MOLECULAR ECOLOGY

Molecular Ecology (2012) **21**, 1567–1582 doi: 10.1111/j.1365-294X.2011.05419.x

Resurrecting an extinct salmon evolutionarily significant unit: archived scales, historical DNA and implications for restoration

ERIC M. IWAMOTO, JAMES M. MYERS and RICHARD G. GUSTAFSON

Conservation Biology Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 2725 Montlake Blvd. E, Seattle, WA 98112, USA

Abstract

Archival scales from 603 sockeye salmon (*Oncorhynchus nerka*), sampled from May to July 1924 in the lower Columbia River, were analysed for genetic variability at 12 microsatellite loci and compared to 17 present-day *O. nerka* populations—exhibiting either anadromous (sockeye salmon) or nonanadromous (kokanee) life histories—from throughout the Columbia River Basin, including areas upstream of impassable dams built subsequent to 1924. Statistical analyses identified four major genetic assemblages of sockeye salmon in the 1924 samples. Two of these putative historical groupings were found to be genetically similar to extant evolutionarily significant units (ESUs) in the Okanogan and Wenatchee Rivers (pairwise F_{ST} = 0.004 and 0.002, respectively), and assignment tests were able to allocate 77% of the fish in these two historical groupings to the contemporary Okanogan River and Lake Wenatchee ESUs. A third historical genetic grouping was most closely aligned with contemporary sockeye salmon in Redfish Lake, Idaho, although the association was less robust (pairwise F_{ST} = 0.060). However, a fourth genetic grouping did not appear to be related to any contemporary sockeye salmon or kokanee population, assigned poorly to the *O. nerka* baseline, and had distinctive early return migration timing, suggesting that this group represents a historical ESU originating in headwater lakes in British Columbia that was probably extirpated sometime after 1924. The lack of a contemporary *O. nerka* population possessing the genetic legacy of this extinct ESU indicates that efforts to reestablish early-migrating sockeye salmon to the headwater lakes region of the Columbia River will be difficult.

Keywords: evolutionarily significant unit, kokanee, microsatellite DNA, *Oncorhynchus nerka*, population extinction, re-introduction

Received 5 August 2011; revision received 15 November 2011; accepted 18 November 2011

Introduction

Evolutionarily significant units (ESUs) are operational conservation units composed of population groups that are distinct from other conspecific populations and possess a shared evolutionary legacy (Ryder 1986; Waples 1995). Describing the spatial structure and genetic and life history diversity of historical ESUs in species that have lost ESUs and suffered substantial abundance declines is crucial to: (i) identification of candidate source populations for potential re-introduction (Schwartz *et al.* 2007; Hoeck *et al.* 2010; Paplinska *et al.* 2011), (ii) confirmation of the integrity of contemporary ESUs (Martinez-Cruz *et al.* 2007; Paplinska *et al.* 2011); and (iii) evaluation of past anthropogenic impacts (Bouzat *et al.* 1998). In the field of fish conservation biology, numerous studies (see review by Nielsen & Hansen 2008) have used DNA from historical fish scales and otoliths to compare genetic diversity of contemporary and historical populations to: (i) evaluate the integrity of contemporary populations (Tessier & Bernatchez 1999; Nielsen *et al.* 2001; Heath *et al.* 2002); (ii) evaluate the impact of hatchery-mediated supportive breeding

Correspondence: Richard G. Gustafson, Fax: (206) 860 3335; E-mail: rick.gustafson@noaa.gov

(Hansen 2002; Eldridge *et al.* 2009; Hansen *et al.* 2010) and translocations (Nielsen *et al.* 2001; Larsen *et al.* 2005; Fraser *et al.* 2007a); (iii) estimate effective population size (Hansen *et al.* 2002; Heath *et al.* 2002; Fraser *et al.* 2007b); (iv) investigate historical population structure (Guinand *et al.* 2003; Palstra & Ruzzante 2010); and (v) detect evidence of selection (Hansen *et al.* 2010).

Lacking historical DNA, some studies have used an estimate of ecological complexity in historical habitats occupied by extirpated Pacific salmon in the Pacific Northwest and California to infer the number of lost genetically distinct populations (Lindley *et al.* 2006; Gustafson *et al.* 2007) and ESUs (Gustafson *et al.* 2007). However, without retrospective genetic analyses of archived salmon specimens, these attempts remain conjectural (Gustafson *et al.* 2007). Fortunately, field notebooks containing sockeye salmon (*Oncorhynchus nerka*) scales obtained from commercial fisheries during the 1923–1925 adult return migrations in the lower Columbia River provided us with an opportunity to examine whether some of the population extirpations and ESU extinctions identified in Gustafson *et al.* (2007) could be verified using historical DNA.

Of the 10 sockeye salmon ESUs that putatively existed historically (prior to Euro-American contact) in the Columbia River Basin (Gustafson *et al.* 2007), only three remain: Lake Wenatchee, Okanogan[1] River and Snake River (also known as the Redfish Lake ESU) (Table S1, Supporting information). Evidence of historical presence and inferences based on genetic, ecological and life history characteristics of existing ESUs were used by Gustafson *et al.* (2007) to putatively identify seven extirpated sockeye salmon ESUs in the Columbia River Basin (Table S1, Supporting information). During the 1920s, sockeye salmon were observed in headwater lake habitats in British Columbia, Canada, above present-day Grand Coulee Dam (Chapman 1943) and may also have had access to Suttle Lake in the Deschutes River Basin in Oregon and to other areas such as Warm Lake on the South Fork Salmon River in Idaho (Table S1, Supporting information; Fig. 1). Although sockeye salmon currently occur in only three extant ESUs in the Columbia River Basin, numerous kokanee populations occupy lacustrine habitats in the Basin that historically contained sockeye salmon. Kokanee are a nonanadromous ecotype of *O. nerka* (*sensu* Wood *et al.* 2008) that have become resident in lake environments, either naturally or via anthropogenic transplant or entrainment (Wood 1995; Wood *et al.* 2008), and typically mature at a much smaller body size than sockeye salmon. Some of these kokanee populations may be descended from historical sockeye salmon populations and as such could potentially serve as source populations for future re-introduction efforts of anadromous sockeye salmon. Recently, a reversion to anadromy by kokanee released from entrainment behind dams after nearly 90 years was documented in two lakes in the lower Fraser River Basin of British Columbia (Godbout *et al.* 2011).

Here, we use a suite of 12 microsatellite DNA loci to compare genetic variation among historical scale samples, existing sockeye salmon populations and selected kokanee populations in the Columbia River Basin and attempt to: (i) identify and genetically characterize population clusters present in the mixed historical sample; (ii) assign historical sockeye salmon genetic clusters to a baseline of contemporary kokanee and sockeye populations; and (iii) determine whether any sockeye salmon present in the historical samples may correspond to extirpated populations or ESUs. We also evaluate historical return migration timing and body size information and compare these to contemporary life history characteristics of Columbia River sockeye salmon.

Methods

Contemporary sample collections and study design

This study was conducted primarily in rivers and lakes in the Columbia River Basin in the states of Washington, Oregon and Idaho, USA, and in the Province of British Columbia, Canada (Fig. 1). Contemporary samples were obtained from three wild sockeye salmon ESUs and 12 wild and two hatchery kokanee populations (Table 1; Fig. 1). Tissue samples, as caudal fin clips, were obtained nonlethally from wild fish on the spawning grounds and from kokanee returning to hatcheries, and preserved in 1.5-mL Eppendorf tubes containing approximately 1.2 mL of 95% ethanol. All sampling occurred in conjunction with agency personnel, or under permit, from the State of Washington or the Province of British Columbia (Table 1). Three contemporary samples (Warm Lake, Fishhook Creek and Redfish Lake) were obtained from −80°C frozen tissue collections held at the Northwest Fisheries Science Center, Seattle, Washington, USA (sampling sites and methods described in Waples *et al.* 2011) (Table 1; Fig. 1). A description of the collections and their rationale follows:

Redfish Lake. Redfish Lake is the only historically occupied sockeye salmon habitat in the Snake River that has supported a contemporary anadromous run in recent years. Previous genetic analyses (Winans *et al.* 1996; Cummings *et al.* 1997; Waples *et al.* 2011) indicate that there are two distinct *O. nerka* gene pools in Redfish

[1] The accepted spelling in Canada is Okanagan, in the USA it is Okanogan.

