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

Christian Michel, Louis Bernatchez, Jasminca Behrmann-Godel

**A R T I C L E   I N F O**

Article history:
Received 26 March 2009
Received in revised form 14 July 2009
Accepted 16 July 2009
Available online 14 August 2009

Keywords:
Gene-conversion
Diversifying selection
Major histocompatibility complex
MHC class II
Recombination
*Perca fluviatilis* L.
Perch

**A B S T R A C T**

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.

© 2009 Elsevier Ltd. All rights reserved.

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 differ 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 MHC complexes (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 to 13 loci in cichlids (Malaga-Trillo et al., 1998).
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 (Martinsohn et al., 2001; Richman et al., 2003; Bos and Waldman, 2006; Schaschl et al., 2006). These studies also indicate that both intra- and inter-locus gene-conversion contribute to the 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 (Höstrand 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 specifically to the allelic diversity of the MHI β1 domain. It has also been suggested that gene-conversion might be an important mechanism for generating new MHI 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 MHI β genes may differ between different vertebrate classes (Edwards et al., 1995) and possibly also among different teleost lineages (Agulgar 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 MHI 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 MHI 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 MHI 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 mm²) 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 MHI 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 (Stzostedtos vitreum, accession no. A1Y58838:1; Cyphophilaia frontosa, accession no. L132321; 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 (StviMHSF), 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 MHI 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 each dNTP-Mix (Peqlab GmbH, Germany), 1 x PCR Buffer (Peqlab) and 1.25 units PeqGold Sawadade Taq DNA polymerase (Peqlab) in 25 μl reactions with approx. 50 ng template. Amplifications were performed in a Preamus 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 E. coli DH5α following the manufacturer’s protocol (Invitrogen). The transformed bacteria were plated onto LB-plates supplemented with X-gal and ampicillin, and colonies grown overnight at 37°C in LB medium. White colonies were picked for plasmid extraction and sequenced (Nationale Sequenzierungszentrum, Heidelberg, Germany).

#### 2.1.3. Data analysis

The nucleotide diversity was calculated using DnaSP version 5.10.01 (Librado and Rozas, 2009) and for the analysis the following parameters were used: the number of observed segregating sites (S), the number of nucleotide differences (k), the number of pairwise nucleotide differences (θ), and the genetic diversity (H). The probability of the observed segregating sites was calculated by the coalescent simulation algorithm of ms (Hudson, 2002) and by the substitution-of-sites algorithm implemented in nucleotide diversity (Nei, 1987). The haplotype diversity and nucleotide diversity were calculated using DnaSP version 5.10.01 (Librado and Rozas, 2009). The total number of codons, the number of synonymous and nonsynonymous substitutions, and the number of codons with nonsynonymous substitutions were calculated using the codeml program of PAML (Yang, 1997). The number of codons and the number of synonymous and nonsynonymous substitutions were calculated using the codeml program of PAML (Yang, 1997). The number of synonymous and nonsynonymous substitutions was calculated using the codeml program of PAML (Yang, 1997). The number of synonymous and nonsynonymous substitutions was calculated using the codeml program of PAML (Yang, 1997).
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® INVoX 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’-ACTGCCCCTGCTTTACA-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 MgCl2, 1 × PCR Buffer Y (Peqlab), and 1.25 units peqGOLD Sawadwe Taq DNA polymerase (Peqlab) in 20 μL reaction volume. All screening reactions were performed in a Perkin 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 Reuilingen, 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 B1 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 1 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 B1 sequences, and for a published B1 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 B1 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 B1 domain of the single MHII β locus in Salmo salar (pM = 0.06 ± 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 Witthaler, 1996, pM = 0.06; Miller et al., 1997, pM = 0.03; Kim et al., 1999, pM = 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.

2.1.3.2. Genetic diversity and identification of putative MHII β gene loci. Phylogenetic reconstruction was used to, firstly, illustrate the genetic diversity of the MHII B1 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 B1 domain sequence of the axolotl Ambystoma mexicanum (Genbank accession no. EF585232).

2.1.3.3. Molecular evolution. The model selection procedure implemented in the Datamonkey webservice (Pond and Frost, 2005; http://www.datamonkey.org) identified the Felstenstein-B1 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 Parrish method (Scheffler et al., 2006) as implemented in the Datamonkey web server was applied. The Parrish 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 χ²-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 dN and three dS 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 dN/dS 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 dN > dS 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 dN > dS to the prior odds for that event and if BF > 30 provides very strong evidence that dN > dS 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 dN > dS exceeded 30. Simulation studies have shown that in phylogenetic REL methods (iii) the intron I sequences all in polymorphism was calculated for (i) all codons of the gene loci A, B, C and E (as shown in the phylogeny Fig. 2B, excluding polymorphism overall and also separately for the putative MHII between exon II and intron I sequences we analysed the nucleotide 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 Mscr 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 dN and dS 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 (dN) with the number of synonymous substitutions per synonymous site (dS).

