ORIGINAL ARTICLE

Detection of outlier loci and their utility for fisheries management

Michael A. Russello[1], Stephanie L. Kirk[1], Karen K. Frazer[1] and Paul J. Askey[2]

1 Department of Biology, University of British Columbia, Okanagan Campus, Kelowna, BC, Canada
2 British Columbia Ministry of Environment, Penticton, BC, Canada

Abstract

Genetics-based approaches have informed fisheries management for decades, yet remain challenging to implement within systems involving recently diverged stocks or where gene flow persists. In such cases, genetic markers exhibiting locus-specific ('outlier') effects associated with divergent selection may provide promising alternatives to loci that reflect genome-wide ('neutral') effects for guiding fisheries management. Okanagan Lake kokanee (*Oncorhynchus nerka*), a fishery of conservation concern, exhibits two sympatric ecotypes adapted to different reproductive environments; however, previous research demonstrated the limited utility of neutral microsatellites for assigning individuals. Here, we investigated the efficacy of an outlier-based approach to fisheries management by screening >11 000 expressed sequence tags for linked microsatellites and conducting genomic scans for kokanee sampled across seven spawning sites. We identified eight outliers among 52 polymorphic loci that detected ecotype-level divergence, whereas there was no evidence of divergence at neutral loci. Outlier loci exhibited the highest self-assignment accuracy to ecotype (92.1%), substantially outperforming 44 neutral loci (71.8%). Results were robust among-sampling years, with assignment and mixed composition estimates for individuals sampled in 2010 mirroring baseline results. Overall, outlier loci constitute promising alternatives for informing fisheries management involving recently diverged stocks, with potential applications for designating management units across a broad range of taxa.

Introduction

Genetic tools have long informed fisheries management (Ferris and Berg 1987; Carvalho and Hauser 1994). Conventional genetics-based fisheries management strategies use a large suite of neutral genetic markers for a range of applications including defining stocks, analyzing mixtures, identifying specimens, monitoring restocking initiatives and guiding aquaculture operations (Ward 2000). Such approaches have been especially informative for guiding management of Pacific salmonids (Small et al. 1998; Beacham et al. 2004, 2005, 2006, 2010). However, traditional approaches may be ineffective when populations are recently isolated, and divergence is not yet reflected at neutral loci. Alternatively, loci influenced by selection may offer valuable population markers on more recent (ecological) timescales and, specifically, in cases where the degree of selection is strong or population sizes are sufficiently large whereby even weak selection may override drift (O'Malley et al. 2007; Westgaard and Fevolden 2007; Hauser and Carvalho 2008; André et al. 2011). Until recently, our capacity to directly investigate adaptive genetic variation has been limited to a handful of model organisms studied under artificial conditions (Vasemägi et al. 2005a; Mitchell-Olds and Schmitt 2006). The emergence of population genomics, which combines genomic technologies with population genetics theory, provides a framework for detecting genome-wide associations between segregating variation and fitness-related traits in natural populations without prior knowledge of genome structure (Black et al. 2001; Luikart et al. 2003).

Okanagan Lake kokanee (*Oncorhynchus nerka*) presents a promising system for exploring the genetic basis of adaptation within a management context. Over a single decade, this population of land-locked sockeye salmon went from being managed as a sport fishery to one of major conservation concern (Shepherd 2000). In recent years, the number of spawners has been reported as <13 000 adults, an estimated 1% of the numbers observed just 25–30 years ago (Shepherd 2000). Despite this marked decline in numbers, Okanagan Lake kokanee continue to display a high degree of variation in their reproductive behavior that is uncommon among salmonids. Specifically, two sympatric reproductive ecotypes have been described, generally termed as 'stream-' and 'shore spawners', with the 'shore' ecotype exhibiting variation from the typical sockeye behavior related to spawning location, substrate, timing and site defense (Table S1; Dill 1996; Shepherd 2000). Outside the spawning season, the two ecotypes are morphologically indistinguishable and reside in mixed schools. Ecotype divergence has only occurred in the last 12 000 years since the Wisconsinan glaciation; nevertheless, low levels of genetic differentiation have been detected between the two reproductive ecotypes in Okanagan Lake based on mitochondrial DNA haplotype data as well as genotypic data at five nuclear microsatellite loci (Taylor et al. 1997, 2000). These findings are in contrast to the fine-scale neutral differentiation found between sockeye salmon ecotypes in Little Togiak Lake, Alaska (Lin et al. 2008).

Fisheries management of Okanagan Lake is dependent upon basic scientific data in the form of stock-specific estimates of abundance, productivity, and harvest. The current lack of genetics-based information means that all in-lake kokanee data (juvenile abundance and angler harvest) cannot be completely interpreted to stock-specific values, as ecotypes are morphologically and behaviorally indistinguishable outside the spawning season (Taylor et al. 2000), unlike what has been found in other sockeye populations (Ramstad et al. 2010). At present, annual spawning ground surveys are the only enumeration method that can produce ecotype-specific abundance estimates for Okanagan Lake kokanee. While visual counts of stream-spawning salmonids are a long-established practice with well-documented assessment protocols (Hilborn et al. 1999; Holt and Cox 2008), evaluating a shore-spawning population in this manner is more complicated and can only be treated as an index (not absolute abundance) owing to logistical complications of visually enumerating shore spawners. In particular, there is a high degree of uncertainty in the accuracy of visual surveys given large spawning aggregations are spread over wide areas (e.g., approximately 100 km of shoreline for Okanagan Lake) and in depths of up to several meters. Fish may also move among spawning areas at unknown spatial and temporal scales, and the open nature of the spawning area makes tagging studies logistically challenging. A genetics-based approach for monitoring Okanagan Lake kokanee has the potential to minimize errors, while allowing relative abundance estimates to be calculated at any time of year and across all life stages. Yet, previous research in this system has demonstrated the limited utility of neutral microsatellite markers for assigning individuals to ecotype and for use in mixed composition (MC) analyses (Taylor et al. 2000). Specifically, mixed population analyses based on allelic variation at five putatively neutral microsatellite resulted in high error rates for the assignment of individuals to their respective ecotypes (i.e., 29% and 24% incorrect assignment of shore- and stream spawners, respectively; Taylor et al. 2000).

The inability of neutral loci to detect population structure among ecotypes is likely related to the recent postglacial isolation of kokanee within Okanagan Lake (Taylor et al. 1997), impeding traditional genetic approaches to fisheries management. However, the identification of loci directly associated with adaptation to markedly different kokanee spawning environments may offer opportunities for genetically identifying ecotypes, upon which past and current fisheries management practices have been based. Population genomics offers an approach for identifying such candidate loci in the form of markers that exhibit locus-specific behavior or patterns of variation that are extremely divergent from the rest of the genome (hereafter referred to as 'outlier' loci; Luikart et al. 2003). Previous studies in other aquatic and terrestrial systems have conducted genomic scans for detecting outlier loci and studying local adaptation using amplified fragment length polymorphisms (AFLPs; Bonin et al. 2006), expressed sequence tag (EST)-linked microsatellite loci (Vasemägi et al. 2005a), and sequenced restriction-site-associated DNA (RAD) tags (Hohenlohe et al. 2010). Recently, Freamo et al. (2011) demonstrated marginally improved accuracy of 14 outlier single nucleotide polymorphisms (SNPs) over 67 neutral SNPs (85% and 75%, respectively) for distinguishing populations of Atlantic salmon. To date, the efficacy of an outlier-based approach to fisheries management has not been explicitly evaluated. In the case of Okanagan Lake kokanee, an outlier-based approach may offer the only alternative to neutral loci for informing fisheries management. More generally, this approach may have broader applications to freshwater and marine fisheries, as well as to terrestrial systems for distinguishing recently diverged populations and/or increasing cost-effectiveness of genetic approaches to management by limiting screenings to substantially fewer, but highly informative loci directly associated with adaptive divergence.

