Diversity and evolution of MHII β genes in a non-model percid species—The Eurasian perch (*Perca fluviatilis* L.)

Christian Michel [a,c], Louis Bernatchez [b], Jasminca Behrmann-Godel [c,*]

[a] Program MGU, Man-Society-Environment, Department of Environmental Sciences, University of Basel, Vesalgasse 1, CH-4051 Basel, Switzerland
[b] Insitut de Biologie Intégrative et des Systèmes (IBIS), Pavillon Charles-Eugène-Marchand, 1030, Avenue de la Médecine, Université Laval, Québec (Québec) G1V 0A6, Canada
[c] Limnological Institute, Department of Biological Sciences, University of Konstanz, 78457 Konstanz, Germany

ABSTRACT

This study provides the first investigation of the diversity, structure, and molecular evolution of MHII β genes in a non-model percid species – the Eurasian perch (*Perca fluviatilis* L.). PCR primers developed here were highly specific, and documented a high diversity of the MHII β1 domain in perch. Our results suggest a minimum of eight MHII β loci in this species – a finding congruent with several studies suggesting that many Euteleostei posses multiple MHII β loci. As for other vertebrates, both positive selection and gene-conversion contribute to the reported high allelic diversity. Similarly, the MHII β1 domain in perch exhibits a characteristic MHC fold known from other vertebrates. In addition, our results suggest some teleost specific differences of the MHII β1 domain, including: differences in chemical properties of specific amino acids in the β1 domain, the absence of the tetrapod specific glycolisation signal, and differences in the positions of some of the positively selected codons in the MHII β1 domain, which are presumably involved in antigen binding. Future studies should investigate the teleost MHII β genes in more details in order to confirm the suggested differences, and to determine the extent to which these differences prevail in different teleost lineages.

1. Introduction

The major histocompatibility complex (MHC) multigene family is functionally involved in the innate and adaptive immune response (Klein, 1986a). In all jawed vertebrates studied to date, specific genes of the MHC encode glycoprotein receptors that constitute a central component of the vertebrate immune system (Flajnik and Du Pasquier, 2004). These highly polymorphic genes consist of two major subclasses – class I and class II genes (Klein, 1986a).

The class II receptors are heterodimers, consisting of two transmembrane proteins (α and β chain), which are encoded by separate genes. They predominantly present antigens from extra-cellular pathogens, such as bacteria or parasites (Trowsdale, 1993). For antigen presentation the antigenic peptides are anchored by specific amino acid (aa) residues ('Peptide Binding Residues', PBR) located in the extra-cellular antigen recognition site (ARS), which is encoded in exon II of the respective gene (Engelhard, 1994). The antigen/receptor complexes are then transported to the cell surface, thereby the antigen is presented to CD4$^+$ bearing T-cells that finally trigger the adaptive immune response. MHC genes that dif-fer at their PBR are able to bind a diverse array of antigens, and, hence, the variability of the PBR critically influences the individual immune response (Hedrick and Kim, 2000).

Homologous genes of both MHC gene classes can be found in representatives of all gnathostomata, and, hence, the proto-MHC likely originated at least 500 million years ago (Flajnik and Kasahara, 2001). The class I and class II genes are tightly linked in humans (The MHC sequencing consortium, 1999), and most other vertebrate classes (Kelley et al., 2005). The only known exception are the teleost fish, where class I and II genes are not linked (Bingulac-Popovic et al., 1997; Malaga-Trillo et al., 1998; Hansen et al., 1999; Sato et al., 2000; Sambrook et al., 2005), and, hence, in teleosts these genes are mostly referred to as major histocompatibility genes (MH genes), rather than MH*Complex* genes (Stet et al., 2003). Furthermore, when more than one class II locus occurs in a given teleost, they can be dispersed throughout the genome and even on separate chromosomes (Sambrook et al., 2002, 2005; Phillips et al., 2003). In general, the genomes of Euteleostei are characterized by an expansion of gene families (Robinson-Rechavi et al., 2001). This trend can also be observed in regard to the MH class II loci, since their number varies substantially among different teleost species, ranging from a single MHII β locus in salmonids (Shum et al., 2001; Stet et al., 2002), to presumably four loci in sticklebacks (Reusch and Langefors, 2005) to up to13 loci in cichlids (Malaga-Trillo et al., 1998).

* Corresponding author. Tel.: +49 07531 88 4536; fax: +49 07531 88 3533.
E-mail address: jasminca.behrmann@uni-konstanz.de (J. Behrmann-Godel).

Doherty and Zinkernagel (1975) first proposed a central role of pathogen mediated positive selection for the maintenance of the MHC diversity. Later, Hughes and Nei (1988) proposed that if the MHC diversity is maintained by positive selection related to antigen recognition such selection should act specifically on the amino acids involved in antigen recognition (i.e. PBR), thereby increasing the PBR variability. Accordingly, a higher number of non-synonymous than synonymous substitutions should be found specifically in those codons encoding the PBR. They confirmed their hypothesis in a study of human and mouse MHC II genes (Hughes and Nei, 1989), and since then their findings were supported by numerous studies of species from all vertebrate classes (Klein et al., 1993; Hughes et al., 1994; Hughes and Hughes, 1995; Apanius et al., 1997; Edwards et al., 1998; Hughes and Yeager, 1998; Figueroa et al., 2000; Bernatchez and Landry, 2003).

Thus, the emerging view is that the high nucleotide diversity observed in the PBR is generated by positive selection, mediated by a host–pathogen arms race (Hedrick and Kim, 2000; Hedrick, 2002; Wegner et al., 2003; Dionne et al., 2007, 2009) Most researchers agree about the central role of positive (diversifying) selection for the maintenance of the MHC diversity (Klein et al., 1993; Hughes et al., 1994; Hughes and Hughes, 1995; Apanius et al., 1997; Bernatchez and Landry, 2003). However, the molecular mechanisms that generate new allelic variants in the MHC genes are still controversial (Klein, 1986b; Hughes et al., 1993; Ohta, 1995; Shum et al., 2001; Richman et al., 2003; Bos and Waldman, 2006; Schaschl et al., 2006). It was originally proposed that the high MHC variation originates solely from point mutations (Klein, 1986a). However, if point mutations are the sole mechanism generating the high allelic diversity observed in the MHC, then the mutation rate should be significantly higher than in other gene loci, which has not been supported by sequence analyses (Satta et al., 1993). Alternatively, repeated non-reciprocal recombination (gene-conversion) could contribute to the observed high allelic diversity (Martinsohn et al., 1999; Ohta, 1999). Evidence for gene-conversion among MHC alleles has been found in several vertebrate species, including teleosts (Högstrand and Böhme, 1994; Zangenberg et al., 1995; Reusch et al., 2004; Reusch and Langefors, 2005; Schaschl et al., 2005). These studies also indicate that both intra- and inter-locus gene-conversion contributes to the allelic diversity observed in the MHC genes (Ohta, 1995; Reusch et al., 2004; Reusch and Langefors, 2005). For example, Reusch and Langefors (2005) suggests that in stickleback *Gasterosteus aculeatus*, a teleost species with multiple MHC class II β loci, inter-locus gene-conversion contributes significantly to the allelic diversity of the MHII β1 domain. It has also been suggested that gene-conversion might be an important mechanism for generating new MHII β alleles, particularly in species where a high number of MHC loci provide a 'broad reservoir' of genetic diversity (Andersson and Mikko, 1995) – such as in teleost fish.

The evolution of the MHII β genes may differ between different vertebrate classes (Edwards et al., 1995) and possibly also among different teleost lineages (Aguilar and Garza, 2007). The current knowledge in teleosts is mostly restricted to a small number of species with only limited information available about non-model teleost species. Hence, a broader taxonomic coverage is warranted to expand our understanding of the diversification of this gene family among different teleost lineages. To this end, we investigated the diversity, evolution, and molecular structure of the MHII β genes in a non-model percid species – the Eurasian perch (*Perca fluviatilis* L.). The specific aims of this study were, firstly, to approximate the putative number of MHII β loci present in perch, and, secondly, to characterize the role of positive selection and gene-conversion for the generation of the allelic diversity. Thirdly, we compared the tertiary structure and amino-acid composition of the β1 domain in perch with other teleost species and tetrapods to investigate possible differences among vertebrate lineages.

2. Experimental

2.1. Materials and methods

2.1.1. Sample collection

To characterize the MHII β1 domain in perch, 58 specimens were sampled from 12 geographic locations of Lake Constance (Germany) with about five individuals per location. Fin clips were taken, and preserved in approximately 10 volumes of p.a. grade ethanol absolute (Riedel de Haen GmbH, Germany). For molecular analyses whole genomic DNA was isolated from a small piece of preserved fin tissue (approx. 1 mm^2) with a modified high-salt DNA extraction protocol following Aljanabi and Martinez (1997).

2.1.2. Molecular methods

To amplify a fragment spanning from intron I to the 3′ end of exon II in a highest possible number of MHII β loci, polymerase chain reaction (PCR) primers were designed in conserved regions of these genes. To identify conserved regions, cDNA sequences of five acanthopterygian species were downloaded from the GenBank database (*Stizostedion vitreum*, accession no. AY158838.1; *Cyphotilapia frontosa*, accession no. L13232.1; *Morone saxatilis*, accession no. L33962.1; *Poecilia reticulata*, accession no. Z54079.1; *Fundulus heteroclitus*, accession no. AF529586.1). These sequences were aligned in BioEdit v7.0.5.3 (Hall, 1999), and PCR primers were designed with the software Primer3 (Rozen and Skaletsky, 2000). The primers are located in conserved regions in the 3′ end of the leader peptide domain of exon I (StviMH5F), the 3′ end of the β1 domain (StviMH5R, PfluMH3R, and PfluMH4R), and the 5′ end of exon III (StviMH6R). Since *S. vitreum*, is the nearest (taxonomically) relative to *P. fluviatilis*, the initial primers (StviMH5F, StviMH5R, and StviMH6R) were based on the *S. vitreum* sequence. During the initial establishing phase, sequences obtained with StviMH5F combined with StviMH6R revealed specific nucleotide substitutions in the StviMH5R binding site. Therefore, two additional reverse primers (PfluMH3R and PfluMH4R; Table 1), specific for *P. fluviatilis*, were designed to complement StviMH5R and to increase the amount of MHII β alleles amplified. Hence, all fragments included in the data analysis cover a part of the leader peptide, the entire intron I and almost the entire exon II.

PCR reactions were performed for each of the three used primer combinations separately (Table 1). Reaction mixtures were composed of 0.3 mM each primer, 200 μM PeqGOLD dNTP-Mix (Peqlab GmbH, Germany), 1 × PCR BufferY (Peqlab), and 1.25 units peqGOLD Sawadee *Taq* DNA polymerase (Peqlab) in 25 μl reactions with approx. 50 ng template. Amplifications were performed in a Primus 96 advanced thermocycler (Peqlab) with the following PCR protocol: one cycle of 5 min initial denaturing at 94 °C followed by 35 cycles of 30 s at 94 °C, 30 s at 55 °C and 30 s at 72 °C, and 10 min of final extension at 72 °C. To check the product sizes, and to quantify the obtained products for cloning, 3 μl of each PCR product was quantified on 1% agarose gels.

Equimolar ratios of all three PCR products each obtained with the reverse primers StviMH5R, PfluMH3R and PfluMH4R were pooled prior to cloning (Addison, 2007). The PCR products were cloned into

Table 1

PCR primers for MHII β genes of *Perca fluviatilis* L. reported here; T_m = melting temperature, GC = GC content.

