Molecular Ecology (2001) **10**, 2525–2539

Comparative analysis of population structure across environments and geographical scales at major histocompatibility complex and microsatellite loci in Atlantic salmon (*Salmo salar*)

CHRISTIAN LANDRY* and LOUIS BERNATCHEZ
Department of Biology, Laval University, Ste-Foy, Québec, Canada, G1K 7P4

Abstract

Evidence of selection acting on major histocompatibility complex (MHC) genes has been illustrated with the analysis of their nucleotide sequences and allele frequency distribution. Comparing the patterns of population differentiation at neutral markers and MHC genes in the wild may provide further insights about the relative role of selection and neutrality in shaping their diversity. In this study, we combine both methods to assess the role of selection on a MHC gene in Atlantic salmon. We compare variation at a MHC class II B locus and microsatellites among 14 samples from seven different rivers and seven subpopulations within a single river system covering a variety of habitats and different geographical scales. We show that diversifying selection is acting on the sites involved in antigen presentation and that balancing selection maintains a high level of polymorphism within populations. Despite important differences in habitat type, the comparison of the population structure at MHC and microsatellites on large geographical scales reveals a correlation between patterns of differentiation, indicating that drift and migration have been more important than selection in shaping population differentiation at the MHC locus. In contrast, strong discrepancies between patterns of population differentiation at the two types of markers provides support for the role of selection in shaping population structure within rivers. Together, these results confirm that natural selection is influencing MHC gene diversity in wild Atlantic salmon although neutral forces may also be important in their evolution.

Keywords: Atlantic salmon, MHC, microsatellite, population, selection

Received 5 April 2001; revision received 11 July 2001; accepted 11 July 2001

Introduction

Genes of the major histocompatibility complex (MHC) encode molecules responsible for the recognition and presentation of foreign antigens in vertebrate genomes. Their primary role is to collect protein fragments in the cells and transport them to the membrane surface where the complex is recognized by T cells that can initiate an immune response (Klein 1986; reviewed by Ploegh & Watts 1998). The region of the molecule responsible for the peptide collection is the peptide-binding region (PBR), the polymorphism of which makes MHC the most variable coding genes of the mammalian, and probably of other vertebrate genomes (Klein 1986). MHC polymorphism is believed to be maintained by balancing selection, which includes overdominant, frequency dependent and diversifying selection across habitats. This has been supported by evidence for the trans-species persistence of allelic lineages (Klein 1987), a more even allele frequency distribution than expected under neutrality (Hedrick & Thomson 1983), and an excess of non-synonymous substitutions over synonymous substitutions in PBR codons (Hughes & Nei 1988, 1989). Because of their immune function, the most obvious agents of selection are pathogens and parasites, as indicated by the association

Correspondence: L. Bernatchez. Fax: 418 656–2043; E-mail: Louis.Bernatchez@bio.ulaval.ca
*Present address: Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA.

© 2001 Blackwell Science Ltd

between pathogen resistance and specific MHC haplotype variants (Hill *et al*. 1991; Paterson *et al*. 1998), and higher parasite resistance in heterozygous relative to homozygous individuals (Doherty & Zinkernagel 1975; Carrington *et al*. 1999). MHC genes are also involved in disassortative mating behaviour and selective abortion, which may contribute to maintaining high polymorphism in mammals (Hedrick 1992; Potts & Wakeland 1993; Edwards & Hedrick 1998). These properties make genes of the MHC among the best candidates for the study of molecular adaptation in vertebrates (Hedrick 1996).

In contrast to the above evidence for the adaptive value of MHC polymorphism, many natural populations exhibiting low levels of variation remain viable, thus raising the possibility that selection for maintaining polymorphism at these loci may be relatively weak (e.g. Slade 1992; Ellegren *et al*. 1993). Relaxed pathogen-driven selection has been proposed to explain the occurrence of low MHC polymorphism (see Slade 1992). Alternatively, Klein (1987) hypothesized that long periods of neutral evolution, during which random genetic drift may be more effective in determining allelic diversity, may alternate with intermittent events of positive selection occurring when populations face new environmental conditions, and thus new parasites. The analysis of genetic variation at MHC among populations that occupy different habitats, which would place MHC genes under differential selective pressure through time, would be particularly relevant to assess this hypothesis empirically (Stet & Egberts 1991).

The immune activity of fish MHC molecules has been investigated for many years and there is increasing evidence that their roles and overall genetic organization are homologous to their mammalian counterparts (reviewed by Stet & Egberts 1991; Ono *et al*. 1993). Extensive MHC gene polymorphism has also recently been characterized in many teleosts, including salmonids (Dixon *et al*. 1996; Miller & Withler 1996; Kim *et al*. 1999). More recently, MHC alleles have been shown to be of importance for individual fitness in farmed Atlantic salmon as some alleles are associated with resistance to furunculous caused by *Aeromonas salmonicida* (Langefors *et al*. 2001). Several salmonid species would be of particular interest in the study of MHC genes in the wild because they occur in anadromous and landlocked forms, which both occupy different feeding habitats, namely marine environments and freshwater lakes. Furthermore, salmonids tend to form genetically differentiated populations that are believed to be locally adapted to their habitat, which may be highly heterogeneous across the species range (Verspoor & Jordan 1989; Taylor 1991; Carvalho 1993). Disease resistance plays a role in salmonid population divergence as it has been shown to vary among populations and to be influenced by genetic factors (Taylor 1991; Bakke & Harris 1998; Van Muiswinkel *et al*. 1999). This aptitude for forming local populations may create genetic differentiation at a geographical scale as fine as several kilometres within a river drainage (Garant *et al*. 2000).

The main objective of this study was to investigate the relative role of selection vs. other evolutionary forces in determining MHC diversity in wild Atlantic salmon (*Salmo salar*). This was achieved by comparing the pattern of genetic differentiation at a MHC class II B gene with that of microsatellite loci among recently diverged populations occupying contrasting habitats (freshwater lake or river and sea) and at different geographical scales (among rivers within habitat types and among spawning areas within rivers). More specifically, we tested the null hypothesis that the patterns of genetic structure at MHC are mainly driven by the effect of neutral mutation, gene flow and random genetic drift. This hypothesis would be supported by the absence of statistical differences between the patterns of population structure at MHC and microsatellites. The alternative hypothesis of a role for selection in shaping MHC genetic structure would in turn be supported by contrasting patterns at both markers. The comparison of these patterns at different geographical scales may also identify the level of environmental heterogeneity at which the role of selection may be more prevalent. This would contribute to identify more specific ecological factors responsible for selection.

Materials and methods

Sampling profile

Atlantic salmon occurs as parapatric anadromous and landlocked forms in the Saguenay-Lac-St-Jean region, central Québec, Canada (Fig. 1a). Lake St-Jean (48°40′ N, 72°00′ W) is connected to the St-Lawrence river estuary by the Saguenay River, three tributaries of which support anadromous populations of Atlantic salmon. The lake harbours four populations of land-locked salmon for which migration to the ocean is replaced by a lacustrine feeding stage (Dahl 1928; Power 1958; Tessier *et al*. 1997). The Saguenay River was inundated by the large periglacial Laflamme Sea between 10 300 and 8700 years BP (Elson 1969) and was the most important colonization route of Lake St-Jean used by anadromous fish following the last glacial retreat (Bernatchez 1997; Tessier & Bernatchez 2000). The main tributary of the Saguenay river is the Ste-Marguerite river, which is subdivided into two main branches (North-East and Principal) and a third smaller tributary (North-West branch) in which the spawning areas have been well characterized (Fig. 1b) (Garant *et al*. 2000). The St-Jean and Petit-Saguenay rivers are the two other tributaries of the Saguenay River harbouring native Atlantic salmon populations.

Samples of the four land-locked populations, which are believed to be derived from the anadromous fish, were obtained in 1994 (see details in Tessier *et al*. 1997 and Tessier & Bernatchez 1999) and 1999 from adult fish caught at