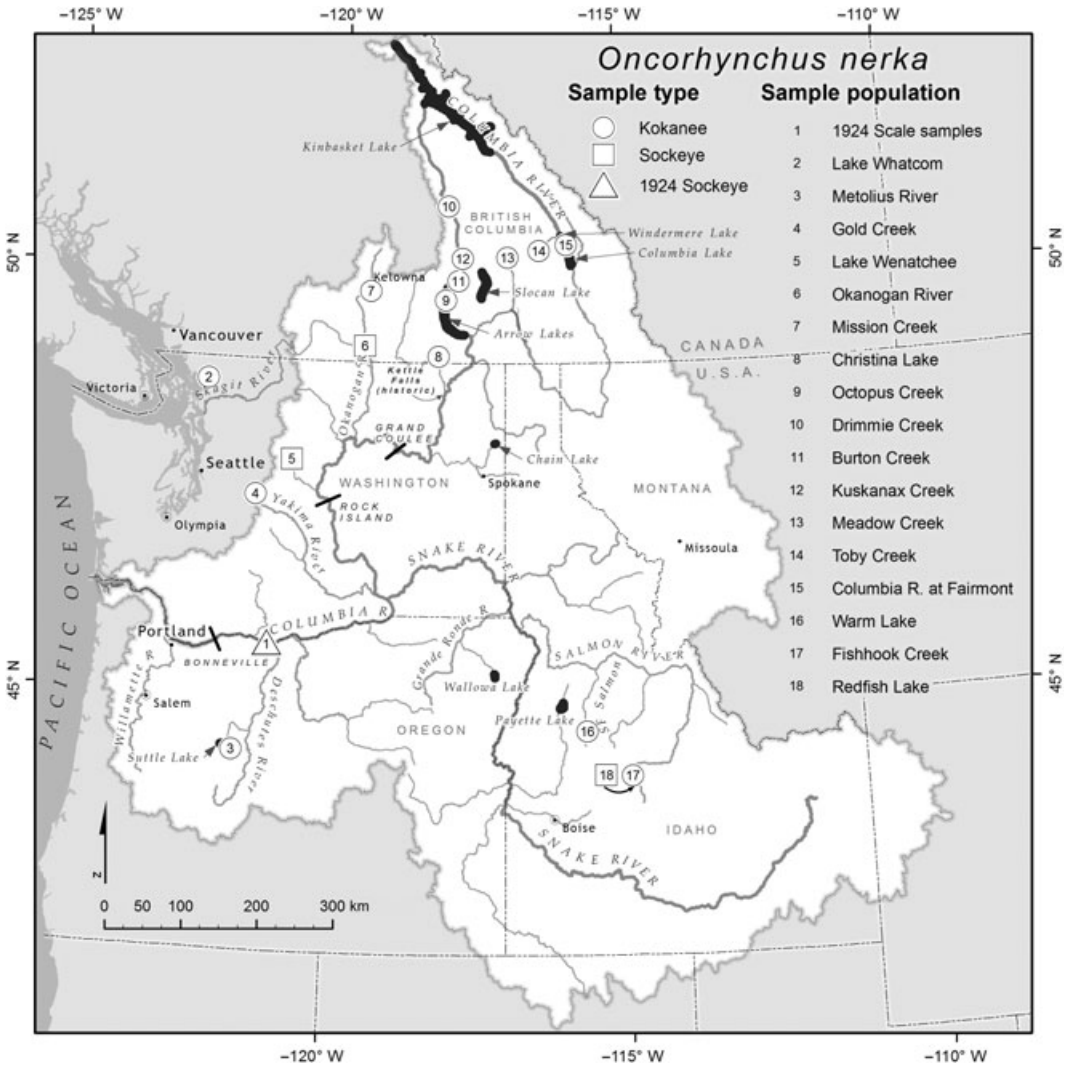

Fig. 1 Locations of *Oncorhynchus nerka* genetic sample collections, dams (indicated by black bars) and geographic features in the U.S. Pacific Northwest and British Columbia, Canada, mentioned in the text. Sample numbers and collection locations correspond with those listed in Table 1.

Lake that represent essentially indigenous populations: sockeye salmon that spawn on the shoreline of Redfish Lake and kokanee that spawn in Fishhook Creek, an inlet to Redfish Lake. Although historical records are equivocal, sockeye salmon may have once spawned in Fishhook Creek (Chapman *et al.* 1995). Therefore, we included both *O. nerka* populations from Redfish Lake in our analyses. Because outmigrating juvenile *O. nerka* from Redfish Lake are genetically indistinguishable from sockeye salmon that returned to Redfish Lake in the early 1990s (see Waples *et al.* 2011), we used a subset of these outmigrants as representative of the contemporary Redfish Lake sockeye salmon gene pool.

Warm Lake. Samples from Warm Lake were included because this lake historically supported a sockeye salmon population and presently contains a presumably indigenous beach-spawning kokanee population that is genetically unlike any other *O. nerka* population, including any populations that may have been planted there (Waples *et al.* 2011). It is uncertain as to whether anadromous access to Warm Lake was available in the 1920s.

Lake Wenatchee and Okanogan River. Currently, sockeye salmon in the Columbia River Basin originate almost exclusively from natural production in these two ESUs. In the 1920s, the Lake Wenatchee and Okanogan River ESUs contained nearly 20% of the accessible lake-rearing habitat for sockeye salmon in the Columbia River Basin (Table S1, Supporting information).

Table 1 Collection data for *Oncorhynchus nerka* samples including sample size (*N*), ecotype (sockeye [S] or kokanee [K]) and *P*-values indicative of deviation from Hardy–Weinberg equilibrium (HWE). Following implementation of the modified false discovery rate correction of Benjamini & Yekutieli (2001), no populations are out of HWE (alpha = 0.05/3.44 = 0.01)

Sample number	Sample name (rearing lake)	Watershed	Collection date	*N*	Ecotype	HWE *P*	Allelic richness
1	1924 scale samples (mixed)	Columbia River, WA and OR	May–July 1924	603	S	N/A	N/A
2	Lake Whatcom Hatchery (Lake Whatcom)*	Puget Sound, WA	8 November 2005	47	K	0.034	10.42
3	Metolius River (Lake Billy Chinook)†	Deschutes River, OR	1 October 2004	46	K	0.358	10.92
4	Gold Creek (Keechelus Lake)*	Yakima River, WA	22 October 2004	44	K	0.723	9.58
5	Lake Wenatchee*	Wenatchee River, WA	21 September 1998	96	S	0.115	8.68
6	Okanogan River (Osoyoos Lake)‡	Okanogan River, WA and BC	14 October 2005	96	S	0.439	8.86
7	Mission Creek (Okanagan Lake)§	Okanogan River, WA and BC	26 October 2005	44	K	0.167	9.50
8	Christina Lake§	Kettle River, WA and BC	1 December 2004	86	K	0.943	7.26
9	Octopus Creek (Lower Arrow Lake)¶	Columbia River, BC	8 September 2005	48	K	0.468	9.17
10	Drimmie Creek (Upper Arrow Lake)¶	Columbia River, BC	8 September 2005	95	K	0.561	8.94
11	Burton Creek (Upper/Lower Arrow Lakes)¶	Columbia River, BC	8 September 2005	94	K	0.485	8.64
12	Kuskanax Creek (Upper Arrow Lake)¶	Columbia River, BC	7 September 2005	71	K	0.772	8.93
13	Meadow Creek (Kootenay Lake)¶	Kootenay River, BC	8 September 2005	48	K	0.517	10.33
14	Toby Creek (Kinbasket Lake)¶	Columbia River, BC	21 October 2005	48	K	0.529	10.58
15	Columbia River at Fairmont (Lake Windermere)¶	Columbia River, BC	21 October 2005	95	K	0.672	10.69
16	Warm Lake	South Fork Salmon River, ID	26 October 1990	48	K	0.554	7.17
17	Fishhook Creek (Redfish Lake)	Salmon River, ID	8 September 1990	48	K	0.391	7.00
18	Redfish Lake	Salmon River, ID	April–May 1991	48	S	0.195	3.92

Allelic richness is the number of alleles for each sample after correction for *N*.
*Samples collected in cooperation with local Washington Department of Fish and Wildlife personnel.
†Samples provided by Oregon Department of Fish and Wildlife.
‡Samples provided by Okanagan Nation Alliance, Fisheries Department, Westbank, BC, Canada.
§Samples provided by BC Ministry of Environment, Penticton, BC, Canada.
¶British Columbia sampling conducted under provincial permit CB/PE05-14310.

Christina Lake. This sample of indigenous kokanee spawns along the shoreline of Christina Lake in late fall and winter, which is atypical for the upper Columbia River region, where other kokanee populations spawn in either August or October (LaCroix & McLean 2005). There were no obvious barriers to anadromy for this stock during the 1920s.

Metolius River. Kokanee that are presumably descended from the extirpated Suttle Lake sockeye salmon ESU are resident in a reservoir on the Deschutes River, Oregon, and spawn in the tributary Metolius River. Records indicate that anadromous access to Suttle Lake was available during the 1920s (Table S1, Supporting information).

Mission Creek. We included this inlet spawning population in our baseline as representative of kokanee from Okanagan Lake, which is the largest lake on the basis of surface area in the Columbia River Basin that may have historically produced sockeye salmon (Table S1, Supporting information).

Gold Creek. This population, which spawns in an inlet to Lake Keechelus on the Yakima River, was included in our baseline because its genetic population structure has not been previously investigated and it is the only available kokanee sample from the Yakima River, which historically produced up to 100 000 sockeye salmon per year.

Arrow Lakes and Upper Columbia River. It has been estimated that 60–85% (Mullan 1986; Chapman *et al.* 1995) of Columbia River sockeye salmon that once migrated above the confluence with the Snake River originated from populations rearing in lake habitats upstream of

Grand Coulee Dam (see Fig. 1), which was completed in 1939 and imposed a total barrier to upstream fish passage. To focus our collection efforts on potential kokanee populations that may have retained a legacy of the extirpated sockeye salmon populations above Grand Coulee Dam, we sampled kokanee solely from basins that, according to available records (see Facchin & King 1983a, b; British Columbia Ministry of Environment 2011), had no known history of *O. nerka* transplantation or artificial propagation. For example, the sample of wild spawning kokanee from Kuskanax Creek (Fig. 1), a tributary of Upper Arrow Lake, is from the same location where sockeye salmon were observed spawning in September 1938 (Chapman 1943). Wild kokanee from regions of the Upper Columbia River Basin are distinguished on the basis of spawn timing: Arrow Lake kokanee spawn in August, whereas Toby Creek and Columbia River at Fairmont kokanee spawn in October.

Hatchery populations. Samples of kokanee from the Lake Whatcom Hatchery and the Meadow Creek Spawning Channel (Table 1; Fig. 1) were included because these stocks have been widely planted throughout the Pacific Northwest (Gustafson *et al.* 1997). If the genetic structure of any of our putatively indigenous *O. nerka* populations was found to resemble either of these two hatchery samples, it would probably indicate that past stocking had impacted their genetic make-up.