Primer	Primer sequence	bp	T_m [°C]	GC [%]
StviMH5F	5′ CCTGCTCTTCATCAGCCTCT 3′	20	55	55
StviMH6R	5′ CGCTGCAGACCAACATGG 3′	18	53	61
StviMH5R	5′ TTTAGTCAGAGCAGCGTGGT 3′	20	54	50
PfluMH3R	5′ TTTAGTCAGAACATTCTGGT 3′	20	46	35
PfluMH4R	5′ TTTAGAGAGGACATGGTCGT 3′	20	50	45

the pCR2.1® vector included in the TA cloning kit (Invitrogen GmbH, Germany) following the manufacturers protocol. Finally, the product/vector construct was transformed into competent One Shot® INVαF Escherichia coli cells (Invitrogen). All transformants were plated on sterile LB agar plates containing 100 μg/ml ampicillin (Sigma–Aldrich GmbH, Germany); positive clones were selected via blue-white selection.

For sequencing, insert sizes of single cell bacterial colonies were screened with vector specific M13 primers (F: 5′-ACTGGCCGTCGTTTTACA-3′; R: 5′-CAGGAAACAGCTATGACC-3′). A single colony scratch was added to a reaction mixture of 0.1 mM each primer, 200 μM PeqGOLD dNTP-Mix (Peqlab), 1.5 mM MgCl$_2$, 1 × PCR Buffer Y (Peqlab), and 1.25 units peqGOLD Sawadee Taq DNA polymerase (Peqlab) in 20 μl reaction volume. All screening reactions were performed in a Primus 96 advanced thermocycler (Peqlab) with the following profile: one cycle of 4 min initial denaturing at 94 °C followed by 27 cycles of 30 s at 94 °C, 30 s at 45 °C and 1 min at 72 °C, and 4 min of final extension at 72 °C. To check the insert sizes, and to quantify the obtained products for subsequent sequencing, 3 μl of each M13-PCR product were quantified on 1% agarose gels. Prior to sequencing, PCR products were cleaned from excess primers and dNTPs by enzymatic degradation with ExoSAP-IT (GE Healthcare, Germany). A mean ± SD of 8.12 ± 7.79 (min = 4, max = 47) clones per individual were submitted for sequencing on an ABI3730 automatic sequencer (4base lab Reutlingen, Germany). All initial sequences were obtained with the M13F primer, additionally all sequences with ambiguous base calls were sequenced from both directions.

2.1.3. Data analyses

2.1.3.1. Processing of sequence files. All sequences were aligned in BioEdit v7.0.5.3 (Hall, 1999) initially with the ClustalW multiple alignment implementation and then unambiguously by eye. For subsequent analysis of the MHII β1 domain all sequences were cropped to 247 base pairs (bp) of exon II starting from aa residue 6 of the mature protein as inferred from Ono et al. (1992). This fragment covers almost the entire β1 domain, with less than 24 bp (primer length plus additional 4 bp) missing at the 3′ end, and contains 13 PBR positions known from humans (Brown et al., 1993). Only sequences that were found in independent PCR reactions from at least two individuals were included in the final dataset. This approach is sufficiently conservative to account for in vitro recombinants or possible substitution errors incorporated by the Taq DNA polymerase (Zylstra et al., 1998). As we cannot unambiguously distinguish alleles from a single locus from alleles originating from different loci, tentative designations based on the MHC nomenclature suggested by Klein et al. (1990) are used. Thus, the MHII β1 alleles are named *MhcPefu*-DXB* followed by the individual allele number (01-28).

2.1.3.2. Genetic diversity and identification of putative MHII β gene loci. Phylogenetic reconstruction was used to, firstly, illustrate the genetic diversity of the MHII β1 domain and intron I in perch, and, secondly, to infer clusters of highly similar sequences. For the exon II sequences a neighbour-joining dendrogram (Saitou and Nei, 1987) was constructed using Mega v4.0 (Tamura et al., 2007) with pairwise distances calculated as p-distance. Gaps were deleted in a pairwise manner, and consistency of clustering was tested by 10,000 bootstrap replicates. The exon dendrogram was rooted with a MHII β1 domain sequence of the axolotl Ambystoma mexicanum (Genbank accession no. EF585232).

Pairwise p-distance matrices were inferred in Mega v4.0 for our dataset of MHII β1 sequences, and for a published β1 domain dataset of the cichlid species *Pseudotropheus fainzilberi* (Blais et al., 2007; N = 219). We selected this dataset because (i) it was obtained

with a similar cloning procedure, (ii) P. fainzilberi is taxonomically the most closely related species to P. fluviatilis among available data and (iii) cichlids have a high number of MHII loci (up to 13; Malaga-Trillo et al., 1998). To compare the diversity in our dataset also with the diversity in a single MHII β locus in teleosts, the β1 domain sequences of the DAB locus of Salmo salar available in the Immuno-Polymorphism-Database (http://www.ebi.ac.uk/ipd; N = 37) were used to obtain a pairwise p-distance matrix. The average p-distances and their standard deviations (SD) were calculated from the pairwise distance matrices. To compare the diversity patterns observed in P. fluviatilis, P. fainzilberi, and S. salar the frequency distributions of the pairwise p-distance matrices were plotted as histograms. The p-distance was chosen for easy comparison with previous studies. Finally, in the exon II dendrogram, clusters of highly similar sequences were defined according to three criteria (i) bootstrap support >75%, (ii) separated from any other cluster by a mean p-distance >0.06, and (iii) mean within cluster p-distance ≤0.06. This threshold was chosen because it represents the magnitude of diversity observed in the β1 domain of the single MHII β locus in *Salmo salar* ($p_m = 0.06 \pm 0.02$, see above). Also, this threshold is conservative; for instance it is equivalent to the maximum value observed in wild salmonid populations (Miller and Withler, 1996, $p_m = 0.06$; Miller et al., 1997, $p_m = 0.03$; Kim et al., 1999, $p_m = 0.02$).

To complement the phylogenetic analysis of the exon II sequences, and to attempt approximating further the putative number of MHII β gene loci present in perch, we used phylogenetic reconstruction of the intron I sequences. Since intron I consisted of multiple repeats (see Section 3 for details) only the last complete 77 bp repeat, the incomplete 77 bp repeat and the adjacent intron/exon boundary was used for phylogenetic reconstruction. To improve the alignment of the intron sequences twenty-nine alignment gaps were introduced. To include the information contained in these gaps they were recoded as parsimony informative sites (Lee, 2001), using modified complex indel coding (MCIC) as implemented in SeqState v1.4 (Müller, 2005). This method was recommended by Simmons et al. (2007) for unambiguously aligned gaps in parsimony based molecular phylogenetic analysis. Therefore, six recoded intron state characters together with 34 parsimony informative point substitutions were included in the maximum parsimony (MP) analysis. The MP analysis was performed in Paup* v4.0b10 (Swofford, 2000) using the parsimony ratchet algorithm to calculate BS (Bremer support) as implemented in PRAP v2.0b3 (Müller, 2004). Robustness of clustering was tested via 1000 bootstrap replicates.

2.1.3.3. Molecular evolution. The model selection procedure implemented in the Datamonkey webserver (Pond and Frost, 2005; http://www.datamonkey.org) identified the Felstenstein-81 model as the best fitting nucleotide substitution model (AIC = 4,488.15), and, hence, this model was used in further analyses. To test for global positive selection the Parris method (Scheffler et al., 2006) as implemented in the Datamonkey web server was applied. The Parris method infers positive selection acting on nucleotide sequences by comparing a null-model (M1, no-selection) where dN and dS are free to vary, but where dN is restrained to be ≤dS, against a full model (M2, selection) where dN and dS are free to vary (i.e. dN is allowed to be >dS). The models M1 and M2 are nested, and, hence, a likelihood ratio test (LRT) based on the chi^2-distribution with $N - 1$ degrees of freedom was used to test whether allowing dN > dS significantly improved the fit of the model.

Individual codons evolving under positive selection were identified with the random effects likelihood (REL) approach implemented in the Datamonkey web server (Kosakovsky Pond and Frost, 2005). In the analysis, dN and dS are free to vary, and the dN and dS values assigned to each codon are drawn from a bivariate general

discrete distribution with three d*N* and three d*S* rate categories. The model fits the parameters of this bivariate general distribution of rates among codons to the dataset yielding a total of nine possible d*N*/d*S* rate classes for each site. To identify codons that evolve under positive selection an empirical Bayes approach based on the maximum likelihood estimates of the rate parameters is applied (Kosakovsky Pond and Frost, 2005). Finally, we used the Bayes factor (BF, cf. Kass and Raftery, 1995) for the event d*N* > d*S* as a criterion to identify sites that evolve under positive selection. The BF is defined as the ratio of the posterior odds for the event d*N* > d*S* to the prior odds for that event and a BF > 30 provides very strong evidence that d*N* > d*S* for the respective codon position (cf. Kass and Raftery, 1995). Hence, we considered a codon to evolve under positive selection when the BF for the event d*N* > d*S* exceeded 30. Simulation studies have shown that in phylogenetic REL methods 1/BF is approximately equivalent to a standard *P*-value (Poon et al., 2009). Hence, our threshold of 30 is approximately equivalent to a standard *P*-value of 0.03, thereby being a conservative criterion for a site to evolve under positive selection.

To analyse the nucleotide polymorphism within exon II and between exon II and intron I sequences we analysed the nucleotide polymorphism overall and also separately for the putative MHII β gene loci A, B, C and E (as shown in the phylogeny Fig. 2B, excluding the two pseudogenes). We restricted the analysis to these four clusters because these contained at least three sequences. Nucleotide polymorphism was calculated for (i) all codons of the β1 domain, (ii) the positively selected sites identified in the REL analysis, and (iii) the intron I sequences all in MEGA v4.0 (Tamura et al., 2007) under the Jukes-Cantor model. Calculations for the exon II followed the method of Nei and Gojobori (1986). Standard errors (SE) were estimated via 10,000 bootstrap replications. Although this method can lead to erroneous values in case of small values for d*N* and d*S* it still provides the best descriptor for the SE for this type of analysis (Hughes, 1999). We tested for positive selection with the codon based *Z*-test implemented in MEGA v4.0 (Tamura et al., 2007) by computing the number of non-synonymous substitutions per non-synonymous site (d*N*) with the number of synonymous substitutions per synonymous site (d*S*).

2.1.3.4. Gene-conversion in the MHII β1 domain. Gene-conversion events in the MHII β1 domain were detected with GENECONV v1.81 (Sawyer, 1999). Its algorithm can be applied in multilocus datasets, and has been reliable in datasets with strong substitution rate variation (Posada and Crandall, 2001; Posada, 2002). During analysis GENECONV scans for significant clustering of shared substitutions, which indicate a mosaic structure such as created by a gene-conversion event. Significance of each cluster was tested via 10,000 non-parametric permutations against the null-hypothesis of randomly distributed substitutions. Finally a global *P*-value that is adjusted to the number of comparisons is given, and a multiple comparisons corrected *P*-value for each pairwise gene-conversion event is calculated.

2.1.3.5. Protein structure. To identify structurally important and conserved aa residues in the β1 domain an aa alignment was created from the aligned nucleotide sequences with the reading frame inferred following Ono et al. (1992). To compare the aa structure of the β1 domain of the Eurasian perch with other teleost fish, a 50% majority rule consensus sequence was created in BIOEDIT v7.0.5.3 (Hall, 1999). All exon II sequences were visually aligned in BIOEDIT v7.0.5.3 (Hall, 1999) with β1 domain aa sequences of MHII alleles of seven teleost species that were obtained from the SwissProt database (http://expasy.org/sprot). Conserved aa residues involved in important structural and functional features were identified following Kaufman et al. (1994) and were complemented with a recent analysis of the human HLA system (Reche and Reinherz, 2003). For

tertiary structure prediction, the β1 protein sequence of the most common allele (*MhcPefu*-DXB*01) was submitted to the 3D-jigsaw server v2.0 (http://bmm.cancerresearchuk.org/∼3djigsaw/). The 3D-jigsaw server reconstructs the tertiary structure of unknown proteins in reference to the most similar protein with known tertiary structure deposited in the Protein Data Bank (www.pdb.org).