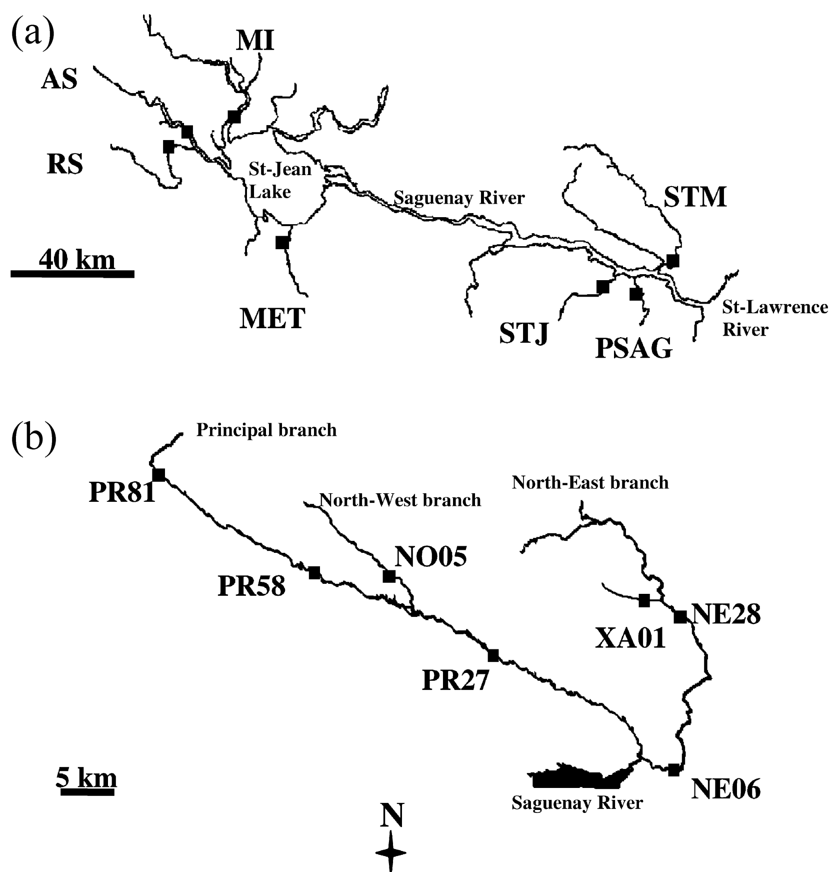

Fig. 1 Location maps of the Saguenay-Lac-St-Jean Region, Central Québec (48°40′ N, 72°00′ W). (a) Land-locked Atlantic salmon populations: Mistassini (MI), Ashuapmushuan (AS), Rivière aux Saumons (RS), Métabetchouane (MET); Anadromous populations: St-Jean (STJ), Petit-Saguenay (PSAG) and Ste-Marguerite River (STM). (b) Spawning areas within the Ste-Marguerite River: PR81, PR58, PR27, NO05, NE06, NE28, XA01.

counting fences within each of the four rivers connected to Lac St-Jean. Samples for the anadromous populations were obtained from sport fishing in the St-Jean and Petit-Saguenay rivers in 1994 (see Tessier & Bernatchez 2000 for details) and 1999. The third anadromous sample came from adult fish of the Ste-Marguerite River caught in 1995 at a counting fence of the North-East branch (see Garant *et al.* 2001). Sampling of fry for the comparison of the within-river population structure was conducted in 1996 on seven large river stretches corresponding to the major emergence or nursery habitats 1 or 2 weeks following emergence as detailed in Garant *et al.* (2000).

Molecular methods

DNA was extracted from adipose fin clips preserved in 95% ethanol as described in Bernatchez *et al.* (1992). A 254-bp fragment of the exon 2 of a MHC class II B gene was obtained using the primers CL007 (5′-GATCTGTATTATGTTTTCC-TTCCAG-3′) and AL1002 (5′-CACCTGTCTTGTCCAGT-ATG-3′) (Olsén *et al.* 1998). Alleles were first identified using radioactive single-strand conformation polymorphism (SSCP) (Hayashi 1991) as this method provides a rapid and sensitive screening for mutations and haplotype polymorphism (Orti *et al.* 1997). The primer AL1002 was end-labelled with [γ^{33}P] and the reaction was carried out in a total volume of 10 µL, containing 20–40 ng of genomic DNA, 0.25 mM dNTPs, 1 µL reaction buffer [10 mM Tris-HCl (pH 9.0), 1.5 mM $MgCl_2$, 0.1% Triton X-100, 50 mM KCl], 1 unit of *Taq* DNA polymerase, 1 pmol of [γ^{33}P]-AL1002, 4 pmol of AL1002 and 5 pmol of CL007. Five microlitres of a denaturing loading buffer were added to the polymerase chain reaction (PCR) products and 6 µL of the mix were loaded on a nondenaturing acrylamide gel (10% 49 : 1 acrylamide: bis-acrylamide, 5% glycerol and 0.5× TBE) for 15 h and 20 W migrations at 4 °C. The dried gels were exposed on an X-ray film for 24–48 h. Because both strands of the PCR product have a different electrophoretic mobility, 11 and 10 gels out of 52 were carried out by labelling, respectively, the other primer or both strands in order to avoid false variant identification or homoplasy. The following PCR profile produced one or two bands for each of the 666 samples analysed: 95 °C 3 min; 94 °C for 30 s, 57 °C for 30 s, 72 °C for 45 s, 32 cycles; 72 °C for 10 min. Allelic segregation was also tested in the progeny of two families. In all genotypes, one allele was from the female and the other was from the male (unpublished data), confirming that a single locus is being screened. This further supports the existence of a single MHC class II B locus in Atlantic salmon, as recently proposed (Langefors *et al.* 1998, 2000).

The reliability of the SSCP gel scoring was confirmed by sequencing analysis of identified SSCP haplotypes. Molecular cloning was performed using the Invitrogen Topo TA cloning kit (Invitrogen). Insertions were sequenced using the BigDye Terminator cycle sequencing kit on an ABI 377 automated sequencer (Applied Biosystems). Two to eight clones were sequenced for each individual representative of different allelic variants ($n = 90$) in order to avoid erroneous allelic identification due to reading errors or recombinant sequences generated by PCR (Ennis *et al*. 1990; Bradley & Hillis 1997).

Microsatellite analysis

Microsatellite analysis of the four land-locked and three anadromous samples of adult salmon had previously been published using seven loci MST-3, MST-79.1, MST-79.2, SFO-23; SSOSL85, Ssa171, Ssa197) (Tessier *et al*. 1997; Tessier & Bernatchez 2000). Within-river population structure was previously performed using five loci (SSOSL85, Ssa85, Ssa171, Ssa197 and Ssa202) (Garant *et al*. 2000). Descriptive statistics and quantification of population divergence are detailed in these studies.

Data analysis

MHC diversity and polymorphism. MHC gene sequences were aligned using the SEQUENCE NAVIGATOR 1.0.1 software (Perkin-Elmer). The relative rates of synonymous (*ds*) and nonsynonymous (*dn*) substitutions were determined according to Nei & Gojobori (1986) and corrected for multiple hits (Jukes & Cantor 1969) using MEGA 2.0 (Kumar *et al*. 2000). MEGA was also used to perform a *Z*-test of selection by comparing *dn* and *ds* as detailed in Nei & Kumar (2000) and a neighbour-joining tree for the MHC alleles (p-distance) based on amino acid sequences. Allele frequencies, allelic diversity and observed and expected heterozygosity were estimated using the GENETIX version 4.0 computer package (Belkhir *et al*. 1998). Comparisons of mean expected heterozygosity between markers and populations were performed using *t*-tests for independent samples. Standardized numbers of MHC class II B alleles (A_{std}) for a sample size of 50 were estimated using equation 11 in Ewens (1972) following the estimation of θ_k, which is an estimator of θ calculated from the infinite allele equilibrium relationship between the expected number of alleles (k), the sample size (n) and θ using the method implemented in ARLEQUIN, version 2.0 (Schneider *et al*. 2000).

Hardy–Weinberg equilibrium and Ewens–Watterson neutrality test. Departure from Hardy–Weinberg equilibrium was tested using the GENEPOP computer package (version 3.1d) (Raymond & Rousset 1995). This involved the use of the Markov chain method to obtain unbiased estimates of Fisher's exact test through 1000 iterations to test the alternative hypotheses of heterozygote excess or deficiency. The effect of balancing selection on sampled populations was tested using the Ewens–Watterson (Watterson 1978) neutrality test implemented in ARLEQUIN 2.0 (Schneider *et al*. 2000). The sum of allele frequencies (F) was compared to a null distribution of F generated by simulating random neutral samples (4000 replicates).

Population structure

Samples of adult fish caught in main rivers across habitats and fry samples from spawning sites within the Ste-Marguerite River were compared separately because they represent two different life stages and had not been genotyped using the same set of microsatellite loci.

Pairwise homogeneity tests of allele frequencies among populations were performed using Fisher's exact test implemented in GENEPOP (Raymond & Rousset 1995). Significance values were adjusted for multiple simultaneous tests using the sequential Bonferroni correction (Rice 1989). The extent of population differentiation was quantified by estimating the pairwise F_{ST} parameter from Weir & Cockerham (1984). For several reasons this was used instead of estimators that take into account the variance in repeat numbers of microsatellite (e.g. R_{ST}). First, these estimators have been shown to have large variance and that they are better estimators of gene flow only when a large number of loci is used (Gaggiotti *et al*. 1999). R_{ST} estimates have also been shown to be more sensitive to differences in sample sizes (Ruzzante 1998). Finally, Slatkin (1995) and Rousset (1996) showed that F_{ST} gave similar information for situations under which differentiation is approximately independent of the mutation process, i.e. large migration rates and/or recent time of divergence among populations, which is the case here.

An increasingly popular approach to assess the role of selection in determining allelic variation is to compare patterns of genetic structure at putative selected loci with those obtained from more neutral loci such as mitochondrial DNA (mtDNA) and microsatellite loci. Gene flow and drift should equally affect neutral loci whereas selection is more likely to be locus specific (Lewontin & Krakauer 1973). Therefore, discordance between potentially selected and neutral loci may be taken as evidence of selection (Spitze 1993; Lynch *et al*. 1999). We compared microsatellite F_{ST} values to those derived from MHC. Since loci are expected to provide nearly independent replicates of the genetic sampling process, we proceeded by resampling over the microsatellite loci (Weir 1990; p. 151). Ninety-five per cent confidence intervals around microsatellite F_{ST} estimates were estimated by bootstrapping (15 000 replicates) over loci using FSTAT, version 2.8 (Goudet 1998). MHC values outside the 95% confidence intervals were considered

significantly different from estimates derived from microsatellites (Weir 1996; p. 175). This was further assessed by performing a Mantel test opposing the matrix of pairwise estimates of genetic differentiation among samples based on MHC and microsatellite variation (Sokal & Rohlf 1995). Five thousand permutations were used in each case, using the method implemented in GENETIX 4.0 (Belkhir *et al*. 1998).

