Population differentiation determined from putative neutral and divergent adaptive genetic markers in Eulachon (Thaleichthys pacificus, Osmeridae), an anadromous Pacific smelt

JOHN R. CANDY,* NATHAN R. CAMPBELL,† MATTHEW H. GRINNELL,* TERRY D. BEACHAM,* WESLEY A. LARSON‡ and SHAWN R. NARUM†
*Fisheries and Oceans Canada, Pacific Biological Station, 3190 Hammond Bay Road, Nanaimo, BC V9T 6N7, Canada, †Columbia River Inter-Tribal Fish Commission, 3059-F National Fish Hatchery Road, Hagerman, ID 83332, USA, ‡School of Aquatic and Fishery Science, University of Washington, 1122 NE Boat Street, Box 355020, Seattle, WA 98195, USA

Abstract

Twelve eulachon (Thaleichthys pacificus, Osmeridae) populations ranging from Cook Inlet, Alaska and along the west coast of North America to the Columbia River were examined by restriction-site-associated DNA (RAD) sequencing to elucidate patterns of neutral and adaptive variation in this high geneflow species. A total of 4104 single-nucleotide polymorphisms (SNPs) were discovered across the genome, with 193 putatively adaptive SNPs as determined by $F_{ST}$ outlier tests. Estimates of population structure in eulachon with the putatively adaptive SNPs were similar, but provided greater resolution of stocks compared with a putatively neutral panel of 3911 SNPs or previous estimates with 14 microsatellites. A cline of increasing measures of genetic diversity from south to north was found in the adaptive panel, but not in the neutral markers (SNPs or microsatellites). This may indicate divergent selective pressures in differing freshwater and marine environments between regional eulachon populations and that these adaptive diversity patterns not seen with neutral markers could be a consideration when determining genetic boundaries for conservation purposes. Estimates of effective population size ($N_e$) were similar with the neutral SNP panel and microsatellites and may be utilized to monitor population status for eulachon where census sizes are difficult to obtain. Greater differentiation with the panel of putatively adaptive SNPs provided higher individual assignment accuracy compared to the neutral panel or microsatellites for stock identification purposes. This study presents the first SNPs that have been developed for eulachon, and analyses with these markers highlighted the importance of integrating genome-wide neutral and adaptive genetic variation for the applications of conservation and management.

Keywords: effective population size, eulachon, genetic stock structure, microsatellites, RAD sequencing, single-nucleotide polymorphisms

Received 30 October 2014; revision received 30 January 2015; accepted 7 February 2015

Introduction

There is growing recognition that population diversity within exploited species can contribute to long-term sustainability and needs to be more explicitly considered in management and conservation schemes (Hilborn et al. 2003; Hutchinson 2008). Biologically and genetically differentiated populations have been credited with a major role in conferring sustainability and buffering the overall productivity of marine fish (Ruzzante et al. 2006) and anadromous Pacific salmon (Schindler et al. 2010). The initial steps in protecting biological diversity are to first identify diversity and then take inventory of the units of diversity that require conservation (Ford 2004; Irvine & Fraser 2006). Until recently, genetic data for this assessment have relied largely on putatively neutral markers such as microsatellites. However, thorough representation of the genome is particularly critical for exploited organisms with relatively high gene flow to distinguish subtle patterns of differentiation that may be associated with local adaptation and important for long-term conservation.

Next-generation sequencing has provided new tools to identify large numbers of informative single-nucleotide
polymorphisms (SNPs) across the genome, which includes surveys of both neutral and adaptive loci used to determine levels of genetic diversity and genetic differentiation. Sequencing of restriction-site-associated DNA (RAD) tags (Miller et al. 2007) simultaneously facilitates both marker discovery and genotyping-by-sequencing (Narum et al. 2013). This approach provides an opportunity to develop large numbers of SNP markers in order to monitor species of conservation concern.

Eulachon (Thaleichthys pacificus, Osmeridae) are anadromous smelts found in the North Pacific Ocean ranging from northern California to the southeastern Bering Sea along the Alaska coast (Gustafson et al. 2012). Eulachon are a small, highly fecund forage fish that mature at three years of age, although older age classes have been observed further north (Rogers et al. 1990; Moody 2010). Females spawn during the spring, upstream of the mouth of large river systems, and fertilized eggs adhere to the substrate, where they hatch in about 20–40 d depending on water temperature (Hay & McCarter 2000). Once hatched, the larvae are immediately flushed to sea where they appear to be dispersed by estuarine and ocean currents (Barradough 1964; Hay & McCarter 2000). Eulachon also known as ‘candlefish’ because once dried, the high oil content is such that the fish burn when ignited, and they remain a prominent source of diet fat for coastal First Nation people by either drying or rendering into a thick oil (Kuhnlein et al. 1982; Joyce et al. 2004).

Prior to this study, studies of population structure in eulachon have used variation in mitochondrial DNA haplotypes (McLean et al. 1999) and microsatellite markers (McLean & Taylor 2001; Beacham et al. 2005; Flannery et al. 2013). Studies with microsatellite markers (McLean & Taylor 2001; Beacham et al. 2005) have found small but statistically significant differences between populations ranging from Columbia River to Cook Inlet. Samples taken from offshore trawl fisheries showed a clear geographical cline of relative abundance of the respective eulachon stocks, where higher representation of northern populations was found in northern samples and higher representation of more southern populations was found in more southern samples (Beacham et al. 2005). Year-to-year variation in eulachon allele frequencies within British Columbia coastal river systems was found to have similar levels of variation between river systems, with pairwise $F_{ST}$ values for microsatellite data ranging between 0.0014 and 0.0130 (Beacham et al. 2005). Cluster analysis showed affinities between populations in the south from the Fraser, Columbia and Cowlitz rivers, a central British Columbia grouping of Kemano, Klinaklini and Bella Coola river populations, and a northern British Columbia group of Nass and Skeena river populations. The limited degree of genetic differentiation observed in eulachon is more typical of marine species in comparison with other anadromous fishes such as Pacific salmon (Beacham et al. 2005). Recently, Flannery et al. (2013) surveyed samples collected at 26 eulachon spawning sites from the Gulf of Alaska with a subset of the same 14 microsatellite loci used by Beacham et al. (2005). Broad-scale regional structure consisted of two groupings, a northern region consisting of Yakutat Forelands, Prince William Sound and Cook Inlet members, and a southern region consisting of upper Lynn Canal, Berners Bay, the Stikine Strait and Behm Canal members. The lack of differentiation within the northern region suggested that the counter-clockwise Alaska gyre may facilitate larval dispersal and promote gene flow among these northern sites, while the Alexander Archipelago shelters the southern region from oceanic currents and restricts gene flow between the two regions (Flannery et al. 2013).

