

## Characterization of a major histocompatibility class II A gene (*Clha-DAA*) with an embedded microsatellite marker in Atlantic herring (*Clupea harengus* L.)

R. J. M. STET\*†, K. MUDDÉ‡, J. W. WYNNE\*§, A. NOOIJEN‡,  
T. G. DAHLGREN||¶, D. E. RUZZANTE# AND C. ANDRÉ\*\*

\*Scottish Fish Immunology Research Centre, School of Biological Sciences, University of Aberdeen, Aberdeen, U.K., ‡Wageningen Institute of Animal Sciences, Wageningen University, Wageningen, The Netherlands, ||Department of Marine Ecology, Göteborg University, Tjärnö Marine Biological Laboratory, S-45296, Strömstad, Sweden and #Department of Biology, Dalhousie University, Halifax NS, B3H 4J1 Canada

(Received 14 June 2007, Accepted 21 February 2008)

An Atlantic herring major histocompatibility class II A (*Clha-DAA*) cDNA sequence has been characterized and was shown to encode a leader peptide, alpha-1 domain, alpha-2 domain, connecting peptide, transmembrane and cytoplasmic region. The *Clha-DAA* protein sequence has all the characteristics of a teleost class II A protein with conserved cysteines in both the alpha-1 and the alpha-2 domains and two potential N-linked glycosylation sites. Exon 2 sequences encoding the polymorphic alpha-1 domain from different individuals were analysed and revealed the presence of at least two loci. The *Clha-DAA* gene consists of four exons and three short introns. Four unique intron 3 sequences from multiple individuals were obtained and were shown to contain a (TG)<sub>n</sub> microsatellite sequence. Primers were optimized such that only a single microsatellite locus designated *Clha-DAA-INTR3* was amplified. Four herring populations from the North Sea and the Baltic Sea were genotyped for *Clha-DAA-INTR3*. In total, 16 *Clha-DAA-INTR3* alleles were detected; the distribution of the alleles showed no deviation from Hardy–Weinberg expectation. Levels of genetic differentiation among samples were of similar magnitude as have been reported earlier for neutral microsatellite loci between northern North Sea and Baltic Sea herring populations. © 2008 The Authors

Journal compilation © 2008 The Fisheries Society of the British Isles

Key words: Baltic Sea; *Clupea harengus*; MHC; microsatellite; North Sea; population genetics.

### INTRODUCTION

The major histocompatibility complex (MHC) genes are found in a complex of tightly linked genes encoding class I and class II cell-surface proteins (Marsh *et al.*, 2000). In addition to the MHC genes proper, the complex also contains class III genes, a number of which are involved in immune responses, *e.g.*

\*\*Author to whom correspondence should be addressed. Tel.: +46 52668600; fax: +46 52668607; email: [carl.andre@marecol.gu.se](mailto:carl.andre@marecol.gu.se)

†Deceased.

§Present address: CSIRO Marine and Atmospheric Research, Hobart, Tasmania, Australia.

¶Present address: Department of Zoology, Göteborg University, P. O. Box 463, S-405 30 Göteborg, Sweden.

complement factors and cytokines. The classical class I and class II proteins present self and non-self peptides to T lymphocytes and thereby initiate adaptive immune responses. An important hallmark of the classical MHC genes is that they encode highly polymorphic MHC class I and class II proteins. Humans are by far the best-studied species with 1839 and 875 alleles for the combined class I (*HLA-A*, *-B*, *-C*, *-E*, *-F* and *-G*) and combined class II *A* and *B* genes (*HLA-DR*, *-DQ*, *-DP*, *-DO* and *-DM*), respectively (<http://www.ebi.ac.uk/imgt/hla/stats.html>).

The first major histocompatibility (MH) genes identified in teleost fish were those in the common carp *Cyprinus carpio* L. (Hashimoto *et al.*, 1990). This was soon followed by many reports on MH class I and class II genes in a variety of fish species (Dixon *et al.*, 1995). In several species, MH class I and class II *B* genes have been studied in much detail, while class II *A* has received far less attention (Dixon *et al.*, 1996; Landry & Bernatchez, 2001; Miller *et al.*, 2001; Kruiswijk *et al.*, 2005). This is probably because of the fact that in mammals the class II *A* genes are less polymorphic than the class II *B* genes. The best example is the *HLA-DR* locus in humans with 577 *HLA-DRB* alleles and only three *HLA-DRA* alleles. MH class II *A* genes and cDNA sequences have been reported in a number of fish species such as zebrafish *Danio rerio* (Hamilton) (Sultmann *et al.*, 1993), common carp (van Erp *et al.*, 1996), cichlids (Murray *et al.*, 1999), striped bass *Morone saxatilis* (Walbaum) (Hardee *et al.*, 1995) channel catfish *Ictalurus punctatus* (Rafinesque) (Godwin *et al.*, 2000), large African barb (Kruiswijk *et al.*, 2004), stickleback *Gasterosteus aculeatus aculeatus* L. (Reusch *et al.*, 2004) and in rainbow trout *Oncorhynchus mykiss* (Walbaum) and Atlantic salmon *Salmo salar* L. (Grimholt *et al.*, 2000). In Atlantic salmon, the MH class II *A* (*Sasa-DAA*) has been studied in great detail and was shown to be almost as polymorphic as the MH class *B* (*Sasa-DAB*) gene (Stet *et al.*, 2002; Consuegra *et al.*, 2005; Wynne *et al.*, 2007).

Unlike all other jawed vertebrates that have the MHC genes located in a single chromosomal region of 4 Mb, teleost fish have class I and class II genes located on different chromosomes. This was reported first in zebrafish (Bingulac-Popovic *et al.*, 1997; Takami *et al.*, 1997) and subsequently confirmed in other species, *e.g.* Atlantic salmon (Grimholt *et al.*, 2002). The unlinked nature of the MH genes in teleost fish has led to the proposal to designate these genes as MH genes (Stet *et al.*, 2003). There are several hypotheses that attempt to explain the unlinked nature of MH genes in teleost fish. Kasahara (1999) suggested that this was because of genome duplications that resulted in the identification of at least three paralogous regions in the genome of jawed vertebrates. Inactivation of either class I genes or class II genes in one of the paralogous region could lead to the unlinked nature of MH genes in teleost fish. This has been substantiated by analyses of class I genes in cyprinid fishes (Stet *et al.*, 2003). The alternative explanation assumes that class II genes have been expelled from the primordial MHC by translocation (Kuroda *et al.*, 2002).

