NPAFC
Doc.1519
Rev. ___

Genetic Variation Among Major Sockeye Salmon Populations in Kamchatka Peninsula Inferred from SNP and Microsatellite DNA Analyses

by

A.M. Khrustaleva[1], M.T. Limborg[2], J.E. Seeb[2]

[1] Federal Research Institute of Fisheries and Oceanography (VNIRO),
17, V.Krasnoselskaya St., Moscow, Russia, 107140;
[2] University of Washington, School of Aquatic and Fishery Sciences,
1122 NE Boat Street, Box 355020, Seattle, WA USA 98195

Submitted to the
NORTH PACIFIC ANADROMOUS FISH COMMISSION
by
RUSSIA and **USA**

April 2014

THIS PAPER MAY BE CITED IN THE FOLLOWING MANNER:
Khrustaleva A.M., M.T. Limborg, and J.E. Seeb. 2014. Genetic variation among major sockeye salmon populations in Kamchatka peninsula inferred from SNP and microsatellite DNA analyses. NPAFC Doc 1519. 17 pp. (Available at http://www.npafc.org).

Genetic variation among major sockeye salmon populations in Kamchatka peninsula inferred from SNP and microsatellite DNA analyses

A.M. Khrustaleva[1], M.T. Limborg[2], J.E. Seeb[2]

[1]Russian Federal Research Institute of Fisheries and Oceanography (VNIRO)
[2]School of Aquatic & Fishery Sciences, University of Washington

Keywords: single nucleotide polymorphism (SNP), microsatellite loci (STR), sockeye salmon, East and West Kamchatka, differentiation

Abstract

Sockeye salmon samples from six populations from Kamchatka Peninsula were tested for polymorphism at six microsatellite (STR) and forty-five single nucleotide polymorphism (SNP) loci. These populations included the five largest populations in the region. Statistically significant genetic differentiation among the local populations from this part of the species range examined was demonstrated. The STR variability points to pronounced genetic divergence of the populations from two geographical regions, Eastern and Western Kamchatka. The results of SNP analysis further revealed that the populations of the two northern Kamchatka rivers (Palana River and Pakhacha River) differed significantly from the other populations studied. We estimated the efficiency for both types of markers for individual assignment of fish taken in mixtures. Accuracy was generally higher for assignment with SNP data; however, pooling of the STR and SNP data sets provided higher accuracy than with either one alone.

Introduction

Studies of structure and genetic differentiation as well as mixed stock analysis of commercial fish species populations represent one of the most useful molecular genetics applications in contemporary ichthyological science. Popular markers include microsatellites (short tandem repeats, STR) and SNPs. So far, both of these methods have been widely used in population studies of pacific salmon.

For example, microsatellite (STR) loci variability was fully examined in both North American (Beacham et al., 2006a, Beacham et al., 2005a, b, Beacham et al., 2011, and many others) and in Asian (Beacham et al., 2006, Khrustaleva, Zelenina, 2008, Khrustaleva et al., 2010, Pilganchuk et al., 2010, Pilganchuk, Shpigalskaya, 2013) sockeye salmon populations. Microsatellite baseline for individual genetic identification and mixed stock analysis was developed for the majority of the largest North American and Asian stocks (Beacham et al., 2005; Beacham et al., 2010). In the

last few years a large-scale study of population structure and differentiation of sockeye salmon from Bristol Bay and southwest coast of Alaska using extensive SNP-loci panel was undertaken (Ackerman et al., 2011; Creelman et. al., 2011, Gomez-Uchida et al., 2011, McGlauflin et al., 2011). The sockeye salmon SNP database developed by Alaska Department of Fish and Game in a partnership with the University of Washington (United States) currently numbers hundreds of loci and it is continually increasing (Habicht et al., 2010, Seeb et al., 2010). These same SNPs are now being used to manage fisheries in Alaska by assigning the catch to stocks of origin (Dann et al. 2013).

At the same time, examination of the genomic polymorphism of Asian sockeye salmon using SNPs is at its beginning (Gritsenko et al., 2007; Khrustaleva et al., 2010; Khrustaleva et al., 2013). Until recently, variations of SNP loci generally used in the international studies for genotyping of North American sockeye salmon were not examined in Russian populations. It is well known that the largest stocks providing more than 95% of the Russian catch of sockeye salmon, reproduce in Kamchatka. Within this territory, sockeye salmon is most abundant in the basins of the Ozernaya River (the size of the Kuril'skoe Lake population constitutes more than 70% of the total number of Asian sockeye salmon), Kamchatka River (where in certain years almost 70 to 80% of the total number of Eastern Kamchatka sockeye salmon is reproduced), rivers of the western coast of the peninsula (Bol'shaya River and Palana River), and the rivers of Koryak Plateau (Pakhacha River and others) (Burgner, 1991; Bugaev, 1995).

The present study was focused on the analysis of STR and SNP polymorphism in the populations of sockeye salmon from East and West Kamchatka, and on the assessment of genetic differentiation of the largest sockeye salmon populations from the Kamchatka Peninsula (Ozernaya River, Bol'shaya River, Palana River, Kamchatka River, and Pakhacha River), as well as on the efficiency of the markers selected for performing population assignment.

Materials and methods

Comparative analysis of sockeye salmon population structure from the five largest river-lake systems of Kamchatka Peninsula was carried out using both types of molecular markers. Specifically, polymorphism at six microsatellite loci (One-108, One-109, One-111, One-114, OtsG253b, and OMM-1082), as well as at forty-five SNP loci (Habicht et al., 2010, Seeb et al., 2010; Smith et al., 2005; Elfstrom et al., 2006; Habicht et al., 2010) was examined. The samples were collected in 2003 during mass spawning run of summer sockeye salmon in the lower reach of Ozernaya, Bol'shaya, and Palana rivers (Kamchatka, western coast), and in 2004 through 2005 in the outlets of Kamchatka River and Pakhacha River (Kamchatka, eastern coast). Brood stock of the spring sockeye salmon from the basin of Kamchatka River, migrating to the spawning grounds in Bushuev River, was sampled in Azabach'e Lake on July 3 and 13, 2004 (Fig. 1, Table 1).