The objective of the current study was to develop the first panels of SNP markers for eulachon in order to compare population differentiation from putatively neutral and adaptive SNP panels. This new set of SNPs also allowed comparisons to the putatively neutral microsatellites previously screened on the same populations. Variation in both the adaptive and neutral panels was evaluated against that for microsatellite loci for the potential to reliably identify individuals to population and region, which is required for accurate genetic stock identification in mixed-stock fisheries.

**Materials and methods**

**Tissue collection and DNA extraction**

Tissue samples from 12 collections ($n = 494$) of Pacific eulachon were obtained from rivers in Alaska, British Columbia and Washington from 2001 to 2012 (Table 1 and Fig. 1). Tissue collections used in the analysis came from the US Fish and Wildlife Service (USFWS), Washington Department of Fish and Wildlife (WDFW), and the Department of Fisheries and Oceans (DFO) laboratory, Pacific Biological Station, Nanaimo, British Columbia. All samples were collected from maturing fish returning to rivers for spawning, except in the Columbia River where out-migrating larvae were collected. The availability and quality of the tissue collections determined the number of individuals analysed per sample, which ranged from 22 to 77 individuals per sample for SNPs. Tissue samples were preserved and stored in ethanol until used for DNA extraction. Extractions were performed with Qiagen DNeasy kits (Qiagen, Valencia, CA, USA). The microsatellite data used in this analysis were from previously surveyed samples described in Beacham et al. (2005).
Library preparation

Libraries for RAD-seq were prepared using methods previously described by Miller et al. (2012) with slight modifications. DNA in the samples was quantified in 96-well assay plates using the Quant-iT dsDNA pico-green assay kit (Life Technologies, Grand Island, NY, USA) and a PerkinElmer Victor 5 plate reader. We used 250 ng of each sample in 100-μL restriction enzyme digests (SbfI, New England Biolabs, Ipswich, MA, USA). Each sample then was tagged by sticky end ligation with one of 96 uniquely barcoded adapters (P1 adapter) to the SbfI cut sites using T4 DNA ligase (New England Biolabs, Ipswich, MA, USA). Once barcoded, the samples were mixed together into libraries of 96 individuals each and approximately 4 μg of each pooled library was sheared to an average size of 500 bp with a Bioruptor UCD-300 sonicator (Diagenode, Denville, NJ, USA). Following sonication, each library was concentrated to a volume of 100 μL using Qiagen MinElute columns. Size selection was performed on each library with Agencourt AMPure XP beads (Beckman-Coulter, Indianapolis, IN, USA) by

---

Table 1

| Region                  | Population      | Sampling Years | Latitude | Longitude | Sample Collection | n SNP | n M\(\mu\) |
|-------------------------|-----------------|----------------|----------|-----------|-------------------|-------|-----------|
| Northern GOA            | Kenai River     | 2004           | 60°32'38" | 151°16'43" | USFWS            | 71    | –         |
|                         | Twenty Mile River | 2001           | 60°58'40" | 128°40'43" | DFO              | 32    | 101       |
| SE Alaska-B.C.          | Stikine River   | 2006           | 56°33'50" | 132°24'16" | USFWS            | 66    | –         |
|                         | Nass River      | 2008           | 54°58'37" | 129°53'22" | DFO              | 41    | 387       |
|                         | Skeena River    | 2010           | 54°08'15" | 130°05'40" | DFO              | 33    | 502       |
|                         | Klinaklini River| 2002           | 51°05'32" | 125°37'34" | DFO              | 41    | 100       |
|                         | Kingcome River  | 2002           | 50°55'38" | 126°11'14" | DFO              | 36    | 128       |
|                         | Keman River     | 2001           | 54°03'17" | 128°39'28" | DFO              | 42    | 140       |
|                         | Bella Coola River | 2003           | 52°23'24" | 126°46'39" | DFO              | 33    | 288       |
| Fraser-Columbia         | Fraser River    | 2009           | 49°07'07" | 123°11'27" | DFO              | 40    | 636       |
|                         | Cowlitz River   | 2002           | 46°05'52" | 122°54'40" | DFO              | 37    | 200       |
|                         | Columbia River  | 2011–12        | 46°14'39" | 124°03'29" | WDFW             | 22    | 70        |

Fig. 1 Location of the twelve eulachon collections used for RAD sequencing. Populations and regional groups labeled as in Table 1. Dotted line indicates boundary between regional groups.
first binding DNA fragments larger than 700 bp with a 2:1 ratio of DNA to beads. The suspension then was cleared with a magnetic stand, and the unbound DNA in the supernatant was transferred to a fresh tube with another 50 µL of AMPure beads, creating a 1:1 ratio of DNA to suspension buffer (binding DNA fragments larger than 200 bp). The suspension was again cleared with a magnetic rack, and the supernatant was removed. The DNA bound to the magnetic beads was then washed twice with 70% ethanol and eluted using TE buffer (the final product contains DNA fragments ranging from 200 to 700 bp). The remaining steps of the RAD library preparation follow the methods outlined in (Miller et al. 2012). Prior to sequencing, a 1:2000 dilution of each library was quantified by qPCR using standard Illumina PCR primers and Power Sybr qPCR master mix (Life Technologies) on an ABI 7900HT instrument (Life Technologies). The libraries then were sequenced by single-end 100 base reads using an Illumina HiSeq1500 sequencer (Illumina Inc., San Diego, CA, USA). One of the libraries was sequenced again using paired-end 100 base reads to generate longer scaffolds for RAD loci of interest; this provides extended sequences to design primers for amplification of target loci for future applications (See data accessibility).

**SNP discovery and genotyping**

Genotyping and SNP discovery were performed using a bioinformatics pipeline provided and detailed in Miller et al. (2012) which incorporates the alignment tool Novoalign (Novocraft, Selangor, Malaysia). Briefly, the 100 base reads were trimmed from the 3’ end to 90 bases to remove the most error-prone portion of each read. Trimmed reads then were filtered using quality scores to eliminate poor quality reads. Following quality filtering, the six base barcode sequence was used to separate reads into a single fastq file for each individual containing 78 base sequence reads after removal of the barcode and partial SbfI restriction site.

