NRC Research Press

ARTICLE

Single-nucleotide polymorphisms (SNPs) identified through genotyping-by-sequencing improve genetic stock identification of Chinook salmon (*Oncorhynchus tshawytscha*) from western Alaska

Wesley A. Larson, James E. Seeb, Carita E. Pascal, William D. Templin, and Lisa W. Seeb

Abstract: Genetic stock identification (GSI), an important tool for fisheries management that relies upon the ability to differentiate stocks of interest, can be difficult when populations are closely related. Here we genotyped 11 850 single-nucleotide polymorphisms (SNPs) from existing DNA sequence data available in five closely related populations of Chinook salmon (*Oncorhynchus tshawytscha*) from western Alaska. We then converted a subset of 96 of these SNPs displaying high differentiation into high-throughput genotyping assays. These 96 SNPs (RAD96) and 191 SNPs developed previously (CTC191) were screened in 28 populations from western Alaska. Regional assignment power was evaluated for five different SNP panels, including a panel containing the 96 SNPs with the highest F_{ST} across the CTC191 and RAD96 panels (F_{ST}96). Assignment tests indicated that SNPs in the RAD96 were more useful for GSI than those in the CTC191 and that increasing the number of reporting groups in western Alaska from one to three was feasible with the F_{ST}96. Our approach represents an efficient way to discover SNPs for GSI and should be applicable to other populations and species.

Résumé : L'identification génétique des stocks (IGS), un important outil de gestion des pêches reposant sur la capacité de distinguer des stocks d'intérêt, peut être difficile si les populations sont étroitement reliées. Nous avons génotypé 11 850 polymorphismes mononucléotidiques (SNPs) tirés de données de séquences d'ADN disponibles pour cinq populations étroitement reliées de saumon quinnat (*Oncorhynchus tshawytscha*) de l'ouest de l'Alaska. Nous avons ensuite converti un sous-ensemble de 96 de ces SNP présentant une forte différentiation en des tests de génotypage de haute capacité. Ces 96 SNP (RAD96) et 191 SNP établis précédemment (CTC191) ont fait l'objet d'un criblage dans 28 populations de l'ouest de l'Alaska. La puissance d'affection régionale a été évaluée pour cinq panels de SNP distincts dont un panel contenant les 96 SNP présentant les F_{ST} les plus élevés des panels CTC191 et RAD96 (F_{ST}96). Les tests d'affectation ont indiqué que les SNP du RAD96 étaient plus utiles pour l'IGS que ceux du CTC191 et qu'une augmentation d'un à trois du nombre de groupes pouvant être distingués dans l'ouest de l'Alaska était possible avec le F_{ST}96. Notre approche constitue une manière efficace de cerner des SNP pour l'IGS et devrait pouvoir s'appliquer à d'autres populations et espèces. [Traduit par la Rédaction]

Introduction

Genetic tools have been used to document biodiversity and to manage wild populations for over four decades (Utter 2004; Waples et al. 2008). These techniques are particularly applicable to Pacific salmon (*Oncorhynchus* spp.); salmon return to their natal streams with high fidelity, promoting local adaptation and the formation of genetically distinct populations (Shaklee et al. 1999; Stewart et al. 2003; Neville et al. 2006). Discrete management of these populations minimizes extirpation of lineages with smaller population sizes and preserves the resiliency of the species as a whole (Hilborn et al. 2003; Schindler et al. 2010).

As genetic techniques improved, genetic stock identification (GSI) became a commonly utilized tool for managing discrete populations of Pacific salmon (Dann et al. 2013). GSI uses the observed allelic frequencies of baseline populations sampled on the spawning grounds to infer the natal origin of fish captured in mixed-stock fisheries (Milner et al. 1985; Utter and Ryman 1993; Beacham et al. 2012). Population-specific assignment is rarely feasible; therefore, baseline datasets are often partitioned into reporting groups composed of genetically similar populations. The proportional contribution of each reporting group to mixed-stock samples is then estimated. GSI has been used to investigate the migration and distribution patterns of many Pacific salmonids (e.g., Habicht et al. 2010; Tucker et al. 2011; Larson et al. 2013) and to inform in-season management of mixed-stock fisheries (e.g., Seeb et al. 2000; Beacham et al. 2008*b*; Dann et al. 2013).

The genetic marker of choice for GSI has evolved dramatically over the past three decades, with allozymes being replaced by microsatellites and, most recently, by single-nucleotide polymorphisms (SNPs; Schlötterer 2004; Hauser and Seeb 2008). Compared with microsatellites, SNPs can be developed and assayed more quickly, and the resulting genotypes are easily transferred among laboratories (Seeb et al. 2011*a*). Recent advances in genomic techniques have made it possible to screen thousands of putative SNPs in hundreds of individuals (reviewed in Allendorf et al. 2010; Narum et al. 2013). Researchers can then select SNPs that display elevated levels of differentiation among populations of interest and convert them to high-throughput genotyping assays for screening thousands of individuals. This type of approach has already been used to assess hybridization between two species of trout (Hohenlohe et al. 2011; Amish et al. 2012) and promises to be

Received 17 September 2013. Accepted 8 January 2014.

Paper handled by Associate Editor Eric Taylor.

W.A. Larson, J.E. Seeb, C.E. Pascal, and L.W. Seeb. School of Aquatic and Fishery Sciences, University of Washington, 1122 NE Boat Street, Box 355020, Seattle, WA 98195-5020, USA.
W.D. Templin. Gene Conservation Laboratory, Alaska Department of Fish and Game, 333 Raspberry Road, Anchorage AK, 99518, USA.
Corresponding author: Lisa W. Seeb (e-mail: lseeb@uw.edu).

Table 1. Collection location, sampling region, reporting group, sampling year, and sample size for each population in the study.

Pop. No.	Location	Region	Reporting group	Sampling year	Sample size
1	Pilgrim River	Norton Sound	Norton Sound	2005, 2006	71
2	**Tubutulik River**	Norton Sound	Norton Sound	2009	85 (**56**)
3	North River	Norton Sound	Norton Sound	2010	60
4	Golsovia River	Norton Sound	Norton Sound	2006	59
5	Andreafsky River	Lower Yukon	Lower Yukon	2003	90
6	**Anvik River**	Lower Yukon	Lower Yukon	2007	52 (**51**)
7	Gisasa River	Lower Yukon	Lower Yukon	2001	81
8	Goodnews River	Kuskokwim Bay	Bristol–Kusk	2006	94
9	Arolik River	Kuskokwim Bay	Bristol–Kusk	2005	52
10	Kanektok River	Kuskokwim Bay	Bristol–Kusk	2005	93
11	Eek River	Kuskokwim: mouth	Bristol–Kusk	2005	76
12	Kisaralik River	Kuskokwim: lower	Bristol–Kusk	2005	94
13	Salmon River	Kuskokwim: middle	Bristol–Kusk	2006	94
14	George River	Kuskokwim: middle	Bristol–Kusk	2005	95
15	Kogrukluk River	Kuskokwim: middle	Bristol–Kusk	2005	49
16	**Kogrukluk River**	Kuskokwim: middle	Bristol–Kusk	2007	94 (**57**)
17	Necons River	Kuskokwim: middle	Bristol–Kusk	2007	94
18	Gagaryah River	Kuskokwim: middle	Bristol–Kusk	2006	94
19	Togiak River	West Bristol Bay	Bristol–Kusk	2009	94
20	Iowithla River	West Bristol Bay	Bristol–Kusk	2010	65
21	Stuyahok River	West Bristol Bay	Bristol–Kusk	2009	93
22	**Koktuli River**	West Bristol Bay	Bristol–Kusk	2010	94 (**56**)
23	Klutuspak Creek	West Bristol Bay	Bristol–Kusk	2009	94
24	Big Creek	East Bristol Bay	Bristol–Kusk	2004	65
25	Henshaw Creek	Middle Yukon	Middle Yukon	2001	88
26	Kantishna River	Middle Yukon	Middle Yukon	2005	94
27	Salcha River	Middle Yukon	Middle Yukon	2005	90
28	**Big Salmon River**	Upper Yukon	Upper Yukon	2007	71 (**47**)
Total					2275 (**267**)

Note: Pop No. corresponds to the numbers in Figs. 1 and 4. Ascertainment populations that were RAD-sequenced are in bold, and sample sizes for RAD sequencing are given in parentheses. RAD data were obtained from Larson et al. (2014). The reporting group is the group that was used for assignment tests. The Bristol Bay and Kuskokwim River reporting group is abbreviated Bristol–Kusk.

extremely applicable to the development of SNP panels for GSI (Storer et al. 2012).

Chinook salmon (*Oncorhynchus tshawytscha*) from western Alaska represent an excellent opportunity to apply genomic techniques towards the development of a SNP panel for GSI. Chinook salmon primarily spawn in drainages in four major regions in western Alaska: Norton Sound, Yukon River, Kuskokwim River, and Bristol Bay (Templin et al. 2011). Recent returns to all four regions have been substantially lower than their long-term average, renewing interest in the migration patterns and relative vulnerability of these stocks to both targeted and bycatch fisheries (Stram and Ianelli 2009; ADF&G 2013). GSI could be used to investigate the above questions, but a lack of substantial genetic differentiation among these regions has prevented its use. Specifically, evidence of shallow genetic structure among regions has been reported with allozyme (Gharrett et al. 1987), SNP (Templin et al. 2011), and microsatellite data (Olsen et al. 2011). The limited discriminatory power of these existing baseline datasets necessitated pooling all four regions in western Alaska into a single reporting group for GSI estimates (Templin et al. 2011; Larson et al. 2013). Nevertheless, given the apparent substructure suggested by previous studies, a search for additional SNPs that can differentiate the major regions in western Alaska is warranted.