Total genomic DNA was extracted from liver and fin samples by standard methods (Maniatis et al., 1982).

Polymorphism of microsatellite DNA loci was examined using the methods described in (Khrustaleva, Zelenina, 2008) and (Zelenina et al., 2008).

Polymorphism of SNP loci was typed using TaqMan-PCR. Molecular genetic analysis was carried out at the Laboratory of Ecological Genomics, School of Aquatic and Fishery Sciences, University of Washington. Tissue samples were transferred to USA under the NPAFC agreement on Sample and Data Exchanges (request № U12-05). The genotyping method is described in detail in (Seeb et al., 2009; Khrustaleva et al., 2013). The One_ prefix accepted for designation of sockeye salmon SNP loci is omitted for brevity.

The main population genetic indices inferred from the analysis of STR and SNP polymorphisms were estimated using the GENEPOP 3.4 (Raymond, Rousset, 1995), FSTAT 2.9.3 (Goudet, 2001), Microsatellite analyzer (MSA) 3.12 (Dieringer, Schlotterer, 2002), Populations 1.2.30 (Langella, 2002), TreeView 1.6.6 (Page, 1996), and Arlequin 2.0 (Schneider et al., 2000) software programs. The test for Hardy–Weinberg equilibrium and linkage disequilibrium test were performed using the Markov Chain method implemented in the GENEPOP 3.4 program. Estimates of allelic diversity corrected for the unified minimal sample size (allelic richness) were calculated in the FSTAT 2.9.3 program. Allele and heterozygosity frequencies in the samples, as well as the conformity between the observed and expected allelic richness in accordance with the two models of mutation processes at microsatellite loci, stepwise mutation model (SMM), and infinite alleles model (IAM), were determined using the MSA 3.12 program. Calculation of Nei's genetic distances and $\delta\mu^2$, as well as the NJ (Neighbor Joining) dendrogram construction were performed in the Populations 1.2.30 program with further visualization in the TreeView 1.6.6. Estimates of inter population differentiation of allelic frequencies Fst (θst), and similar estimates accounting for microsatellite allele sizes Rst (ρst) were calculated using the GENEPOP 3.4 program. The test for population assignment (likelihood ratio test) was carried out in the Arlequin 2.0 program.

Results

Microsatellite (STR) Markers

All microsatellite loci examined were highly polymorphic, and number of alleles per locus varied from 8 to 28 (16.7, on average, Table 2). The observed number of alleles per locus at the OtsG253b, One-108, and One-109 loci in all samples, as well as at the One-114 locus in five samples, was more consistent with the expected number of alleles per locus based on the stepwise mutation model (SMM). However, description of the mutation process at the One-111 and OMM-1082 loci in 70% of tests fitted the infinite alleles model (IAM). Thus, further description of the microsatellite loci polymorphism implies the use of statistics, based on the analysis of variance of allele sizes (SMM model), and allelic frequencies (IAM model). Furthermore, the latter model can be also applied to the analysis of SNP loci.

The expected heterozygosities varied in the range from 0.77 to 0.95 (Table 2). The mean expected heterozygosity and allelic diversity estimates were the lowest in the samples from Palana and Kamchatka rivers (He = 0.877; allelic richness (minimum sample size n = 52), 14.0 and 15.0, respectively). At the same time, sockeye salmon from Ozernaya River were characterized by highest genetic polymorphism (He = 0.896; number of alleles per locus, 17.0). In the samples from Pakhacha and Bol'shaya rivers, and from Azabach'e Lake the values of mean expected heterozygosity of 0.889, 0.890, and 0.894, respectively. The mean number of alleles per locus was 15.7, 15.4, and 15.6, respectively. In most of the samples examined the observed heterozygosity was not higher than the expected one. One exception was the sample from Pakhacha River, where at six loci the opposite tendency was observed (Table 2). Comparison of observed and expected genotypic distributions revealed statistically significant deficit of heterozygotes at the One-114 and One-111 loci in the populations of the Ozernaya and Palana rivers, and at the One-109 locus in the populations from Azabach'e Lake. The test for linkage disequilibrium revealed no correlation between the genotypes at any of the loci.

The probability tests for genetic differentiation between the samples revealed statistically significant inter population differences ($p << 0.003$ after correction for multiple comparisons). Pairwise comparisons of the genotypic frequencies also pointed to statistically significant genotypic differentiation of the Kamchatka populations ($p << 0.003$); statistically significant differences were observed at all of the loci examined. Differences in allelic and genotypic frequencies were not observed between the seasonal races of sockeye salmon from Kamchatka River (the samples from Kamchatka River and Azabach'e Lake).

Fst varied from 0.006 to 0.027, and Rst from –0.001 (i.e., allelic variants from different populations were more close to each other, than within one population) to 0.041 (Table 3). Based on these estimates (Table 3), it is suggested that the pattern of population genetic differentiation globally correlates with spatial geographic structure of the species within the section of the range examined. To determine whether the correlation between genetic and geographic distances was statistically significant, the hypothesis of isolation by distance was tested using Mantel's test. The testing demonstrated that gene migration between local populations of Kamchatka sockeye salmon decreased with the increase of the distance between spawning grounds ($p = 0.010/p = 0.012$; respectively, for $F_{st}(D)/R_{st}(D)$). Global differentiation level among all the samples (F_{st}/R_{st}) over all loci examined constituted 0.017/0.018 ($p = 0.0001$), and at individual locus, from 0.008 (OMM-1082)/0.0009 (One-111) to 0.033 (One-109)/0.061 (OtsG253b). We note that more prominent differences between the two regions East and West Kamchatka were revealed by use of R-statistics, whereas F-statistics disclosed sufficient inter population variability in sockeye salmon stocks within the regions (Table 4).

