ORIGINAL ARTICLE

MHC genetic structure and divergence across populations of Chinook salmon (*Oncorhynchus tshawytscha*)

ML Evans[1], BD Neff[1] and DD Heath[2]

[1]*Department of Biology, University of Western Ontario, London, Ontario, Canada and* [2]*Great Lakes Institute of Environmental Research, University of Windsor, Windsor, Ontario, Canada*

The major histocompatibility complex (MHC) is thought to be under strong selection pressure because of its integral role in pathogen recognition. Consequently, patterns of MHC genetic variation should reflect selection pressures across the landscape. We examined genetic variation and population genetic structure at the MHC class I-A1 and class II-B1 exons in five Chinook salmon (*Oncorhynchus tshawytscha*) populations from two geographic regions in British Columbia, Canada. We then compared estimates of population structure at the MHC genes with neutral estimates based on microsatellites to examine the potential for local adaptation at the MHC. Chinook salmon are in decline throughout much of their native range and understanding the degree of local adaptation exhibited by the MHC may be important in conservation planning. Comparisons among populations yielded higher G'_{ST} estimates for the MHC class I than expected under neutrality based on the microsatellites. In contrast, the MHC class II tended to exhibit lower G'_{ST} values than did the microsatellites. These results suggest that across populations unique selection pressures are driving allele frequency differences at the MHC class I but that the MHC class II may be the subject of homogenizing selection. Rates of nonsynonymous versus synonymous substitutions found in codons associated within the MHC class I and II peptide-binding regions provided strong evidence of positive selection. Together, these results support the hypothesis that selection is influencing genetic variation at the MHC, but suggest that selection pressures may vary at the two classes of loci both at the sequence and population levels.
Heredity (2010) **104,** 449–459; doi:10.1038/hdy.2009.121; published online 23 September 2009

Introduction

Population genetic structure and variation are determined by the interaction of mutation, genetic drift, gene flow, and selection (Wright, 1931). The relative contribution of these processes to genetic structure within and among populations is a fundamental question in evolutionary biology. Most population genetic studies have used putatively neutral genetic markers to describe relationships between populations. However, the population genetic structure of functionally important genetic regions may differ from that of neutral genetic regions because of the influence of differential selection across the landscape (van Tienderen *et al.*, 2002; Sommer, 2005). In non-model organisms, it has been difficult to investigate adaptive variation and divergence across populations because relatively few functional genes have been identified (Vasemagi and Primmer, 2005). An exception lies in the genes of the major histocompatibility complex (MHC). The MHC represents a functionally important region of the vertebrate genome that has been well-characterized across many

taxa (Klein *et al.*, 1997; Sommer, 2005; Piertney and Oliver, 2006). Several recent studies have shown that the MHC provides an ideal system in which to examine the importance of selection, relative to other evolutionary processes, in population genetic structure (for example, Landry and Bernatchez, 2001; Heath *et al.*, 2006; Ekblom *et al.*, 2007).

The MHC encodes for proteins that bind to pathogen-derived antigens (Matsumura *et al.*, 1992). An immune response is activated when foreign-derived peptides (antigens) are presented to T cells by MHC molecules (Klein, 1986). MHC class I proteins generally function in the recognition of intracellular pathogens, whereas MHC class II proteins function in the recognition of extracellular pathogens (Klein, 1986; Nei and Hughes, 1991). The peptide-binding region (PBR) of the MHC is involved in pathogen recognition, and tends to be the region of highest genetic variability (Hughes and Hughes, 1995; Hughes and Yeager, 1998). Changes to amino acids of the PBR may result in altered binding capability and subsequent pathogen recognition abilities of MHC molecules (Brown *et al.*, 1993). For example, studies of the MHC in humans, birds, and fishes have shown that individual alleles are associated with disease resistance (for example, Hill *et al.*, 1991, 1994; Kean *et al.*, 1994; Langefors *et al.*, 2001; Grimholt *et al.*, 2003; Miller *et al.*, 2004; Johnson *et al.*, 2008). Pathogen-mediated selection can be assessed by comparing the ratio of

nonsynonymous (D_n) to synonymous (D_s) nucleotide polymorphisms in the PBR versus the non-PBR of MHC molecules, and, indeed, most studies of nucleotide substitution rates within the MHC PBR have indicated that this region is the subject of positive selection (Hill and Hastie, 1987; Hughes and Nei, 1988; reviewed by Bernatchez and Landry, 2003; Sommer, 2005; Piertney and Oliver, 2006).

The MHC encompasses one of the most variable genetic regions ever reported (Bernatchez and Landry, 2003; Sommer, 2005; Piertney and Oliver, 2006). For example, various loci of the human MHC (human leukocyte antigen or HLA) exhibit several hundred alleles (Solberg *et al.*, 2008). It has been suggested that the extensive genetic variation at the MHC is maintained through balancing selection, mediated through overdominance effects (for example, Langefors *et al.*, 1998; Arkush *et al.*, 2002; Wegner *et al.*, 2003) or negative frequency-dependent selection (Hedrick, 1999; Bernatchez and Landry, 2003; Sommer, 2005). If overdominance is maintaining genetic variation, then heterozygous individuals should be capable of mounting an immune response against a wider range of pathogens than homozygous individuals (Doherty and Zinkernagel, 1975). In the case of negative frequency-dependent selection, rare alleles will be associated with relatively high fitness if pathogens are able to exploit the most common host MHC genotype (Hughes and Nei, 1988; Paterson *et al.*, 1998). Temporal or spatial fluctuations in pathogen communities may also play important roles in maintaining population genetic variation (Hedrick, 2002; Summers *et al.*, 2003).

The MHC often displays considerable differentiation across geographical landscapes as has been found in populations of mammals, birds, and fishes (for example, Miller and Withler, 1997; Landry and Bernatchez, 2001; Miller *et al.*, 2001; Wegner *et al.*, 2003; Bryja *et al.*, 2007; Dionne *et al.*, 2007; Ekblom *et al.*, 2007). Population subdivision at the MHC may arise because of mutation, gene flow, genetic drift, or local adaptation (Schierup *et al.*, 2000). The role of selection in the evolution of MHC allele differences across populations can be assessed by comparing patterns of genetic variation at the MHC with that found in putatively neutral genetic markers such as microsatellites (Schierup *et al.*, 2000; Bernatchez and Landry, 2003). Population structure at microsatellite loci provides a baseline estimate of population subdivision under neutral evolutionary processes. Under balancing selection, it is expected that MHC genetic divergence across populations will be lower than at microsatellites because MHC alleles are less likely to be lost due to random processes (Bernatchez and Landry, 2003). Conversely, if spatially variable selection is driving populations to become uniquely adapted to the environment, populations will exhibit higher levels of divergence at the MHC than at microsatellite markers (for example, Landry and Bernatchez, 2001).

In this study, we examine genetic variation at the MHC in five populations of Chinook salmon (*Oncorhynchus tshawytscha*) occurring in British Columbia, Canada. Salmonid populations have faced significant declines and even extirpation over the past 50 years (Nehlsen *et al.*, 1991; Levin and Schiewe, 2001), and the degree of local adaptation exhibited by functional loci, such as the MHC, may be important when population conservation

and plans for re-introduction are considered (Meyers and Bull, 2002; van Tienderen *et al.*, 2002; Sommer, 2005). Chinook salmon are anadromous, semelparous breeders, in which individuals typically return to their natal freshwater stream to breed after a period of 2–6 years in salt water. Natal philopatry limits gene flow among populations, and thus, Chinook salmon populations are expected to show high levels of local adaptation (Taylor, 1991; Waples *et al.*, 2004). Chinook salmon occur in all major river systems in the north Pacific Rim and the genetic relationships between these populations have been documented using neutral genetic markers (Beacham *et al.*, 2006; Heath *et al.*, 2006). However, studies of population structure and divergence at the MHC in Chinook salmon are limited. Miller *et al.* (1997) reported extensive genetic variation at the MHC class I and II of Chinook salmon; however, the study documented variation in only two populations. In a more extensive study, Miller and Withler (1997) examined variation at the MHC class I across 15 populations, but used a denaturing gradient gel electrophoresis approach that was able to distinguish only about half of the alleles, and thereby significantly underestimated genetic variation within populations and the extent of population genetic structure. Similarly, Heath *et al.* (2006) examined population structure of the MHC class I and II across 10 populations of Chinook salmon; however, the authors used a coarse PCR-RFLP approach capable of detecting variation in only a few common alleles. Given the potential importance of negative frequency-dependent selection in the maintenance of genetic variation at the MHC, an examination of all alleles is clearly important. In this study, we expand on these earlier studies by examining genetic variation across populations of Chinook salmon at the MHC class I-A1 (exon two) and class II-B1 (exon two) using a sequencing and single-strand conformation polymorphism approach capable of detailing all alleles. We also investigate selection at the sequence level by examining patterns of sequence substitutions. Finally, we assess the potential for local adaptation at MHC loci by examining patterns of MHC genetic variation relative to microsatellite genetic variation across populations. Our data add significantly to the understanding of evolution at both class I and II loci of the MHC, and provide important data on five threatened Chinook salmon populations.