Pathogen-mediated selection has been hypothesized to be the driving force for maintaining MHC polymorphism within populations (Prugnolle *et al.*, 2005). However, few empirical studies underpin this notion (Sommer, 2005). Empirical studies have been performed in Atlantic salmon in both farmed (Grimholt *et al.*, 2003; Kjøglum *et al.*, 2006) and wild populations (De Eyto

*et al.*, 2007). The unlinked nature of the MH genes in teleost fishes has made it possible to investigate the role of each class of MH genes in disease resistance or susceptibility independently. Large-scale experiments using bacterial (*Aeromonas salmonicida*) and viral (infectious salmon anaemia virus) challenges in farmed Atlantic salmon have shown statistically significant correlations between disease resistance and particular MH class I and class II alleles (Grimholt *et al.*, 2003; Kjøglum *et al.*, 2006). De Eyto *et al.* (2007), however, performed a field experiment and compared observed MH genotype frequencies of Atlantic salmon surviving in a river 6 months after their introduction as eggs with frequencies expected from parental crosses. Significant differences were observed for the class II locus but not for the class I locus. Analysis of the data showed that survival was associated with additive allelic effects at the MH class II A locus. No differences were observed at the class I locus or at seven non-MH-linked microsatellite loci. Collectively, the data from the Atlantic salmon studies suggest that the MH class I and class II loci are under selection. In both studies, MH class I and class II genotyping were facilitated by the presence of a minisatellite marker in the 3' untranslated region (UTR) of the MH class II A locus and a microsatellite marker in the 3' UTR of the class I locus.

In the present study, the MH class II A locus was characterized in Atlantic herring *Clupea harengus* L., a highly abundant and migratory pelagic marine fish, and a microsatellite marker embedded in this locus was determined. Furthermore, the differentiation was investigated in the microsatellite marker among four field samples from the North Sea–Baltic Sea region.

## MATERIALS AND METHODS

### SAMPLE COLLECTION

Atlantic herring were collected on board research vessels and a commercial trawler. Tissue samples from the liver, spleen, muscle and fins were dissected from fresh fish and stored in RNAlater (Ambion, Austin, TX, U.S.A.) at 4° C for future use. Fish were targeted in spawning condition (maturity stage six), and in total, 336 individuals were obtained from their spawning grounds in the year 2002 at four geographically distinct locations: Berwick (56°18' N 0°58' W), Karmøy (59°14' N 05°10' E), Tjøme (59°35' N 10°55' E) and Rügen (54°23' N 13°43' E). These sample locations cover the North Sea, Skagerrak and the low-salinity western Baltic Sea.

### DNA AND RNA ISOLATION

Genomic DNA was isolated from the liver of selected individuals using the Wizard Genomic DNA purification kit according to the manufacturer's specification (Promega, Madison, WI, U.S.A.). Genomic DNA from all the sample collection individuals was prepared by the hotshot DNA isolation method (Truett *et al.*, 2000) using fin clips or muscle. Subsequently, the DNA was purified over S-400 spin columns (Amersham Pharmacia Biotech, Little Chalfont, U.K.). RNA was isolated from liver and spleen samples using the method described by Dixon *et al.* (1996). DNA and RNA concentrations were determined with the GeneQuant system (Amersham Pharmacia Biotech).

### cDNA LIBRARY CONSTRUCTION

A cDNA library was constructed with mRNA isolated from the liver of a single fish. An mRNA was converted into cDNA using the SMART cDNA library kit and ligated

into the Triplex Lambda vector following the manufacturer's specification (Clontech, Mountain View, CA, U.S.A.). The primary library with a titre of  $1.5 \times 10^6$  PFU was amplified to  $10^9$  PFU ml<sup>-1</sup>. The library was stored at 4° C for further use.

### RAPID AMPLIFICATION OF cDNA ENDS

Overlapping 5' and 3' cDNA ends of the Atlantic herring class II A gene sequences were obtained by rapid amplification of cDNA ends (RACE) using the GeneRacer kit according to the manufacturer's specifications (Invitrogen, Carlsbad, CA, U.S.A.). The template RNA used was isolated from the same individual that was used to construct the cDNA library. The resulting cDNA fragments were cloned and sequenced as described below.

### POLYMERASE CHAIN REACTION AMPLIFICATIONS OF *MHCCLHA-DAA* SEQUENCES

Herring sequences were obtained by anchored polymerase chain reaction (PCR) using aliquots of cDNA library as template. Degenerate primers were designed based on alignments of MHC class I and class II sequences or were available from other studies (Hashimoto *et al.*, 1990) and combined with Triplex reverse and forward primers in an anchored PCR (Table I). The reactions were performed in *Taq* buffer using 1 unit of *Taq* polymerase (Eurogentec, Herstal, Belgium) supplemented with 1.5 mM MgCl<sub>2</sub>, 200 µM dNTPs and 0.2 µM each of the forward and reverse primers under the following PCR conditions: denaturation at 94° C for 4 min followed by 30 cycles consisting of denaturation at 94° C for 30 s, annealing at 50–55° C (depending on the type of amplification) for 30 s, followed by polymerization at 72° C for 60 s and final single step of polymerization at 72° C for 10 min. Templates consisted of genomic DNA, cDNA or aliquots of the cDNA library. The PCRs were performed using the GeneAmp PCR system 9700 (Applied Biosystems, Foster City, CA, U.S.A.).

TABLE I. Polymerase chain reaction (PCR) primers

Primers	Sequence 5'–3'
1. TLAR	AGR CTK GKR TGC TCC ACI TGR CA
2. Triplex5	CTC GGG AAG CGC GCC ATT GTG TTG GT
3. Triplex3	ATA CGA CTC ACT ATA GGG CGA ATT GGC C
4. Clha-DA267F	GAA GTT AGC AGC AGT CAG CTC CGC
5. Clha-DA307R	CCT CTC CTG AGG GGT GAA CTT CAG
6. Clha-DALP1	CTT ACT GGG ATT ATC TGC ACA GAG
7. Clha-DA-3uR	CAG AGG GTT TAT TGT GAA GTT TCA
8. Clha-DAA1F1	CAG ATG TGG ATA TTG CTC TG
9. Clha-DAA1R1	CTG CTT CTG CTG GAC TTT TGT A
10. Clha-DAA1R2	ATC TGA GTC TGT GCA CCC GAC CAG
11. Clha-DAA1F2	AGG TGA AGG TGG GCA CTG TCA ACA CG
12. Clha-DAA2F	TGC ACT GTG GAG CAC CAG GCA CTG
13. Clha-DAA2R7	GAT GTA GCA GAT GAG CGT GTT GA
14. ClhaDA-TmR	AGC CAC TCC CAG CAG TCC CAG AGT
15. Clha-DAinR1	GCA CAA GAC TGG GCT CTG AC
16. Clha-DAinF10	GTA CTG GTA GGT TTG TCT ATA A