Historical samples. Sockeye salmon scales were obtained from field notebooks that had been in storage for several decades at the National Archives Regional Facility in Seattle, Washington, USA. United States Bureau of Fisheries (USBF) staff obtained these samples from fish wheels operated by commercial fishers near the Cascades of the Columbia—a six-mile-long rapids that is now submerged, but which historically existed at the approximate present-day location of Bonneville Dam (at river km 235) (Fig. 1). This location was downstream of any known historical sockeye salmon spawning location. Samples were taken in 1923–1925; however, we chose to limit our analysis to scales collected in 1924, as they were the most numerous of the three available return years and their sampling dates spanned the entire migration period (May–July) for this species. We have no record of the actual sampling technique; however, clumps of scales were placed into numbered sections on each page of USBF scale sample books together with records on each fish's date of capture, sex, length and weight. The bottom of each page was folded over the scale samples and apparently left to air dry. Only scales that remained firmly attached to the notebooks were utilized (i.e. only those we could clearly associate with sample information). Sub-samples of 3–6 scales were taken from each of 665 individual sockeye salmon for DNA extraction.

Sample preparation and microsatellite analysis

DNA samples from all tissues were extracted using Qiagen DNeasy 96 kits (Qiagen Corp[2]) following the manufacturer's animal tissue protocol. Scale extractions required the removal of the undigested scale from the lysate prior to passage through the DNeasy filter.

In accordance with Nielsen *et al.* (1999) and our preliminary findings, we identified microsatellite loci with allele lengths under 300 base pairs (bp) from the literature and screened them for allelic variation. Twelve microsatellite loci (*Oke*2, *Omm*1085, *One*18, *Ots*10, *Ots*100, *One*13, *Omm*1159, *Omy*77, *Ots*103, *One*21, *Omm*1068 and *Oki*29) were ultimately selected based on allelic variation, ease of scoring and compatibility for multiplexing (Table S2, Supporting information).

In accordance with recent technical advances in multiplex 'preamplification' PCR (Morin & McCarthy 2007; Smith *et al.* 2011) that serve to increase available template DNA, we incorporated a preamplification step into our PCR protocol. Multiplex PCR preamplification reduces problems associated with low concentrations of template DNA of historical and ancient DNA samples, such as high genotyping error rates, allelic dropout and high rates of PCR failure (Smith *et al.* 2011). Our multiplex PCR preamplification procedure had two steps: an initial large volume (50 μL) preamplification step using unlabelled primers meant to amplify all microsatellite loci, followed by a secondary conventional amplification with labelled primers to genotype the PCR products of the initial preamplification. For each multiplex PCR preamplification, aliquots of all forward and reverse primer pairs were combined into a single 40-μL reaction mix along with 10 μL of template DNA. The reaction mix consisted of 0.01 μM forward primer, 0.01 μM reverse primer, 0.2 mM dNTP, 2.0 mM MgCl, 1.0 units Taq (Promega), 1× buffer and 1× BSA. After an initial denaturation at 95 °C for 2 min, PCR preamplification was performed under the following conditions: (i) 94 °C for 40 s; (ii) 50 °C for 60 s; and (iii) 72 °C for 40 s, cycled 25 times, with a final extension of 72 °C for 5 min.

DNA extracted from contemporary samples and product of preamplified scale DNA was then used as the template in standard PCRs, which consisted of a 10-μL reaction mix containing 3 μL of DNA, 0.4 μM forward primer, 0.4 μM reverse primer, 0.2 mM dNTP,

[2] The use of any branded item in this study does not constitute an endorsement by the National Marine Fisheries Service.

2.0 mM MgCl, 0.5 units Taq (Promega), 1× buffer and 1× BSA (see Table S2, Supporting information, for specific PCR conditions). Electrophoresis was conducted on an ABI Prism 3100 automated sequencer (Applied Biosystems), and sizing was performed with Genescan ver. 3.7 (Applied Biosystems). Size data were output from the ABI Prism 3100 and binned using Genotyper 3.7 (Applied Biosystems). Prior to statistical analysis, input files were reformatted for GenePop using the Microsatellite Toolkit (Park 2001) and for Biosys using GenePop on the Internet (available online at http://genepop.curtin.edu.au/; Raymond & Rousset 1995; Rousset 2008). Microsatellite loci were successfully amplified in 603 of the initial 665 scale samples.

Extraction and PCR preparation of historical samples were conducted in a facility specifically designated for use with archived or forensic samples and physically separated from the facility where contemporary samples were processed. Scales samples were run in conjunction with negative (no template controls) and positive (known microsatellite DNA profile) controls. Approximately 8% of all samples were re-amplified to check for reproducibility with <1% genotyping error rate.

Statistical analyses and assignment tests

For contemporary populations, genotypic frequencies for each locus were examined for departures from Hardy–Weinberg expectations using chi-square goodness-of-fit tests, and F_{ST} values were obtained using FSTAT (Goudet 1995). We used MICRO-CHECKER v. 2.2.3 (Van Oosterhout *et al.* 2004) to test each locus within each sample for genotyping errors (stuttering, allelic dropout and null alleles). For all samples, allelic richness and private allelic richness were calculated with HP-Rare (Kalinowski 2005) using the suggested default of 100 genes in each sample to account for rarefaction. Cavalli-Sforza and Edwards (CSE) chord distances (Cavalli-Sforza & Edwards 1967) were generated and bootstrapped using the statistical package POPULATIONS v. 1.2.31 (available online at http://www.bioinformatics.org/~tryphon/populations/#ancre_bibliographie). Consensus neighbour-joining trees were visualized in Treeview (Page 1996), and bootstrap values were generated in POPULATIONS. Reynold's coancestry distances (Reynolds *et al.* 1983) and Nei's D (which accounts for different sample sizes; Nei 1978) were also calculated in POPULATIONS but are not herein reported as they yielded similar results to CSE distances and produced trees with similar topologies. The level of significance (0.05) was corrected for all analytical results involving multiple simultaneous tests by applying the modified false discovery rate (FDR) procedure of Benjamini & Yekutieli (2001), as suggested by Narum (2006).

Population structure of the archived scale sample and the contemporary populations were evaluated with STRUCTURE 2.2 (Pritchard *et al.* 2000; Falush *et al.* 2003). This Bayesian clustering technique identifies the number of genetically similar groups or clusters (k) in a mixed sample, based on individual genotypes at all known loci, and identifies the k affiliation of each individual. Historical genotypes were initially analysed in STRUCTURE with the admixture model, 50 000 burn-in iterations, 500 000 Markov chain Monte Carlo iterations (post–burn-in), and correlated allele frequencies to define the number of clusters (k = 1–12), with 20 independent runs for each k. As suggested by Evanno *et al.* (2005), we performed a second round of STRUCTURE analysis to detect additional within-group differentiation. Contemporary populations were similarly evaluated with STRUCTURE using k = 1–17. Finally, we evaluated the combined data set of historical and contemporary samples with STRUCTURE using k = 7, based on results of the two previous STRUCTURE analyses.

In the absence of samples from historical spawning populations, contemporary *O. nerka* samples were used to form a baseline of populations. GeneClass2 (Piry *et al.* 2004)-based assignment tests of individuals in the historical mixed scale sample (unknowns) were conducted utilizing contemporary populations as a baseline (knowns). Contemporary populations were assumed to be indigenous to the location from which they were sampled. We used two different assignment testing methods from GeneClass2 to assign individuals to baseline populations. The first method produces a likelihood score that an individual belongs to a particular population in the baseline (Rannala & Mountain 1997). The second method produces an assignment probability that the individual observed is from a specific population (Paetkau *et al.* 2004). The Rannala & Mountain (1997) likelihood score presents a standardized assignment, whereas the Paetkau *et al.* (2004) probability does not assume that individuals come from a population in the baseline.

Fish were considered assigned to a baseline population when the assignment score exceeded 0.95 (Hauser *et al.* 2006) and those that failed to assign to any single population with at least this level of probability were considered as not assigned. One weakness of these assignment tests is that the assignment probabilities are relative to the other potential sources in the baseline. To obtain an absolute measure of compatibility of individual genotypes with particular populations, we also used the 'enable probability computation' option in GeneClass2 using the Paetkau *et al.* (2004) algorithm, incorporating a 10% missing allele level, and 10 000 simulations. We used a 10% missing allele level because we assumed that contemporary samples have

experienced loss of alleles owing to genetic drift and population bottlenecks, relative to historical samples.

To validate our results, we also conducted assignment tests of individuals in our contemporary population samples (knowns) to the four historical scale groups (as inferred by STRUCTURE) using GeneClass2 under the Paetkau *et al.* (2004) algorithm, incorporating a 1% missing allele level, and 10 000 simulations. We used a 1% missing allele level because historical groups probably maintained the full complement of ancestral alleles.

Statistical analyses (ANOVA, Tukey's test, discriminant analysis, etc.) of life history characters were computed with JMP, version 5.1.2 (SAS Institute). Migration timing of sockeye salmon in the lower Columbia River was derived from mean daily commercial fishery landings in fish wheels for 1883–1893 (Smith 1895) and from fish passage records at Bonneville Dam for 1938–1947 (US Army Corps of Engineers, http://www.nwp.usace.army.mil/environment/fishdata.asp).

