

Lack of parallel genetic patterns underlying the repeated ecological divergence of beach and stream-spawning kokanee salmon

K. K. FRAZER & M. A. RUSSELLO

Department of Biology, The University of British Columbia, Okanagan Campus, Kelowna, BC, Canada

Keywords:

adaptation;
expressed sequence tag-linked
microsatellites;
life history evolution;
population genomics;
sockeye salmon.

Abstract

Recent progress in methods for detecting adaptive population divergence *in situ* shows promise for elucidating the conditions under which selection acts to generate intraspecific diversity. Rapid ecological diversification is common in fishes; however, the role of phenotypic plasticity and adaptation to local environments is poorly understood. It is now possible to investigate genetic patterns to make inferences regarding phenotypic traits under selection and possible mechanisms underlying ecotype divergence, particularly where similar novel phenotypes have arisen in multiple independent populations. Here, we employed a bottom-up approach to test for signatures of directional selection associated with divergence of beach- and stream-spawning kokanee, the obligate freshwater form of sockeye salmon (*Oncorhynchus nerka*). Beach- and stream-spawners co-exist in many post-glacial lakes and exhibit distinct reproductive behaviours, life-history traits and spawning habitat preferences. Replicate ecotype pairs across five lakes in British Columbia, Canada were genotyped at 57 expressed sequence tag-linked and anonymous microsatellite loci identified in a previous genome scan. Fifteen loci exhibited signatures of directional selection (high F_{ST} outliers), four of which were identified in multiple lakes. However, the lack of parallel genetic patterns across all lakes may be a result of: 1) an inability to detect loci truly under selection; 2) alternative genetic pathways underlying ecotype divergence in this system; and/or 3) phenotypic plasticity playing a formative role in driving kokanee spawning habitat differences. Gene annotations for detected outliers suggest pathogen resistance and energy metabolism as potential mechanisms contributing to the divergence of beach- and stream-spawning kokanee, but further study is required.

Introduction

There has been recent progress in our understanding of the formative role of selection driving population divergence, even in the face of on-going gene flow (Via & West, 2008; Wolf *et al.*, 2010; Feder *et al.*, 2012). Technological advances in data collection paired with the analytical framework of population genomics have

sparked increasing refinement of our ability to detect adaptive population divergence *in situ*, even without prior knowledge of the phenotypes under selection or the availability of extensive genome maps (Luikart *et al.*, 2003; Nielsen, 2005; Storz, 2005; Stinchcombe & Hoekstra, 2008). Yet, many questions remain unresolved regarding the number, arrangement and genome-wide influence of ecologically relevant loci under selection that are required to overcome the homogenizing effects of gene flow to establish the reproductive barriers ultimately required for speciation (Via, 2009). This recognition has led to an increasing number of studies of locally diverging populations where whole-genome isolation has not been achieved (Turner *et al.*,

Correspondence: Michael Russello, Department of Biology, The University of British Columbia, Okanagan Campus, 3333 University Way, Kelowna, BC V1V 1V7, Canada. Tel.: +1 250 807 8762; fax: +1 250 807 8005; e-mail: michael.russello@ubc.ca

2005; Savolainen *et al.*, 2006; Nosil *et al.*, 2009; Stölting *et al.*, 2013) and the development of models describing the genomics of speciation-with-gene flow (Feder *et al.*, 2012).

Salmonids offer ample opportunities for studying the genetic basis of local adaptation within recently diverged populations where gene flow persists. They inhabit a wide range of habitats and exhibit extensive phenotypic diversity, often predicted to be the result of local adaptation (reviewed in Hendry & Stearns, 2004; Quinn, 2005). In many cases, precise natal homing has led to isolation and subsequent emergence of new ecological forms in multiple, independent populations over a short period of time (< 10 000 years; Taylor, 1991; Dittman & Quinn, 1996). Currently, a better understanding of the mechanisms driving ecotype divergence is still needed. In the absence of conspicuous drivers, a bottom-up approach may be fruitful for detecting adaptive variation and identifying traits that are strongly linked to fitness in particular environments (e.g. behavioural, life-history or physiological traits).

Parallel patterns of genotypic and phenotypic divergence can provide strong evidence for the action of natural selection, allowing insights into the underlying processes involved in speciation (Butlin, 2010; Schluter *et al.*, 2010). Post-glacial fishes have played a prominent role within such studies, including parallel adaptation and phenotypic evolution in the three-spine stickleback (Colosimo *et al.*, 2004; Hohenlohe *et al.*, 2010; Deagle *et al.*, 2012), and divergence of lake whitefish morphotypes (Rogers & Bernatchez, 2005). Variation in spawning habitat preference and reproductive strategies has also been observed in other post-glacial fishes, particularly sockeye salmon (*Oncorhynchus nerka*; Burgner, 1991; Quinn, 1999). However, processes underlying this divergence can be difficult to identify in populations that exploit distinct environments, but show little difference in morphological features.

Kokanee is an obligate freshwater form of sockeye salmon, with parallel evolutionary origins from anadromous sockeye throughout the North Pacific since the last (Wisconsinan) glaciation (~15 000 years ago; Rickler, 1940; Taylor *et al.*, 1996). There are two reproductive ecotypes of kokanee that alternatively spawn in inlet streams or along lake beaches (Burgner, 1991). Stream-spawners exhibit a reproductive strategy more typical of anadromous sockeye, whereas beach-spawners form large spawning aggregations along the shoreline (Shephard, 2000) weeks or months following stream-spawners likely due to the varying temperature regimens of the spawning substrate (Burgner, 1991; Burger *et al.*, 1995; Winans *et al.*, 2003). In that regard, beach habitats are typically deeper (0.1–10 metres) and warmer, characterized by low flow regimes and larger rocky substrates (Hassemer & Rieman, 1981; Dill,

1996). Usually beach-spawners do not exhibit courting behaviour, mate selection, redd excavation or nest defence (Dill, 1996; Ashley *et al.*, 1998; Shephard, 2000); however, a targeted behavioural study has not been published to date. In some lakes, the ecotypes co-occur with each other and, in other cases, with anadromous sockeye as well (Burgner, 1991). Although kokanee ecotypes often display different reproductive traits and experience varying incubation environments associated with their corresponding stream and shoreline spawning areas (Ashley *et al.*, 1998; Shephard, 2000; Andrusak & Andrusak, 2011), they show no morphological or ecological differences between emergence and maturation (Taylor *et al.*, 1997, 2000; Winans *et al.*, 2003).

The specific fitness-related traits involved in generating reproductive barriers among kokanee ecotypes are currently unknown, but are likely related to habitat preference, mating behaviour, energy metabolism and/or life-history ecology that increase reproductive success in adult spawners and/or survival at early life stages (Lecomte & Dodson, 2004). Previous studies have demonstrated weak neutral differentiation among kokanee ecotypes in Okanagan Lake in British Columbia, Canada (Taylor *et al.*, 1997, 2000). More recently, a population genomics study of Okanagan Lake kokanee screened over 11 000 expressed sequence tag (EST)-linked microsatellites, resulting in a panel of 57 markers (49 EST-linked and 8 anonymous; Russello *et al.*, 2012). Eight loci exhibited patterns of genetic variation that deviated from expectations under a neutral model of evolution (i.e. outlier loci) representing candidate gene regions exhibiting signatures of selection or regions that are genetically linked to selected loci. However, this study focused on a single lake and was unable to validate the role of selection in ecotype divergence.

Here, we tested for evidence of directional selection driving the adaptive divergence of beach- and stream-spawning kokanee salmon in five lakes in British Columbia, Canada. We used the panel of EST-linked microsatellites developed in Russello *et al.* (2012) and four conceptually different outlier detection tests to reduce false positives. Neutral loci were used to determine whether ecotypes represent distinct genetic lineages (i.e. monophyletic groups with a common ancestor) or have arisen multiple times (i.e. polyphyletic origins). If a common genetic basis (e.g. same outlier loci) can be identified across multiple, independent ecotype pairs, we can infer preliminary evidence for natural selection driving parallel divergence in this system (Johannesson, 2001; McKinnon *et al.*, 2004; Hendry, 2009). Based on this approach, we describe a number of candidate loci and infer potential mechanisms associated with the divergence of beach- and stream-spawning kokanee.

Materials and methods

Study sites and sample collection

Between 2007 and 2011, we sampled kokanee from five British Columbian lakes where beach- and stream-spawners co-exist: Duncan, Kootenay, Okanagan, Wood and Tchesinkut Lakes (Fig. 1). These lakes have minimal or no documented stocking history, little or no overlap with anadromous sockeye (Table S1) and are distributed across two major drainage systems (Fraser and Columbia Rivers).

A total of 536 live and recently moribund individuals caught in gillnets and dip nets were sampled post-spawn at two to seven different spawning sites (Table S2) within each lake during the peak of their respective spawning periods. The adipose fin or operculum tissue was collected from each fish and preserved in 2 mL vials containing 100% ethanol and stored at $-20\text{ }^{\circ}\text{C}$ for subsequent DNA analysis.

Data collection

Total genomic DNA was extracted with the NucleoSpin Tissue kit (Macherey Nagel) according to the manufacturer's suggested protocol for 96-well plates. Samples collected from Duncan, Kootenay, Tchesinkut and Wood Lakes were amplified via polymerase chain



Fig. 1 Geographic locations of the five lakes sampled in British Columbia, Canada. Tchesinkut Lake is part of the Fraser River Drainage system. Okanagan and Wood Lakes in the Okanagan River Chain, and Duncan and Kootenay Lakes in the Kootenay River Chain, all of which are tributary to the Columbia River in Washington, USA.

reaction (PCR) at 49 EST-linked microsatellite loci and eight anonymous microsatellite loci. These loci were selected because they demonstrated consistent amplification and polymorphism among ecotypes in Okanagan Lake kokanee in a previous study that screened 11 488 EST-linked loci, of which 243 were characterized (Russello *et al.*, 2012).