## CLONING AND SEQUENCING

Single PCR fragments were cloned directly into pGEMT-easy according to the manufacturer's instructions (Promega). Alternatively, when multiple fragment sizes were obtained, bands of the expected sizes were excised from the agarose gel and DNA was isolated using the Qiagen Quick Gel Extraction kit (Qiagen, Valencia, CA, U.S.A.). Fragments obtained were ligated into pGEMT-easy and transformed into competent JM109 bacteria using standard protocols. Plasmid DNA was isolated from single colonies grown overnight using the QIAprep Spin Miniprep kit (Qiagen). Sequences were determined for both the DNA strands with the BigDye Terminator Cycle Sequencing kit (Applied Biosystems) on an ABI 377 automated sequencer. Ten individual clones were sequenced for each amplicon. Sequences were further analysed with Sequencher software (Gene Codes, Ann Arbor, MI, U.S.A.). Identity to database sequences was assessed by basic local alignment search tool (BLAST); National Center for Biotechnology Information, MD, U.S.A. and sequences were aligned using CLUSTAL X. A neighbour-joining tree was constructed using MEGA 2.1 and confidence in individual nodes was assessed by 1000 bootstrap replications (Kumar *et al.*, 2001).

## *CLHA-DAA-INTR3* EMBEDDED MICROSATELLITE MARKER

Primers were designed based on the available *Clha-DAA* cDNA sequence with each primer pair spanning a putative intron. Resulting single PCR fragments were cloned and sequenced. Individual bands obtained from intron 3 amplifications from multiple individuals were excised from agarose gels and subjected to cloning and sequence analyses. Multiple alignments of the intron 3 sequences were constructed and used to design primers that only amplify a single intron 3 sequence containing the informative microsatellite repeat sequence.

The optimized primers (forward primer *Clha-DAinF10* and reverse primer *Clha-DAinR1*; Table I) were used to genotype the four populations. PCR amplifications were performed using hotshot isolated DNA that was purified over Sephacryl-400 spin columns using the standard cycling conditions as described above with the annealing temperature set at 50° C. PCR products were analysed with an ABI 377 automated sequencer (GMI, Ramsey, MN, U.S.A.) and fragment sizes were assessed by Genotyper software (Applied Biosystems).

## POPULATION GENETIC ANALYSIS

Amounts of genetic variation within herring samples were estimated as expected and observed heterozygosity. Departure from Hardy–Weinberg expectation (HWE) was examined with exact tests using the software GENEPOP 3.4. (Rousset & Raymond, 1995). Population differentiation,  $F_{ST}$ , was estimated per sample pair and overall using the unbiased estimator  $\theta$  (Weir & Cockerham, 1984). Statistical significance of  $F_{ST}$  was examined using a permutation test (10 000 randomizations) implemented in FSTAT (Goudet, 2001). The presence of null alleles and scoring errors because of stuttering and large allele drop-out was assessed by MICRO-CHECKER (Van Oosterhout *et al.*, 2004).

## RESULTS

### CHARACTERIZATION OF MHC CLASS II ALPHA GENE

A highly degenerate primer designed around the conserved second cysteine in the MHC class I alpha-3 and MHC class II beta-2 domains of genes of human, mouse and chicken (Hashimoto *et al.*, 1990) was combined with a Triplex5 primer in an anchored PCR with the herring cDNA library as template (Table I: primers 1 and 2). Bands of the expected size (*c.* 600 bp) were excised from

agarose gel, cloned and sequenced. One of the clones had a high score with teleost MH class II A sequences when analysed using BLAST. The sequence encoded the 5' end of a herring class II A cDNA and was used to design specific primers (Table I: primers 4 and 5), allowing the amplification of overlapping 5'- and 3'-ends encoding the complete cDNA sequence using 5' and 3' RACE. Subsequently, primers were designed in the leader peptide and 3' UTR (Table I: primers 6 and 7) to obtain the full-length cDNA sequence. A single sequence was obtained encoding the complete MHC class II A chain and was designated *Clha-DAA\*0101*, according to standard nomenclature (Ellis *et al.*, 2006). The *Clha-DAA\*0101* cDNA sequence has been deposited in the EMBL database and assigned the accession number AM492999. The cDNA sequence encodes the 5' UTR, leader peptide, alpha-1 and alpha-2 domains, connecting peptide,

**5' UT**

-60

acatggactgaaggagttagaaaattcaacactgagacccttcggttcctgcacccaaag

**Leader Peptide**

ATG ACT TTA ACA TGG ATT TTA CTG CTC CTT ACT GGG ATT ATC TGC ACA GAG ACT 54  
M T L T W I L L L L T G I I C T E T 18

**Alpha-1 domain**

AAG ATT GTA CAT GTG GAT ATT GCT CTG GTC GGG **TGC** ACA GAC TCA GAT GGA GAG AAG ATG TTT 117  
K I V H V D I A L V G **C** T D S D G E K M F 39

GGA CTG GAT GGG GAA GAG AAG GGC CAC GCA GAC TTC ACC AAG GGC AAA TTC ATT ATG ACA CTG 180  
G L D G E E K G H A D F T K G K F I M T L 60

CCT GAG TTT GCT GAT CCG TTT AAA TAT GAG GAA GGC GCA TAT GAG GGT GCA GTT CGT GAC AAG 243  
P E F A D P F K Y E E G A Y E G A V R D K 81

GAG GTC **TGC** AAA CAG AAC TTG CAA GTA GCC ATA CAA GCA TAC AAA AGT CCA GCA GAA GCA GAG 306  
E V **C** K Q N L Q V A I Q A Y K S P A E A E 102

**Alpha-2 domain**

GCC CCC CCC ATG AGC TCC ATC TAC CCC AGA CAT GAG GTG AAG GTG GGC ACT GTC AAC ACG CTC 369  
A P P P M S S I Y P R H E V K V G T V N T L 123

ATC **TGC** TAC ATC GCT GGG TTC TAC CCT CCC CGG CTC ACC GTG AGG TGG ACC AGG AAC AAC AAG 432  
I **C** Y I A G F Y P P R L T V R W T R N N K 144

AAC GTG ACC CAG GGA GTG AGC AGC AGT CAG TTG CGC TTG AAT GTT AAT GAC CTC TCC TTC AAC 495  
**N V T** Q G V S S S Q L R L N V N D L S F N 165

