Assessment of genetic variation between reproductive ecotypes of Klamath River steelhead reveals differentiation associated with different run-timings

By R. Papa[1,2], J. A. Israel[2], F. Nonnis Marzano[1] and B. May[2]

[1]*Dipartimento di Biologia Evolutiva e Funzionale, University of Parma, Parma, Italy;* [2]*Genomic Variation Laboratory, Department of Animal Science, University of California, Davis, CA, USA*

Summary

Seven microsatellite DNA loci were optimized to assess genetic differentiation in coastal steelhead (*Oncorhynchus mykiss irideus*) sampling groups from the lower Klamath River (California, USA). Genetic relationships among three winter-run and two summer-run groups were investigated. The different groups displayed high levels of allelic variation. Pairwise F_{ST} comparisons and Nei's genetic distance supported low, yet significant, genetic differentiation between summer and winter run-timings similar to other studies of temporal variation in salmonids. Analysis of molecular variance showed that most of the genetic variation was at the individual level (97.9%), although significant genetic variation existed between timing of runs (2.59%). Additionally, at least one locus in each group was out of Hardy–Weinberg equilibrium due to a deficiency in heterozygotes, and significant F_{IS} values were observed in three temporal collections. Together, these results suggest stock admixture, caused by multiple populations of origin in each sampling group, better known as the Wahlund Effect. These observations are preliminary evidence for isolation by time between Klamath River steelhead runs during distinct periods.

Introduction

The scale of reproductive isolation among Klamath River coastal steelhead stocks is unclear. The National Marine Fisheries Service (NMFS, Busby et al., 1996) recognizes two distinct reproductive ecotypes, based upon sexual maturation, that migrate into the basin during the summer or winter. The summer ecotype enters freshwater sexually immature, and requires several months for eggs to ripen to spawning condition (Burgner et al., 1992). Peak migration period in these fish occurs in June, however, the spawning season remains unknown. The winter ecotype matures sexually in the ocean and enters the river between November and March (Hopelain, 2001). These fish find suitable spawning habitat relatively quickly, with a peak in spawning activity during January. The existence of a third group, called the fall run (Hopelain, 2001), of sexually mature steelhead migrates in the Klamath mainstem between July and October, but may be an extension of a summer ecotype. Changes in the temporal and spatial variability of river flows and water quality, as well as a number of other causes, have led to a serious decline in the Klamath River wild steelhead population, such that these wild stocks no longer support an in-river sportfishery. Despite this, the population has not been listed under the Endangered Species Act (NMFS, 2001).

The Klamath River Stock Identification Committee (KRSIC) identifies a summer and winter metapopulation, based upon the fishes' time of river entry. However, a potential overlap in migration and spawning periods makes differentiating these ecotypes difficult (Roelofs, 1983) and confounds their population structure. Microsatellite DNA markers have been particularly useful for studying population structure and life-history variability in salmonids (Garrant et al., 2000; Neraas and Spruell, 2001). This study was designed to describe the genetic variation associated with winter and summer coastal summer ecotypes entering the Klamath River and to determine if these ecotypes are reflective of isolation by time. The null hypothesis evaluated by this study is that Klamath River steelhead, from different run-timing periods, are genetically similar and form a single population.

Materials and methods

Study area

The Klamath River basin is located in southern central Oregon and northwestern California. The 40 468-km^2 basin drains six major rivers before emptying into the Pacific Ocean. The Trinity River, the largest Klamath River tributary, joins the Klamath approximately 90 km from the ocean. Other tributaries upstream of this confluence include the Shasta, Scott, South Fork Trinity, and Salmon rivers, and the drainage system includes about 200 minor tributaries. Two recently constructed dams block the migration of anadromous fish into the upper reaches of their respective rivers, which historically supported steelhead. Lewiston Dam, completed in 1963, blocks migration into the Upper Trinity River basin, while Iron Gate Dam on the Klamath has blocked migration into the Upper Klamath River tributaries since 1962. Terminal hatcheries are maintained at both dams.