For each PCR, we combined 1.25 μL of 10 \times PCR buffer, 1.25 μL of 2 mM dNTP mix, 0.5 μL of 1 μM forward primer, 0.5 μL of 10 μM M13 fluorescent-labelled primer, 0.5 μL of 10 μM reverse primer, 0.5 units of *Taq* polymerase and 20–80 ng of DNA template for a total reaction volume of 12.5 μL . Forward primers were modified to include an M13 tail [5'-TCCCAGTCAC GACGT -3'] (Schuelke, 2000), which allowed for a fluorescent-labelled universal primer (6-FAM, VIC, NED, or PET; Applied Biosystems) to be incorporated into PCR amplicons. Reverse primers were modified to include a GTTT-tail to improve scoring quality (Brownstein *et al.*, 1996). All reactions used KAPA *Taq* DNA polymerase (Kapa Biosystems), except for markers *EVI70*, *OMM5008*, *OMM5067* and *One14*. AmpliTaq Gold DNA polymerase (Applied Biosystems) was used to promote amplification of these markers. Amplification was achieved using a touchdown cycling programme, which started with an initial denaturation at 94 $^{\circ}\text{C}$ for 2 min (or 10 min for reactions using AmpliTaq Gold), followed by 20 cycles at 94 $^{\circ}\text{C}$ for 30 s, 60 $^{\circ}\text{C}$ for 30 s and 72 $^{\circ}\text{C}$ for 30 s, with the annealing temperature decreasing 0.5 $^{\circ}\text{C}$ per cycle. For 15 more cycles, the annealing temperature was held at 50 $^{\circ}\text{C}$ and final extension was 2 min at 72 $^{\circ}\text{C}$. DNA fragments were analysed on an Applied Biosystems 3130XL DNA automated sequencer using the GS500 LIZ size standard to determine fragment length. Two independent investigators manually scored all alleles in GENEMAPPER 4.0 (Applied Biosystems). New genotypic data were then combined with that previously collected for the Okanagan Lake kokanee by Russello *et al.* (2012).

Data quality and definition of genetic units

Loci were tested for null and false alleles using MICROCHECKER (Van Oosterhout *et al.*, 2004). Deviation from Hardy–Weinberg Equilibrium (HWE) was assessed using the Markov chain Monte Carlo (MCMC) approximation of Fisher's exact test (using 1000 batches with 1000 iterations; Guo & Thompson, 1992), and linkage disequilibrium (LD) was assessed for all possible marker combinations using simulated exact tests as implemented in GENEPOP 3.3 (Raymond & Rousset, 1995). A sequential Bonferroni procedure (Rice, 1989) was used to correct for multiple comparisons ($P < 0.05$) in tests of LD and departure from HWE. Because selection violates a critical assumption of HWE and can generate patterns of LD, beach- and stream-spawners were evaluated separately in each lake and a locus was only

removed if such patterns were detected in both ecotypes for at least one lake. Pairwise F_{ST} comparisons were conducted in ARLEQUIN 3.5 (Excoffier & Lischer, 2010) using the anonymous microsatellites to ensure that spawning sites at similar habitats within a single lake were not significantly differentiated prior to pooling individuals by ecotype for all downstream analyses.

Outlier locus detection

In each lake, putative outliers were identified based on patterns of genetic diversity (lnRH; Kauer *et al.*, 2003), differentiation (DetSel; Vitalis *et al.*, 2003) or both (Lositan Selection Workbench; Antao *et al.*, 2008; BayeScan; Foll & Gaggiotti, 2008). The lnRH test (Kauer *et al.*, 2003) detects loci departing from the genomic average by calculating the ratio of expected heterozygosity (H_E) for each locus. Monomorphic loci were assumed to have one allele that differed from the others to avoid dividing by zero. Estimates were standardized to obtain a mean of 0 and a standard deviation of 1, so that 90%, 95% and 99% of the loci were expected to have values within ± 1.64 , ± 1.96 and ± 2.58 , respectively. Loci with values outside these boundaries were considered putative outliers at the respective levels. Second, we employed the probabilistic approach implemented in DETSEL 1.0 (Vitalis *et al.*, 2003), which assumes a pure divergence model. Coalescent simulations generated a joint distribution of expected F_{ST} values under a neutral model. For post-glacial lakes, this null distribution (i.e. confidence envelope) was generated using the following parameters following Russello *et al.* (2012): $N_e = 500, 1000, 10000$; $N_o = 500$; $T_o = 50, 100, \text{ or } 1000$; $\mu = 0.0001, 0.00001$; $t = 100$. For Duncan Lake, we used modified parameters ($N_e = 100, 500, \text{ and } 5000$; $N_o = 50$; $T_o = 5, 10, \text{ or } 20$; $t = 1$) given its recent formation 47 years ago following the construction of Duncan Dam. In two-dimensional arrays of 50×50 square cells, outliers were identified as loci falling outside of the confidence envelope with an empirical P -value of $> 90\%$, $> 95\%$ and $> 99\%$ (Vitalis *et al.*, 2001). Finally, the F_{DIS}T2 approach of Beaumont & Nichols (1996) was implemented in both LOSITAN and BAYESCAN. The LOSITAN test (Antao *et al.*, 2008) assumes an island model of migration and uses coalescent simulations to generate the null distribution of F_{ST} values. After an initial 50 000 simulations, loci outside the 95% confidence interval (CI) were removed to calculate a more accurate mean neutral F_{ST} , which was forced in a second set of 50 000 simulations to calculate the probability of each locus being under selection at 90%, 95% and 99% CIs. In BAYESCAN 2.0 (Foll & Gaggiotti, 2008), the F_{DIS}T2 approach is modified to use Bayesian-based simulations. For each locus, posterior probabilities were estimated for two alternative models (one with and one without locus-specific effects) using a reversible jump MCMC. For the MCMC algorithm, 20 pilot runs of 5000

iterations were conducted followed by 100 000 iterations with a burn-in of 50 000. Posterior probabilities of > 0.76 , > 0.91 and > 0.97 were interpreted as 'substantial', 'strong' and 'very strong' support for the action of selection, respectively (Foll & Gaggiotti, 2008).

Potential false outliers were eliminated based on a lack of concordance across tests in each lake (Vasemägi & Primmer, 2005; Bonin *et al.*, 2006; Foll & Gaggiotti, 2008; Shikano *et al.*, 2010) rather than using a Bonferroni correction or a false discovery rate (Benjamini & Hochberg, 1995; Schlotterer, 2003). Therefore, only loci detected by two or more tests above the 90% CI in a single lake are referred to as 'outliers' hereon and included in the 'outlier' data set. Any locus detected by multiple tests in two or more lakes was also included in the 'repeat outlier' data set. Finally, any locus not detected by any test in any lake, and that was not monomorphic in any lake, was included in the 'neutral' data set.

Population genetic analyses

Mean number of alleles (N_A), observed heterozygosity (H_o), expected heterozygosity (H_e ; Nei, 1987) and the percentage of polymorphic loci were calculated for each ecotype in each lake using GENALEX version 6.2 (Peakall & Smouse, 2006). We tested for evidence of a recent demographic contraction within each lake because some populations have undergone substantive declines (Shephard, 2000) and bottlenecks can significantly bias outlier detection results (Teshima *et al.*, 2006). We used the neutral loci and the one-tailed Wilcoxon (Cornuet & Luikart, 1996) and mode-shift indicator tests implemented in BOTTLENECK (Piry *et al.*, 1999) under a two-phase model (TPM) of mutation (80% stepwise mutation model; Dirienzo *et al.*, 1994). We also calculated M -ratio using M_P_VAL.EXE (Garza & Williamson, 2001), a marker mutation rate μ of 5×10^{-4} and a pre-bottleneck N_e ranging from 500 to 5000. These parameters resulted in a value of θ ($4 N_e \mu$) ranging from 1 to 10 and 3.5 base steps for multistep mutations. For the TPM, the amount of single step mutations was $p_s = 0.9$ as recommended by Garza & Williamson (2001).

To ensure that each ecotype pair could be considered an independent sample, we tested for correspondence of geographically separated populations and ecotypes as discrete genetic units at neutral loci using STRUCTURE 2.3.3 (Pritchard *et al.*, 2000). Run length was set to 1 000 000 MCMC steps after a burn-in period of 500 000 using correlated allele frequencies under an admixture model using the LOCPRIOR option (i.e. ecotype and lake were used as prior information to assist the clustering when structuring is relatively weak; Hubisz *et al.*, 2009). Clusters (K) from 1 to 10 were evaluated, with 5 iterations per K value. The most likely number of clusters was determined using the method of Pritchard *et al.* (2000) and the ΔK approach (Evanno *et al.*,

2005) as implemented in STRUCTURE HARVESTER (Earl & vonHoldt, 2012). The level of genetic similarity within and among ecotype pairs from the five lakes was visualized using a discriminant analysis of principal components (DAPC) conducted in R (R Development Core Team, 2012) with the ADEGENET package version 1.3.0 (Jombart *et al.*, 2010). DAPC is a model-free approach that produces synthetic variables to maximize the among-population variation visible in two dimensions whereas minimizing the variation within predefined groups. We retained 90% of the variation and plotted all individuals on the first two principal components.

Divergence at neutral and outlier loci

Because loci truly under selection are expected to show distinct patterns of genetic variation compared with neutral loci (Storz, 2005), global and lake-by-lake estimates of population differentiation based on all outliers, outliers identified in each respective lake (referred to as 'lake-specific outliers'), repeat outliers and neutral loci data sets were compared. An analysis of molecular variance (AMOVA; Excoffier *et al.*, 1992) was used to assess the hierarchical organization of genetic variation within and among ecotypes and across lakes, and pairwise F_{ST} comparisons (Weir & Cockerham, 1984) were used to assess genetic differentiation among ecotypes within each lake, as implemented in ARLEQUIN 3.5 (Excoffier & Lischer, 2010). The 95% CIs were estimated by bootstrapping over all loci. We used the probabilistic approach of STRUCTURE 2.3.3 to determine the number of genetically distinct clusters within each lake. As above, run length was set to 1 000 000 MCMC steps after a burn-in period of 500 000 using correlated allele frequencies under an admixture model using the LOCPRIOR option (i.e. ecotype). Clusters (K) from 1 to 10 were evaluated, with 5 iterations per K value. The most likely number of clusters was determined using the method of Pritchard *et al.* (2000) and the ΔK approach (Evanno *et al.*, 2005) as implemented in STRUCTURE HARVESTER (Earl & vonHoldt, 2012).