Our goals were (*i*) to use genotyping-by-sequencing to develop a new set of 96 information-rich SNPs for western Alaska, (*ii*) to compare the resolving power of these 96 new SNPs with 191 existing SNPs, and (*iii*) to construct the best panel of 96 SNPs for GSI from all available SNPs. Panel sizes of 96 were selected because this represents the maximum number of SNPs that can be assayed simultaneously using the most prevalent genotyping platform for Pacific salmon management, the Fluidigm 96.96 dynamic array (Fluidigm, South San Francisco, California).

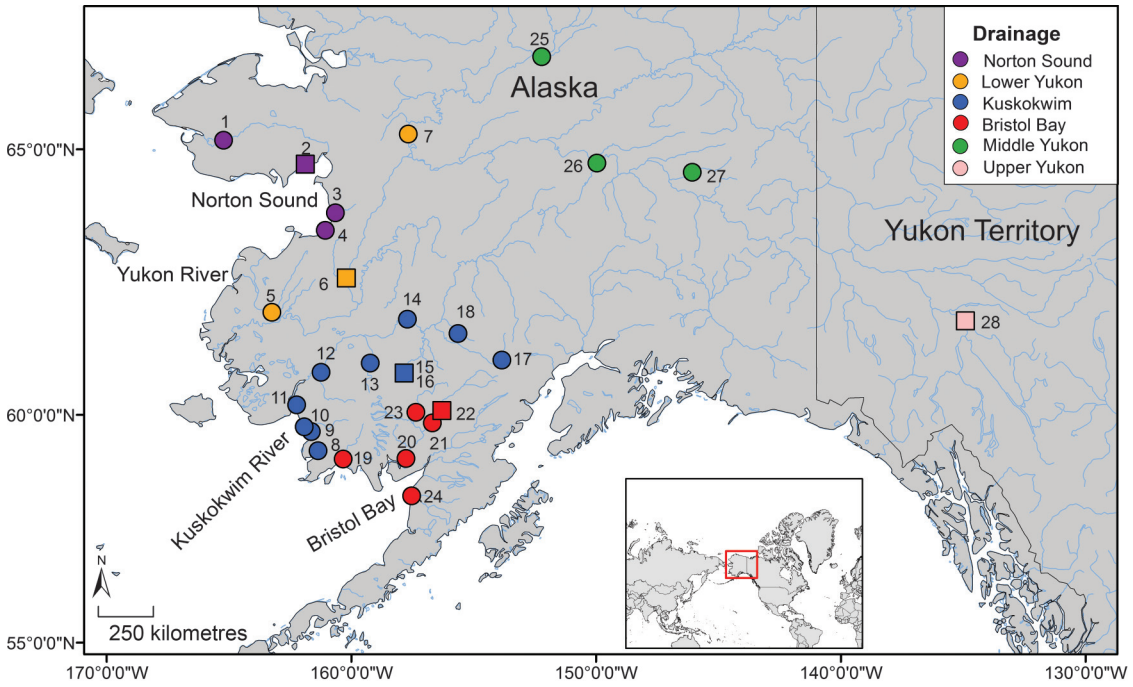
We identified 11 850 putative SNPs in five populations of Chinook salmon from western Alaska using data from restriction-site-associated DNA (RAD) sequencing obtained by Larson et al. (2014). We then developed high-throughput assays for 96 RAD-derived SNPs showing high levels of differentiation. The 96 RAD-derived SNPs along with 191 SNPs developed previously for Chinook salmon were genotyped in 28 populations from across western Alaska. From these data, we compared the resolving power of the 96 RAD-derived SNPs with the 191 previously developed SNPs, identified the 96 SNPs that displayed the highest levels of differentiation across these two panels, and tested the utility of the top 96 SNPs for GSI. Using the top 96 SNPs, we were able to increase the number of reporting groups for GSI in western Alaska from one to three. Based on these results, we believe that SNP discovery using genomic techniques can improve GSI in populations characterized by low genetic divergence.

Materials and methods

Tissue sampling

Fin clips preserved in 100% ethanol were available from 28 populations of Chinook salmon collected throughout coastal western Alaska and the middle and upper Yukon River (2275 fish total, 21 populations shared with Templin et al. 2011; Table 1; Fig. 1). Five populations that spanned the study area were RAD-sequenced by Larson et al. (2014). These five ascertainment populations did not have unusually small census sizes and were genetically similar to proximate populations (Templin et al. 2011). All 28 populations were then used to evaluate the resolving power of the RAD-derived and previously developed SNPs. Chinook salmon from the upper and to a lesser extent middle Yukon River are highly differentiated from those of coastal western Alaska (Smith et al. 2005c;

Fig. 1. Sampling locations for the 28 populations of Chinook salmon. Ascertainment populations that were RAD-sequenced are denoted by squares. Table 1 provides additional details about each sampling site.



Templin et al. 2011). We included populations from this region to anchor inferences of population structure and ensure that GSI outside of coastal western Alaska was feasible with the SNPs discovered in this study. Collections from multiple years were pooled if sample sizes were <48 following recommendations of Waples (1990).

Quality filtering and SNP discovery

Raw RAD sequence data (single-end, 100 base pair target length) were available from Larson et al. (2014). Quality filtering, SNP discovery, and genotyping were performed on these data using a modified version of the pipeline first described in Miller et al. (2012) and adapted by Everett et al. (2012). The last base pair of each read was trimmed, and reads with <90% chance of being error-free were discarded. A separate file was then created for each individual containing all of their unique sequences and the number of times they occurred. Sequences occurring <6 or >200 times were removed. We only used the 16 individuals with the most data from each population for SNP discovery to reduce the frequency of false positives in our dataset. Putative SNPs within each individual were identified with the program NOVOALIGN 2.07 (http://www.novocraft.com) using the following alignment parameters: maximum of 10 alignments returned per unique sequence and a maximum alignment score of 245. Alignments for each individual were filtered using the methods described in Miller et al. (2012) to retain RAD tags with a single putative SNP that did not align closely to any other sequence. Polymorphism data from each individual were combined to form a catalog of RAD tags, each containing a single, bi-allelic putative SNP. This catalog was aligned to each individual using Bowtie version 0.12.9 (Langmead et al. 2009), and sequence counts for each allele were tabulated using the methods of Miller et al. (2012). Genotypes were obtained from allele counts using a two-allele maximum likelihood approach following the framework of Hohenlohe et al. (2010), with a static error rate based on the published value for Illumina HiSeq data (0.0016; Minoche et al. 2011).

As an initial screen for paralogous sequence variants (PSVs), we genotyped 33 individuals from a haploid family, available from another study (Everett and Seeb 2014), and removed loci with >10% heterozygosity. PSVs are closely related sequences from different genomic locations that are abundant in salmonids as a result of a whole-genome duplication event (Allendorf and Thorgaard 1984; Seeb et al. 2011*b*). Although PSVs are difficult to genotype accurately because they do not segregate as single loci (Gidskehaug et al. 2011), haploid individuals can be used to differentiate true SNPs from PSVs because true SNPs will be homozygous in all haploid individuals, whereas PSVs will often be heterozygous (Hecht et al. 2013).

Allele frequencies and sample sizes for each putative SNP were calculated using GENEPOP 4 (Rousset 2008) to enable the removal of uninformative or unreliable loci. Putative SNPs that failed to genotype in >80% of individuals and those with minor allele frequencies <0.1 in all populations were removed. As a final filtration step, we removed individuals with <10× average coverage across the filtered SNPs because these individuals likely contained a substantial amount of missing data that could cause genotyping errors.

It is important to note that SNPs discovered in this study are not necessarily the same as those discovered in Larson et al. (2014) because Larson et al. (2014) used the STACKS software package (Catchen et al. 2011, 2013) for SNP discovery.

Paired-end assembly and BLAST annotation

Paired-end data (100 × 2 base pair target length) were available from eight Chinook salmon collected in coastal western Alaska (Larson et al. 2014). Paired-end assemblies for each locus were conducted using the methods of Etter et al. (2011) and adapted by Everett et al. (2012) to increase query lengths for BLAST annotation and template length for assay design. We used the program VELVET 1.1.06 (Zerbino and Birney 2008) to create a consensus sequence for each locus using all the paired and single-end reads that aligned to that locus. Consensus sequences for each locus were aligned to the Swiss-Prot database using the BLASTX search algorithm. Alignments with E values of $\leq 10^{-4}$ were retained. If multiple alignments had E values of $\leq 10^{-4}$ for the same locus, then the alignment with the lowest E value was retained.

Table 2. Pairwise F_{ST} values for the five ascertainment populations calculated with 11 850 RAD-derived SNPs (overall F_{ST} = 0.041).

	Tubutulik River	Anvik River	Kogrukluk River	Koktuli River
Anvik River	0.030			
Kogrukluk River	0.026	0.005		
Koktuli River	0.028	0.006	0.002	
Big Salmon River	0.097	0.077	0.075	0.077

Construction of high-throughput assays from RAD data

We selected 150 RAD-derived SNPs that displayed high levels of differentiation in our ascertainment populations for conversion to the 5′-nuclease reaction (Holland et al. 1991) with TaqMan chemistry (Life Technologies, Grand Island, New York), a chemistry commonly used on high-throughput genotyping platforms (Seeb et al. 2009*a*). Genetic differentiation among our ascertainment populations was estimated across all RAD-derived loci with overall F_{ST} values (Weir and Cockerham 1984) calculated in GENEPOP. We also calculated pairwise F_{ST} values and conducted exact tests for deviations from Hardy–Weinberg equilibrium for each locus in GENEPOP.