The unrooted NJ-tree built based on the $\delta\mu^2$ reflected differentiation of Kamchatka sockeye salmon into two large regional complexes of Eastern and Western Kamchatka (Fig. 2).

Population assignment of individuals from Eastern and Western Kamchatka was performed using the likelihood ratio test (population assignment test). The data on

samples assignment are demonstrated in Table 5. On average, brood stocks of Kamchatka sockeye salmon were accurately assigned to native populations only in 74% of the tests. The proportion of accurate population assignments was higher in the samples of Eastern Kamchatka, while probability of population discrimination from the southwest of the peninsula was rather low (66 to 67%).

SNP Markers

From 45 SNP loci examined, only four loci (p53-576, ctgf-301, RAG1-103, U404-229) were excluded from the analysis as monomorphic. Whereas in Pakhacha River sample RAG1-103 and U404-229 were found to have one variant allele for each locus, p53-576 and ctgf-301 loci were fixed for one of the alleles in all populations. Minor allele frequencies in the remaining SNPs (41 loci) were all above 1%.

Test for linkage disequilibrium revealed a correlation between the genotypes of the mitochondrial SNPs (CO1, Cytb_26, Cytb_17). Moreover the tests on loci independence were not significant ($p< 0.00007$ after Bonferroni correction) for the loci pairs: zP3b - MARCKS-241, ctgf-30 - Tf_ex3-182 in KB-03 sample, GPDH2 - GPDH in KK sample, and MHC2_190v2 and MHC2_251v2 – in the following samples: KB-03, KPh, and Kka. Thus we have a good reason to consider the three mitochondrial SNPs as a combined haplotype (one locus) - *Cytb_CO1*. There were no sufficient grounds for combining *MHC2_190v2* and *MHC2_251v2* in a linkage group because of marked differences in exon and inron evolution of MHC gene complex as well as high probability of location of recombination "hot spot" between them as previously demonstrated (Gomez-Uchida et al., 2011). Thus, after combining the linked loci, there are 39 SNP loci for further analysis (38 nuclear loci and 1 mitochondrial locus).

The detection of the loci as candidates for selection revealed five extreme F_{st} values. It was ascertained by the results of the test that the five loci (GPH-414, MHC2_251v2, MHC2_190v2, pIns-107, ALDOB-135) were candidates for diversifying selection ($p<0.01$).

Estimates of intra population genetic diversity in six sockeye salmon populaiton are given in Table 6.

Out of 228 tests for deviation from Hardy–Weinberg equilibrium (mitochondrial markers excluded), 18 tests were significant ($p< 0.05$). After Bonferroni correction the probability test revealed statistically significant departure of genotype distributions from Hardy–Weinberg proportions ($p< 0.0013$) in Kamchatka River sample on MHC2_190v2 and MHC2_251v2 loci associated with the heterozygote deficiency.

The sample examined demonstrated statistically significant heterogeneity relative to allelic and genotypic frequencies ($p << 0.003$ after Bonferroni correction). The tests for gene and genotypic population differentiation were statistically significant at almost all the loci tested with the exception of MARCKS-241 ($p= 0.412$).

The mean inter-population genetic diversity measured by the Fst value constituted 0.106 ($p = 0.0001$). The largest genetic differences were revealed between the Palana River population and local populations of southwest (Ozernaya River, Bolshaya River) and east (Kamchatka River, Pakhacha River) of Kamchatka (Table 3). Sockeye salmon from Pakhacha River differed sufficiently from the other samples and

foremost from Kamchatka River early runners. The differences between sockeye salmon samples from Southwest Kamchatka were the least (Table 3). At the individual loci, F_{st} values varied from 0.0002 (*MARCKS-241*) to 0.374 (*GPH-414*). This means that the largest inter population differentiation was detected at the latter locus. Isolation by distance of local sockeye salmon populations from Kamchatka at the SNP loci was not confirmed by Mantel's test ($p = 0.156$). Moreover it was ascertained from the results of the SNP analysis that differences among sockeye salmon populations within the regions of origin sufficiently exceeded the differences between regions (Table 4).

In NJ-tree the two clades are distinguished: Southwest Kamchatka and Kamchatka River drainage. Samples from Palana River (northwest coast of the peninsula) and Pakhacha River (the northeast coast) stand apart from the others.

The percentage of accurate population assignments of individual sockeye salmon based on the results of SNP analysis was on average higher than those inferred from six microsatellite loci (Table 5). But for neighbor populations from Southwest Kamchatka the percentage was also rather low and did not exceed 80%.

Discussion

The patterns of spatial genetic differentiation of sockeye salmon from the largest Asian population systems inferred from SNP and microsatellite DNA analyses did not correspond to each other. STR-loci polymorphism in the samples analyzed point to strong genetic divergence of the populations from two geographic regions, Eastern and Western Kamchatka. For example the dendrogram reflecting population genetic differentiation of sockeye salmon based on six microsatellite loci illustrated partitioning of Kamchatka sockeye salmon into two large regional complexes (Figure 2). On the other hand, according to the SNP data, populations of northern rivers of Kamchatka - Palana R. and Pakhacha R. considerably differed from the others (Figure 3). Estimates of inter population genetic divergence based on microsatellite loci data correlated with spatial geographic structure of the species in the part of the area examined; the degree of population differentiation based on SNP markers was not associated with geographic distance and may be associated with other factors.