Outlier annotation

Sequence similarity searches were conducted for all outlier loci within the consortium for Genomics Research on All Salmon (cGRASP) and the salmonidae database in BLAST to identify candidate genes and investigate functional annotations (Salem *et al.*, 2010). Avoiding the repeat region, primers were designed for repeat outliers and two strong outliers (Table S3) using Primer3 (Rozen & Skaletsky, 2000). Approximately 300–600 basepairs of the candidate genes were Sanger sequenced for four stream- and four beach-spawning individuals. Sequences were aligned with the reference nucleotide sequence in GENEIOUS 5.6 (Drummond *et al.*, 2011) and screened for single nucleotide polymorphisms (SNPs). If SNPs were identified, the locus was sequenced at an additional eight beach- and eight stream-spawners to better resolve population-level patterns of differentiation.

Results

Data quality and defining genetic units

Eleven individuals with > 20% missing data were removed from the data set. Three loci that exhibited false alleles (*EV103*, *EV626* and *One109*) and deviations from HWE (*EV626* and *One109*) in Okanagan Lake were removed. Three pairs of loci were in LD in Okanagan and Wood Lake (*OMM5099* & *Ots29*, *Ca687* & *EV712* and *Ca613* & *Ots14*), so the locus with greater missing data within each pair was removed from the data set (*Ca613*, *Ca687*, and *Ots29*). Locus *Ssa85* was also removed due to inconsistent genotypic data. As a result, 525 individuals and 50 loci were retained in the final data set, which had 1.4% missing data overall. No lake showed evidence for a recent population bottleneck regardless of approach (heterozygote excess, mode-shift, M -ratio; Table 1). Spawning sites of the same habitat type were pooled in each lake, as estimates of neutral genetic differentiation within each group were low on average

Table 1 Estimates of population genetic parameters for each ecotype within each lake using all 50 loci including: sample size (N), mean number of alleles per locus (N_A), range in the number of alleles, mean expected heterozygosity (H_E), mean observed heterozygosity (H_o) and the percentage of polymorphic loci (%poly).

Lake	Ecotype	N	N_A	N_A range	H_E	H_o	% poly	Wilcoxon TPM	Mode-shift	M -ratio (P -value)
Okanagan	Beach	72	7.451	2–23	0.507	0.513	100%	ns	Normal	0.725 (0.377)
	Stream	69	7.196	2–23	0.516	0.507	100%	ns	Normal	0.783 (0.897)
Wood	Beach	39	5.196	2–19	0.505	0.525	100%	ns	Normal	0.709 (0.725)
	Stream	38	5.137	1–17	0.526	0.535	96%	ns	Normal	0.675 (0.392)
Duncan	Beach	47	6.000	1–21	0.515	0.527	90%	ns	Normal	0.756 (0.901)
	Stream	50	6.280	1–22	0.529	0.538	92%	ns	Normal	0.727 (0.644)
Kootenay	Beach	27	4.320	1–14	0.438	0.449	92%	ns	Normal	0.739 (0.982)
	Stream	41	5.220	1–17	0.467	0.474	98%	ns	Normal	0.776 (0.980)
Tchesinkut	Beach	48	3.451	1–13	0.326	0.337	80%	ns	Normal	0.874 (1.00)
	Stream	94	3.843	1–14	0.317	0.326	80%	ns	Normal	0.811 (0.948)

($F_{ST} = 0.008$), ranging from -0.003 to 0.021 (Table S4). Only two of 12 intra-ecotype comparisons were significant, both of which occurred in Okanagan Lake (NE beach/NW beach: $F_{ST} = 0.012$; Penticotin creek/Mission Creek: $F_{ST} = 0.021$; Table S4). Given the extremely low F_{ST} values and the results of previous detailed analyses focusing on Okanagan Lake that revealed no site-specific patterns (Russello *et al.*, 2012), we proceeded to pool individuals by ecotype within each lake for all downstream analyses.

Within-population diversity at all loci

The level of polymorphism varied across lakes, but was high overall (80–100%; Table 1). Polymorphism was lower in lakes outside of the Okanagan River Chain: Tchesinkut Lake kokanee had eight monomorphic loci

(EV291, EV691, OMM5032, OMM5037, OMM5099, OMM5121, One8, Ots06), Duncan Lake had three (EV723, EV911, Ots06) and Kootenay Lake had one (Ots06). No trends were observed in N_A or H_E among beach- and stream-spawners, although H_E was slightly higher for the stream-spawners in four of five lakes. Overall, the N_A ranged from 1 to 23, with a mean of 5.35, and H_E ranged widely from 0 to 0.93 across loci with a mean of 0.46.

Outlier locus detection and dataset definition

Outlier locus detection rates varied substantially across the four methods. BAYESCAN was the most conservative whereas LOSITAN and DETSEL had the highest outlier detection rates (Table 2). In total, 30 of 43 EST-linked microsatellite loci (69.8%) and three of seven anonymous

Table 2 Loci detected as outliers by four different algorithms (BAYESCAN/LOSITAN/DETSEL/*ln*RH) in each of five British Columbian Lakes. Loci detected by only one approach are listed as false outliers (False). Loci detected by at least two approaches in at least one lake are outliers (Outlier; light shading), and those outliers that are detected in two or more lakes are repeat outliers (R outlier; dark shading). Each marker is identified as either an EST-linked microsatellite marker (EST) or an anonymous microsatellite marker (Anon).

Marker type	Locus	Okanagan	Wood	Duncan	Kootenay	Tchesinkut	Status
EST	TAP2	-/-/*/-			-/*/*/*/-	-/*/*/*/-	R Outlier
EST	EV358	-/*/*/*/*	-/*/*/*/-	-/*/*/*/*		-/*/*/*/-	R Outlier
EST	OMM5003		-/*/*/*/-	-/*/*/*/-		*/*/*/*/*/-	R Outlier
EST	OMM5067		-/*/*/*/*	-/*/*/*/-	-/*/*/*/*	-/*/*/*/-	R Outlier
EST	OMM5125	***/*/*/*/*/-					Outlier
EST	Ots06	-/-/*/*/*	-/*/*/*/*/*	MONO	MONO	MONO	Outlier
EST	EV862	-/*/*/*/-	-/*/*/*/-				Outlier
EST	EV170	-/*/*/*/*			-/*/*/*/*		Outlier
EST	EV642	-/*/*/*/*/-	-/*/*/*/-		-/*/*/*/-		Outlier
EST	EV685		-/*/*/*/*		-/*/*/*/*/-		Outlier
EST	OMM5037		-/*/*/*/*			MONO	Outlier
EST	EV691		-/*/*/*/*/-	-/*/*/*/*		MONO	Outlier
EST	EV740			-/*/*/*/*/*			Outlier
EST	OMM5033			-/*/*/*/*	-/*/*/*/*		Outlier
EST	OMM5121				-/*/*/*/*/-	MONO	Outlier
EST	EV536	-/*/*/*/*					False
EST	EV188	-/*/*/*/*	-/*/*/*/*/-				False
EST	OMM5124		-/*/*/*/*/-				False
EST	Ca983		-/*/*/*/*/-				False
EST	EV291		-/*/*/*/*			MONO	False
EST	EV597		-/*/*/*/*/-				False
Anon	One8		-/*/*/*/*/-	-/*/*/*/*/-		MONO	False
EST	OMM5058		-/*/*/*/*/-	-/*/*/*/*/-			False
EST	EV475			-/*/*/*/*/-			False
EST	OMM5053		*/*/*/*/*/-	-/*/*/*/*	-/*/*/*/*/-		False
EST	OMM5032				-/*/*/*/*/-	MONO	False
EST	EV365				-/*/*/*/*/-		False
EST	EV634				-/*/*/*/*/-		False
Anon	One108				-/*/*/*/*/-		False
EST	EV220				-/*/*/*/*		False
Anon	Ots14				-/*/*/*/*		False
EST	Ev484					-/*/*/*/*	False
EST	EV911			MONO		-/*/*/*/*	False

MONO indicates that the locus is monomorphic in that lake.

* $P < 0.10$, ** $P < 0.05$, *** $P < 0.01$.

microsatellite loci (42.9%) were detected as putative outliers by at least one method in at least one lake. Based on the patterns in positive detections across the four methods, 15 outliers were retained (30% overall) and 18 were discarded as false outliers (Table 2). Four of the 15 outliers were repeatedly detected in multiple lakes ('repeat outliers'), although no locus was identified as an outlier in all five lakes. Of the 17 loci not detected as outliers by any test in any lake, 15 were polymorphic in all five lakes and retained in the neutral data set.

Among-population differentiation at neutral loci

The Bayesian clustering analysis revealed two discrete genetic clusters (primary ΔK peak at $K = 2$) corresponding to kokanee native to the Columbia River drainage (Okanagan, Wood, Kootenay, Duncan) and Tchesinkut Lake, the sole representative from the Fraser River drainage (Fig. 2a). A secondary ΔK peak at $K = 6$ further indicated discrete clusters by lake, with Wood Lake stream- and beach-spawners constituting separate units (Fig. 2b). Similarly, the DAPC plot clustered groups geographically, such that the first principal component emphasized genetic differences among populations from the Columbia River drainage and Tchesinkut Lake, and the second principal component reflected differences among and within the Okanagan and Kootenay River Chains (Fig. 3). Considerable overlap was observed for Okanagan and Wood Lake kokanee, as well as Kootenay and Duncan Lake kokanee (Fig. 3).

Neutral vs. adaptive genetic variation

The amount of genetic variation occurring among ecotypes putatively due to selection (outlier loci) exceeded the variation due to drift (neutral loci) as revealed by a global AMOVA (Table 3). A significant difference was detected between the four repeat outliers and 15 neutral loci (chi-squared test, $P = 0.002$) in the AMOVA,

but not between all 15 outliers and 15 neutral loci (chi-squared test, $P = 0.332$). Within each lake, ecotype differentiation was higher at outlier loci (average $F_{ST} = 0.074 \pm 0.034$ 95% CI) compared with neutral loci (average $F_{ST} = 0.009 \pm 0.012$ 95% CI; Table 4). Furthermore, the level of ecotype differentiation in Okanagan, Kootenay and Tchesinkut Lakes was significantly greater at lake-specific outliers than repeat outliers (paired t -test; $P = 0.018$). Still, among-lake divergence substantially exceeded ecotype divergence within lake across all comparisons (Table 4).