CAG TTC TTC ACC CTG AAT **TTC ACC** CCT CAG GAG GGG GAC ATG TAC ACC **TGC** ACT GTG GAG CAC 558  
Q F F T L **N F T** P Q E G D M Y T **C** T V E H 176

CAG GCA CTG GAG GGG CCC ATG ACA AGA GAG TTT GAT 594  
Q A L E G P M T R E F D 198

< **Connecting peptide**

&gt;&lt;

**Transmembrane region**

GTT GAG GTG TCA GAG CCC AGT CTT GGT CCC TCA GTG TTC TGT GGA GTG GGT CTG ACT CTG GGA 657  
V E V S E P S L G P S V F C G V G L T L G 219

>> **Cytoplasmic region** >

CTG CTG GGA GTG GCT ACT GGA ACA TTC TTC CTC GTC AAG GGA AAC CAG TGC AAC TGA 714  
L L G V A T G T F F L V K G N Q C N \* 237

**3' UT**

cattctctgtctctctgcctcagcctttatggtgaaagttaaatggaaga  
atgaaacttcacataaaaccctctgtaaatc(a)<sub>n</sub>

795

FIG. 1. Complete cDNA sequence of the Atlantic herring major histocompatibility class II A gene (*Clha-AA\*0101*). Conserved cysteines in the alpha-1 and alpha-2 domains are highlighted in bold. Potential N-linked glycosylation sites are boxed and the polyadenylation site is underlined.

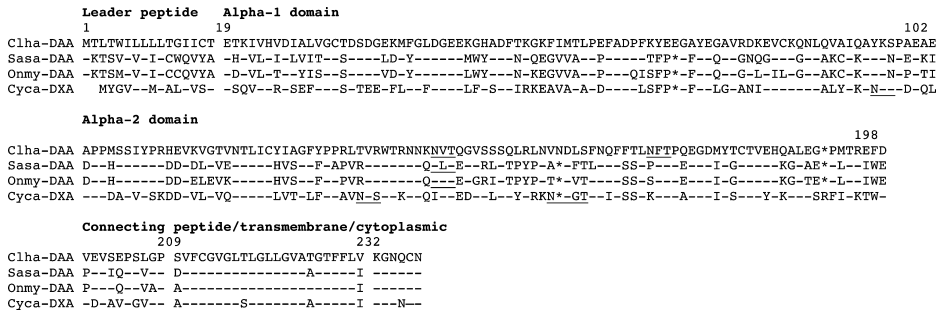


FIG. 2. Multiple alignment of Atlantic herring major histocompatibility class II A protein (*Clha-DAA*, accession number AM492999) with Atlantic salmon (*Sasa-DAA*, accession number L77086), rainbow trout (*Onmy-DAA*, accession number AJ251432) and common carp (*Cyca-DXA*, accession number X95432) class II A proteins. Potential N-linked glycosylation sites are underlined.

transmembrane and cytoplasmic regions (Fig. 1). The 5' UTR is 60 bp in length, and the 3' UTR is very short with only 85 bp but contains the canonical polyadenylation site and poly-A tail. The 3' UTR of *Clha-DAA*\*0101 does not contain any minisatellite or microsatellite repeat sequences.

The deduced protein sequence [237 amino acids (aa)] contains all known features of a teleost MHC class II alpha chain (Fig. 2). The alpha-1 domain has the characteristic additional cysteines, at positions 30 and 84, not found in other vertebrate species (Stet *et al.*, 1996). The alpha-2 domain contains two putative N-linked glycosylation sites at positions 145–147 and 171–173. The remainder of the alpha chain consists of the connecting peptide, transmembrane region and cytoplasmic region. The transmembrane and cytoplasmic regions are highly conserved and almost identical to those of carp, rainbow trout and Atlantic salmon (Fig. 2). The percentage identity and similarity between the herring class II A aa sequence and those of other teleost species ranges from 39 to 51% identity and 58 to 68% similarity, with the highest scores with Atlantic salmon *Sasa-DAA*.

GENOMIC ORGANIZATION

To investigate the presence of embedded repeat sequences, *i.e.* polymorphic microsatellite or minisatellite, the intron sequences were amplified using primers based on the available cDNA sequence (Table I: primers 6, 8–14). The organization of the class II A genes in teleost fish is highly conserved and based on the high identity and similarity of the *Clha-DAA* sequence with that of Atlantic salmon (*Sasa-DAA*) the authors were able to predict accurately the position of these introns (Dixon *et al.*, 1995; Stet *et al.*, 2002). Genomic DNA of the same individual used to construct the cDNA library was used as template in PCR amplification with primers spanning each putative intron. The PCRs resulted in a single sequence when amplifying intron 1 and intron 2 but revealed multiple sequences for intron 3 (data not shown). Sequence analyses of the fragment-spanning intron 1 showed that exon 1 encodes the leader peptide and 3 aa of the alpha-1 domain (Fig. 3). Intron 1 is 516 bp long and does not contain any repeat sequences. Similar analysis of the fragment-spanning intron 2 showed

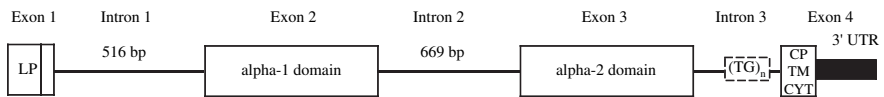
*Clha-DAA\*0101*

FIG. 3. Genomic organization of the Atlantic herring major histocompatibility class A gene *Clha-DAA\*0101* determined by sequence analysis of sequential polymerase chain reaction (PCR) products derived from multiple clones. Exons encoding leader peptide (LP) and four amino acids of the alpha-1 domain, remainder of the alpha-1 domain, alpha-2 domain, connecting peptide (CP), transmembrane (TM), cytoplasmic (CYT) region and 3' untranslated region (3' UTR).

that this intron is 669 bp long and also has no repeat sequences. Intron 3 analyses were more complicated as four fragments were obtained after amplification using primers located in the exons encoding the alpha-2 domain and connecting peptide (Table I: primers 12 and 15). Single PCR fragments were excised from the agarose gel and cloned and sequenced. Sequence analyses revealed four unique intron 3 sequences, which all contained a dinucleotide repeat (TG)<sub>n</sub> sequence (data not shown). The introns differed in sequence, presence of indels and length of the repeat. This potentially interesting intron 3 was chosen for further investigations because of the presence of the (TG)<sub>n</sub> repeat.