To explicitly examine the genetic basis of kokanee adaptation to varying spawning environments and to evaluate the efficacy of an outlier-based approach to fisheries management, we screened over 11 000 EST-linked microsatellite loci and conducted genomic scans for individuals sampled across seven Okanagan Lake spawning sites (Fig. 1). Subsequently, we employed multiple approaches to detect outlier loci and to explore the ability of outlier loci to improve upon previous approaches for differentiating kokanee ecotypes in both individual assignment (IA) and MC analyses.

Materials and methods

Study site and tissue collection
Okanagan Lake is located between the Monashee and Cascade mountain ranges in the south-central interior of British Columbia (Fig. 1). It spans 351 km^2, has an average depth of 76 m, and supports an estimated 22
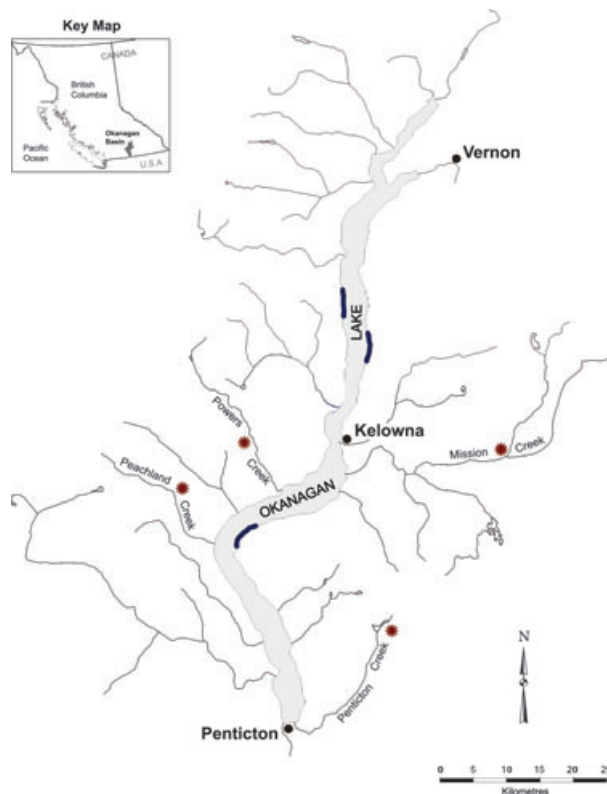

Figure 1 Map of Okanagan Lake showing locations of kokanee sampling sites analyzed in this study. Darkly shaded shorelines denote shore-spawning sites in the northeast, northwest, and southeast quadrants of Okanagan Lake. Stars denote stream-spawning sites including Powers Creek, Peachland Creek, Penticton Creek, and Mission Creek.

freshwater fish species including native kokanee (*BC Ministry of Environment Fish Inventory Data Queries* 2010).

Operculum punches were collected from mature Okanagan Lake kokanee at the time of spawning in the fall of 2007 ($n = 136$). Samples were obtained from seven sampling sites spanning broad geographic coverage of Okanagan Lake including the four main tributaries [Peachland Creek ($n = 15$), Penticton Creek ($n = 16$), Mission Creek ($n = 22$), and Powers Creek ($n = 15$)] and three shore-spawning localities in the northeast ($n = 20$), northwest ($n = 29$), and southeast ($n = 19$) quadrants (Fig. 1). In addition, tissue samples were collected from mature Okanagan Lake kokanee at the time of spawning in the fall of 2010 ($n = 28$), representing four individuals from each of the seven sampling sites (Fig. 1).

Marker selection

A literature search for EST-linked microsatellite markers described for *Salmo* or *Oncorhynchus* species was conducted, from which 99 were selected for subsequent testing (Miller et al. 1997; Grimholt et al. 2002; Ng et al. 2005; Rexroad et al. 2005; Vasemägi et al. 2005b; Wright et al. 2008). In addition, 11 389 ESTs described for *O. nerka* were identified from GenBank (date of inspection 01/08/2008), examined for microsatellites containing uninterrupted dinucleotide repeats using Tandem Repeats Finder (Benson 1999), and subsequently cross-referenced with the consortium for Genomics Research on All Salmon (cGRASP) to target those that have known functional annotations. Using this approach, polymerase chain reaction (PCR) primers were designed for an additional 125 EST-linked microsatellites using PRIMER3 software (Rozen and Skaletsky 2000). Lastly, we targeted 19 putatively neutral, non-EST-linked microsatellite loci developed specifically for *Oncorhynchus* spp. and used in previously published population genetic analyses of Pacific salmonids (Morris et al. 1996; O'Reilly et al. 1996; Olsen et al. 1996; Scribner et al. 1996; Banks et al. 1999; Nelson and Beacham 1999; Lin et al. 2008). In total, we tested 243 loci comprising 224 EST-linked microsatellites and 19 putatively neutral, non-EST-linked microsatellites (Table S2).

Marker characterization and genetic data collection

Genomic DNA was extracted from all 164 tissue samples using the NucleoSpin Tissue kit (Macherey Nagel, Düren, Germany), following the manufacturer's suggested protocol. All PCRs were carried out on a Veriti® thermal cycler in (Applied Biosystems, Foster City, CA, USA) 12.5-μL reactions containing: ~20–50 ng of DNA, 10 mM Tris–HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl2, 200 μM

dNTPs, 0.08 μM of the M13-tailed forward primer, 0.8 μM of each of the reverse primer and the M13 fluorescent dye-labeled primer, and 0.5 U of *Taq* polymerase in a total volume of 12.5 μL. KAPA *Taq* DNA polymerase (Kapa Biosystems, Woburn, MA, USA) was used for the majority of PCRs; however, AmpliTaq Gold DNA polymerase (Applied Biosystems) was used occasionally as a secondary measure to promote PCR amplification. All forward primers were 5′-tailed with an M13 sequence [5′-TCCCAGTCACGA-CGT-3′] to facilitate automated genotyping. Specifically, the M13-tailed forward primer was used in combination with an M13 primer of the same sequence 5′-labeled with one of four fluorescent dyes (6-FAM, VIC, NED, PET; Applied Biosystems), effectively incorporating the fluorescent label into the resulting PCR amplicon (Schuelke 2000). Likewise, reverse primers were modified following Brownstein et al. (1996) to improve scoring quality.

A touchdown PCR was used with an initial denaturation at 94°C for 2–10 min depending on the manufacturer's recommendation for the *Taq* DNA polymerase being used. This was followed by 20 cycles at 94°C for 30 s, 60°C for 30 s, and 72°C for 30 s with the annealing temperature decreasing by 0.5°C each cycle to 50°C. The annealing temperature was maintained at 50°C for another 15 cycles followed by a final extension at 72°C for 2 min. If the initial PCR amplification failed, a second touchdown PCR was attempted with the annealing temperatures lowered by 5°C to a range of 55–45°C, which is similar to the PCR optimization strategy used by Vasemägi et al. (2005b). As a final strategy for amplification, the PCR program was modified to maintain the annealing temperature at 5°C below the lowest melting temperature of the two primers. PCR products of four representative samples were electrophoresed in a 1.5% agarose gel and visualized on the gel documentation system Red (Alpha Innotech, San Leandro, CA, USA). Fragment analysis was performed on an Applied Biosystems 3130XL DNA automated sequencer using GS500 LIZ size standard to determine fragment length. Alleles were scored based on their consistent pattern of stutter peaks and peak intensity for individuals at each locus using GENEMAPPER 4.0 (Applied Biosystems, Foster City, CA, USA).