STRUCTURE analyses revealed distinct patterns of within-lake structuring in three (Okanagan, Wood and Kootenay) of the five lakes when using all loci, corresponding to ecotype ($K = 2$; Fig. 4). Complete admixture ($K = 1$) was inferred for all lakes when using the neutral loci except Wood Lake, which was highly differentiated. No evidence of genetic structure within ecotype pairs was detected using the four repeat outliers alone (except Wood Lake; data not shown).

Functional annotations of outlier loci

Annotations for nine of the 15 outlier loci were recovered via BLAST database searches (using 'salmonidae'), including three of the four repeat outliers (Table 5). Two of the loci are associated with genes coding for transmembrane receptors (*EV358*, *OMM5003*), three with enzymes (*EV170*, *EV685*, *EV740*), two with protein transporters (*EV691*, *TAP2*) and one with a transcription regulator (*EV642*) (Table 5). The percentage identity scores for four of the loci (*EV685*, *EV691*, *EV7540*, *EV862*) were exceedingly low, suggesting their annotations should be interpreted with caution. Annotations were not found for *Ots06* and five of the *OMM* markers originally described in rainbow trout (*Oncorhynchus mykiss*) by Rexroad *et al.* (2005).

All five reliably annotated outlier loci (*EV170*, *EV358*, *EV642*, *OMM5003*, *TAP2*) and three strong lake-specific

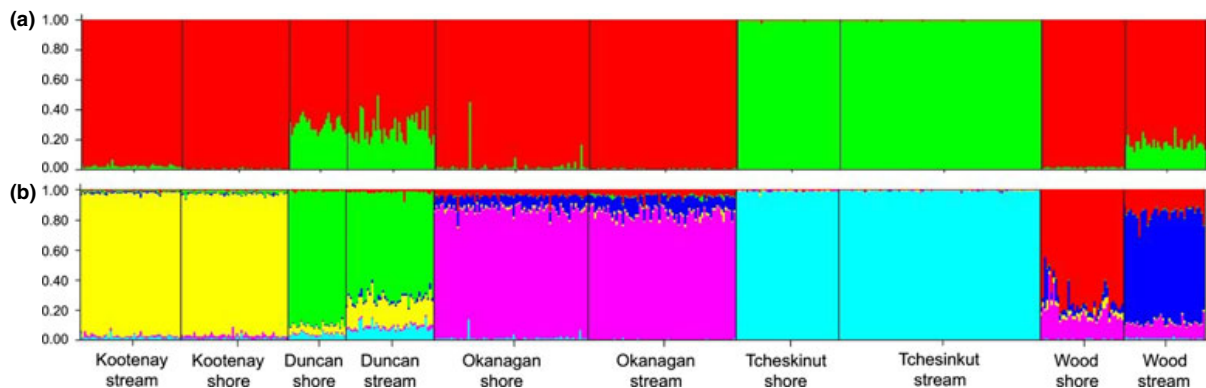


Fig. 2 STRUCTURE plots generated using neutral loci ($n = 15$) for ecotype pairs sampled from Duncan, Kootenay, Okanagan, Tchesinkut and Wood Lakes for: A. $K = 2$; and B. $K = 6$.

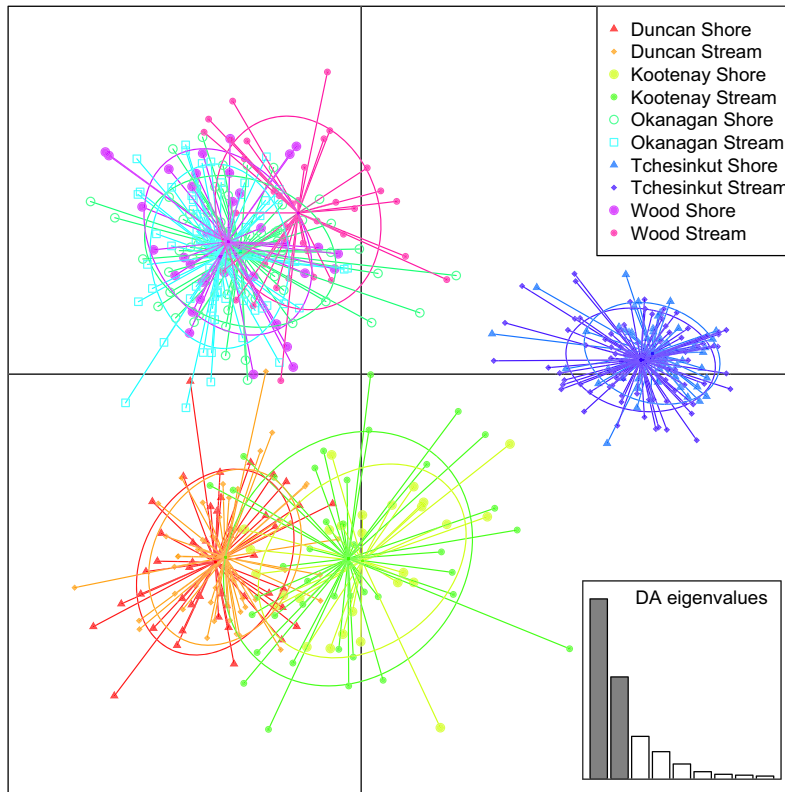


Fig. 3 Discriminate analysis of principal components (DAPC) depicting the relationships among individuals from each ecotype in each lake based on genetic similarity. Similarity is estimated from polymorphism frequency data at 15 neutral loci and displayed on the first two principal components (x- and y-axes).

Table 3 The percentage of genetic variation occurring among lakes and among ecotypes within lakes as assessed by a hierarchical analysis of molecular variance (AMOVA). Estimates were obtained separately for repeat outliers, all outliers and neutral loci.

Level	4 Repeat outliers (%)	15 Outliers (%)	15 Neutral loci (%)
Among lakes	12.71*	19.14*	14.98*
Among ecotypes	3.33*	2.49*	0.72*
Within ecotypes	83.96	78.37	84.30

*Indicates a significant level of differentiation ($P < 0.05$).

outliers (*OMM5067*, *OMM5125*, *Ots06*) were Sanger sequenced at adjacent regions. Two loci produced unreliable sequence (*EV170*, *EV642*). Of the remaining loci, only the *TAP2* locus exhibited variation: two distinct haplotypes differing in 22 SNPs and a 12 basepair insertion. One haplotype was shared by all eight stream-spawners and five beach-spawners and the other highly divergent haplotype was unique to three beach-spawning individuals from Tchesinkut Lake.

Table 4 The proportion of genetic variation occurring among ecotype pairs from five lakes as assessed by pairwise F_{ST} using repeat outliers, outliers identified in each respective lake and neutral loci.

Lake	4 Repeat outliers	Lake-specific outliers (number of loci)	15 Neutral loci
Duncan	0.013*	0.017* (4)	-0.001
Kootenay	0.044*	0.071* (5)	0.008*
Okanagan	0.017*	0.081* (3)	0.005*
Wood	0.104*	0.126* (6)	0.032*
Tchesinkut	0.041*	0.074* (3)	0.000

*Indicates a significant level of differentiation ($P < 0.05$).

Discussion

Phylogeographic patterns

When using population genomic approaches to detect divergent selection, sampled populations should be closely related, recently diverged and subject to ongoing gene flow, all conditions that were met in the current