Methods

Fish sampling
In the spring of 2006 and 2007, we collected tissue samples from Chinook salmon fry in three rivers located on the east coast of Vancouver Island, British Columbia. These rivers included the Quinsam, Puntledge, and Big Qualicum rivers (Figure 1). In the Puntledge River, there are two distinct Chinook salmon runs (fall and summer). We collected tissue samples from fall-run fry, as the summer run is an endangered population numbering only 400 spawning adults and is maintained primarily through hatchery enhancement (Guimond and Withler, 2006). The Puntledge River fall run was also considerably depleted in the mid-1980s but was subsequently re-established using individuals from the Quinsam and Big Qualicum rivers (Heath *et al.*, 2006). We also

collected samples in the fall of 2006 and 2007 from the Kitsumkalum River (a tributary of the lower Skeena River) in northwestern BC, and in 2007, from the main stem of the lower Skeena River, approximately 40 km west of the confluence of the Skeena and Kitsumkalum rivers (Figure 1). The east coast of Vancouver Island and the Skeena represent two distinct regions, and were expected to exhibit significant differentiation (see Beacham et al., 2006). All fry were captured using a combination of baited gee traps, live traps, or beach seining. Fry were sampled over a 1-week period in May (east coast Vancouver Island) or September (northwestern BC) of each year and were anaesthetized in MS-222 (Tricaine Methanesulfonate; Argent labs, Redmond, WA, USA) so that a tissue sample could be obtained. The tissue samples were stored in 95% ethanol for later genetic analyses.

MHC genotyping

DNA was extracted from tissue using a DNA Wizard Extraction kit (Promega Corp., Madison, WI, USA).
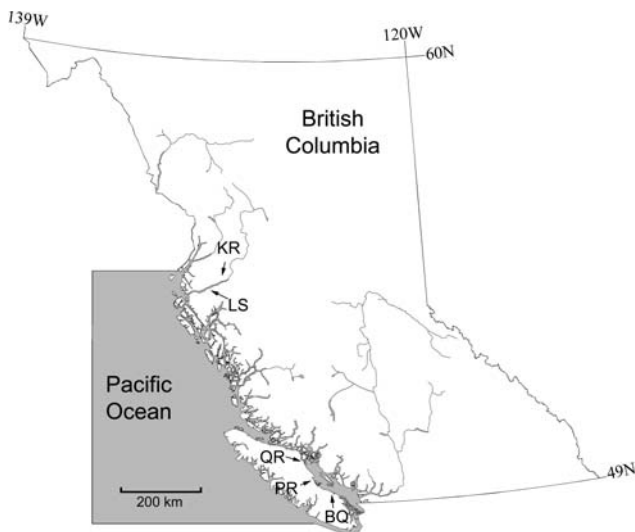

Figure 1 Map of British Columbia indicating the locations of the two northwestern British Columbia and three east coast Vancouver Island populations of Chinook salmon (*Oncorhynchus tshawytscha*). The northwestern British Columbia populations are KR, Kitsumkalum River; and LS, lower Skeena River. The Vancouver Island populations are QR, Quinsam River; PR, Puntledge River; and BQ, Big Qualicum River. Figure modified from Heath *et al.* (2006).

We examined genetic variation within a portion of the MHC class I-A1, and MHC class II-B1 using primer sets described in Grimholt et al. (1993), Hordvik et al. (1993), and Miller et al. (1997). Primers were designed to amplify the exons encoding the PBR located on the α_1 chain of the MHC class I and the β_1 chain of the MHC class II. Both the MHC class I and II studied are likely functional because they are transcribed in closely related species including the Atlantic salmon (Grimholt et al., 1993; Hordvik et al., 1993). We used polymerase chain reaction (PCR) to amplify each locus following the protocols outlined in Miller et al. (1997) and Docker and Heath (2002). The MHC class I-A1 primers amplify either 222 or 228 bp, whereas the MHC class II-B1 primers amplify 213 bp. For both loci, PCR products were visualized using SSCP (single-strand conformation polymorphism). Amplicons were electrophoresed through GeneGels on the Amersham-Biosciences SSCP system using buffer system 'C' at 12 °C following the manufacturer's protocols and gels were fixed and stained using a silver stain (GE Healthcare Bio-Sciences Corp., Piscataway, NJ, USA). Samples that exhibited unique conformations were cloned using the Promega pGEM T-easy vector kit following manufacturer's instructions (Promega Corp.). Between four and eight clones containing inserts were sequenced each time a conformation was sequenced (see Table 1 for breakdown of sequencing sample sizes). Unique single-strand conformations and alleles that were observed only once were verified by retyping the individuals containing the allele. Amplification products were sequenced by Genome Quebec (McGill University, Montreal, Canada).

Sequences were aligned using molecular evolutionary genetics analysis (MEGA) software 4 (http://www.megasoftware.net/; Tamura et al., 2007) and chromatograms were read in FinchTV (Geospiza Inc., Seattle, WA, USA, http://www.geospiza.com). MHC class I sequences were aligned to the human HLA-A1 (Bjorkman et al., 1987) and MHC class II sequences were aligned to the human HLA DRB1 (Brown et al., 1993) to identify the nucleotides that encode the putative PBR for each locus. Sequences were also aligned to previously reported Chinook salmon MHC class I-A1 or II-B1 alleles (Miller et al., 1997).

Sequence-level analyses

For both the MHC class I and II, we calculated rates of nonsynonymous (D_n) and synonymous (D_s) substitutions in MEGA 4 (Tamura et al., 2007). Nonsynonymous

Table 1 Sample sizes of five populations of Chinook salmon (*Oncorhynchus tshawytscha*)

Region	Population	Number of individuals genotyped		Number of clones sequenced	
		MHC class I	*MHC class II*	*MHC class I*	*MHC class II*
East coast Vancouver Island	Big Qualicum	85	90	201	98
	Quinsam	95	104	216	96
	Puntledge	88	91	243	17
Northwestern British Columbia	Kitsumkalum	54	59	155	91
	Lower Skeena	20	21	80	21
	Total	342	365	895	323

Abbreviation: MHC, major histocompatibility complex.
Numbers of individuals and clones sequenced for MHC class I and II markers are indicated.

substitutions involve a change in amino acid sequence and are more likely to undergo selection than are synonymous substitutions. Purifying selection results in D_n/D_s ratios of <1, whereas D_n/D_s ratios >1 indicate that nonsynonymous substitutions are being maintained through processes such as balancing selection (Nei and Hughes, 1991; Ohta, 1991). Standard errors for the rate estimates were based on 1000 bootstrap replications. Z-tests for selection (using the Nei–Gojorbori method applying the Jukes–Cantor correction) were also conducted in MEGA 4. To investigate selection across codons of the MHC loci, we used omegaMap (Wilson and McVean, 2006). OmegaMap uses a Bayesian approach combined with a population genetics approximation of the coalescent with recombination to examine positive selection across DNA sequences (Wilson and McVean, 2006). The model was run twice on population allele frequencies at each MHC locus. In each simulation we ran 500 000 Markov-chain Monte Carlo iterations and the results were thinned every 1000 iterations to obtain the posterior distribution (see Miyake et al., 2009). We discarded the first 10 000 iterations and 20 000 iterations for the MHC class II and I, respectively, as 'burn-in'. We did not have prior knowledge of codon frequencies across the Chinook salmon genome, so it was assumed that all codons had equal equilibrium frequencies in the model. The parameters estimated by omegaMap include the D_n/D_s ratio (ω), recombination rate (ρ), transition transversion rate ratio (κ), the rate of sysnonymous transversion (μ), and the rate of insertion/deletion (φ). We allowed ω and ρ to vary across codons and we used the inverse prior for these variables; average block size at which ω and ρ were estimated was set at 10 and 30, respectively. The remaining variable priors were set to improper_inverse. All priors were set as recommended by Wilson and McVean (2006) for circumstances in which knowledge of parameter distributions is unknown for the study organism.