Sample collection and preparation

Scale and fin clip samples of 238 coastal steelhead were collected during net harvest monitoring by the Yurok Tribal Fisheries Program and dried in envelopes. These samples were collected between January 2000 and September 2002 between the estuary and confluence of Klamath and Trinity rivers. Samples corresponded to the summer and winter ecotypes based on the classification of Busby et al. (1996). All samples were taken below the confluence of the Klamath–Trinity rivers, thus precluding our ability to detect genetic variation attributable to Klamath–Trinity sub-basin spatial differences.

Locus	Source	MgCl$_2$ concentration (mM)	Initial annealing temperature (°C)	PCR product dilution
OtsG 85	Williamson et al. (2002)	1.5	57	1–60
OtsG 253c	C. Garza (pers. comm.)	1.5	68	1–40
OtsG 83b	Williamson et al. (2002)	1.5	65	1–32
OtsG 249b	Williamson et al. (2002)	1.5	57	1–48
Omy 1101	C. Garza (pers. comm.)	0.75	62	1–32
Omm 1082	Rexroad et al. (2002)	1.5	57	1–54
Omm 1087	Rexroad et al. (2002)	2.0	57	1–32

Sampling above the main confluence and in discrete spawning locations was logistically unfeasible and beyond the scope of our hypothesis. Genomic DNA was extracted from dry fin clip and scale samples, using the Promega Wizard SV 96 Genomic DNA Purification System^{TM}.

Microsatellite DNA experiment

Seven microsatellite DNA primer sets were optimized for use in this study by adjusting magnesium chloride (MgCl$_2$) concentration, initial annealing temperature, and DNA concentration (Table 1). Polymerase chain reaction (PCR) conditions started with a 3-min denaturing step at 94°C, followed by 10 cycles of 94°C for 45 s, a locus-specific starting annealing temperature for 45 s (decreasing 0.5°C/cycle) and 72°C for 1 min; followed by 33 cycles of 94°C for 30 s, a locus-specific annealing temperature (5°C less than the starting annealing temperature) for 30 s, and 72°C for 1 min; and a final extension at 60°C for 3 min. These reactions included 1 μl of 10X PCR buffer, 1.5 or 0.75 μl 50 mM MgCl$_2$ depending on optimal conditions for each loci, 0.10 μg BSA, 0.80 μl 2 mM dNTP mixture, 0.6 μM forward primer labeled with one of three fluorescent dyes (NED, VIC or 6FAM), 0.60 μM unlabeled reverse primer, 0.075 μl FASTSTART *Taq* polymerase (0.375 U total), approximately 50 ng of DNA, and sterile dH$_2$O to reach a 10-μl volume. PCR products were diluted (Table 1) and separated electrophoretically on a 5.5% polyacrylamide gel using the MJ Research BaseStation gel analysis system. Allele sizes were designated using a Rox-labeled Genescan 500 size standard (MJ Research, Inc., San Francisco, CA, USA) run in each lane. A number of the same individuals were included on each gel to ensure consistent scoring of unknown individuals across all gels. Gel images were analyzed with the Cartographer® software package (MJ Research, Inc.).

Population genetic analyses

Heterozygosity was calculated using Tools for Population Genetic Analysis (TFPGA 1.3, Miller, 1997). Tests for Hardy–Weinberg and genotypic pairwise disequilibrium were performed with the software package Genetic Data Analysis (GDA, Lewis and Zaykin, 2002). P-values were estimated by 3200 random permutations setting the significance level (α) at

0.05. F_{IS} (Nei, 1978) was computed in GDA, and 1000 bootstrap resamplings yielded 95% confidence intervals to assess statistical significance over loci. Genepop on the Web (GENEPOP, Raymond and Rousset, 1995) was used to calculate genotypic differentiation among runs with 20,000 permutations. Significance of observed differentiation was tested with an unbiased estimate of the P-value of a log-likelihood (*G*)-based exact test (Goudet et al., 1996). A matrix of pairwise F_{ST} values was estimated between years (Weir and Cockerham, 1984) with GENETIX (Belkhir et al., 2000), and bootstrap values determined following 2000 permutations. An analysis of molecular variance (AMOVA) was used to partition the allelic variance and determine divergence within and among populations with the software packet ARLEQUIN 2.0 (Excoffier et al., 1992).