Successfully amplified markers were tested for genetic variability in 16 individuals including 2–3 from each of the seven Okanagan Lake shore and stream-spawning localities (Fig. 1). Candidate loci for further characterization exhibited clean peak topography with few or no stutter bands and conformed to expectations of the documented microsatellite motif. Loci that were monomorphic or nearly monomorphic in all individuals (i.e., loci with dominant allele frequencies >0.95 in both ecotypes) were excluded to account for allele call and

fragment analysis errors (Pompanon et al. 2005). Once polymorphic markers were identified, the sample size was expanded to a larger screen of 136 individuals ($n = 68$ shore spawners; $n = 68$ stream spawners) to identify candidate loci exhibiting outlier behavior. Based on results generated for this baseline sampling, the 28 individuals sampled in 2010 were genotyped at identified outlier loci as well as at the retained non-EST-linked microsatellite loci (see Results).

Outlier loci detection

Outlier loci were detected using three different approaches. In all cases, the input data set was split by ecotype. First, a coalescent-based simulation approach was used to identify outlier loci displaying unusually high and low values of F_{ST} by comparing observed F_{ST} values with values expected under neutrality (Beaumont and Nichols 1996) as implemented in LOSITAN Selection Workbench (Antao et al. 2008). We performed an initial run with 50 000 simulations and all loci, using the mean neutral F_{ST} as a preliminary value. A more accurate estimate of the mean neutral F_{ST} was obtained following the first run by excluding all loci lying outside the 99% confidence interval, as their distribution could be the result of selection rather than neutral evolution. This refined estimate was used for a final set of 50 000 simulations over all loci. Second, we employed the approach of Vitalis et al. (2001), which investigates outliers in a pairwise fashion based on population-specific F-statistics. The coalescent simulations were performed with DETSEL 1.0 (Vitalis et al. 2003). Null distributions were generated using the following parameters: population size before the split $N_0 = 500$; mutation rate $\mu = 0.0001$ and 0.00001; ancestral population size $N_e = 500$, 1000 and 10 000; time since bottleneck $T_0 = 50$, 100 and 1000; and time since population split $t = 100$. Outliers were determined based on an empirical P value for each locus at the 95% and 99% levels using the two-dimensional arrays of 50×50 square cells (Vitalis et al. 2001). Lastly, we used the Bayesian simulation-based test of Beaumont and Balding (2004) that has been further refined and implemented in the software BAYESCAN 2.0 (Foll and Gaggiotti 2008). We based our analyses on 10 pilot runs each consisting of 5000 iterations, followed by 100 000 iterations with a burn-in of 50 000 iterations.

Population genetic analysis

The data set was screened for null alleles using MICRO-CHECKER (Van Oosterhout et al. 2004). Allelic diversity, observed (H_O) and expected heterozygosity (H_E) were calculated at each locus for each ecotype and spawning locality

using ARLEQUIN 3.11 (Excoffier et al. 2005). Deviation from Hardy–Weinberg (H-W) equilibrium was assessed using exact tests based on the Markov chain method of Guo and Thompson (1992) as implemented in GENEPOP 3.3 (1000 dememorization, 1000 batches and 10 000 iterations; Raymond and Rousset 1995). Linkage disequilibrium was investigated for all pairs of loci using GENEPOP 3.3 (Raymond and Rousset 1995). Type I error rates for tests of linkage disequilibrium and departure from H-W expectations were corrected for multiple comparisons using the sequential Bonferroni procedure (Rice 1989).

The hierarchical organization of genetic variation was assessed using an analysis of molecular variance (AMOVA; Excoffier et al. 1992) based on F_{ST} comparisons within and among reproductive ecotypes as implemented in ARLEQUIN 3.11 (Excoffier et al. 2005). Likewise, genetic differentiation among each spawning locality was calculated by pairwise F_{ST} (Weir and Cockerham 1984), for which 95% confidence intervals were estimated by bootstrapping over loci, all of which were implemented in ARLEQUIN 3.11 (Excoffier et al. 2005).

Correspondence of geographically separated spawning sites and ecotypes as discrete genetic units was further tested using the Bayesian method of Pritchard et al. (2000) as implemented in STRUCTURE 2.3.3. Run length was set to 1 000 000 Markov chain Monte Carlo (MCMC) replicates after a burn-in period of 500 000 using correlated allele frequencies under an admixture model using the LOCPRIOR option. The LOCPRIOR option uses sampling locations as prior information to assist the clustering for use with data sets where the signal of structure is relatively weak (Hubisz et al. 2009). The most likely number of clusters inferred from the different data sets was determined using the ΔK approach (Evanno et al. 2005), by varying the number of clusters K from 1 to 10 with 20 iterations per value of K.

IA and MC analysis

The ability of different sets of markers to assign individuals to their most likely source ecotype was assessed using two approaches. First, we employed the leave-one-out test implemented in the genetic stock identification program ONCOR (Kalinowski et al. 2007). This approach sequentially removes each fish in each population from the baseline, and its origin is estimated using the rest of the baseline. Second, we performed realistic fishery simulations by randomly generating 200 fish from the baseline data using the actual mixture proportions (Anderson et al. 2008) as well as a range of possible simulated mixture proportions using ONCOR (Kalinowski et al. 2007) including (shore:stream): 10:90, 25:75, 50:50, 75:15, and 90:10. The performance of each baseline data set to assign

individuals to their ecotype of origin across this range of mixture proportions was estimated as the proportion of individuals accurately assigned.

We conducted a series of analyses for assessing the performance of different sets of markers for estimating stock composition as implemented in ONCOR (Kalinowski et al. 2007). First, we used the 100% simulation method of Anderson et al. (2008) to simulate mixture genotypes and to estimate their probability of occurrence in the baseline population. For this analysis, we used the same sample size as in the original baseline ($n = 136$), simulating 10 000 mixture samples of 200 individuals. Second, we performed realistic fishery simulations by randomly generating 200 fish from the baseline data using the actual mixture proportions (Anderson et al. 2008) as well as a range of possible simulated mixture proportions using ONCOR (Kalinowski et al. 2007) including (shore:stream): 10:90, 25:75, 50:50, 75:15, and 90:10. Estimated ecotypes proportions and their standard deviations were compared with the actual (simulated) proportions for each data set.

High-grading bias may result from using the same set of individuals for both selecting a panel of loci and assessing their informativeness (Anderson 2010). To assess the potential for high-grading bias, the accuracy of among-year IA and MC analyses was also tested using the 28 individuals sampled in 2010 from the same shore ($n = 12$ individuals) and stream ($n = 16$ individuals) localities as the baseline sampling from 2007. For all analyses, the 2007 samples ($n = 136$ from seven localities) were used as the baseline file. The 2010 samples ($n = 28$) were assigned to ecotype based on the highest probability of producing the observed genotype in the mixture as implemented in ONCOR (Kalinowski et al. 2007). This approach uses both genotype frequencies and mixture proportions when estimating the origin of individuals, incorporating the method of Rannala and Mountain (1997) to estimate assignment probability. Ecotype mixture proportions were estimated in the 2010 sample using the conditional maximum likelihood approach implemented in ONCOR (Kalinowski et al. 2007).

Results

Marker characterization
PCR amplification was successful for 204 of the 243 EST-linked and non-EST-linked microsatellite markers evaluated. Following initial tests of polymorphism at these 204 loci based on 16 representative individuals, multilocus genotypic data were generated for 136 individuals at 57 candidate loci (i.e., 48 EST-linked and nine non-EST-linked microsatellite loci) that were used for further analyses (Table S3). None of the 57 candidate loci had missing data >5%, averaging 1.6% per locus.

Evidence for null alleles was detected at two loci (EV103, One109) for both the combined data set ($n = 136$), as well as when shore ($n = 68$) and stream ($n = 68$) ecotypes were considered separately. In addition, three pairs of loci (Ca687/EV712; Ca613/Ots14; Omm5099/Ots29) exhibited evidence of linkage disequilibrium within both ecotypes. Based on these results, five loci (Ca687, EV103, One109, Ots14, Ots29) were removed from any further analyses. None of the remaining 52 loci (i.e., 44 EST-linked and eight non-EST-linked microsatellite loci) showed evidence for departure from H-W expectations within ecotype.