Population analyses

Estimates of allele and genotype frequencies, observed and expected heterozygosities, F_{ST}, and tests of Hardy–Weinberg equilibrium were calculated using Genepop 3.4 (Raymond and Rousset, 1995). Allele richness, based on a sample size of 20, was calculated in FSTAT 2.9.3 (Goudet, 2002). We tested for allele frequency differences between years using a generalized linear model (GLM) in SPSS 17. We specified the GLM to use a normal distribution, a nominal response variable, and year and population as factors in the model. Mantel tests were performed in Genepop to examine isolation-by-distance relationships between estimates of MHC $F_{ST}/(1-F_{ST})$ and the natural logarithm of the geographic distance along the river and coast (in km) between the collection locations in each population. The significance of this relationship was determined using 1000 permutations in the subroutine Isolde. Analysis of molecular variance across geographic regions, across populations, and within populations was conducted in Arlequin 3.1, using the Jukes–Cantor correction. The statistical significance of the observed variance was determined over 1023 haplotype permutations (Excoffier et al., 2006). Calculation of nucleotide diversity was also conducted in Arlequin 3.1.

We used Micro-Checker (van Oosterhout et al., 2004) to test for the presence of null alleles at the MHC class I and II. MHC class I allele frequencies indicated the likely presence of a null allele in the Vancouver Island and lower Skeena populations, so we used Genepop 3.4 to generate maximum likelihood corrected allele frequencies across these populations (Dempster et al., 1977). Corrected allele frequencies were subsequently used in estimates of population genetic differentiation following Chapius and Estoup's (2007) ENA method (Supplementary information Table S1).

When comparing genetic divergence at microsatellites and MHC loci, we converted our estimates of F_{ST} to G'_{ST} because the magnitude of F_{ST} is dependent on the heterozygosity of the locus examined; G'_{ST} scales estimates of F_{ST} to the maximum amount of genetic variation each locus exhibits and thus accounts for differences in variation among genetic markers (Hedrick, 2005). Thus, we calculated G'_{ST} as $= F_{ST}/F_{ST}(max)$. $F_{ST}(max)$ was defined as $1-Hs$ (see equation (3) in Hedrick, 2005), where Hs is the mean heterozygosity across all subpopulations examined (Hedrick, 2005). Relationships between MHC and 13 microsatellite markers were initially examined by performing a Mantel test that compares estimates of pairwise population divergence for each type of marker. Microsatellite allele frequencies were obtained for individuals sampled over multiple years by the Molecular Genetics Lab of Fisheries and Oceans, Pacific Region (see Beacham et al., 2006; http://www.pac.dfo-mpo.gc.ca/sci/mgl/data_e.htm). A summary of the microsatellite dataset used is provided in Supplementary information Table S2. Correlation coefficients were calculated in SPSS version 16 (SPPS Inc., Chicago, IL, USA), and Mantel tests were conducted in Genepop 3.4, using 1000 permutations to determine statistical significance. We also generated 95% confidence intervals (CIs) for pairwise microsatellite G'_{ST} estimates by bootstrapping over loci 1000 times in FSTAT 2.9.3 (Goudet, 2002). We then compared G'_{ST} estimates for the MHC loci to the 95% CIs, and identified outliers as populations undergoing selection at the MHC (Weir, 1996).

Results

MHC diversity

For both the MHC class I and II, typically 1–2 alleles were identified per individual, suggesting that neither locus is duplicated. For the MHC class I, 49 of the 228 (21%) nucleotide sites were variable, and we identified 37 distinct alleles (Supplementary information Figure S1a). Ten of the 37 (27%) alleles were found in only one population and 5 (14%) of the alleles were found in all populations (Supplementary information Table S1). Seventeen (46%) of the MHC class I alleles were identical to alleles published in Genbank. For the MHC class II, 38 of 213 (18%) nucleotide sites were variable and we found 17 distinct alleles (Supplementary information Figure S1b). Eight (47%) of these alleles were found in only one population, whereas five (29%) alleles were found in all populations (Supplementary information Table S1). Ten (59%) of the MHC class II alleles were identical to alleles published in Genbank. Examining the amino acid sequences, for MHC class I, 28 of 76 (37%) codons were variable. All 37 alleles coded for unique

amino acid sequences. For MHC class II, 25 of 72 (35%) codons were variable and all 17 alleles coded for unique amino acid sequences. Neither locus showed signs of frameshift mutations that would cause the alleles to become non-functional.

Historical selection

For the putative PBR of the MHC class I, the rate of nonsynonymous substitutions was almost twice as high as that of synonymous substitutions; however, the Z-test for selection indicated that this difference was not statistically significant (Table 2). Codons outside of the putative PBR also showed higher rates of nonsynonymous substitutions than synonymous substitutions but again this difference was not statistically significant (Table 2). In contrast, the MHC class II showed a significantly higher rate of nonsynonymous substitutions than synonymous substitutions in the putative PBR (Table 2). However, a statistically higher ratio of nonsynonymous substitutions to synonymous substitutions was not observed for codons outside of the putative PBR for the MHC class II (Table 2).

The omegaMap analysis showed evidence of positive selection (that is $D_n/D_s > 1$) across several MHC classes I and II codons. Within the MHC class I (Figure 2a), codons 54–64 and 69–73 exhibited a posterior probability of positive selection > 95%. All of these codons are associated with the putative PBR of the MHC class I molecule. Codons 58–71 also showed evidence of recombination (lower 95% CI $\rho > 1$). The omegaMap analysis also identified several regions of the MHC class II molecule exhibiting significant evidence of positive selection (Figure 2b). Specifically, codons 10–15, 30–32, 40–42, 50–51, and 54–65 exhibited significant posterior probabilities of positive selection. Codons 10, 15, 31, 56–60, and 63–64 are part of the putative PBR of the MHC class II. Some of the codons exhibiting evidence of positive selection were not identified as part of the putative PBR; however, these codons were located immediately adjacent to the putative PBR (that is codons 11–14, 30, 32, 40–42, 50–51, 54–55, 65). No codons within the MHC class II exhibited evidence of recombination.

Population variation

Allele frequencies did not differ significantly between years for the MHC class I (GLM: $\chi^2_1 = 0.295$, $P = 0.587$, $n = 668$) or the MHC class II (GLM: $\chi^2_1 = 3.183$, $P = 0.074$, $n = 709$). Within populations, the number of alleles at the MHC class I ranged from 15 to 27 (Table 3). Most populations, except for the Kitsumkalum River, exhibited a significant heterozygote deficit (Table 3). In contrast,

the MHC class II exhibited a significant excess of heterozygotes in three populations (Table 3). The MHC class II also exhibited a lower allelic richness than did the MHC class I between 8 and 10 alleles found within populations (Table 3). Mean nucleotide diversity across populations was 0.041 ± 0.011 for the MHC class I and 0.104 ± 0.046 for the MHC class II (Table 3). No single MHC class I allele was most common in all populations. Allele 12 was most common in the Big Qualicum and Puntledge rivers, at 15% and 14% of alleles, respectively, allele 3 was most common in the Quinsam River at 28% of alleles, allele 32 was most common in the Kitsumkalum River at 40% of alleles, and allele 32, at 14% of alleles, was most common in the lower Skeena River (Supplementary information Table S1). However, the MHC class II allele 2 was the most common allele found in all five of our study populations, though the frequency of the allele varied between 32% and 73% of alleles among populations (Supplementary information Table S1).

For the MHC class I, the analysis of molecular variance indicated that the majority of the variance occurred within populations but significant differentiation occurred among populations within each region (Table 4). Near-significant geographic differentiation also occurred between regions for the MHC class I. A similar analysis of molecular variance result was shown for the MHC class II, in which the majority of the variance occurred within populations (Table 4). However, near-significant geographic differentiation also occurred among regions and among populations within regions.

Among population divergence

Pairwise population genetic differentiation at the MHC class I, as estimated by F_{ST}, ranged from –0.002 to 0.155 (Table 5). Estimates of F_{ST} for the MHC class II ranged from –0.002 to 0.167 (Table 5). We did not detect a significant isolation-by-distance relationship for either MHC class I (Mantel test: $r = 0.38$, $P = 0.388$, Figure 3) or the MHC class II (Mantel test: $r = 0.58$, $P = 0.109$, Figure 3).

Pairwise estimates of MHC class I G'_{ST} and the G'_{ST} of microsatellites were only nearly significantly correlated (Mantel test: $r = 0.46$, $P = 0.065$) whereas those for the MHC class II and microsatellites were significantly correlated (Mantel test: $r = 0.80$, $P = 0.03$). However, we found that the overall MHC class I G'_{ST} (0.424) was significantly higher than the G'_{ST} for microsatellites (microsatellite $G'_{ST} = 0.183$, 95% CI = 0.125–0.258). In contrast, the overall MHC class II G'_{ST} (0.144) fell within the lower 95% CI for the microsatellite G'_{ST}. Moreover, the majority (60%) of the pairwise G'_{ST} estimates for the

Table 2 Summary of nucleotide substitution rates in the MHC of Chinook salmon (*Oncorhynchus tshawytscha*)

	PBR				Non-PBR			
	D_s	D_n	Z	P	D_s	D_n	Z	P
MHC class I	0.058 ± 0.035	0.112 ± 0.036	1.035	0.151	0.015 ± 0.008	0.025 ± 0.009	1.073	0.143
MHC class II	0.002 ± 0.002	0.094 ± 0.024	3.833	< 0.001	0.027 ± 0.016	0.043 ± 0.013	0.905	0.184

Abbreviation: MHC, major histocompatibility complex.
Rates of synonymous (D_s) and nonsynonymous (D_n) substitutions (with a Jukes–Cantor correction) across codons of the putative peptide-binding region (PBR) and non-PBR in the Chinook salmon MHC class I-A1 and class II-B1 are indicated. Results of the Z-test for selection across codons of the PBR and non-PBR are also indicated. All error terms are standard errors.