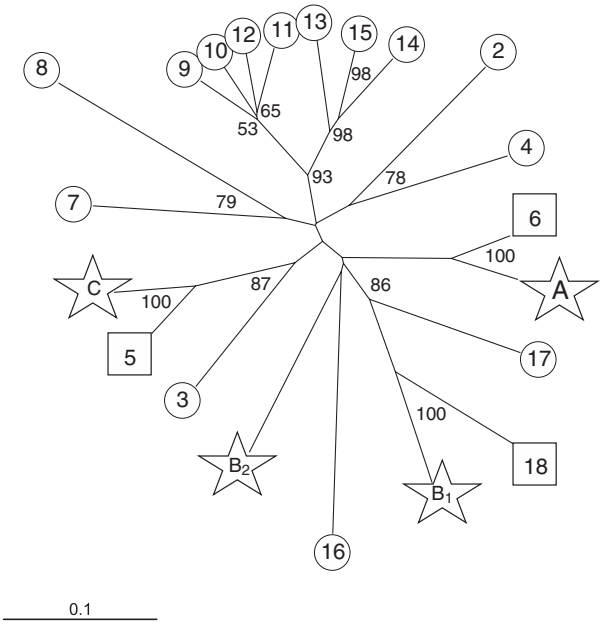

Fig. 2 Consensus neighbour-joining tree of Cavalli-Sforza and Edwards chord distances of historical scale sample groups and contemporary *Oncorhynchus nerka* populations in the Columbia River Basin. Population numbers correspond to those in Table 1. Population 1 (1924 sockeye salmon scales) has been divided into genetic groupings: A, B_1, B_2 and C based on results from STRUCTURE analysis. Stars (☆) indicate samples from the 1924 Columbia River fishery, squares (□) from contemporary sockeye salmon populations and circles (○) from contemporary kokanee populations. Numbers at the nodes (only values >50% are shown) indicate the percentage of 10 000 bootstrap trees where collections beyond the nodes grouped together.

Results

Microsatellite validation and variation

Genotypes were determined in more than 91% of historical and 97% of contemporary samples (based on number of samples that amplified for each population at each locus). The number of alleles per locus ranged from 12 to 29 (Table S2, Supporting information). None of the 17 contemporary population samples deviated from Hardy–Weinberg equilibrium (HWE) over all loci, after correction for multiple tests (Table 1). Before modified FDR correction, five deviations from HWE of individual loci within samples were observed, fewer than would be expected by chance alone at the 5% significance level: *Oke*2 and *Omm*1085 in the Lake Whatcom Hatchery sample, *Omm*1068 in Lake Wenatchee, *Omy*77 in the Meadow Creek Spawning Channel sample and *Ots*103 in Octopus Creek. After applying the modified FDR correction for multiple tests, only *Omm*1068 in Lake Wenatchee and *Oke*2 in Lake Whatcom Hatchery continued to have a significant excess of homozygotes. It is unlikely that null alleles would give rise to these results as no single locus was affected across populations. Failure to amplify large alleles was not the source of homozygote excess because the excess was distributed across allele size classes.

Population structure of contemporary populations

A consensus neighbour-joining tree of CSE chord distances (Fig. 2) indicated that contemporary Columbia River kokanee samples in British Columbia formed two distinct clusters consisting of: (i) an Upper Columbia River cluster composed of Toby Creek (sample 14), Columbia River at Fairmont (sample 15) and Meadow Creek Spawning Channel (sample 13); and (ii) an Arrow Lakes cluster composed of Octopus, Drimmie, Burton and Kuskanax creeks (samples 9–12) (Fig. 2). Within each of these two clusters, pairwise multilocus comparisons of F_{ST} were not statistically different from zero (Table S3, Supporting information). All other pairwise F_{ST} comparisons of contemporary samples were significantly different from zero (Table S3, Supporting information).

STRUCTURE indicated that our 17 contemporary samples consisted of $k = 6$ clusters (Figs S1 and S2, Supporting information). These results supported the pooling of samples 13–15 and of samples 9–12, similar to results from analysis of CSE chord distances and pairwise F_{ST} comparisons. Therefore, in subsequent analyses, we pooled individuals from samples 13–15 into a single Upper Columbia River kokanee sample and samples 9–12 were likewise pooled into a single Arrow Lake kokanee sample.

It was apparent from the consensus neighbour-joining tree of CSE chord distances (Fig. 2) that both Christina Lake (sample 8) and Warm Lake (sample 16) kokanee samples were genetic outliers, which was also indicated by the high pairwise F_{ST} values for all contemporary populations compared to Christina Lake ($F_{ST} \geq 0.148$) and Warm Lake ($F_{ST} \geq 0.171$) (Table S3, Supporting information). All three contemporary sockeye salmon populations were genetically distinct as indicated by their pairwise F_{ST} comparisons (Okanogan vs. Wenatchee, 0.093; Okanogan vs. Redfish, 0.116; Wenatchee vs. Redfish, 0.156) (Table S3, Supporting information; Fig. 2).

Population structure of historical samples

The first round of STRUCTURE analysis identified three genetically distinct clusters or groups ($k = 3$) at the highest hierarchical level among the 1924 sockeye salmon scale samples (Fig. 3a; Fig. S3, Supporting information). These groups are herein designated as Group A (351 fish), Group B (71 fish) and Group C (181 fish).

Further analyses of each of the three cluster groups using STRUCTURE indicated that Group B could be further subdivided into two smaller components (Fig. 3b; Figs S4–S6, Supporting information), designated Group B_1 and Group B_2, consisting of 26 and 45 fish, respectively. However, these analyses did not support further subdivision of either Group A (Fig. S4, Supporting information) or Group C (Fig. S6, Supporting information).

The four historical scale groups and the contemporary Okanogan River and Lake Wenatchee sockeye salmon populations exhibited equivalent levels of allelic richness (Table S4, Supporting information). Additionally, private allelic richness for Groups A, B_2 and C were substantially higher than that observed in Group B_1. Private allelic richness was also higher in Groups A, B_2 and C than in contemporary sockeye salmon populations (Table S4, Supporting information).

A consensus neighbour-joining tree of CSE chord distances (Fig. 2) illustrates clear affinities between Group A and contemporary Okanogan River sockeye salmon (sample 6), Group B_1 and Redfish Lake sockeye salmon (sample 18), and Group C and Lake Wenatchee sockeye salmon (sample 5). In all three cases, these population pairs formed branches with 100% bootstrap support. Pairwise F_{ST} values were quite low between Group A and Okanogan River sockeye (pairwise $F_{ST} = 0.004$) and between Group C and Lake Wenatchee sockeye (pairwise $F_{ST} = 0.002$); however, these values were significantly different from zero (Table S3, Supporting information). Group B_1 and Redfish Lake sockeye salmon (sample 18) clustered together on the CSE neighbour-joining tree (Fig. 2); however, the high pairwise F_{ST} value (0.064) indicated that these two samples were genetically divergent. The consensus neighbour-joining tree (Fig. 2) and high pairwise F_{ST} values (pairwise $F_{ST} \geq 0.067$; Table S3, Supporting information) indicate that Group B_2 was not closely related genetically to any contemporary population of sockeye salmon or kokanee in the Columbia River Basin. Our analysis of the combined historical and contemporary samples using STRUCTURE further illustrates the close relationship between both Group A and contemporary Okanogan River sockeye salmon and Group C and Lake Wenatchee sockeye salmon (Fig. S7, Supporting information).

Using the Rannala & Mountain (1997) standardized assignment scores, GeneClass2 assigned a majority of

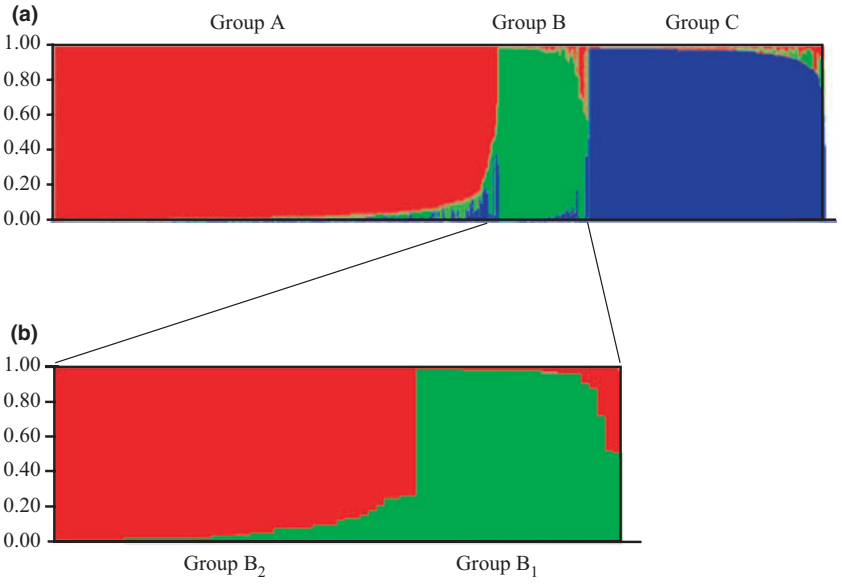

Fig. 3 STRUCTURE results for the individual genotype admixture analysis of the 1924 Columbia River sockeye salmon scale samples. (a) Results of initial STRUCTURE analysis. Colours correspond to the three initial groups (A, B and C), $N = 603$. (b) Results of secondary STRUCTURE analysis of Group B. Colours correspond to the two Groups B_1 and B_2, $N = 71$. Each vertical bar represents a single scale sample. Individuals with more than one colour have admixed genotypes from multiple groups.

sockeye salmon from historical Groups A, B_1 and C—representing over 79% of the historical population admixture—to the three extant sockeye salmon ESUs in our contemporary baseline: Group A to the Okanogan River, Group B_1 to Redfish Lake and Group C to Lake Wenatchee (Table 2). Of the 351 fish in Group A, 89% were assigned to the contemporary Okanogan River sockeye salmon population (10% were unassigned using the 0.95 threshold score) (Table 2). Similarly, 83% of the 181 fish in Group C assigned to the Lake Wenatchee sockeye salmon population (16% were unassigned). About 42% of sockeye salmon in Group B_1 assigned to the contemporary Redfish Lake sockeye salmon population (38% of Group B_1 were unassigned) (Table 2). In contrast, 58% of the 45 Group B_2 sockeye salmon could not be assigned to any population in our baseline, and assignment of the remaining fish was spread among four contemporary populations, including 11 fish (22%) that were assigned to the Okanogan River sample.