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 of 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 (wwwpdb.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 StviMHSF was found at the beginning of every repeat, indicating multiple binding sites for StviMHSF 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. 1, 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 1, supplementary material). This is comparable in magnitude to the number of alleles found in twoichid 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. 1, supplementary material). All nucleotide sequence data reported here is available in the GenBank database (GenBank accession nos. FN 293111–29, 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 fluviatilis* 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 fluviatilis* 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 (1 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 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 (group 1) 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 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 95% 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 MHL 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 ≤ dS). In addition, the REL analysis identified 17 codons in the MHL 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 I 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 ± 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 sixfold 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>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>0.342</td>
<td>5.636</td>
<td>5.295</td>
<td>0.99</td>
<td>&gt;100</td>
<td>0.005</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>11</td>
<td>0.733</td>
<td>6.336</td>
<td>5.603</td>
<td>1.00</td>
<td>&gt;100</td>
<td>0.000</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>13</td>
<td>0.248</td>
<td>6.326</td>
<td>6.078</td>
<td>1.00</td>
<td>&gt;100</td>
<td>0.000</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>16</td>
<td>0.252</td>
<td>1.804</td>
<td>1.552</td>
<td>0.98</td>
<td>&gt;50</td>
<td>0.014</td>
<td>×</td>
<td>#</td>
</tr>
<tr>
<td>29</td>
<td>0.516</td>
<td>6.104</td>
<td>5.587</td>
<td>0.99</td>
<td>&gt;100</td>
<td>0.006</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>31</td>
<td>0.351</td>
<td>6.334</td>
<td>5.984</td>
<td>1.00</td>
<td>&gt;100</td>
<td>0.000</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>36</td>
<td>0.353</td>
<td>2.891</td>
<td>2.537</td>
<td>0.96</td>
<td>36</td>
<td>0.028</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>0.386</td>
<td>6.332</td>
<td>5.946</td>
<td>1.00</td>
<td>&gt;100</td>
<td>0.000</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>56</td>
<td>0.326</td>
<td>1.857</td>
<td>1.530</td>
<td>0.96</td>
<td>34</td>
<td>0.003</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>0.236</td>
<td>6.220</td>
<td>5.983</td>
<td>1.00</td>
<td>&gt;100</td>
<td>0.000</td>
<td>×</td>
<td>59</td>
</tr>
<tr>
<td>66</td>
<td>0.194</td>
<td>1.798</td>
<td>1.604</td>
<td>1.00</td>
<td>&gt;100</td>
<td>0.002</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td>68</td>
<td>0.413</td>
<td>6.334</td>
<td>5.921</td>
<td>1.00</td>
<td>&gt;100</td>
<td>0.000</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>1.202</td>
<td>6.320</td>
<td>5.118</td>
<td>1.00</td>
<td>&gt;100</td>
<td>0.002</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>0.336</td>
<td>1.839</td>
<td>1.503</td>
<td>0.96</td>
<td>31</td>
<td>0.032</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>73</td>
<td>0.303</td>
<td>1.876</td>
<td>1.573</td>
<td>0.97</td>
<td>41</td>
<td>0.025</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>0.348</td>
<td>2.142</td>
<td>1.794</td>
<td>0.95</td>
<td>31</td>
<td>0.032</td>
<td>78</td>
<td></td>
</tr>
<tr>
<td>87</td>
<td>1.647</td>
<td>6.331</td>
<td>4.685</td>
<td>1.00</td>
<td>&gt;100</td>
<td>0.001</td>
<td>×</td>
<td>×</td>
</tr>
</tbody>
</table>

# 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.
Table 3

Nucleotide polymorphism of the MHII β1 domain and intron 1 in *Perca fluviatilis* L. Seq. clusters A, B, C and E refer to putative loci as identified in the intron 1 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 *P*-value 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.

<table>
<thead>
<tr>
<th>Seq. cluster</th>
<th>Gene region</th>
<th>Codon type</th>
<th>dN</th>
<th>dS (d introns)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Exon 2 (N = 4)</td>
<td>All codons</td>
<td>0.029 ± 0.011**</td>
<td>0.000 ± 0.000</td>
</tr>
<tr>
<td></td>
<td>Exon 2 (N = 4)</td>
<td>REL</td>
<td>0.058 ± 0.051*</td>
<td>0.000 ± 0.000</td>
</tr>
<tr>
<td></td>
<td>Intron 1 (N = 5)</td>
<td>All codons</td>
<td>0.008 ± 0.005</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>Exon 2 (N = 4)</td>
<td>All codons</td>
<td>0.018 ± 0.008</td>
<td>0.023 ± 0.017</td>
</tr>
<tr>
<td></td>
<td>Exon 2 (N = 4)</td>
<td>REL</td>
<td>0.066 ± 0.034</td>
<td>0.050 ± 0.059</td>
</tr>
<tr>
<td></td>
<td>Intron 1 (N = 6)</td>
<td>All codons</td>
<td>0.013 ± 0.006</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>Exon 2 (N = 6)</td>
<td>All codons</td>
<td>0.108 ± 0.023</td>
<td>0.111 ± 0.037</td>
</tr>
<tr>
<td></td>
<td>Exon 2 (N = 6)</td>
<td>REL</td>
<td>0.370 ± 0.120***</td>
<td>0.031 ± 0.026</td>
</tr>
<tr>
<td></td>
<td>Intron 1 (N = 5)</td>
<td>All codons</td>
<td>0.021 ± 0.009</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>Exon 2 (N = 4)</td>
<td>All codons</td>
<td>0.224 ± 0.038</td>
<td>0.137 ± 0.041</td>
</tr>
<tr>
<td></td>
<td>Exon 2 (N = 4)</td>
<td>REL</td>
<td>1.207 ± 0.217***</td>
<td>0.085 ± 0.053</td>
</tr>
<tr>
<td></td>
<td>Intron 1 (N = 4)</td>
<td>All codons</td>
<td>0.049 ± 0.017</td>
<td></td>
</tr>
</tbody>
</table>