Hierarchical cluster analysis determined using the unweighted pair group method using arithmetic averages (UPGMA) algorithm (Sneath and Sokal, 1973) implemented in TFPGA 1.3 was calculated using Nei's unbiased genetic distance (1978) with 1000 bootstrapped permutations. Nei's unbiased genetic distance (1978) was computed with GENETIX, because this program allows calculation of bootstrap values of genetic distances based on 2000 permutations. A consensus UPGMA diagram was then generated with the original branch lengths, and all bootstrap values were plotted on to the dendrogram to indicate stability of the nodes.

Additional tests for population differentiation were performed using a Bayesian genetic clustering procedure implemented in STRUCTURE (Pritchard et al., 2000). This program assumes complete linkage and Hardy–Weinberg equilibrium (HWE) within populations, and allocates individuals into *k* groups that minimize departures from HWE and maximise LE (linkage equilibrium) without considering when or where the individuals were collected. Individuals are assigned probabilistically to populations (groups), or jointly to two or more groups if their genotypes indicate that they are admixed.

In this study we used the STRUCTURE program to determine the likely number of populations that captured the major genetic structure of the data for each summer and winter of our collection (*k* = 1–5) by averaging results from 100 000 iterations (burn-in 500 000 replicates).

Results

Observed and expected heterozygosity at the seven microsatellite DNA loci, allelic number, and the sample sizes (n) are included in Table 2. Of 147 pairwise locus combinations, 65 had significant levels (44.2%, α = 0.05) of linkage disequilibrium. When the same statistical analysis was used while preserving genotypes, thus removing within-locus disequilibrium, pairwise linkage disequilibrium was reduced in all collections except for winter 2001, and a remaining 44 of 147 pairwise comparisons (29.9%, α = 0.05) showed significant departure from HWE. Except for the summer 2001 sample, each collection contained at least one locus with allelic proportions that deviated from Hardy–Weinberg expectations. *OtsG* 253c, displayed the greatest departure from expectation (of equilibrium in four of six collections), while the other six loci did not show any particular departure from Hardy–Weinberg expectations. F_{IS} values ranged from 0.003 (summer 2001) to 0.131 (winter 2000). Significant departures from zero of the F_{IS} coefficient were detected in the 2000 and 2001 winter and 2001 summer collections.

Exact tests (20 000 permutations) of allelic differentiation showed significant differences [either with P < 0.001 or

Table 2
Sample size (n), allelic numbers (A), observed and expected heterozygosity (H_o and H_e, respectively) for each of the sample collections

	Year:	2000	2001		2002	
	Run time:	Winter	Winter	Summer	Winter	Summer
Locus	n:	42	45	32	52	40
OtsG 85	A	19	20	19	20	21
	H_o	0.93	0.91	0.94	0.86	0.87
	H_e	0.93	0.93	0.92	0.92	0.93
OtsG 253c	A	20	19	15	20	19
	H_o	**0.62**	**0.67**	0.81	**0.59**	**0.77**
	H_e	0.92	0.94	0.89	0.92	0.91
OtsG 83b	A	19	18	18	18	15
	H_o	0.85	0.91	0.97	0.85	**0.71**
	H_e	0.92	0.92	0.93	0.92	0.91
OtsG 249b	A	16	16	13	15	18
	H_o	0.86	0.95	0.94	0.91	**0.82**
	H_e	0.89	0.89	0.87	0.89	0.9
Omy 1101	A	16	13	14	13	12
	H_o	**0.79**	0.79	0.87	0.86	**0.81**
	H_e	0.88	0.85	0.89	0.88	0.86
Omm 1082	A	14	12	12	15	13
	H_o	**0.83**	0.86	0.94	0.88	0.9
	H_e	0.9	0.86	0.87	0.9	0.9
Omm 1087	A	18	15	15	16	12
	H_o	**0.67**	0.91	0.87	0.8	0.8
	H_e	0.85	0.87	0.89	0.89	0.88
Total mean	A	17.43	16.14	15.14	16.71	15.71
	H_o	0.79	0.86	0.91	0.82	0.81
	H_e	0.9	0.89	0.89	0.9	0.9

Bold values denote significant departures of observed heterozygosity from expected values.