Figure 2 Codon-level analysis of positive selection across MHC sequences of Chinook salmon (*Oncorhynchus tshawytscha*). Analyses comprise the MHC class I-A1 exon (**a**) and MHC class II-B1 exon (**b**) with upper plots within each figure showing estimates of positive selection and lower plots showing the posterior probability of positive selection. Mean omegaMap estimates of positive selection (ω) are indicated by the solid black line. The dotted line in the upper plots corresponds to neutral estimates of nonsynonymous/synonymous substitutions (that is $\omega = 1$) and in the lower plots correspond to the 95% posterior probability. The grey shaded area shows the 95% CI surrounding the estimates of positive selection. Stars indicate the locations of the putative peptide-binding regions (PBR) for each molecule. Putative PBRs were identified by aligning the MHC sequences in MEGA with human HLA molecules.

MHC class I fell above the 95% CIs estimated from microsatellites (Figure 4a) showing that some populations of Chinook salmon are more divergent at the MHC class I than expected based on neutral estimations. The Puntledge-Big Qualicum MHC class I G'_{ST} was slightly lower than the G'_{ST} of microsatellites, whereas only the Big Qualicum-lower Skeena, and Quinsam-lower Skeena, and Puntledge-lower Skeena G'_{ST} estimates were not significantly different than the G'_{ST} of microsatellites (Figure 4a). At the MHC class II, 50% of the pairwise G'_{ST} estimates were significantly lower than at the microsatellites, and the majority of these values occurred in pairwise comparisons within geographic regions (Figure 4b).

Discussion

The MHC represents one of the most variable genetic regions in vertebrate genomes. In this study, we detected high allelic diversity at both the MHC class I and II loci across five populations of Chinook salmon. We found 37 alleles at the MHC class I and 17 alleles at the MHC class II in 342 and 365 individuals, respectively. These levels of allelic variation exceed the levels of variation that have been reported earlier for Chinook salmon (Miller *et al.*, 1997; Miller and Withler, 1997; Kim *et al.*, 1999; Garrigan and Hedrick, 2001); although this difference could reflect our relatively large sample size or increased allele detection sensitivity obtained through an
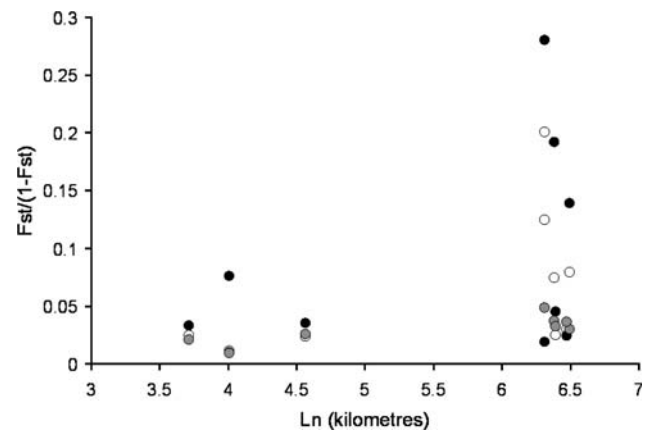
Table 3 Summary of genetic variation across five populations of Chinook salmon (*Oncorhynchus tshawytscha*)

Locus	East coast Vancouver Island			Northwestern British Columbia	
	Big Qualicum	Puntledge	Quinsam	Kitsumkalum	Lower Skeena
Class I					
H_E	0.928	0.925	0.856	0.792	0.925
H_O	0.706[a]	0.704[a]	0.695[a]	0.839	0.500[a]
A	27	25	21	15	15
AR	16.5	15.5	13.2	10.2	15.0
π	0.039	0.036	0.035	0.060	0.035
Class II					
H_E	0.628	0.618	0.449	0.739	0.805
H_O	0.778[b]	0.714[b]	0.471	0.915[b]	0.714
A	10	8	9	8	10
A_R	6.3	5.6	5.1	5.6	9.9
π	0.066	0.096	0.057	0.161	0.141

Expected (H_E) and observed (H_O) heterozygosity, number of alleles (A), allelic richness (A_R), and nucleotide diversity (π) of MHC class I-A1 and class II-B1 are indicated.
[a]Significant excess of homozygotes at $P<0.05$.
[b]Significant excess of heterozygotes at $P<0.05$.

Figure 3 Isolation-by-distance for pairwise comparisons of five populations of Chinook salmon (*Oncorhynchus tshawytscha*). Each point represents a pairwise comparison (F_{ST}) between the MHC class I (black circles), MHC class II (open circles), and microsatellites (grey circles) and natural log of the geographic distance.

Table 4 Summary of analysis of molecular variance (AMOVA) results based on five populations of Chinook salmon (*Oncorhynchus tshawytscha*)

Locus	Source of variation	d.f.	SS	Variance components	Percentage of variation	Fixation index	P
MHC class I	Between regions	1	98.15	0.33	6.68	0.067	0.10
	Among populations within regions	3	70.55	0.14	2.78	0.029	<0.001
	Within populations	675	3059.33	4.48	90.54	0.095	<0.001
MHC class II	Between regions	1	337.08	1.27	11.48	0.115	0.10
	Among populations within regions	3	70.79	0.09	0.86	0.010	0.08
	Within populations	731	7003.41	9.66	87.66	0.123	<0.001

Sources of genetic variation at the MHC class I and II comprise within population, among population, and between regions. Significance tests are the result of 1023 permutations.

Table 5 Summary of genetic differentiation occurring among five populations of Chinook salmon (*Oncorhynchus tshawytscha*)

	Big Qualicum		Puntledge		Quinsam		Kitsumkalum		Lower Skeena	
	Class I	Class II	Class I	Class II	Class I	Class II	Class I	Class II	Class I	Class II
Big Qualicum			0.369	0.892	<0.001	0.004	<0.001	<0.001	0.002	<0.001
Puntledge	−0.002	−0.002			<0.001	0.001	<0.001	<0.001	0.002	<0.001
Quinsam	0.028	0.023	0.029	0.024			<0.001	<0.001	<0.001	<0.001
Kitsumkalum	0.123	0.073	0.123	0.070	0.155	0.167			<0.001	0.029
Lower Skeena	0.025	0.029	0.024	0.025	0.046	0.111	0.104	0.011		

Pairwise F_{ST} values for the MHC class I and II are shown below the diagonal and significance tests for the corresponding F_{ST} estimates are shown above the diagonal.
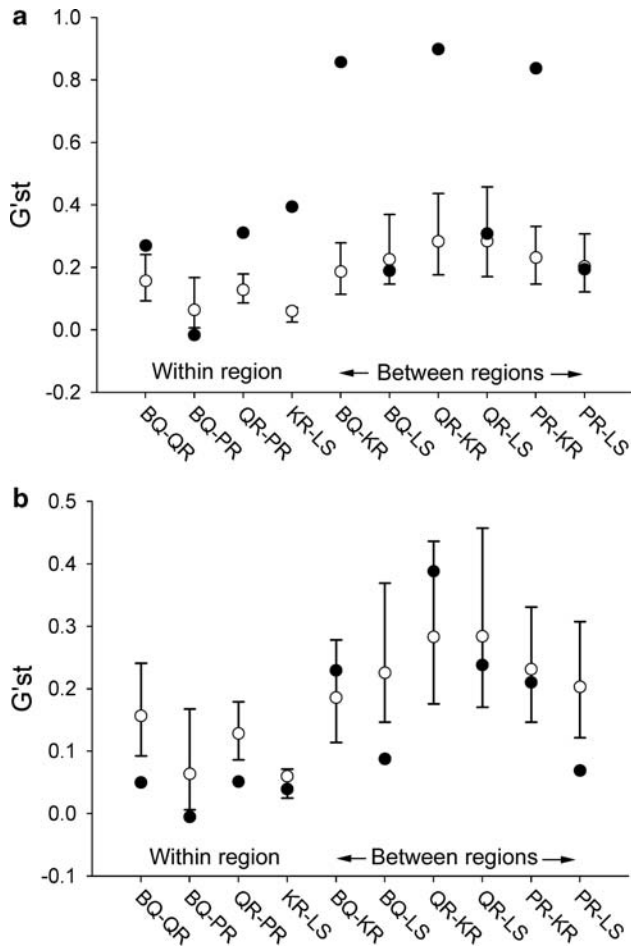
Figure 4 Relationship between genetic divergence (G'_{ST}) at MHC and microsatellite loci in five populations of Chinook salmon (*Oncorhynchus tshawytscha*). Closed circles indicate pairwise population divergence estimates at MHC class I (**a**) and class II (**b**). Population genetic divergence at microsatellite loci is indicated by open circles and error bars indicate 95% CIs. Populations are KR, Kitsumkalum River; LS, lower Skeena River; QR, Quinsam River; PR, Puntledge River; and BQ, Big Qualicum River. Note that the scales differ on the y-axes.