#### VALIDATION OF THE MICROSATELLITE REPEATS IN INTRON 3 OF *CLHA-DAA*

Genomic DNA from eight randomly chosen individuals from the Karmøy and Berwick samples was used to analyse intron 3 sequences in more detail. Primers were designed in conserved regions of the intron 3 and connecting peptide, and *Clha-DAA* intron 3 was amplified from different individuals. The PCR amplifications resulted in multiple bands per individual. In total, 25 bands were excised from agarose gel and cloned and sequenced. The sequence analyses revealed the presence of four different intron sequences. Multiple alignments of all the available intron 3 sequences allowed to design specific forward primers that would amplify intron 3 sequences containing one or two alleles of the (TG)<sub>n</sub> microsatellite marker embedded in the *Clha-DAA* locus. The reverse primer was the same in each combination and was designed in the region encoding the connecting peptide (Table I: primer 15). Each primer set was tested on multiple individuals from the different geographical regions. Only the primer combination *Clha-DAA*inF10 and *Clha-DAA*inR1 (Table I: primers 15 and 16) consistently amplified a fragment around 100 bp from all individuals tested as visualized on agarose gels. The optimal annealing temperature was 50° C, and the genomic DNA template isolated by the hotshot method was purified over a Sephacryl S-400 spin column for consistent results. This new microsatellite marker has been designated *Clha-DAA-INTR3* to reflect the fact that the marker is located in intron 3 of the Atlantic herring *C. harengus* MH class II A gene (*Clha-DAA*).

#### MICROSATELLITE ANALYSES

*Clha-DAA-INTR3* microsatellite genotypes of 336 Atlantic herring from four different locations (Berwick, Karmøy, Tjøme and Rügen) were determined.

TABLE II. Genetic variation within samples of herring

Sample	<i>n</i>	<i>A</i>	<i>A<sub>r</sub></i>	<i>H<sub>e</sub></i>	<i>H<sub>o</sub></i>	<i>P</i>
Berwick	89	11	10.4	0.725	0.809	0.073
Karmøy	81	12	11.3	0.764	0.741	0.239
Tjøme	98	14	12.9	0.780	0.714	0.184
Rügen	68	11	11.0	0.775	0.794	0.473

*A*, number of alleles; *A<sub>r</sub>*, allelic richness; *H<sub>e</sub>*, expected heterozygosity; *H<sub>o</sub>*, observed heterozygosity; *n*, number of individuals; *P*, probability of departure from Hardy–Weinberg expectations.

The number of alleles varied from 11 to 14 among samples (Table II); there were in total 16 alleles ranging in size from 109 to 139 bp and forming a perfect 2 bp ladder (Table III). Allelic richness was similar in the four samples (Table II). Allele frequencies in all samples agreed with HWEs (Table II), and there were no indications of scoring errors or null alleles. There was an overall genetic heterogeneity among samples ( $F_{ST} = 0.0060$ ;  $P = 0.020$ ), and the pair-wise  $F_{ST}$  values indicate that Baltic herring from Rügen may be genetically distinct from North Sea herring (Table IV).

#### *CLHA-DAA* SEQUENCES

The analyses of intron 3 sequences suggested that there could be more than one MHC class II *A* locus in Atlantic herring. To further analyse the class II *A* locus, primers were designed at the beginning and end of exon 2, encoding the alpha-1 domain (Table I: primers 8 and 9). The region on which the primers were designed does not contain polymorphic peptide-binding residues (PBRs) and allows amplification of all class II *A* sequences (Stet *et al.*, 2002). The PCR resulted in a single band of the expected size, which was cloned and

TABLE III. *Clha-DAA-INTR3* allele frequencies per sample

Allele (bp)	Berwick	Karmøy	Tjøme	Rügen
109		0.62		
111				0.74
113	1.12	0.62	1.02	0.74
115	23.60	17.90	28.57	30.15
117	7.87	14.20	8.66	11.76
119	1.12	2.47	3.57	
121	44.38	41.36	34.69	33.09
123	13.48	11.11	8.67	6.62
125	3.37	4.94	7.14	9.56
127	1.12	4.32	2.04	5.15
129		1.23	1.53	0.74
131	2.81		2.04	0.74
133		0.62	1.02	
135	0.56	0.62		
137	0.56		0.51	
139			0.51	0.74

```

10
DAA*0101 VGCTDS DGEK MFGLDGE EKG HADFTKGKFI MTLPEFADPF KYBEGAYEGA VRDKEVCKQN LQVAIQ A 76
DAA*0102 -----M- -----R-MV -----Q- -----T-----
DAA*0201 T-----D -H-----MA ----F-R-YV ----Q- --V--T--Q- --AGQQI---- --TI----
DAA*0301 Q-----Y LY-----MA -----V -----QL T-P--E--Q- --GQEQI---- --N-K--
DBA*0101 T-----D -Y-----IA -T-----L- G--K-*--QS -F--T--P- --VMQQG---D ---D-ES
DBA*0102 T-----E-D -Y-----IA -T-----LV G--K-*--QS NF--T--Q- --VMQQG---D ---D-ES
DBA*0201 S-----D -Y-----MA -----L- GK--K-*--QS -FAKAT-KQ- ATMQQG---D ---D-ES
DBA*0202 S-----D -Y-----MA Y-----L- GK--K-*--QS -FAKDT-KQ- ATMQQG---D ---D-ES

```

FIG. 4. Multiple alignment of partial Atlantic herring major histocompatibility class II A exon 2 translated sequences (accession numbers AM492999–AM4933006). Numbering is based on the mature *Clha-DAA\*0101* protein.

sequenced. Eight MHC class II A sequences were obtained from four individuals with one sequence, *Clha-DAA\*0201*, shared between all individuals, and one or two additional unique sequences (Fig. 4). The exon 2 sequences have been deposited in the EMBL database and were assigned the accession numbers AM493000–AM493006. *Clha-DAA\*0101* was the sequence obtained from the cDNA library. Phylogenetic analyses of the sequences revealed the presence of at least two loci, designated *Clha-DAA* and *Clha-DBA* (Fig. 5). The *Clha-DBA* locus, apart from being divergent from the *Clha-DAA* locus, is characterized by the presence of a single aa deletion in the alpha-1 domain (Fig. 4). The sequences *Clha-DAA\*0201* and *\*0301* cluster with the *Clha-DAA\*0101* and *\*0102* sequences, but their position is not supported by a high bootstrap value. It is unclear whether these sequences belong to separate loci.