As expected, using the Paetkau *et al*. (2004) method in GeneClass2, the individuals from the contemporary populations self-assigned to their respective populations of origin at a mean probability (*P*) near 0.500 (range 0.430–0.504; Table 3). More significantly, some of the historical samples assigned to the contemporary populations at levels comparable to self-assignment of contemporary populations: Group A with Okanogan River ($P = 0.472$) and Group C with Lake Wenatchee

Table 2 Standardized assignment scores for the four 1924 Columbia River sockeye salmon genetic groups to the 17 contemporary Columbia River *Oncorhynchus nerka* baseline populations using Rannala & Mountain (1997) in GeneClass2 (number of fish assigned with a score >0.95 and proportional assignment in parenthesis). There were no assignments with a score >0.95 of any historical individuals to population numbers 7 through 16

Group/ population	Total N	*N* not assigned	(2) Lake Whatcom (K)	(3) Metolius River (K)	(4) Gold Creek (K)	(5) Lake Wenatchee (S)	(6) Okanogan River (S)	(17) Fishhook Creek (K)	(18) Redfish Lake (S)
Group A	351	37 (0.105)	0 (0.000)	1 (0.003)	0 (0.000)	1 (0.003)	312 (0.889)	0 (0.000)	0 (0.000)
Group B_1	26	10 (0.385)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	3 (0.115)	2 (0.077)	11 (0.423)
Group B_2	45	26 (0.578)	1 (0.022)	0 (0.000)	5 (0.111)	2 (0.044)	11 (0.244)	0 (0.000)	0 (0.000)
Group C	181	29 (0.160)	0 (0.000)	1 (0.006)	0 (0.000)	151 (0.834)	0 (0.000)	0 (0.000)	0 (0.000)

Sample sizes for baseline populations are given in Table 1. K, kokanee; S, sockeye salmon.

Table 3 Mean probability assignments of four historical scale groups and contemporary samples to contemporary Columbia River *Oncorhynchus nerka* baseline populations using the GeneClass2 assignment probability method (Paetkau *et al*. 2004). Arrow Lakes and Upper Columbia River represent pooled kokanee samples 9–12 and 13–15, respectively

Sample number	Sample name (ecotype)	2	3	4	5	6	7	8	9–12	13–15	16	17	18
	Group A (S)	0.014	0.104	0.058	0.065	0.472	0.033	0.001	0.025	0.051	0.003	0.045	0.011
	Group B_1 (S)	0.002	0.051	0.036	0.115	0.197	0.021	0.001	0.006	0.080	0.015	0.180	0.164
	Group B_2 (S)	0.007	0.072	0.132	0.080	0.120	0.045	0.011	0.016	0.025	0.016	0.020	0.016
	Group C (S)	0.023	0.413	0.166	0.524	0.035	0.010	0.002	0.011	0.043	0.002	0.019	0.005
2	Lake Whatcom Hatchery (K)	**0.501**	0.045	0.100	0.001	0.001	0.004	0.000	0.005	0.018	0.000	0.000	0.000
3	Metolius River (K)	0.009	**0.493**	0.056	0.025	0.007	0.002	0.000	0.002	0.014	0.000	0.000	0.000
4	Gold Creek (K)	0.038	0.089	**0.479**	0.005	0.002	0.003	0.000	0.002	0.017	0.000	0.000	0.000
5	Lake Wenatchee (S)	0.005	0.269	0.032	**0.484**	0.047	0.002	0.000	0.004	0.018	0.000	0.002	0.000
6	Okanogan River (S)	0.002	0.109	0.013	0.049	**0.495**	0.007	0.000	0.009	0.031	0.000	0.005	0.000
7	Mission Creek (K)	0.011	0.065	0.021	0.001	0.004	**0.507**	0.000	0.004	0.028	0.000	0.000	0.000
8	Christina Lake (K)	0.000	0.010	0.002	0.000	0.001	0.008	**0.509**	0.000	0.011	0.000	0.000	0.000
9–12	Arrow Lakes (K)	0.099	0.192	0.087	0.010	0.027	0.028	0.000	**0.479**	0.417	0.000	0.002	0.000
13–15	Upper Columbia River (K)	0.036	0.102	0.041	0.005	0.007	0.026	0.000	0.070	**0.480**	0.000	0.003	0.000
16	Warm Lake (K)	0.003	0.050	0.138	0.012	0.030	0.000	0.000	0.001	0.004	**0.480**	0.029	0.000
17	Fishhook Creek (K)	0.002	0.134	0.048	0.024	0.088	0.010	0.000	0.004	0.119	0.000	**0.529**	0.005
18	Redfish Lake (S)	0.000	0.174	0.020	0.219	0.454	0.009	0.001	0.007	0.111	0.001	0.336	**0.506**

Sample sizes for baseline populations are given in Table 1. Values for self-assignment are in bold. K, kokanee; S, sockeye salmon.

(P = 0.524). In contrast, individuals from Group B_1 and B_2 did not assign to any one contemporary population at the levels observed for self-assignment, and no single population had a value of $P > 0.200$, with multiple populations exhibiting values of $P > 0.100$ (Table 3).

It is not surprising that assignment of contemporary population samples (knowns) to the baseline composed of the four scale groups (unknowns) identified by STRUCTURE can result in overall lower assignment probabilities (Table S5, Supporting information) than assignment of our scale groups to a baseline of contemporary populations (Table 3), considering that: (i) genetic ancestors of most of the contemporary populations were not present in the scale groups; (ii) scale groups do not represent 'pure' populations; and (iii) it is known that the three extant sockeye populations have lost varying degrees of allelic diversity (Waples *et al.* 2011).

Life history attributes of sockeye salmon in the Columbia River in 1924

Each of the four population groups identified by STRUCTURE was associated with somewhat distinct run timing or body size characteristics. Holmes (1924) identified two distinct size classes of Columbia River sockeye salmon, those less than and those more than 46 cm in length. Our Group A included both large- and small-bodied sockeye salmon; only 42% were longer than 46 cm (Fig. 4). Both Groups B_1 and C were composed of predominately larger bodied fish (96% and 98% larger than 46 cm, respectively). Similarly, Group B_2 sockeye salmon were predominately large bodied fish (only one B_2 fish was <46 cm in length) (Fig. 4). Mean length of sockeye salmon in Group B_2 was significantly greater than the mean length of sockeye salmon in Groups A or C (56.7 cm vs. 46.3 and 53.7, respectively; Tukey's *post hoc* test, $P < 0.05$) (Table S6, Supporting information). Discriminant analysis using life history traits alone (migration date and weight [JMP]) correctly assigned fish to the four STRUCTURE groups in 66% of the cases, with Group B_2 having a 82% correct assignment rate (data not shown).

Holmes (1924) also identified two distinct 'races' of Columbia River sockeye salmon that he thought could be separated on the basis of run timing in the vicinity of the Cascades of the Columbia, those arriving prior to 16 June and those arriving after this date. For comparative purposes with Holmes (1924) data, we have also chosen 16 June to discriminate early- from late-arriving portions of the sockeye run. Many of the Group B_2 sockeye were caught in May (49%) and most were caught prior to 16 June (84%), although a few Group B_2 fish continued to be caught through July (Fig. 4). Group A was sampled primarily from mid-June through July. Only a single fish in Group A was caught in May and just 5% were caught prior to 16 June. Similarly, no fish from Group B_1 or Group C were caught in May, and both groups occurred in the fishery predominately after mid-June (Fig. 4). ANOVA and Tukey's *post hoc* test ($P < 0.05$) indicated that sockeye salmon from Group B_2 migrated significantly earlier in the year than did the other three groups, whereas Group A migrated significantly later than other groups (Table S6, Supporting information). Comparison of run timing during two decade-long periods (1883–1893 and 1938–1947) indicated that a substantial portion of the sockeye salmon stock historically migrated through the lower Columbia River during the month of May and that this portion of the run is no longer present (Fig. 5).

Discussion

Contemporary sockeye salmon and kokanee populations

Similar to findings reported herein, Winans *et al.* (1996) and Waples *et al.* (2011) found the three contemporary populations of sockeye salmon in the Columbia River to be substantially differentiated using variation at allozyme loci. Beacham *et al.* (2006) reported a significant pairwise F_{ST} value of 0.097 for Okanogan and

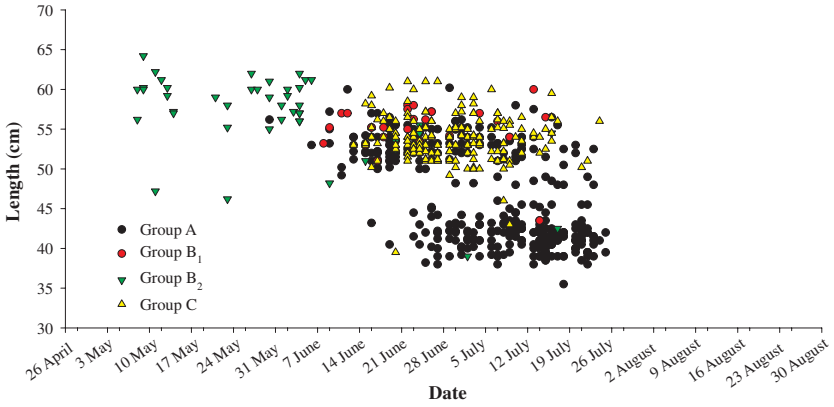

Fig. 4 Length (cm), date of collection (proxy for run timing) and genetic group assignment for individual sockeye salmon captured in the 1924 Columbia River fishery near The Dalles, Oregon. Individuals were sorted into genetically similar groupings as inferred by STRUCTURE.