The 150 SNPs were chosen in an iterative fashion. First we choose the 150 SNPs with the highest overall F_{ST} across the Bristol Bay (Koktuli River), Kuskokwim River (Kogrukluk River), and lower Yukon River (Anvik River) populations and tested this panel's assignment power with 100% simulations conducted in the program ONCOR (see below for further details). We then modified the panel by adding and removing SNPs until we were able to find the 150 SNPs that achieved the highest possible assignment accuracies for all ascertainment populations. We did not choose SNPs that differentiated the upper Yukon River (Big Salmon River) or Norton Sound (Tubutulik River) populations because these populations were highly differentiated from all others and could likely be resolved with any SNP panel (Table 2).

We limited our selections for conversion to the 5′-nuclease reaction to SNPs that were in Hardy–Weinberg equilibrium in at least three of the five populations ($P > 0.05$). Also, we chose SNPs that were located past base pair 34 of the RAD tag to accommodate the primer–probe configuration of the 5′-nuclease reaction. Paired-end data were used to increase template length for assay design if no primer–probe configuration was feasible with the single-end reads.

Successfully designed assays were genotyped on 24 fish from each of the four ascertainment populations from coastal western Alaska (populations 2, 6, 16, and 22; 96 fish total). Genotyping was conducted with preamplification according to the methods of Smith et al. (2011). Assays that did not amplify or produce consistent results were discarded, and the 96 assays with the highest overall F_{ST} across the Bristol Bay, Kuskokwim River, and lower Yukon River populations based on the RAD data were retained to form a panel of 96 RAD-derived SNPs, hereafter referred to as the RAD96.

Selection and evaluation of SNP panels

Two major goals of this study were (*i*) to evaluate the resolving power of the RAD96 compared with a panel of previously developed SNPs and (*ii*) to construct the best possible panel of 96 SNPs to discriminate stocks in western Alaska from all SNPs available. To achieve these goals, we genotyped 2275 fish from 28 populations throughout western Alaska for the RAD96 and 191 SNPs previously developed for Chinook salmon. The 191 previously developed SNPs (hereafter referred to as CTC191) were mainly chosen for applications south of Alaska as part of a project funded by the Pacific Salmon Commission's Chinook Technical Committee (Warheit et al. 2013); these originated primarily from expressed sequence tags (Smith et al. 2005*a*, 2005*b*, 2007; Campbell and Narum 2008, 2009; Clemento et al. 2011; Warheit et al. 2013).

SNPs from the CTC191 and RAD96 were genotyped using the 5′-nuclease reaction with pre-amplification (Smith et al. 2011), and the reproducibility of our results was quantified by regenotyping four of every 95 (4%) fish at all loci. Individuals with >5% missing genotypes were excluded from further analyses. Tests for deviation from Hardy–Weinberg and linkage equilibrium were conducted for each locus across all 28 populations in GENEPOP, and loci out of equilibrium in >50% of the populations ($P < 0.05$) were removed. Observed and expected heterozygosities for each locus were calculated in GenAlEx 6.5 (Peakall and Smouse 2012) and overall F_{ST} (Weir and Cockerham 1984) for each locus was calculated in GENEPOP. Calculations of locus-specific heterozygosity and F_{ST} were conducted using populations 1–24 (excluding populations 2 and 25–28, see below for justification).

Genetic differentiation across all 28 populations was estimated separately for the CTC191 and RAD96 panels with pairwise F_{ST} values (Weir and Cockerham 1984) calculated in GENEPOP to compare the patterns of population structure resolved by each panel. We then conducted principal coordinate analysis (PCoA) in GenAlEx for each panel to visualize patterns of population structure. Populations 25–28 from the middle and upper Yukon River were not included in the PCoA because these populations are highly differentiated from those of coastal western Alaska according to previous studies (Smith et al. 2005*c*; Templin et al. 2011). Including these populations may have prevented us from detecting signals of differentiation among the remaining populations.

After comparing the CTC191 and RAD96 panels, all SNPs were ranked by overall F_{ST} across populations 1–24 (excluding population 2), and the 96 SNPs with the highest F_{ST} were chosen to create the final panel for GSI, referred to hereafter as the F_{ST}96. The Tubutulik River (population 2) was excluded from this analysis because it was a genetic outlier (Fig. 2). SNPs were ranked by F_{ST} because this method produced the highest performing panels for population assignment in recent analyses of multiple ranking methods (Storer et al. 2012; Warheit et al. 2013).

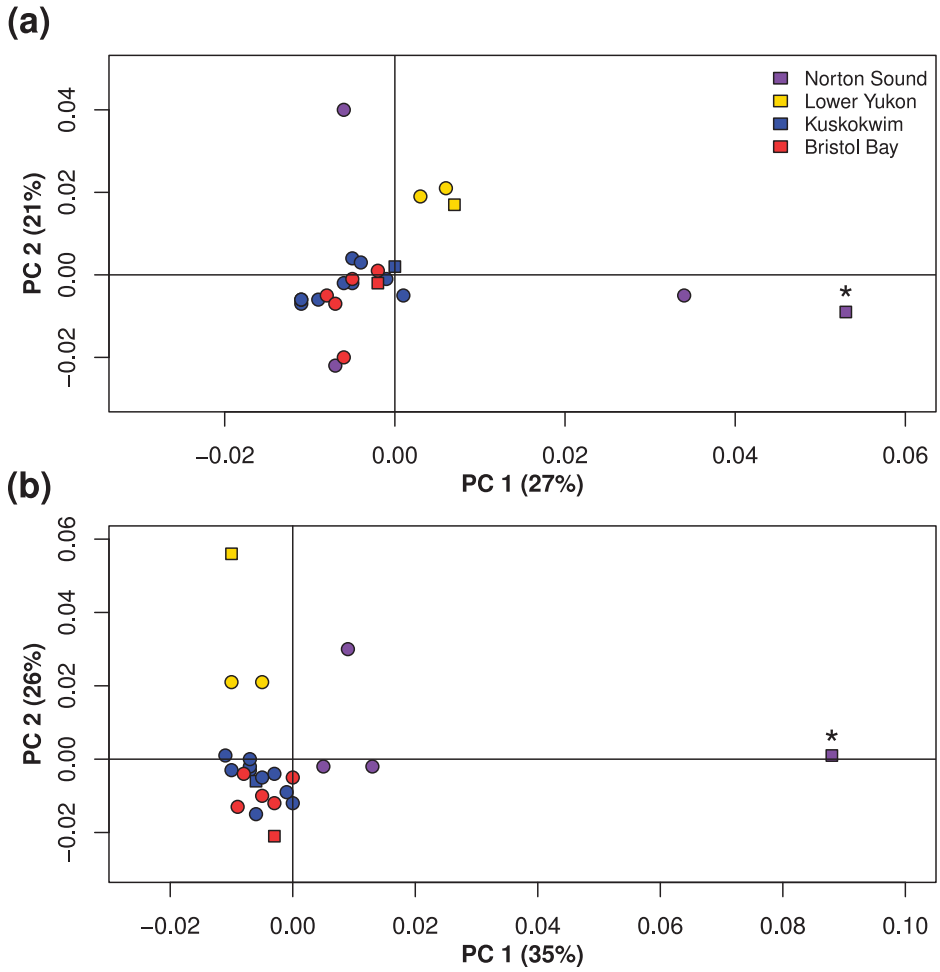
The assignment accuracy of the complete dataset (281 SNPs), the F_{ST}96, the CTC191, the RAD96, and 96 randomly chosen SNPs from the complete dataset was evaluated with the 100% simulation method described in Anderson et al. (2008) and implemented in ONCOR (http://www.montana.edu/kalinowski/Software.htm) with the default parameters. The simulation method implemented in ONCOR simulates a mixture sample where all individuals are from the same population and then uses maximum likelihood to determine the percentage of the sample that is correctly allocated back to the population and reporting group of origin. A minimum value of 90% correct assignment is typically required for a reporting group to be considered identifiable and robust for management applications (Seeb et al. 2000). Regional aggregations (reporting groups) for this analysis were Norton Sound, lower Yukon River, Bristol Bay – Kuskokwim River, middle Yukon River, and upper Yukon River (Table 1). These groups were similar to the fine-scale reporting groups presented in Templin et al. (2011) with one exception; populations from Bristol Bay and the Kuskokwim River regions were combined into one reporting group because preliminary assignment tests were generally unable to differentiate these two regions. Assignment success between panels was compared with a Student's *t* test. The small sample sizes for the Norton Sound and lower Yukon River collections prevented dividing datasets into separate training and holdout sets as suggested by Anderson (2010).

Results

RAD sequencing and SNP discovery

Sequence data from 284 Chinook salmon available from Larson et al. (2014) were used to discover 26 567 putative SNPs. Filtration steps eliminated 1602 potential PSVs and 13 115 loci with low minor allele frequencies and genotyping rates. Seventeen individu-

Fig. 2. Principal coordinate analysis (PCoA) of 24 populations from coastal western Alaska with (*a*) CTC191 and (*b*) RAD96. Only SNPs that were in linkage and Hardy–Weinberg equilibrium were used in this analysis. The PCoA is based on pairwise F_{ST} values. Squares are ascertainment populations that were RAD-sequenced. Population 2 (Tubutulik River) is labeled with an asterisk (*) because it was a genetic outlier and was removed from some analyses (see text).



als with <10× coverage across all filtered SNPs were removed (adjusted sample sizes in Table 1). The final filtered dataset consisted of 267 individuals genotyped at 11 850 SNPs. The mean depth of coverage across these individuals for the filtered SNPs was 29.1 (range 10.5–70.6).