SNPs were identified using the first 500 K sequencing reads from a subset of 12 individuals that were selected to represent genetic variation from each collection site. Reads within each of these files were collapsed into a FASTA file such that each unique sequence is represented only once, and a header line contains the name of the collection, a sequence identification number and the number of occurrences for that particular sequence. Sequences occurring three or fewer times were trimmed from the files since these would not pass filtering criteria in later steps and excluding them speeds up sequence alignment. These 12 files for the 12 individuals then were combined to yield a single file containing unique sequences occurring four or more times within each individual. This concatenated file was used to generate sequence alignments using the program Novoalign (Novocraft). These sequence alignment files then were analysed using a custom perl script (described in Miller et al. 2012) and available in Hecht et al. (2012) to identify biallelic RAD tag sequences. Sequence alignments with more than two SNPs were excluded, while those with up to two SNPs were allowed if no more than two haplotypes were identified (the SNP sites were in phase). Once an index of putative SNP loci was created, exact matches for each allele sequence were counted within the reads from each individual. Genotypes were collected for loci with read depths of five or greater. If the read depth at a given locus was less than five, the genotype was scored as missing. At this level of read depth, the probability of genotyping a true heterozygote as a homozygote due to a sampling error given our genotype scoring parameters was 3.1%. Heterozygous genotypes were assigned if the ratio of reads between allele 1 and allele 2 (A1/A2) was between 10:1 and 1:10. If the ratio was >10:1, the genotype was scored as being homozygous for the A1 allele. If the ratio was below 1:10, the genotype was scored as being homozygous for the A2 allele.

**Filtering SNP loci**

The SNP discovery process resulted in the identification and genotyping of 7968 biallelic loci for 593 individuals (data accessibility). These data then were filtered by removing loci that produced genotypes in <70% of the samples and/or had an overall minor allele frequency of <3.0%. Likewise, samples with >70% missing genotypes at the remaining loci were removed from the data set. Finally, the remaining loci were tested for deviation from Hardy–Weinberg expectations (HWE) within each of the 12 collections with GENEPOP (Raymond & Rousset 1995). To filter out markers with technical problems such as null alleles (i.e. heterozygote deficit loci) and potential paralogous sequence variants (i.e. heterozygote excess loci), loci with significant Bonferroni-corrected P-values (α = 0.05; P-value < 0.000012) for deviation from HWE in six or more of the collections tested also were excluded. After applying these filters, the final data set included 4104 SNP loci genotyped in 494 individuals, with an average genotype percentage of 96.5% among both SNP loci and individuals.

**Detection of loci under putative selection**

An $F_{ST}$ outlier approach was used to determine a set of candidate SNPs that had significantly higher $F_{ST}$ values than expected under a neutral model of selection. A set of 4104 SNPs from the 12 collections was analysed with LOSITAN (Antao et al. 2008) with parameter settings of
50,000 simulations, confidence interval of 0.95, false discovery rate of 0.1, attempted FST of 0.009, subsample size of 50 and simulated FST of 0.008. Of the 4104 loci considered candidates, a panel of 193 SNPs was found with a high FST, greater than the probability of 0.975 as putatively under divergent selection (Fig. 2 and data accessibility). A panel of putative neutral SNPs consisted of the remaining 3911 SNPs with nonsignificant FST values determined with LOSITAN.

Genetic diversity between populations
We calculated expected heterozygosity (H_e) and allelic richness (A_{R}) in each population for the neutral and adaptive panels using the package HIERFSTAT (Goudet 2005, 2013) implemented in R (R_Core_Team 2013). To compare mean diversity measures between populations, we used the pairwise Wilcoxon rank sum tests with Bonferroni-corrected P-values for multiple tests (Bauer 1972; Holander et al. 2013) implemented in R (R_Core_Team 2013).

Population structure
Cavalli-Sforza & Edwards’ (1967) CSE chord distance was calculated from SNP allelic frequency data with PHYLIP 3.5.7 (Felsenstein 1985) for the putatively neutral and adaptive SNP panels. A neighbour-joining tree (Saitou & Nei 1987) was used to visualize the CSE distances for each of these loci sets using FIGTREE 1.4.0 (Rambaut 2012). Node support for tree structure was generated by bootstrapping across loci using 1000 replicate trees (Felsenstein 1985).

In order to assess the extent of genetic divergence at different levels of geographic hierarchy, we partitioned the overall molecular variance into components corresponding to the population divergence within and among predefined regional groups by the analysis of molecular variance (AMOVA) (Excoffier et al. 1992) using ARLEQUIN 3.5.1.2 (Excoffier & Lischer 2010). Region membership was determined by the tree structure and the DAPC analysis described below. Negative variance estimates were set to zero. Default settings in ARLEQUIN were also used to calculate FST values between pairs of samples and level of statistical significance between population pairs for both the neutral and adaptive panels.

We used discriminant analysis of principal components (DAPC; Jombart et al. 2010) conducted in R with the package ADEGENET 1.3.0 (Jombart 2008) to identify and describe clusters of genetically related individuals.

Estimation of effective population size
Estimates of contemporary effective population size (N_e) for each of the 12 populations were performed with the linkage disequilibrium method using N_eESTIMATOR (Do et al. 2014) with the following parameter settings: (i) a random mating model; (ii) a P_{crit} value, the criterion for excluding rare alleles, of 0.02 for microsatellites and 0.05 for the SNPs. Estimates of contemporary N_e are considered one or more generations that span the time frame represented by the samples (Waples 2005). The method assumes that all loci in the analysis are physically unlinked. To test the effects of physical linkage on the N_e estimates, pairwise comparisons with R^2 ≤ 0.5 were removed before calculating N_e using custom R scripts (see data accessibility) following the method of Grunetal. (2014). The 193 adaptive SNPs were also removed from this analysis because potential biases from loci under selection have not been fully evaluated using the linkage disequilibrium method for estimating N_e (Waples 2006).

Assignment test
ONCOR (Kalinowski et al. 2007) implements a leave-one-out (LOO) test, sequentially removing each fish from the baseline data set and estimating its origin using the rest of the baseline. Complete multilocus genotypes are required for this assignment test, so missing SNP loci were imputed using the R program ADEGENET 1.3.0 (Jombart 2008), replacing missing loci with a mean value for the population (see data accessibility for custom scripts). A total of 66 124 genotypes, or 3.4%, were imputed for the neutral panel and 4607 genotypes, or 4.8%, were imputed for the adaptive panel. For comparison, the assignment tests were run for the 12 populations using both SNP panels and the 10 populations where microsatellite genotypes were available. The effects of
the imputed data on assignment estimation were tested by replacing missing data with a standard genotype of ‘0102’ in each of the panels analysed.