SSCP and sequencing approach. In a study of approximately 40 individuals from two populations, Miller *et al.* (1997) described 22 alleles at the Chinook salmon MHC class I-A1 locus and 3 alleles at the MHC class II-B1 locus, whereas Garrigan and Hedrick (2001) detected 12 MHC class I-A1 alleles in 44 individuals. Studies of other fishes have found comparatively high numbers of alleles at the MHC class II as were identified in this study. Landry and Bernatchez (2001) reported 18 alleles across 14 populations of Atlantic salmon (*Salmo salar*) and Miller *et al.* (2001) reported 11 alleles in 31 populations of Sockeye salmon (*O. keta*). The allelic diversity we observed in Chinook salmon provides additional evidence that the MHC is one of the most polymorphic regions of vertebrate genomes.

Positive selection is frequently invoked to explain the high levels of polymorphism observed at the genes of the MHC (Hughes and Nei, 1988). Evidence for positive selection can be assessed through the comparison of synonymous (D_s) and nonsynonymous (D_n) substitution

rates within the PBR and also by comparing substitution rates within and outside of the PBR (Hill and Hastie, 1987; Hughes and Nei, 1988). We found a $D_n/D_s>1$ in the putative PBR of both MHC loci, which indicates that nonsynonymous substitutions are being maintained at a rate that is higher than expected under neutral evolution. Moreover, the D_n/D_s ratio in the putative PBR was higher than the D_n/D_s ratio observed in the non-PBR for both loci, albeit it was only significantly higher for the MHC class II. We also tested for positive selection across regions of the MHC molecules using a codon model that incorporates recombination in omegaMap. For both the MHC class I and II we found several regions of the molecules exhibiting evidence of positive selection. For the MHC class I, all of the codons exhibiting evidence of positive selection were associated with the putative PBR, which provides strong evidence that amino acids associated with this region are undergoing selection for diversity. We also found evidence that the putative PBR is undergoing recombination, suggesting that recombination may also be an important mechanism generating genetic diversity within the MHC. For the MHC class II, we found that codons located within or immediately adjacent to the putative PBR appear to be undergoing selection for amino acid diversity. One of those codons is not part of the HLA PBR (codon 31 here and codon 53 in the HLA) but has previously been shown to exhibit evidence of positive selection, leading to the suggestion that the codon may actually be part of the salmon MHC PBR (Aguilar and Garza, 2007). However, a structural model of the Chinook salmon MHC class II molecule will be necessary to confirm this suggestion. Together our analyses indicate that positive selection is stronger in the PBR of both the MHC class I and II in Chinook salmon, as has been shown in other studies (reviewed in Bernatchez and Landry, 2003; van Oosterhout *et al.*, 2006; Dionne *et al.*, 2007), and suggest that historical positive selection is maintaining amino acid diversity within the MHC PBR.

Under balancing selection, heterozygous individuals are selectively favoured either through increased ability to mount an immune response against pathogens (Doherty and Zinkernagel, 1975). For the MHC class II, we found that three populations, the Big Qualicum, Puntledge, and Kitsumkalum, exhibited heterozygosity that significantly exceeded Hardy–Weinberg expectations. Other studies of the MHC have also found higher levels of observed heterozygosity than expected, suggesting that selection for heterozygotes is occurring in many species (for example, Seddon and Baverstock, 1999; van Haeringen *et al.*, 1999; Peters and Turner, 2008; Oliver *et al.*, 2009; see Garrigan and Hedrick, 2003 for review). In a recent study, we have shown that MHC class II heterozygous Chinook salmon fry exhibit fewer bacterial infections that do homozygotes (Evans and Neff, 2009). Moreover, a study by Arkush *et al.* (2002) on Chinook salmon showed that heterozygotes at the MHC class II exhibited higher survival than homozygotes when challenged with infectious haematopoietic necrosis virus (but see Langefors *et al.*, 2001). Given the potential prevalence of infectious haematopoietic necrosis virus in aquatic systems (Meyers, 1998), it is a possible source of selection pressure on heterozygotes in wild Chinook salmon populations. Overall, our population genetic data, in combination with studies that have examined

infections among MHC class II heterozygous and homo-zygous Chinook salmon, suggest that balancing selection is operating at this locus.

In contrast to the heterozygote advantage hypothesis, we detected a heterozygote deficit for the MHC class I in all of our study populations except for the Kitsumkalum. This deficit could be mediated by underdominance genetic effects, in which heterozygous genotypes exhibit lower survival than do homozygotes. Such effects have been shown in one study of Chinook salmon, albeit at the MHC class II locus (Pitcher and Neff, 2006). In Atlantic salmon it has been recently shown that fertilization success is higher between MHC class I similar mates (Yeates et al., 2009), which may reflect underdominance. Alternatively, it is possible that the observed hetero-zygote deficit reflects genetic drift occurring within populations or the presence of a null allele (Manwell and Baker, 1970). In the Malagasy giant jumping rat (Hypogeomys antimena), genetic drift was implicated in heterozygote deficits observed at the MHC (Sommer, 2003). Miller and Withler (1997) detected a heterozygote deficit at the MHC class I in the Big Qualicum Chinook salmon population, as well as in a population in the Skeena River and the authors speculated that a null allele was the cause. A study of MHC class I inheritance will be necessary to determine whether a null allele is indeed the cause of the heterozygote deficit.

Allele frequencies within populations reflect the out-come of selection or neutral processes such as gene flow, and drift. Estimates of genetic divergence at microsatellites and the MHC class II were significantly correlated, and near significantly correlated with the MHC class I, which suggests that population genetic divergence at the MHC may, in part, be driven by neutral evolutionary forces. In contrast, there was no apparent isolation-by-distance at the MHC, which is unexpected if neutral processes are important. Furthermore, several population pairwise comparisons of divergence exhib-ited higher differentiation at the MHC class I and lower differentiation at the class II relative to microsatellite loci. These data suggest that the class I region is subject to differential selection pressures across populations whereas the class II region is subject to homogenizing or balancing selection across populations. Differential selection at MHC classes I and II loci is highly plausible given that the genetic regions encoding for these loci are unlinked in fishes (Klein et al., 1997). Interestingly, Heath et al. (2006) instead reported higher genetic differ-entiation at the MHC class II than at microsatellite loci in the same species as we studied. However, as mentioned previously, their study used a PCR-RFLP approach to examine population genetic variation and thus, is not directly comparable to our results. Similar to our findings, a study of the guppy (Poecilia reticulata) found lower population differentiation at the MHC class II than at microsatellites (Fraser et al., 2009). The authors of that study suggest that the genetic similarity across popula-tions is due to homogenizing directional selection driven by a common parasite across populations (Fraser and Neff, 2009). Yet studies on other fish species (for example, O. keta, Miller et al., 2001; S. salar, Landry and Bernatchez, 2001; O. mykiss, Aguilar and Garza, 2006), birds (for example, Gallinago media, Ekblom et al., 2007), and mammals (for example, Arvicola terrestris, Bryja et al., 2007) have shown higher levels of divergence at the MHC when compared with microsatellites, or in a few cases similar levels of genetic divergence at MHC markers and at microsatellite markers (for example, Boyce et al., 1997; Peters and Turner, 2008). Overall divergence trends at the MHC class I, in combination with the maintenance of high levels of genetic variability, suggest that spati-ally variable balancing selection is an important factor driving population differences in allele frequencies. At the MHC class II, selection appears to also play a role, but is driving some populations to exhibit similar popu-lation genetic structure.

For species in decline, such as Chinook salmon, it is clear that we must consider adaptive divergence at functionally important genes in recovery efforts. Our results indicate that selection is driving population variation at the Chinook salmon MHC. This variation may relate to differences in pathogen communities across regions, but further studies are required to address the importance of specific pathogens in the evolution of population differences.

Acknowledgements

References

Aguilar A, Garza JC (2006). A comparison of variability and population structure for major histocompatibiltiy complex and microsatellite loci in California coastal steelhead (Oncorhynchus mykiss Walbaum). Mol Ecol 15: 923–937.