Paired-end assembly and BLAST annotation

Paired-end assemblies produced 12 016 contigs with a mean length of 268 bp (minimum 150 bp, maximum 565 bp). BLAST annotation of these contigs yielded significant hits for 1466 of 11 850 SNPs, representing a 12% success rate. Of these hits, 547 (37%) aligned to transposable elements. Other common functional groups included DNA polymerases and structural proteins (see online supplementary material, Table S1[1]).

Construction of high-throughput assays from RAD data

Assay design for the 5′-nuclease reaction was successful for 128 of the 150 assays attempted (Table 3). These 128 assays were tested in 96 fish, and 101 of them successfully amplified. The top 96 assays were retained to form the RAD96 panel (see Materials and methods). Paired-end data were required to design 47 of the 96 assays (49%), and BLAST annotations were successful for 9 of 96 assays (9%; see Table S2[1] for primer and probe sequences and BLAST annotations for the RAD96). A comparison of genotypes derived from RAD and 5′-nuclease data revealed 99% concordance between chemistries (Table 4). The most common type of error was a heterozygous 5′-nuclease genotype that was called a homozygote from RAD data, an expected result for data from next-generation genotyping (Nielsen et al. 2011).

Table 3. Number of SNPs at each stage of SNP discovery.

Dataset	No. of putative SNPs
Unfiltered RAD	26 567
Filtered RAD	11 850
5′-nuclease assays attempted	150
5′-nuclease assays designed	128
5′-nuclease assays validated	101
Top 96 assays	96

Note: Validated 5′-nuclease assays are those that successfully amplified and produced clean scatter plots.

Selection and evaluation of SNP panels

Our genotyping success rate for the 5′-nuclease reaction was 97% (2275 of 2355 samples), and our genotyping discrepancy rate,

[1]Supplementary data are available with the article through the journal Web site at http://nrcresearchpress.com/doi/suppl/10.1139/cjfas-2013-0502.

Table 4. Number of discrepancies between 5′-nuclease and RAD genotypes across 254 individuals that were genotyped for both chemistries.

5'-nuclease genotype	RAD genotype	Number	Proportion
Concordance	—	23 955	0.990
Discrepancies			
AA	BB	0	0.000
AA or BB	AB	62	0.003
AB	AA or BB	182	0.007
Total discrepancies		244	0.010

Note: The table is based on a bi-allelic locus with allele one designated by A and allele two designated by B.

calculated from regenotyping 4% of samples, was 0.03%. Four locus pairs were significantly out of linkage equilibrium in greater than half of the populations ($P < 0.05$). These marker pairs were *Ots_FGF6A* and *Ots_FGF6B_1* (28/28 populations), *Ots_RAD8200-45* and *Ots_RAD9480-51* (28/28 populations), *Ots_HSP90B-100* and *Ots_HSP90B-385* (24/28 populations), and *Ots_RAD11821* and *Ots_RAD3703* (16/28 populations). The marker with the highest F_{ST} for each pair was retained, resulting in the removal of *Ots_FGF6A*, *Ots_RAD9480-51*, *Ots_HSP90B-100*, and *Ots_RAD11821* from further analyses. Significant deviations from Hardy–Weinberg equilibrium ($P < 0.05$) in more than half of the populations occurred for two loci, *Ots_111084b-619* (28/28 populations) and *Ots_111666-408* (28/28 populations); these loci were removed from further analyses. After removing SNPs that were out of Hardy–Weinberg and linkage equilibrium, 186 SNPs were retained from the CTC191 and 95 were retained from the RAD96 (see Table S3[1] for summary statistics for each locus).

Patterns of population structure were similar between the CTC191 and RAD96 panels, with populations from the Bristol Bay and Kuskokwim River regions forming a discrete cluster and populations from the lower Yukon River forming another cluster (Fig. 2). Populations from Norton Sound, however, did not form a single cluster and were generally distinct from all other populations. Populations from the middle and upper Yukon River (not shown in Fig. 2) were extremely differentiated from those of coastal western Alaska with both panels and displayed pairwise F_{ST} values that were at least two times larger than any within coastal western Alaska comparison (Tables S4, S5[1]). Although the CTC191 and RAD96 panels showed similar patterns of population structure, the mean F_{ST} and H_O were significantly higher for markers in the RAD96 compared with the CTC191 (CTC191: $H_O = 0.24$, $F_{ST} = 0.006$; RAD96: $H_O = 0.34$, $F_{ST} = 0.008$; $P < 0.0001$ for both Student's t tests; Fig. 3).

After evaluating the CTC191 and RAD96 panels separately, we ranked all SNPs by overall F_{ST} across populations 1–24 (excluding population 2, F_{ST} ranks in Table S3[1]). We then choose the top 96 to form the F_{ST}96 panel: 49 SNPs from the CTC191 and 47 SNPs from the RAD96. The F_{ST}96 panel was composed of 49% RAD-derived SNPs, while RAD-derived SNPs composed 33% of the full dataset.

Assignment accuracies calculated with GSI simulations in ONCOR varied across panels but were generally highest with the F_{ST}96 and the complete dataset (Fig. 4; Table S6[1]). The F_{ST}96 panel produced assignment accuracies >90% to reporting group for 26 of 28 populations (88% for population 1, 89% for population 7), and the complete dataset produced accuracies >90% for 25 of 28 populations (89% for population 1, 88% for population 5, 87% for population 7), while all other panels produced accuracies >90% for fewer than 24 populations. Additionally, the F_{ST}96 and the full dataset significantly outperformed the panel of 96 randomly chosen SNPs, the CTC191, and the RAD96 ($P < 0.05$; Fig. 4; Table S6[1]). Assignment rates were slightly higher for the complete dataset compared with the F_{ST}96 panel ($P = 0.04$), but the F_{ST}96 panel did outperform the complete dataset in three populations (3, 5, and 7). The CTC191 and RAD96 panels performed similarly ($P = 0.29$) despite the fact that the CTC191 panel contained almost twice as many SNPs.

Discussion

RAD sequencing for SNP development

We efficiently developed 96 novel high-throughput assays for GSI in western Alaska using data from RAD sequencing. Compared with previous methods for SNP discovery in Pacific salmon, mining RAD sequence data was quicker, required fewer validation steps, and facilitated directed SNP discovery for markers showing high levels of differentiation among populations. Mining RAD sequence data was much less time-consuming than methods mining expressed sequence tag (EST) databases for putative SNPs (e.g., Smith et al. 2005*b*) and achieved an approximately 30% higher conversion rate to the 5′-nuclease reaction (Smith et al. 2005*a*; Amish et al. 2012). This approach also represented a major improvement over transcriptome-based methods, which require multiple validation steps and still achieve a conversion rate to the 5′-nuclease reaction of less than 50% (Everett et al. 2011; Seeb et al. 2011*b*). Additionally, the discrepancy rate between genotypes obtained from RAD and 5′-nuclease data was extremely low (1%).

Population structure

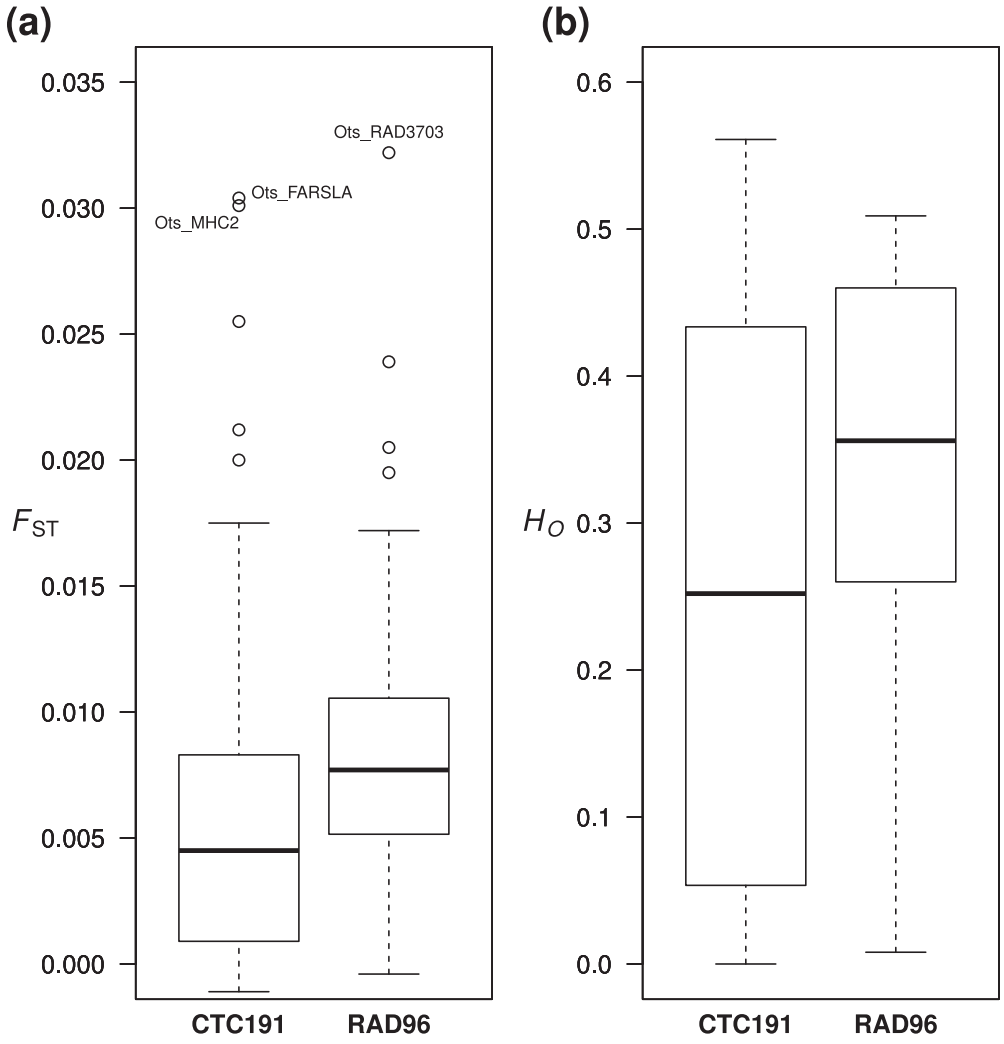
General patterns of population structure in western Alaska were similar among the 11 850 RAD SNPs, the CTC191, and the RAD96 and are consistent with results from previous studies in the region (Olsen et al. 2011; Templin et al. 2011). The largest differentiation in all three datasets existed between populations from coastal western Alaska (populations 1–24) and those from the middle and upper Yukon River (populations 25–28; Tables S4, S5[1]). This pattern has been documented in numerous studies (e.g., Gharrett et al. 1987; Smith et al. 2005*c*; Beacham et al. 2008*a*) and is consistent with isolation during the last glacial maximum (Olsen et al. 2011). Within coastal western Alaska, populations from Norton Sound and the lower Yukon River displayed the highest levels of differentiation, while populations from the Bristol Bay and Kuskokwim River regions appeared to be closely related. It is likely that the observed structure is the result of genetic drift in the lower Yukon River and Norton Sound facilitated by relatively small census sizes. Populations in the lower Yukon River and Norton Sound regions generally contain fewer than 2000 spawners, whereas many populations in the Bristol Bay and Kuskokwim River regions contain greater than 10 000 spawners (Molyneaux and Dubois 1999; Baker et al. 2006; Banducci et al. 2007; Howard et al. 2009). Different levels of effective migration within regions may also influence this pattern.