Outlier loci detection and dataset definition

A total of eight outlier loci were identified using three different approaches. Two outlier loci (EV358, Omm5125) were identified by all three algorithms (Figs 2, S1 and S2). One additional locus (EV642) was identified
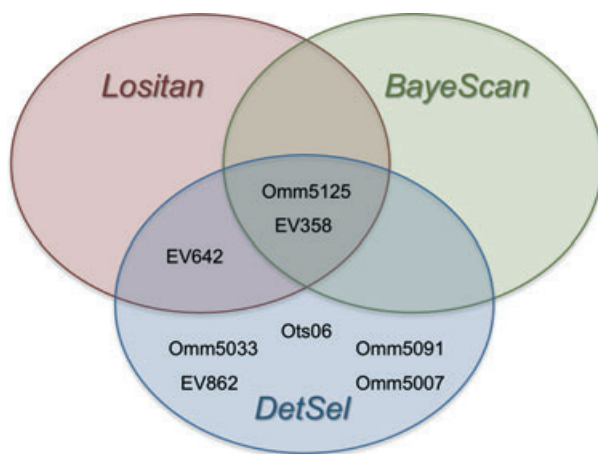

Figure 2 Outlier loci identified by the approaches of Beaumont and Nichols (1996), Vitalis et al. (2001) and Beaumont and Balding (2004) as implemented in LOSITAN (Antao et al. 2008), DETSEL 1.0 (Vitalis et al. 2003), and BAYESCAN 2.0 (Foll and Gaggiotti 2008), respectively.

by both LOSITAN and DETSEL, while the remaining loci (EV862, OMM5007, Omm5033, Omm5091, and Ots06) were identified by DETSEL only (Fig. 2). All eight loci were subjected to sequence similarity searches via BLAST as well as localized within the rainbow trout transcriptome (Salem et al. 2010). Two loci exhibited similarity to an annotated protein-coding gene within GenBank. Specifically, EV358 showed similarity to a CXC chemokine receptor type 4 putative mRNA cloned in *Osmerus mordax* (Genbank accession no BT075426) while EV642 exhibited partial overlap with *Salmo salar* transcription elongation factor B polypeptide 2 (Genbank accession no BT049838).

Based on the results from the outlier detection analyses, three data sets were constructed from the overall pool of 52 candidate loci for exploring patterns of population differentiation and outlier behavior. One data set comprised only the eight outlier loci (EV358, EV642, EV862, Omm5007, Omm5033, Omm5091, Omm5125, and Ots06). A second data set comprised the 44 putatively neutral EST-linked and non-EST-linked loci. A third data set contained the eight non-EST-linked loci (One8, One14, One102, One105, One108, One110, One112, and Ssa85) conventionally used in population genetic analyses.

Neutral and adaptive population divergence

A high degree of genetic variation was detected within this system (Table 1 and S4). Overall, mean number of alleles per locus was 5.4 with a mean expected heterozygosity of 0.53. For shore spawners, the number of alleles per locus ranged from 2 to 21 with a mean of 5.7. Likewise for stream-spawning individuals, the number of alleles ranged from 2 to 25 per locus with a mean of 5.1.

The degree of ecotype divergence varied among the three data sets. Based on the eight outlier loci only, the AMOVA revealed that 6.31% of the variation occurred among ecotypes, which was highly significant. These patterns were further reflected by the results from the STRUCTURE analysis, which inferred two clusters corresponding

Table 1. Okanagan Lake kokanee sampling information.

Site	Abbreviation	Sample size	Allelic richness*	Expected heterozygosity*	Latitude	Longitude	Ecotype
Mission Creek	MC	22	5.56	0.53	49.877027	−119.429492	Stream
Peachland Creek	PAC	15	4.75	0.52	49.784735	−119.714618	Stream
Penticton Creek	PNC	16	5.19	0.53	49.493774	−119.579110	Stream
Powers Creek	POC	15	4.71	0.53	49.833474	−119.644761	Stream
Northeast Shore	NE	20	5.49	0.53	50.034793	−119.446837	Shore
Northwest Shore	NW	29	6.08	0.52	50.068750	−119.495884	Shore
Southeast Shore	SE	19	5.39	0.53	49.746542	−119.715971	Shore

*Mean across 52 loci.

to ecotype based on the eight outlier loci (Fig. 3A). In contrast, there was no evidence of among-ecotype divergence when using the two data sets composed of varying numbers of putatively neutral loci in the AMOVA, resulting in 0.16% and 0.49% among-ecotype variation for the eight and 44 neutral loci, respectively. The absence of signal for ecotype divergence at neutral loci was likewise reflected in the STRUCTURE analyses (Fig. 3B).

Finer-scale analyses investigating differentiation among the seven spawning localities were largely consistent with the ecotype-level results. All 12 pairwise comparisons of inter-ecotype spawning localities exhibited significant F_{ST} values based on the eight outlier loci, while only one of the intra-ecotype comparisons was significant (Peachland Creek/Mission Creek; Table 2 below diagonal). In contrast, only four of a total of 21 pairwise comparisons were significant using the 44 putatively neutral loci, all of which

included inter-ecotype comparisons (Table 2 above diagonal). The large number of loci used in this latter analysis ($n = 44$) likely contributed to the findings of significant differentiation despite the low F_{ST} values (0.008–0.010; Table 2 above diagonal), as no pairwise comparisons between spawning localities were significant based on the eight non-EST-linked neutral loci (data not shown).

IA and MC analysis

Consistent with the results from the population genetic analyses, outlier loci substantially outperformed the data sets composed of eight and 44 neutral loci for IA to ecotype and in MC analyses (Table 3; Table S5). Overall, the eight outlier loci correctly self-assigned 92.0% of all individuals to ecotype based on the leave-one-out test implemented in ONCOR (Kalinowski et al. 2007), while the
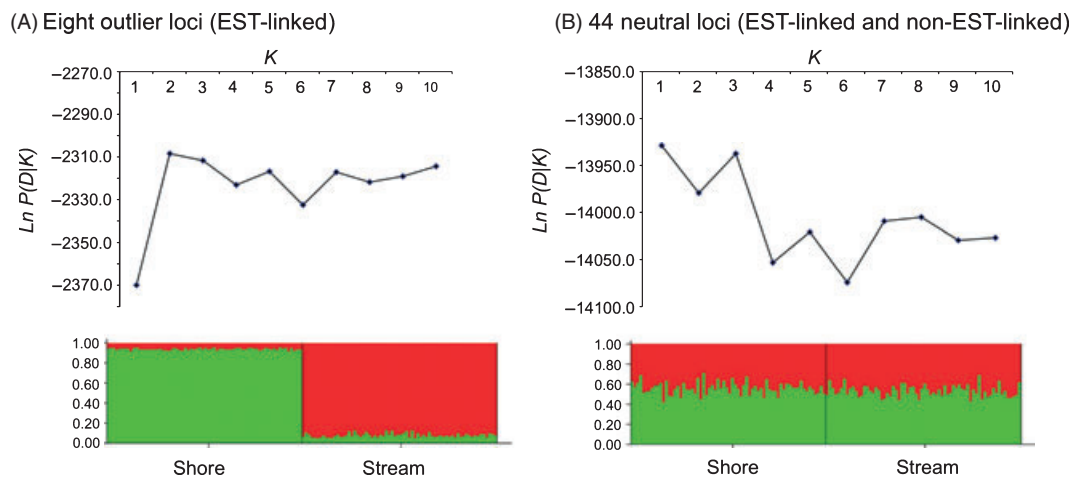

Figure 3 Bayesian clustering according to the approach of Pritchard et al. (2000) as implemented in STRUCTURE. Log-likelihood profiles for a given number of putative populations (*K*) ranging from 1 to 10, averaged over 20 independent runs and population assignment to shore (green)- and stream (red)-spawning kokanee are shown, based on (A) eight expressed sequence tag (EST)-linked outlier loci; (B) 44 EST-linked and non-EST-linked neutral microsatellite loci.