The discrepancy revealed can be caused by differences in evolution patterns of the markers selected. Microsatellites localized mainly in noncoding genome regions are commonly considered as neutral markers while, in the case of SNP, their neutrality should be checked for each individual locus, taking into account that many SNPs were detected in structural genes or EST-sequences. Non-synonymous substitutions, as well as SNPs associated with adaptively important genes or localized in non-translated regulatory DNA regions are more likely to be affected by selection. This fact can cause a significant distortion in phylogenetic reconstructions and inter-population differentiation. This is because the estimates of divergence time tend to decrease when working with over dominant genes and increase if loci that experience the pressure of disruptive selection prevail (Altukhov, Salmenkova, 2002). Moreover microsatellites are characterized by a high frequency of mutations with estimates of 10^{-2}–10^{-4}, as compared to 10^{-8}–10^{-9} for SNPs; therefore, the latter are more frequently used for the

study of populations with a more prolonged period of divergence (e.g. populations of geographically distant river-lake systems). For this problem microsatellites are often inappropriate because of allelic homoplasy, i.e. fragment identity by length but not by origin. For instance the highest contribution to differentiation of Asian and American sockeye salmon at the level of large population systems, associated with lake or river basins, was made by MHC; these differences may reflect adaptive divergence of the MHC loci as has been suggested elsewhere (Ackerman et al. 2011, Gomez-Uchida et al. 2011, McLaughlin et al. 2011).

It is important to note that the analysis of polymorphism at microsatellite loci is combined with several technical difficulties related, for instance, to the determination of allele variants. The problems of choice of an adequate mathematical model and interpretation of results of microsatellite analysis are mainly determined by a complex mutation behavior of loci, homoplasy, and null-alleles.

For assessment of efficiency of Kamchatka sockeye salmon individual identification for both marker types assignment tests were carried out. According to the results of classification samples consisting of 100% individuals from a single population, the resolution of the baseline containing allele frequencies of 6 microsatellite loci and 45 SNP loci was determined. Tests showed that the identification accuracy achieved by using 45 SNP baseline was much higher than for 6 microsatellites baseline. Individual assignment success depends on the number of markers and their variation (Morin et al., 2004). Bi-allelic markers are less informative compared to multi-allelic and consequently their number should be much bigger. However, use of such markers together with microsatellites as pooled baseline for population assignment of sockeye salmon provided substantial increase of the assignment accuracy, up to 95% as in our experiment, while even a 90% level is considered rather satisfactory (Smith, Seeb, 2008).

References

Ackerman, M.W., C. Habicht, and L.W. Seeb 2011. Single-nucleotide polymorphisms (SNPs) under diversifying selection provide increased accuracy and precision in mixed-stock analyses of sockeye salmon from the Copper River. Alaska. Trans. Amer. Fish. Soc. 140(3): 865-881.

Altukhov, Yu.P. and Salmenkova E.A. 2002. DNA polymorphism in population genetics. Russ. J. Genet. 38(9): 989–1008.

Beacham, T.D., B. McIntosh, and C. Wallace. 2010. A comparison of stock and individual identification for sockeye salmon (*Oncorhynchus nerka*) in British Columbia provided by microsatellites and single nucleotide polymorphisms. Canadian Journal of Fisheries and Aquatic Sciences. 67(8): 1274-1290.

Beacham, T.D., J.R. Candy, E. Porszt, S. Sato, and S. Urawa. 2011. Microsatellite Identification of Canadian Sockeye Salmon Rearing in the Bering Sea. Trans. Amer. Fish. Soc. 140: 296-306.

Beacham, T.D., J.R. Candy, B. McIntosh, C. MacConnachie, A. Tabata, K.M. Miller, and R.E. Withler. 2005a. DNA-Level variation of sockeye salmon in Southeast Alaska and the Nass and Skeena Rivers, British Columbia, with applications to

stock identification. North American Journal of Fisheries Management. 25(3):763-776.

Beacham, T.D., J.R. Candy, B. McIntosh, C. MacConnachie, A. Tabata, K. Kaukinen, L.T. Deng, K.M. Miller, R.E. Withler, and N. Varnavskaya. 2005b. Estimation of stock composition and individual identification of sockeye salmon on a Pacific Rim basis using microsatellite and major histocompatibility complex variation. Transactions of the American Fisheries Society 134(5):1124-1146.

Beacham, T.D., B. McIntosh, C. MacConnachie, K.M. Miller, and R.E. Withler. 2006a. Pacific Rim population structure of sockeye salmon as determined from microsatellite analysis. Transactions of the American Fisheries Society 135(1):174-187.

Beacham, T.D., N.V. Varnavskaya, B.McIntosh, and C. MacConnachie. 2006b. Population structure of sockeye salmon from Russia determined with microsatellite DNA variation. Transactions of the American Fisheries Society 135(1):97-109.

Bugaev, V.F., Aziatskaya nerka (presnovodnyi period zhizni, struktura lokal'nykh stad, dinamika chislennosti) (Asian Sockeye Salmon (Freshwater Period of Life, Local Stock Structure, and Population Number Dynamics)), Moscow: Kolos, 1995. 464 p.

Burgner, R.L. 1991. Life History of Sockeye Salmon (*Oncorhynchus nerka*), Pacific Salmon Life Histories, Groot, C. and Margolis, L., Eds. UBC Press: 3–117.