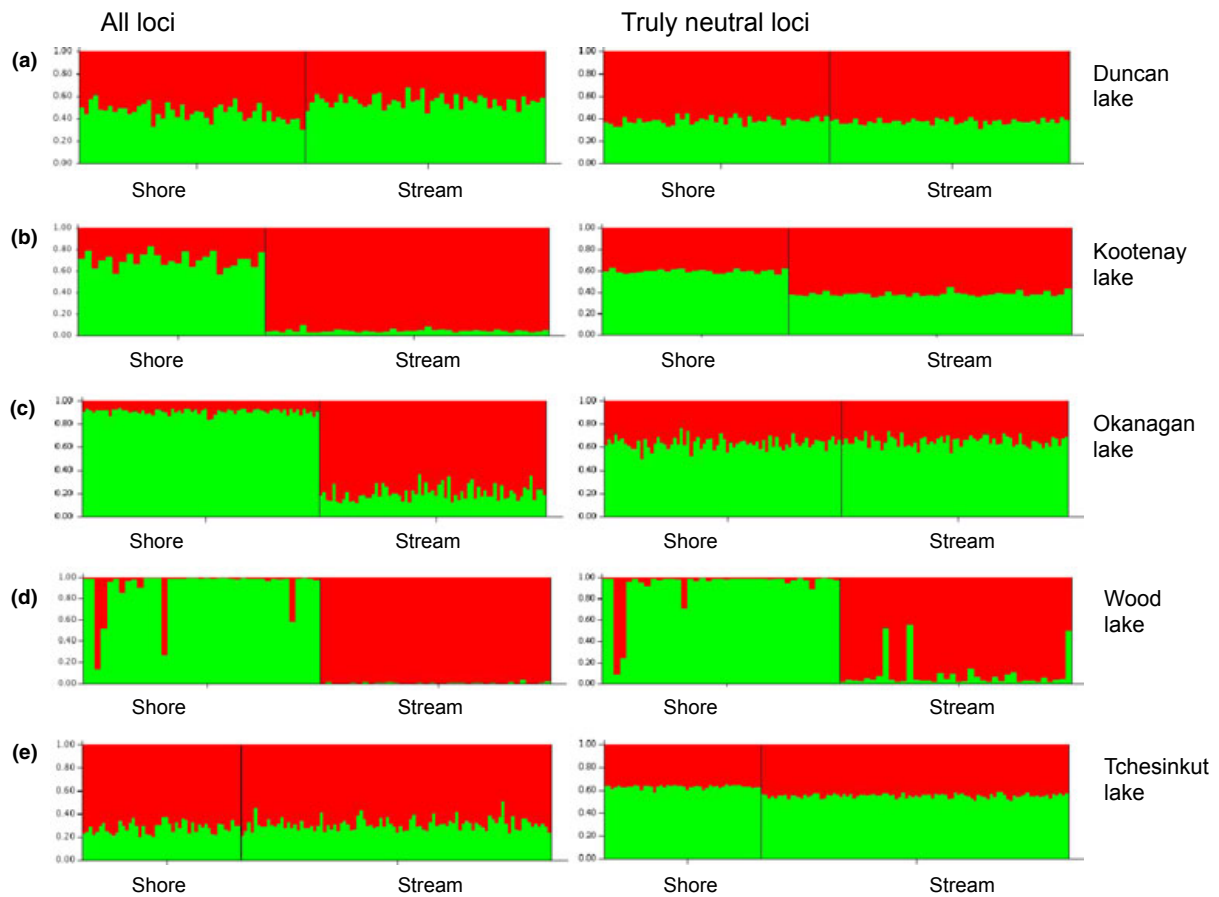


Fig. 4 STRUCTURE plots indicating the probability of membership to a cluster for each individual (vertical bars) using all loci ($n_{\text{loci}} = 48\text{--}50$; left) and neutral loci only ($n_{\text{loci}} = 15$; right). The numbers of clusters inferred using all loci were as follows: (a) Duncan ($K = 1$); (b) Kootenay ($K = 2$); (c) Okanagan ($K = 2$); (d) Wood ($K = 2$); (e) Tchesinkut ($K = 1$). The number of clusters inferred using neutral loci was $K = 1$ for all lakes but (d) Wood ($K = 2$). For display purposes, $K = 2$ plots are shown.

study (Beaumont & Nichols, 1996; Schlötterer, 2002; Stinchcombe & Hoekstra, 2008). We found evidence for neutral differentiation among lakes (Table 3), including those adjacent to one another, but none among ecotypes within each lake other than in Wood (Table S4). Trends in polymorphism (Table 1) and clustering analyses (Fig. 3) suggest that Tchesinkut Lake kokanee are genetically isolated from the other four populations, consistent with the proposed colonization and phylogeographic history of *Oncorhynchus nerka* in British Columbia (Foote *et al.*, 1989; Taylor *et al.*, 1996). Specifically, sockeye from the Bering Refugium likely colonized post-glacial lakes in northern BC and Alaska, whereas sockeye from the Columbia Refugium colonized southern BC and northern USA (Foote *et al.*, 1989; Taylor *et al.*, 1996). Weak neutral differentiation among populations sampled from adjacent lakes in the Kootenay (i.e. Duncan and Kootenay) and Okanagan River Chains (i.e. Okanagan and Wood Lakes) further

reflects the recent connectivity among them. After water levels reached their present level < 10 000 years ago, migration among these lakes would have been possible up until the construction of several dams in the 1910s and 1960s (Table S1). Overall, these results indicate that kokanee in BC share a recent common lineage and ecotype divergence was likely initiated and maintained within each lake, suggesting that neutral evolutionary processes are likely not confounding our interpretation of putative signatures of selection detected in this study.

Wood Lake kokanee are unique because they exhibited much greater differentiation at neutral and putatively adaptive loci than any other ecotype pair. The annual supplementation of ~50 000 captive reared fry spawned by Wood Lake stream-spawners may be producing a bottlenecking effect, although no signatures of demographic contraction were detected. Alternatively, the stocking programme artificially imposes a

Table 5 Gene annotations for nine EST-linked microsatellite loci identified as outliers. Each annotated gene had the largest similarity score in either the cGRASP or BLAST database and was originally described in Atlantic salmon (*Salmo salar*). The name, accession number, similarity score (e-value), type of protein, location of expression and function are given if known. Annotated repeat outliers shaded in grey.

Locus	Accession number	Similarity score (% coverage / % identity)	Gene	Protein type	Location of expression	Protein function
EV358	NP_001158765.1	1e-43 (56/54)	Chemokine receptor 4 (CXCR4)	Receptor	Immune system	Transmembrane receptor
OMM5003	BT044770	0.0 (88/94)	Plasminogen activator inhibitor 1 RNA-binding protein (PAIRB)	Receptor	Blood	n/a
TAP2	CAB05916.1	1e-62 (66/68)	ATP-binding cassette transporter (TAP2B)	Transporter	Immune system	Antigen peptide transmembrane transporter; exports noxious substances across cellular membranes
EV170	BT060423.1	0.0 (96/95)	Malate dehydrogenase (MDHC)	Enzyme	Cytoplasm	Enzyme in mitochondria that is part of the citric acid cycle
EV642	NM_001141486.2	5e-45 (14/98)	Transcription elongation factor B (ELOB)	Transcription regulator	n/a	RNA transcription regulator
EV685	NM_001173960.1	0.0 (99/96)	Proline 4-hydroxylase, alpha (P4HA2)	Enzyme	n/a	n/a
EV691	ACN58615.1.1	0.002 (11/38)	Ras-related protein Rab-14	Transporter	n/a	Protein transporter
EV740	ACM08278.1	0.013 (27/35)	Steroid 5-alpha-reductase (S5A2)	Enzyme	n/a	Enzyme responsible for the formation of dihydrotestosterone
EV862	ACI67592.1	0.02 (17/26)	Limb-bud heart (LBH) protein	Transcription regulator	Heart	Transcription activator for mitogen-activated protein kinase-signalling pathway

barrier to interhabitat migration, effectively rendering a subset of the stream-spawning population allopatric to the beach-spawning fish each year. From an ecological perspective, density-dependent selection may be stronger during the spawning or incubation period if abundance is not otherwise limited by lake productivity and predation. This lake has none of the usual predators and is eutrophic, allowing it to support a high density of kokanee that reach sizes much larger than those in the other four oligotrophic lakes. The high outlier detection rate and number of potential false positives identified in this lake emphasizes the importance of using divergent populations with ongoing gene flow and/or using multiple populations to reduce Type I error in studies of adaptive population divergence.

In contrast, Tchesinkut Lake exhibits substantially lower levels of genetic variation than all other lakes included in the analysis. There was no evidence for population contraction in this system that could account for this finding (Table 1). Rather, we suggest that ascertainment bias related to initial marker choice is the likely cause of the lower levels of variation detected in Tchesinkut Lake. The loci used were based on a previous study of Okanagan Lake kokanee (Russello *et al.*, 2012) which, along with Wood, Kootenay and Duncan, are all part of the Columbia River system. As the sole representative of the Fraser River drainage, Tchesinkut Lake has an independent history that may

have contributed to ascertainment bias and associated lower levels of detected genetic variation.

Outlier locus detection

Accurately estimating parameters for natural populations with complex demographic histories is a challenge for all outlier detection methods, contributing to high Type I and II error rates, especially when selection is weak (Storz, 2005; Excoffier *et al.*, 2009; Narum & Hess, 2011). Although the implementation of stringent significance criteria and corrections for multiple testing are commonly used to reduce the number of false positives (Storz & Nachman, 2003; Egan *et al.*, 2008), they also increase the risk of Type II error (Moran, 2003). Here, we combined the results of four conceptually different tests to minimize Type I error, as positive detection by methods that use different algorithms and assumptions (i.e. regarding the metric used, model of evolution, equal or constant population sizes) are more likely to be indicative of truly adaptive genomic regions (Oetjen & Reusch, 2007; Shikano *et al.*, 2010). Although many loci detected at a 90–95% CI were later identified as putative false outliers, these loci were also detected by other algorithms and/or in other lakes at a higher CI in many instances (e.g. OMM5003 in Tchesinkut Lake was detected at > 0.76 in BAYESCAN, > 95% in DETSEL and > 99% in LOSITAN). This result demonstrates that

the frequency and consistency of parallel patterns in outlier detection across lakes reported here might be an underestimate, influenced by Type II error associated with our conservative approach for identifying outliers. Similarly, the genomic baseline was estimated using a combination of coding (EST-linked) and noncoding (anonymous) markers, which may have elevated the 'background' divergence and increased the probability of false negatives. That said, our final detection rate (30%) was relatively high compared with other studies of similar scope (13–20%; Vasemagi *et al.*, 2005; Namroud *et al.*, 2008; Shikano *et al.*, 2010), likely due to the initial screening for polymorphic EST-linked microsatellites that preceded this study (Russello *et al.*, 2012) and the broad geographic scale of the analysis.

The repeat outliers represent the most promising candidate loci because they were detected in several ecotype pairs from different lakes (Table 2), suggesting that sampling bias did not generate the observed departure from the genomic average (Foll & Gaggiotti, 2008). Evidence of greater ecotype differentiation (Tables 3 and 4) and better resolution of population structuring at outlier loci compared with neutral loci suggests that selection may be acting on gene regions linked to the outlier loci identified here (Table 5). Duncan Lake kokanee showed very low levels of differentiation and structure regardless of the data set used, but this is likely because the lake was only formed 47 years ago following the construction of Duncan Dam.

Lack of parallel genetic patterns

Some studies have revealed compelling evidence for a common genetic basis underlying the repeated evolution of adaptive traits in discrete populations (Colosimo *et al.*, 2004). Here, we found consistent patterns in outlier detection across more than one lake, but identified loci were never coincident in more than three of five lakes. It is unclear whether the lack of concordance across all lakes is a result of: 1) an inability to detect loci truly under selection; 2) alternative genes or genetic pathways underlie ecotype divergence in this system; and/or 3) phenotypic plasticity plays a formative role in driving kokanee spawning habitat differences. A simulation study by Teshima *et al.* (2006) showed that many loci under selection may be missed by population genomic approaches, especially if the locus under selection was previously neutral. Also, if a soft selective sweep has taken place, the gene associated with the favoured trait would be present in the population at a low frequency for some time, allowing polymorphisms to accumulate at the linked microsatellite marker through neutral processes (i.e. mutation and drift). In such cases, many alleles may have hitchhiked when the population encountered a new selection regime and the selected gene moved towards fixation (Barrett & Hoekstra, 2011). Signatures of selection may

be weak if the environment is heterogeneously distributed, traits under selection are not pleiotropic (Liao *et al.*, 2010), or linkage between the neutral marker, and the fitness-relevant mutation has been broken up by recombination (Nielsen, 2005; Teshima *et al.*, 2006; Stinchcombe & Hoekstra, 2008). Likewise, signatures of selection may be obscured by other confounding factors including inflated estimates of neutral differentiation due to random chance, sampling bias (i.e. ascertainment bias; Thornton & Jensen, 2007; Deagle *et al.*, 2012; Roesti *et al.*, 2012) or the inclusion of unidentified strays (Excoffier *et al.*, 2009).