Comparison of panels for GSI

Previous studies demonstrate that the level of polymorphism (H_O) and differentiation (F_{ST}) of SNPs is positively correlated with their value for GSI (Ackerman et al. 2011; Bradbury et al. 2011; Storer et al. 2012). We observed a significantly higher mean H_O and F_{ST} for the RAD96 panel compared with the CTC191 panel, and a higher proportion of SNPs from the RAD96 were chosen for the final F_{ST}96 panel. These results indicate that, on average, the SNPs in the RAD96 panel are likely to be more useful for GSI in populations from western Alaska than the SNPs in the CTC191 panel.

Assignment accuracies for all populations with both the full dataset and the F_{ST}96 panel were close to or above the 90% threshold necessary for management applications (Seeb et al. 2000). Assignment rates were lower for the CTC191 and RAD96, implying that GSI with our reporting groups would be less powerful with only one of these panels. Although both the CTC191 and RAD96 panels displayed similar assignment accuracies overall, there were major differences between the two panels for specific populations in the Norton Sound and lower Yukon River regions (e.g., populations 1, 3, 6). These differences demonstrate the impor-

Fig. 3. Box and whisker plots of locus-specific overall (*a*) F_{ST} and (*b*) H_O for two SNP datasets. Datasets are CTC191 (mean H_O = 0.24, mean F_{ST} = 0.006) and RAD96 (mean H_O = 0.34, mean F_{ST} = 0.008). Only SNPs that were in linkage and Hardy–Weinberg equilibrium were used in this analysis. A Student's *t* test indicated that the two datasets have significantly different distributions of H_O and F_{ST} ($P < 0.0001$). Loci with F_{ST} values above 0.03 are labeled in panel (*a*). Each dataset includes populations 1–24 (excluding population 2, see text).



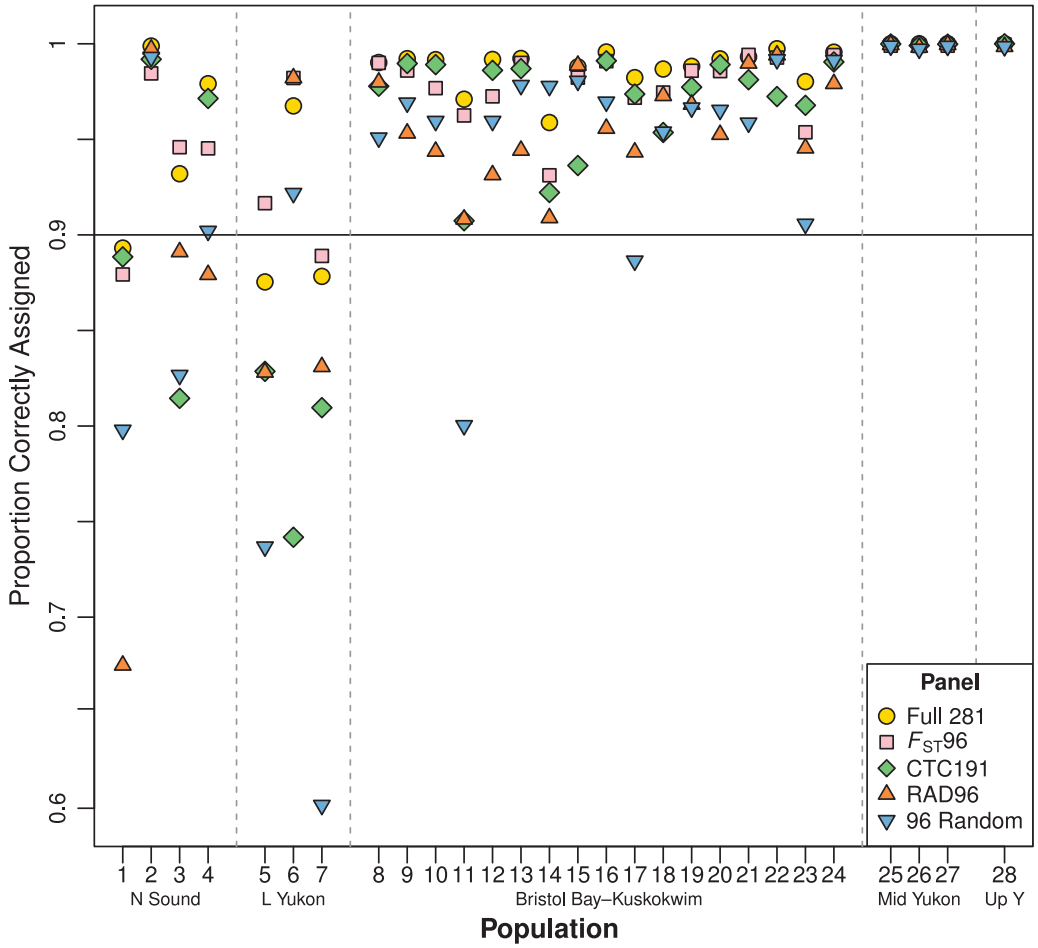
tance of obtaining a representative set of ascertainment populations when attempting to create a SNP panel for GSI.

SNP discovery and evaluation conducted in this study has increased the number of feasible reporting groups for GSI in western Alaska from one to three, but accuracy could be further improved by sampling additional populations from the lower Yukon River and Norton Sound regions. It is especially important to sample throughout Norton Sound because these populations were each genetically distinct from each other and all others in the study. Norton Sound is composed of many small, unconnected rivers with census sizes that are often under 1000 (Banducci et al. 2007). These populations may have quickly diverged from each other because of greater genetic drift in small populations and (or) regional landscape features restricting gene flow. Dense sampling is therefore necessary to accurately characterize genetic variation in this region. Any additional populations could also be used as a holdout set to assess the assignment accuracy of our panels as suggested by Anderson (2010).

Ascertainment bias

Both the CTC191 and RAD96 panels exhibited similar patterns of population structure but also displayed evidence of ascertainment bias. Ascertainment bias occurs when genetic markers are chosen such that they are unrepresentative of genetic variation in all populations or regions of interest (Smith et al. 2007). Ascertainment bias can distort estimates of population structure but can also increase assignment power in the region of interest (Bradbury et al. 2011). Two major sources of ascertainment bias were present in our data: (*i*) regional ascertainment bias in the CTC191 and RAD96 panels and (*ii*) population-specific bias in the populations that were RAD-sequenced. The regional ascertainment bias in the CTC191 and RAD96 panels occurred because these two panels were largely created for regional applications (CTC191: south of Alaska; RAD96: western Alaska), a common occurrence with SNP panels developed for salmonid management (e.g., Seeb et al. 2011*c*). In this case, the regional bias in the RAD96 is helpful because it likely increases our power to differentiate populations in western Alaska. However, this bias may also decrease the power of these SNPs to differentiate populations outside the region of interest, possibly reducing the utility of the RAD96 across the species range (Smith et al. 2007). Assignment accuracies from the middle and upper Yukon River populations suggest that the SNPs developed in this study should be useful outside of the ascertainment area, but further testing is needed to fully validate this assumption.

Fig. 4. Assignment probabilities to reporting group for the full dataset of 281 SNPs (Full 281), the 96 SNPs with the highest overall F_{ST} (F_{ST}96), the CTC191, the RAD96, and 96 randomly chosen SNPs (96 Random). Only SNPs that were in linkage and Hardy–Weinberg equilibrium were used in this analysis. Population numbers correspond to those in Table 1. Reporting groups (*X* axis) are separated by gray dashed lines. Abbreviations are Norton Sound (N Sound), lower Yukon River (L Yukon), Bristol Bay and Kuskokwim River (Bristol Bay–Kuskokwim), middle Yukon River (Mid Yukon), and upper Yukon River (Up Y). Bristol Bay and Kuskokwim River populations were combined into a single reporting group for this analysis (see text). The line at 0.9 represents a common value used to consider an assignment robust for management applications (Seeb et al. 2000). Confidence intervals for each assignment probability are reported in Table S6[1].