Aguilar A, Garza JC (2007). Patterns of historical balancing selection on the salmonid major histocompatibility complex class II β gene. J Mol Evol 65: 34–43.

Arkush KD, Giese AR, Mendonca HL, McBride AM, Marty GD, Hedrick PW (2002). Resistance to three patho-gens in the endangered winter-run Chinook salmon (Oncorhynchus tshawytscha): effects of inbreeding and major histocompatibility complex genotypes. Can J Fish Aquat Sci 59: 966–975.

Beacham TD, Jonson KL, Supernault J, Wetklo M, Deng LT, Varnavskaya N (2006). Pacific Rim population structure of Chinook salmon as determined from microsatellite analysis. Trans Am Fish Soc 135: 1604–1621.

Bernatchez L, Landry C (2003). MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? J Evol Biol 16: 363–377.

Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, Wiley DC (1987). The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature* **329**: 512–518.

Boyce WM, Hedrick PW, Muggli-Cockett NE, Kalinowski S, Penedo MCT, Ramey III RR (1997). Genetic variation of major histocompatibility complex and microsatellite loci: a comparison in bighorn sheep. *Genetics* **145**: 421–433.

Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL *et al.* (1993). Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* **364**: 33–39.

Bryja J, Charbonnel K, Berthier K, Galan M, Cosson J-F (2007). Density-related changes in selection pattern for major histocompatibility complex genes in fluctuating populations of voles. *Mol Ecol* **16**: 5084–5097.

Chapius M-P, Estoup A (2007). Microsatellite null alleles and estimation of population differentiation. *Mol Biol Evol* **24**: 621–631.

Dempster AP, Laird NM, Rubin DB (1977). Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc B* **39**: 1–38.

Dionne M, Miller KM, Dodson JJ, Caron F, Bernatchez L (2007). Clinal variation in MHC diversity with temperature: evidence for the role of host-pathogen interaction on local adaptation in Atlantic salmon. *Evolution* **61**: 2154–2164.

Docker MF, Heath DD (2002). PCR-based markers detect genetic variation at growth and immune function-related loci in Chinook salmon (*Oncorhynchus tshawytscha*). *Mol Ecol Notes* **2**: 606–609.

Doherty PC, Zinkernagel RM (1975). Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature* **256**: 50–52.

Ekblom R, Saether SA, Jacobsson P, Fiske P, Sahlman T, Grahn M *et al.* (2007). Spatial pattern of MHC class II variation in the great snipe (*Gallingo media*). *Mol Ecol* **16**: 1439–1451.

Evans ML, Neff BD (2009). MHC heterozygote advantage and widespread bacterial infections in populations of Chinook salmon (*Oncorhynchus tshawytscha*). *Mol Ecol* (in press).

Excoffier L, Laval G, Schneider S (2006). Arlequin v 3.1: an integrated software package for population genetics data analysis. University of Berne, Switzerland.http://cmpg.unibe.ch/software/arlequin3.

Fraser BA, Neff BD (2009). Parasite mediated homogenizing selection at the MHC in guppies. *Genetica* (in press; doi:10.1007/s10709-009-9402-y).

Fraser BA, Ramnarine IW, Neff BD (2009). Selection at the MHC class IIB locus across populations in the guppy (*Poecilia reticulata*). *Heredity* (in press; doi:10.1038/hdy.2009.99).

Garrigan D, Hedrick PW (2001). Class I MHC polymorphism and evolution in endangered California Chinook and other pacific salmon. *Immunogenetics* **53**: 483–489.

Garrigan D, Hedrick PW (2003). Perspective: detecting adaptive molecular polymorphism: lessons from the MHC. *Evolution* **57**: 1707–1722.

Goudet J (2002). Fstat version 2.9.3.2.. UNIL, Switzerland. http://www2.unil.ch/popgen/softwares/fstat.htm.

Grimholt U, Hordvik I, Fosse VM, Olsaker I, Endresen C, Lie O (1993). Molecular cloning of major histocompatibility complex class I cDNAs from Atlantic salmon (*Salmo salar*). *Immunogenetics* **37**: 469–473.

Grimholt U, Larsen S, Nordmo R, Midtlyng P, Kjoeglum S, Storset A *et al.* (2003). MHC polymorphism and disease resistance in Atlantic salmon (*Salmo salar*); facing pathogens with single expressed major histocompatibility class I and class II loci. *Immunogenetics* **55**: 210–219.

Guimond E, Withler R (2006). *Puntledge River Summer Chinook DNA Analyses 2006*. Comox Valley Project Watershed Society: Courtney, BC.

Heath DD, Shrimpton JM, Hepburn RI, Jamieson SK, Brode SK, Docker MF (2006). Population structure and divergence using microsatellite and gene locus markers in Chinook salmon (*Oncorhynchus tshawytscha*) populations. *Can J Fish Aquat Sci* **63**: 1370–1383.

Hedrick PW (1999). Balancing selection and MHC. *Genetica* **104**: 207–214.

Hedrick PW (2002). Pathogen resistance and genetic variation at MHC loci. *Evolution* **56**: 1902–1908.

Hedrick PW (2005). A standardized genetic differentiation measure. *Evolution* **59**: 1633–1638.

Hill AVS, Allsopp CEM, Kwiatkowski D, Anstey NM, Twumasi P, Rowe PA *et al.* (1991). Common west African HLA antigens are associated with protection from severe malaria. *Nature* **352**: 595–600.

Hill AVS, Yates SNR, Allsopp CEM, Guptra S, Gilbert SC, Lalvani A *et al.* (1994). Human leukocyte antigens and natural selection by malaria. *Philos Trans R Soc Lond B Biol Sci* **346**: 379–385.

Hill RE, Hastie ND (1987). Accelerated evolution in the reactive centre regions of serine protease inhibitors. *Nature* **326**: 96–99.

Hordvik I, Grimholt U, Fosse VM Lie O, Endersen C (1993). Cloning and sequence analysis of cDNAs encoding the MHC class II β chain in Atlantic salmon (*Salmo salar*). *Immunogenetics* **37**: 437–441.

Hughes AL, Hughes MK (1995). Self peptides bound by HLA class I molecules are derived from highly conserved regions of a set of evolutionarily conserved proteins. *Immunogenetics* **41**: 257–262.

Hughes AL, Nei M (1988). Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335**: 167–170.

Hughes AL, Yeager M (1998). Natural selection at major histocompatibility complex loci of vertebrates. *Annu Rev Genet* **32**: 415–435.

Johnson NA, Vallejo RL, Silverstein JT, Welch TJ, Wiens GD, Hallerman EM *et al.* (2008). Suggestive association of major histocompatibility IB genetic markers with resistance to bacterial coldwater disease in rainbow trout (*Oncorhynchus mykiss*). *Mar Biotechnol* **10**: 429–437.

Kean RP, Briles WE, Cahaner A, Freeman AE, Lamont SJ (1994). Differences in major histocompatibility complex frequencies after multitrait, divergent selection for immunocompetence. *Poult Sci* **73**: 7–17.

Kim TJ, Parker KM, Hedrick PW (1999). Major histocompatibility complex differentiation in Sacramento River Chinook salmon. *Genetics* **151**: 1115–1122.

Klein J (1986). *Natural History of the Major Histocompatibility Complex*. Wiley: New York.

Klein J, Figueroa F, Klein D, Sato A, O'hUigin C (1997). Major histocompatibility complex genes in the study of fish phylogeny. In: Kocher TD, Stepian CA (eds). *Molecular Systematics of Fishes*. Academic Press: London. pp 271–283.

Landry C, Bernatchez L (2001). Comparative analysis of population structure across environments and geographical scales at major histocompatibility complex and microsatellite loci in Atlantic salmon (*Salmo salar*). *Mol Ecol* **10**: 2525–2539.

Langefors A, Lohm J, Grahn M, Andersen O, von Schantz T (2001). Association between major histocompatibility complex class IIB alleles and resistance to *Aeromonas salmonicida* in Atlantic salmon. *Proc Biol Sci* **268**: 479–485.

Langefors A, von Schantz T, Widegren B (1998). Allelic variation of MHC class II in Atlantic salmon: a population genetic analysis. *Heredity* **80**: 568–575.

Levin PS, Schiewe MH (2001). Preserving salmon biodiversity. *Am Sci* **89**: 220–227.

Manwell C, Baker CMA (1970). *Molecular Biology and the Origin of Species*. University of Washington Press: Seattle, WA.

Matsumura M, Fremont DH, Peterson PA, Wilson IA (1992). Emerging principles for the recognition of peptide antigens by MHC class I molecules. *Science* **257**: 927–934.