3. Results

3.1. Processing of sequence files

Intron I and most of the β1 domain of the MHII β genes of 58 specimens were amplified by combining one specific forward primer with one of three specific reverse primers (Table 1). Except for PfluMH4R, the other two reverse primers amplified multiple bands ranging in size from 300 bp to 700 bp. Our cloning and sequencing procedure detected sequences in the size range of all the bands (data not shown). A mean ± SD of 8.12 ± 7.79 clones (min = 4, max = 47) were sequenced per individual, and a total of 471 sequences were obtained. These sequences represented 61 different MHII β1 domain sequences and 61 different intron I sequences.

A close inspection of the obtained sequences by eye indicated that intron I mainly consisted of a tandem array, comprising a 77 bp DNA fragment, occurring in different numbers (1–7 repeats) with slight sequence variation (Fig. I supplementary material). A putative binding site for StviMH5F was found at the beginning of every repeat, indicating multiple binding sites for StviMH5F in intron I. Thus, the observed length variation and the differences in repeat numbers are partly attributable to amplification artefacts. Two intron I sequences occurred with a 130 bp indel in one of the repeats (Fig. I, supplementary material).

3.2. Genetic diversity

From the 61 β1 domain sequences identified, 35 were found in only one individual. However, two of these alleles, *MhcPefu*-DXB*27 and *MhcPefu*-DXB*28, were by now verified as proper alleles via sequencing of SSCP bands amplified from additional individuals (data not shown). Thus, after elimination of 33 sequences for further analyses, the final dataset consisted of 28 different β1 domain sequences (Fig. 1). A mean of 3.9 ± 1.4 (SD) alleles per individual were found with a maximum of 10 alleles found in a single individual, (Table I, supplementary material). This is comparable in magnitude to the number of alleles found in two cichlid species (2.31), for which 5.74 ± 2.22 SD clones per individual were sequenced (Blais et al., 2007). One allele (*MhcPefu*-DXB*07), with a five bp frame-shift mutation in the 3′-end of the β1 domain, most likely represented a non-functional pseudogene. Moreover, we found one allele (*MhcPefu*-DXB*28) with a nine bp insertion in the 3′-end of the β1 domain (Fig. 1). This mutation is located in the α-helical part of the mature protein, likely affecting the functionality of the mature protein, and, hence, we also considered this allele as non-functional.

From the 61 intron I sequences identified, 31 were found in only one individual. Thus after elimination of these 31 sequences, the final intron I dataset consisted of 30 sequences (Fig. I, supplementary material). All nucleotide sequence data reported here is available in the GenBank database (GenBank accession nos. FN 293111–27, 29–31, 34, 37, 39–42, 45–50, 52–55).

The phylogenetic reconstruction of the MHII β1 domain indicated substantial sub-structuring in our dataset (Fig. 2A). In general, the reconstructed dendrogram exhibited higher levels of bootstrap support (BS) towards the terminal nodes. The average mean *p*-distance in the entire exon II dataset was $p_m = 0.19 \pm 0.01$. The majority of the exon II alleles (23 out of 28; 82%) could be assigned to eight distinct clusters consisting of two to four sequences based
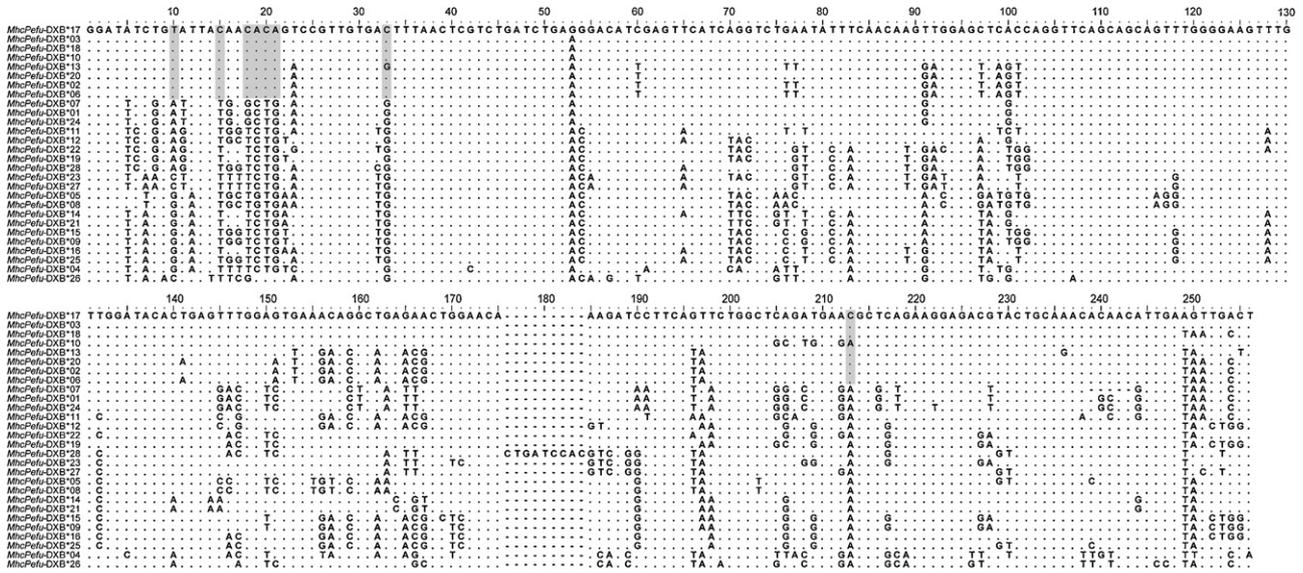
Fig. 1. Alignment of MHII β1 domain sequences in *Perca fluviatlis* L. Sequence names refer to different exon II alleles found (*MhcPefu*-DXB-01-28); dots = identical sequence, bars = alignment gaps, grey blocks = fixed nucleotide substitutions that separate group 1 sequences from all others (for details see text).
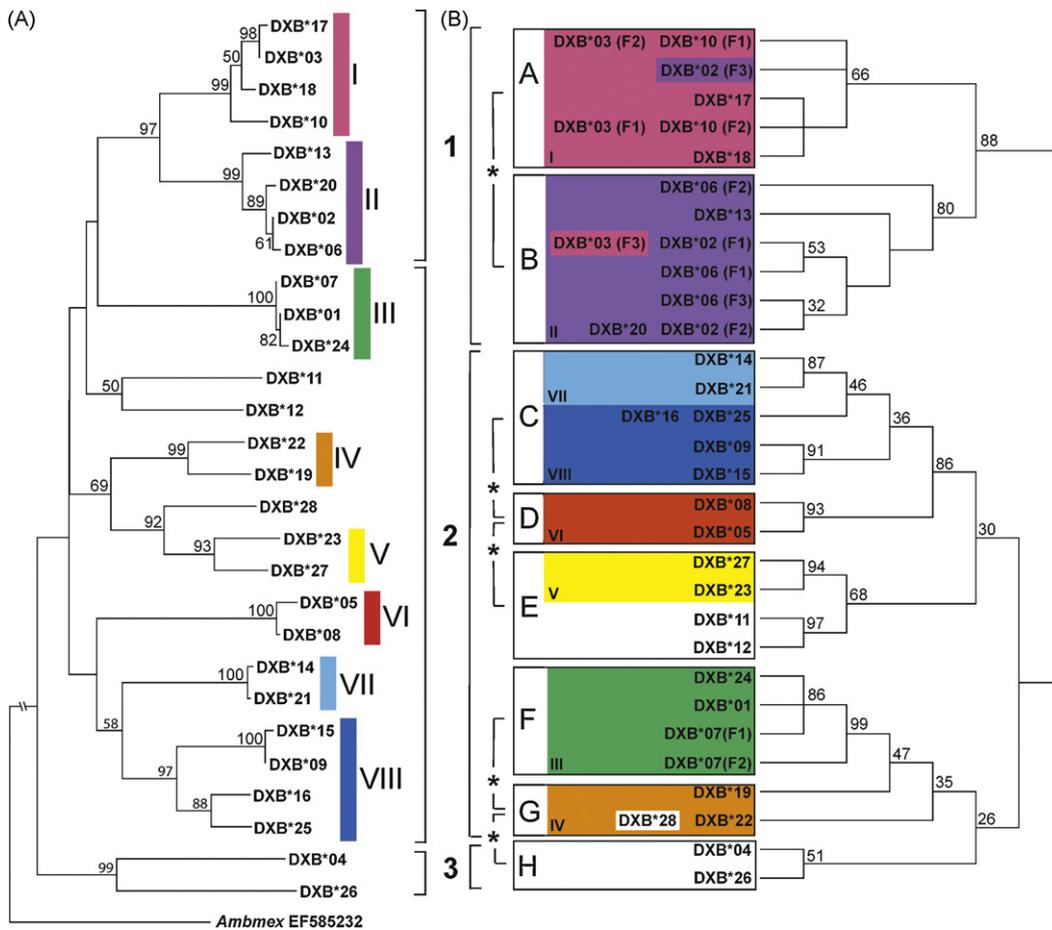
Fig. 2. Phylogenetic trees of MHII exon II (β1 domain) sequences (A) and intron I sequences (B) from *Perca fluviatlis* L. Sequence names refer to different exon II alleles found (*MhcPefu*-DXB*01-28) Brackets define sequence groups 1, 2 and 3, on the basis that sequences belonging to these groups were successfully amplified with three different reverse primers (see text for details) A: NJ tree, *p*-distance, 10,000 BS. Root: *Ambystoma mexicanum*. Colored bars (I–VIII) represent clusters with high bootstrap support (>75%) B: MP dendrogram of single nucleotide polymorphisms and indel patterns for intron I sequences, 1000 BS. Colored bars refer to clusters of the exon II phylogeny. If more than one sequence name appears at the tip of the branches, all these different exon II sequences share the identical intron I sequence. Extensions (F1, F2 and F3) indicate that identical exon II sequences were found with different intron I sequences. Black boxes around sequence clusters A–H indicate different MHII loci. Stars indicate that the locus status of the respective clusters (A vers. B, C vers. D, D vers. E, F vers. G and G vers. H) was unambiguously confirmed by individual allele patterns (see text for further details).

on criteria defined in Section 2.1 (Fig. 2A, clusters I to VIII). Each of these clusters was separated from any other cluster by a p_m of at least 0.09; no more than two alleles in any of these clusters were found in a single individual. The only two alleles amplified with the reverse primer PfluMH4R group in a basal position of the dendrogram, albeit being highly divergent (Fig. 2A, group 3, within group $p_m = 0.18$). Similarly, all exon II alleles amplified with the reverse primer PfluMH3R formed a monophyletic group in the upper distal part of the dendrogram (Fig. 2A, group 1, BS = 97%, within group $p_m = 0.07 \pm 0.01$). Group 1 mainly consisted of two distinct clusters (I and II, each with BS = 99% each), with four sequences in each cluster. The alleles in group 1 were separated from all other alleles by seven fixed nucleotide substitutions, which separate group 1 from all other alleles (Fig. 1). The remaining sequences, amplified with the reverse primer StviMH5R, did not form a monophyletic group (Fig. 2A, referred to as group 2). The phylogenetic structure in that group was more heterogeneous, which translated into a high average p-distance ($p_m = 0.18 \pm 0.02$). Three sequences in group 2 (*MhcPefu*-DXB*11, *MhcPefu*-DXB*12 and *MhcPefu*-DXB*28) could not be attributed to a cluster based on our criteria.