Table 2. Pairwise comparisons of F_{ST} values between spawning localities based on outlier and neutral microsatellite loci.

	Shore			Stream			
	NE	SE	NW	PAC	PNC	MC	POC
NE	–	0.006	−0.002	0.002	0.002	0.008*	0.009*
SE	0.008	–	−0.001	0.003	0.003	0.009*	0.006
NW	0.007	0.002	–	0.010*	0.005	0.005	0.003
PAC	0.118*	0.220*	0.145*	–	−0.005	0.007	0.007
PNC	0.068*	0.058*	0.090*	0.002	–	−0.001	−0.001
MC	0.035*	0.040*	0.067*	0.045*	0.008	–	−0.003
POC	0.033*	0.033*	0.056*	0.022	−0.007	−0.002	–

Above diagonal results based on 44 expressed sequence tag (EST)-linked and non-EST-linked neutral loci; below diagonal based on eight EST-linked outlier loci. Abbreviations for spawning localities are as in Table 1.
*$P < 0.05$.

Table 3. Individual assignment (IA) to ecotype (Shore/Stream) and mixed composition (MC) analyses (Stream%) across different mixture proportions using outlier and neutral microsatellite loci.

Ecotype	Simulated proportions	Performance*		
		Outlier (eight loci)†	Neutral (eight loci)‡	Neutral (44 loci)[†,‡]
Shore	Self-assign[§]	**0.940**[¶]	0.622	0.721
Stream	Self-assign	**0.900**	0.559	0.714
Shore	0.10	0.659	0.558	**0.703**
Stream	0.90	**0.951**	0.765	0.812
Stream%	0.90	**0.866 (0.011)**[**]	0.703 (0.020)	0.750 (0.015)
Shore	0.25	**0.755**	0.608	0.736
Stream	0.75	**0.916**	0.726	0.787
Stream%	0.75	**0.728 (0.012)**	0.625 (0.018)	0.649 (0.014)
Shore	0.50	**0.844**	0.676	0.775
Stream	0.50	**0.856**	0.664	0.748
Stream%	0.50	**0.503 (0.013)**	0.496 (0.019)	0.489 (0.015)
Shore	0.75	**0.907**	0.740	0.812
Stream	0.25	**0.768**	0.595	0.704
Stream%	0.25	**0.277 (0.013)**	0.367 (0.019)	0.324 (0.015)
Shore	0.90	**0.944**	0.778	0.835
Stream	0.10	**0.675**	0.550	0.671
Stream%	0.10	**0.140 (0.011)**	0.286 (0.019)	0.227 (0.014)

*IA and MC analyses based on realistic fisheries simulations conducting according to the method of Kalinowski et al. (2007). IA performance is based on assignment accuracy to ecotype. Performance of MC analyses assessed how closely estimated mixture proportions match actual mixture proportions.

†Expressed sequence tag (EST)-linked microsatellite loci.

‡Non-EST-linked microsatellite loci.

§Self-assignment according to the leave-one-out test implemented in ONCOR (Kalinowski et al. 2007).

¶Bolded values indicate highest assignment accuracy among data sets (1.0 = perfect assignment). For MC analyses (Stream%), bolded values are closest to the simulated mixture proportions. Note there is no best performing data set for a 50:50 mix, given overlapping standard deviations.

**Standard deviations indicated in parentheses.

neutral eight and neutral 44 loci data sets correctly assigned 59.1% and 71.8%, respectively (data not shown). Similarly, outlier loci self-assigned a substantially higher proportion of individuals than the two neutral data sets when ecotypes were analyzed separately (Table 3). Over a range of simulated mixture proportions, the outlier loci also assigned the highest proportion of individuals to ecotype in all but one instance (Table 3).

In terms of MC analysis, outlier loci consistently recovered estimated stock proportions that most closely mirrored simulated proportions without exception (Table 3; Table S5). In addition, the outliers substantially outperformed the neutral loci in 100% simulation, recovering 95.6% and 96.4% probability of occurrence of shore- and stream spawners, respectively (Table S5).

Individual assignment and MC analyses using the 28 individuals sampled in 2010 and baseline data from 2007 produced similar results. The outlier loci substantially outperformed the neutral data set, misassigning only one individual per ecotype, with a combined assignment accuracy of 92.9%. The high assignment accuracy using the outlier loci was in stark contrast to results based on the

eight neutral loci, which only correctly assigned 57.1% of individuals to ecotype (Table 4). These patterns were mirrored in MC analyses, where the results based on outlier loci closely tracked actual proportions, while estimated mixtures based on neutral loci deviated substantially (Table 4).

Discussion

Genetics-based approaches have informed fisheries management for decades, yet they remain challenging to implement within systems involving recently diverged stocks or in cases where gene flow persists. Okanagan Lake kokanee exhibits both of these confounding factors, which has prevented incorporation of genetic tools into fisheries management strategies. Previous genetic studies have revealed that kokanee ecotype differentiation has likely occurred in Okanagan Lake since the last glaciation (<12 000 ybp; Taylor et al. 1997). Likewise, the lack of population genetic structure among ecotypes as inferred from neutral loci (Taylor et al. 2000; current study) is indicative of a measurable level of gene flow (Fig. 3B).

Table 4. Individual assignment (IA) and mixed composition (MC) analyses for the Okanagan Lake kokanee 2010 sampling using outlier and neutral microsatellite loci.

Ecotype	Actual proportions	Proportion accurately assigned*		Estimated stock proportions*	
		Outlier (eight loci)†	Neutral (eight loci)‡	Outlier (eight loci)†	Neutral (eight loci)‡
Shore	12/28 (0.43)	**0.917**§	0.750	**0.451**	0.622
Stream	16/28 (0.57)	**0.944**	0.438	**0.549**	0.378

*IA accuracy and mixed stock proportions estimated according to algorithms implemented in ONCOR (Kalinowski et al. 2007).
†Expressed sequence tag (EST)-linked microsatellite loci.
‡Non-EST-linked microsatellite loci.
§Bolded values indicate highest assignment accuracy among data sets (1.0 = perfect assignment). For MC analyses, bolded values are closest to the actual mixture proportions.

Indeed, phenotypically identified stream spawner males have been observed within shore-spawning aggregations later in the season, suggesting that stream-to-shore straying may occur (P. Askey and J. Webster, personal observations). Nevertheless, the ability to distinguish stream- and shore-spawning kokanee remains important to fisheries managers, as ecotypes may be differentially impacted by alternative management regimes with regard to water use, spawning habitat protection/enhancement, and recreational harvest regulations. To date, IA and MC analyses based on neutral microsatellite loci have produced unacceptably high error rates for improving upon more traditional approaches to stock assessment, namely visual counts. Outlier loci, however, may constitute promising alternatives to more conventional genetic tools, as they may be directly linked to gene regions associated with ecotype divergence over very recent (ecological) timescales.

Outlier loci and divergent selection

In the current study, the eight loci identified as outliers represent candidate gene regions exhibiting signatures of divergent selection or regions that are genetically linked to loci associated with the distinct reproductive strategies displayed by Okanagan Lake kokanee ecotypes. While the nature of outlier detection allows for the likelihood of type I errors, it is possible to guard against false positives by relying on the comparison of multiple statistical tests for outlier behavior (Luikart et al. 2003; Bonin et al. 2006). Here, we employed three different algorithms for outlier detection that produced partially overlapping results (Figs 2, S1 and S2). Similarly, outlier behavior was explored using a variety of approaches including F_{ST}-based estimates of population structure (Tables 2 and 3) and Bayesian clustering analyses (Fig. 3).