Results

MHC diversity

Forty nucleotide positions out of 254 were polymorphic, which resolved a total of 18 alleles (GenBank accession numbers: AF373692–AF373709) among the 666 salmon analysed (Fig. 2), the alleles having a mean pairwise number of nucleotide differences of 14.5. Comparisons with other salmonid sequences available in GenBank confirmed that the amplified sequences are part of the exon 2 of a MHC class II B gene and revealed that four alleles, Sasa-sm8, Sasa-ch2, Sasa-sm12 and Sasa-sm7, corresponded to previously identified sequences among European populations of Atlantic salmon; Sasa-c22 (Hordvik *et al*. 1993), Sasa-db06, Sasa-db08 and Sasa-db14 (Grimholt *et al*. 1993).

Alleles also differed in amino acid composition by one to 17 substitutions out of 84 possible sites (Fig. 3). The fact that neither indels nor stop codons were observed along with the high polymorphism suggests that the segment analysed was unlikely to be part of a pseudogene. The pattern of nucleotide substitution was also compatible with that expected under the influence of past or contemporary diversifying selection as the rate of nonsynonymous substitutions was significantly higher than that of the synonymous substitution

```
     1          11         21         31         41         51        60
me2  GATGGATATT TTGAACAGGT TGTGAGACAG TGCCGATACT CCTCAAAGGA CCTGCAGGGT
rs1  .......... ..T.T..... .......... .......... .......... ..........
sm11 .......... .......... ......T... .......T.. .......... ..........
me1  .......... ..T.T..TA. GA........ .......... .......... ..........
sm12 .......... ..TTT..... .......... .......... .......... ..........
sm10 .......... ..T.T..TA. GA........ .......... .......... ..........
sm9  .......... ..T.T..TA. GA........ .......... .......... ..........
sm8  .......... .......... .......... .......... .......... ..........
sm7  .......... ..T.T...AG G...TC.G.. .......... .......... ..........
sm6  .......... ..T.T..... .......... .......... .......... ..........
sm5  .......... ..TTT..... .......... .......... .......... ..........
sm4  .......... ..T.T...AG G...TC.G.. .......... .......... ..........
sm3  .......... .......... .......... .......... .......... ..........
sm2  .......... ..T.T..TA. GA........ .......... .......... ..........
sm1  .......... .......... ......T... .......T.. .......... ..........
mi1  .......... ..T.T..... .......... .......... .......... ..........
ch2  .......... ..TTT..... .......... .......... .......... ..........
ch1  .......... ..T.T..... .......... .......... .......... ..........

     61         71         81         91         101        111       120
me2  ATAGAGTTTA TAGACTCTTA TGTTTTCAAT AAGGCTGAAT ATATCAGATT CAACAGCACT
rs1  .......... .......... .......... C......... .......... ..........
sm11 .......... ..C....... .......... .......... .......... ..........
me1  ........G. .TAC...G.. .......... C........A .......... ..........
sm12 .......... ..C....... .......... C........A .......... ..........
sm10 ........G. .TAC...G.. .......... C........A .......... ..........
sm9  ........G. .TAC...G.. .......... C......... .......... ..........
sm8  .......... .......... .......... .......... ..G....... ..........
sm7  .......... .......... .......... .......... ..G....... ..........
sm6  .......... .......... .......... C......... .......... ..........
sm5  .......... .......... .......... C........A .......... ..........
sm4  .......... .......... .......... C........A .......... ..........
sm3  .......... .......... .......... C......... .......... ..........
sm2  ........G. .TAC...G.. .......... C........A .......... ..........
sm1  .......... ..C....... .......... .......... .......... ..........
mi1  .......... .......... .......... .......... ..G....... ..........
ch2  .......... ..C....... .......... C........A .......... ..........
ch1  .......... .......... .......... C......... .......... ..........

     121        131        141        151        161        171       180
me2  GTGGGGAAGT ATGTTGGATA CACTGAGTAT GGAGTGAAGA ATGCAGAAGC CTGGAACAAA
rs1  .......... T......... .......C.. .......... .......... ..........
sm11 .......... .......... .......CTG .......... .......... ..........
me1  .......... T......... .......C.. .......... .......... ..........
sm12 .......... .......... .......CTG .......... .......... ..........
sm10 .......... T......... .......C.. .......... .......... ..........
sm9  .......... .......... .......... .......... .......... ..........
sm8  .......... .......... .......CTG .......... .......... ..........
sm7  .......... .......... .......... .......... .......... ..........
sm6  .......... T......... .......C.. .......... .......... ..........
sm5  .......... T......... .......CTG .......... .......... ..........
```

Fig. 2 Exon 2 nucleotide sequences of the MHC class II B alleles of Atlantic salmon. Dots indicate identity with the first sequence. GenBank accession numbers: AF373692–AF373709.

```
sm4   .......... T......... .......CTG .......... .......... ..........
sm3   .......... T......... .......CTG .......... .......... ..........
sm2   .......... T......... .......C.. .......... .......... ..........
sm1   .......... .......... .......CTG .......... .......... T.........
mi1   .......... .......... .......... .......... .......... ..........
ch2   .......... .......... .......CTG .......... .......... ..........
ch1   .......... T......... .......C.. .......... .......... ..........

      181        191        201        211        221        231      240
me2   GGTCCTGAGC TGGCTAGAGC GCTAGGGGAG CTGGAGCGTT ACTGTAAGTT TAACGCTCCT
rs1   .......... .....GT..A .......... .......... T.......CA ..........
sm11  .......... .....G...A .......... .......... T.......CA .......GA.
me1   .......... .....G...A .......... .......... T.......CA .......GA.
sm12  .......... .....GT..A .......... .......... ........CA .......GA.
sm10  .......... .....G...A .......... .......... T.......CA ..........
sm9   .......... .....G...A ........T. .......... ........CA ..........
sm8   .......... .....GT..A .......... .......... T.......CA .......GA.
sm7   .......... .....G...A .......... .........G T.......CA ..........
sm6   .......... .....G...A .......... .......... ........C. ..........
sm5   .......... .......... .......... .......... ........C. ..........
sm4   .......... .....G...A .......... .......... T.......CA .......GA.
sm3   .......... .....GT..A .......... .......... ........C. ..........
sm2   .......... .....G...A .......... .........G T.......CA .......GA.
sm1   .......... .....G...A .......... .......... T.......CA ..........
mi1   .......... .....G...A .......... .......... ........CA ..........
ch2   .......... .....G...A .......... .......... T.......CA .......GA.
ch1   .......... .....GT..A .......... .......... ........CA .......GA.

      241           254
me2   ATCTACTACA GCGC
rs1   ...G...... ....
sm11  C..C...... ..A.
me1   C..C...... ..A.
sm12  ...G...... ....
sm10  ...G...... ....
sm9   .......... ....
sm8   C..C.....C ....
sm7   ...G...... ....
sm6   ...G...... ....
sm5   .......... ....
sm4   C..C...... ..A.
sm3   ...G...... ....
sm2   C..C...... ..A.
sm1   ...G...... ....
mi1   .......... ....
ch2   C..C...... ..A.
ch1   C..C...... ..A.
```

Fig. 2 *Continued*

in the putative PBR region (Table 1) (Hughes & Nei 1988, 1989). This trend, although weaker, was also observed for the sites outside the putative PBR. This could be because the model commonly used to identify the PBR is derived from human molecules and that other sites may be involved in peptide presentation in salmonids (Brown *et al*. 1993; Dixon *et al*. 1996).

Within-population diversity

Mean standardized numbers of MHC alleles for a sample size of 50 were 6.02 and 10.61 in land-locked and anadromous samples, respectively (Table 2). The overall pattern of within-population diversity observed at the MHC, as revealed by the heterozygosities, was very similar to that observed at microsatellite loci. In land-locked samples, the mean expected heterozygosity (H_E) for the MHC was 0.68 and did not differ significantly ($P = 0.52$) from the mean value of 0.62 observed at microsatellites. Similarly, the mean H_E value of 0.82 for MHC among the anadromous samples did not differ ($P = 0.34$) from that observed at microsatellites (0.77). As also reported for microsatellites, the mean expected heterozygosity was significantly lower in land-locked than in anadromous populations for MHC ($P < 0.001$). Absolute values of population differentiation parameters (e.g. F_{ST}) strongly depend upon within population polymorphism of the markers used (Charlesworth 1998; Hedrick 1999). Consequently, the similar genetic diversity at MHC and microsatellite loci indicated that the comparisons of inter-population divergence were not biased by this potential problem in this study.