The histograms created from the pairwise p-distance matrices suggested a similar frequency distribution in the two perciform species *P. fluviatilis* and *P. fainzilberi*, which were both clearly distinct from the distribution observed at the single MHII β locus in *Salmo salar* (Fig. 3).

Overall, the dendrogram inferred with the intron I sequences was comparable to the exon II dendrogram (Fig. 2B). Most clusters with a bootstrap support >75% in the exon II dendrogram were also well supported in the intron I phylogeny (including clusters II/B, III/F, V/E, VI/D, VII/C and VIII/C, Fig. 2). Only three clusters with high bootstrap support in the exon II phylogeny (clusters I, IV, and group 3) had either an intermediate bootstrap support in the intron I dendrogram (cluster I/A and group 3) or were not well supported (cluster IV/G). However, sequences *MhcPefu*-DXB*11 and *MhcPefu*-DXB*12 that clustered with a low bootstrap support in the exon II phylogeny formed a well supported cluster (97% BS) in the intron I dendrogram. Only the pseudogene *MhcPefu*-DXB*28 that clusters with cluster V in the exon II phylogeny was attributed to a different cluster (cluster G) in the intron I tree.

Similar to exon II, the intron I sequences amplified with the reverse primer PfluMH3R formed a distinct group (group 1, BS = 89%), and this group was also split into two distinct clusters A and B. Interestingly, we found that identical exon II sequences could have different intron I sequences (compare clusters I and II from Fig. 2A with clusters A and B form Fig. 2B), including sequences *MhcPefu*-DXB*02, *MhcPefu*-DXB*03, *MhcPefu*-DXB*06, *MhcPefu*-DXB*07 and *MhcPefu*-DXB*10. The two exon sequences *MhcPefu*-DXB*03 and *MhcPefu*-DXB*02 occurred with three different intron I sequences (F1, F2, F3), which in both cases originated from the group 1 clusters A and B. Exon II sequences *MhcPefu*-DXB*10 and *MhcPefu*-DXB*07 appeared with two (F1, F2), *MhcPefu*-DXB*06 with three (F1, F2, F3) different intron I sequences but stemming from the same cluster (cluster A for *MhcPefu*-DXB*10, cluster B for *MhcPefu*-DXB*06 and cluster F for *MhcPefu*-DXB*07). However, more often different exon II alleles were found to share identical intron I sequences (Fig. 2B).

3.3. Identification of putative MHII β gene loci

In the phylogenetic analysis all intron I sequences were assigned to eight clusters (Fig. 2B, clusters A–H). For the following reasons, we hypothesize that these clusters suggest at least eight different MHII β loci for perch. Firstly, a high bootstrap support (BS > 75%) in both phylogenetic trees (Fig. 2, A and B), and, secondly the individual allele patterns (Table I, supplementary material). Accordingly, a
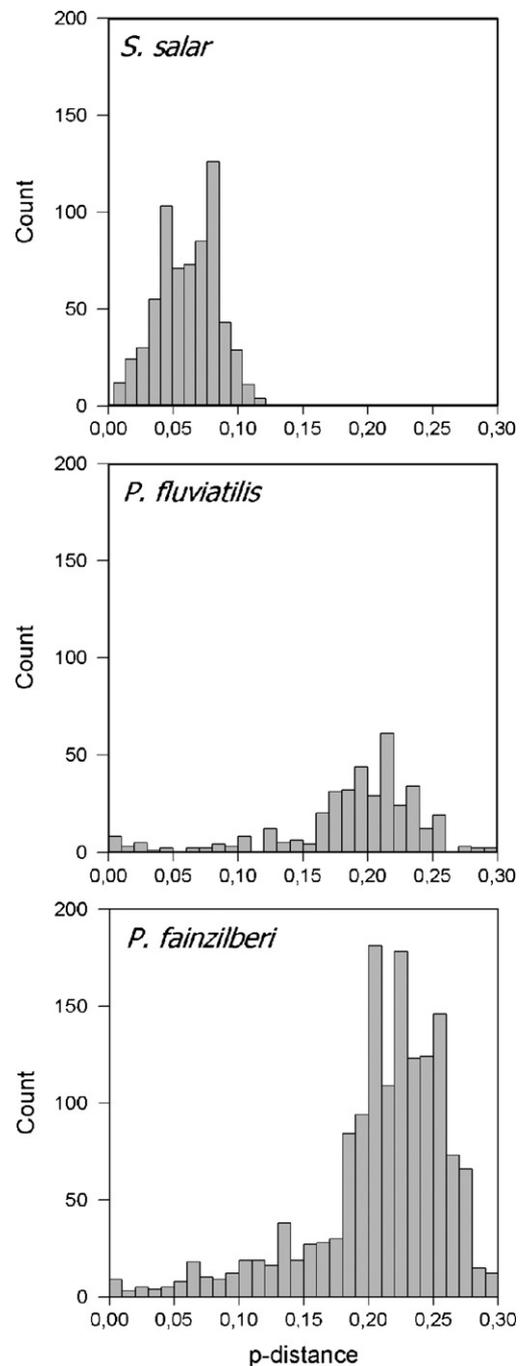
Fig. 3. Frequency distributions of pairwise p-distances inferred for all MHII β1 domain sequences of *Perca fluviatilis* L. and in published datasets from *Pseudotropheus fainzilberi* (Blais et al., 2007) and *Salmo salar* (http://www.ebi.ac.uk/ipd).

well supported cluster likely represents two loci if more than two alleles from that cluster can be found in a single individual.

More specifically, sequences from cluster A/B (group1) were all amplified with the reverse primer PfluMH3R. This group is separated from all other sequences by a bootstrap support of 97% in the exon II and 88% in the intron I dendrogram. However, clusters A and B most likely represent two different loci because: (i) the intron I sequences from cluster A shared a 33 bp deletion and a 91 bp insertion clearly separating them from all other intron I sequences (Fig. I, supplementary material), (ii) both clusters are well defined by a bootstrap support of 99% in the exon II phylogeny and 66% resp. 80% in the intron I phylogeny (compare Fig. 2A and B), (iii) based on the individual allele patterns, we found numerous individuals with

two alleles from cluster A and another one or two from cluster B (Table I, supplementary material).

The sequences in clusters C, D, E. F and G (group 2) were all amplified with the reverse primer StviMH5F. Clusters C/D are grouped with a BS of 86% in the intron I phylogeny and consist of sequences from the well supported clusters VI, VII and VIII (BS 100% and 97%, respectively, see Fig. 2A). However, the clusters C and D most likely also represent at least two different loci due to individual allele patterns (four individuals carry more than two alleles from clusters C and D while no individual carries more than two alleles from either cluster C or D, see Table I, supplementary material). Cluster E has a medium BS of 68% in the intron I phylogeny, it comprises sequences from the good supported cluster V of the exon II phylogeny and the sequences *MhcPefu*-DXB*11 and *MhcPefu*-DXB*12. Because cluster V sequences appear far apart in the exon II phylogeny, this cluster may also represent two loci, however, no individual was found to have more than two alleles from cluster E. Cluster F is well supported in both phylogenies and no individual carries more than two alleles from this cluster. Cluster G is not well supported in the intron I phylogeny but is composed of one well supported cluster (IV) from the exon II phylogeny with a BS of 99% and includes sequence *MhcPefu*-DXB*28. The remaining locus is represented by cluster H (50% BS). It contains the two sequences *MhcPefu*-DXB*04 and *MhcPefu*-DXB*26. These are the only two sequences that are amplified with the reverse primer PfluMH4R and form a well supported cluster in the exon II phylogeny (BS 99%).

The hypothesis of at least eight MHII β loci in perch is also compatible with the maximum number of 10 alleles (minimum of 5 maximum of 10 loci supposing complete hetero- or homozygosity, respectively) that we have found in a single individual (Table I, supplementary material, individual J1).

3.4. Molecular evolution

Allowing $dN > dS$ in the full model M2 revealed clear evidence for global positive selection in the exon II dataset since this model had a significantly increased likelihood (LRT = 29.04, d.f. = 2, $P < 0.001$) compared to the null model M1 (i.e. constraint $dN \leq dS$). In addition, the REL analysis identified 17 codons in the MHII β1 domain with a Bayes factor >30, and a posterior probability of at least 0.95 for the event $dN > dS$ (Table 2).

The sequence clusters A, B, C and E resolved from the intron I phylogeny were used for comparison of nucleotide polymorphism within exon II and between exon II and intron I sequences (Table 3). The magnitude of replacement mutations (dN) in the positively selected codons (identified in the REL analysis) that are putatively involved in peptide binding was particularly high and reached a value of $dN = 1.207 \pm 0.217$ in sequence cluster E. A significant excess of dN over dS in the positively selected codons was found in three of the four clusters but only in one cluster considering all codons. The nucleotide polymorphism (d) in intron I was on average six-fold lower than the synonymous nucleotide polymorphism (dS) in exon II (Table 3).

The substitution model implemented in GENECONV provided some evidence for gene-conversion ($P < 0.001$). In addition to this significant global test, 13 pairwise gene-conversion events were detected at a multiple comparisons corrected significance level of $P < 0.05$ (Table 4). In addition 148 pairwise gene-conversion events were detected that were significant prior to the multiple comparisons correction. Out of the 28 sequences compared 14 (50%) were involved in at least one pairwise gene-conversion event (Table 4). All gene-conversion events were detected between different clusters in the dendrograms (Table 4), which was also reflected in high pairwise p-distances between alleles that were involved (mean = 0.18 ± 0.02 (SD), min = 0.15, max = 0.20).

Table 2
REL analysis showing positively selected sites in the MHII β1 domain of *Perca fluviatilis* L. BF = Bayes factor for the event $dN > dS$, 1/BF = approximately equivalent to a standard *P*-value (Poon et al., 2009). PP = posterior probability for the event $dN > dS$. C = cichlid (*Pseudotropheus fainzilberi* and *P. emmiltos*, Blais et al., 2007), cross indicates that this position is positively selected in cichlids (see text), H = human, cross indicates that this position agrees with human PBR (see text), Number = indicates position of positively selected residue close to positively selected residue in *Perca fluviatilis*.

Codon	E [dS]	E [dN]	E [dN−dS]	PP	BF	1/BF	C	H
9	0.342	5.636	5.295	0.99	>100	0.005	×	×
11	0.733	6.336	5.603	1.00	>100	0.000	×	×
13	0.248	6.326	6.078	1.00	>100	0.000	×	×
16	0.252	1.804	1.552	0.98	>50	0.014	×	#
29	0.516	6.104	5.587	0.99	>100	0.006	×	×
31	0.351	6.334	5.984	1.00	>100	0.000	×	×
36	0.353	2.891	2.537	0.96	36	0.028		38
39	0.386	6.332	5.946	1.00	>100	0.000	×	×
56	0.326	1.857	1.530	0.96	34	0.030		
61	0.236	6.220	5.983	1.00	>100	0.000	×	59
66	0.194	1.798	1.604	1.00	>100	0.002	65	
68	0.413	6.334	5.921	1.00	>100	0.000	×	
71	1.202	6.320	5.118	1.00	>100	0.002	×	
72	0.336	1.839	1.503	0.96	31	0.032	×	×
73	0.303	1.876	1.573	0.97	41	0.025	75	75
79	0.348	2.142	1.794	0.95	31	0.032	78	×
87	1.647	6.331	4.685	1.00	>100	0.001	×	×

PBR in human HLA-DQ β1 domain according to Reche and Reinherz (2003).