Results

Twelve eulachon populations distributed from Alaska to the Columbia River were sequenced and genotyped using RAD sequencing. A total of 4104 SNPs were discovered and genotyped for 494 individuals. Of the 4104 SNPs, 193 were found to be putatively under selection using an $F_{ST}$ outlier test.

Sequencing data

A total of 1.5B reads of single-end sequencing data were collected from nine Illumina HiSeq lanes. However, technical issues with library construction resulted in poor barcode recoveries in two of those lanes, and approximately 425M reads were discarded after multiple attempts to generate data. Of the remaining seven lanes of raw sequence reads, an average of 82% passed quality filtering and 70% of the reads were retained after barcode splitting. Among the 593 individual samples attempted, the average number of quality filtered reads was 1.2M and ranged from zero to 3.6M. The average number of reads among the 494 individual samples retained after all filtering steps was 1.4M and ranged from 0.5 to 3.6M. A single lane of paired-end sequencing was carried out using a library containing 94 individual samples resulting in 142 M reads. Paired-end data were not used in genotyping but rather collected for later assembly of longer sequence scaffolds for the purposes of designing primers flanking selected SNP loci.

Genetic diversity

For the putative neutral panel and microsatellites, genetic diversity estimates for $A_R$ and $H_E$ displayed a narrow range of diversity values, with no significant trends across the geographic range (Fig. 3). For the neutral SNP panel, mean $A_R$ by population ranged from 1.27 to 1.28 and mean $H_E$ by population ranged from 0.26 to 0.29. However, for the putative adaptive SNPs, $A_R$ and $H_E$ both displayed a trend of increasing values from the southern Fraser-Columbia populations to the northern GOA populations (Fig. 3A,B). Mean $A_R$ by population ranged from 1.63 in the south to 1.84 in the north, and mean $H_E$ by population ranged from 0.23 in the south to 0.31 in the north. There was a significant pairwise difference for $A_R$ between all northern GOA and Fraser-Columbia populations, except for the Twenty Mile River and Columbia River populations ($P < 0.05$). The comparison between Kingcome River and Kenai River populations was also significant.

![Fig. 3](image-url)

**Fig. 3** Box plots of genetic diversity statistics for eulachon populations for the putatively adaptive SNPs (A and B), putatively neutral SNPs (C and D), and the microsatellites (E and F) where: A, C and E is the allelic richness average number of alleles per locus ($A_R$) and B, D, and F show the gene diversity or expected heterozygosity ($H_E$). Green boxes represent Fraser-Columbia populations, blue boxes SE Alaska-BC populations, and red boxes northern GOA populations. Each Tukey box plot shows median (notch), upper and lower quartiles (upper and lower box boundaries), and 95% interval for (box whiskers). Populations Kenai and Stikine were not screened for microsatellites.

© 2015 The Authors. *Molecular Ecology Resources* published by John Wiley & Sons Ltd.
(P < 0.05). Additionally, there was a significant difference for \( H_E \) between all northern GOA and Fraser-Columbia populations (P < 0.05). Significant differences for \( H_E \) (P < 0.05) also were observed between Kingcome River—Kenai River and Bella Coola River—Kenai River populations.

**Population structure**

Population genetic structure was examined at both the putatively neutral and putatively adaptive SNP panels. The analysis of genetic stock structure using the SNP's loci was compared to patterns of stock structure seen with the set of microsatellite markers.

Pairwise \( F_{ST} \) values between populations ranged from 0.0000 to 0.0131 (mean = 0.0034) for the neutral panel, and ranged from 0.0000 to 0.1041 (mean = 0.0416) for the adaptive panel (Table 2). The adaptive panel displayed more significant pairwise \( F_{ST} \) values when compared to the neutral panel at the highest significance level of \( P < 0.001 \) (52 vs. 35 of 66 comparisons, respectively). An AMOVA using the three geographic groupings (Columbia-Fraser, SE Alaska-BC, and northern GOA) yielded an overall \( F_{ST} \) of 0.00381 (P < 0.001) for the neutral panel and an overall \( F_{ST} \) of 0.0195 (P < 0.001) for the adaptive panel (Table 3). The largest component of total genetic variation was observed within collections (99.6% for neutral and 93.13% for adaptive panel). The three geographic groups accounted for 0.66% neutral panel and 6.47% adaptive panel of total observed variation (\( F_{CT} = 0.00659 \) for neutral and \( F_{CT} = 0.01510 \) for the adaptive panel, P < 0.001), and the remaining variation among populations within groups was small (0.00% neutral and 0.40% adaptive) but significant (\( F_{SC} = 0.00279 \) and \( F_{SC} = 0.00450 \), P < 0.001 for neutral and adaptive panels, respectively).

**Table 2** Pairwise \( F_{ST} \) values for the putative neutral SNPs (top triangle) and putative SNPs under selection (bottom triangle)