Resource-based natural selection driving population divergence has been demonstrated in a range of species

displaying similar adaptations to ecological niches, such as the Lake whitefish species complex (Lu and Bernatchez 1999) and the rough periwinkle snail (Wilding et al. 2001). However, it remains uncertain whether divergent selection is driving kokanee ecotype differentiation as the specific genomic location and putative function of the eight outlier loci remain largely unknown. The inability to directly link candidate loci with fitness-related genotypes has proven to be a common limitation of population genomics studies conducted in nonmodel organisms, as recently reviewed by Bonin (2008) and Stinchcombe and Hoekstra (2008). In the case of kokanee, there are several lakes across the Pacific Northwest known to contain apparently similar sympatric ecotype pairs. Future research aimed at investigating whether the same candidate adaptive gene regions are common among all ecotype pairs may provide strong evidence for divergent selection. Despite our uncertainty regarding underlying mechanism, evidence for adaptive population divergence, both genetic (current study) and ecological (Dill 1996; Shepherd 2000), supports the need for unique conservation and management efforts for shore- and stream-spawning kokanee in Okanagan Lake to conserve maximum evolutionary potential (Fraser and Bernatchez 2001).

IA and MC analysis

The power of outlier loci to identify source ecotypes and differentiate between populations was superior to putatively neutral loci and non-EST-linked loci used in this study. Using simulation approaches, outlier loci were able to correctly self-assign individuals to their source ecotype in 94% and 90% of shore- and stream-spawning kokanee, respectively, outperforming previous studies based on neutral loci (Taylor et al. 2000). Similarly, MC analyses conducted with outlier loci were substantially improved, achieving 95.6% (shore) and 96.4% (stream) correct

estimation of simulated mixture samples containing 100% of each ecotype. Yet, these results, in and of themselves, do not alleviate concerns related to high-grading bias associated with marker selection. Anderson (2010) has recently highlighted the problems associated with using the same set of individuals for both selecting a panel of loci and assessing their informativeness. This bias can be particularly severe when sample sizes of individuals are small, large numbers of markers are used, and true differences between groups are small (Anderson 2010; Waples 2010), all of which may apply to the current study.

To cross-validate the informativeness of our panel of loci, we included 28 individuals that were not used in initial marker selection and were sampled over a very different time period (2010 rather than the 2007 baseline). In this case, the panel of eight outlier loci correctly assigned 92.9% of 2010 individuals to ecotype and outperformed other marker types in MC analyses (Table 4), levels on par to those revealed by simulation approaches based on the 2007 baseline alone (Table 3). Although these results provide a measure of confidence in the informativeness of the outlier loci for ecotype assignment, the small sample size does not allow us to completely rule out issues associated with high-grading bias. Enhanced sampling in future years will allow us to more comprehensively assess this potential bias.

Despite the improved accuracy achieved using outlier loci for IA and MC analyses, assignment accuracy still remains at unacceptably low levels at skewed proportions of shore- and stream-spawning kokanee (Table 3; Table S5). This is particularly relevant to Okanagan Lake fisheries management, as estimated visual enumerations of shore/stream kokanee more closely reflect the extreme proportions simulated in this study (approximately 85:15 in 2007 and 72:28 in 2008; P. Askey, unpublished data). A conventional approach for improving IA and MC accuracy and precision when using neutral loci would be to increase the number of loci (Bernatchez and Duchesne 2000). However, previous work by Scribner et al. (1998) and Winans et al. (2004) did not match these expectations and, in the current study, the eight outliers outperformed a substantially greater number of neutral loci ($n = 44$). Another option would be to increase the sample size of the reference baseline. In fact, initial analyses relying on a baseline of 96 individuals (48 shore-spawning and 48 stream spawners) produced substantially lower assignment accuracy (80% overall; data not shown) and greater deviation in MC analyses (83.7% and 84.3% probability of occurrence of shore- and stream spawners; data not shown). The modest 42% increase in baseline sampling between initial and reported results resulted in a substantial improvement across all analyses. As the bias in stock proportions is consistent (favoring 50:50) and

dependent on sample size, there may be additional value in future research that investigates improved/corrected statistical procedures for estimating stock proportions.

Management implications for the Okanagan Lake kokanee and other inland fisheries

Although many studies report levels of population genetic diversity, population structure and identify genes under selection, few bridge the gap between acquisition and application of this knowledge to stock assessment and management of populations under threat. Here, the ability to distinguish stream- and shore-spawning Okanagan Lake kokanee afforded by outlier loci and the concomitant lack of population structure among corresponding sampling localities suggest that management strategies should focus at the ecotype level, treating all spawning sites within ecotype as a single population. Still, there remains the possibility that our study design associated with the initial screening of loci for polymorphism may have led to ascertainment bias, potentially limiting our ability to detect variation among sampling sites within ecotype. We took the following precautions to minimize this potential bias by means of (i) including multiple individuals per spawning site in the original sample of 16 individuals for which polymorphism was assessed to retain the potential for detecting site-specific differences, and (ii) screening the larger sample ($n = 136$) at non-EST-linked microsatellite loci that have been used in other studies of Pacific salmonids to enable population genetic analyses at the site- and ecotype-specific levels.

At a finer level, kokanee fisheries management will be greatly improved by the enhanced specimen identification afforded by outlier loci. Specifically, Okanagan Lake kokanee has undergone a dramatic decline in abundance in recent decades owing to several concurrent anthropogenic factors including spawning habitat degradation, competition with introduced species, declines in nutrient loading, and angler impact (Shepherd 2000). However, it is very difficult to tease apart the relative role of these impacts (and prioritize future management actions) when data are observational, and there is no definitive method for ecotype identification within the lake. For example, annual acoustic and trawl monitoring only allow for an overall in-lake estimate of kokanee fry abundance. Alternatively, accurate methods for specimen identification would allow for estimates of ecotype-specific fry abundance and corresponding egg-to-fry survival rates. These early life survival rates are the key performance measures needed to assess impacts of water and habitat management in ecotype-specific spawning areas. Similarly, specimen identification is a critical input parameter for understanding fishery catch data and partitioning ecotype-specific harvest rates. Such

data could lead to specific spatial and temporal fishing regulations aimed at stock-specific sustainable harvest rates. Finally, specimen identification will allow for a basic calibration of the current visual shore-spawner survey index to an estimate of absolute abundance. The calibration can be performed by comparing the ratio of ecotypes in the lake, as estimated from genetic assignment of trawl samples, with that observed on the spawning grounds (as stream spawner abundance can be accurately estimated by way of spawner fences).

The apparent reproductive isolation observed in space and time (Table S1) provides an intuitive separation of kokanee ecotypes for fisheries managers. Yet, the recent postglacial divergence negates traditional genetic stock identification tools, a problem common among inland fisheries of the Pacific Northwest. The use of outlier loci constitutes a promising alternative to neutral loci for informing fisheries management of recently diverged stocks, particularly for enhancing stock assessment, long-term ecological monitoring, and evaluation of current management practices. Additionally, an outlier-based approach may be useful at even finer spatial and temporal scales for in-season monitoring, distinguishing drainage-specific stocks (Ackerman et al. 2011), and run types (Hess and Narum 2011; Hess et al. 2011) as recently suggested in anadromous sockeye and Chinook salmon, respectively. More generally, an outlier-based approach may prove informative for designating management units across a wide variety of terrestrial and aquatic systems, providing an avenue for directly integrating processes underlying biodiversity generation into conservation prioritization strategies.

Acknowledgements

Data archiving

Data for this study are available at Dryad: doi:10.5061/dryad.5bk66.

Literature cited

Ackerman, M. W., C., Habicht, and L. W. Seeb. 2011. Single-Nucleotide Polymorphisms (SNPs) under Diversifying Selection Provide Increased Accuracy and Precision in Mixed-Stock Analyses of Sockeye Salmon from the Copper River, Alaska. Transactions of the American Fisheries Society **140**:865–881.

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

Anderson, E. C., R. S. Waples, and S. T. Kalinowski. 2008. An improved method for predicting the accuracy of genetic stock identification. Canadian Journal of Fisheries and Aquatic Sciences **65**:1475–1486.