3.5. Protein structure

The predicted tertiary structure (Fig. 4) suggests a characteristic 'MHC-fold' of the β1 domain in perch that agrees with the structure of the human β1 subunit (Brown et al., 1993). The modelled structure indicated a floor of three β-sheets that were flanked by a wall of two α-helices. Similar to the human HLA-DR β1 subunit, the majority of the positively selected residues were located in the β-sheets, which form the floor of the ARS, or in the flanking α-helices (Fig. 4). The alignment of the β1 domain protein sequences with other teleost sequences is shown in Fig. 5. All polymorphic residues known to be involved in antigen recognition in humans (Brown et al., 1993) were polymorphic among all species included. The protein alignment indicated that most structurally important aa residues that are conserved in vertebrates were also conserved
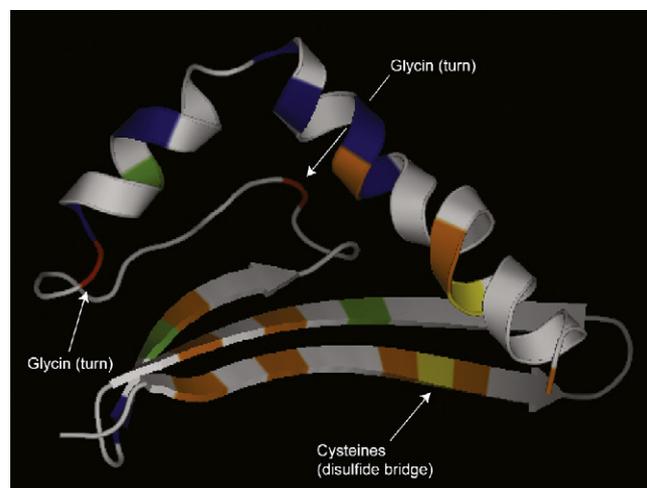
Fig. 4. Predicted tertiary structure of the MHII β1 domain in *Perca fluviatilis* L. Colors: blue = residue under positive selection in *P. fluviatilis*, orange = residue under positive selection in *P. fluviatilis* that agrees with human PBR, green = human PBR not positively selected in *P. fluviatilis*, yellow = cysteine residues forming disulfide bridge, red = glycins in turns. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Table 3
Nucleotide polymorphism of the MHII β1 domain and intron I in *Perca fluviatilis* L. Seq. clusters = A, B, C and E refer to putative loci as identified in the intron I phylogenetic analysis; REL = positively selected codons identified in the REL analysis; dN, dS and d values were inferred in MEGA v4.0 with the method of Nei and Gojobori (1986) with the Jukes Cantor 69 model. The Z-test implemented in MEGA v4.0 was used to test for an excess of dN over dS; statistical significance: ***P < 0.001; **0.001 < P < 0.01; *0.01 < P < 0.05.

Seq. cluster	Gene region	Codon type	dN	dS (d introns)
A	Exon 2 (N = 4)	All codons	0.029 ± 0.011**	0.000 ± 0.000
	Exon 2 (N = 4)	REL	0.098 ± 0.051*	0.000 ± 0.000
	Intron 1 (N = 5)	All codons		0.008 ± 0.005
B	Exon 2 (N = 4)	All codons	0.018 ± 0.008	0.023 ± 0.017
	Exon 2 (N = 4)	REL	0.066 ± 0.034	0.050 ± 0.059
	Intron 1 (N = 6)	All codons		0.013 ± 0.006
C	Exon 2 (N = 6)	All codons	0.108 ± 0.023	0.111 ± 0.037
	Exon 2 (N = 6)	REL	0.370 ± 0.120***	0.031 ± 0.026
	Intron 1 (N = 5)	All codons		0.021 ± 0.009
E	Exon 2 (N = 4)	All codons	0.224 ± 0.038	0.137 ± 0.041
	Exon 2 (N = 4)	REL	1.207 ± 0.217***	0.085 ± 0.053
	Intron 1 (N = 4)	All codons		0.049 ± 0.017
Over all mean[a]	Exon 2 (N = 26)	All codons	0.223 ± 0.032	0.213 ± 0.038
Over all mean[a]	Exon 2 (N = 26)	REL	0.942 ± 0.117***	0.213 ± 0.079
Over all mean	Intron 1 (N = 30)	All codons		0.071 ± 0.014

[a] Including all clusters; excluding pseudogenes.

Table 4
Detailed results of the GENECONV analysis showing gene-conversion events between MHII β1 domain sequences from *Perca fluviatilis* L. Listed are only global inner fragments, Sim P = multiple comparisons corrected simulated P-value; p-dist = p-distance between involved alleles, Begin/End = nucleotide position of begin and end of inferred fragment; Length = length of inferred fragment in bp; Poly = Number of polymorphic sites in the fragment, Dif = nucleotide differences within fragment, Difs = total nucleotide differences between both sequences, Cluster = gene-conversion events between intron clusters defined in Fig. 2B.

MHII β alleles [MhcPefu-]	Sim P	p-dist.	Begin	End	Length	Poly	Dif	Difs	Cluster
DXB*08/DXB*02	0.022	0.194	168	250	83	40	3	48	D/A(B)
DXB*08/DXB*06	0.016	0.198	168	250	83	40	3	49	D/B
DXB*08/DXB*20	0.016	0.198	168	250	83	40	3	49	D/B
DXB*19/DXB*03	0.022	0.186	152	196	45	23	0	46	G/A(B)
DXB*19/DXB*10	0.044	0.178	152	196	45	23	0	44	G/A
DXB*19/DXB*15	0.042	0.146	207	256	50	29	0	36	G/C
DXB*19/DXB*17	0.014	0.190	152	196	45	23	0	47	G/A
DXB*19/DXB*18	0.032	0.182	152	196	45	23	0	45	G/A
DXB*22/DXB*03	0.038	0.186	152	195	44	22	0	46	G/A(B)
DXB*22/DXB*17	0.029	0.190	152	195	44	22	0	47	G/A
DXB*22/DXB*18	0.038	0.186	152	195	44	22	0	46	G/A
DXB*25/DXB*05	0.025	0.166	210	256	47	26	0	41	C/D
DXB*28/DXB*11	0.001	0.194	1	75	75	32	1	49	G/E

in the β1 domain of perch. However, there appear to be noteworthy differences in the chemical properties of the amino acids in some of these positions. To avoid redundancy in repeating these results the detailed interpretation of the individual aa positions is given in the discussion below.

4. Discussion

4.1. Genetic diversity of the β1 domain

This study provides evidence for the occurrence of multiple MHII β loci in perch. This was namely reflected in a three- to eight-fold higher p_m in our dataset (0.19 ± 0.01) as compared to the variation found in population samples of the β1 domain in salmonids, which posses a single MHII β locus (Miller and Withler, 1996: p_m = 0.06; Miller et al., 1997: p_m = 0.03; Kim et al., 1999: p_m = 0.02). Moreover, the phylogenetic analysis of the exon II sequences identified eight distinct clusters (Fig. 2A, clusters I–VIII) that had a within cluster $p_m \leq 0.06$, compared to a separation between any two clusters $p_m \geq 0.10$. Some researchers interpreted phylogenetic clusters in multilocus datasets of MHII β1 domain sequences as homologous allelic lineages (Dixon et al., 1996), and the presence of similar clustering of some alleles in both phylogenetic analyses (Fig. 2) in our study might support this to some degree. However, our results also showed that the clustering can be different when comparing

exon II and intron I phylogenies. This finding agrees with Reusch and Langefors (2005) who warranted that paralogous β1 domain alleles can cluster together with high bootstrap support in bifurcating phylogenetic trees, especially when gene-conversion is present in a dataset (Edwards et al., 1998). In such cases the putative number of loci can be further approached via phylogenetic analysis of intron sequences (Reusch and Langefors, 2005) and visual inspection of the sequences for locus specific nucleotide substitutions that can distinguish paralogous alleles (Parham et al., 1989).