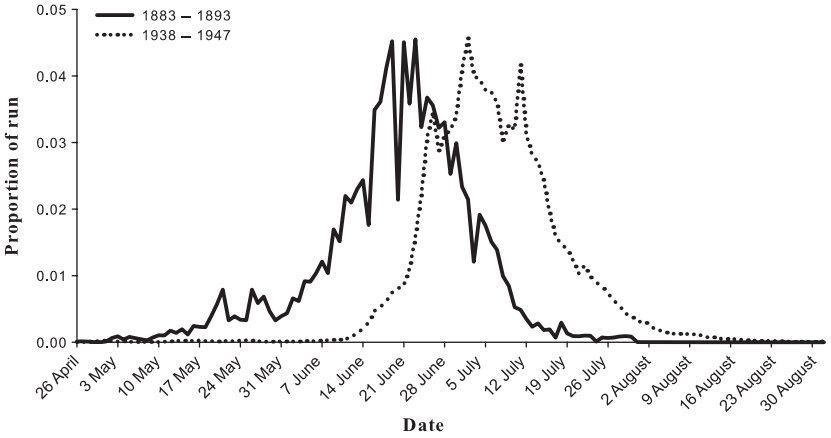

Fig. 5 Comparison of sockeye salmon run timing in the lower Columbia River, based on mean daily commercial landings from 1883 to 1893 at two fish wheels near the present site of Bonneville Dam (data from Smith 1895) and mean daily counts of returning adults at Bonneville Dam from 1938 to 1947 (U.S. Army Corps of Engineers data online at: http://www.nwp.usace.army.mil/environment/fishdata.asp [accessed January 2011]).

Wenatchee sockeye salmon over 14 microsatellite loci, similar to our pairwise value for these populations of 0.093 (Table S3, Supporting information). Consistent with previous findings that the Redfish Lake sockeye salmon population has probably experienced previous population bottlenecks (Winans *et al.* 1996; Waples *et al.* 2011), the Redfish Lake samples exhibited a very low level of allelic richness (Table 1), indicative of past genetic diversity losses.

The lack of genetic differentiation among Meadow Creek, Toby Creek and Columbia River at Fairmont samples suggests a common origin for kokanee above the Arrow Lakes. Furthermore, the observed genetic differences between the Meadow Creek Spawning Channel and Arrow Lakes tributaries (Octopus, Drimmie, Burton and Kuskanax creeks) kokanee (Table S3, Supporting information; all pairwise $F_{ST} \geq 0.031$), indicate that it is unlikely that these later populations originated from Meadow Creek kokanee planted in tributaries of the Arrow Lakes.

Population structure of Columbia River sockeye salmon in 1924

Our data indicate that the genetic population structure of Okanogan River and Lake Wenatchee sockeye salmon ESUs has changed surprisingly little over the past 87 years in spite of substantial anthropogenic perturbations and large fluctuations in abundance, and over 20 generations of genetic drift (Table 2; Fig. 2). Temporal changes in the fine-scale structure of Okanogan River and Lake Wenatchee sockeye salmon will be more fully explored in a forthcoming study and are not further dealt with herein.

Although the closest genetic neighbour of Group B_1 was the contemporary Redfish Lake sockeye salmon population (Fig. 2), the low assignment certainty of this group to our baseline (Tables 2 and 3) indicated that this group is probably composed of individuals from multiple populations or that genetic composition of Redfish Lake sockeye salmon has changed dramatically since 1924, or both.

The poor assignment of samples from Group B_2 to any contemporary sockeye salmon or kokanee population (Tables 2 and 3) would be expected if their historical source population(s) were absent from the baseline. At the time the 1924 sample was taken, the only sockeye salmon populations in the Columbia River were Lake Wenatchee, Okanogan River, Redfish Lake (the three contemporary populations), and an unknown number of populations that utilized headwater lakes in British Columbia (Table S1, Supporting information), and perhaps Suttle Lake on the Metolius River in Oregon, and Warm Lake in the South Fork Salmon River Basin in Idaho (Fulton 1970; Fryer 1995; Gustafson *et al.* 2007). However, sockeye salmon in Group B_2 were genetically unrelated to contemporary populations of what are presumed to be indigenous kokanee descended from sockeye salmon in Warm Lake and the Metolius River or to contemporary kokanee sampled in the Columbia River headwaters. Therefore, we believe that the fish in Group B_2 probably belonged to one or more anadromous ESUs of sockeye salmon that disappeared subsequent to 1924.

Alternatively, the sockeye salmon in Group B_2 may belong to a contemporary population of *O. nerka* that we were unable to sample. Currently, the only known population of indigenous kokanee in the Columbia River Basin that is lacking from our baseline occurs in the Chain Lake section of the Little Spokane River (NPCC 2011); however, construction of Little Falls Dam on the Little Spokane River in 1911 effectively entrained this population (Table S1, Supporting information). In addition, we have been unable to retrieve a genetic sample from a possibly indigenous beach-spawning population of kokanee, which spawns in late December, often under ice, in Payette Lake (D. Anderson, personal communication). However, anadromous access to the

Payette lakes was effectively blocked prior to the early 1920s (Table S1, Supporting information), making it unlikely that any remnant *O. nerka* stocks in the Payette River could have contributed to the 1924 return migration of sockeye salmon. It is also unlikely that any members of Group B_2 were progeny of hatchery-reared sockeye salmon juveniles from Alaska released into the Columbia River in the early 1920s, because Rich (1926) reported that no marked adult returns of Alaskan sockeye from this release experiment were encountered either at the release hatchery or in the commercial fishery landings in the Columbia River in 1924.

The levels of allelic and private allelic richness (i.e. genetic diversity) found in Group B_2 were similar to that found in Groups A and C (Table S4, Supporting information). Given that Groups A and C correspond to contemporary Okanogan River and Lake Wenatchee ESUs, respectively, it is probable that Group B_2 with equivalent levels of genetic diversity, represented an equivalent unit, at least an ESU. Furthermore, much of the allelic diversity observed in Group B_2 no longer exists; at least 16 alleles unique to this historical group were not observed in the contemporary sockeye salmon samples. Given that Group B_2 was only composed of 45 sockeye salmon, the total loss in diversity was probably much greater. It is probable that the measures of private allelic richness in the historical samples are an underestimate, given that the original sample came from a mixed fishery. Any mis-assignment of individuals by STRUCTURE to the scale groups could result in a reduction in the number of private alleles (Table S4, Supporting information).

Concordance of life history information with genetic findings

Consistent with our findings on run timing and body size of sockeye salmon sampled in the 1924 fishery (Fig. 4), the Columbia River was historically known to support both an early run of large sockeye salmon and a later run of both large and small sockeye salmon (Holmes 1924; Seufert 1980). The size dimorphism of Group A suggests the presence of two age classes and is consistent with both present-day life history characteristics of Okanogan River sockeye salmon (Gustafson *et al.* 1997) and with our genetic assignments. By contrast, only about 2% ($n = 181$) of sockeye salmon in Group C were under 46 cm in length indicative of a 3-year-old life history, the rest being 4- and 5-year-olds. The present-day Lake Wenatchee population also exhibits a very low frequency of 3-year-old adults (Gustafson *et al.* 1997), which is also in accord with the genetic relatedness we found between Group C and the Lake Wenatchee ESU.

Our findings that Group B_2 sockeye salmon migrated significantly earlier in the year than the other three groups (Table S6, Supporting information; Fig. 4) allows us to speculate as to this groups historical abundance. Historical run timing and abundance—based on commercial landings from 1883 to 1893 (Smith 1895)—show that this early run (prior to 16 June) represented 32% of overall historical landings. Comparison with more recent run timing—based on fish passage at Bonneville Dam from 1938 to 1947—shows that about 0.6% of sockeye salmon returned prior to 16 June (Fig. 5). Assuming fish wheel harvests were proportional to total harvest, and assuming a 73% historical harvest rate as calculated by Chapman (1986), the total run size of these early returning fish from 1890 to 1893—the first years that have total harvest estimates (Beiningen 1976)—would have ranged from about 76 000 in 1892 to over 508 000 in 1890, with an average of over 249 000 fish between the years 1890 and 1893. However, by 1916–1924, the proportion of sockeye salmon landed prior to 16 June had dropped to 4.2% (Holmes 1924), similar to the 7.5% per cent of our scale sample that fell within Group B_2.

Conservation implications

Although we were unable to identify a contemporary *O. nerka* population that reliably assigned to our historical Group B_2, we believe the Columbia River headwater lake system in British Columbia to be the most probable origin of fish in Group B_2. The only portion of the Columbia River Basin still accessible during the 1920s with sufficient habitat to produce upwards of half a million adult sockeye (our historical abundance estimate for this early-run component) occurred in the headwater lakes system in British Columbia. Collectively, these lakes had approximately 75% of the available rearing lake surface area at the time (Table S1, Supporting information).

Historically, sockeye salmon in the Columbia River undertook freshwater migrations of up to 1900 km (Windermere Lake, British Columbia) and ascended to elevations over 2000 m—the highest spawning grounds known for the species—at Alturas Lake in Idaho (Gustafson *et al.* 2007). This level of biocomplexity in spawning and rearing habitat by an aggregate of diverse locally adapted spawning populations is thought to offset environmentally induced abundance variation in individual populations and acts to sustain overall basin-wide productivity (aka 'portfolio effects;' Hilborn *et al.* 2003; Greene *et al.* 2010; Schindler *et al.* 2010). These portfolio effects were probably central to the maintenance of an estimated 2.2–2.6 million sockeye salmon in the Columbia River prior to Euro-American

contact (Chapman 1986). Current mean annual run sizes have averaged <124 000 over the most recent 10 years (2001–2010). Management efforts are currently underway to recover some of these lost 'portfolio effects' by re-introducing sockeye salmon to areas where they were historically extirpated in the Salmon, Yakima, Deschutes and upper Okanogan rivers (Table S1, Supporting information). In addition, one of the stated goals of the 'Tribal Vision for the Future of the Columbia River Basin' is to re-introduce anadromous salmon above Grand Coulee Dam (Heinith 1999). For these efforts to be successful, it is crucial to identify the most appropriate locally adapted source populations for potential re-introduction (Fraser *et al.* 2007a).