Table 3
Unbiased genetic distance (Nei, 1978) above diagonal and pairwise F_{ST} estimates for differentiation at seven loci below diagonal (*P < 0.05; **P < 0.01; ***P < 0.001)

		2000	2001		2002	
		Winter	Winter	Summer	Winter	Summer
2000	Winter		0.014	**0.020****	0.016	**0.021****
2001	Winter	0.003		0.016	0.016	**0.022*****
	Summer	**0.007****	0.004		0.020*	0.017
2002	Winter	0.003	0.004	0.005*		**0.024*****
	Summer	**0.009*****	**0.011*****	0.004	**0.011*****	

Significant values (after Bonferroni correction) are in bold.

P < 0.01 after sequential Bonferroni correction (Rice, 1989)] for a majority of pairwise comparisons between winter and summer runs. A genotypic comparison between these same winter and summer sample groupings at all loci showed high statistical support (P < 0.001). Significant but low F_{ST} values were detected between each collection of winter and summer runs except in the 2001 samples, and ranged from 0.0016 to 0.0271 (Table 3). A comparison of these values with significance tests of Nei's (1978) unbiased genetic distance shows similar results. Between years, F_{ST} values were not significantly greater than zero for winter samples from 2000 and 2002 or for summer samples. UPGMA cluster analysis using 1000 permutations of unbiased genetic distance showed strong support for differentiation between all pairwise comparisons of summer and winter runs. However, differentiation among summer or winter runs from 1 year to the next was weak (Fig. 1). AMOVA results attributed greater variance to differences among winter
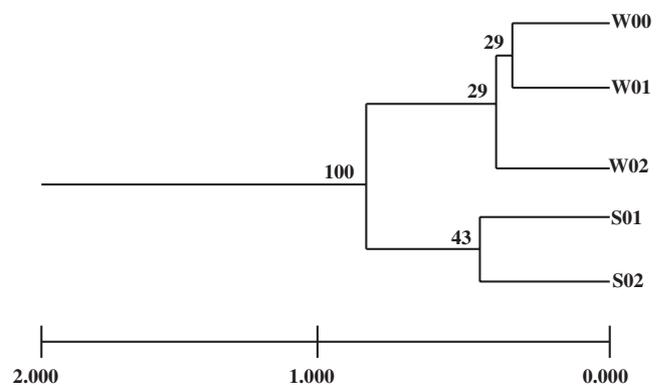

Fig. 1. Unweighted pair group method using arithmetic averages tree with bootstrap values for five collections of steelhead samples during 3 years (2000, 00; 2001, 01; 2002, 02) of summer (S) and winter (W) migration periods

and summer runs in a single year (2.59%) than differences among both winter and summer samples combined between 2001 and 2002. The majority (97.9%) of allelic variance was attributed to individuals.

The results of the STRUCTURE analysis for each run-timing sample revealed than Bayesian posterior probabilities were highest for $k = 2$, in each summer and winter collection, with the exclusion of summer 2002.

Discussion

Despite the high levels of overall genetic diversity, significant deviations from Hardy–Weinberg proportions, due to a deficiency of heterozygous genotypes, were observed at multiple loci in all the collection, with the exception of summer 2001. In addition, a large portion of pairwise comparisons of loci showed linkage disequilibrium. F_{IS} values evidenced a general low inbreeding and a departure from zero of the same collections out of HWE, suggesting that the individuals with sampling groups do not represent single reproductively isolated stocks at equilibrium, but that admixed populations within the individual temporal sample collections and the results are characteristics of the Wahlund effect.

With the only exception of summer 2002, STRUCTURE supported the evidence that two clusters captured most of the genetic structure found in each run-timing, confirming a possible composition of different populations in our collection.