Hardy–Weinberg equilibrium and Ewens–Watterson test

The null hypothesis of Hardy–Weinberg equilibrium was not rejected at MHC for any sample following corrections for multiple tests, except for the Mistassini River (MI), NO05 and NE06 spawning sites. In these latter samples, a significant excess of heterozygotes was observed. These samples were in Hardy–Weinberg equilibrium at microsatellite loci, which allowed us to reject limited number of families as a potential explanation for heterozygote excess

```
     1          11         21         31         41         51      60
           * *               * *       **          *          *   **
me2  DGYFEQVVRQ CRYSSKDLQG IEFIDSYVFN KAEYIRFNST VGKYVGYTEY GVKNAEAWNK
rs1  ....Y..... .......... .......... Q......... ...F.....H ..........
sm11 ........S. ..F....... ....H..... .......... .........L ..........
me1  ....YHMM.. .......... ..L.T..... Q..N...... ...F.....H ..........
sm12 ....F..... .......... ....H..... Q..N...... .........L ..........
sm10 ....YHMM.. .......... ..L.T..... Q..N...... ...F.....H ..........
sm9  ....YHMM.. .......... ..L.T..... Q......... ......... ..........
sm8  .......... .......... .......... ....V..... .........L ..........
sm7  ....Y.R.SE .......... .......... ....V..... ......... ..........
sm6  ....Y..... .......... .......... Q......... ...F.....H ..........
sm5  ....F..... .......... .......... Q..N...... ...F.....L ..........
sm4  ....Y.R.SE .......... .......... Q..N...... ...F.....L ..........
sm3  .......... .......... .......... Q......... ...F.....L ..........
sm2  ....YHMM.. .......... ..L.T..... Q..N...... ...F.....H ..........
sm1  ........S. ..F....... ....H..... .......... .........L ..........
mi1  ....Y..... .......... .......... ....V..... ......... ..........
ch2  ....F..... .......... ....H..... Q..N...... .........L ..........
ch1  ....Y..... .......... .......... Q......... ...F.....H ..........

     61         71         81
      * * **  *    *  *    ** *
me2  GPELARALGE LERYCKFNAP IYYS
rs1  .....VE... ...F..H... .D..
sm11 .....GE... ...F..H..D LH..
me1  .....GE... ...F..H..D LH..
sm12 .....VE... ......H..D .D..
sm10 .....GE... ...F..H... .D..
sm9  .....GE..V ......H... ....
sm8  .....VE... ...F..H..D LH.R
sm7  .....GE... ...V..H... .D..
sm6  .....GE... ......L... .D..
sm5  .......... ......L... ....
sm4  .....GE... ...F..H..D LH..
sm3  .....VE... ......L... .D..
sm2  .....GE... ...V..H..D LH..
sm1  .....GE... ...F..H... .D..
mi1  .....GE... ......H... ....
ch2  .....GE... ...F..H..D LH..
ch1  .....VE... ......H..D LH..
```

Fig. 3 Exon 2 amino-acid sequences of the MHC class II B alleles of Atlantic salmon. Dots indicate identity with the first sequence. The putative peptide binding region is indicated with asterisks (Brown *et al*. 1993; Dixon *et al*. 1996).

Table 1 Mean pairwise numbers of nonsynonymous substitutions per nonsynonymous sites (*dn*) and synonymous substitutions per synonymous sites (*ds*) observed among alleles at the peptide binding region (PBR) and non-PBR of exon 2 of the Sasa class II B gene. Standard errors are given in parenthesis. *P*-values are the probability of rejecting the null hypothesis of neutrality ($dn = ds$) (Nei and Kumar 2000)

Region	Codons	*ds*	*dn*	*dn*/*ds*	*P*
Non-PBR	20	0.024 (0.015)	0.038 (0.014)	1.58	0.45
PBR	64	0.029 (0.021)	0.179 (0.043)	6.17	0.001
Total	84	0.025 (0.012)	0.068 (0.016)	2.72	0.01

(Tessier *et al*. 1997; Tessier and Bernatchez 2000; Garant *et al*. 2000).

In all samples but the Ashuapmushuan (AS), the observed homozygosity was less than expected assuming neutrality for a given number of alleles and a given sample size (Ewens–Watterson test), reflecting a more even allele frequency distribution than for neutral markers with the same polymorphism (Table 3). This pattern was statistically significant ($P < 0.01$) for two samples (PSAG and PR27) and was marginally significant ($P < 0.1$) for four others (NO05, NE06, XA01 and MI). Combining probabilities of independent tests to generate an overall test for significance using Fisher's method (Sokal & Rohlf 1995) revealed that balancing selection might be acting to maintain a higher genetic variability than under neutrality in these samples ($P < 0.001$).

Comparisons between habitats and among rivers

Population structure at MHC. A strong pattern of genetic differentiation among land-locked and anadromous samples was first reflected by the heterogeneity in allele frequencies distribution (Table 4). Nine and two alleles were exclusive to anadromous and land-locked populations, respectively. The allele Sasa-me2 was present at a relatively high and similar frequency among all land-locked populations (mean: 33.96 range: 29.4–41.7%), whereas it was found at low and variable frequencies among anadromous samples (mean: 4.06 range: 0–15.1%). In contrast, the alleles Sasa-sm6, Sasa-sm7 and Sasa-sm11 were ubiquitous and relatively evenly distributed among anadromous samples but absent or at low frequencies among land-locked populations. However, based on amino acid sequences, no llelic lineages were exclusive to a particular form (land-locked or anadromous)

Table 2 Samples sizes (n), number of alleles (A), number of alleles at MHC standardized for a sample size of 50 (A_{std}) (Ewens 1972), expected (H_E) and observed (H_O) heterozygosity at microsatellite and MHC loci

	Land-locked				Anadromous			Spawning sites						
Locus	MI	AS	RS	MET	STJ	PSAG	STM	PR81	PR58	PR27	NO05	NE06	NE28	XA01
MHC														
n	63	35	46	60	35	46	76	43	43	42	45	46	45	41
H_E	0.76	0.58	0.70	0.67	0.82	0.86	0.85	0.81	0.79	0.88	0.78	0.81	0.79	0.81
H_O	0.92*	0.57	0.61	0.65	0.83	0.78	0.88	0.81	0.81	0.83	1.00*	1.00*	0.89	0.71
A	7	6	6	5	11	10	16	11	10	11	5	9	11	8
A_{std}	6.67	6.48	6.11	4.84	12.18	10.22	14.25	11.46	10.40	11.54	7.17	9.19	11.32	8.39
MST79.1														
n	36	39	35	41	40	40	38	50	46	50	48	49	49	48
H_E	0.60	0.46	0.65	0.29	0.56	0.47	0.46	0.87	0.83	0.84	0.83	0.89	0.90	0.89
H_O	0.61	0.41	0.57	0.32	0.58	0.45	0.42	0.88	0.83	0.84	0.81	0.86	0.88	0.92
A	4	5	4	4	6	5	4	14	16	14	15	18	17	18
Ssa-171														
n	36	41	37	40	40	40	40	50	46	50	48	49	49	48
H_E	0.88	0.90	0.80	0.78	0.85	0.88	0.91	0.83	0.81	0.84	0.84	0.82	0.87	0.74
H_O	0.92	0.95	0.89	0.85	0.83	0.88	0.98	0.86	0.91	0.84	0.79	0.76	0.84	0.75
A	13	14	10	9	15	17	20	11	11	12	11	12	12	12
Ssa-197														
n	36	41	37	42	40	40	40	50	46	50	48	49	49	48
H_E	0.67	0.59	0.72	0.57	0.90	0.87	0.87	0.85	0.87	0.85	0.84	0.84	0.87	0.83
H_O	0.64	0.61	0.70	0.43	0.88	0.88	0.93	0.88	0.91	0.96	0.96	0.92	0.92	0.94
A	5	7	6	5	15	13	13	11	11	11	13	14	13	12
MST-3														
n	36	40	36	41	40	40	40							
H_E	0.57	0.49	0.67	0.73	0.74	0.76	0.72							
H_O	0.42	0.45	0.67	0.63	0.93	0.65	0.78							
A	5	3	4	4	6	6	6							
MST-79.2														
n	36	39	36	39	40	40	40							
H_E	0.46	0.46	0.50	0.44	0.14	0.22	0.24							
H_O	0.44	0.36	0.58	0.54	0.15	0.20	0.28							
A	2	2	2	2	2	2	2							
SFO-23														
n	36	39	30	37	40	40	40							
H_E	0.75	0.73	0.66	0.05	0.76	0.83	0.81							
H_O	0.70	0.77	0.53	0.03	0.78	0.68	0.85							
A	9	9	8	3	11	12	11							
SSOSL85														
n	35	41	35	40	40	40	40							
H_E	0.77	0.80	0.70	0.58	0.87	0.82	0.88							
H_O	0.80	0.73	0.77	0.60	0.90	0.78	0.90							
A	8	8	9	5	13	10	13							
Ssa-85														
n								50	46	50	48	49	49	48
H_E								0.71	0.77	0.66	0.67	0.74	0.76	0.67
H_O								0.66	0.74	0.66	0.73	0.80	0.74	0.67
A								10	8	7	6	8	9	8
Ssa-202														
n								50	46	50	48	49	49	48
H_E								0.81	0.79	0.86	0.74	0.80	0.80	0.85
H_O								0.84	0.89	0.92	0.73	0.92	0.84	0.94
A								10	10	12	6	9	9	10

Population names and locations are described in Fig. 1. *Denote samples with significant excess in heterozygotes.