Creelman, E. K., L. Hauser, R. K. Simmons, W. D. Templin, and L. W. Seeb. 2011. Temporal and geographic genetic divergence: characterizing sockeye salmon populations in the Chignik Watershed, Alaska, using single-nucleotide polymorphisms. Transactions of the American Fisheries Society 140(3):749-762.

Dann, T. H., C. Habicht, T. T. Baker, and J. E. Seeb. 2013. Exploiting genetic diversity to balance conservation and harvest of migratory salmon. Canadian Journal of Fisheries and Aquatic Sciences 70(5):785-793.

Dieringer, D., and C. Schlotterer. 2002. Microsatellite Analyzer (MSA): A Platform Independent Analysis Tool for Large Microsatellite Data Sets. Mol. Ecol. Notes. 3(4): 167–169.

Elfstrom, C. M., C. T. Smith, and J. E. Seeb. 2006. Thirty-two single nucleotide polymorphism markers for high-throughput genotyping of sockeye salmon. Molecular Ecology Notes 6(4):1255-1259.

Gomez-Uchida, D., J. E. Seeb, M. J. Smith, C. Habicht, T. P. Quinn, and L. W. Seeb. 2011. Single nucleotide polymorphisms unravel hierarchical divergence and signatures of selection among Alaskan sockeye salmon populations. BMC Evolutionary Biology 11:48. http://www.biomedcentral.com

Goudet, J. 2001 FSTAT, a Program to Estimate and Test Gene Diversities and Fixation Indices: Version 2.9.3. http://www.unil.ch/izea/softwares/fstat.html.

Gritsenko, O. F., N. V. Klovach, D. A. Zelenina, A. M. Khrustaleva, and N. S. Muge. 2007. Population structure of immature sockeye salmon in the western Bering Sea in autumn. Izvestia TINRO. 151:206–213.

Habicht, C., L. W. Seeb, K. W. Myers, E. V. Farley, and J. E. Seeb. 2010. Summer-fall distribution of stocks of immature sockeye salmon in the Bering Sea as revealed by single-nucleotide polymorphisms (SNPs). Transactions of the American Fisheries Society 139:1171-1191.

Khrustaleva, A.M. 2010. Variability of size-age indices and genetic variability of sockeye salmon *Oncorhynchus nerka* from the Western Coast of Kamchatka: comparative analysis of methods of differentiation of local populations. Vopr. Ikhtiol. 50(3): 305–317.

Khrustaleva, A.M. and D.A. Zelenina. 2008. Seasonal and interannual variations of sockeye salmon (*Oncorhynchus nerka*) microsatellite DNA in two Kamchatka lake–river systems. Russ. J. Genet. 44(7): 826–832.

Khrustaleva, A.M., O.F. Gritsenko, and N.V. Klovach. 2013. Single-nucleotide polymorphism in populations of sockeye salmon *Oncorhynchus nerka* from Kamchatka Peninsula. Russ. J. Genet. 49(11): 1155-1167.

Langella, O. 2002. Population genetic software, POPULATIONS 1.2.30. http://bioinformatics.org/~tryphon/populations

McGlauflin, M. T., D. Schindler, C. Habicht, L. W. Seeb, and J. E. Seeb. 2011. Influences of spawning habitat and geography: population structure and juvenile migration timing of sockeye salmon in the Wood River lakes. Alaska. Transactions of the American Fisheries Society 140(3):763-782.

Morin, P. A., G. Luikart, and R. K. Wayne. 2004. SNPs in ecology, evolution and conservation. Trends in Ecology & Evolution. 19(4):208-216.

Page, R.D.M. 1996. TREEVIEW: An Application to Display Phylogenetic Trees on Personal Computers. Comp. Appl. Biosci. 12:357–358.

Pilganchuk, O. A., and N. Yu. Shpigalskaya. 2013. Genetic differentiation of sockeye salmon *Oncorhynchus nerka* (Walbaum, 1792) populations of eastern Kamchatka. Russian Journal of Marine Biology. 39(5):371-379.

Pilganchuk, O. A., N. V. Varnavskaya, and T. D. Bichem 2010. Characterization of population structure on the nuclear microsatellite DNA variations for sockeye salmon from Kurilskoye lake and Kamchatka river. The researches of the aquatic biological resources of Kamchatka and of the north-west part of the Pacific Ocean. 18: 28-37.

Raymond, M. and F. Rousset. 1995. GENEPOP (Version 1.2): Population Genetics Software for Exact Tests and Ecumenicism. J. Heredity. 86: 248–249.

Schneider, S., D. Roessli, and L. Excoffier. 2000. ARLEQUIN, Version 2.000: A Software for Population Genetics Data Analysis. Geneva: Univ. Geneva.

Seeb, J. E., C. E. Pascal, R. Ramakrishnan, and L. W. Seeb. 2009. SNP genotyping by the 5'-nuclease reaction: advances in high throughput genotyping with non-model organisms. Pages 277-292 in A. Komar, editor. Methods in Molecular Biology, Single Nucleotide Polymorphisms, 2d Edition. Humana Press.

Seeb, L. W., J. E. Seeb, C. Habicht, E. V. Farley, and F. M. Utter. 2011. Single-nucleotide polymorphic genotypes reveal patterns of early juvenile migration of sockeye salmon in the Eastern Bering Sea. Transactions of the American Fisheries Society 140(3):734-748.

Smith, C. T., and L. W. Seeb. 2008. Number of alleles as a predictor of the relative assignment accuracy of short tandem repeat (STR) and single-nucleotide-polymorphism (SNP) baselines for chum salmon. Transactions of the American Fisheries Society 137:751-762.

Smith, C. T., C. M. Elfstrom, L. W. Seeb, and J. E. Seeb. 2005. Use of sequence data from rainbow trout and Atlantic salmon for SNP detection in Pacific salmon. Molecular Ecology 14(13):4193-4203.