Meyers LA, Bull JJ (2002). Fighting change with change. *Trends Ecol Evol* **17**: 551–557.

Meyers TR (1998). Healthy juvenile sockeye salmon reared in virus-free hatchery water return as adults infected with infectious hematopoietic necrosis virus (IHNV): a case report and review of controversial issues in the epizootiology of IHNV. *J Aquat Anim Health* **10**: 172–181.

Miller KM, Kaukinen KH, Beacham TD, Withler RE (2001). Geographic heterogeneity in natural selection on an MHC locus in sockeye salmon. *Genetica* **111**: 237–257.

Miller KM, Winton JR, Schulze AD, Purcell MK, Mling TJ (2004). Major histocompatibility complex loci are associated with susceptibility of Atlantic salmon to infectious hematopoietic necrosis virus. *Environ Biol Fishes* **69**: 307–316.

Miller KM, Withler RE (1997). MHC diversity in Pacific salmon: population structure and trans-species allelism. *Hereditas* **127**: 83–95.

Miller KM, Withler RE, Beacham TD (1997). Molecular evolution at MHC genes in two populations of Chinook salmon *Oncorhynchus tshawytscha*. *Mol Ecol* **6**: 937–954.

Miyake T, Takebayashi N, Wolf DE (2009). Possible diversifying selection in the imprinted gene, *MEDEA*, in *Arabidopsis*. *Mol Biol Evol* **26**: 843–857.

Nehlsen W, Williams JE, Lichatowich JA (1991). Pacific salmon at the crossroads: stocks at risk from California, Oregon, Idaho, and Washington. *Fisheries* **16**: 4–21.

Nei M, Hughes AL (1991). Polymorphism and evolution of the major histocompatibility complex loci in mammals. In: Selander RK, Clark RG, Whittam TS (eds). *Evolution at the Molecular Level*. Sinauer Associates Inc.: Sunderland, MA. pp 222–247.

Ohta T (1991). Role of diversifying selection and gene conversion in evolution of major histocompatibility complex loci. *Proc Natl Acad Sci USA* **88**: 6716–6720.

Oliver MK, Lambin X, Cornulier T, Piertney SB (2009). Spatio-temporal variation in the strength and mode of selection acting on major histocompatibility complex diversity in water vole (*Arvicola terrestris*) metapopulations. *Mol Ecol* **18**: 80–92.

Paterson S, Wilson K, Pemberton JM (1998). Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population (*Ovis aries* L). *Proc Natl Acad Sci USA* **95**: 3714–3719.

Peters MB, Turner TF (2008). Genetic variation of the major histocompatibility complex (MHC class II beta gene) in the threatened Gila trout, *Oncorhynchus gilae gilae*. *Conserv Genet* **9**: 257–270.

Piertney SB, Oliver MK (2006). The evolutionary ecology of the major histocompatibility complex. *Heredity* **96**: 7–21.

Pitcher TE, Neff BD (2006). MHC Class IIB alleles contribute to both additive and nonadditive genetic effects on survival in Chinook salmon. *Mol Ecol* **15**: 2357–2365.

Raymond M, Rousset F (1995). Genepop (version 1.2): population genetics software for exact tests and ecumenicism. *J Hered* **86**: 248–249.

Schierup M, Vekemans X, Charlesworth D (2000). The effect of subdivision on variation at multi-allelic loci under balancing selection. *Genet Res* **76**: 51–62.

Seddon JM, Baverstock PR (1999). Variation on islands: major histocompatibility complex (MHC) polymorphism in populations of the Australian bush rat. *Mol Ecol* **8**: 2071–2079.

Solberg OD, Mack SJ, Lancaster AK, Single RM, Tsai Y, Sanchez-Mazas A *et al.* (2008). Balancing selection and heterogeneity across the classical human leukocyte antigen loci: a meta-analytic review of 497 population studies. *Hum Immunol* **69**: 443–464.

Sommer S (2003). Effects of habitat fragmentation and changes of dispersal behaviour after a recent population decline on the genetic variability of noncoding and coding DNA of a monogamous Malagasy rodent. *Mol Ecol* **12**: 2845–2851.

Sommer S (2005). The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Front Zool* **2**: 16–34.

Summers K, McKeon S, Sellars J, Keusenkothen M, Morris J, Gloeckner D *et al.* (2003). Parasitic exploitation as an engine of diversity. *Biol Rev* **78**: 639–675.

Tamura K, Dudley J, Nei M, Kumar S (2007). MEGA 4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**: 1596–1599.

Taylor EB (1991). A review of local adaptation in Salmonidae, with particular reference to Pacific and Atlantic salmon. *Aquaculture* **98**: 185–207.

van Haeringen WA, Gwakisa PS, Mikko S, Eythorsdottir E, Holm LE, Olsaker I *et al.* (1999). Heterozygosity excess at the cattle DRB locus revealed by large scale genotyping of two closely linked microsatellites. *Anim Genet* **30**: 169–176.

van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004). Micro-Checker: software for identifying and correcting geno-typing errors in microsatellite data. *Mol Ecol Notes* **4**: 535–538.

van Oosterhout C, Joyce DA, Cummings SM (2006). Balancing selection, random genetic drift, and genetic variation at the major histocompatibility complex in two wild populations of guppies (*Poecilia reticulata*). *Evolution* **60**: 2562–2574.

van Tienderen PH, de Haan AA, van der Linden CG, Vosman B (2002). Biodiversity assessment using markers for ecologically important traits. *Trends Ecol Evol* **17**: 577–582.

Vasemagi A, Primmer CR (2005). Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Mol Ecol* **14**: 3623–3642.

Waples RS, Teel DJ, Myers JM, Marshall AR (2004). Life-history divergence in Chinook salmon: historic contingency and parallel evolution. *Evolution* **58**: 386–403.

Wegner KM, Reusch TBH, Kalbe M (2003). Multiple parasites are driving major histocompatibility complex polymorphism in the wild. *J Evol Biol* **16**: 224–232.

Weir BS (1996). *Genetic Data Analysis II: Methods for Discrete Population Genetic Data*. Sinauer Associates Inc.: Sunderland, MA.

Wilson DJ, McVean G (2006). Estimating diversifying selection and functional constraint in the presence of recombination. *Genetics* **172**: 1411–1425.

Wright S (1931). Evolution in Mendelian populations. *Genetics* **16**: 97–159.

Yeates SE, Einum S, Fleming IA, Megens H-J, Stet RJM, Hindar K *et al.* (2009). Atlantic salmon eggs favour sperm in competition that have similar major histocompatibility alleles. *Proc Biol Sci* **276**: 559–566.

Supplementary information accompanies the paper on Heredity website (http://www.nature.com/hdy)

Supplemental Materials
Table S1: Allele frequencies at the MHC class I-A1 and MHC class II-B1 across five populations of Chinook salmon (*Oncorhynchus tshawytscha*)

MHC class I	East coast Vancouver Island			North coast British Columbia	
Alleles	Big Qualicum	Puntledge	Quinsam	Kitsumkalum	Lower Skeena
1	0.047	0.039	0.016	0	0.119
2	0.087	0.073	0.156	0.009	0.048
3	0.087	0.096	0.281	0.009	0.119
4	0.012	0.006	0.031	0	0
5	0.012	0.067	0.042	0.027	0.024
6	0.070	0.101	0.031	0.009	0
7	0.087	0.079	0.047	0	0.024
9	0.006	0.028	0.042	0.054	0
10	0.017	0.022	0.016	0	0.048
11	0	0.006	0.005	0	0
12	0.145	0.140	0.052	0	0.024
14	0.006	0	0.010	0	0
15	0	0	0.010	0	0.024
16	0	0	0.010	0	0
17	0.058	0.022	0.073	0	0
18	0.012	0	0	0	0
19	0.023	0.017	0.010	0	0
20	0.012	0.006	0.010	0.152	0
23	0.035	0.028	0	0	0
24	0.017	0.006	0	0	0
25	0.023	0.006	0.026	0.107	0.048
26	0.017	0.011	0.005	0	0.024
28	0.006	0	0	0	0
29	0.035	0.034	0	0	0.048
30	0	0.011	0	0	0
32	0.023	0.028	0.005	0.402	0.143
38	0	0.006	0	0.107	0.024
40	0	0	0	0.045	0.024
41	0	0.017	0	0	0
42	0.006	0.006	0.005	0.009	0
43	0.006	0	0	0	0
45	0.006	0	0	0	0
46	0.012	0	0	0.009	0.024
47	0	0.006	0	0.027	0
48	0.006	0	0	0	0
50	0	0	0	0.027	0
51	0	0	0	0.009	0
NULL	0.128	0.140	0.115	0	0.238

S1

MHC class II					
2	0.572	0.577	0.726	0.322	0.381
3	0.178	0.165	0.144	0.322	0.167
4	0.011	0.011	0.019	0.017	0.048
5	0	0	0.005	0	0
6	0.011	0	0.005	0	0
7	0.067	0.038	0.014	0	0
8	0.05	0.055	0.01	0.008	0.048
9	0.006	0.005	0	0.008	0
12	0.006	0.005	0	0.085	0.071
13	0.094	0.143	0.067	0.229	0.167
19	0.006	0	0	0	0
20	0	0	0	0.008	0
21	0	0	0	0.008	0
22	0	0	0	0	0.024
23	0	0	0	0	0.048
24	0	0	0	0	0.024
25	0	0	0.01	0	0

Supplementary Materials Table S2: Summary of microsatellite loci used to derive neutral estimates of genetic differentiation among Chinook salmon (*Oncorhynchus tshawtyscha*) populations. The name of each microsatellite locus is indicated in addition to the locus genetic diversity (*H*), number of alleles (*A*), and sample size (*N*) within each population.