## DISCUSSION

In teleost fish, the characterization and polymorphism of the MH class II A genes have received far less attention than those of the class II B genes. However, studies in Atlantic salmon *S. salar* have shown that the class II A genes are almost as polymorphic as the class II B genes (Stet *et al.*, 2002; Consuegra *et al.*, 2005; Wynne *et al.*, 2007). In the present study, a class II A cDNA

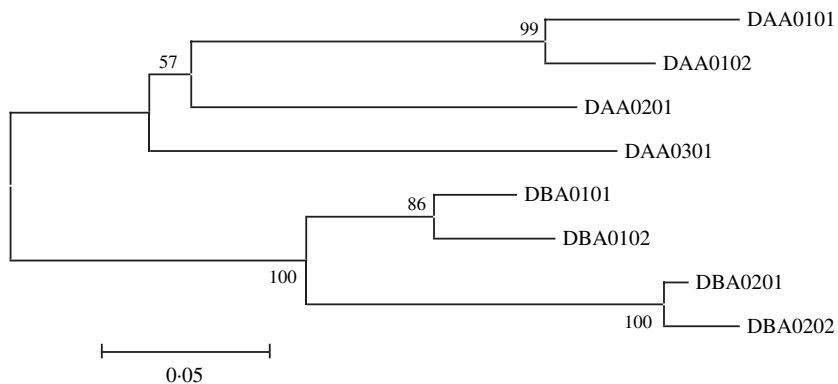


FIG. 5. Phylogenetic analysis of Atlantic herring major histocompatibility class II A sequences. Analysis was performed using the neighbour-joining method based on partial amino acid sequences of exon 2. Topology was evaluated by 1000 bootstrap replications. Accession numbers of the sequences included are depicted in Fig. 4.

TABLE IV. Pair-wise comparison of genetic differentiation

$F_{ST}/P$	Berwick	Karmøy	Tjøme	Rügen
Berwick		0.059	0.164	<b>0.004</b>
Karmøy	0.002		0.120	0.061
Tjøme	0.006	0.008		0.331
Rügen	0.014	0.011	-0.003	

$F_{ST}$  below and  $P$  above diagonal. Statistically significant differentiation after correction for multiple tests is given in bold.

sequence were isolated from Atlantic herring using an approach similar to that of Hashimoto *et al.* (1990). A degenerate reverse primer located in the sequence encoding the class II alpha-2-beta-2 domains were combined with lambda primers in an anchored PCR. After initial 5' and 3' amplifications, the complete class II A sequence was obtained by combining forward and reverse primers in the 5' and 3' UTRs. The sequence was designated *Clha-DAA\*0101* according to standard nomenclature (Ellis *et al.*, 2006). The complete cDNA encoding the class II A protein has all the features of a teleost class II A, including additional cysteines in the alpha-1 domain and glycosylation sites in both the alpha-1 and the alpha-2 domains as described previously for teleost class II A proteins (Dixon *et al.*, 1995; van Erp *et al.*, 1996; Grimholt *et al.*, 2000). Comparison of the herring *Clha-DAA* protein with those of other teleost species revealed highly conserved transmembrane and cytoplasmic regions. Although in the present study, the degenerate reverse primer designed on the coding region surrounding the second cysteine in the alpha-2 domain was successfully used, the highly conserved transmembrane and cytoplasmic regions could be used to design additional degenerate primers for isolating MH class II A sequences in other teleost species.

The availability of the complete *Clha-DAA* cDNA sequence facilitated the establishment of genomic organization of the *Clha-DAA* gene. The *Clha-DAA* gene organization is very similar to those in Atlantic salmon, zebrafish *D. rerio* and channel catfish *I. punctatus* and is characterized by short introns 1, 2 and 3 (Sultmann *et al.*, 1993; Godwin *et al.*, 2000; Stet *et al.*, 2002). Neither *Clha-DAA* introns 1 and 2 nor the 3' UTR contain a microsatellite or minisatellite, while intron 3 has a (TG)<sub>n</sub> repeat similar to intron 3 of Atlantic salmon and channel catfish. Intron 3 was further investigated by amplifying the intron from several individuals. This revealed the presence of two to four fragment sizes per individual and suggested that there are more class II A loci present in herring and that the loci probably show haplotype variation, *i.e.* different gene number per locus. The presence of multiple class II A loci and haplotype variation is consistent with the situation observed in carp (van Erp *et al.*, 1996), cichlids (Murray *et al.*, 2000) and zebrafish (Sultmann *et al.*, 1993). In contrast, salmonids seem to possess only a single major expressed class II A locus (Grimholt *et al.*, 2000; Stet *et al.*, 2002). The presence of multiple loci prompted the authors to amplify exon 2 sequences from a number of individuals. Phylogenetic analysis of the sequences obtained revealed the presence of at least two loci designated *Clha-DAA* and *Clha-DBA*. The assignment of the latter locus was

based both on phylogenetic analysis and on the fact that the *Clha-DBA* alpha-1 domain is 1 aa shorter compared with the *Clha-DAA* alpha-1 domain. The sequence *Clha-DAA\*0201* was present in all individuals analysed and could, based on the phylogenetic analysis, represent a separate locus.

The presence of such a single monomorphic class II A locus is novel to teleost fish but is frequently observed in mammalian species. Selection, whether it be through overdominance or negative frequency dependent, acts to maintain and promotes polymorphism at the MHC. The fixation of *Clha-DAA\*0201* at a monomorphic locus may reflect a recent and strong frequency-dependent selection event. The *Clha-DAA\*0201* allele may have conferred increased resistance to a specific and common pathogen, which led to its fixation across multiple populations. Future research examining the variation at the MH class II B locus could provide valuable insight into this hypothesis.

At present, there are not sufficient numbers of allelic sequences representing a single locus to perform an analysis of peripheral benzodiazepine receptors (PBR) as has been reported for Atlantic salmon (Consuegra *et al.*, 2005). In Atlantic salmon, 17 allelic *Sasa-DAA* sequences were analysed, and it was shown that 10–14 PBRs were under positive selection. Comparison of the *Clha-DAA* and *Clha-DBA* allelic sequences separately revealed only 2–6 polymorphic sites that could potentially be under positive selection. This comparison is, however, hampered by the small number of sequences available for analyses.