Zelenina, D.A., N.S. Myuge, A.A. Volkov, and V.I. Sokolov. 2008. Red King Crab (*Paralithodes camtschaticus*) in the Barents Sea : A Comparative Study of Introduced and Native Populations. Russ. J. Genet. 44(7): 983–991.

Table 1. The samples of sockeye salmon from Eastern and Western Kamchatka

Population	Label	Catch dates	Sample size	
			Analysis of STR loci polymorphism	Analysis of SNP
Eastern Kamchatka, Pakhacha River	KPh	June 17–27, 2005	52	59
Eastern Kamchatka, Kamchatka River (mouth)	KK	June 29–July 9, 2004	51	95
Eastern Kamchatka, Kamchatka River basin, Azabach'e Lake	KKa	July 3, 13, 2004	58	81
Western Kamchatka, Ozernaya River	KO	August 4–7, 2003	91	95
Western Kamchatka, Bol'shaya River	KB	July 23–30, 2003	93	91
Western Kamchatka, Palana River	KP	July 10–21, 2003	96	87

Table 2. Characteristics of the STR and SNP loci examined (observed (H_o) and expected (H_e) heterozygosity, data of the probability test for fit to Hardy–Weinberg equilibrium (p), and departure of the observed genotype ratios from Hardy–Weinberg equilibrium (F_{is}))

Population	Locus	Allele size (limits)	*n* of alleles	Expected allele number		H_o	H_e	p	F_{is}
				SMM	IAN				
KB	OtsG253b	126-166	11	8.5	19.4	0.78	0.846	0.617	0.075
	One-111	186-318	22	11.3	25.9	0.928	0.893	0.294	-0.042
	OMM-1082	198-298	20	12.1	27	0.863	0.901	0.584	0.04
	One-108	180-236	15	10.4	24.1	0.871	0.881	0.123	0.009
	One-109	124-180	12	10.3	23.8	0.857	0.88	0.62	0.023
	One-114	192-292	26	18.6	39.4	0.919	0.94	0.173	0.02
KK	OtsG253b	130-162	8	7.1	13.9	0.76	0.807	0.481	0.053
	One-111	198-310	19	9.1	17.9	0.92	0.86	0.248	-0.076
	OMM-1082	234-274	11	9.8	19.3	0.857	0.874	0.784	0.014
	One-108	176-240	15	13.2	23.7	0.864	0.912	0.08	0.048
	One-109	120-176	14	10.2	19.4	0.756	0.879	0.623	0.137
	One-114	188-288	24	17.2	29.1	0.933	0.936	0.052	-0.003
KKa	OtsG253b	130-166	10	7.8	16	0.737	0.829	0.583	0.108
	One-111	198-306	22	12.3	24.4	0.875	0.903	0.204	0.027
	OMM-1082	202-278	14	10.9	21.8	0.87	0.888	0.989	0.016
	One-108	184-240	14	13.8	25.9	0.941	0.916	0.331	-0.033
	One-109	124-184	13	9.8	19.4	0.765	0.873	0.0009***	0.12
	One-114	184-288	24	23.5	37.6	0.941	0.955	0.857	0.009
KO	OtsG253b	118-162	11	7.7	16.9	0.747	0.824	0.149	0.091
	One-111	194-318	27	12	26	0.875	0.9	0.335	0.025
	OMM-1082	202-294	20	12	26.2	0.836	0.901	0.076	0.069
	One-108	180-236	15	13.1	29.9	0.852	0.91	0.221	0.061
	One-109	124-176	14	11.1	25.3	0.855	0.89	0.129	0.036
	One-114	200-316	28	22.9	46.1	0.852	0.952	0***	0.103
KP	OtsG253b	122-162	10	7.6	17.3	0.77	0.822	0.558	0.06
	One-111	198-298	18	12.7	29.1	0.885	0.907	0.0175*	0.021
	OMM-1082	202-302	15	8.5	19.7	0.805	0.846	0.766	0.047
	One-108	192-232	11	8.6	20.1	0.87	0.848	0.384	-0.029
	One-109	124-180	15	10.8	25.5	0.815	0.887	0.338	0.078
	One-114	196-296	25	22.9	46.9	0.88	0.952	0***	0.073
KPh	OtsG253b	130-162	9	6.2	11.9	0.837	0.769	0.158	-0.094
	One-111	186-306	26	22.2	36.1	0.96	0.952	0.809	-0.014
	OMM-1082	210-282	17	10.3	20.3	0.82	0.881	0.244	0.065
	One-108	188-228	11	11.5	22.4	0.961	0.896	0.47	-0.079
	One-109	124-168	10	10.7	21	0.922	0.885	0.296	-0.046
	One-114	188-288	24	22.4	36.5	0.922	0.952	0.731	0.028

Table 3. Estimates of pairwise genetic differentiation indices F_{st} (above the line) and R_{st} (below the line) over all microsatellite loci (above the diagonal) and Fst over all SNP loci (below the diagonal)

Sample	KB	KK	KKa	KO	KP	KPh
KB	-	0.0194 0.0067	0.0158 0.0298	0.0061 -0.0011	0.0179 0.0165	0.0168 0.0157
KK	0.0648	-	0.0015 0.0074	0.0171 0.0140	0.0243 0.0341	0.0124 0.0023
KKa	0.0611	0.0875	-	0.0147 0.0290	0.028 0.0512	0.0115 0.0027
KO	0.0289	0.0578	0.0838	-	0.0132 0.0008	0.0178 0.0169
KP	0.1640	0.1755	0.2024	0.1270	-	0.0269 0.0410
KPh	0.0811	0.0871	0.1416	0.0787	0.1390	-