The majority of the pairwise F_{ST} and Nei's unbiased genetic distance values support small, significant genetic differences among summer and winter run-timings and tests of genotypic differentiation show highly significant differences over all loci among these two groups. This result is similar to those exhibited in other studies demonstrating restricted gene flow between salmonids with a bimodal escapement pattern (Nielsen and Fountain, 1999; Ramstad et al., 2003). While our results could be caused by the sampling of multiple geographic stocks within each temporal collection of samples, another possible explanation points to the extraordinary plasticity of *Oncorhynchus mykiss* life-history characteristics. The Klamath–Trinity River basin covers a heterogeneous landscape, shaped by active geologic and hydrologic processes, which has forced aquatic organisms to develop behavioral and physiologic mechanisms for coping with an increased level of disturbance. Thus, these data suggest that Klamath River coastal steelhead have evolved multiple temporal stocks for persisting in a system where critical habitat parameters are highly variable.

Another result suggesting time has been an isolating mechanism upon Klamath River steelhead is that the amount of genetic variation attributed, by AMOVA, to differences between run-timings was considerably larger (2.69%) than the non-significant variance attributable to differences between different years. This result suggests temporal differences are scaled within a year and not within age-classes. These results are in concordance with other studies, which sampled seasonal runs from multiple years, where clustering by reproductive time rather than by year was observed (Fillatre et al., 2003; Ramstad et al., 2003).

These results are in concordance with scientific knowledge that salmonids migrate into their natal rivers for spawning and have behavioral traits, which reinforce spatial isolation, instead of regular gene flow and a homogenous metapopulation structure. While temporal populations were recognized by the KRSIC (unpublished data) more than a decade ago,

this genetic study is the first on the Klamath River to provide data supporting that steelhead population structuring is occurring due to temporal and spatial scales. In fact, given the high probability that our collections contained samples below the confluence of two major river drainages, it seems likely that there is spatial genetic structuring between sub-basins in the Klamath–Trinity basin. The results of this study suggest that the Klamath River contains at least two, and perhaps more, discrete populations that migrate at different times. The role that spatial and temporal isolation plays in steelhead wild populations is difficult to separate. Our results reflect the confounding effect these isolating barriers have on each other (Hendry and Day, 2005), and thus we cannot evaluate the relative importance of each isolating barrier.

The observed Wahlund effect and significant values for population differentiation suggests a persistent overlapping of seasonal migrants originating from distinct population of origin. Our results suggest the need for a complete investigation of population structure in Klamath steelhead, including samples from discrete breeding aggregations of multiple years as well as sampling of fish throughout the migratory period. In this way, it would be possible to quantify differences in both time and space between paired genetic samples (Hendry and Day, 2005). While this type of sampling is logistically difficult, particularly when breeding locations are unknown or difficult to access as in the case of Klamath River, it will allow one to control for either isolating mechanisms to describe the variation associate with either temporal or spatial structuring.

Management plans to implement conservation strategies related to migratory timing must take into account evidence for multiple populations occurring temporally and spatially across landscapes. Further studies are necessary in tributaries and sub-basins to determine the spatial component of the Klamath steelhead population structure and confirm the identification of reproductively isolated breeding populations at microgeographic scale.

References

Belkhir, K.; Borsa, P.; Chikni, L.; Raufaste, N.; Bonhomme, F., 2000: GENETIX 4.0, logiciel sous windows TM Pour la genetique des populations. Laboratoire Genome, Population, Interactions, CNRS UMR 5000, University of Montpellier II, Montpellier.

Burgner, R. L.; Light, J. T.; Margolis, L.; Okazaki, T.; Tautz, A.; Ito, S., 1992: Distribution and origins of steelhead trout (*Oncorhynchus mykiss*) in offshore waters of the north Pacific Ocean. Int. N. Pac. Fish. Comm. Bull. **51**, 1–92.

Busby, P. J.; Wainwright, T. C.; Bryant, G. J.; Lierheimer, L. J.; Waples, R. S.; Waknitz, F. W.; Lagomarsino, I. L., 1996: Status review of west coast steelhead from Washington, Idaho, Oregon, and California. National Marine Fisheries Technical Memorandum NMFS-NWFSC-27, Seattle, WA.

Excoffier, L. P.; Smouse, P.; Quattro, J., 1992: Analysis of molecular variance inferred from metric distances among DNA haplotypes: applications to human mitochondrial DNA restriction data. Genetics **131**, 479–491.