On the other hand, focusing on repeat outliers may result in overlooking loci of interest (Stinchcombe & Hoekstra, 2008). Even if selection favours the same phenotype in multiple isolated populations, mutations in different genes or different pathways could produce similar phenotypes. A recent meta-analysis estimated a 32% probability of reuse of the same genes underlying repeated phenotypic evolution in natural populations, a larger than expected proportion yet still suggestive of a high frequency of paralogy (Conte *et al.*, 2012). Many other population genomic studies have failed to detect parallel patterns (Campbell & Bernatchez, 2004; Egan *et al.*, 2008; DeFaveri *et al.*, 2011; Renaut *et al.*, 2011; Tice & Carlon, 2011), and others show that mutations in different genes have produced the same phenotype in closely related populations (Hoekstra *et al.*, 2006). Here, incongruent patterns in loci detected as outliers across lakes suggest that there may not be a common genetic basis underlying ecotype divergence in different lakes. For example, some loci were detected as strong outliers (e.g. > 99% CI) by two or three different methods in only a single lake (i.e. *OMM5125* in Okanagan, and *Ots06* in Wood Lake) and some loci identified as outliers in one lake were monomorphic in others. Imperfect parallelism in habitat features (e.g. water depth, substrate type, flow regimes) and/or phenotypic traits across the five lakes may explain these patterns. Another potentially confounding factor was the comparatively few loci sampled ($n = 57$), yet recent characterization of the kokanee transcriptome holds promise for identifying SNPs at a broader genomic scale for further investigating a common genetic basis for ecotype divergence (Lemay *et al.*, 2013).

We were unable to identify polymorphic SNPs in the annotated genes associated with the outliers, except for *TAP2B*. For this locus, only two alleles were present across our entire sampling, corresponding to two distinct haplotypes differing at 22 SNPs and a 12 basepair insertion. Interestingly, there is some preliminary indication of ecotype-specific haplotype frequencies in Okanagan, Wood and Kootenay lakes at *TAP2B*. At this biallelic locus, the major allele in stream-spawners included the 12 basepair insertion (ranging from 0.54–0.62), whereas the inverse pattern was detected in beach-spawners (ranging from 0.33–0.40). *TAP2B* was

detected as a high F_{ST} outlier as has been reported previously in brown trout (*Salmo trutta*, L.; Jensen *et al.*, 2008), and exhibited no significant difference in observed heterozygosity between ecotypes. Although no similar insertions have been documented in the literature, the high conservation of alleles across the landscape was similar to that found in brown trout (Keller *et al.*, 2011). Further study is required involving larger sample sizes and additional lakes to investigate whether this pattern is biologically meaningful in kokanee.

For all candidate genes, we only sequenced 300–600 basepairs in immediate proximity to the microsatellite repeat, and empirical studies have shown that a hitchhiking effect may be observed in 20–300 kilobasepair regions around the selected gene (Nash *et al.*, 2005; Teschke *et al.*, 2008). Consequently, our inability to detect additional informative SNPs does not necessarily invalidate these regions (or linked regions) as strong candidates for divergent selection. A more concerted effort to sequence adjacent exons and explore associated linkage regions may yield additional insights.

Traits putatively under selection

Currently, little is known about the specific adaptations that differentially influence individual fitness in a stream vs. beach environment. We have identified several candidate genes that allow us to make preliminary inferences about the traits potentially involved based on their annotated function. In particular, locus *EV358* is linked to an immunological gene, *CXCR4*. In Atlantic salmon, *CXCR4* is differentially expressed when exposed to sea lice (Skugar *et al.*, 2008) and saprolegniasis, a fungal infection (Roberge *et al.*, 2007). *CXCR4* has potent chemotactic activity for lymphocytes and has been shown to inhibit haematopoietic stem cell proliferation (Nie *et al.*, 2008). The *TAP2B* gene is also associated with immune response, encoding an antigen-transporter protein that hydrolyses ATP to export noxious substances and extracellular toxins across the membrane of the endoplasmic reticulum where it associates with MHC class I molecules (Monaco, 1992). The functional role of TAP2 polymorphism is poorly understood even in humans (McCluskey *et al.*, 2004), but some studies have found differential ability of allelic products to transport specific peptides (Monaco, 1992; Powis *et al.*, 1992; Heemels & Ploegh, 1994) and strong LD between MHC I and TAP2 (Ohta *et al.*, 2003). The functioning of this protein can be suppressed by some viruses in brown trout and European trout (*Salmo trutta*), which suggests that selection may be pathogen-mediated (Abele & Tampe, 2006; Jensen *et al.*, 2008; Keller *et al.*, 2011). Moreover, studies of Alaskan sockeye have correlated habitat types and spawning sites with variation at SNPs in the MHC that are linked to immune function and disease resistance (Creelman *et al.*, 2011; Gomez-Uchida *et al.*, 2011). In the case of

kokanee, stream-spawners typically defend territories, mates (Morbey, 2002, 2003) and nests to a significantly greater extent than beach-spawners (Dill, 1996), making them more prone to physical injury and potentially conferring increased susceptibility to infection. Interestingly, preliminary evidence of differential pathogen exposure of beach- and stream-spawners has been indirectly detected, but additional studies are required to explicitly test this hypothesis (Lemay *et al.*, 2013).

Locus *EVI70* is associated with an enzyme in the citric acid cycle and is therefore critically involved in energy metabolism throughout the body. Selection on this gene may reflect the reduced metabolic demands of beach-spawners, because the movement to beach sites is less taxing than up-stream migrations. An over-expression of genes with key roles in the citric acid cycle (including malate dehydrogenase) has also been implicated in the divergence of dwarf and normal lake whitefish (St-Cyr *et al.*, 2008). Although these candidate genes offer some insight into plausible mechanisms driving ecotype divergence in kokanee, the loci used here represent a very limited portion of the genome. For example, there is a temporal separation in stream- and beach-spawning in many lakes that may have a genetic basis, but no loci linked to spawning time were identified or explicitly included in this study (e.g. CLOCK genes; O'Malley *et al.*, 2007). On-going genotyping-by-sequencing efforts to characterize genome-wide variation within and among kokanee ecotypes may reveal additional candidate genes within this broader system (M. Lemay and M. A. Russello, unpublished).

Conclusions

Loci exhibiting parallel patterns of outlier behaviour across multiple replicate ecotype pairs have the greatest potential to be under the action of natural selection and involved in generating barriers to gene flow among divergent ecotypes. Here, we provide evidence for the independent colonization of beach habitats by stream-spawning counterparts in each lake. The lack of parallel patterns in outlier locus detection suggests that each beach-spawning population may be uniquely adapted to their habitat in each lake. We identified several candidate gene regions that putatively exhibit signatures of directional selection, which require further validation. Based on these initial findings, we can infer pathogen resistance and energy metabolism as potential and highly plausible mechanisms contributing to the independent divergence of beach- and stream-spawning kokanee, yet we cannot rule out a role for plasticity and nonadaptive processes in contributing to spawning habitat differences. Overall, these insights demonstrate the utility of a population genomic approach for identifying candidate loci underlying fine-scale ecotype

divergence and punctuate the need for integrating multiple approaches when investigating causal links between phenotype, genotype and fitness in natural populations.

Acknowledgments

Paul Askey (BC Ministry of Forests, Lands and Natural Resource Operations) offered valuable background information regarding this system and, together with Jason Webster (Chara Consulting), was instrumental in providing samples from Okanagan and Wood Lakes. Jeff Burrows and Joe DeGisi (BC Ministry of Environment) facilitated sampling in Kootenay and Tchesinkut Lakes, respectively, and all samples from Duncan Lake were provided by Ico de Zwart (Masse Environmental Consultants). Matt Lemay provided helpful comments on an earlier version of this manuscript and Mark Rheault offered valuable advice regarding gene annotations. Funding was provided by the Natural Sciences and Engineering Research Council of Canada (NSERC; Discovery grant # 341711-07 to MR) and KF was partially supported by an NSERC Canada Graduate Scholarship and a Pacific Century Award.