|        | Twenty | Kenai | Stikine | Nass | Skeena | Kemano | Bella | Coola | Kingcome | Klinaklini | Fraser | Columbia | Cowltitz |
|--------|--------|-------|---------|------|--------|--------|-------|-------|----------|------------|--------|----------|---------|
| Twenty |        | 0.0000| 0.0080  | 0.0099| 0.0111 | 0.0112 | 0.0107| 0.0131| 0.0124   | 0.0098     | 0.0000 | 0.0128   |         |
| Kenai  | 0.0010 |       | 0.0077  | 0.0035| 0.0083 | 0.0054 | 0.0086| 0.0098| 0.0093   | 0.0066     | 0.0000 | 0.0103   |         |
| Stikine| 0.0840 | 0.0732| 0.0000  | 0.0000| 0.0000 | 0.0000 | 0.0000| 0.0000| 0.0000   | 0.0000     | 0.0000 | 0.0028   |         |
| Nass   | 0.0882 | 0.0682| 0.0000  | 0.0000| 0.0000 | 0.0000 | 0.0000| 0.0000| 0.0000   | 0.0000     | 0.0000 | 0.0026   |         |
| Skeena | 0.0903 | 0.0722| 0.0000  | 0.0043| 0.0011 | 0.0007 | 0.0014| 0.0018| 0.0018   | 0.0007     | 0.0000 | 0.0052   |         |
| Kemano | 0.0903 | 0.0700| 0.0003  | 0.0000| 0.0064 | 0.0014 | 0.0023| 0.0016| 0.0032   | 0.0025     | 0.0000 | 0.0025   |         |
| Bella | 0.0932 | 0.0793| 0.0033  | 0.0078| 0.0067 | 0.0108 | 0.0012| 0.0009| 0.0022   | 0.0000     | 0.0000 | 0.0046   |         |
| Coola  | 0.1041 | 0.0890| 0.0101  | 0.0181| 0.0134 | 0.0139 | 0.0110| 0.0005| 0.0024   | 0.0000     | 0.0000 | 0.0044   |         |
| Klinaklini | 0.1033 | 0.0852| 0.0037  | 0.0057| 0.0126 | 0.0071 | 0.0074| 0.0026 | 0.0013 | 0.0000 | 0.0053 |
| Fraser | 0.0946 | 0.0837| 0.0085  | 0.0577| 0.0392 | 0.0423 | 0.0446| 0.0373| 0.0346 | 0.0000 | 0.0000 |
| Columbia | 0.0715 | 0.0599| 0.0333  | 0.0347| 0.0302 | 0.0249 | 0.0450| 0.0470| 0.0306 | 0.0163 | 0.0000 |
| Cowltitz | 0.0980 | 0.0869| 0.0476  | 0.0468| 0.0519 | 0.0526 | 0.0548| 0.0491| 0.0498 | 0.0053 | 0.0000 |

Values in bold font are significant at P < 0.05, and values with grey background are significant at P < 0.001. Negative \( F_{ST} \) values were set to zero.

For (Table 3A): Average overall loci: \( F_{ST} = 0.0195, F_{SC} = 0.0045, F_{CT} = 0.0151, *P < 0.001. \)

For (Table 3B): Average overall loci: \( F_{ST} = 0.00381, F_{SC} = -0.00279, F_{CT} = 0.00659, *P < 0.001. \)
Tree structure topology using CSE chord distances and a neighbour-joining clustering algorithm was almost identical for both the neutral and adaptive panels of SNPs (Fig. 4A,B). These results indicate that there is a three-population southern Columbia-Fraser group (Cowlitz, Columbia, and Fraser rivers), a seven-population British Columbia (BC) – SE Alaska group (Stikine, Nass, Skeena, Klinaklini, Kingcome, Kemano and Bella Coola rivers) and a two-population northern Gulf of Alaska (GOA) group (Twenty Mile and Kenai rivers). Branch lengths were approximately ten times shorter for the neutral panel (Fig. 4A) compared with the adaptive panel (Fig. 4B). There was 100% bootstrap support from consensus trees for nodes between regional groupings for both panels.

Results of the DAPC analysis indicated that the optimal number of clusters of individual genotypes was three according to the Bayesian information criterion, corresponding to the inferred group membership determined from the tree topology (Fig. 5). The first principal axis differentiated the northern GOA populations from SE Alaska-BC and Columbia-Fraser populations, and the second axis expressed the variability between SE Alaska-BC and Columbia-Fraser. For the neutral panel with three clusters of populations, the DAPC explained 47% of the variance with 150 principal component eigenvalues and three discriminant analysis functions; DAPC with the adaptive panel and three clusters explained 86% of the variance with 100 principal component eigenvalues and three discriminant analysis functions.

**Estimation of \( N_e \)**

Estimates of \( N_e \) for the neutral panel were quite variable, ranging from \(<1000\) for Kenai to infinity for Stikine, Kingcome and Fraser River populations (Table 4). Estimates of \( N_e \) were smaller using no \( R^2 \) cut-off, compared to \( R^2 \leq 0.5 \) cut-off, and this effect was more pronounced for large population estimates. For example, Klinaklini River \( N_e \) values increased from \( 14 \, 126 \) to \( 46 \, 021 \) when the cut-off was applied, while estimates for populations with small \( N_e \) such as the Kenai, Nass, Kemano and Cowlitz river populations changed by only a few individuals when the \( R^2 \) cut-off was applied.

Estimates of \( N_e \) from microsatellite markers were variable, ranging from \( 1601 \) for Twenty Mile River to infinity for the Nass, Skeena, Kingcome and Cowlitz rivers. For the Bella Coola and Kemano rivers, where there were both SNP and microsatellite-based estimates, \( N_e \) estimates ranged from \( 3998 \) to \( 5738 \) for SNPs to \( 1411 \) to \( 1656 \).
for microsatellites encompassing the 95% confidence limits. Estimates of \( N_e \) between marker types were much larger for Twenty Mile River (8984 for SNPs and 1601 for \( M_L \)) and for Columbia River (2250 for SNPs and 13 688 for \( M_L \)).

**Assignment test**

Per cent correct assignment of the LOO algorithm for the neutral panel ranged from 6% to 94% (mean = 53%) for assignment to population and 89% to 100% (mean = 98%) for assignment to region (Table 5). Per cent correct assignment for the adaptive panel a ranged from 22% to 71% (mean = 64%) for assignment to population and 97% to 100% (mean = 99%) for assignment to region. The per cent correct assignment using fixed value alternative of ‘0102’ used to replace missing data compared to the imputed data had mean assignment accuracy of 63% to population and 98% to region and 61% to population and 98% to region for the neutral and adaptive panels, respectively. Accuracy of the LOO algorithm for the populations using microsatellite panel was considerably poorer, ranging from 10% to 47% (mean = 19%) for assignment to population and from 47% to 80% (mean = 67%) for assignment to region.

**Discussion**

We used RAD sequencing to discover putatively neutral and adaptive SNP variation in eulachon. Analyses of this data set provide important information required to identify conservation units, assess risk of extinction and set recovery goals for extremely depressed populations over the southern portion of the eulachon’s range (Gustafson et al. 2012; Schweigert et al. 2012). As with other recent studies using RAD sequencing (Hess et al. 2014; Larson et al. 2014), adaptive differentiation uncovered in this study has implications for refining conservation units which may have been previously determined using only neutral genetic markers. These new informative loci also will be useful for increasing accuracy when assigning individual eulachon to stock-of-origin in bycatch fisheries. The patterns of genetic stock structure using RAD-SNPs were similar to that found in a previous microsatellite study (Beacham et al. 2005), but yielded higher resolution genetic information to infer genetic stock structure, calculate diversity measures and estimate contemporary effective population size.