* Over all mean

* Over all mean

| Over all mean | All codons | 0.223 ± 0.032 | 0.213 ± 0.038 |

| Over all mean | All codons | 0.942 ± 0.117*** | 0.213 ± 0.079 |

| Over all mean | All codons | 0.071 ± 0.014 |                  |

Table 4

Detailed results of the GenoConv 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.

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

* Including all clusters; excluding pseudogenes.

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*-values in our dataset (0.19 ± 0.01) as compared to the variation found in population samples of the β1 domain in salmonids, which possesses a single MHII β locus (Miller and Withiler, 1996: *p* = 0.06; Miller et al., 1997: *p* = 0.03; Kim et al., 1999: *p* = 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*-value ≤ 0.06, compared to a separation between any two clusters *p*-value ≥ 0.10. Some researchers interpreted phylogenetic clusters in multilocus datasets of MHII β 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 exon 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
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 positions 46 and 55). Similarly the two glycins that provide flexibility in structurally important turns of the human HLA-DR1 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 an Ig-like domain in the vertebrate 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; Dixon et al., 1996). In perch there is a putative glycolisation site (‘N-S-S’) in aa positions 18 to 20 of the β1 domain – by the uncharged aa serin (S) that agrees with human peptide binding residue.

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 \( \beta \)1 domain diversity in perch. The number of codons evolving under positive selection in perch (17 codons) is of comparable magnitude as in the MHII \( \beta \)1 domain of cichlids. 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 \( \beta \)1 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 \( \beta \)1 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 \( \beta \)1 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 \( \beta \)1 domain (Ohta, 1995; Martinsohn et al., 1999). Our results indicate that gene-conversion also contributes to the diversity of the MHII \( \beta \)1 domain in perch. A high frequency of gene-conversion events among MHII \( \beta \) 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 Langeffors, 2005). The number of 13 gene-conversion events among 28 MHII \( \beta \)1 alleles detected here is slightly lower than reported by Reusch and Langeffors (2005), who observed 22 pairwise gene-conversion events among 31 MHII \( \beta \)1 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 \( \beta \) 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 Langeffors, 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 Langeffors (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 \( \beta \) loci provide a broad ‘reservoir’ of genetic diversity (Andersson and Mikko, 1995), such as in perch and other teleost species.

Our result provide some evidence for the hypothesis that the evolution and the structure of the MH(C) II \( \beta \) 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 \( \beta \) 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 \( \beta \) genes depending on the specific evolutionary history of the species under consideration. Hence, species with a single MHII \( \beta \) locus like the Atlantic salmon (Langeffors et al., 2001), which cannot undergo inter-locus recombination, may have taken another evolutionary “route” to maintain their MHII \( \beta \) gene variability than species with multiple loci like sticklebacks (Reusch and Langeffors, 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 \( \beta \) gene loci that exhibit a high allelic variability (28 alleles in 58 individuals analysed). The tertiary structure of the \( \beta \)1 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 \( \beta \) genes in this species. Further studies could elaborate several intriguing aspects of the MHII \( \beta \) gene evolution in perch, such as investigating the number of MHII \( \beta \) loci expressed in living fish and the genetic architecture of the MHII \( \beta \) genes in this non-model percid species. Finally future studies will focus on
the role of the MHLI 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 MHLI β gene variability in teleosts.

Acknowledgements

The authors would like to thank Reiner Eckmann, Elisabeth Gross, Melanie Hempel, Arne Nolte, and Jonatan Blais for help and support. Moreover, Dave Gerrard and Kai Müller are acknowledged for help with the data analysis. The corrections and comments of Andrew Clarke substantially improved earlier versions of this manuscript. Additionally we want to thank two anonymous reviewers for their comments on the manuscript. This project was financially supported by the Deutsche Forschungsgemeinschaft (DFG BE 3830/1-1).

Appendix A. Supplementary data


References


