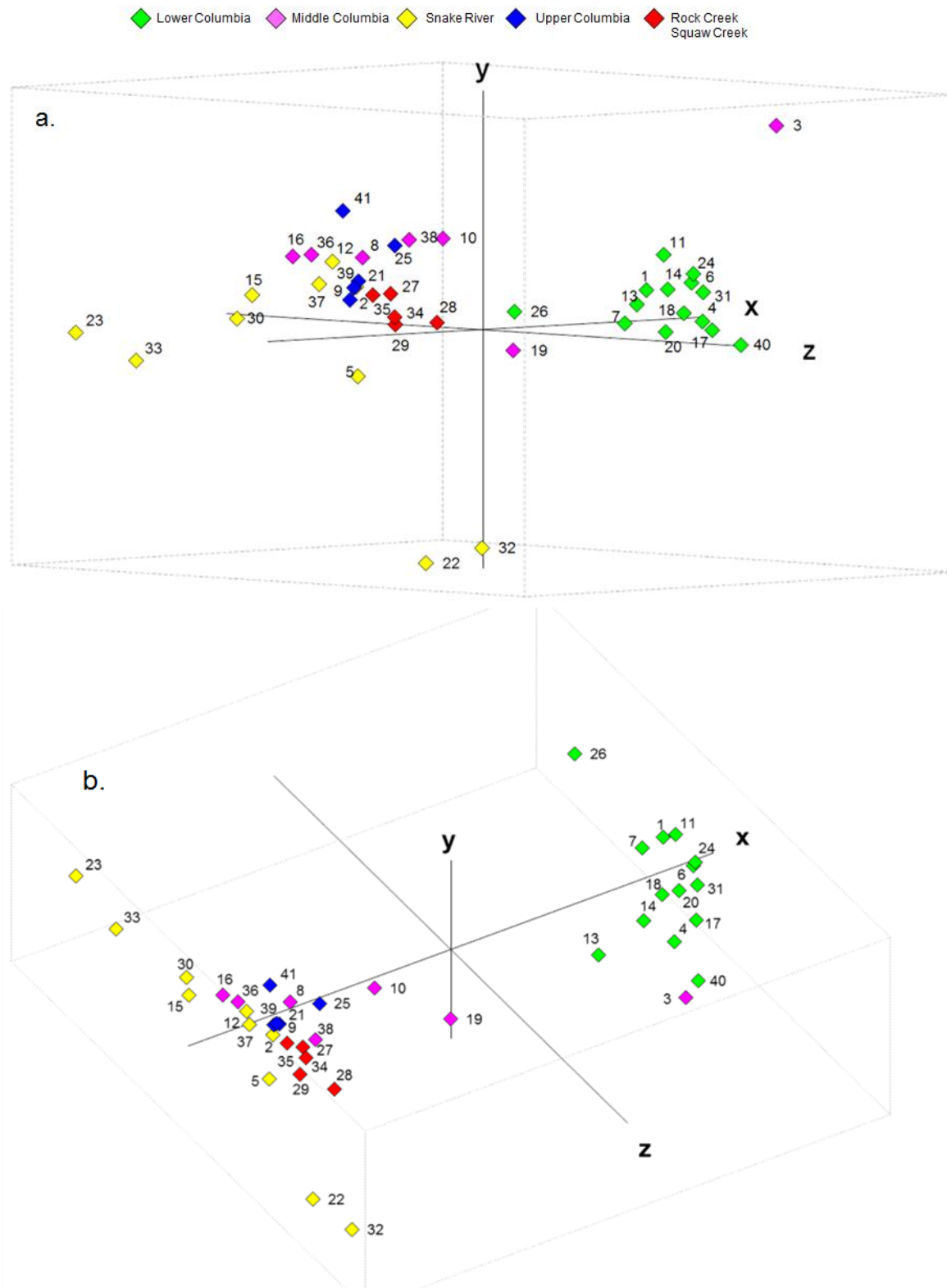
**Figure 2.** Comparison of among-group genetic variation (pairwise  $F_{st}$ ). Results are the mean of all pairwise comparisons for each of the 41 putative populations: 5 RSC collections and 36 collections from the Columbia River SNP baseline for *O. mykiss* (red text). Note that white circles represent the mean of pairwise comparisons between each collection and only the five RSC collections.