Table 3 Ewens–Watterson test of neutrality based on MHC class II B allele frequency distribution within populations summarized by the observed homozygosity (F) and the expected F (Σp_i^2) for the same sample sizes and numbers of alleles under neutrality

	Land-locked				Anadromous			Spawning sites						
	MI	AS	RS	MET	STJ	PSAG	STM	PR81	PR58	PR27	NO05	NE06	NE28	XA01
Observation F	0.238	0.422	0.302	0.326	0.178	0.142	0.151	0.193	0.225	0.131	0.224	0.191	0.214	0.196
Expected F	0.392	0.403	0.421	0.496	0.223	0.267	0.185	0.236	0.288	0.256	0.369	0.296	0.242	0.322
P value	< 0.1	NS	NS	NS	NS	< 0.01	NS	NS	NS	< 0.01	< 0.1	< 0.1	NS	< 0.1

Table 4 Allele frequencies distribution of 18 identified Sasa MHC class II B alleles among samples of anadromous and land-locked samples of Atlantic salmon. GenBank accession numbers: AF373692–AF373709

	Land-locked				Anadromous			Spawning sites						
	MI	AS	RS	MET	STJ	PSAG	STM	PR81	PR58	PR27	NO05	NE06	NE28	XA01
Sasa-														
me2	0.294	0.300	0.348	0.417	0.057	—	0.033	0.151	0.023	0.119	—	—	0.022	—
rs1	0.032	0.057	0.391	0.292	—	—	0.013	0.116	—	0.119	—	—	—	—
sm11	0.167	0.029	0.152	—	0.357	0.196	0.237	0.326	0.302	0.191	0.267	0.283	0.222	0.110
me1	—	—	—	0.258	0.029	0.011	0.040	0.012	0.047	0.143	—	0.022	0.022	0.085
sm12	—	—	—	—	0.100	0.109	0.026	—	—	0.036	—	—	—	—
sm10	—	—	—	—	—	—	0.013	0.012	—	—	0.022	—	0.022	—
sm9	—	—	—	—	0.057	0.065	0.020	0.093	0.151	0.107	—	0.141	0.044	0.024
sm8	0.008	—	—	0.025	0.071	0.076	0.040	—	0.012	—	—	—	—	—
sm7	—	—	—	—	0.100	0.228	0.250	0.047	0.291	0.107	0.311	0.217	0.367	0.342
sm6	—	—	0.033	—	0.114	0.130	0.138	0.198	0.070	—	0.133	0.185	0.033	0.134
sm5	—	—	—	—	0.043	0.033	0.053	0.023	0.035	0.036	0.078	0.011	0.111	0.134
sm4	—	—	—	—	0.057	0.076	0.033	—	0.012	0.060	0.011	0.076	0.022	0.146
sm3	—	—	—	—	0.014	0.076	0.046	—	0.058	—	0.178	0.054	0.111	0.024
sm2	—	—	—	—	—	—	0.013	0.012	—	0.071	—	—	0.022	—
sm1	—	—	—	—	—	—	0.033	0.012	—	0.012	—	0.011	—	—
mi1	0.214	0.014	—	0.008	—	—	—	—	—	—	—	—	—	—
ch2	0.278	0.571	0.054	—	—	—	—	—	—	—	—	—	—	—
ch1	0.008	0.029	0.022	—	—	—	0.013	—	—	—	—	—	—	—

(Fig. 4). The heterogeneity in allele distribution was confirmed by the test of population differentiation. Indeed, of the 21 pairwise homogeneity tests of allele frequency distribution, all but three (comparisons among the anadromous PSAG, STJ and STM samples) were significant following sequential corrections for multiple tests. Therefore, this allowed us to reject the allele frequency homogeneity for 17 of the comparisons (initial $\alpha = 0.0024$).

MHC vs. microsatellite population differentiation. The patterns of population differentiation at MHC across and within habitats were overall similar to that observed at microsatellite loci. Namely, land-locked populations were much more differentiated from each other than were the anadromous samples. There was also a significant correlation (Mantel test: $Z = 0.96$, $P = 0.03$) between the matrices of pairwise F_{ST} observed for the two types of markers (Fig. 5).

In general, the F_{ST} values we estimated were in the same range as those observed by Langefors *et al*. (1998) between salmon populations of the Baltic Sea (0.01–0.235). The overall F_{ST} value observed at the MHC locus (0.157) among populations and habitats was not significantly different from the overall F_{ST} estimate at microsatellite loci (0.124 C.I. 95%: 0.079–0.179). There was, however, a significant trend for the pairwise MHC estimates to be higher than those derived from microsatellites. All values but three were above the 1 : 1 relationship (Fig. 5), whereas a random distribution should result in an even number of points above and below 1 : 1 (t-test on proportion, $P = 0.008$). Large variance among loci in the estimation of the global F_{ST} may partly explain the inconsistency between these results. All comparisons across habitats were above the 1 : 1 relationship. Three of them (AS-PSAG, AS-STJ, AS-STM) and two comparisons within each habitat showed a significantly higher F_{ST} estimate at the MHC locus than at microsatellites. Because the Rivière-aux-Saumons River (RS) is a tributary of the Ashuapmushuan River (AS) and thus these could be linked by a high level of gene flow, the

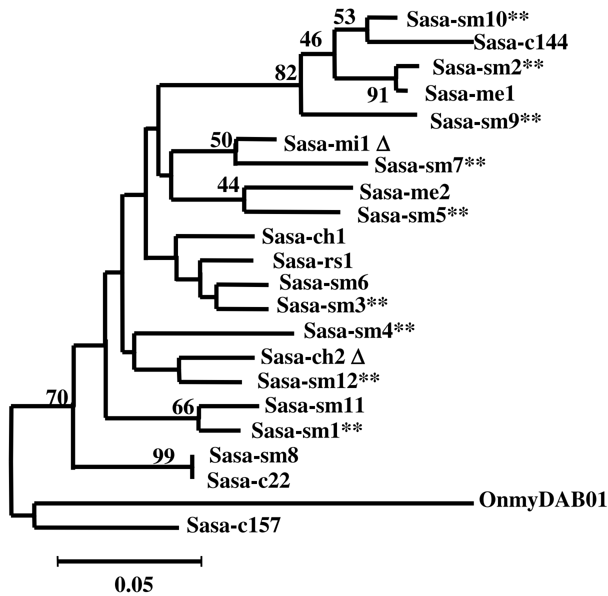

Fig. 4 Neighbour-joining tree based on the class II B exon 2 amino acid sequences (p-distance). Numbers indicate bootstrap percentages (2000 iterations). The tree also includes sequences from European Atlantic salmon (Sasa-c144, Sasa-c22, Sasa-c157; Hordvik *et al*. 1993) and Rainbow trout (OnmyDAB01; Glaman 1995). Double asterisks indicate allele exclusively found in the anadromous form and the triangles indicate alleles found in the land-locked form only.

much higher MHC F_{ST} estimate (0.225) observed between samples AS and RS compared to the low estimate based on microsatellites (F_{ST}: 0.068 CI 95%: 0.058–0.084) was of particular interest.

Comparisons within rivers

Population structure at MHC. Exact tests of population differentiation revealed significant differences in allele frequencies in all but three pairwise comparisons (NE28-NO05, NE28-PR58, and PR58-NE06) following corrections for multiple tests (initial α = 0.0024).

MHC vs. microsatellite population differentiation. A more pronounced discrepancy in the extent of genetic differentiation at MHC and microsatellite loci was observed within rivers relative to comparisons between habitats and among rivers. First, there was no correlation between the matrices of pairwise F_{ST} estimates at the two types of markers (Mantel test: Z = 0.05, $P = 0.29$, Fig. 6). Thirteen out of 21 pairwise F_{ST} estimates were significantly higher for MHC than microsatellites, whereas only four were significantly lower. Furthermore, the overall extent of population differentiation at the MHC locus ($F_{ST} = 0.047$) was significantly higher than that observed for microsatellite loci (0.028 CI 95%: 0.019–0.035).