Based on the combined information from exon II and intron I sequences, we hypothesize that there are at least eight different MHII β loci in perch. Admittedly, however, more detailed investigations will be required in order to determine the exact number of MHII β gene loci in perch, as well as to understand how many of these loci are actually expressed. Indeed, the genes of the major histocompatibility complex are thought to evolve according to a 'birth-and-death' like process (Nei and Rooney, 2005), and the occurrence of pseudogenes is an inherent part of the major histocompatibility multigene family (McConnell et al., 1998; Gu and Nei, 1999). Here, we found two putatively non-functional alleles (*MhcPefu*-DXB*07 and *MhcPefu*-DXB*28), indicating the presence of at least one MHII β pseudogene in perch. Since all sequences were amplified directly from whole genomic DNA it cannot be presumed *a priori* that all reported alleles are expressed in live fish. However, we found no indication for additional non-functional
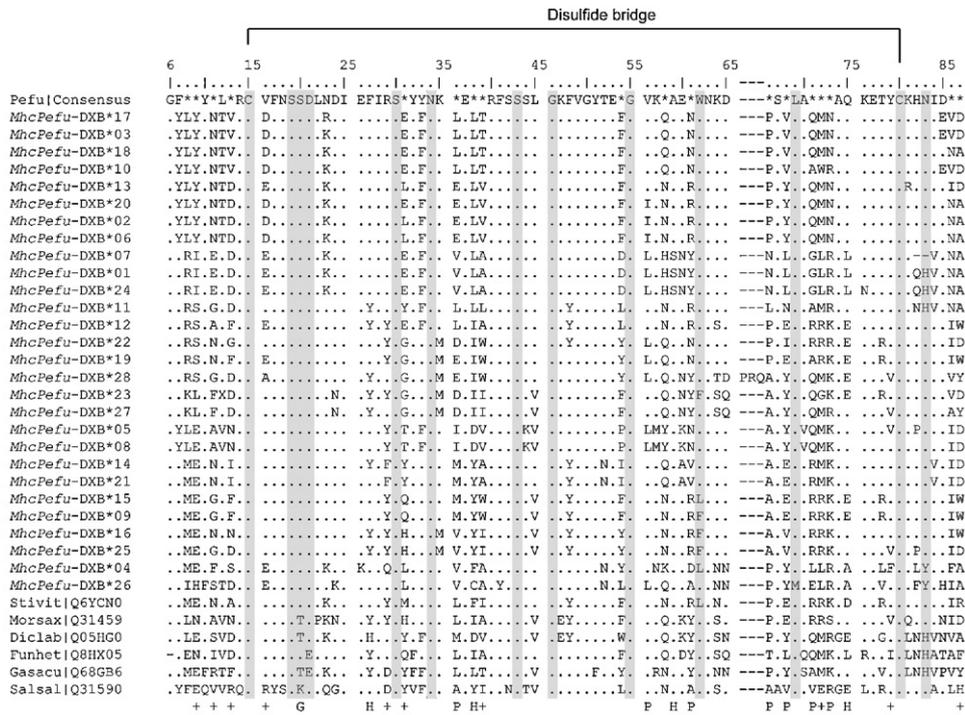
```
                                                    Disulfide bridge
                                  ┌──────────────────────────────────────────────────────┐
                 6     15        25        35        45        55        65        75        85
                 |...|....| ....|....| ....|....| ....|....| ....|....| ....|....| ---....|....| ....|....|.|
Pefu|Consensus   GF**Y*L*RC VFNSSDLNDI EFIRS*YYNK *E**RFSSSL GKFVGYTE*G VK*AE*WNKD ---*S*LA***AQ KETYCKHNID**
MhcPefu-DXB*17   .YLY.NTV.. D......R.. .....E.F.. L.LT...... ........F. ..Q..N.... ---P.V..QMN.. .........EVD
MhcPefu-DXB*03   .YLY.NTV.. D......K.. .....E.F.. L.LT...... ........F. ..Q..N.... ---P.V..QMN.. .........EVD
MhcPefu-DXB*18   .YLY.NTV.. D......K.. .....E.F.. L.LT...... ........F. ..Q..N.... ---P.V..QMN.. ..........NA
MhcPefu-DXB*10   .YLY.NTV.. D......K.. .....E.F.. L.LT...... ........F. ..Q..N.... ---P.V..AWR.. .........EVD
MhcPefu-DXB*13   .YLY.NTD.. E......K.. ....L.F.. E.LV...... ........F. ..N..R.... ---P.Y..QMN.. .....R....ID
MhcPefu-DXB*20   .YLY.NTD.. D......K.. .....E.F.. E.LV...... ........F. I.N..R.... ---P.Y..QMN.. ..........NA
MhcPefu-DXB*02   .YLY.NTD.. D......K.. .....L.F.. E.LV...... ........F. I.N..R.... ---P.Y..QMN.. ..........NA
MhcPefu-DXB*06   .YLY.NTD.. D......K.. .....L.F.. E.LV...... ........F. I.N..R.... ---P.Y..QMN.. ..........NA
MhcPefu-DXB*07   ..RI.E.D.. E......K.. .....E.F.. V.LA...... ........D. L.HSNY.... ---N.L..GLR.L ...--V.NA
MhcPefu-DXB*01   ..RI.E.D.. E......K.. .....E.F.. V.LA...... ........D. L.HSNY.... ---N.L..GLR.L ....QHV.NA
MhcPefu-DXB*24   ..RI.E.D.. E......K.. .....E.F.. V.LA...... ........D. L.HSNY.... ---N.L..GLR.L N...QHV.NA
MhcPefu-DXB*11   ..RS.G.D.. .......... .Y...Y.F.. L.LL...... .Y.....L. ..N..R.... ---L.N..AMR.. .....NHV.NA
MhcPefu-DXB*12   ..RS.A.F.. E......... .Y.Y.E.F.. L.IA...... .Y.....L. ..N..R..S. ---P.E..RRK.E .........IW
MhcPefu-DXB*22   ..RS.N.G.. .......... .Y..G..M D.IW...... .Y.....Y. L.Q..N.... ---P.I..RRR.E ..R......ID
MhcPefu-DXB*19   ..RS.N.F.. E......... .Y..G..M E.IW...... .......Y. L.Q..N.... ---P.E..ARK.E ..R......IW
MhcPefu-DXB*28   ..RS.G.D.. A......... .Y...G..M E.IW...... .......Y. L.Q.NY..TD PRQA.Y..QMK.E ...V.....VY
MhcPefu-DXB*23   ..KL.FXD.. ........N. .Y.Y.G..M D.II....V ........F. ..Q..NY.SQ ---A.Y..QGK.E ..R......VD
MhcPefu-DXB*27   ..KL.F.D.. .......N. .Y..G..M D.II....V ........F. ..Q.NY.SQ ---A.Y..QMR.. ...V.....AY
MhcPefu-DXB*05   .YLE.AVN.. .......... .Y.T.F.. I.DV...KV .......P. LMY.KN.... ---A.Y..VQMK.. ...V.P...ID
MhcPefu-DXB*08   .YLE.AVN.. .......... .Y.T.F.. I.DV...KV .......P. LMY.KN.... ---A.Y..VQMK.. .........ID
MhcPefu-DXB*14   ..ME.N.I.. .......... .Y.F.Y.. M.YA...... .Y...N.I. .Q.AV.... ---A.E..RMK.. ........V.ID
MhcPefu-DXB*21   ..ME.N.I.. .......... ..F.Y.. M.YA...... .Y...N.I. .Q.AV.... ---A.E..RMK.. ........V.ID
MhcPefu-DXB*15   ..ME.G.F.. .......... .Y.Q... M.YW....V .Y.....F. ..N..RL.. ---A.E..RRK.E ..R......IW
MhcPefu-DXB*09   ..ME.G.F.. .......... .Y.Q... M.YW....V .Y.....F. ..N..RF... ---A.E..RRK.E ..R......IW
MhcPefu-DXB*16   ..ME.N.N.. .......... .Y.Y.H.. M V.YI....V .Y.....Y. ..N..RF... ---A.V..RRK.. .........IW
MhcPefu-DXB*25   ..ME.G.D.. .......... .Y.Y.H.. M V.YI....V .Y.....Y. ..N..RF... ---A.V..RRK.. ...V.P...ID
MhcPefu-DXB*04   ..ME.F.S.. E......K. K..Q.L.. V.FA...... .......N.Y .NK..DL.NN ---P.Y..LLR.A .LF..LY..FA
MhcPefu-DXB*26   ..IHFSTD.. E......K. .....L... V.CA.Y.. .......N.L L.Q..A..NN ---P.YM.ELR.A ..V...FY.HIA
Stivit|Q6YCN0    ..ME.N.A.. .......K. ...Y.M.. L.FI...... .Y.....F. ..N..RL.N. ---P.E..RRK.D ..R......IR
Morsax|Q31459    ..LN.AVN.. ....T.PKN. .Y.Y.H.. L.IA....V .EY....F. ..Q.KY..S. ---P.E..RRS.. ...V.Q...NID
Diclab|Q05HG0    ..LE.SVD.. .....T..K. .H...Y.F.. M.DV....V .EY....W. ..Q.KY..SN ---P.Y..QMRGE ..G..LNHVNVA
Funhet|Q8HX05    -.EN.IVD.. .....E.... .Y...QF.. L.IA...... ........F. ..Q.DY.SQ ---T.L.QQMK.L R..I.LNHATAF
Gasacu|Q68GB6    ..MEFRTF.. ....TE.K. .Y.D.YFF.. L.LT....V ......F..Y .RN..Y..NN ---P.Y.SAMK.. ..V..LNHVPVY
Salsal|Q31590    .YFEQVVRQ. RYS.K..QG. ...D.YVF.. A.YI..N.TV ........L. .N..A..S. ---AAV..VERGE L.R....A.LH
                 + + +      +    G      H + +       P   H+               P   H  P        P  P P+P H      +        +
```

Fig. 5. Protein alignment of reported and published MHII β1 domain sequences for *Perca fluviatilis* L. Pefu-Consensus = 50% majority rule consensus sequence of all *P. fluviatilis* sequences, asterics indicates that this aa position is conserved in less than 50% of the sequences. Shaded areas indicate conserved positions as mentioned in the text. Letters below alignment: G = glycolisation site in mammals, P = position under positive selection in *P. fluviatilis*, H = human peptide binding residue not positively selected in *P. fluviatilis*, + = position under positive selection in *P. fluviatilis* that agrees with human peptide binding residue.

alleles in our dataset; since all remaining alleles possess most structurally important characteristics of functional MHII β alleles (see below).

4.2. Protein structure

The modelled tertiary structure indicates a characteristic 'MHC-fold' of the β1 domain of perch (Fig. 4). Like in other vertebrates, the majority of the positively selected sites were either located in the β-sheets ('floor') or in the α-helices ('walls'). As expected in the β1 domain of functional MHII alleles, many structurally important aa positions that are conserved among tetrapods (Kaufman et al., 1994) were also conserved in perch (Fig. 5). This can be exemplary seen in the two cysteine residues that form a disulfide bridge in vertebrates stabilizing the β1 subunit (Figs. 4 and 5, aa position 15 and 80). As in the human HLA-DR β1 domain these residues are in contiguous positions in the modelled tertiary structure (Fig. 4). Similarly, two glycins that provide flexibility in structurally important turns of the human HLA-DR β1 subunit are conserved, and also located in turns (Figs. 4 and 5, aa positions 46 and 55). Similarly the positions 34 (N), 43 (S) and 69 (L) that are conserved among vertebrates are also conserved in perch, with the asparagine in position 34 being involved in a Ig-like domain in the vertebrate β1 domain. Finally, a highly conserved asparagine (N) is found in aa position 63, this position is also highly conserved in vertebrates, and important for T-cell receptor binding in humans (Reche and Reinherz, 2003).

While the position and the specific aa residue of many conserved positions in the β1 domain of perch agrees with tetrapods (Kaufman et al., 1994; Reche and Reinherz, 2003), our results suggest some differences in the chemical properties of the amino acids in some of these conserved positions in some alleles. These include the substitution of the charged aspartic acid (D) in position 30 – a residue that is a conserved part of an immune-globulin like domain in the mammalian β1 domain – by the uncharged aa serin (S). Moreover, the positions 62 and 83 that are involved in peptide main chain binding and conserved in mammals (Reche and Reinherz, 2003) are more variable in perch (Fig. 5). In some alleles, the aromatic tryptophan (W) found in position 62 is replaced by an aromatic phenylalanine (F) or an aliphatic Leucin (L). Similarly, the uncharged asparagine (N) in position 83 is replaced by a basic and positively charged histidine (H) in some alleles. Interestingly, both alleles in group 3 (Fig. 2) have a large aromatic tyrosine (Y) in position 83. This substitution might be quite essential given the largely different chemical properties of tyrosine as compared to asparagine. Both latter positions (i.e. 62 and 83) are involved in peptide main chain binding, and contribute to the binding properties and specificity in the human HLA system (Reche and Reinherz, 2003). Hence, some of these substitutions might affect the stability of different antigenic peptides in the PBR.

The presence of an N-linked glycan at the asparagin residue in position 19 of the β1 domain is a common feature of all tetrapod class II β molecules (Kaufman et al., 1994). Hence, another structural difference of the β1 domain in perch was the absence of the tetrapod specific glycolisation consensus signal 'N-G-T' in aa positions 19 to 21 (Fig. 5). This tetrapod consensus sequence is also absent in other teleost species (Ono et al., 1993a,b; Grimholt et al., 1994; Dixon et al., 1996). In perch there is a putative glycolisation site ('N-S-S') in aa position 18 to 20 of the β1 domain that is conserved among all alleles, which also corroborates results obtained in cichlids (Ono et al., 1993a; Figueroa et al., 2000). In addition, McConnell et al. (1998) points out that many expressed teleost MHII β genes have a glycolisation signal encoded in this position.

Collectively, our results indicate that the tertiary structure of the β1 domain in perch is similar to that reported in other vertebrates, and, that the amplified alleles show most features of functional MHII β alleles, except for the two putative pseudogenes. In addition, we suggest that the protein structure of the amplified alleles might indicate some teleost specific differences as compared to tetrapods,

such as previously suggested for teleost fish (Kaufman et al., 1994; Stet et al., 2003).