Table 4. Genetic differentiation indices in sockeye salmon at different levels of population hierarchy inferred from STR and SNP loci

Index	STR loci (No. of different alleles method / Sum of squared size difference method)	SNP
Differentiation level (proportion of variance, %):		
Among populations (F_{st}(/R_{st}), %)	1.67/1.84	10.64
Within populations	98.33/98.16	89.36
Differentiation level with subdivision into regions of origin (proportion of variance, %):		
Between regions	0.95/2.32	0.07
Among populations within regions	1.12/0.52	10.60
Within populations	97.92/97.16	89.33
F_{st}(/R_{st})	0.021/0.028	0.107

Table 5. Population assignment test for sockeye salmon form Eastern and Western Kamchatka

Region	Locality	Proportion of accurate assignments, %		
		analysis of STR polymorphism	analysis of SNP polymorphism	pooled SNP and STR data
Western Kamchatka	Ozernaya River	70.3	72.6	94.2
	Bol'shaya River	63.4	78.0	91.6
	Palana River	77.1	95.4	98.8
Eastern Kamchatka	Kamchatka River (late run)	76.5	84.2	93.8
	Kamchatka River (early run)	75.9	91.4	98.2
	Pakhacha River	84.6	96.6	96.0
On average		74.6	86.4	95.4

Table 6. Indices of genetic diversity in sockeye salmon populations inferred from analysis of 38 nuclear SNP loci.

Sample	H_o(s.d.)	H_e(s.d.)	n_a(s.d.)	*a.r.*
KPh	0.267(0.168)	0.280(0.174)	1.98(0.16)	1.97
KK	0.255(0.163)	0.254(0.186)	1.87(0.34)	1.87
Kka	0.299(0.170)	0.255(0.185)	1.87(0.34)	1.86
KP	0.292(0.149)	0.249(0.184)	1.83(0.39)	1.82
KB	0.257(0.152)	0.261(0.177)	1.92(0.27)	1.91
KO	0.275(0.185)	0.259(0.182)	1.95(0.23)	1.95

Bottom note: H_o – observed heterozygosity, H_e - expected heterozygosity, s.d. – standard deviation, n_a – mean allelic number per locus; *a.r.* – Allelic richness (minimum sample size n = 56).

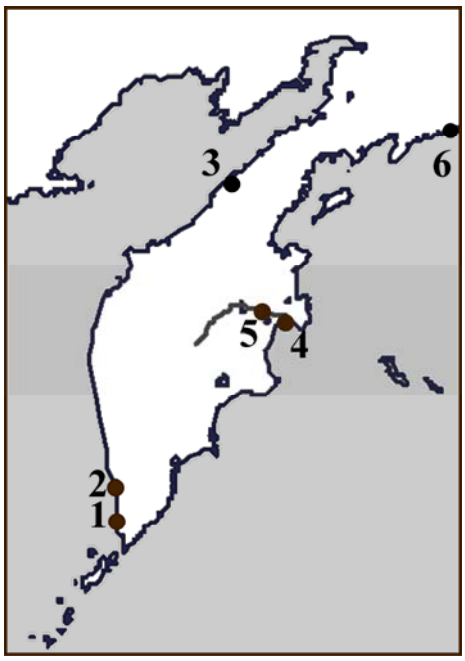

Figure 1. Schematic map of sampling location. 1—Ozernaya River; 2—Bol'shaya River; 3—Palana River; 4—Kamchatka River, mouth; 5—Kamchatka River, Azabach'e Lake; 6—Pakhacha River.