Fillatre, E. K.; Etherton, P.; Heath, D. D., 2003: Bimodal run distribution in northern population of sockeye salmon (*Oncorhynchus nerka*): life history and genetic analysis on temporal scale. Mol. Ecol. **12,** 1793–1805.

Garrant, D.; Dodson, J. J.; Bernatchez, L., 2000: Ecological determinants and temporal stability of the within-river population structure in Atlantic salmon (*Salmo salar* L.). Mol. Ecol. **9,** 615–628.

Goudet, J.; Raymond, M.; De Meeues, T.; Rousset, F., 1996: Testing differentiation in diploid populations. Genetics **144,** 1933–1940.

Hendry, A. P.; Day, T., 2005: Population structure attributable to reproductive time: isolation by time and adaptation by time. Mol. Ecol. **14,** 901–916.

Hopelain, J. S., 2001: Lower Klamath River angler creel census with emphasis on upstream migrating fall Chinook salmon, coho salmon, and steelhead trout during July through October, 1983 through 1987. Cal. Dept. Fish Game, Inl. Fish. Ad. Rep. 01-1.

Lewis, P. O.; Zaykin, D., 2002: Genetic Data Analysis: computer program for the analysis of allelic data, version 1.1. Available at http://hydrodictyon.eeb.uconn.edu/people/plewis/software.php (accessed 22 November 2006)

Miller, M. P., 1997: Tools for population genetics analyses (TFPGA). A Windows program for the analysis of allozyme and molecular population genetic data. Version 1.3, Department of Biological Sciences, Northern Arizona University: Flagstaff, AZ.

Nei, M., 1978: Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics **89,** 583–590.

Neraas, L. P.; Spruell, P., 2001: Fragmentation of riverine systems: the genetic effects of dams on bull trout (*Salvelinus confluentus*) in the Clark Fork River system. Mol. Ecol. **10,** 1153–1164.

Nielsen, J. L.; Fountain, M. C., 1999: Microsatellite diversity in sympatric reproductive ecotypes of Pacific steelhead (*Oncorhynchus mykiss*) from the Middle Fork Eel River, California. Ecol. Freshw. Fish. **8,** 159–168.

NMFS (National Marine Fisheries Service), 2001: Notice of Determination; endangered and threatened species: final listing determination for Klamath Mountains Province steelhead. Fed. Regist. **66,** 17845–17856.

Pritchard, J. K.; Stephens, M.; Donnelly, P., 2000: Inference of population structure using multilocus genotype data. Genetics **155,** 945–959.

Ramstad, K. M.; Foote, C. J.; Olsen, J. B.; Rogers, D., 2003: Genetic and phenotypic evidence of reproductive isolations between seasonal runs of sockeye salmon in Bear Lake, Alaska. Trans. Am. Fish. Soc. **132,** 997–1013.

Raymond, M.; Rousset, F., 1995: GENEPOP (version 1.2): population genetic software for exact tests and ecumenism. J. Hered. **86,** 248–249.

Rexroad, C. E.; Coleman, R. L.; Hershberger, W. K.; Killefer, J., 2002: Rapid communications: thirty-eight polymorphic microsatellite markers for mapping in rainbow trout. J. Anim. Sci. **80,** 541–542.

Rice, W. R., 1989: Analysing tables of statistical tests. Evolution **43,** 223–225.

Roelofs, T. D., 1983: Current status of California summer steelhead (*Salmo gairdnerii*) stocks and habitat, and recommendations for their management. Report to the U.S. Forest Service, Region 5.

Sneath, P. H. A.; Sokal, R. R., 1973: Numerical taxonomy. Freeman, San Francisco, CA.

Weir, B. S.; Cockerham, C. C., 1984: Estimating F-statistics for the analysis of population structure. Evolution **38,** 1358–1370.

Williamson, K. S.; Cordes, J. F.; May, B. P., 2002: Characterization of microsatellite loci in Chinook salmon (*Oncorhynchus tshawytscha*) and cross species amplification in other salmonids. Mol. Ecol. Notes **2,** 17–19.

Author's address: Riccardo Papa, Department of Biology, University of Puerto Rico – Rio Piedras, PO Box 23360, San Juan 00931-3360, Puerto Rico.
E-mail:papa@biol.unipr.it