Locus[a]	Big Qualicum		Quinsam		Puntledge		Kitsumkalum		Lower Skeena	
	H	*A/N*	*H*	*A/N*	*H*	*A/N*	*H*	*A/N*	*H*	*A/N*
Ots 2	0.849	17/339	0.840	16/416	0.853	18/497	0.675	14/420	0.688	14/667
Ots 9	0.531	4/350	0.558	6/430	0.501	7/542	0.558	5/423	0.521	6/686
Ots 101	0.925	25/287	0.923	26/409	0.921	25/489	0.947	33/441	0.953	31/672
Ots 100	0.954	42/291	0.967	40/408	0.963	47/520	0.953	34/435	0.947	35/670
Ots 102	0.927	34/310	0.936	27/371	0.926	36/505	0.943	42/333	0.955	46/610
Ots 104	0.933	30/307	0.913	28/416	0.940	31/512	0.953	29/425	0.957	32/636
Ots 107	0.925	27/311	0.909	25/415	0.903	32/519	0.935	36/405	0.917	33/661
Ssa 197	0.905	21/336	0.924	22/417	0.920	29/517	0.947	30/445	0.945	35/271
Ogo 2	0.732	12/346	0.733	10/417	0.747	11/508	0.750	14/422	0.796	17/688
Ogo 4	0.807	16/329	0.851	14/420	0.858	13/426	0.832	18/501	0.840	16/571
Oke 4	0.727	8/348	0.553	7/413	0.713	7/491	0.614	6/325	0.603	5/595
Oki 100	0.947	23/335	0.930	28/391	0.945	30/519	0.946	27/414	0.952	27/654
Omy 325	0.816	15/333	0.760	9/424	0.805	12/509	0.864	17/411	0.864	20/681

[a]see Beacham et al. 2006 for full genotyping protocol.

Figure S1a: MHC class I amino acid sequences for 37 alleles from Chinook salmon (*Oncorhynchus tshawytscha*). Alignment with a human leukocyte antigen A1 amino acid sequence is shown. The locations of amino acids associated with peptide binding region (Bjorkman et al. 1987) are indicated by the letter P.

```
HLA-A1   FFTSVSRPGR GEPRFIAVGY VDDTQFVRFD SDAASQRMEP RAPWIEQE?- ?RSIGTRRHG M*RPSHRLTE WTWGPC
PBR      P               P P P                                       P PP PPPPPPP PPPPPPPPPP  PPP
1        FYTASSEVPN FPEFVVVGMV DGVQMVHYDS NSQRAVPKQD WINKAAETLP QYWERETGNF KGDQQTFKAN IDIAKQ
2        .F........ .......... .....F.... .......... .M........ .....N...C .......... ...V..
3        ...T...... .......... .....F.... .......... .M........ .........C .......... ......
4        .......... .....I.... .......... .......... .......... ....S..... ..A....... ...V..
5        .......... .......... .......... .......... .......... ....S..... ..A....... ...V..
6        .......... ........T. .....F.... .......... .M........ .........ID ..A....... ...V..
7        .......... ........T. .....F.... .......... .M....D--. .....N...C ..T....... ...V..
9        .......... .......... .......... .......... .V....D--. .....N.... .......... ......
10       .F........ .......I.... .......... .......... .V....D--. .....N...G ..A....... ......
11       .......... ........T. .....F.... .......... .V........ .....N...C ..T....... ...V..
12       .F........ ........T. .....F.... .......... .......... .........IC ..T....... ...V..
14       .......... .......... .......... .......... .......... .....N.... ..A....... ......
15       .......... .....A... .......... .......... .M........ .........W ..T....... ...V..
16       .......... .....A... .......... .......... .V....D--. .....N.ELG ..A....... ......
17       .......... .......... .......... .......... .V....D--. .....N.I. ..A....... ......
18       .F........ .......... .....F.... .......... .M........ .........ID .......... ...V..
19       .......... .....I.... .......... .......... .V....D--. .....N...G ..A.R..... ...V..
20       ......G.T. .......... .......... ......S... .M..QTD--- ....QN..IL L.S.R....G V.TL..
23       .F........ .......... .....F.... .......... .......... .........IC ..A....... ......
24       .F........ .......... .......... .......... .V....D--. .....N.... .......... ......
25       .......... .......... .......... .......... .V....D--. .....N.... ..A....... ......
26       .......... .......... .....F.... .......... .......... .........IC .......... ......
28       .......... ........T. .....F.... .......... .......... .......... .......... ......
29       .......... .......... .......... .......... .V....D--. .....N.I. ..N....... ......
30       .F........ ........T. .....F.... .......... .......... .......... .......... ......
32       .......... ........T. .....F.... .......... .......... ....S...IC .......... ......
38       .......... .......... .......... .......... .M........ ....S...ID ..N....... ...V..
40       .......... .......... .......... .......... .......... .......... ..T....... ......
41       .F........ .......... .......... .......... .V....D--. .....N...C .......... ...V..
42       .......... .......... .......... .......... .......... ....S..... ..N....... ......
43       .......... .....I.... .....F.... .......... .V....D--. .....N.EIG ..A....... ......
45       .F........ .......... .....F.... .......... .......... .........IC ..A....... ..TP..
46       .F........ .......?T. .....F.... ..R....... .......... .........IC ..T....... ...P..
47       .......... ........T. .....F.... .......... .......... ....S...IC .......... ..TL..
48       ...T...... .......... .......... .......... .......... .....N.I. ..A....... ......
50       .H........ .S......T. .....F.... .......... .......... ....S...IC .......... ...L..
51       .......... .......... .......... .....A.... .M........ ....S...ID ..N....... ...V..
```

Figure S1b: MHC class II-B1 amino acid sequences for 17 alleles from Chinook salmon (*Oncorhynchus tshawytscha*). Alignment with a human leukocyte antigen B1 amino acid sequence is shown. The locations of amino acids associated with peptide binding region (Brown et al. 1993) are indicated by the letter P.

```
HLA-B1   RVRFLDRYFY NQEESVRFDS DVGEFRAVTE LGRPDAEYWN SQKDILEQAR AAVDTYCRHN YGVVESFTVQ RR
PBR          P P P        PP          P          P   PP     P   P PP    P    P   PP    PP PP
2        GIEFIDSYVF NKAEYIRFNS TVGRYVGYTE LGVKNAEAWN KGP-QLGQEQ AELERFCKPN AALHYRAILD KT
3        .....H.... ..V.H..... .......... ..L....... ...-...... .....V.... ...E...... ..
4        .......... ..V.H..... .......... ..L....... ...-...... .......... .......... ..
5        .......... .......... .......... ..L....... ...-...... G.......R. .D.D...... ..
6        .......... .RV.H..... .......... ..L....... ...-...... G....V..R. ..NA...... ..
7        .......... ..V.N..... .......... .......... ...-...... .......... .......... ..
8        .......... ..V.H..... .......... H......... ...-...... .......... .......... ..
9        .......... .......... .......... H......... ...-...... .......... .......... ..
12       ...L...... ..V.D..... ....F..... H.L....... SDAG-..... G....V..H. .D...S.... ..
13       .......... .....V.... .......... Y......... SDAGI..... .Q..SY..SH .DIL.S.... ..
19       .......... ..V.N..... .......... .......... ...-...... .......... .......... ..
20       .....H.... ..V.H..... .......... .......... ...-...... .......... .......... ..
21       .......... .RV.H..... .......... ..L....... ...-...... G.......H. .D........ ..
22       .......... .......... .......... Y......... ...-...... G.......H. .D.D...... ..
23       .......... ..V.H..... .......... ..L....... ...-...... .....V.... ...E...... ..
24       .......... ..V.H..... .......... H......... ...-...... .......... ...E...... ..
25       ...L...... S......... ....F..... H.L....... SDAGL..... G....V..H. .D..CS.... ..
```