The second type of ascertainment bias present in our data was population-specific bias in the populations that were RAD-sequenced. This bias likely occurred because dozens of SNPs showing high differentiation were chosen from thousands, causing the populations that were RAD-sequenced to appear more differentiated than expected. This type of bias was especially apparent in the Anvik River (population 6), which clustered tightly with the two other lower Yukon River populations using the CTC191 but was highly diverged with the RAD96.

Population-specific ascertainment bias could lead to upwardly biased estimates of assignment accuracy and could distort phylogenetic relationships among populations. To reduce unwanted population-specific bias, we suggest that future studies with similar objectives sequence at least two ascertainment populations from each drainage or region. Hierarchical *F* statistics could then be used to discover SNPs that are similar within but divergent among regions. For example, SNPs with high values of F_{CT} (variation among reporting groups) and small values of F_{SC} (variation among populations within reporting groups) could be chosen.

Use of adaptively important markers for GSI

The accuracy of GSI in poorly differentiated populations can often be improved by including adaptively important markers that are undergoing divergent natural selection (Nielsen et al. 2012). For example, Ackerman et al. (2011) found that the addition of adaptively important markers to a panel of neutral markers significantly improved assignment accuracy, and Russello et al. (2012) showed that assignment accuracies were much higher with a panel of adaptively important markers compared with a panel of neutral markers. Multiple studies using RAD sequencing have found signatures of natural selection (e.g., Hohenlohe et al. 2010; Gagnaire et al. 2013), but strong signatures of selection were not apparent in our data. Specifically, patterns of population structure were similar with the RAD96 and the primarily neutral CTC191, and we observed relatively small F_{ST} values across most loci, indicating that the majority of loci in our dataset were probably neutral. It is interesting to note that one locus from the CTC191, *Ots_MHC2*, had an overall F_{ST} of 0.431 in Chinook salmon from the Copper River and was found to be under strong divergent selection in this environment (Seeb et al. 2009*b*; Ackerman et al. 2013). *Ots_MHC2* also had one of the highest overall F_{ST} values in our study, indicating that it may be adaptively important in western Alaska. Future studies attempting to improve GSI in our study region would likely benefit from the inclusion of additional adaptively important markers such as *Ots_MHC2*. For example, adaptively important markers might be useful for differentiating popula-

tions from the Kuskokwim River and Bristol Bay regions, something that was not possible with our current set of SNPs.

Management applications

The precipitous decline of Chinook salmon in western Alaska has prompted multiple fisheries closures, causing extensive economic hardship and threatening subsistence catches for natives of the western Alaska region (ADF&G 2013). Increased resolution for GSI facilitated by our study has the potential to substantially improve fisheries management in this region. Specifically, GSI can be used to monitor the contribution of different stocks in mixed-stock fisheries, informing fisheries management and preventing unnecessary fishery closures (Shaklee et al. 1999; Smith et al. 2005*c*; Dann et al. 2013). Additionally, SNPs developed in this study can be used to improve resolution in studies of migration and distribution patterns of Chinook salmon on the high seas (c.f., Tucker et al. 2009; Guthrie et al. 2013; Larson et al. 2013). The ability to measure stock-specific abundance on the high seas can provide important information for stock assessment models that is currently unavailable.

Conclusions

We increased the number of feasible reporting groups for GSI in coastal western Alaska from one to three using directed SNP discovery. The SNPs we developed from RAD data displayed higher levels of polymorphism and differentiation compared with many previously developed SNPs and were more useful for GSI. RAD sequence data therefore provided an excellent tool for discovering high-resolution SNPs that can differentiate closely related populations. The increased resolution for GSI in coastal western Alaska provided by this study will facilitate research into migration patterns and vulnerability to fisheries of Chinook salmon in this region, aiding in the conservation of an extremely important economic and cultural resource.

Data accessibility

Illumina RAD-tag sequences have been submitted to the NCBI SRA database (project accession ID: SRP034950). Genotype data for the 287 5′-nuclease assays and the 11 850 SNPs genotyped using RAD as well as sequences for the 11 850 RAD tags are available as supplementary material in addition to Tables S1–S6[1].

Acknowledgements

We thank Nick DeCovich and Andrew Barclay from the Alaska Department of Fish and Game for providing the samples for this project and providing valuable local knowledge. We also thank Ryan Waples, University of Washington, for assistance with bioinformatics aspects of this project. Additionally, we thank Kristen Gruenthal and Fred Utter for commentary on this manuscript. This research was partially funded by the Alaska Sustainable Salmon Fund under study Nos. 44515 and 44522 from NOAA, US Department of Commerce, administered by the ADF&G. Funding was also provided by a grant from the Gordon and Betty Moore Foundation. Additional support for WAL was provided by the H. Mason Keeler Endowment for Excellence and a National Science Foundation Graduate Research Fellowship (grant No. DGE-0718124).

References

Ackerman, M.W., Habicht, C., and Seeb, L.W. 2011. Single-nucleotide polymorphisms (SNPs) under diversifying selection provide increased accuracy and precision in mixed-stock analyses of sockeye salmon from the Copper River, Alaska. Trans. Am. Fish. Soc. **140**(3): 865–881. doi:10.1080/00028487.2011.588137.

Ackerman, M.W., Templin, W.D., Seeb, J.E., and Seeb, L.W. 2013. Landscape heterogeneity and local adaptation define the spatial genetic structure of Pacific salmon in a pristine environment. Conserv. Genet. **14**(2): 483–498. doi:10.1007/s10592-012-0401-7.

ADF&G. 2013. Chinook salmon stock assessment and research plan [online]. Alaska Department of Fish and Game, Special Publication No. 13-01. Available from http://www.adfg.alaska.gov/FedAidPDFs/SP13-01.pdf [accessed 7 March 2013].

Allendorf, F., and Thorgaard, G.H. 1984. Polyploidy and the evolution of salmonid fishes *In* The evolutionary genetics of fishes. *Edited by* B.J. Turner. Plenum Press, New York. pp. 1–53.

Allendorf, F.W., Hohenlohe, P.A., and Luikart, G. 2010. Genomics and the future of conservation genetics. Nat. Rev. Genet. **11**(10): 697–709. doi:10.1038/nrg2844. PMID:20847747.

Amish, S.J., Hohenlohe, P.A., Painter, S., Leary, R.F., Muhlfeld, C., Allendorf, F.W., and Luikart, G. 2012. RAD sequencing yields a high success rate for westslope cutthroat and rainbow trout species — diagnostic SNP assays. Mol. Ecol. Resour. **12**(4): 653–660. doi:10.1111/j.1755-0998.2012.03157.x. PMID:22672623.

Anderson, E.C. 2010. Assessing the power of informative subsets of loci for population assignment: standard methods are upwardly biased. Mol. Ecol. Resour. **10**(4): 701–710. doi:10.1111/j.1755-0998.2010.02846.x. PMID:21565075.

Anderson, E.C., Waples, R.S., and Kalinowski, S.T. 2008. An improved method for predicting the accuracy of genetic stock identification. Can. J. Fish. Aquat. Sci. **65**(7): 1475–1486. doi:10.1139/F08-049.

Baker, T.T., Fair, L.F., Clark, R.A., and Hasbrouck, J.J. 2006. Review of salmon escapement goals in Bristol Bay, Alaska, 2006 [online]. Alaska Department of Fish and Game, Fishery Manuscript No. 06-05. Available from http://www.sf.adfg.state.ak.us/FedAidPDFs/fms06-05.pdf [accessed 6 May 2012].

Banducci, A., Kohler, T., Soong, J., and Menard, J. 2007. 2005 Annual management report for Norton Sound, Port Clarence, and Kotzebue [online]. Alaska Department of Fish and Game, Fishery Management Report No. 07-32. http://www.adfg.alaska.gov/FedAidPDFs/fmr07-32.pdf [accessed 30 January 2013].

Beacham, T.D., Wetklo, M., Wallace, C., Olsen, J.B., Flannery, B.G., Wenburg, J.K., Templin, W.D., Antonovich, A., and Seeb, L.W. 2008*a*. The application of microsatellites for stock identification of Yukon River Chinook salmon. N. Am. J. Fish. Manage. **28**(1): 283–295. doi:10.1577/M06-253.1.

Beacham, T.D., Winther, I., Jonsen, K.L., Wetklo, M., Deng, L., and Candy, J.R. 2008*b*. The application of rapid microsatellite-based stock identification to management of a Chinook salmon troll fishery off the Queen Charlotte Islands, British Columbia. N. Am. J. Fish. Manage. **28**(3): 849–855. doi:10.1577/M06-167.1.

Beacham, T.D., Candy, J.R., Wallace, C., Wetklo, M., Deng, L., and MacConnachie, C. 2012. Microsatellite mixed-stock identification of coho salmon in British Columbia. Mar. Coast. Fish. **4**(1): 85–100. doi:10.1080/19425120.2012.661393.

Bradbury, I.R., Hubert, S., Higgins, B., Bowman, S., Paterson, I.G., Snelgrove, P.V.R., Morris, C.J., Gregory, R.S., Hardie, D.C., Borza, T., and Bentzen, P. 2011. Evaluating SNP ascertainment bias and its impact on population assignment in Atlantic cod, *Gadus morhua*. Mol. Ecol. Resour. **11**: 218–225. doi:10.1111/j.1755-0998.2010.02949.x. PMID:21429176.

Campbell, N.R., and Narum, S.R. 2008. Identification of novel single-nucleotide polymorphisms in Chinook salmon and variation among life history types. Trans. Am. Fish. Soc. **137**(1): 96–106. doi:10.1577/T07-011.1.