**Population structure and differentiation**

Eulachon genetic stock differentiation was found to be relatively weak compared to other anadromous fishes, such as Pacific salmonids (e.g. Narum et al. 2008; Beacham et al. 2012; Larson et al. 2014; Seeb et al. 2014). The high level of gene flow among eulachon stocks may be due to a life history strategy of limited juvenile freshwater residence, larval seaward migration and passive transport by current at a larval life stage precluding imprinting on natal streams such as that observed in salmon species (Hay & McCarter 2000). This life history may lead to limited philopatry and low differentiation among eulachon stocks in comparison with salmonids. Although overall stock differentiation was weak compared to salmonids, the putatively adaptive SNP panel demonstrated considerably higher levels of differentiation at both the population and regional levels.
demonstrated by higher mean and overall \( F_{ST} \) values and greater assignment success than the neutral panel. Both SNP panels maintained the same regional structure, and results were consistent with other studies using microsatellites markers. Proportionally, the adaptive panel showed more variation between regional groups and less
variation within populations compared to the neutral panel.

The BC-SE Alaska group determined from the SNP survey has common group membership (Kemano, Klinaklini, Bella Coola, Nass, Skeena Rivers) with the microsatellite central mainland and Nass-Skeena groups of Beacham et al. (2005) and the southern group (Stikine River) of Flannery et al. (2013). The BC-SE Alaska group identified in this study could extend as far north as Upper Lynn Canal, the northern extent of the Flannery et al.’s (2013) Alaskan southern group. Finally, the genetic signature of the northern GOA group corresponded with the northern Alaskan group, with Kenai River as a common population in both analyses (Flannery et al. 2013; current study). Using the SNP panels, we were unable to define the boundary between the BC-SE Alaska group and the northern GOA group because of a gap in sampling north of the Stikine River and south of Cook Inlet. Flannery et al. (2013) placed the boundary between the Alaskan northern and the southern groups at Yakutat Bay (Fig. 1). Further studies with adaptive markers and dense sampling may provide a more precise location for this boundary break.

The Columbia-Fraser group determined from the SNP survey corresponded with the same sampled populations used in the microsatellite data, with both analyses displaying the same high degree of bootstrap consensus tree support of 100% and 97%, respectively, for this group (Beacham et al. 2005; current study). The marked genetic boundary outlined by both SNP and microsatellite data sets between Fraser-Columbia and other northerly populations in British Columbia was consistent with variation in meristic characters of eulachon. The average number of vertebrae in Fraser River eulachon sampled in the early 1940s differed significantly from those of eulachon sampled in more northerly parts of British Columbia, indicative of low levels of mixing between the Fraser River and more northerly runs (Hart & McHugh 1944). Hay & McCarter (2000) reanalysed the vertebrae counts, included data from the Columbia River (DeLacy & Batts 1963) and found no significant difference between Columbia River and Fraser River populations, but significant differences were detected in vertebrae counts between Fraser-Columbia populations and more northerly populations in British Columbia. Both the genetic and meristic data suggested that the southern distinct population segment (DPS) extends only as far north as the Fraser River, instead of possibly the Nass River as proposed by Gustafson et al. (2012). The samples screened in this study did not allow us to evaluate the level of interannual variability compared to regional variability which was found to be quite high for microsatellites (Beacham et al. 2005).

**Regional diversity**

Despite the relatively low levels of divergence of eulachon stocks with putatively neutral markers, the presence of loci under putative divergent selection suggests that gene flow has not completely eroded adaptive divergence among populations. This result was similar to those for other species such as anadromous Pacific salmon (Hess et al. 2013) and marine fishes (Atlantic Herring, Limborg et al. 2012; Atlantic Cod, Nielsen et al. 2009) that experience high gene flow but retain adaptive genetic variation among stocks. However, it is unknown whether the adaptive loci inferred in this study were directly related to local adaptation of eulachon populations to specific marine and freshwater environments or were linked to genes underlying specific traits.

Analysis of the putatively adaptive SNPs indicated that there was a gradient of lower genetic diversity ($A_R$ and $H_E$) in the southern populations (Columbia-Fraser) to higher genetic diversity in the northern populations (northern GOA), with the SE Alaska-BC populations being intermediate. This gradient may be related to the respective effective population sizes of each group, but a gradient of genetic diversity may also be indicative of a demographic expansion from a refuge where leading-edge colonizers following glacial retreat would have both lower $A_R$ and $H_E$ (Hewitt 1996). However, if increased diversity in the north could be attributed to effective population size or to expansion from northern refugia, then we would have expected to also see this same gradient in the putatively neutral SNPs, but this was not observed. If eulachon expanded northward from the southern Pacific refuge, thought to be near the Columbia River, as proposed by Flannery et al. (2013), it is possible that large effective population sizes prevented founder effects and a cline in genetic diversity for neutral markers, while differences in selective pressures among recently colonized environments created the patterns of diversity observed among the putatively adaptive SNPs.

However, a northern eulachon refuge cannot be ruled out as eulachon spawn in the lower reaches of rivers and have limited freshwater requirements. It is known that a refuge for plants and animals persisted at least as far north as the Alexander Archipelago in southeast Alaska, which served as a centre of biotic dispersal upon regional deglaciation (Heaton et al. 1996; Carrara et al. 2007). Additionally, eulachon are found only in large rivers systems which have spring freshets characteristic of headwaters draining large snow packs or glaciers (Hay & McCarter 2000). During the fullest extent of glaciation, these large glacial river systems would likely have occurred along the coast where the continental shelf was exposed by lower ocean levels possibly, providing suitable habitat for spawning eulachon. It is possible that the
high diversity observed at putatively adaptive SNPs in northern populations of eulachon could have been a result of a northern glacial refuge; however, we do not have the data to definitively support this hypothesis.