References

- Abele, R. & Tampe, R. 2006. Modulation of the antigen transport machinery TAP by friends and enemies. *FEBS Lett.* **580**: 1156–1163.
- Andrusak, G. & Andrusak, H. 2011. *Observations and Analysis of Shore Spawning Kokanee (Oncorhynchus Nerka) in the West Arm of Kootenay Lake BC Hydro, Fortis BC and Columbia Power Corporation*. Redfish Consulting Ltd, Nelson, BC.
- Antao, T., Lopes, A., Lopes, R.J., Beja-Pereira, A. & Luikart, G. 2008. LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics* **9**: 323.
- Ashley, K., Shepherd, B., Sebastian, D., Thompson, L., Vidmanic, L., Ward, P. *et al.* 1998. Okanagan Lake Action Plan Year 1 (1996-1997) and Year 2 (1997-1998) Report. Fisheries Project Report No. RD 73, Province of British Columbia, Ministry of Fisheries, Fisheries Management Branch.
- Barrett, R.D.H. & Hoekstra, H.E. 2011. Molecular spandrels: tests of adaptation at the genetic level. *Nat. Rev. Genet.* **12**: 767–780.
- Beaumont, M.A. & Nichols, R.A. 1996. Evaluating loci for use in the genetic analysis of population structure. *Proc. Biol. Sci.* **263**: 1619–1626.
- Benjamini, Y. & Hochberg, Y. 1995. Controlling the false discovery rate - A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B* **57**: 289–300.
- Bonin, A., Taberlet, P., Miaud, C. & Pompanon, F. 2006. Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Mol. Biol. Evol.* **23**: 773–783.
- Brownstein, M.J., Carpten, J.D. & Smith, J.R. 1996. Modulation of non-templated nucleotide addition by *Taq* DNA polymerase: primer modifications that facilitate genotyping. *Biotechniques* **20**: 1004.
- Burger, C.V., Finn, J.E. & Hollandbartels, L. 1995. Pattern of shoreline spawning by sockeye-salmon in a glacially turbid lake - evidence for subpopulation differentiation. *Trans. Am. Fish. Soc.* **124**: 1–15.
- Burgner, R.L. 1991. Life history of sockeye salmon *Oncorhynchus nerka*. In: *Pacific Salmon Life Histories* (C. Groot & L. Margolis, eds), pp. 3–118. UBC Press, Vancouver.
- Butlin, R.K. 2010. Population genomics and speciation. *Genetica* **138**: 409–418.
- Campbell, D. & Bernatchez, L. 2004. Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. *Mol. Biol. Evol.* **21**: 945–956.
- Colosimo, P.F., Peichel, C.L., Nereng, K., Blackman, B.K., Shapiro, M.D., Schluter, D. *et al.* 2004. The genetic architecture of parallel armor plate reduction in threespine sticklebacks. *PLoS Biol.* **2**: 635–641.
- Conte, G.L., Arnegard, M.E., Peichel, C.L. & Schluter, D. 2012. The probability of genetic parallelism and convergence in natural populations. *Proc. Biol. Sci.* **279**: 5039–5047.
- Cornuet, J.M. & Luikart, G. 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**: 2001–2014.
- Creelman, E.K., Hauser, L., Simmons, R.K., Templin, W.D. & Seeb, L.W. 2011. Temporal and geographic genetic divergence: characterizing sockeye salmon populations in the Chignik Watershed, Alaska, using single-nucleotide polymorphisms. *Trans. Am. Fish. Soc.* **140**: 749–762.
- Deagle, B.E., Jones, F.C., Chan, Y.F., Absher, D.M., Kingsley, D.M. & Reimchen, T.E. 2012. Population genomics of parallel phenotypic evolution in stickleback across stream-lake ecological transitions. *Proc. Biol. Sci.* **279**: 1277–1286.
- DeFaveri, J., Shikano, T., Shimada, Y., Goto, A. & Merila, J. 2011. Global analysis of genes involved in freshwater adaptation in three-spined stickleback (*Gasterosteus aculeatus*). *Evolution* **65**: 1800–1807.
- Dill, P. 1996. *A Study of Shore-Spawning Kokanee Salmon (Oncorhynchus Nerka) at Bertram Creek Park Okanagan Lake, B.C., 1992–1996*, pp. 23. B.C. Ministry of Environment, Penticton, B.C.
- Dirienzo, A., Peterson, A.C., Garza, J.C., Valdes, A.M., Slatkin, M. & Freimer, N.B. 1994. Mutational processes of simple-sequence repeat loci in human populations. *Proc. Natl. Acad. Sci. USA* **91**: 3166–3170.
- Dittman, A.H. & Quinn, T.P. 1996. Homing in Pacific salmon: mechanisms and ecological basis. *J. Exp. Biol.* **199**: 83–91.
- Drummond, A., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Duran, C. *et al.* 2011. Geneious v5.4. Available from <http://www.geneious.com/>
- Earl, D.A. & vonHoldt, B.M. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* **4**: 359–361.
- Egan, S.P., Nosil, P. & Funk, D.J. 2008. Selection and genomic differentiation during ecological speciation: isolating the contributions of host association via a comparative genome scan of *Neochlamisus bebbianae* leaf beetles. *Evolution* **62**: 1162–1181.
- Evanno, G., Regnaut, S. & Goudet, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* **14**: 2611–2620.

- Excoffier, L. & Lischer, H.E.L. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10**: 564–567.
- Excoffier, L., Smouse, P.E. & Quattro, J.M. 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., Hofer, T. & Foll, M. 2009. Detecting loci under selection in a hierarchically structured population. *Heredity* **103**: 285–298.
- Feder, J.L., Egan, S.P. & Nosil, P. 2012. The genomics of speciation-with-gene-flow. *Trends Genet.* **28**: 342–350.
- Foll, M. & Gaggiotti, O. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180**: 977–993.
- Footo, C.J., Wood, C.C. & Withler, R.E. 1989. Biochemical genetic comparison of anadromous sockeye and kokanee, the anadromous and non-anadromous forms of *Oncorhynchus nerka*. *Can. J. Fish. Aquat. Sci.* **46**: 149–158.
- Garza, J.C. & Williamson, E.G. 2001. Detection of reduction in population size using data from microsatellite loci. *Mol. Ecol.* **10**: 305–318.
- Gomez-Uchida, D., Seeb, J.E., Smith, M.J., Habicht, C., Quinn, T.P. & Seeb, L.W. 2011. Single nucleotide polymorphisms unravel hierarchical divergence and signatures of selection among Alaskan sockeye salmon (*Oncorhynchus nerka*) populations. *BMC Evol. Biol.* **11**: 48.
- Guo, S.W. & Thompson, E.A. 1992. A Monte-Carlo method for combined segregation and linkage analysis. *Am. J. Hum. Genet.* **51**: 1111–1126.
- Hassemer, P.F. & Rieman, B.E. 1981. Observations of deep-spawning kokanee on artificially created spawning habitat. *North Am. J. Fish. Manag.* **1**: 173–176.
- Heemels, M.T. & Ploegh, H.L. 1994. Substrate-specificity of allelic variants of the TAP peptide transporter. *Immunity* **1**: 775–784.
- Hendry, A.P. 2009. Ecological speciation! Or the lack thereof? *Can. J. Fish. Aquat. Sci.* **66**: 1383–1398.
- Hendry, A.P. & Stearns, S.C. 2004. *Evolution Illuminated: Salmon and Their Relatives*. Oxford University Press, New York.
- Hoekstra, H.E., Hirschmann, R.J., Bunde, R.A., Insel, P.A. & Crossland, J.P. 2006. A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* **313**: 101–104.
- Hohenlohe, P.A., Bassham, S., Etter, P.D., Stiffler, N., Johnson, E.A. & Cresko, W.A. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet.* **6**: e1000862.
- Hubisz, M.J., Falush, D., Stephens, M. & Pritchard, J.K. 2009. Inferring weak population structure with the assistance of sample group information. *Mol. Ecol. Resour.* **9**: 1322–1332.
- Jensen, L.F., Hansen, M.M., Mensberg, K.L. & Loeschcke, V. 2008. Spatially and temporally fluctuating selection at non-MHC immune genes: evidence from TAP polymorphism in populations of brown trout (*Salmo trutta*, L.). *Heredity* **100**: 79–91.
- Johannesson, K. 2001. Parallel speciation: a key to sympatric divergence. *Trends Ecol. Evol.* **16**: 148–153.
- Jombart, T., Devillard, S. & Balloux, F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* **11**: 94.
- Kauer, M.O., Dieringer, D. & Schlötterer, C. 2003. A microsatellite variability screen for positive selection associated with the “Out of Africa” habitat expansion of *Drosophila melanogaster*. *Genetics* **165**: 1137–1148.
- Keller, I., Taverna, A. & Seehausen, O. 2011. Evidence of neutral and adaptive genetic divergence between European trout populations sampled along altitudinal gradients. *Mol. Ecol.* **20**: 1888–1904.
- Lecomte, F. & Dodson, J.J. 2004. Role of early life-history constraints and resource polymorphism in the segregation of sympatric populations of an estuarine fish. *Evol. Ecol. Res.* **6**: 631–658.
- Lemay, M., Donnelly, D. & Russello, M.A. 2013. Transcriptome-wide comparison of sequence variation informs SNP discovery in divergent ecotypes of kokanee salmon. *BMC Genomics* **14**: 12.
- Liao, B.-Y., Weng, M.-P. & Zhang, J. 2010. Contrasting genetic paths to morphological and physiological evolution. *Proc. Natl. Acad. Sci. USA* **107**: 7353–7358.
- Luikart, G., England, P.R., Tallmon, D., Jordan, S. & Taberlet, P. 2003. The power and promise of population genomics: from genotyping to genome typing. *Nat. Rev. Genet.* **4**: 981–994.
- McCluskey, J., Rossjohn, J. & Purcell, A.W. 2004. TAP genes and immunity. *Curr. Opin. Immunol.* **16**: 651–659.
- McKinnon, J.S., Mori, S., Blackman, B.K., David, L., Kingsley, D.M., Jamieson, L. et al. 2004. Evidence for ecology’s role in speciation. *Nature* **429**: 294–298.
- Monaco, J.J. 1992. Antigen presentation not so groovy after all? *Curr. Biol.* **2**: 433–435.
- Moran, M.D. 2003. Arguments for rejecting the sequential Bonferroni in ecological studies. *Oikos* **100**: 403–405.
- Morbey, Y.E. 2002. The mate-guarding behaviour of male kokanee *Oncorhynchus nerka*. *Behaviour* **139**: 507–528.
- Morbey, Y.E. 2003. Pair formation, pre-spawning waiting, and protandry in kokanee, *Oncorhynchus nerka*. *Behav. Ecol. Sociobiol.* **54**: 127–135.
- Namroud, M.-C., Beaulieu, J., Juge, N., Laroche, J. & Bousquet, J. 2008. Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Mol. Ecol.* **17**: 3599–3613.
- Narum, S.R. & Hess, J.E. 2011. Comparison of F(ST) outlier tests for SNP loci under selection. *Mol. Ecol. Resour.* **11**: 184–194.
- Nash, D., Nair, S., Mayxay, M., Newton, P.N., Guthmann, J., Nosten, F. et al. 2005. Selection strength and hitchhiking around two anti-malarial resistance genes. *Proc. Biol. Sci.* **272**: 1153–1161.
- Nei, M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- Nie, Y.C., Han, Y.C. & Zou, Y.R. 2008. CXCR4 is required for the quiescence of primitive hematopoietic cells. *J. Exp. Med.* **205**: 777–783.
- Nielsen, R. 2005. Molecular signatures of natural selection. *Annu. Rev. Genet.* **39**: 197–218.
- Nosil, P., Funk, D.J. & Ortiz-Barrientos, D. 2009. Divergent selection and heterogeneous genomic divergence. *Mol. Ecol.* **18**: 375–402.
- Oetjen, K. & Reusch, T.B.H. 2007. Genome scans detect consistent divergent selection among subtidal vs. intertidal populations of the marine angiosperm *Zostera marina*. *Mol. Ecol.* **16**: 5156–5167.