The detailed analysis of intron 3 sequences allowed to design forward primers specific for each of the four intron 3 sequences. The 5' end of the four intron sequences after the TG repeat were identical, and to be able to amplify a fragment of around 100 bp, the reverse primer was designed in exon 4, which is identical in all potential class II A loci. The primer pair *Clha-DAinF10* and *Clha-DAinR1* consistently amplified a fragment of the predicted size and was shown, by genotyping a number of individuals, to amplify one or two alleles of the microsatellite marker designated *Clha-DAA-INTR3*. Genotyping MH loci using embedded microsatellite or minisatellite have been successfully used in Norwegian and Irish Atlantic salmon (Grimholt *et al.*, 2002; Stet *et al.*, 2002; Grimholt *et al.*, 2003; De Eyto *et al.*, 2007). In these fish, a one-to-one relationship was observed between the minisatellite marker and MH class II A–class II B haplotype. However, this was not the case for the MH class I embedded minisatellite marker where a single microsatellite allele could be found in several MH class I alleles. In subsequent studies of domesticated Australian populations of Atlantic salmon, the observed one-to-one relationship between microsatellite marker and MH polymorphism also seems to break down with two *Sasa-DAA-3UTR* alleles each linked to two allelic MH class II *Sasa-DAA* sequences (Wynne *et al.*, 2007). Irrespective of this, the microsatellite marker embedded in the MH class II A genes seems to be a good proxy for actual MH class II polymorphism and can facilitate the MH typing of large number of individuals.

In total, 16 *Clha-DAA-INTR3* alleles were detected in the four herring samples, and together they formed a perfect 2 bp allele ladder ranging from 109 to 139 bp. This underpins the fact that the primers used amplify a microsatellite marker from a single *Clha-DAA* locus. In addition, no departure from HWEs was observed in the four populations. The allelic richness in each sample

was extremely similar and ranged from 10 to 12.9. These values are higher than the number of *Sasa-DAA-3UTR* alleles observed in wild Canadian and domesticated Australian Atlantic salmon for which values ranged from 6.8 to 8.0 (Wynne *et al.*, 2007). Currently, no data on MH-embedded microsatellite markers are available from other abundant pelagic fish species that allows a comparison of the allelic richness of MH-embedded microsatellite markers among different species. However, it is possible to compare the MH-embedded microsatellite marker *Clha-DAA-INTR3* with other putatively neutral microsatellite markers. Pair-wise  $F_{ST}$  estimates for the *Clha-DAA-INTR3* marker range from  $-0.003$  to  $0.014$ . These values are similar or somewhat lower than pair-wise  $F_{ST}$  values between North Sea and Western Baltic samples of herring using neutral microsatellite markers (Bekkevold *et al.*, 2005; Ruzzante *et al.*, 2006). A more detailed comparison between the performance of the *Clha-DAA-INTR3* with neutral microsatellites and the allozymes to detect genetic differentiation among populations of Atlantic herring is under preparation (unpubl. data). In conclusion, the MHC class II *Clha-DAA* cDNA sequence and genomic organization have been isolated and characterized and a microsatellite marker (*Clha-DAA-INTR3*) embedded in intron 3 of this locus validated. The *Clha-DAA-INTR3* microsatellite marker is polymorphic with 16 alleles detected in the populations studied. The microsatellite shows no deviation from HWES and is able to detect genetic differentiation among populations.

This work was supported by the European Commission as part of a Framework 5 Programme HERGEN Q5RS-2001-01370 (<http://www.hull.ac.uk/hergen>). We thank E. Hatfield (FRS Marine laboratory, Aberdeen, U.K.), E. Torstensen (Institute of Marine Research, Flødevigen, Norway) and L. Clausen (Department of Marine Fisheries, Danish Institute for Fisheries Research, Charlottenlund, Denmark) for providing the herring samples, and the HERGEN consortium for valuable discussions.

## References

- Bekkevold, D., André, C., Dahlgren, T. G., Clausen, L. A., Torstensen, E., Mosegaard, H., Carvalho, G. R., Christensen, T. B., Norlinder, E. & Ruzzante, D. E. (2005). Environmental correlates of population differentiation in Atlantic herring. *Evolution* **59**, 2656–2668.
- Bingulac-Popovic, J., Figueroa, F., Sato, A., Talbot, W. S., Johnson, S. L., Gates, M., Postlethwait, J. H. & Klein, J. (1997). Mapping of Mhc class I and class II regions to different linkage groups in the zebrafish, *Danio rerio*. *Immunogenetics* **46**, 129–134.
- Consuegra, S., Megens, H. J., Leon, K., Stet, R. J. M. & Jordan, W. C. (2005). Patterns of variability at the major histocompatibility class II alpha locus in Atlantic salmon contrast with those at the class I locus. *Immunogenetics* **57**, 16–24.
- De Eyto, E., McGinnity, P., Consuegra, S., Coughlan, J., Tufto, J., Farrell, K., Megens, H. J., Jordan, W., Cross, T. & Stet, R. J. M. (2007). Natural selection acts on Atlantic salmon major histocompatibility (MH) variability in the wild. *Proceedings of the Royal Society of London B* **274**, 861–869.
- Dixon, B., van Erp, S. H., Rodrigues, P. N., Egberts, E. & Stet, R. J. M. (1995). Fish major histocompatibility complex genes: an expansion. *Developmental and Comparative Immunology* **19**, 109–133.