Discussion

The role of selection on MHC variation has generally been investigated from patterns of nucleotide substitution and allelic diversity within populations (Hedrick & Thomson 1983; Hughes & Nei 1988, 1989; Paterson 1998). Comparing patterns of genetic polymorphism among populations at different types of loci for investigating the selective effects on other genes has also proven to be an efficient and illustrative method (Lewontin & Krakauer 1973; Karl & Avise 1992; Spitze 1993; Pogson *et al*. 1995; Lynch *et al*. 1999). Although this latter method has been criticized (Robertson 1975; Beaumont & Nichols 1996), it nevertheless provides a baseline to discuss the relative contribution of gene flow and local adaptation to overall differentiation (Lemaire *et al*. 2000). In this study, we combined both approaches to investigate the relative role of selection vs. other evolutionary forces (neutral mutation, gene flow and genetic drift) in determining the patterns of MHC allelic diversity in natural populations (see also Boyce *et al*. 1997; Madsen *et al*. 2000). This was achieved by comparing the patterns of genetic

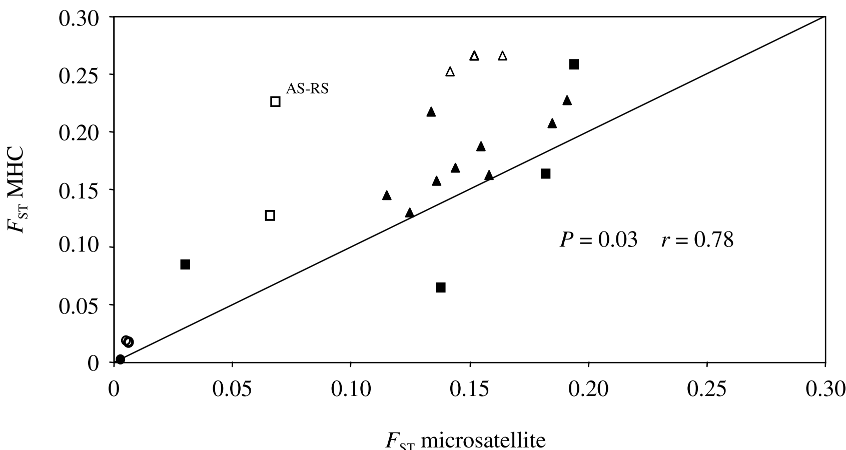

Fig. 5 Correlation between the extent of genetic differentiation (F_{ST}) at microsatellite and MHC loci in anadromous and land-locked samples of Atlantic salmon: (closed triangle) between habitats, anadromous vs. land-locked populations; (closed square) within habitat, land-locked populations; (closed circle) within habitat, anadromous populations. Open symbols represent MHC F_{ST} values outside the 95% confidence intervals of the microsatellite F_{ST} estimates. The P-value represents the outcome of the Mantel test. The value corresponding to the differentiation between the sample AS and RS is highlighted.

© 2001 Blackwell Science Ltd, *Molecular Ecology*, 10, 2525–2539

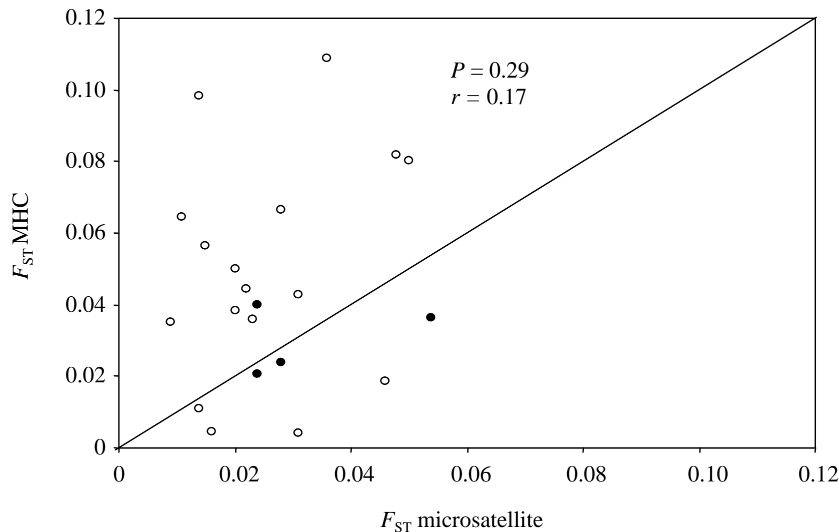

Fig. 6 Correlation between the extent of genetic differentiation (F_{ST}) at microsatellite and MHC loci among salmon samples from different spawning areas within the Ste. Marguerite River. Empty circles represent MHC F_{ST} values outside the 95% confidence intervals of the microsatellite F_{ST} estimates. The *P*-value represents the outcome of the Mantel test.

differentiation at a MHC class II B gene with that of microsatellite loci among recently diverged Atlantic salmon populations occupying contrasting habitats with different life histories and at different geographical scales. More specifically, we tested the null hypothesis that genetic diversity at MHC could be accounted for by neutral mutations, gene flow and random genetic drift.

Globally, the null hypothesis of neutrality of the MHC gene was rejected, as we found an excess of nonsynonymous substitutions at the sites involved in the antigen presentation. Furthermore, there were significantly different patterns of population differentiation between microsatellites and MHC in several comparisons. In other comparisons, however, patterns of differentiation at both markers were correlated. These results indicated altogether that migration and genetic drift have had a prevalent role in shaping the patterns of MHC genetic diversity observed among recently diverged salmon populations but they were not sufficient to account for the overall patterns observed. We propose that this implies a role for selection in shaping MHC genetic diversity in natural populations of Atlantic salmon.

Selection vs. nucleotide substitution and allelic diversity

The observed pattern of nucleotide substitution and allelic diversity indicated that selection plays a role in maintaining polymorphism within and/or among populations, although its intensity appeared to be weak. A prerequisite for inferring the role natural selection has on MHC is to confirm that the allelic nucleotide composition is compatible with the effect of selection (Hughes & Nei 1988; Paterson 1998). This was supported in our study by a high rate of nonsynonymous substitutions relative to synonymous substitutions in the PBR, suggesting selection for increased diversity at sites involved in antigen presentation. All alleles differed in amino acid composition, which, along with the observed heterogeneity of allelic distribution among samples, potentially implies a functional or phenotypic pattern of population differentiation. Furthermore, several alleles were identical to those reported among European populations, despite evidence for highly restricted if not complete absence of gene flow between both continents for several thousands of generations (Bermingham *et al*. 1991; McConnell *et al*. 1995). This may be indicative of a balancing selection effect, as generally inferred for explaining transpecific allelic identity of MHC genes (Klein 1987).

Balancing selection at MHC genes may be the outcome of heterozygote superiority and rare allele advantage against pathogens. A strong selective advantage of heterozygous over homozygous individuals should also result in an excess of heterozygotes (Doherty & Zinkernagel 1975; Hedrick & Thomson 1983; Hughes & Nei 1988). This was only partially supported since the null hypothesis of Hardy–Weinberg equilibrium was rejected for only three out of 14 samples. This could hypothetically be related to pathogen resistance since weaker disease resistance for homozygotes relative to heterozygotes has been inferred to explain differential patterns of survival to pathogens in salmon from Sweden (Langefors *et al*. 1998). Other mechanisms, such as dis-assortative mating at MHC, may also create excess in heterozygotes (Potts *et al*. 1991), and such a mechanism has recently been documented by Landry *et al*. (2001). Rare allele advantage has also recently been proposed to be an important factor in the maintenance of MHC variability in Atlantic salmon facing specific parasites (Langefors *et al*. 2001). These three mechanisms, heterozygote superiority, dissasortative mating and frequency-dependent selection, if maintained over generations, may act to preserve a high and evenly distributed allelic diversity within populations, resulting in lower homozygosity than expected under neutrality (Watterson 1978). This was supported in this study by the overall statistically significant trend for reduced

homozygosity within samples. By reducing the effect of bottlenecks and genetic drift, balancing selection is expected to undermine the pace of population differentiation (Takahata & Nei 1990), even under restricted migration (Schierup *et al*. 2000). However, the fact that the allelic diversity and heterozygosity we observed at MHC and microsatellite loci were almost identical suggests that the overall effect of balancing selection on the extent of population differentiation observed at MHC may have been limited.

Patterns of MHC and microsatellite diversity across habitats

After having left their natal river, anadromous salmon spend from 1 to 3 years in the ocean before returning to reproduce. In the land-locked form, this ocean phase is replaced by a lacustrine one (Tessier *et al*. 1997). Freshwater parasite communities affecting salmon differ from marine ones (Bouillon & Dempson 1989; Bakke & Harris 1998), which potentially represents a selective force in differentially shaping MHC variation across habitats. In such a case, one would predict that the extent of genetic differentiation between anadromous and land-locked populations should be more pronounced at MHC than at microsatellite loci.

Our results were supportive of a predominant effect of genetic drift and migration on population differentiation although they also indicated that selection has played a role in shaping the pattern of MHC variation across habitats. A correlation between the patterns of genetic differentiation at MHC and microsatellites was revealed and the global F_{ST} estimated from MHC was not higher than that estimated from microsatellites. Comparable results have been found in Bighorn sheep, were MHC genes and microsatellite gave similar F_{ST} estimates within and across regions (Boyce *et al*. 1997). Conversely, there was an overall statistically significant trend for pairwise F_{ST} estimates derived from MHC to be higher than those obtained with microsatellites. As expected, the discrepancy between MHC and microsatellites was particularly pronounced in comparisons involving land-locked and anadromous samples, all F_{ST} values estimated from MHC were higher than those derived from microsatellite loci. As stated earlier, it is unlikely that this was due to differential allelic diversity or heterozygosity within populations since these parameters were identical for both markers.

Differential selection across habitats should also result in habitat-specific allelic composition at MHC. This was suggested by the differential distribution of individual alleles between anadromous or land-locked populations, which appeared incompatible with the sole influence of random drift and migration, and current knowledge on the origin of land-locked populations. Namely, allele Sasa-me2 occurred in relatively high and even frequencies in all four land-locked populations, as was Sasa-rs1 in the Métabetchouane and Rivière aux Saumons, despite evidence for highly restricted gene flow among land-locked populations (*Nm* ranging from 0.539 to 5.848) and their small effective population sizes (Tessier *et al*. 1997). Furthermore, these abundant alleles are almost absent in anadromous populations from the Saguenay River, that must represent the ancestral populations from which land-locked populations originated less than 2000 generations ago (Tessier & Bernatchez 2000). Admittedly, we cannot refute the possibility that the contemporary allelic composition among anadromous populations from the Saguenay River sharply differs from that of ancestral populations. However, microsatellites revealed an almost complete absence of genetic differentiation among anadromous salmon from different rivers. This suggests a very limited effect of genetic drift in this anadromous system.