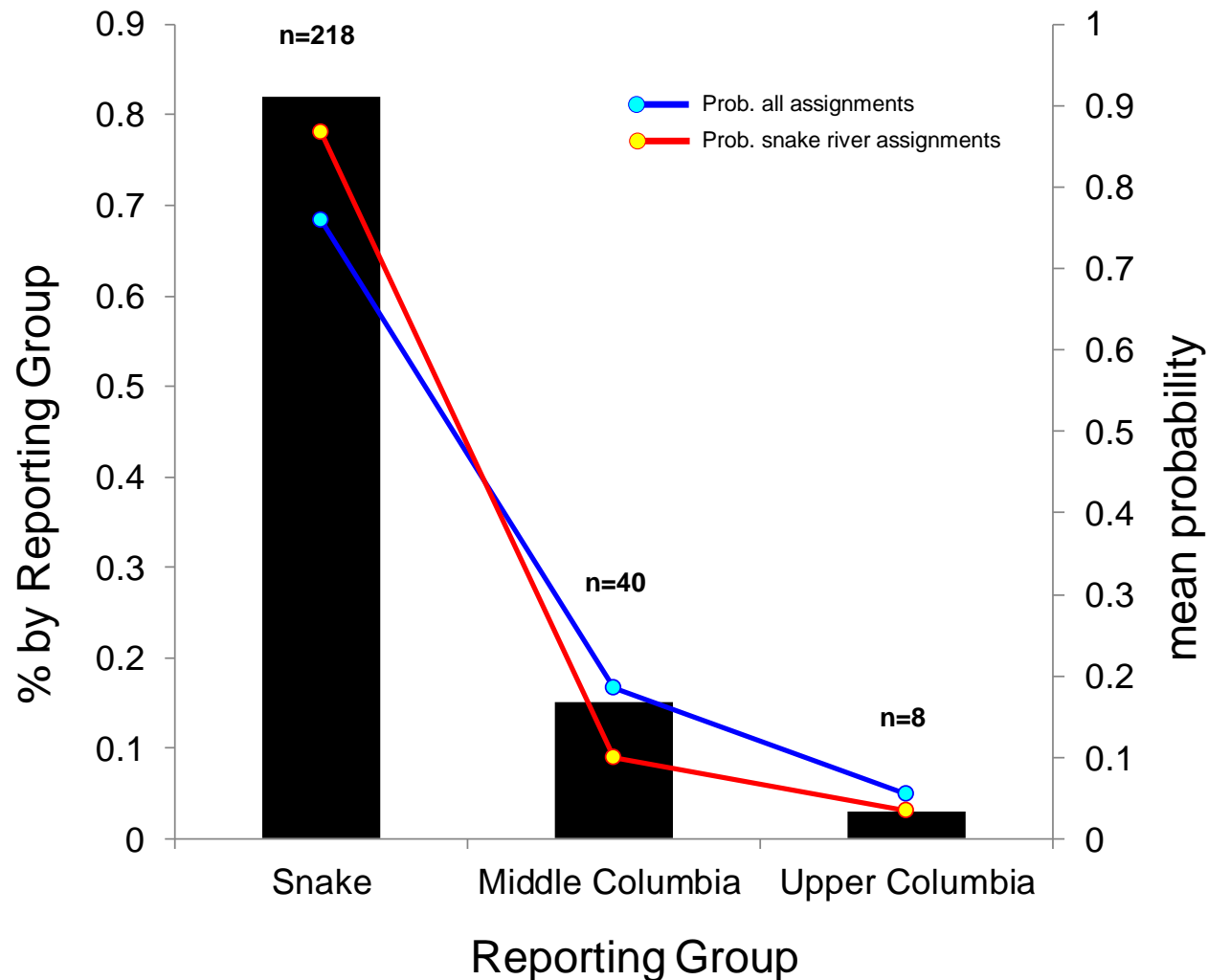
**Figure 3.** Neighbor-Joining (radial) tree topology identifying the genetic distance relationship (Nei 1987) between Rock and Squaw Creek collections and populations from the SNP baseline. Significant consensus topology across 1000 replicate data sets is shown with bootstrap support at branch nodes (red text).