Campbell, N.R., and Narum, S.R. 2009. Identification and characterization of heat shock response related single nucleotide polymorphisms in *O. mykiss* and *O. tshawytscha*. Mol. Ecol. Resour. **9**: 1450–1559.

Catchen, J.M., Amores, A., Hohenlohe, P., Cresko, W., and Postlethwait, J.H. 2011. Stacks: building and genotyping loci *de novo* from short-read sequences. G3: Genes Genomes Genet. **1**: 171–182. doi:10.1534/g3.111.000240.

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., and Cresko, W.A. 2013. Stacks: an analysis tool set for population genomics. Mol. Ecol. **22**(11): 3124–3140. doi:10.1111/mec.12354. PMID:23701397.

Clemento, A.J., Abadía-Cardoso, A., Starks, H.A., and Garza, J.C. 2011. Discovery and characterization of single nucleotide polymorphisms in Chinook salmon, *Oncorhynchus tshawytscha*. Mol. Ecol. Resour. **11**: 50–66. doi:10.1111/j.1755-0998.2010.02972.x. PMID:21429162.

Dann, T.H., Habicht, C., Baker, T.T., and Seeb, J.E. 2013. Exploiting genetic diversity to balance conservation and harvest of migratory salmon. Can. J. Fish. Aquat. Sci. **70**(5): 785–793. doi:10.1139/cjfas-2012-0449.

Etter, P.D., Preston, J.L., Bassham, S., Cresko, W.A., and Johnson, E.A. 2011. Local *de novo* assembly of RAD paired-end contigs using short sequencing reads. PLoS ONE, **6**(4): e18561. doi:10.1371/journal.pone.0018561. PMID:21541009.

Everett, M.V., and Seeb, J.E. 2014. Detection and mapping of QTL for temperature tolerance and body size in Chinook salmon (*Oncorhynchus tshawytscha*) using genotyping by sequencing. Evol. Appl. [In press.] doi:10.1111/eva.12147.

Everett, M.V., Grau, E.D., and Seeb, J.E. 2011. Short reads and nonmodel species: exploring the complexities of next-generation sequence assembly and SNP discovery in the absence of a reference genome. Mol. Ecol. Resour. **11**: 93–108. doi:10.1111/j.1755-0998.2010.02969.x. PMID:21429166.

Everett, M.V., Miller, M.R., and Seeb, J.E. 2012. Meiotic maps of sockeye salmon derived from massively parallel DNA sequencing. BMC Genomics, **13**: 521. doi:10.1186/1471-2164-13-521. PMID:23031582.

Gagnaire, P.-A., Pavey, S.A., Normandeau, E., and Bernatchez, L. 2013. The genetic architecture of reproductive isolation during speciation with gene flow in lake whitefish assessed by RAD sequencing. Evolution, **67**(9): 2483–2497. doi:10.1111/evo.12075. PMID:24033162.

Gharrett, A.J., Shirley, S.M., and Tromble, G.R. 1987. Genetic relationships among populations of Alaskan chinook salmon (*Oncorhynchus tshawytscha*). Can. J. Fish. Aquat. Sci. **44**(4): 765–774. doi:10.1139/f87-093.

Gidskehaug, L., Kent, M., Hayes, B.J., and Lien, S. 2011. Genotype calling and

mapping of multisite variants using an Atlantic salmon iSelect SNP array. Bioinformatics, **27**(3): 303–310. doi:10.1093/bioinformatics/btq673. PMID: 21149341.

Guthrie, C.M., Nguyen, H.T., and Guyon, J.R. 2013. Genetic stock composition analysis of Chinook salmon bycatch samples from the 2011 Bering Sea and Gulf of Alaska trawl fisheries. US Dept. Commer., NOAA Tech. Memo. NMFS-AFSC-244.

Habicht, C., Seeb, L.W., Myers, K.W., Farley, E.V., and Seeb, J.E. 2010. Summer–fall distribution of stocks of immature sockeye salmon in the Bering Sea as revealed by single-nucleotide polymorphisms. Trans. Am. Fish. Soc. **139**(4): 1171–1191. doi:10.1577/T09-149.1.

Hauser, L., and Seeb, J.E. 2008. Advances in molecular technology and their impact on fisheries genetics. Fish. Fish. **9**(4): 473–486. doi:10.1111/j.1467-2979.2008.00306.x.

Hecht, B.C., Campbell, N.R., Holecek, D.E., and Narum, S.R. 2013. Genome-wide association reveals genetic basis for the propensity to migrate in wild populations of rainbow and steelhead trout. Mol. Ecol. **22**(11): 3061–3076. doi:10.1111/mec.12082. PMID:23106605.

Hilborn, R., Quinn, T.P., Schindler, D.E., and Rogers, D.E. 2003. Biocomplexity and fisheries sustainability. Proc. Natl. Acad. Sci. U.S.A. **100**(11): 6564–6568. doi:10.1073/pnas.1037274100. PMID:12743372.

Hohenlohe, P.A., Bassham, S., Etter, P.D., Stiffler, N., Johnson, E.A., and Cresko, W.A. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet. **6**(2): e1000862. doi:10.1371/journal.pgen.1000862. PMID:20195501.

Hohenlohe, P.A., Amish, S.J., Catchen, J.M., Allendorf, F.W., and Luikart, G. 2011. Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. Mol. Ecol. Resour. **11**: 117–122. doi:10.1111/j.1755-0998.2010.02967.x. PMID:21429168.

Holland, P.M., Abramson, R.D., Watson, R., and Gelfand, D.H. 1991. Detection of specific polymerase chain-reaction product by the 5′–3′ exonuclease activity of *Thermus aquaticus* DNA polymerase. Proc. Natl. Acad. Sci. **88**(16): 7276–7280. doi:10.1073/pnas.88.16.7276. PMID:1871133.

Howard, K.G., Hayes, S.J., and Evenson, D.F. 2009. Yukon River Chinook salmon stock status and action plan 2010; a report to the Alaska Board of Fisheries [online]. Alaska Department of Fish and Game, Special Publication No. 09-26. Available from http://www.sf.adfg.state.ak.us/FedAidpdfs/Sp09-26.pdf [accessed 6 May 2012].

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. **10**(3): R25. doi:10.1186/gb-2009-10-3-r25.

Larson, W.A., Utter, F.M., Myers, K.W., Templin, W.D., Seeb, J.E., Guthrie, C.M., III, Bugaev, A.V., and Seeb, L.W. 2013. Single-nucleotide polymorphisms reveal distribution and migration of Chinook salmon (*Oncorhynchus tshawytscha*) in the Bering Sea and North Pacific Ocean. Can. J. Fish. Aquat. Sci. **70**(1): 128–141. doi:10.1139/cjfas-2012-0233.

Larson, W., Seeb, L.W., Everett, M.V., Waples, R.K., Templin, W.D., and Seeb, J.E. 2014. Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (*Oncorhynchus tshawytscha*). Evol. Appl. [online ahead of print.] doi:10.1111/eva.12128.

Miller, M.R., Brunelli, J.P., Wheeler, P.A., Liu, S., Rexroad, C.E., III, Palti, Y., Doe, C.Q., and Thorgaard, G.H. 2012. A conserved haplotype controls parallel adaptation in geographically distant salmonid populations. Mol. Ecol. **21**(2): 237–249. doi:10.1111/j.1365-294X.2011.05305.x. PMID:21988725.

Milner, G.B., Teel, D.J., Utter, F.M., and Winans, G.A. 1985. A genetic method of stock identification in mixed populations of Pacific Salmon, *Oncorhynchus* spp. Mar. Fish. Rev. **47**(1): 1–8.

Minoche, A.E., Dohm, J.C., and Himmelbauer, H. 2011. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and Genome Analyzer systems. Genome Biol. **12**(11): R112. doi:10.1186/gb-2011-12-11-r112.

Molyneaux, D.B., and Dubois, L. 1999. Salmon age, sex and length catalog for the Kuskokwim area, 1998 progress report [online]. Alaska Department of Fish and Game, Regional Information Report No. 3A99-15. Available from http://www.sf.adfg.state.ak.us/FedAidpdfs/RIR.3A.1999.15.pdf [accessed 30 January 2013].

Narum, S.R., Buerkle, C.A., Davey, J.W., Miller, M.R., and Hohenlohe, P.A. 2013. Genotyping-by-sequencing in ecological and conservation genomics. Mol. Ecol. **22**(11): 2841–2847. doi:10.1111/mec.12350. PMID:23711105.

Neville, H.M., Isaak, D.J., Dunham, J.B., Thurow, R.F., and Rieman, B.E. 2006. Fine-scale natal homing and localized movement as shaped by sex and spawning habitat in Chinook salmon: insights from spatial autocorrelation analysis of individual genotypes. Mol. Ecol. **15**(14): 4589–4602. doi:10.1111/j.1365-294X.2006.03082.x. PMID:17107485.

Nielsen, E.E., Cariani, A., Mac Aoidh, E., Maes, G.E., Milano, I., Ogden, R., Taylor, M., Hemmer-Hansen, J., Babbucci, M., Bargelloni, L., Bekkevold, D., Diopere, E., Grenfell, L., Helyar, S., Limborg, M.T., Martinsohn, J.T., McEwing, R., Panitz, F., Patarnello, T., Tinti, F., Van Houdt, J.K.J., Volckaert, F.A.M., Waples, R.S., FishPopTrace, C., and Carvalho, G.R. 2012. Gene-associated markers provide tools for tackling illegal fishing and false eco-certification. Nat. Commun. **3**: 851. doi:10.1038/ncomms1845. PMID:22617291.