Patterns of MHC and microsatellite diversity within rivers

The most salient result of this study was the more pronounced discrepancy between MHC and microsatellites in comparisons involving salmon from different spawning areas within a river relative to those quantified between habitats or among rivers within a habitat. At microsatellites, allelic distributions were more homogeneous among sites, which led to a significantly reduced population differentiation relative to MHC. Furthermore, there was no correlation between pairwise F_{ST} estimates of both markers. These results therefore suggest that, in contrast to comparisons among rivers and habitats, selective factors may predominate over drift and migration in shaping the population structure at this much smaller geographical scale. This may be not specific to the anadromous form, since a much more pronounced level of differentiation at MHC was also observed between the land-locked Ashuapmushuan and Rivière-aux-Saumons populations within the same river system.

The more pronounced differentiation at MHC may be the result of local adaptation at this locus, as the population structure observed at microsatellite loci is temporally stable (Garant *et al*. 2000), which would allow natural selection to modulate the allele frequency among sites. Furthermore, it is known that parasites and pathogens can have a very local distribution and spread slowly, such that spatially isolated subpopulations (e.g. parr in a riffle) may be heavily infected while neighbouring populations remain uninfected (Bakke & Harris 1998). The observed pattern could reflect the differential effect of natural selection among sites in the weeks (within cohort selective mortality) between hatching and sampling. However, it is still unknown whether the early life history stages of salmon are able to mount specific immune responses (Bakke & Harris 1998). Also, fish were sampled within 2 weeks following emergence, providing little time for selection (unless very pronounced) to eliminate less fit genotypes.

© 2001 Blackwell Science Ltd, *Molecular Ecology*, 10, 2525–2539

Such a fine scale of differentiation at MHC raises the issues of its involvement in homing and kin recognition. Indeed, the freshwater phase of homing in salmonids is hypothesized to be governed by olfactory recognition of homestream water using odour cues experienced by fish either during the outward journey or during earlier stages of development (reviewed in Dittman & Quinn 1996). It has long been proposed that olfactory cues used by homing fish come from metabolic products released by salmon parr residents (the so-called pheromone hypothesis; Solomon 1973; Nordeng 1977). This population recognition mechanism has been proposed to be a broader scale case of kin recognition, which is influenced by MHC genes (Brown & Brown 1992; Moore *et al*. 1994; Olsén 1998; Olsén *et al*. 1998). The capacity of Atlantic salmon to chose their mates in their natural environment according to their MHC genotypic makeup has also recently been demonstrated (Landry *et al*. 2001). Consequently, the fine-scale genetic differentiation we described might allow the MHC based odours released by resident parr to be used by migrating fish in order to recognize their natal sites once they have entered their river and thus, MHC genes could count among the most likely candidate polymorphic genes that are concerned by the still debated pheromone hypothesis proposed 30 years ago by Solomon (1973) and Nordeng (1977).

Acknowledgements

We acknowledge N. Aubin-Horth, J.J. Dodson, W.C. Jordan and V. Castric and two anonymous referees for useful comments on the manuscript and E. St-Michel for laboratory assistance. C. Landry was financially supported by a postgraduate fellowship from the Natural Sciences and Engineering Research Council of Canada (NSERC) and Gérard D. Lévesque Funds (Gaspésie, Québec, Canada). This study was financially supported by an operating NSERC grant to L.B.

References

Bakke TA, Harris PD (1998) Diseases and parasites in wild Atlantic salmon (*Salmo salar*) populations. *Canadian Journal of Fisheries and Aquatic Sciences*, **55**, 247–266.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London B*, **263**, 1619–1626.

Belkhir K, Borsa P, Goudet J, Chikhi L, Bonhomme F (1998) *GENETIX, logiciel sous Windows TM pour la génétique des populations*. Laboratoire Génome et Populations, CNRS UPR 9060, Université de Montpellier II, Montpellier, France.

Bermingham E, Forbes SH, Friedland K, Pla C (1991) Discrimination between Atlantic salmon (*Salmo salar*) of North American and European origin using restriction analysis of mitochondrial DNA. *Canadian Journal of Fisheries and Aquatic Sciences*, **48**, 884–893.

Bernatchez L (1997) Mitochondrial DNA analysis confirms the existence of two glacial races of rainbow smelt *Osmerus mordax* and their reproductive isolation in the St Lawrence River estuary (Québec, Canada). *Molecular Ecology*, **6**, 73–83.

Bernatchez L, Guyomard R, Bonhomme F (1992) DNA sequence variation of the mitochondrial control region among geographically and morphologically remote European brown trout *Salmo trutta* populations. *Molecular Ecology*, **1**, 161–173.

Bouillon DR, Dempson JB (1989) Metazoan parasite infections in landlocked and anadromous Arctic charr (*Salvelinus alpinus Linnaeus*), and their use as indicators of movement to sea in young anadromous charr. *Canadian Journal of Zoology*, **67**, 2478–2485.

Boyce WM, Hedrick PW, Muggli-Cockett NE *et al.* (1997) Genetic variation of major histocompatibility complex and microsatellite loci: a comparison in bighorn sheep. *Genetics*, **145**, 421–433.

Bradley RD, Hillis DM (1997) Recombinant DNA sequences generated by PCR amplification. *Molecular Biology and Evolution*, **14**, 592–593.

Brown GE, Brown JA (1992) Do rainbow trout and Atlantic salmon discriminate kin? *Canadian Journal of Zoology*, **70**, 1636–1640.

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

Carrington M, Nelson GW, Martin MP *et al.* (1999) HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. *Science*, **283**, 1748–1752.

Carvalho GR (1993) Evolutionary aspects of fish distribution: genetic variability and adaptation. *Journal of Fish Biology*, **43**, 53–73.

Charlesworth B (1998) Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution*, **15**, 538–543.

Dahl (1928) The dwarf salmon of lake Byglandsfjord. A landlocked salmon from Norway. *Salmon Trout Magazine*, **51**, 108–112.

Dittman AH, Quinn TP (1996) Homing in Pacific salmon: Mechanisms and ecological basis. *Journal of Experimental Biology*, **199**, 83–91.

Dixon B, Nagelkerke LAJ, Sibbing FA, Egberts E, Stet RJM (1996) Evolution of MHC class II B hain-encoding genes in the Lake Tana barbel species flock (*Barbus intermedius* complex). *Immunogenetics*, **44**, 419–431.

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

Edwards SV, Hedrick PW (1998) Evolution and ecology of MHC molecules: from genomics to sexual selection. *Trends in Ecology and Evolution*, **13**, 305–311.

Ellegren H, Hartman G, Johansson M, Andersson L (1993) Major histocompatibility complex monomorphism and low levels of DNA fingerprinting variability in a reintroduced and rapidly expanding population of beavers. *Proceedings of the National Academy of Sciences of the USA*, **90**, 8150–8153.

Elson JA (1969) Late quaterny marine submergence of Quebec. *Revue de Géographie de Montréal*, **23**, 247–258.

Ennis PD, Zemmour J, Salter RD, Parham P (1990) Rapid cloning of HLA-A,B complementary DNA by using the polymerase chain reaction: Frequency and nature of errors produced in amplification. *Proceedings of the National Academy of Sciences of the USA*, **87**, 2833–2837.

Ewens WJ (1972) The sampling theory of selectively neutral alleles. *Theoretical Population Biology*, **3**, 87–112.

Gaggiotti OE, Lange O, Rassmann Gliddon K (1999) A comparison of two indirect methods for estimating average levels of gene flow using microsatellite data. *Molecular Ecology*, **8**, 1513–1520.

Garant D, Dodson JJ, Bernatchez L (2000) Ecological determinants and temporal stability of whitin-river population structure in Atlantic salmon (*Salmo salar* L.). *Molecular Ecology*, **9**, 615–628.

Garant D, Dodson JJ, Bernatchez L (2001) A genetic evaluation of mating system and determinants of individual reproductive success in Atlantic salmon (Salmo salar L.). *Journal of Heredity*, **92**, 137–145.

Glamann J (1995) Complete coding sequence of Rainbow trout Mhc II B chain. *Scandinavian Journal of Immunology*, **41**, 365–372.

Goudet J (1998) *FSTAT, Version 2.8*. Institute of Ecology, University of Lauzanne, Lausanne, Switzerland.

Grimholt U, Hordvik I, Fosse VM *et al.* (1993) Molecular cloning of major histocompatibility complex class I cDNAs from Atlantic salmon (*Salmo salar*). *Immunogenetics*, **37**, 469–473.

Hayashi K (1991) PCR–SSCP: a simple and sensitive method for detection of mutations in the genomic DNA. *PCR Methods and Applications*, **1**, 34–38.

Hedrick PW (1992) Female choice and variation in the major histocompatibility complex. *Genetics*, **132**, 575–581.

Hedrick PW (1996) Conservation genetics and molecular techniques: a perspective. In: *Molecular Genetic Approaches in Conservation* (eds, Smith TB, Wayne RK), pp. 459–477. Oxford University Press, New York.