Estimation of \( Ne \)

Estimates of contemporary effective population sizes were quite variable, especially for larger values of \( Ne \). There was moderate concordance between microsatellite and SNP estimates of \( Ne \), especially when the effective population size was smaller. This trend is consistent with previous findings, where estimates of \( Ne \) were most precise for small populations such as endangered species or salmon subpopulations (Waples & Teel 1990). The confidence intervals were smaller for the SNPs compared with the microsatellites. Estimates of \( Ne \) without an \( R^2 \) cut-off were lower in all cases, and it has been found that estimates are biased downward when data from physically linked loci are not removed (Larson et al. 2014; current study). Where sample sizes are small for populations, estimates of \( Ne \) are more likely to be infinite, which indicates that there is no detectable signal of linkage disequilibrium due to genetic drift after correcting for the expected disequilibrium caused by sampling error (Do et al. 2014). The linkage disequilibrium method appears to be fairly robust to violations of closed populations unless migration rates are higher than 5–10%; with increased migration, \( Ne \) should converge on an estimate of global (metapopulation) effective size (Waples & England 2011). Relatively weak genetic differentiation, possibly caused by high immigration rates between populations, might contribute to the large number of infinite estimates of \( Ne \). It is not unreasonable to estimate effective population size much smaller than the census size (Frankham 1996), but unfortunately, there is little census information available for eulachon as they are difficult to locate and count in large river systems. In these circumstances where census information is hard to obtain, periodic re-estimation of effective population size may be a useful tool for monitoring population status for conservation purposes.

Applications for fisheries management

High \( F_{ST} \) markers have been shown to provide increased discrimination power for genetic stock identification applications to determine the origin of unknown fish (Ackerman et al. 2011; Hess et al. 2011) and may provide strong utility for this purpose in mixed-stock eulachon bycatch fisheries. The adaptive SNP panel showed greater assignment success than the neutral SNP panel, and both panels were considerably more powerful than microsatellites. The adaptive panel increased accuracy by about 10–20% on average for assignment to population compared with the neutral SNP panel. The adaptive panel also required far fewer SNPs to genotype for characterizing mixed-stock fisheries samples (193 vs. 3911). However, as in this study, the expected assignment power may be overestimated if the same samples used to select a panel of particularly divergent loci then are used to estimate assignment accuracy (Anderson 2010). Screening of further samples collected but not used for SNP discovery will be required to assess the accuracy reported in this study. Replacing missing data for the LOO algorithm with either imputed data (mean genotype per population) or a fixed value made little difference to the accuracy estimates where the total number of replaced genotypes was <5%. The SNP-based assignment accuracy was much higher than the microsatellites, possibly caused by overly optimistic SNP assignment success as the same samples were used for both discovery and assignment testing. However, increased assignment accuracy with SNPs compared to microsatellites might be due to improved loci selection. Recent advances in genomic techniques (Narum et al. 2013; current study) make it possible to screen thousands of SNPs in hundreds of individuals, selecting for those SNPs that display elevated levels of differentiation between populations.

Summary

Patterns of regional stock structure for eulachon are robust, being conserved across neutral and adaptive variation and between marker types (microsatellites and SNPs). Population-specific differentiation of eulachon is low, with regional groups well differentiated, possibly due to the life history of eulachon, or possibly due to the lingering historical effects of demographic radiation from glacial refugia. Gene flow between populations may be affected by limited freshwater requirements, limited ability to home to natal stream and spatial patterns of passive larval dispersion. Our study highlights the importance of integrating both neutral and adaptive genetic variation to infer the evolutionary mechanisms involved in shaping population diversity. Further, this study demonstrates that conservation and management of eulachon can be enhanced by the use of genome-wide selection of highly differentiating SNPs.

The resources presented in this study were as follows: (i) RAD-sequencing information for 593 (494 after filtering) individual eulachon samples, (ii) a list of thousands of RAD loci with SNP variations identified, (iii) paired-end sequencing data for extending the sequence at interesting RAD loci and (iv) GENEPOP files of the screened
loci. In addition, custom R scripts used in this study were provided as data accessibility.

Acknowledgements

We thank First Nations and Fisheries and Oceans Canada staff for the eulachon collections along the coast of British Columbia. Washington Department of Fisheries staff provided samples from Washington, and Alaska Department of Fish & Game, US Fish & Wildlife Service and University of Alaska Fairbanks staff provided samples from Alaska. We appreciate assistance from Travis Jacobson in the genetics laboratory. Funding was provided by Fisheries and Oceans Canada. WAL was supported by a National Science Foundation Graduate Research Fellowship (Grant # DGE-0718124).

References

Ackerman MW, Habicht C, Seeb LW (2011) Single-nucleotide polymorphisms (SNPs) under diversifying selection provide increased accuracy and precision in mixed-stock analyses of sockeye salmon from the Copper River, Alaska. Transactions of the American Fisheries Society, 140, 865–883.

Anderson EC (2010) Assessing the power of informative subsets of loci for population assignment: standard methods are upwardly biased. Molecular Ecology Resources, 10, 701–710.

Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: a workbench to detect molecular adaptation based on a FST-outlier method. BMC Bioinformatics, 9, 323.

Barraclough WE (1964) Contribution to the marine life history of the Eulachon Thaleichthys pacificus. Journal of the Fisheries Research Board of Canada, 21, 1333–1337.

Bauer DF (1972) Constructing confidence sets using rank statistics. Journal of the American Statistical Association, 67, 687–690.

Beacham TD, Hay DE, Le KD (2005) Population structure and stock identification of Eulachon (Thaleichthys pacificus), an anadromous smelt, in the pacific northwest. Marine Biotechnology (NY), 7, 363–372.

Beacham T, Jonsen KL, Wallace C (2012) A comparison of stock and individual identification for Chinook salmon in British Columbia provided by microsatellites and single-nucleotide polymorphisms. Marine and Coastal Fisheries, 4, 1–22.

Carrara PE, Ager TA, Baichlal JF (2007) Possible refugia in the Alexander Archipelago of southeastern Alaska during the late Wisconsin glaciation. Canadian Journal of Earth Sciences, 44, 229–244.

Cavalli-Sforza LL, Edwards AW (1967) Phylogenetic models. Models and estimation procedures. American Journal of Human Genetics, 19, 223–257.

DeLacy AC, Batts BS (1963) Possible population heterogeneity in the Columbia River smelt. Fisheries Research Institute, College of Fisheries, University of Washington.

Do C, Waples RS, Peel D et al. (2014) NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. Molecular Ecology Resources, 14, 209–214.

Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources, 10, 564–567.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics, 131, 479–491.

Felsenstein J (1985) Confidence limits on phylogenies: an approach using bootstrap. Evolution, 39, 783–791.

Flannery BG, Spangler RE, Norcross BL, Lewis CJ, Wenburg JK (2013) Microsatellite analysis of population structure in Alaska Eulachon with application to mixed-stock analysis. Transactions of the American Fisheries Society, 142, 1036–1048.

Ford MJ (2004) Conservation units and preserving diversity. Evolution illuminated: salmon and their relatives. Oxford University Press, New York, 338-357.