- Dixon, B., Nagelkerke, L. A., Sibbing, F. A., Egberts, E. & Stet, R. J. M. (1996). Evolution of MHC class II beta chain-encoding genes in the Lake Tana barbell species flock (*Barbus intermedius* complex). *Immunogenetics* **44**, 419–431.
- Ellis, S. A., Bontrop, R. E., Antczak, D. F., Ballingall, K., Davies, C. J., Kaufman, J., Kennedy, L. J., Robinson, J., Smith, D. M., Stear, M. J., Stet, R. J. M., Waller, M. J., Walter, L. & Marsh, S. G. (2006). ISAG/IUIS-VIC Comparative MHC Nomenclature Committee report, 2005. *Immunogenetics* **57**, 953–958.
- van Erp, S. H., Egberts, E. & Stet, R. J. M. (1996). Characterization of class II A and B genes in a gynogenetic carp clone. *Immunogenetics* **44**, 192–202.
- Godwin, U. B., Flores, M., Quiniou, S., Wilson, M. R., Miller, N. W., Clem, L. W. & McConnell, T. J. (2000). MHC class II A genes in the channel catfish (*Ictalurus punctatus*). *Developmental and Comparative Immunology* **24**, 609–622.
- Grimholt, U., Getahun, A., Hermsen, T. & Stet, R. J. M. (2000). The major histocompatibility class II alpha chain in salmonid fishes. *Developmental and Comparative Immunology* **24**, 751–763.
- Grimholt, U., Drablos, F., Jorgensen, S. M., Hoyheim, B. & Stet, R. J. M. (2002). The major histocompatibility class I locus in Atlantic salmon (*Salmo salar* L.): polymorphism, linkage analysis and protein modelling. *Immunogenetics* **54**, 570–581.
- Grimholt, U., Larsen, S., Nordmo, R., Midtlyng, P., Kjoeglum, S., Storset, A., Saebo, S. & Stet, R. J. M. (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.
- Hardee, J. J., Godwin, U., Benedetto, R. & McConnell, T. J. (1995). Major histocompatibility complex class II A gene polymorphism in the striped bass. *Immunogenetics* **41**, 229–238.
- Hashimoto, K., Nakanishi, T. & Kurosawa, Y. (1990). Isolation of carp genes encoding major histocompatibility complex antigens. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 6863–6867.
- Kasahara, M. (1999). Genome dynamics of the major histocompatibility complex: insights from genome paralogy. *Immunogenetics* **50**, 134–145.
- Kjoeglum, S., Larsen, S., Bakke, H. G. & Grimholt, U. (2006). How specific MHC class I and class II combinations affect disease resistance against infectious salmon anaemia in Atlantic salmon (*Salmo salar*). *Fish & Shellfish Immunology* **21**, 431–441.
- Kruiswijk, C. P., Hermsen, T., Fujiki, K., Dixon, B., Savelkoul, H. F. & Stet, R. J. M. (2004). Analysis of genomic and expressed major histocompatibility class Ia and class II genes in a hexaploid Lake Tana African ‘large’ barb individual (*Barbus intermedius*). *Immunogenetics* **55**, 770–781.
- Kruiswijk, C. P., Hermsen, T., van Heerwaarden, J., Dixon, B., Savelkoul, H. F. & Stet, R. J. M. (2005). Major histocompatibility genes in the Lake Tana African large barb species flock: evidence for complete partitioning of class II B, but not class I, genes among different species. *Immunogenetics* **56**, 894–908.
- Kumar, S., Tamura, K., Jakobsen, I. B. & Nei, M. (2001). MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**, 1244–1245.
- Kuroda, N., Figueroa, F., O’Hugin, C. & Klein, J. (2002). Evidence that the separation of Mhc class II from class I loci in the zebrafish, *Danio rerio*, occurred by translocation. *Immunogenetics* **54**, 418–430.
- 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*). *Molecular Ecology* **10**, 2525–2539.
- Marsh, S. G., Parham, P. & Barber, L. D. (2000). *The HLA Facts Book*. San Diego, CA: Academic Press.
- Miller, K. M., Kaukinen, K. H., Beacham, T. D. & Withler, R. E. (2001). Geographic heterogeneity in natural selection on an MHC locus in sockeye salmon. *Genetica* **111**, 237–257.

- Murray, B. W., Sultmann, H. & Klein, J. (1999). New family of Mhc class II A genes identified from cDNA sequences in the cichlid fish *Aulonocara hansbaenschi*. *Immunogenetics* **49**, 544–548.
- Murray, B. W., Shintani, S., Sultmann, H. & Klein, J. (2000). Major histocompatibility complex class II A genes in cichlid fishes: identification, expression, linkage relationships, and haplotype variation. *Immunogenetics* **51**, 576–586.
- Prugnolle, F., Manica, A., Charpentier, M., Guegan, J. F., Guernier, V. & Balloux, F. (2005). Pathogen-driven selection and worldwide HLA class I diversity. *Current Biology* **15**, 1022–1027.
- Reusch, T. B., Schaschl, H. & Wegner, K. M. (2004). Recent duplication and inter-locus gene conversion in major histocompatibility class II genes in a teleost, the three-spined stickleback. *Immunogenetics* **56**, 427–437.
- Rousset, F. & Raymond, M. (1995). Testing heterozygote excess and deficiency. *Genetics* **140**, 1413–1419.
- Ruzzante, D. E., Mariani, S., Bekkevold, D., André, C., Mosegaard, H., Clausen, L. A., Dahlgren, T. G., Hutchinson, W. F., Hatfield, E. M., Torstensen, E., Brigham, J., Simmonds, E. J., Laikre, L., Larsson, L. C., Stet, R. J. M., Ryman, N. & Carvalho, G. R. (2006). Biocomplexity in a highly migratory pelagic marine fish, Atlantic herring. *Proceedings of the Royal Society of London B* **273**, 1459–1464.
- Sommer, S. (2005). The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Frontiers in Zoology* **2**, 16.
- Stet, R. J. M., Dixon, B., van Erp, S. H., van Lierop, M.-J. M. C., Rodrigues, P. N. & Egberts, E. (1996). Inference of structure and function of fish major histocompatibility complex (MHC) molecules from expressed genes. *Fish & Shellfish Immunology* **6**, 305–318.
- Stet, R. J. M., de Vries, B., Mudde, K., Hermsen, T., van Heerwaarden, J., Shum, B. P. & Grimholt, U. (2002). Unique haplotypes of co-segregating major histocompatibility class II A and class II B alleles in Atlantic salmon (*Salmo salar*) give rise to diverse class II genotypes. *Immunogenetics* **54**, 320–331.
- Stet, R. J. M., Kruiswijk, C. P. & Dixon, B. (2003). Major histocompatibility lineages and immune gene function in teleost fishes: the road not taken. *Critical Reviews in Immunology* **23**, 441–471.
- Sultmann, H., Mayer, W. E., Figueroa, F., O'hUigin, C. & Klein, J. (1993). Zebrafish Mhc class II alpha chain-encoding genes: polymorphism, expression, and function. *Immunogenetics* **38**, 408–420.
- Takami, K., Zaleska-Rutczynska, Z., Figueroa, F. & Klein, J. (1997). Linkage of LMP, TAP, and RING3 with Mhc class I rather than class II genes in the zebrafish. *Journal of Immunology* **159**, 6052–6060.
- Truett, G. E., Walker, A. A., Truett, A. A., Mynatt, R. L., Heeger, P. & Warman, M. (2000). Preparation of PCR-quality DNA with hot sodium hydroxide and tris (HOTSHOT). *Biotechniques* **29**, 52–54.
- Van Oosterhout, C., Hutchinson, W. F., Mills, D. P. M. & Shipley, P. (2004). Microchecker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**, 535–538.
- Weir, B. S. & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370.
- Wynne, J. W., Cook, M. T., Holmes, B. H. & Elliott, N. G. (2007). Allelic and haplotypic diversity at the major histocompatibility class II within domesticated Australian Atlantic salmon (*Salmo salar* L.). *Journal of Fish Biology* **70**, 45–49.

### Electronic Reference

- Goudet, J. (2001). *FSTAT, a Program to Estimate and Test Gene Diversities and Fixation Indices (Version 2.9.3.2)*. Available at <http://www.unil.ch/izea/software/fstat.html>