Hedrick PW (1999) Perspective: highly variable loci and their interpretation in evolution and conservation. *Evolution*, **53**, 313–318.

Hedrick PW, Thomson G (1983) Evidence for balancing selection at HLA. *Genetics*, **104**, 449–456.

Hill AVS, Allsopp EM, Kwiatkowski D *et al.* (1991) Common West African HLA antigens are associated with protection from severe malaria. *Nature*, **352**, 595–600.

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

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, Nei M (1989) Nucleotide substitution at major histocompatibility complex class II loci: Evidence for overdominant selection. *Proceedings of the National Academy of Sciences of the USA*, **86**, 958–962.

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: *Mammalian Protein Metabolism* (ed. Munro HN), pp. 21–132. Academic Press, New York.

Karl SA, Avise JC (1992) Balancing selection at allozyme loci in oysters: implications from nuclear RFLPs. *Science*, **256**, 100–102.

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

Klein J (1986) *The Natural History of the Major Histocompatibility Complex*. John Wiley & Sons, New York.

Klein J (1987) Origin of Major Histocompatibility Complex Polymorphism: The trans-species hypothesis. *Human Immunology*, **19**, 155–162.

Kumar S, Tamura K, Jakobsen I, Nei M (2000) *MEGA: Molecular Evolutionary Genetics Analysis, Version 2.0*. Pennsylvania State University, University Park, PA.

Landry C, Garant D, Duchesne P, Bernatchez L. (2001) 'Good genes as heterozygosity': the major histocompatibility complex and mate choice in Atlantic salmon (Salmo salar). *Proceedings of the Royal Society B*, **268**, 1279–1286.

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.

Langefors A, Lohm J, Von Schantz T, Grahn M (2000) Screening of Mhc variation in Atlantic salmon (*Salmo salar*): a comparison of restriction fragment length polymorphism (RFLP), denaturing gradiant gel electrophresis (DGGE) and sequencing. *Molecular Ecology*, **9**, 215–219.

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. *Proceedings of the Royal Society B*, **268**, 479–485.

Lemaire C, Allegrucci G, Naciri M *et al.* (2000) Do discrepancies between microsatellite and allozyme variation reveal differential selection between sea and lagoon in the sea bass (*Dicentrarchus labrax*)? *Molecular Ecology*, **9**, 457–467.

Lewontin RC, Krakauer J (1973) Distribution of gene frequencies as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, **74**, 175–195.

Lynch M, Perender M, Spitze K *et al.* (1999) The quantitative and molecular genetic architecture of a subdivided species. *Evolution*, **53**, 100–110.

Madsen T, Olsson M, Wittzell H *et al.* (2000) Population size and genetic diversity in sand lizards (*Lacerta agilis*) and adders (*Vipera berus*). *Biological Conservation*, **94**, 257–262.

McConnell SK, O'Reilly P, Hamilton L, Wright JN, Bentzen P (1995) Polymorphic microsatellite loci from Atlantic salmon (*Salmo salar*): genetic differentiation of North American and European populations. *Canadian Journal of Fisheries and Aquatic Sciences*, **52**, 1863–1872.

Miller KM, Withler RE (1996) Sequence analysis of a polymorphic Mhc class II gene in Pacific salmon. *Immunogenetics*, **43**, 337–351.

Moore A, Ives MJ, Kell LT (1994) The role of urine in sibling recognition in Atlantic salmon *Salmo salar* (L.) parr. *Proceeding of the Royal Society of London B*, **255**, 173–180.

Nei M, Gojobori T (1986) Simple method for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution*, **3**, 418–426.

Nei M, Kumar S (2000) *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.

Nordeng H (1977) A pheromone hypothesis for homeward migration in anadromous salmonids. *Oikos*, **28**, 155–159.

Olsén H (1998) Present knowledge of kin discrimination in salmonids. *Genetica*, **104**, 295–299.

Olsén KH, Grahn M, Lohm J, Langefors A (1998) MHC and kin discrimination in juvenile Artic Charr, *Salvelinus alpinus* (L.). *Animal Behaviour*, **56**, 319–327.

Ono H, O'Huigin C, Vincek V, Klein J (1993) Exon-intron organization of fish major histocompatibility complex class II beta genes. *Immunogenetics*, **38**, 223–234.

Orti G, Hare MP, Avise JC (1997) Detection and isolation of nuclear haplotypes by PCR–SSCP. *Molecular Ecology*, **6**, 575–580.

Paterson S (1998) Evidence for balancing selection at the major histocompatibility complex in a free-living ruminant. *Journal of Heredity*, **89**, 289–294.

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.). *Proceedings of the National Academy of Sciences of the USA*, **95**, 3714–3719.

Ploegh H, Watts C (1998) Antigen recognition. *Current Opinion in Immunology*, **10**, 57–58.

Pogson GH, Mesa KA, Boutilier RG (1995) Genetic population structure and gene flow in the Atlantic cod *Gadus morhua*: a comparison of allozyme and nuclear RFLP loci. *Genetics*, **139**, 375–385.

Potts WK, Wakeland EK (1993) Evolution of MHC genetic diversity: a tale of incest, pestilence and sexual preference. *Trends in Genetics*, **9**, 408–412.

Potts WK, Manning CJ, Wakeland EK (1991) Mating patterns in seminatural populations of mice influenced by MHC genotype. *Nature*, **352**, 619–621.

Power (1958) The evolution of freshwater races of the Atlantic salmon (*Salmo salar* L.) in eastern North America. *Arctic*, **11**, 86–92.

Raymond M, Rousset F (1995) GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 448–249.

Rice WR (1989) Analysing tables of statistical tests. *Evolution*, **43**, 223–225.

Robertson A (1975) Remarks on the Lewontin-Krakauer test. *Genetics*, **80**, 396.

Rousset F (1996) Equilibrium values of measures of population subdivision for stepwise mutation processes. *Genetics*, **142**, 1357–1362.

Ruzzante DE (1998) A comparison of several measures of genetic distance and population structure with microsatellite data: bias and sampling variance. *Canadian Journal of Fisheries and Aquatic Sciences*, **55**, 1–14.

Schierup MH, Vekemans X, Charlesworth D (2000) The effect of subdivision variation at multi-allelic loci under balancing selection. *Genetical Research*, **76**, 51–62.

Schneider S, Roessli D, Excoffier L (2000) *ARLEQUIN, Version 2.000: A software for population genetics data analysis*. Genetic and Biometry Laboratory, University of Geneva, Geneva.

Slade RW (1992) Limited MHC polymorphism in the southern elephant seal: implications for MHC evolution and marine mammal biology. *Proceeding of the Royal Society of London B*, **249**, 163–171.

Slatkin M (1995) A measure of population subdivision based on allele frequencies. *Genetics*, **139**, 457–462.

Sokal RR, Rohlf FJ (1995) *Biometry. The Principles and Practice of Statistics in Biological Research*. Freeman, New York.

Solomon DJ (1973) Evidence for pheromone-influenced homing by migrating Atlantic salmon, *Salmo salar* (L.). *Nature*, **244**, 231–232.

Spitze K (1993) Population structure in Daphnia obtusa: quantitative genetic and allozymic variation. *Genetics*, **135**, 367–374.

Stet RJM, Egberts E (1991) The histocompatibility system in teleostean fishes: from multiple histocompatibility loci to a major histocompatibility complex. *Fish and Shellfish Immunology*, **1**, 1–16.

Takahata N, Nei M (1990) Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics*, **124**, 967–978.

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

Tessier N, Bernatchez L (1999) Stability of population structure and genetic diversity across generations assessed by microsatellites among sympatric populations of landlocked Atlantic salmon (*Salmo salar* L.). *Molecular Ecology*, **8**, 169–179.

Tessier N, Bernatchez L (2000) A genetic assessment of single versus double origin of landlocked Atlantic salmon (*Salmo salar*) from Lake Saint-Jean, Québec, Canada. *Canadian Journal of Fisheries and Aquatic Sciences*, **57**, 197–804.

Tessier N, Bernatchez L, Wright JM (1997) Population structure and impact of supportive breeding inferred from mitochondrial and microsatellite DNA analyses in landlocked Atlantic salmon *Salmo salar* L. *Molecular Ecology*, **6**, 735–750.

Van Muiswinkel W, Wiegertjes G, Stet R (1999) The influence of environmental and genetic factors on the disease resistance of fish. *Aquaculture*, **172**, 103–110.

Verspoor E, Jordan WC (1989) Genetic variation at the *Me*-2 locus in the Atlantic salmon within and between rivers: evidence for its selective maintenance. *Journal of Fish Biology*, **35**, 205–213.

Watterson GA (1978) The homozygosity test of neutrality. *Genetics*, **88**, 405–417.

Weir BS (1990) *Genetic Data Analysis*. Sinauer Associates, Sunderland, MA.

Weir BS (1996) *Genetic Data Analysis II*. Sinauer Associates, Sunderland, MA.

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

This study is the main part of C. Landry's MSc thesis on the role of Mhc variation in population divergence and mating patterns in Atlantic salmon supervised by L. Bernatchez. Louis Bernatchez's major interests are in the understanding of patterns and processes of molecular and organismal evolution, as well as their significance to conservation.