Frankham R (1996) Relationship of genetic variation to population size in wildlife. Conservation Biology, 10, 1500–1508.

Goudet J (2005) PROGRAM NOTE: hierfstat, a package for R to compute and test hierarchical F-statistics. Molecular Ecology Notes, 5, 184–186.

Goudet J (2013) Hierfstat: Estimation and tests of hierarchical F-statistics, p. 46. Available: http://www.r-project.org, http://www.unil.ch/popgen/software/hierfstat.htm (June 2014).

Gruenthal KM, Witting DA, Ford T et al. (2014) Development and application of genomic tools to the restoration of green abalone in southern California. Conservation Genetics, 15, 109–121.

Gustafson RG, Ford MJ, Adams PB et al. (2012) Conservation status of eulachon in the California Current. Fish and Fisheries, 13, 121–138.

Hart JL, McHugh JL (1944) The smelts (Osmeridae) of British Columbia. Fisheries Research Board of Canada Bulletin, 64, 1–27.

Hay D, McCarter PB (2000) Secretariat CSA (2000) Status of the eulachon Thaleichthys pacificus in Canada. Canadian Stock Assessment Secretariat Research Document, 145, 1–92.

Heaton TH, Tabbot SL, Shields GF (1996) An ice age refugium for large mammals in the Alexander Archipelago, southeastern Alaska. Quaternary Research, 46, 186–192.

Hecht BC, Thrower FP, Hale MC, Miller MR, Nichols KM (2012) Genetic architecture of migration-related traits in rainbow and steelhead trout, Oncorhynchus mykiss. G3: Genes|Genomes|Genetics, 2, 1113–1127.

Hess JE, Matala AP, Narum SR (2011) Comparison of SNPs and microsatellites for fine-scale application of genetic stock identification of Chinook salmon in the Columbia River Basin. Molecular Ecology Resources, 11, 137–149.

Hess JE, Campbell NR, Close DA, Docker MF, Narum SR (2013) Population genomics of Pacific lamprey: adaptive variation in a highly dispersive species. Molecular Ecology, 22, 2898–2916.

Higuchi RE, Caudill CC, Keefer ML et al. (2014) Genes predict long distance migration and large body size in a migratory fish, Pacific lamprey. Evolutionary Applications, 7, 1192–1208.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. Biological Journal of the Linnean Society, 58, 247–276.

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

Holland M, Wolfe DA, Chicken E (2013) Nonparametric Statistical Methods. John Wiley & Sons, New York.

Hutchinson WF (2008) The dangers of ignoring stock complexity in fisheries management: the case of the North Sea cod. Biology Letters, 4, 693–695.

Irvin JR, Fraser GA (2006) Canada’s wild Pacific salmon policy and the maintenance of diversity. American Fisheries Society Symposium, 49, 567–594.

Jombart T (2008) adegenet: A R package for the multivariate analysis of genetic markers. Bioinformatics, 24, 1403–1405.

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC genetics, 11, 94.

Joyce TL, Lambert MB, Moffitt S (2004) Eulachon subsistence harvest opportunities. Final Report No. FIS02-075-1, Alaska Department of Fish and Game, Anchorage, Alaska.

Kalnowski ST, Manlove KR, Taber ML (2007) ONCOR: a computer program for genetic stock identification. Montana State University, Bozeman. Available: www.montan.edu/kalinowski/Software/ONCOR.htm (September 2008).

Kuhlin HV, Chan AC, Thompson JN, Nakai S (1982) Ooligan grease: a nutritious fat used by native people of coastal British Columbia. Journal of Ethnobiology, 2, 154–161.
Larson WA, Seeb LW, Everett MV et al. (2014) Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (Oncorhynchus tshawytscha). Evolutionary Applications, 7, 355–369.

Limborg MT, Helyar SJ, De Bruyn M et al. (2012) Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (Clupea harengus). Molecular Ecology, 21, 3686–3703.

McLean J, Taylor E (2001) Resolution of population structure in a species with high gene flow: microsatellite variation in the eulachon (Osmeridae: Thaleichthys pacificus). Marine Biology, 139, 411–420.

McLean J, Hay D, Taylor E (1999) Marine population structure in an anadromous fish: life-history influences patterns of mitochondrial DNA variation in the eulachon, Thaleichthys pacificus. Molecular Ecology, 8, S143–S158.

Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Research, 17, 240–248.

Miller MR, Brunelli JP, Wheeler PA et al. (2012) A conserved haplotype controls parallel adaptation in geographically distant salmonid populations. Molecular Ecology, 21, 237–249.

Moody MP (2010) Eulachon (Thaleichthys pacificus): past and present. p. 197. University of British Columbia, Vancouver, BC (Canada).

Narum SR, Banks M, Beacham TD et al. (2008) Differentiating salmon populations at broad and fine geographical scales with microsatellites and single nucleotide polymorphisms. Molecular Ecology, 17, 3464–3477.

Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA (2013) Genotyping-by-sequencing in ecological and conservation genomics. Molecular Ecology, 22, 2841–2847.

Nielsen EE, Hemmer-Hansen J, Poulsen NA et al. (2009) Genomic signatures of local directional selection in a high gene flow marine organism: the Atlantic cod (Gadus morhua). BMC Evolutionary Biology, 9, 276.

R Core Team, (2013) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rambaut A (2012) FigTree v1.4.0 2006-2012., Institute of Evolutionary Genetics software for exact tests and ecumenicism. At Http://tree.bio.ed.ca.uk/software/figtree/

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

Rogers JH, Birtwell IK, Kruzyński GM (1990) The pacific eulachon (Thaleichthys pacificus) as a pollution indicator organism in the Fraser River estuary, Vancouver, British Columbia. Science of the Total Environment, 97, 713–727.

Ruzzante DE, Mariani S, Bekkevold D et al. (2006) Biocomplexity in a highly migratory pelagic marine fish, Atlantic herring. Proceedings of the Royal Society B: Biological Sciences, 273, 1459–1464.

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution, 4, 406–425.

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

Schweigert J, Wood C, Hay D et al. (2012) Recovery potential assessment of eulachon (Thaleichthys pacificus) in Canada. p. 128. Department of Fisheries and Ocean, Ottawa, ON (Canada).

Seeb LW, Waples RK, Limborg MT et al. (2014) Parallel signatures of selection in temporally isolated lineages of pink salmon. Molecular Ecology, 23, 2473–2485.

© 2015 The Authors. Molecular Ecology Resources published by John Wiley & Sons Ltd.