4.3. Molecular evolution

There is a general consensus about the central role of positive selection for the maintenance of the high allelic diversity observed in the MH genes (Klein et al., 1993; Apanius et al., 1997; Bernatchez and Landry, 2003). Here we documented that positive selection also contributes to the MHII β1 domain diversity in perch. The number of codons evolving under positive selection in perch (17 codons) is of comparable magnitude as in the MHII β1 domain of cichlids were 19 codons were found to evolve under positive selection (Blais et al., 2007). In addition, the number of 17 residues appears to be slightly higher as the number of PBRs in humans, where crystallographic analysis has identified 13 PBR in the HLA-DR β1 domain (Brown et al., 1993). Twelve (71%) of the positively selected sites identified in our study agree with cichlids, and ten (59%) with humans. Altogether, nine of the residues under selection in perch are also positively selected in cichlids and agree with human PBRs (Table 2). Only three codons (aa residues 36, 56, and 66) that are positively selected in perch were neither evolving under positive selection in cichlids nor identified as human PBR. Two of these residues are adjacent to residues that evolve under positive selection in cichlids (aa position 65 and 75; Blais et al., 2007). Accordingly, our analysis might suggest some differences in the exact positions of the codons evolving under positive selection among teleosts, such as also reported for cyprinids (Ottova et al., 2005) and salmonids (Aguilar and Garza, 2007). Additionally, Aguilar and Garza (2007) suggested that the position of positively selected codons in the MHII β1 domain differs even among eleven salmonid species from the genera *Oncorhynchus*, *Salmo*, and *Salvelinus*. Hence, our results together with the aforementioned findings suggest that although the functional structure of the mammalian and fish MHII β1 domain is similar, there might be differences in the precise position of the aa positions evolving under positive selection.

To a lesser degree than positive selection, gene-conversion has also been identified as a possible mechanism generating new allelic variants in the MHII β1 domain (Ohta, 1995; Martinsohn et al., 1999). Our results indicate that gene-conversion also contributes to the diversity of the MHII β1 domain in perch. A high frequency of gene-conversion events among MHII β genes in the deer mouse has led Richman et al. (2003) to conclude that intragenic recombination is an important feature of MHII polymorphism. Our dataset did not allow determining the frequency of gene-conversion events, since suitable coalescent approaches require a data set consisting of homologous alleles (e.g. Reusch and Langefors, 2005). The number of 13 gene-conversion events among 28 MHII β1 alleles detected here is slightly lower than reported by Reusch and Langefors (2005), who observed 22 pairwise gene-conversion events among 31 MHII β1 domain sequences of *Gasterosteus aculeatus*. However, similar to these authors we warrant that the actual number of gene-conversion events in our dataset might be higher, since the high number of pairwise comparisons in the GENECONV analysis results in a drastic reduction of the significance level following corrections for multiple comparisons (148 significant gene-conversion events prior correction; data not shown).

The excess of synonymous diversity in exon II over the diversity in intron I, both within and between sequence clusters, is also congruent with recombination acting on the MHII β genes. Diversifying selection is only acting on PBR codons in exon II (and exon III in class I genes). Hence, with free recombination across the exon–intron boundary, genetic drift could lead to fixation of introns while exon polymorphism is maintained (Hughes, 2000; Reusch and Langefors, 2005). This is namely reflected in a higher synonymous divergence in the exon compared to the intron (Cereb et al., 1997), which finally

results in a high similarity of intron sequences whereby the diversity in the exon is maintained.

Interestingly, all gene-conversion events (Table 4) were detected between otherwise divergent alleles (Table 4, *p*-distance ≥ 0.15) from distinct phylogenetic clusters or even loci (cf. Fig. 2). This may partly reflect less resolution power to detect gene-conversion between more similar sequences. However, GENECONV attempts to correct for such differences at least for long sequences (Sawyer, personal communication). Moreover, some of the sequences involved in gene-conversion events detected in our analysis originate from different phylogenetic clusters that likely represent different loci. Hence, we suggest that this result might reflect a dominance of inter-locus gene-conversion, a result consistent with Reusch and Langefors (2005). Ohta (1999) predicted that gene-conversion enhances the allelic diversity mostly among divergent alleles, such as for example between paralogous allelic lineages. The rationale for this finding is that a new allelic variant created from dissimilar alleles has a higher chance of being a new and unique allele, which might increase pathogen resistance, and will be selected for. Hence, inter-locus gene-conversion can be particularly effective in increasing the allelic repertoire, and might be important in species where a high number of MHII β loci provide a broad 'reservoir' of genetic diversity (Andersson and Mikko, 1995), such as in perch and other teleost species.

Our results provide some evidence for the hypothesis that the evolution and the structure of the MH(C) II β genes might differ between vertebrate classes (Edwards et al., 1995) and possibly also between different teleost lineages (Aguilar and Garza, 2007). Several molecular mechanisms act jointly to generate the high diversity of the MH(C) II β genes: (i) new gene loci are generated by gene duplication (Wolfe, 2001; Hokamp et al., 2003). (ii) New allelic variants can be generated by point mutations and/or via inter- and intragenic recombination or gene-conversion (Ohta, 1995). However, Richman et al. (2003) have pointed out that in terms of maintaining or increasing MHC allele variability, point mutations may play a rather inferior role as compared to intragenic recombination. (iii) Finally, positive selection acting on the PBR is crucial for maintaining the previously generated diversity. All these mechanisms interact in maintaining the variability of MH(C) II β genes depending on the specific evolutionary history of the species under consideration. Hence, species with a single MHII β locus like the Atlantic salmon (Langefors et al., 2001), which cannot undergo inter-locus recombination, may have taken another evolutionary "route" to maintain their MHII β gene variability than species with multiple loci like sticklebacks (Reusch and Langefors, 2005), cichlids (Malaga-Trillo et al., 1998) or the Eurasian perch (this study). Altogether we suggest that care has to be taken when transferring conclusions between vertebrate lineages and data interpretation should also consider the evolutionary history of the species under consideration.

5. Conclusions

This study provides evidence that the non-model percid species Eurasian perch (*Perca fluviatilis* L.) has at least eight MHII β gene loci that exhibit a high allelic variability (28 alleles in 58 individuals analysed). The tertiary structure of the β1 domain in perch is similar to other vertebrates. However, our results indicate some teleost specific differences in the amino acid composition. Finally, both positive selection and gene-conversion events (i.e. intra- and intergenic sequence exchanges) contribute to the demonstrated high allelic diversity of MHII β genes in this species. Further studies could elaborate several intriguing aspects of the MHII β gene evolution in perch, such as investigating the number of MHII β loci expressed in living fish and the genetic architecture of the MHII β genes in this non-model percid species. Finally future studies will focus on

the role of the MHII genes in pathogen resistance, kin recognition and mate choice in perch (Behrmann-Godel et al., 2006) to further elaborate our understanding of the selective forces that generate the MHII β gene variability in teleosts.

Acknowledgements

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.molimm.2009.07.012.

References

Addison, J.A., 2007. Simultaneous cloning of multiple nuclear genes by pooling PCR products of variable size: a cost-effective method of improving efficiency in large-scale genetic analyses. Mol. Ecol. Notes 7, 389–392.

Aguilar, A., Garza, J.C., 2007. Patterns of historical balancing selection on the Salmonid major histocompatibility complex class II beta gene. J. Mol. Evol. 65, 34–43.

Aljanabi, S.M., Martinez, I., 1997. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. Nucleic Acids Res. 25, 4692–4693.

Andersson, L., Mikko, S., 1995. Generation of MHC class II diversity by intra- and intergenic recombination. Immunol. Rev. 143, 5–12.

Apanius, V., Penn, D., Slev, P.R., Ruff, L.R., Potts, W.K., 1997. The nature of selection on the major histocompatibility complex. Crit. Rev. Immunol. 17, 179–224.

Behrmann-Godel, J., Gerlach, G., Eckmann, R., 2006. Kin and population recognition in sympatric Lake Constance perch (*Perca fluviatilis* L.): can assortative shoaling drive population divergence? Behav. Ecol. Sociobiol. 59, 461–468.

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

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.

Blais, J., Rico, C., van Oosterhout, C., Cable, J., Turner, G.F., Bernatchez, L., 2007. MHC adaptive divergence between closely related and sympatric African cichlids. PLoS ONE 2 (8), e734.

Bos, D.H., Waldman, B., 2006. Evolution by recombination and transspecies polymorphism in the MHC class I gene of *Xenopus laevis*. Mol. Biol. Evol. 23, 137–143.

Brown, J.H., Jardetzky, T.S., Gorga, J.C., Stern, L.J., Urban, R.G., Strominger, J.L., Wiley, D.C., 1993. Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. Nature 364, 33–39.

Cereb, N., Hughes, A.L., Yang, S.Y., 1997. Locus-specific conservation of the HLA class I introns by intra-locus homogenization. Immunogenetics 47, 30–36.

Dionne, M., Miller, K.M., Dodson, J.J., Bernatchez, L., 2009. MHC standing genetic variation and pathogen resistance in wild Atlantic salmon. Philos. Trans. R. Soc. Lond. B: Biol. Sci. 364, 1555–1565.

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

Dixon, B., Nagelkerke, L.A.J., Sibbing, F.A., Egberts, E., Stet, R.J.M., 1996. Evolution of MHC class II B chain-encoding genes in the Lake Tana barbel species flock (*Barbus intermedius* complex). Immunogenetics 44, 419–431.

Doherty, P.C., Zinkernagel, R.M., 1975. Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. Nature 256, 50–52.

Edwards, S.V., Wakeland, E.K., Potts, W.K., 1995. Contrasting histories of avian and mammalian Mhc genes revealed by class II B sequences from songbirds. Proc. Natl. Acad. Sci. U.S.A. 92, 12200–12204.

Edwards, S.V., Gasper, J., March, M., 1998. Genomics and polymorphism of Agph-DAB1, an MHC class II B gene in red-winged blackbirds (*Agelaius phoeniceus*). Mol. Biol. Evol. 15, 236–250.

Engelhard, V.H., 1994. Structure of peptides associated with class I and class II MHC molecules. Annu. Rev. Immunol. 12, 181–207.

Figueroa, F., Mayer, W.E., Sultmann, H., O'HUigin, C., Tichy, H., Satta, Y., Takezaki, N., Takahata, N., Klein, J., 2000. Mhc class II B gene evolution in East African cichlid fishes. Immunogenetics 51, 556–575.

Flajnik, M.F., Du Pasquier, L., 2004. Evolution of innate and adaptive immunity: can we draw a line? Trends Immunol. 25, 640–644.

Flajnik, M.F., Kasahara, M., 2001. Comparative genomics of the MHC: glimpses into the evolution of the adaptive immune system. Immunity 15, 351–362.

Grimholt, U., Olsaker, I., de Vries Lindstrom, C., Lie, O., 1994. A study of variability in the MHC class II beta 1 and class I alpha 2 domain exons of Atlantic salmon, *Salmo salar* L. Anim. Genet. 25, 147–153.

Gu, X., Nei, M., 1999. Locus specificity of polymorphic alleles and evolution by a birth-and-death process in mammalian MHC genes. Mol. Biol. Evol. 16, 147–156.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids Symp. Ser. 41, 95–98.

Hansen, J.D., Strassburger, P., Thorgaard, G.H., Young, W.P., Du Pasquier, L., 1999. Expression, linkage, and polymorphism of MHC-related genes in rainbow trout, *Oncorhynchus mykiss*. J. Immunol. 163, 774–786.

Hedrick, P.W., 2002. Pathogen resistance and genetic variation at MHC loci. Evol. Int. J. Org. Evol. 56, 1902–1908.

Hedrick, P.W., Kim, T.J., 2000. Genetics of complex polymorphisms: parasites and maintenance of major histocompatibility complex variation. In: Singh, R.S., Krimbas, C.B. (Eds.), Evolutionary Genetics: From Molecules to Morphology. Cambridge University Press, pp. 204–234.

Högstrand, K., Böhme, J., 1994. A determination of the frequency of gene conversion in unmanipulated mouse sperm. Proc. Natl. Acad. Sci. U.S.A. 91, 9921–9925.

Hokamp, K., McLysaght, A., Wolfe, K.H., 2003. The 2R hypothesis and the human genome sequence. J. Struct. Funct. Genomics 3, 95–110.