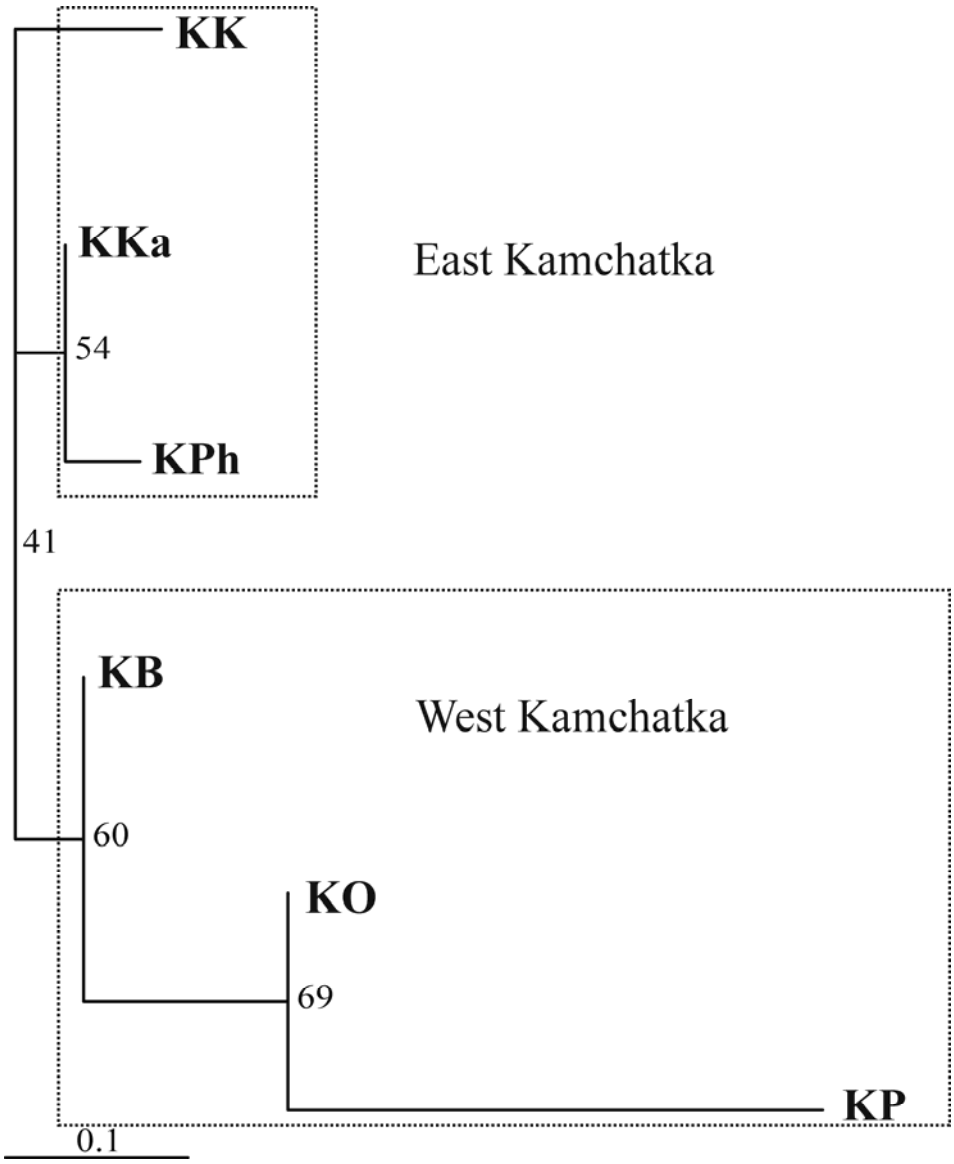

Figure 2. NJ-dendrogram built using $\delta\mu^2$ distances, based on the analysis of microsatellite loci polymorphism (in the nodes are the bootstrap support indices, 1000 iterations).

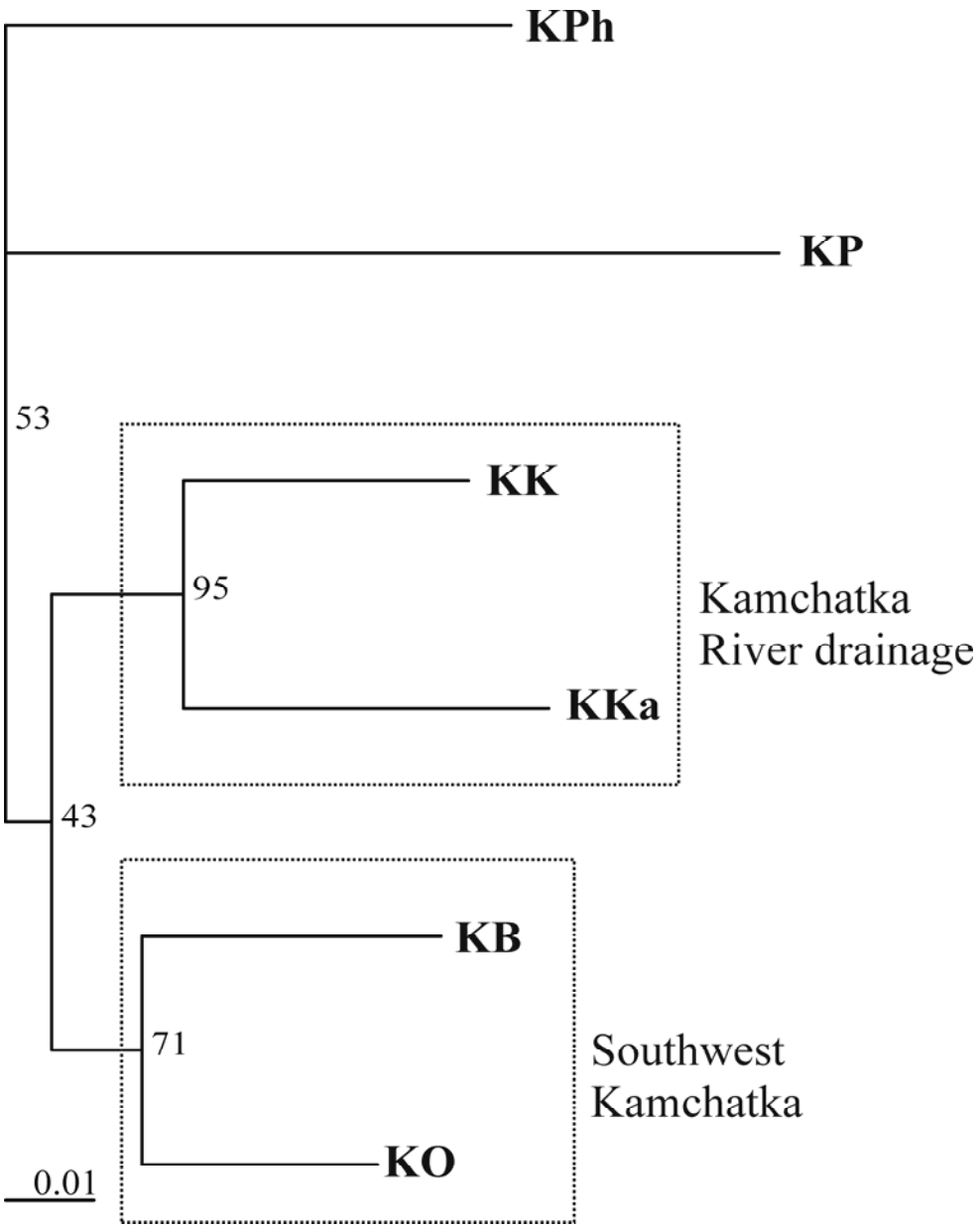

Figure 3. NJ-dendrogram constructed using chord distances, based on the analysis of SNP loci polymorphism.