Nielsen, R., Paul, J.S., Albrechtsen, A., and Song, Y.S. 2011. Genotype and SNP calling from next-generation sequencing data. Nat. Rev. Genet. **12**(6): 443–451. doi:10.1038/nrg2986. PMID:21587300.

Olsen, J.B., Crane, P.A., Flannery, B.G., Dunmall, K., Templin, W.D., and Wenburg, J.K. 2011. Comparative landscape genetic analysis of three Pacific salmon species from subarctic North America. Conserv. Genet. **12**(1): 223–241. doi:10.1007/s10592-010-0135-3.

Peakall, R., and Smouse, P.E. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research — an update. Bioinformatics, **28**(19): 2537–2539. doi:10.1093/bioinformatics/bts460.

Rousset, F. 2008. GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. Mol. Ecol. Resour. **8**(1): 103–106. doi:10.1111/j.1471-8286.2007.01931.x. PMID:21585727.

Russello, M.A., Kirk, S.L., Frazer, K.K., and Askey, P.J. 2012. Detection of outlier loci and their utility for fisheries management. Evol. Appl. **5**(1): 39–52. doi:10.1111/j.1752-4571.2011.00206.x.

Schindler, D.E., Hilborn, R., Chasco, B., Boatright, C.P., Quinn, T.P., Rogers, L.A., and Webster, M.S. 2010. Population diversity and the portfolio effect in an exploited species. Nature, **465**(7298): 609–612. doi:10.1038/nature09060. PMID:20520713.

Schlötterer, C. 2004. The evolution of molecular markers — just a matter of fashion? Nat. Rev. Genet. **5**(1): 63–69. doi:10.1038/nrg1249. PMID:14666112.

Seeb, J.E., Pascal, C.E., Ramakrishnan, R., and Seeb, L.W. 2009*a*. SNP genotyping by the 5′-nuclease reaction: advances in high-throughput genotyping with nonmodel organisms. *In* Single nucleotide polymorphisms: methods and protocols, 2nd ed. *Edited by* A.A. Komar. Methods Mol. Biol. **578**: 277–292.

Seeb, J.E., Carvalho, G., Hauser, L., Naish, K., Roberts, S., and Seeb, L.W. 2011*a*. Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. Mol. Ecol. Resour. **11**: 1–8. doi:10.1111/j.1755-0998.2010.02979.x. PMID:21429158.

Seeb, J.E., Pascal, C.E., Grau, E.D., Seeb, L.W., Templin, W.D., Harkins, T., and Roberts, S.B. 2011*b*. Transcriptome sequencing and high-resolution melt analysis advance single nucleotide polymorphism discovery in duplicated salmonids. Mol. Ecol. Resour. **11**(2): 335–348. doi:10.1111/j.1755-0998.2010.02936.x.. PMID:21429141.

Seeb, L.W., Habicht, C., Templin, W.D., Tarbox, K.E., Davis, R.Z., Brannian, L.K., and Seeb, J.E. 2000. Genetic diversity of sockeye salmon of Cook Inlet, Alaska, and its application to management of populations affected by the *Exxon Valdez* oil spill. Trans. Am. Fish. Soc. **129**(6): 1223–1249. doi:10.1577/1548-8659(2000)129<1223:GDOSSO>2.0.CO;2.

Seeb, L.W., DeCovich, N.A., Barclay, A.W., Smith, C., and Templin, W.D. 2009*b*. Timing and origin of Chinook salmon stocks in the Copper River and adjacent ocean fisheries using DNA markers [online]. Annual report for Study 04-507 USFWS Office Subsistence Management Fisheries Resource Monitoring Program. Available from http://www.adfg.alaska.gov/FedAidpdfs/fds09-58.pdf [accessed 6 December 2013].

Seeb, L.W., Templin, W.D., Sato, S., Abe, S., Warheit, K., Park, J.Y., and Seeb, J.E. 2011*c*. Single nucleotide polymorphisms across a species' range: implications for conservation studies of Pacific salmon. Mol. Ecol. Resour. **11**: 195–217. doi:10.1111/j.1755-0998.2010.02966.x. PMID:21429175.

Shaklee, J.B., Beacham, T.D., Seeb, L., and White, B.A. 1999. Managing fisheries using genetic data: case studies from four species of Pacific salmon. Fish Res. **43**(1–3): 45–78. doi:10.1016/S0165-7836(99)00066-1.

Smith, C.T., Elfstrom, C.M., Seeb, L.W., and Seeb, J.E. 2005*a*. Use of sequence data from rainbow trout and Atlantic salmon for SNP detection in Pacific salmon. Mol. Ecol. **14**(13): 4193–4203. doi:10.1111/j.1365-294X.2005.02731.x. PMID:16262869.

Smith, C.T., Seeb, J.E., Schwenke, P., and Seeb, L.W. 2005*b*. Use of the 5′-nuclease reaction for single nucleotide polymorphism genotyping in Chinook salmon. Trans. Am. Fish. Soc. **134**(1): 207–217. doi:10.1577/T04-019.1.

Smith, C.T., Templin, W.D., Seeb, J.E., and Seeb, L.W. 2005*c*. Single nucleotide polymorphisms provide rapid and accurate estimates of the proportions of U.S. and Canadian Chinook salmon caught in Yukon River fisheries. N. Am. J. Fish. Manage. **25**(3): 944–953. doi:10.1577/M04-143.1.

Smith, C.T., Antonovich, A., Templin, W.D., Elfstrom, C.M., Narum, S.R., and Seeb, L.W. 2007. Impacts of marker class bias relative to locus-specific variability on population inferences in Chinook salmon: a comparison of single-nucleotide polymorphisms with short tandem repeats and allozymes. Trans. Am. Fish. Soc. **136**(6): 1674–1687. doi:10.1577/T06-227.1.

Smith, M.J., Pascal, C.E., Grauvogel, Z., Habicht, C., Seeb, J.E., and Seeb, L.W. 2011. Multiplex preamplification PCR and microsatellite validation enables accurate single nucleotide polymorphism genotyping of historical fish scales. Mol. Ecol. Resour. **11**(s1): 268–277. doi:10.1111/j.1755-0998.2010.02965.x.

Stewart, I.J., Quinn, T.P., and Bentzen, P. 2003. Evidence for fine-scale natal homing among island beach spawning sockeye salmon, *Oncorhynchus nerka*. Environ. Biol. Fishes, **67**(1): 77–85. doi:10.1023/A:1024436632183.

Storer, C.G., Pascal, C.E., Roberts, S.B., Templin, W.D., Seeb, L.W., and Seeb, J.E. 2012. Rank and order: evaluating the performance of SNPs for individual assignment in a non-model organism. PLoS ONE, **7**(11): e49018. doi:10.1371/journal.pone.0049018. PMID:23185290.

Stram, D.L., and Ianelli, J.N. 2009. Eastern Bering Sea pollock trawl fisheries: variation in salmon bycatch over time and space. *In* Pacific salmon: ecology and management in western Alaska's populations. *Edited by* C.C. Krueger and C.E. Zimmerman. Am. Fish. Soc. Symp. No. 70, Bethesda, Md. pp. 827–850.

Templin, W.D., Seeb, J.E., Jasper, J.R., Barclay, A.W., and Seeb, L.W. 2011. Genetic differentiation of Alaska Chinook salmon: the missing link for migratory studies. Mol. Ecol. Resour. **11**: 226–246. doi:10.1111/j.1755-0998.2010.02968.x. PMID:21429177.

Tucker, S., Trudel, M., Welch, D.W., Candy, J.R., Morris, J.F.T., Thiess, M.E., Wallace, C., Teel, D.J., Crawford, W., Farley, E.V., and Beacham, T.D. 2009. Seasonal stock-specific migrations of juvenile sockeye salmon along the west coast of North America: implications for growth. Trans. Am. Fish. Soc. **138**(6): 1458–1480. doi:10.1577/T08-211.1.

Tucker, S., Trudel, M., Welch, D.W., Candy, J.R., Morris, J.F.T., Thiess, M.E., Wallace, C., and Beacham, T.D. 2011. Life history and seasonal stock-specific ocean migration of juvenile Chinook salmon. Trans. Am. Fish. Soc. **140**(4): 1101–1119. doi:10.1080/00028487.2011.607035.

Utter, F. 2004. Population genetics, conservation and evolution in salmonids and other widely cultured fishes: some perspectives over six decades. Rev. Fish Biol. Fish. **14**(1): 125–144. doi:10.1007/s11160-004-3768-9.

Utter, F., and Ryman, N. 1993. Genetic markers and mixed stock fisheries. Fisheries, **18**(8): 11–21. doi:10.1577/1548-8446(1993)018<0011:GMAMSF>2.0.CO;2.

Waples, R.S. 1990. Temporal changes of allele frequency in Pacific salmon: implications for mixed-stock fishery analysis. Can. J. Fish. Aquat. Sci. **47**(5): 968–976. doi:10.1139/f90-111.

Waples, R.S., Dickhoff, W.W., Hauser, L., and Ryman, N. 2008. Six decades of fishery genetics: taking stock. Fisheries, **33**(2): 76–79.

Warheit, K., Seeb, L.W., Templin, W.D., and Seeb, J.E. 2013. Moving GSI into the Next Decade: SNP Coordination for Pacific Salmon Treaty Fisheries. Chinook Technical Committee. Project number N10-8.

Weir, B.S., and Cockerham, C.C. 1984. Estimating *F*-statistics for the analysis of population structure. Evolution, **38**(6): 1358–1370. doi:10.2307/2408641.

Zerbino, D.R., and Birney, E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res. **18**(5): 821–829. doi:10.1101/gr.074492.107. PMID:18349386.