- Ohta, Y., Powis, S.J., Lohr, R.L., Nonaka, M., Du Pasquier, L. & Flajnik, M.F. 2003. Two highly divergent ancient allelic lineages of the transporter associated with antigen processing (TAP) gene in *Xenopus*: further evidence for co-evolution among MHC class I region genes. *Eur. J. Immunol.* **33**: 3017–3027.
- O'Malley, K.G., Camara, M.D. & Banks, M.A. 2007. Candidate loci reveal genetic differentiation between temporally divergent migratory runs of Chinook salmon (*Oncorhynchus tshawytscha*). *Mol. Ecol.* **16**: 4930–4941.
- Peakall, R. & Smouse, P.E. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* **6**: 288–295.
- Piry, S., Luikart, G. & Cornuet, J.M. 1999. BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. *J. Hered.* **90**: 502–503.
- Powis, S.J., Deverson, E.V., Coadwell, W.J., Ciruela, A., Huskisson, N.S., Smith, H. *et al.* 1992. Effect of polymorphism of an MHC-linked transporter on the peptides assembled in a class-I molecule. *Nature* **357**: 211–215.
- Pritchard, J.K., Stephens, M. & Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.
- Quinn, T.P. 1999. Variation in Pacific salmon reproductive behaviour associated with species, sex and levels of competition. *Behaviour* **136**: 179–204.
- Quinn, T.P. 2005. *The Behaviour and Ecology of Pacific Salmon and Trout*. University of Washington Press, Vancouver.
- Raymond, M. & Rousset, F. 1995. An exact test for population differentiation. *Evolution* **49**: 1280–1283.
- R Development Core Team 2012. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Renaud, S., Nolte, A.W., Rogers, S.M., Derome, N. & Bernatchez, L. 2011. SNP signatures of selection on standing genetic variation and their association with adaptive phenotypes along gradients of ecological speciation in lake whitefish species pairs (*Coregonus spp.*). *Mol. Ecol.* **20**: 545–559.
- Rexroad, C.E., Rodriguez, M.F., Coulibaly, I., Gharbi, K., Danzmann, R.G., DeKoning, J. *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.
- Ricker, W.E. 1940. On the origin of kokanee, a fresh-water type of sockeye salmon. *Trans. R. Soc. Can.* **34**: 121–135.
- Roberge, C., Paez, D.J., Rossignol, O., Guderley, H., Dodson, J. & Bernatchez, L. 2007. Genome-wide survey of the gene expression response to saprolegniasis in Atlantic salmon. *Mol. Immunol.* **44**: 1374–1383.
- Roesti, M., Hendry, A.P., Salzburger, W. & Berner, D. 2012. Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. *Mol. Ecol.* **21**: 2852–2862.
- Rogers, S.M. & Bernatchez, L. 2005. Integrating QTL mapping and genome scans towards the characterization of candidate loci under parallel selection in the lake whitefish (*Coregonus clupeaformis*). *Mol. Ecol.* **14**: 351–361.
- Rozen, S. & Skaletsky, H.J. 2000. Primer3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (S. Krawetz & S. Misener, eds), pp. 365–386. Humana Press, Totowa, NJ.
- Russello, M.A., Kirk, S., Frazer, K.K. & Askey, P. 2012. Detection of outlier loci and their utility for fisheries management. *Evol. Appl.* **5**: 39–52.
- Salem, M., Rexroad, C.E., Wang, J., Thorgaard, G.H. & Yao, J. 2010. Characterization of the rainbow trout transcriptome using Sanger and 454-pyrosequencing approaches. *BMC Genomics* **11**: 564.
- Savolainen, V., Anstett, M.C., Lexer, C., Hutton, I., Clarkson, J.J., Norup, M.V. *et al.* 2006. Sympatric speciation in palms on an oceanic island. *Nature* **441**: 210–213.
- Schlotterer, C. 2003. Hitchhiking mapping - functional genomics from the population genetics perspective. *Trends Genet.* **19**: 32–38.
- Schlötterer, C. 2002. A microsatellite-based multilocus screen for the identification of local selective sweeps. *Genetics* **160**: 753–763.
- Schluter, D., Marchinko, K.B., Barrett, R.D.H. & Rogers, S.M. 2010. Natural selection and the genetics of adaptation in threespine stickleback. *Philos. Trans. R. Soc. B* **365**: 2479–2486.
- Schuelke, M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nat. Biotechnol.* **18**: 233–234.
- Shephard, B.G. 2000. A case history: the kokanee stocks of Okanagan Lake. Proceedings of a conference on the biology and management of species and habitats at risk. L. M. Darling. Kamloops, B.C.: 609–616.
- Shikano, T., Ramadevi, J. & Merila, J. 2010. Identification of local- and habitat-dependent selection: scanning functionally important genes in nine-spined sticklebacks (*Pungitius pungitius*). *Mol. Biol. Evol.* **27**: 2775–2789.
- Skugar, S., Glover, K.A., Nilsen, F. & Krasnov, A. 2008. Local and systemic gene expression responses of Atlantic salmon (*Salmo salar* L.) to infection with the salmon louse (*Lepeophtheirus salmonis*). *BMC Genomics* **9**: 498–516.
- St-Cyr, J., Derome, N. & Bernatchez, L. 2008. The transcriptomics of life-history trade-offs in whitefish species pairs (*Coregonus sp.*). *Mol. Ecol.* **17**: 1850–1870.
- Stinchcombe, J.R. & Hoekstra, H.E. 2008. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* **100**: 158–170.
- Stölting, K., Nipper, R., Lindtke, D., Caseys, C., Waeber, S., Castiglione, S. *et al.* 2013. Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species. *Mol. Ecol.* **22**: 842–855.
- Storz, J.F. 2005. Using genome scans of DNA polymorphism to infer adaptive population divergence. *Mol. Ecol.* **14**: 671–688.
- Storz, J.F. & Nachman, M.W. 2003. Natural selection on protein polymorphism in the rodent genus *Peromyscus*: evidence from interlocus contrasts. *Evolution* **57**: 2628–2635.
- Taylor, E.B. 1991. A review of local adaptation in Salmonidae, with particular reference to Pacific and Atlantic salmon. *Aquaculture* **98**: 185–207.
- Taylor, E.B., Foote, C.J. & Wood, C.C. 1996. Molecular genetic evidence for parallel life-history evolution within a Pacific salmon (sockeye salmon and kokanee, *Oncorhynchus nerka*). *Evolution* **50**: 401–416.

- Taylor, E.B., Harvey, S., Pollard, S. & Volpe, J. 1997. Postglacial genetic differentiation of reproductive ecotypes of kokanee *Oncorhynchus nerka* in Okanagan Lake, British Columbia. *Mol. Ecol.* **6**: 503–517.
- Taylor, E.B., Kuiper, A., Troffe, P.M., Hoysak, D.J. & Pollard, S. 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. *Conserv. Genet.* **1**: 231–249.
- Teschke, M., Mukabayire, O., Wiehe, T. & Tautz, D. 2008. Identification of selective sweeps in closely related populations of the house mouse based on microsatellite scans. *Genetics* **180**: 1537–1545.
- Teshima, K.M., Coop, G. & Przeworski, M. 2006. How reliable are empirical genomic scans for selective sweeps? *Genome Res.* **16**: 702–712.
- Thornton, K.R. & Jensen, J.D. 2007. Controlling the false-positive rate in multilocus genome scans for selection. *Genetics* **175**: 737–750.
- Tice, K.A. & Carlon, D.B. 2011. Can AFLP genome scans detect small islands of differentiation? The case of shell sculpture variation in the periwinkle *Echinolittorina hawaiiensis*. *J. Evol. Biol.* **24**: 1814–1825.
- Turner, T.L., Hahn, M.W. & Nuzhdin, S.V. 2005. Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biol.* **3**: 1572–1578.
- Van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M. & Shipley, P. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes* **4**: 535–538.
- Vasemägi, A. & Primmer, C.R. 2005. Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Mol. Ecol.* **14**: 3623–3642.
- Vasemagi, A., Nilsson, J. & Primmer, C.R. 2005. Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in Atlantic salmon (*Salmo salar* L.). *Mol. Biol. Evol.* **22**: 1067–1076.
- Via, S. 2009. Natural selection in action during speciation. *Proc. Natl. Acad. Sci. USA* **106**: 9939–9946.
- Via, S. & West, J. 2008. The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Mol. Ecol.* **17**: 4334–4345.
- Vitalis, R., Dawson, K. & Boursot, P. 2001. Interpretation of variation across marker loci as evidence of selection. *Genetics* **158**: 1811–1823.
- Vitalis, R., Dawson, K., Boursot, P. & Belkhir, K. 2003. DetSel 1.0: a computer program to detect markers responding to selection. *J. Hered.* **94**: 429–431.
- Weir, B.S. & Cockerham, C.C. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.
- Winans, G.A., Pollard, S. & Kuligowski, D.R. 2003. Two reproductive life history types of kokanee, *Oncorhynchus nerka*, exhibit multivariate morphometric and protein genetic differentiation. *Environ. Biol. Fishes* **67**: 87–100.
- Wolf, J.B.W., Lindell, J. & Backstrom, N. 2010. Speciation genetics: current status and evolving approaches. *Philos. Trans. R. Soc. B* **365**: 1717–1733.

Supporting information

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

Table S1 The physical and biological attributes of each British Columbian Lake included in this study.

Table S2 Details of kokanee sample collection are listed, including the number of sites and total number of samples collected for each ecotype group. The capture method used to obtain samples, the type of tissue collected and the timing of tissue collection are also provided for each ecotype group in each lake.

Table S3 Primers designed for Sanger sequencing the top seven candidate genes.

Table S4 Population pairwise F_{ST} comparisons among sampling sites in (a) Duncan, (b) Kootenay, (c) Okanagan, (d) Tchesinkut and (e) Wood Lake, using seven anonymous microsatellite markers. Intra-ecotype comparisons are bolded.

Data deposited at Dryad: doi:10.5061/dryad.38c7

Received 2 January 2013; revised 21 July 2013; accepted 25 August 2013