The present study is the first we are aware of that uses historical samples to inform potential re-introduction plans for an extirpated Pacific salmon ESU, although similar retrospective analyses of historical DNA have led to recommendations for ESU level re-introductions of the wolverine (*Gulo gulo*) in California (Schwartz *et al.* 2007), brush-tailed rock-wallaby (*Petrogale penicillata*) in Australia (Paplinska *et al.* 2011) and the Floreana mockingbird (*Mimus trifasciatus*) in the Galapagos Islands (Hoeck *et al.* 2010). Unfortunately, the vast majority of past efforts to transplant anadromous sockeye salmon have resulted in failure (Withler 1982; Wood 1995). In only three instances have self-perpetuating sockeye salmon runs been established, and in each case, the donor populations originated <100 km from the transplant site, which 'in many ways resembled a natural colonization of adjacent habitat' (Wood 1995, p. 205). As the recent documentation of reversion to anadromy by *O. nerka* populations entrained behind dams in the lower Fraser River Basin illustrates (Godbout *et al.* 2011), the most appropriate source population for re-introduction, when available, may well be the indigenous kokanee. However, our inability to identify a contemporary kokanee population closely related to Group B_2 indicates that any future effort to reestablish this early-migrating sockeye salmon component above Grand Coulee Dam may be difficult and that selection of a potential donor population for this effort must rely on information outside of the field of population genetics. This study does not preclude the potential for existing kokanee above Grand Coulee Dam in British Columbia to give rise to an anadromous run of *O. nerka* but does underscore the need to preserve all remaining *O. nerka* biodiversity in the Columbia River Basin.

Acknowledgements

We thank Sue Karren and the staff at the National Archives, Pacific Alaska Region, for their assistance in locating and retrieving historical scale notebooks from the USBF files. We also thank David Teel, Robin Waples, Gary Winans, Mike Ford, Jeff Hard, Linda Park, Professor Michael Hansen (Subject Editor) and three anonymous peer reviewers for providing valuable comments on earlier versions of the manuscript and Damon Holzer for assistance with the figures. We are grateful to various personnel that assisted in collecting contemporary genetic samples from the ODFW, WDFW, Okanagan Nation Alliance and Province of British Columbia. This research was supported by separate Northwest Fisheries Science Center Internal Grants to the first and second authors. Finally, we wish to acknowledge the work of Harlan B. Holmes (1898–1975), U.S. Bureau of Fisheries, who oversaw the collection of the scales in the 1920s and whose numerous field notes and reports provided us with considerable insight into the biology of sockeye salmon in a free-flowing Columbia River.

References

Beacham TD, McIntosh B, MacConnachie C, Miller KM, Withler RE, Varnavskaya NV (2006) Pacific Rim population structure of sockeye salmon as determined from microsatellite analysis. *Transactions of the American Fisheries Society*, **135**, 174–187.

Beiningen DT (1976) *Fish Runs. Investigative Reports of Columbia River Fisheries Project*, pp. E1–E65. Pacific Northwest Regional Commission, Vancouver, Washington.

Benjamini Y, Yekutieli D (2001) The control of false discovery rate under dependency. *Annals of Statistics*, **29**, 1165–1188.

Bouzat JL, Lewin HA, Paige KN (1998) The ghost of genetic diversity past: historical DNA analysis of the greater prairie chicken. *American Naturalist*, **152**, 2–6.

British Columbia Ministry of Environment (2011) Fisheries inventory data queries, fish stocking query. Online database. Available from http://a100.gov.bc.ca/pub/fidq/stockedSpeciesSelect.do

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: models and estimation procedures. *Evolution*, **21**, 550–570.

Chapman WM (1943) The spawning of Chinook salmon in the main Columbia River. *Copeia*, **1943**, 168–170.

Chapman DW (1986) Salmon and steelhead abundance in the Columbia River in the nineteenth century. *Transactions of the American Fisheries Society*, **115**, 662–670.

Chapman D, Peven C, Giorgi A *et al.* (1995) *Status of Sockeye Salmon in the Mid-Columbia Region*. Don Chapman Consultants, Boise, Idaho.

Cummings SA, Brannon EL, Adams KJ, Thorgaard GH (1997) Genetic analyses to establish captive breeding priorities for endangered Snake River sockeye salmon. *Conservation Biology*, **11**, 662–669.

Eldridge WH, Myers JM, Naish KA (2009) Long-term changes in the fine-scale population structure of coho salmon populations (*Oncorhynchus kisutch*) subject to extensive supportive breeding. *Heredity*, **103**, 299–309.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Facchin A, King G (1983a) Lake survey and stocking records for the Kootenay Region of British Columbia. Volume 1 – management units 4-1 to 4-22. British Columbia Ministry of Environment, Fisheries Technical Circular No. 54. Available

from http://www.for.gov.bc.ca/hfd/LIBRARY/MOE/BOSZ/BOSZ.pdf.

Facchin A, King G (1983b) Lake survey and stocking records for the Kootenay Region of British Columbia. Volume 2 – management units 4-23 to 4-40. British Columbia Ministry of Environment, Fisheries Technical Circular No. 55. Available from http://www.for.gov.bc.ca/hfd/LIBRARY/MOE/BOTA/BOTA.pdf.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Fraser DJ, Jones MW, McParland TL, Hutchings JA (2007a) Loss of historical immigration and the unsuccessful rehabilitation of extirpated salmon populations. *Conservation Genetics*, **8**, 527–546.

Fraser DJ, Hansen MM, Østergaard S, Tessier N, Legault M, Bernatchez L (2007b) Comparative estimation of effective population sizes and temporal gene flow in two contrasting population systems. *Molecular Ecology*, **16**, 3866–3889.

Fryer JK (1995) *Columbia Basin sockeye salmon: causes of their past decline, factors contributing to their present low abundance, and future outlook*. PhD thesis, University of Washington, Seattle, Washington.

Fulton LA (1970) *Spawning Areas and Abundance of Steelhead Trout and Coho, Sockeye, and Chum Salmon in the Columbia River Basin – Past and Present*. Special Scientific Reports – Fisheries No. 618, 37 pp. National Marine Fisheries Service, Washington, District of Columbia.

Godbout L, Wood CC, Withler RE *et al*. (2011) Sockeye salmon (*Oncorhynchus nerka*) return after an absence of nearly 90 years: a case of reversion to anadromy. *Canadian Journal of Fisheries and Aquatic Sciences*, **68**, 1590–1602.

Goudet J (1995) FSTAT (version 1.2): a computer program to calculate F-statistics. *Journal of Heredity*, **86**, 485–486.

Greene CM, Hall JE, Guilbault KR, Quinn TP (2010) Improved viability of populations with diverse life-history portfolios. *Biology Letters*, **6**, 382–386.

Guinand B, Scribner KT, Page KS, Burnham-Curtis MK (2003) Genetic variation over space and time: analyses of extinct and remnant lake trout populations in the Upper Great Lakes. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **270**, 425–433.

Gustafson RG, Wainwright TC, Winans GA, Waknitz FW, Parker LT, Waples RS (1997) Status review of sockeye salmon from Washington and Oregon. NOAA Technical Memorandum NMFS-NWFSC-33.

Gustafson RG, Waples RS, Myers JM *et al*. (2007) Pacific salmon extinctions: quantifying lost and remaining diversity. *Conservation Biology*, **21**, 1009–1020.

Hansen MM (2002) Estimating the long-term effects of stocking domesticated trout into wild brown trout (*Salmo trutta*) populations: an approach using microsatellite DNA analysis of historical and contemporary samples. *Molecular Ecology*, **11**, 1003–1015.

Hansen MM, Ruzzante DE, Nielsen EE, Bekkevold D, Mensberg KLD (2002) Long-term effective population sizes, temporal stability of genetic composition and potential for local adaptation in anadromous brown trout (*Salmo trutta*) populations. *Molecular Ecology*, **11**, 2523–2535.

Hansen MM, Meier K, Mensberg KLD (2010) Identifying footprints of selection in stocked brown trout populations: a spatio-temporal approach. *Molecular Ecology*, **19**, 1787–1800.

Hauser L, Seamons TR, Dauer M, Naish KA, Quinn TP (2006) An empirical verification of population assignment methods by marking and parentage data: hatchery and wild steelhead (*Oncorhynchus mykiss*) in Forks Creek, Washington, USA. *Molecular Ecology*, **15**, 3157–3173.

Heath DD, Busch C, Kelly J, Atagi DY (2002) Temporal change in genetic structure and effective population size in steelhead trout (*Oncorhynchus mykiss*). *Molecular Ecology*, **11**, 197–214.

Heinith B (1999) The tribal vision for the future of the Columbia River Basin and how to achieve it. Columbia River Inter-Tribal Fish Commission. Available from http://www.critfc.org/legal/vision.htm.

Hilborn R, Quinn TP, Schindler DE, Rogers DE (2003) Biocomplexity and fisheries sustainability. *Proceedings of the National Academy of Sciences USA*, **100**, 6564–6568.

Hoeck PEA, Beaumont MA, James KE, Grant RB, Grant PR, Keller LF (2010) Saving Darwin's muse: evolutionary genetics for the recovery of the Floreana mockingbird. *Biology Letters*, **6**, 212–215.

Holmes H (1924) A report upon an investigation of blueback salmon in the Columbia River. U.S. Bureau of Fisheries Report.

Kalinowski ST (2005) HP-Rare: a computer program for performing rarefaction on measures of allelic diversity. *Molecular Ecology Notes*, **5**, 187–189.

LaCroix B, McLean R (2005) Christina Lake Management Plan, a community based plan. Available from http://www.lakesteward.ca/pdfs/clmp2005.pdf.

Larsen PF, Hansen MM, Nielsen EE, Jensen LF, Loeschcke V (2005) Stocking impact and temporal stability of genetic composition in a brackish northern pike population (*Esox lucius* L.), assessed using microsatellite DNA analysis of historical and contemporary samples. *Heredity*, **95**, 136–143.