André, C., L. C. Larsson, L. Laikre, D. Bekkevold, J. Brigham, G. R. Carvalho, T. G. Dahlgren et al. 2011. Detecting population structure in a high gene-flow species, Atlantic herring (Clupea harengus): direct, simultaneous evaluation of neutral vs putatively selected loci. Heredity **106**:270–280.

Antao, T., A. Lopes, R. J. Lopes, A. Beja-Pereira, and G. Luikart. 2008. LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. BMC Bioinformatics **9**:323.

Banks, M. A., M. S. Blouin, B. A. Baldwin, V. K. Rashbrook, H. A. Fitzgerald, S. M. Blankenship, and D. Hedgecock. 1999. Isolation and inheritance of novel microsatellites in chinook salmon (Oncorhynchus tschawytscha). Journal of Heredity **90**:281–288.

BC Ministry of Environment Fish Inventory Data Queries 2010. http://a100.gov.bc.ca/pub/fidq/fissSpeciesSelect.do [cited 2010] (accessed on August 13, 2010).

Beacham, T. D., M. Lapointe, J. R. Candy, K. M. Miller, and R. E. Withler. 2004. DNA in action: rapid application of DNA variation to sockeye salmon fisheries management. Conservation Genetics **5**:411–416.

Beacham, T. D., B. McIntosh, and C. MacConnachie. 2005. Population structure and stock identification of sockeye salmon (Oncorhynchus nerka) in coastal lakes in British Columbia, Canada. Canadian Journal of Zoology-Revue Canadienne De Zoologie **83**:834–844.

Beacham, T. D., J. R. Candy, K. L. Jonsen, J. Supernault, M. Wetklo, L. T. Deng, K. M. Miller et al. 2006. Estimation of stock composition and individual identification of Chinook salmon across the Pacific Rim by use of microsatellite variation. Transactions of the American Fisheries Society **135**:861–888.

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

Beaumont, M. A., and D. J. Balding. 2004. Identifying adaptive genetic divergence among populations from genome scans. Molecular Ecology **13**:969–980.

Beaumont, M. A., and R. A. Nichols. 1996. Evaluating loci for use in the genetic analysis of population structure. Proceedings of the Royal Society of London Series B-Biological Sciences **263**:1619–1626.

Benson, G. 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Research **27**:573–580.

Bernatchez, L., and P. Duchesne. 2000. Individual-based genotype analysis in studies of parentage and population assignment: how many loci, how many alleles? Canadian Journal of Fisheries and Aquatic Sciences **57**:1–12.

Black, W. C, C. F. Baer M. F. Antolin, and N. M. DuTeau. 2001. Population genomics: genome-wide sampling of insect populations. Annual Review of Entomology **46**:441–469.

Bonin, A. 2008. Population genomics: a new generation of genome scans to bridge the gap with functional genomics. Molecular Ecology **17**:3583–3584.

Bonin, A., P. Taberlet, C. Miaud, and F. Pompanon. 2006. Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). Molecular Biology and Evolution **23**:773–783.

Brownstein, M. J., J. D. Carpten, and J. R. Smith. 1996. Modulation of non-templated nucleotide addition by *Taq* DNA polymerase: primer modifications that facilitate genotyping. BioTechniques **20**:1004.

Carvalho, G. R., and L. Hauser. 1994. Molecular genetics and the stock concept in fisheries. Reviews in Fish Biology and Fisheries **4**:326–350.

Dill, P. 1996. A Study of Shore-Spawning Kokanee Salmon (*Oncorhynchus nerka*) at Bertram Creek Park, Okanagan Lake, BC 1992–1996. BC Ministry of Environment, Penticton, BC.

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

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

Excoffier, L., G. Laval, and S. Schneider. 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evolutionary Bioinformatics **1**:47–50.

Ferris, S. D., and W. J. Berg. 1987. The utility of mitochondrial DNA in fish genetics and fishery management. In N. Ryman, and F. M. Utter, eds. Population Genetics and Fishery Management, pp. 227–229. University of Washington Press, Seattle.

Foll, M., and O. Gaggiotti. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a bayesian perspective. Genetics **180**:977–993.

Fraser, D. J., and L. Bernatchez. 2001. Adaptive evolutionary conservation: towards a unified concept for defining conservation units. Molecular Ecology **10**:2741–2752.

Freamo, H., P. O'Reilly, P. R. Berg, S. Lien, and E. G. Boulding. 2011. Outlier SNPs show more genetic structure between two Bay of Fundy metapopulations of Atlantic salmon than do neutral SNPs. Molecular Ecology Resources **11**:254–267.

Grimholt, U., F. Drablos, S. M. Jorgensen, B. Hoyheim, and R. J. M. Stet. 2002. The major histocompatibility class I locus in Atlantic salmon (*Salmo salar L.*): polymorphism, linkage analysis and protein modelling. Immunogenetics **54**:570–581.

Guo, S. W., and E. A. Thompson. 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics **48**:361–372.

Hauser, L., and G. R. Carvalho. 2008. Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. Fish and Fisheries **9**:333–362.

Hess, J. E., and S. R. Narum. 2011. Single-Nucleotide Polymorphism (SNP) Loci Correlated with Run Timing in Adult Chinook Salmon from the Columbia River Basin. Transactions of the American Fisheries Society **140**:855–864.

Hess, J. E., A. P. Matala, and S. R. Narum. 2011. Comparison of SNPs and microsatellites for fine-scale application of genetic stock identifi-

cation of Chinook salmon in the Columbia River Basin. Molecular Ecology Resources **11**:137–149.

Hilborn, R., B. G. Bue, and S. Sharr. 1999. Estimating spawning escapements from periodic counts: a comparison of methods. Canadian Journal of Fisheries and Aquatic Sciences **56**:888–896.

Hohenlohe, P. A., S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson, and W. A. Cresko. 2010. Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. Plos Genetics **6**:e1000862.

Holt, K. R., and S. P. Cox. 2008. Evaluation of visual survey methods for monitoring pacific salmon (*Oncorynchus* spp.) escapement in relation to conservation guidelines. Canadian Journal of Fisheries and Aquatic Sciences **65**:212–226.

Hubisz, M. J., D. Falush, M. Stephens, and J. K. Pritchard. 2009. Inferring weak population structure with the assistance of sample group information. Molecular Ecology Resources **9**:1322–1332.

Kalinowski, S. T., K. R. Manlove, and M. L. Taper. 2007. ONCOR: A Computer Program for Genetic Stock Identification. Department of Ecology, Montana State University, Bozeman, MT. http://www.montana.edu/kalinowski/Software/ONCOR.htm.

Lin, J., T. P. Quinn, R. Hilborn, and L. Hauser. 2008. Fine-scale differentiation between sockeye salmon ecotypes and the effect of phenotype on straying. Heredity **101**:341–350.

Lu, G. Q., and L. Bernatchez. 1999. Correlated trophic specialization and genetic divergence in sympatric lake whitefish ecotypes (*Coregonus clupeaformis*): support for the ecological speciation hypothesis. Evolution **53**:1491–1505.

Luikart, G., P. R. England, D. Tallmon, S. Jordan, and P. Taberlet. 2003. The power and promise of population genomics: from genotyping to genome typing. Nature Reviews Genetics **4**:981–994.

Miller, K. M., R. E. Withler, and T. D. Beacham. 1997. Molecular evolution at MHC genes in two populations of chinook salmon *Oncorhynchus tshawytscha*. Molecular Ecology **6**:937–954.

Mitchell-Olds, T., and J. Schmitt. 2006. Genetic mechanisms and evolutionary significance of natural variation in *Arabidopsis*. Nature **441**:947–952.

Morris, D. B., K. R. Richard, and J. M. Wright. 1996. Microsatellites from rainbow trout (*Oncorhynchus mykiss*) and their use for genetic study of salmonids. Canadian Journal of Fisheries and Aquatic Sciences **53**:120–126.