Hughes, A.L., 1999. Adaptive Evolution of Genes and Genomes. Oxford University Press, New York.

Hughes, A.L., 2000. Evolution of introns and exons of class II major histocompatibility complex genes of vertebrates. Immunogenetics 51, 473–486.

Hughes, A.L., Hughes, M.K., 1995. Natural selection on the peptide-binding regions of major histocompatibility complex molecules. Immunogenetics 42, 233–243.

Hughes, A.L., Hughes, M.K., Howell, C.Y., Nei, M., 1994. Natural selection at the class II major histocompatibility complex loci of mammals. Philos. Trans. R. Soc. Lond. B: Biol. Sci. 346, 359–367.

Hughes, A.L., Hughes, M.K., Watkins, D.I., 1993. Contrasting roles of interallelic recombination at the HLA-A and HLA-B loci. Genetics 133, 669–680.

Hughes, A.L., Nei, M., 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335, 167–170.

Hughes, A.L., Nei, M., 1989. Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. Proc. Natl. Acad. Sci. U.S.A. 86, 958–962.

Hughes, A.L., Yeager, M., 1998. Natural selection at major histocompatibility complex loci of vertebrates. Annu. Rev. Genet. 32, 415–435.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90, 773–795.

Kaufman, J., Salomonsen, J., Flajnik, M., 1994. Evolutionary conservation of MHC class I and class II molecules—different yet the same. Semin. Immunol. 6, 411–424.

Kelley, J., Walter, L., Trowsdale, J., 2005. Comparative genomics of major histocompatibility complexes. Immunogenetics 56, 683–695.

Kim, T.J., Parker, K.M., Hedrick, P.W., 1999. Major histocompatibility complex differentiation in Sacramento River chinook salmon. Genetics 151, 1115–1122.

Klein, J., 1986a. The Natural History of the Major Histocompatibility Complex. John Wiley and Sons, New York, USA.

Klein, J., 1986b. The continuation of the debate on gene conversion. Transplantation 42, 102–105.

Klein, J., Satta, Y., O'HUigin, C., Takahata, N., 1993. The molecular descent of the major histocompatibility complex. Annu. Rev. Immunol. 11, 269–295.

Klein, J., Bontrop, R.E., Dawkins, R.L., Erlich, H.A., Gyllensten, U.B., Heise, E.R., Jones, P.P., Parham, P., Wakeland, E.K., Watkins, D.I., 1990. Nomenclature for the major histocompatibility complexes of different species: a proposal. Immunogenetics 31, 217–219.

Kosakovsky Pond, S.L., Frost, S.D., 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol. Biol. Evol. 22, 1208–1222.

Langefors, A., Lohm, J., von Schantz, T., 2001. Allelic polymorphism in MHC class II B in four populations of Atlantic salmon (*Salmo salar*). Immunogenetics 53, 329–336.

Lee, M.S., 2001. Unalignable sequences and molecular evolution. Trends Ecol. Evol. 16, 681–685.

Malaga-Trillo, E., Zaleska-Rutczynska, Z., McAndrew, B., Vincek, V., Figueroa, F., Sultmann, H., Klein, J., 1998. Linkage relationships and haplotype polymorphism among cichlid MHC class II B loci. Genetics 149, 1527–1537.

Martinsohn, J.T., Sousa, A.B., Guethlein, L.A., Howard, J.C., 1999. The gene conversion hypothesis of MHC evolution: a review. Immunogenetics 50, 168–200.

McConnell, T.J., Godwin, U.B., Cuthbertson, B.J., 1998. Expressed major histocompatibility complex class II loci in fishes. Immunol. Rev. 166, 294–300.

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

Miller, K.M., Withler, R.E., 1996. Sequence analysis of a polymorphic MHC class II gene in Pacific salmon. Immunogenetics 43, 337–351.

Müller, K.F., 2005. SeqState: primer design and sequence statistics for phylogenetic DNA datasets. Appl. Bioinformatics 4, 65–69.

Müller, K.F., 2004. PRAP-computation of Bremer support for large data sets. Mol. Phylogenet. Evol. 31, 780–782.

Nei, M., Rooney, A.P., 2005. Concerted and birth-and-death evolution of multigene families. Annu. Rev. Genet. 39, 121–152.

Nei, M., Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3, 418–426.

Ohta, T., 1999. Effect of gene conversion on polymorphic patterns at major histocompatibility complex loci. Immunol. Rev. 167, 319–325.

Ohta, T., 1995. Gene conversion vs point mutation in generating variability at the antigen recognition site of major histocompatibility complex loci. J. Mol. Evol. 41, 115–119.

Ono, H., O'HUigin, C., Tichy, H., Klein, J., 1993a. Major-histocompatibility-complex variation in two species of cichlid fishes from Lake Malawi. Mol. Biol. Evol. 10, 1060–1072.

Ono, H., O'HUigin, C., Vincek, V., Stet, R.J., Figueroa, F., Klein, J., 1993b. New beta chain-encoding Mhc class II genes in the carp. Immunogenetics 38, 146–149.

Ono, H., Klein, D., Vincek, V., Figueroa, F., O'HUigin, C., Tichy, H., Klein, J., 1992. Major histocompatibility complex class II genes of zebrafish. Proc. Natl. Acad. Sci. U.S.A. 89, 11886–11890.

Ottova, E., Simkova, A., Martin, J.F., de Bellocq, J.G., Gelnar, M., Allienne, J.F., Morand, S., 2005. Evolution and trans-species polymorphism of MHC class II beta genes in cyprinid fish. Fish Shellfish Immunol. 18, 199–222.

Parham, P., Lawlor, D.A., Lomen, C.E., Ennis, P.D., 1989. Diversity and diversification of HLA-A,B,C alleles. J. Immunol. 142, 3937–3950.

Phillips, R.B., Zimmerman, A., Noakes, M.A., Palti, Y., Morasch, M.R., Eiben, L., Ristow, S.S., Thorgaard, G.H., Hansen, J.D., 2003. Physical and genetic mapping of the rainbow trout major histocompatibility regions: evidence for duplication of the class I region. Immunogenetics 55, 561–569.

Pond, S.L., Frost, S.D., 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21, 2531–2533.

Poon, A.F.Y., Frost, S.D., Kosakovsky Pond, S.L., 2009. Detecting signatures of selection from DNA sequences using Datamonkey. In: Posada, D. (Ed.), Methods in Molecular Biology. Humana Press, pp. 163–183.

Posada, D., 2002. Evaluation of methods for detecting recombination from DNA sequences: empirical data. Mol. Biol. Evol. 19, 708–717.

Posada, D., Crandall, K.A., 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proc. Natl. Acad. Sci. U.S.A. 98, 13757–13762.

Reche, P.A., Reinherz, E.L., 2003. Sequence variability analysis of human class I and class II MHC molecules: functional and structural correlates of amino acid polymorphisms. J. Mol. Biol. 331, 623–641.

Reusch, T.B.H., Langefors, A., 2005. Inter- and intralocus recombination drive MHC class IIB gene diversification in a teleost, the three-spined stickleback *Gasterosteus aculeatus*. J. Mol. Evol. 61, 531–541.

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.

Richman, A.D., Herrera, L.G., Nash, D., Schierup, M.H., 2003. Relative roles of mutation and recombination in generating allelic polymorphism at an MHC class II locus in *Peromyscus maniculatus*. Genet. Res. 82, 89–99.

Robinson-Rechavi, M., Marchand, O., Escriva, H., Bardet, P.L., Zelus, D., Hughes, S., Laudet, V., 2001. Euteleost fish genomes are characterized by expansion of gene families. Genome Res. 11, 781–788.

Rozen, S., Skaletsky, H.J., 2000. Primer3 on the www for general users and for biologist programmers. In: Krawetz, S., Misener, S. (Eds.), Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, pp. 365–386.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.

Sambrook, J.G., Figueroa, F., Beck, S., 2005. A genome-wide survey of major histocompatibility complex (MHC) genes and their paralogues in zebrafish. BMC Genomics 6, 152.

Sambrook, J.G., Russell, R., Umrania, Y., Edwards, Y.J., Campbell, R.D., Elgar, G., Clark, M.S., 2002. Fugu orthologues of human major histocompatibility complex genes: a genome survey. Immunogenetics 54, 367–380.

Sato, A., Figueroa, F., Murray, B.W., Malaga-Trillo, E., Zaleska-Rutczynska, Z., Sultmann, H., Toyosawa, S., Wedekind, C., Steck, N., Klein, J., 2000. Nonlinkage of major histocompatibility complex class I and class II loci in bony fishes. Immunogenetics 51, 108–116.

Satta, Y., O'HUigin, C., Takahata, N., Klein, J., 1993. The synonymous substitution rate of the major histocompatibility complex loci in primates. Proc. Natl. Acad. Sci. U.S.A. 90, 7480–7484.

Sawyer, S.A., 1999. GENECONV: a computer package for the statistical detection of gene conversion. Distributed by the author, Department of Mathematics, Washington University in St. Louis, available at http://www.math.wustl.edu/~sawyer.

Schaschl, H., Wandeler, P., Suchentrunk, F., Obexer-Ruff, G., Goodman, S.J., 2006. Selection and recombination drive the evolution of MHC class II DRB diversity in ungulates. Heredity 97, 427–437.

Schaschl, H., Suchentrunk, F., Hammer, S., Goodman, S.J., 2005. Recombination and the origin of sequence diversity in the DRB MHC class II locus in chamois (*Rupicapra* spp.). Immunogenetics 57, 108–115.

Scheffler, K., Martin, D.P., Seoighe, C., 2006. Robust inference of positive selection from recombining coding sequences. Bioinformatics 22, 2493–2499.

Shum, B.P., Guethlein, L., Flodin, L.R., Adkison, M.A., Hedrick, R.P., Nehring, R.B., Stet, R.J., Secombes, C., Parham, P., 2001. Modes of salmonid MHC class I and II evolution differ from the primate paradigm. J. Immunol. 166, 3297–3308.

Simmons, M.P., Müller, K., Norton, A.P., 2007. The relative performance of indel-coding methods in simulations. Mol. Phylogenet. Evol. 44, 724–740.

Stet, R.J., Kruiswijk, C.P., Dixon, B., 2003. Major histocompatibility lineages and immune gene function in teleost fishes: the road not taken. Crit. Rev. Immunol. 23, 441–471.

Stet, R.J., 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.

Swofford, D.L., 2000. PAUP: Phylogenetic Analysis Using Parsimony. Smithsonian Institution and Sinauer, Sunderland, MA.

Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. Mol. Biol. Evol. 24, 1596–1599.

The MHC sequencing consortium, 1999. Complete sequence and gene map of a human major histocompatibility complex. The MHC sequencing consortium. Nature 401, 921–923.

Trowsdale, J., 1993. Genomic structure and function in the MHC. Trends Genet. 9, 117–122.

Wegner, K.M., Kalbe, M., Kurtz, J., Reusch, T.B., Milinski, M., 2003. Parasite selection for immunogenetic optimality. Science 301, 1343.

Wolfe, K.H., 2001. Yesterday's polyploids and the mystery of diploidization. Nat. Rev. Genet. 2, 333–341.

Zangenberg, G., Huang, M.M., Arnheim, N., Erlich, H., 1995. New HLA-DPB1 alleles generated by interallelic gene conversion detected by analysis of sperm. Nat. Genet. 10, 407–414.

Zylstra, P., Rothenfluh, H.S., Weiller, G.F., Blanden, R.V., Steele, E.H., 1998. PCR amplification of murine immunoglobulin germline V genes: strategies for minimization of recombination artefacts. Immunol. Cell Biol. 76, 395–405.