Lindley ST, Schick RS, Agrawal A *et al*. (2006) Historical population structure of Central Valley steelhead and its alteration by dams. *San Francisco Estuary and Watershed Science*. Available from http://repositories.cdlib.org/jmie/sfews/vol4/iss1/art3.

Martinez-Cruz B, Godoy JA, Negro JJ (2007) Population fragmentation leads to spatial and temporal genetic structure in the endangered Spanish imperial eagle. *Molecular Ecology*, **16**, 477–486.

Morin PA, McCarthy M (2007) Highly accurate SNP genotyping from historical and low-quality samples. *Molecular Ecology Notes*, **7**, 937–946.

Mullan JW (1986) Determinants of sockeye salmon abundance in the Columbia River, 1880–1982: a review and synthesis. U.S. Fish and Wildlife Service Biological Report No. 86(12).

Narum SR (2006) Beyond Bonferroni: less conservative analyses for conservation genetics. *Conservation Genetics*, **7**, 783–787.

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

Nielsen EE, Hansen MM (2008) Waking the dead: the value of population genetic analyses of historical samples. *Fish and Fisheries*, **9**, 450–461.

Nielsen EE, Hansen MM, Loeschcke V (1999) Analysis of DNA from old scale samples: technical aspects, applications and perspectives for conservation. *Hereditas*, **130**, 265–276.

Nielsen EE, Hansen MM, Bach LA (2001) Looking for a needle in a haystack: discovery of indigenous Atlantic salmon (*Salmo salar* L.) in stocked populations. *Conservation Genetics*, **2**, 219–232.

Northwest Power and Conservation Council (NPCC) (2011) Spokane subbasin assessment – aquatic, section 22, Intermountain Province Subbasin Plan. Prepared for the Northwest Power and Conservation Council. Available from http://www.nwcouncil.org/fw/subbasinplanning/admin/level2/intermtn/plan/22_spokane_assess_aquatic.pdf

Paetkau D, Slade R, Burden M, Estoup A (2004) Direct, real-time estimation of migration rate using assignment methods: a simulation-based exploration of accuracy and power. *Molecular Ecology*, **13**, 55–65.

Page RDM (1996) TREEVIEW: an application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences*, **12**, 357–358.

Palstra FP, Ruzzante DE (2010) A temporal perspective on population structure and gene flow in Atlantic salmon (*Salmo salar*) in Newfoundland, Canada. *Canadian Journal of Fisheries and Aquatic Sciences*, **67**, 225–242.

Paplinska JZ, Taggart DA, Corrigan T, Eldridge MDB, Austin JJ (2011) Using DNA from museum specimens to preserve the integrity of evolutionarily significant unit boundaries in threatened species. *Biological Conservation*, **144**, 290–297.

Park SDE (2001) *Trypanotolerance in West African cattle and the population genetic effects of selection*. PhD thesis, University of Dublin, Dublin, Ireland. Available from http://animalgenomics.ucd.ie/sdepark/ms-toolkit/.

Piry S, Alapetite A, Cornuet J-M, Paetkau D, Baudouin L, Estoup A (2004) GeneClass2: a software for genetic assignment and first-generation migrant detection. *Journal of Heredity*, **95**, 536–539.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences USA*, **94**, 9197–9201.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Reynolds J, Weir BS, Cockerham CC (1983) Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics*, **105**, 767–779.

Rich W (1926) Progress in biological inquiries, July 1 to December 31, 1924. *Report of the United States Commissioner of Fisheries for the Fiscal Year 1925, Appendix III*, pp. 37–64. Department of Commerce, Washington, District of Columbia.

Rousset F (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.

Ryder OA (1986) Species conservation and systematics: the dilemma of subspecies. *Trends in Ecology and Evolution*, **1**, 9–10.

Schindler DE, Hilborn R, Chasco B *et al.* (2010) Population diversity and the portfolio effect in an exploited species. *Nature*, **465**, 609–612.

Schwartz MK, Aubry KB, McKelvey KS *et al.* (2007) Inferring geographic isolation of wolverines in California using historical DNA. *Journal of Wildlife Management*, **71**, 2170–2179.

Seufert F (1980) *Wheels of Fortune*. Oregon Historical Society, Portland.

Smith HM (1895) Notes on a reconnaissance [sic] of the fisheries of the Pacific coast of the United States in 1894. *Bulletin of the U.S. Fish Commissioner*, **XIV**, 223–288.

Smith MJ, Pascal CE, Grauvogel Z, Habicht C, Seeb JE, Seeb LW (2011) Multiplex preamplification PCR and microsatellite validation enables accurate single nucleotide polymorphism genotyping of historical fish scales. *Molecular Ecology Resources*, **11**(Suppl. 1), 268–277.

Tessier N, Bernatchez L (1999) Stability of population structure and genetic diversity across generations assessed by microsatellites among sympatric populations of landlocked Atlantic salmon (*Salmo salar* L.). *Molecular Ecology*, **8**, 169–179.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) Micro-Checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

Waples RS (1995) Evolutionary significant units and the conservation of biological diversity under the Endangered Species Act. *American Fisheries Society Symposium*, **17**, 8–27.

Waples RS, Aebersold PB, Winans GA (2011) Population genetic structure and life history variability in *Oncorhynchus nerka* from the Snake River Basin. *Transactions of the American Fisheries Society*, **140**, 716–733.

Winans GA, Aebersold PB, Waples RS (1996) Allozyme variability of *Oncorhynchus nerka* in the Pacific Northwest, with special consideration to populations of Redfish Lake, Idaho. *Transactions of the American Fisheries Society*, **125**, 645–663.

Withler FC (1982) Transplanting Pacific salmon. Canadian Technical Report of Fisheries and Aquatic Sciences, No. 1079. Available from http://www.dfo-mpo.gc.ca/Library/18172.pdf.

Wood CC (1995) Life history variation and population structure in sockeye salmon. *American Fisheries Society Symposium*, **17**, 195–216.

Wood CC, Bickman JW, Nelson RJ, Foote CJ, Patton JC (2008) Recurrent evolution of life history ecotypes in sockeye salmon: implications for conservation and future evolution. *Evolutionary Applications*, **1**, 207–221.

E.I. research is focused on the identification of genetic population structure in marine and anadromous fishes. J.M. research efforts focus on characterizing existing and historical salmon populations and studying the effects of selective breeding, inbreeding and outbreeding on salmon in captive culture. R.G. current research involves identification of genetic population structure in marine and anadromous fishes and using historical data sources to characterize past biodiversity in marine animal populations.

Data accessibility

Sample locations and microsatellite data: DRYAD data identifier: doi: 10.5061/dryad.3nb259pq.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Physical attributes of historical rearing-lake habitat and ESU status for sockeye salmon in the Columbia River Basin.

Table S2 Microsatellite loci, primer sequences, product size range, number of alleles detected, and PCR thermocycling conditions for *Oncorhynchus nerka* samples.

Table S3 Pairwise F_{ST} values of contemporary *Oncorhynchus nerka* populations and genetic groupings from the 1924 historical Columbia River sockeye salmon run (Groups A, B_1, B_2, and C).

Table S4 Allelic richness and private allelic richness among sockeye salmon samples from the Columbia River Basin.

Table S5 Mean probabilities from genetic assignment tests of contemporary Columbia River *Oncorhynchus nerka* populations to the four historical genetic groups of sockeye salmon as identified in STRUCTURE, using the GeneClass2 assignment probability method (Paetkau *et al.* 2004).

Table S6 Group means for sockeye salmon migration day of year, length, and weight (±SE). Groups with different letter designations are significantly different from one another ($P < 0.05$, Tukey's *post hoc* test).

Fig. S1 Probability of contemporary population samples of *Oncorhynchus nerka* representing 1–17 clusters determined from 12 replicate analyses using STRUCTURE 2.2 (Pritchard *et al.* 2000; Falush *et al.* 2003) and the associated ΔK statistic, which measures the steepness of increase of P(D) and indicates the most likely number of clusters (Evanno *et al.* 2005).

Fig. S2 STRUCTURE results for the individual genotype analysis of the 17 contemporary *Oncorhynchus nerka* samples assuming six different clusters of individuals ($k = 6$) using an admixture model and no prior population information.

Fig. S3 Probability of mixed historical sample of 603 sockeye salmon from 1924 representing 1–12 clusters determined from 20 replicate analyses using STRUCTURE 2.2 (Pritchard *et al.* 2000; Falush *et al.* 2003) and the associated ΔK statistic, which measures the steepness of increase of P(D) and indicates the most likely number of clusters (Evanno *et al.* 2005).

Fig. S4 Probability of Group A (as identified in initial STRUCTURE run) representing 1–12 clusters determined from 20 replicate analyses using STRUCTURE 2.2 (Pritchard *et al.* 2000; Falush *et al.* 2003) and the associated ΔK statistic, which measures the steepness of increase of P(D) and indicates the most likely number of clusters (Evanno *et al.* 2005).

Fig. S5 Probability of Group B (as identified in initial STRUCTURE run) representing 1–12 clusters determined from 20 replicate analyses using STRUCTURE 2.2 (Pritchard *et al.* 2000; Falush *et al.* 2003) and the associated ΔK statistic, which measures the steepness of increase of P(D) and indicates the most likely number of clusters (Evanno *et al.* 2005).

Fig. S6 Probability of Group C (as identified in initial STRUCTURE run) representing 1–12 clusters determined from 20 replicate analyses using STRUCTURE 2.2 (Pritchard *et al.* 2000; Falush *et al.* 2003) and the associated ΔK statistic, which measures the steepness of increase of P(D) and indicates the most likely number of clusters (Evanno *et al.* 2005).

Fig. S7 STRUCTURE results for the individual genotype analysis of contemporary *Oncorhynchus nerka* samples ($N = 1103$) combined with scale samples from 1924 (Groups A, B_1, B_2, and C; $N = 603$) assuming seven different clusters of individuals ($k = 7$) using an admixture model and no prior population information.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.