**Figure 4.** Principle Coordinates Analysis (PCA) plot. Results are presented in two perspectives (a. b.), with rotation on the x- and y-axes to show multi-dimensional clustering; see appendix 1 for collection designation by number.



**Figure 5.** Mixed stock assignment results. The Rock Creek and Squaw Creek collections (n=266) were treated as a mixture of unknown origin, where each individual in the mixture was assigned to the most likely baseline population of origin using ONCOR (see Table 2). No lower Columbia (coastal lineage) populations were represented in the baseline. Results of assignment to population were grouped by GSI reporting group (region), consisting of three options: Snake River, middle Columbia and upper Columbia.



**Appendix 1.** Steelhead (*O. mykiss*) baseline groups used in these analyses. Collections in the Snake River that clustered closest with RSC in PCA and genetic distance analyses are identified with (\*).

ID#	Location / Collection	Region	Lineage	(n)
1	Abernathy Creek	lower Columbia	coastal	165
*2	Asotin Creek	Snake	Inland	49
3	Big White Salmon River	middle Columbia	Inland	81
4	Clackamas River	lower Columbia	coastal	245
*5	Clearwater River mainstem	Snake	Inland	200
6	Coweeman River	lower Columbia	coastal	47
7	Cowlitz River	lower Columbia	coastal	94
8	Deschutes River	middle Columbia	Inland	260
9	Entiat River	upper Columbia	Inland	238
10	Fifteenmile Creek	middle Columbia	Inland	91
11	Germany Creek	lower Columbia	coastal	48
*12	Grande Ronde River	Snake	Inland	426
13	Hood River (winter)	middle Columbia	Inland	54
14	Hood River (summer)	middle Columbia	Inland	35
15	Imnaha River	Snake	Inland	186
16	John Day River	middle Columbia	Inland	367
17	Kalama River (summer)	lower Columbia	coastal	94
18	Kalama River (winter)	lower Columbia	coastal	94
19	Klickitat River	middle Columbia	Inland	438
20	Lewis River	lower Columbia	coastal	173
21	Methow River	upper Columbia	Inland	92
22	Middle Fork Clearwater River	Snake	Inland	358
23	Middle Fork Salmon River	Snake	Inland	322
24	Mill Creek	lower Columbia	coastal	44
25	Okanogan River	upper Columbia	Inland	294
26	Quinault River	Washington Coast	coastal	92
<b>27</b>	<b>Rock Creek 08 (Bickleton)</b>	<b>middle Columbia</b>	<b>Inland</b>	<b>51</b>
<b>28</b>	<b>Rock Creek 08 (Hwy. 8)</b>	<b>middle Columbia</b>	<b>Inland</b>	<b>33</b>
<b>29</b>	<b>Rock Creek 09</b>	<b>middle Columbia</b>	<b>Inland</b>	<b>43</b>
30	Salmon River mainstem	Snake	Inland	548
31	Sandy River	lower Columbia	coastal	28
32	South Fork Clearwater River	Snake	Inland	167
33	South Fork Salmon River	Snake	Inland	137
<b>34</b>	<b>Squaw Creek 08</b>	<b>middle Columbia</b>	<b>Inland</b>	<b>70</b>
<b>35</b>	<b>Squaw Creek 09</b>	<b>middle Columbia</b>	<b>Inland</b>	<b>69</b>
36	Touchet River	middle Columbia	Inland	89
*37	Tucannon River	Snake	Inland	95
38	Umatilla River	middle Columbia	Inland	34
39	Wenatchee River	upper Columbia	Inland	147
40	Willamette River	lower Columbia	coastal	267
41	Yakima River	upper Columbia	Inland	262