Nelson, R. J., and T. D. Beacham. 1999. Isolation and cross species amplification of microsatellite loci useful for study of Pacific salmon. Animal Genetics **30**:228–229.

Ng, S. H. S., A. Chang, G. D. Brown, B. F. Koop, and W. S. Davidson. 2005. Type I microsatellite markers from Atlantic salmon (*Salmo salar*) expressed sequence tags. Molecular Ecology Notes **5**:762–766.

Olsen, J. B., J. K. Wenburg, and J. Bentzen. 1996. Semi-automated multilocus genotyping of Pacific salmon (*Oncorhynchus* spp.) using microsatellites. Molecular Marine Biology and Biotechnology **5**:259–272.

O'Malley, K. G., M. D. Camara, and M. A. Banks. 2007. Candidate loci reveal genetic differentiation between temporally divergent migratory runs of Chinook salmon (*Oncorhynchus tshawytscha*). Molecular Ecology **16**:4930–4941.

O'Reilly, P. T., L. C. Hamilton, S. K. McConnell, and J. M. Wright. 1996. Rapid analysis of genetic variation in Atlantic salmon (*Salmo salar*) by PCR multiplexing of dinucleotide and tetranucleotide microsatellites. Canadian Journal of Fisheries and Aquatic Sciences **53**:2292–2298.

Pompanon, F., A. Bonin, E. Bellemain, and P. Taberlet. 2005. Genotyping errors: causes, consequences and solutions. Nature Reviews Genetics **6**:847–859.

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

Ramstad, K. M., C. A. Woody, and F. W. Allendorf. 2010. Recent local adaptation of sockeye salmon to glacial spawning habitats. Evolutionary Ecology **24**:391–411.

Rannala, B., and J. L. Mountain. 1997. Detecting immigration by using multilocus genotypes. Proceedings of the National Academy of Sciences of the United States of America **94**:9197–9201.

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

Rexroad, C. E., M. F. Rodriguez, I. Coulibaly, K. Gharbi, R. G. Danzmann, J. DeKoning, R. Phillips *et al.* 2005. Comparative mapping of expressed sequence tags containing microsatellites in rainbow trout (*Oncorhynchus mykiss*). BMC Genomics **6**:54.

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

Rozen, S., and H. Skaletsky. 2000. Primer3 on the WWW for general users and for biologist programmers. Methods in Molecular Biology **132**:365.

Salem, M., C. E. Rexroad, J. N. Wang, G. H. Thorgaard, and J. B., Yao. 2010. Characterization of the rainbow trout transcriptome using Sanger and 454-pyrosequencing approaches. BMC Genomics **11**:564.

Schuelke, M. 2000. An economic method for the fluorescent labeling of PCR fragments. Nature Biotechnology **18**:233–234.

Scribner, K. T., J. R. Gust, and R. L. Fields. 1996. Isolation and characterization of novel salmon microsatellite loci: cross-species amplification and population genetic applications. Canadian Journal of Fisheries and Aquatic Sciences **53**:833–841.

Scribner, K. T., P. A. Crane, W. J. Spearman, and L. W. Seeb. 1998. DNA and allozyme markers provide concordant estimates of population differentiation: analyses of US and Canadian populations of Yukon River fall-run chum salmon (*Oncorhynchus keta*). Canadian Journal of Fisheries and Aquatic Sciences **55**:1748–1758.

Shepherd, B. G. 2000. A case history: the kokanee stocks of Okanagan Lake. In L. M. Darling, ed. Proceedings of a Conference on the Biology and Management of Species and Habitats at Risk, Kamloops, B.C., 15–19 Feb., 1999, Vol. 2, pp. 520. B.C. Ministry of Environment, Lands and Parks, Victoria, BC and University College of the Cariboo, Kamloops, BC.

Small, M. P., R. E. Withler, and T. D. Beacham. 1998. Population structure and stock identification of British Columbia coho salmon, Oncorhynchus kisutch, based on microsatellite DNA variation. Fishery Bulletin **96**:843–858.

Stinchcombe, J. R., and H. E. Hoekstra. 2008. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. Heredity **100**:158–170.

Taylor, E. B., S. Harvey, S. Pollard, and J. Volpe. 1997. Postglacial genetic differentiation of reproductive ecotypes of kokanee *Oncorhynchus nerka* in Okanagan Lake, British Columbia. Molecular Ecology **6**:503–517.

Taylor, E. B., A. Kuiper, P. M. Troffe, D. Hoysak, and S. Pollard. 2000. Variation in developmental biology and microsatellite DNA in reproductive ecotypes of kokanee, *Oncorhynchus nerka*: implications

for declining populations in a large British Columbia lake. Conservation Genetics **1**:231–249.

Van Oosterhout, C., W. F. Hutchinson, D. P. M. Wills, and P. Shipley. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes **4**:535–538.

Vasemägi, A., R. Gross, T. Paaver, M. L. Koljonen, M. Saisa, and J. Nilsson. 2005a. Analysis of gene associated tandem repeat markers in Atlantic salmon (*Salmo salar L.*) populations: implications for restoration and conservation in the Baltic Sea. Conservation Genetics **6**:385–397.

Vasemägi, A., J. Nilsson, and C. R. Primmer. 2005b. Seventy-five EST-linked Atlantic salmon (*Salmo salar L.*) microsatellite markers and their cross-amplification in five salmonid species. Molecular Ecology Notes **5**:282–288.

Vitalis, R., K. Dawson, and P. Boursot. 2001. Interpretation of variation across marker loci as evidence of selection. Genetics **158**:1811–1823.

Vitalis, R., K. Dawson, P. Boursot, and K. Belkhir. 2003. DetSel 1.0: a computer program to detect markers responding to selection. Journal of Heredity **94**:429–431.

Waples, R. S. 2010. High-grading bias: subtle problems with assessing power of selected subsets of loci for population assignment. Molecular Ecology **19**:2599–2601.

Ward, R. D. 2000. Genetics in fisheries management. Hydrobiologia **420**:191–201.

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

Westgaard, J. I., and S. E. Fevolden. 2007. Atlantic cod (*Gadus morhua L.*) in inner and outer coastal zones of northern Norway display divergent genetic signature at non-neutral loci. Fisheries Research **85**:306–315.

Wilding, C. S., R. K. Butlin, and J. Grahame. 2001. Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. Journal of Evolutionary Biology **14**:611–619.

Winans, G. A., M. M. Paquin, D. M. Van Doornik, B. M. Baker, P. Thornton, D. Rawding, A. Marshall *et al.* 2004. Genetic stock identification of steelhead in the Columbia River Basin: an evaluation of different molecular markers. North American Journal of Fisheries Management **24**:672–685.

Wright, J. J., K. P. Lubieniecki, J. W. Park, S. H. S. Ng, R. H. Devlin, J. Leong, B. F. Koop *et al.* 2008. Sixteen Type 1 polymorphic microsatellite markers from Chinook salmon (*Oncorhynchus tshawytscha*) expressed sequence tags. Animal Genetics **39**:84–85.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. Genomic scan results as implemented in BAYESCAN 2.0.

Figure S2. Genomic scan results as implemented in LOSITAN Selection Workbench.

Table S1. Kokanee ecotype variation in Okanagan Lake, BC including environmental, morphological, life history and behavioural characteristics.

Table S2. Description of 243 genetic markers screened for use in this study listing primer name, primer sequence, marker type and references for each locus.

Table S3. Description of 57 microsatellite loci comprising the baseline dataset of this study listing primer name, working name, Genbank

accession, microsatellite motif, annealing temperature (T_a), range of allele size and locus behaviour for each locus.

Table S4. Expanded results of genetic variation at 57 polymorphic microsatellite loci in three shore spawning localities, four stream spawning localities and total populations in two ecotypes of Okanagan Lake kokanee.

Table S5. Performance of outlier and neutral loci for mixed composition analyses based on simulation of a range of realistic fishery sce-

narios and according to actual and simulated baselines, including standard deviations.

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