

RESEARCH ARTICLE

Open Access

High intralocus variability and interlocus recombination promote immunological diversity in a minimal major histocompatibility system

Anthony B Wilson^{1,2,5*}, Camilla M Whittington^{1,3} and Angela Bahr^{1,4}

Abstract

Background: The genes of the major histocompatibility complex (MHC/MH) have attracted considerable scientific interest due to their exceptional levels of variability and important function as part of the adaptive immune system. Despite a large number of studies on MH class II diversity of both model and non-model organisms, most research has focused on patterns of genetic variability at individual loci, failing to capture the functional diversity of the biologically active dimeric molecule. Here, we take a systematic approach to the study of MH variation, analyzing patterns of genetic variation at MH class II α and II β loci of the seahorse, which together form the immunologically active peptide binding cleft of the MH class II molecule.

Results: The seahorse carries a minimal class II system, consisting of single copies of both MH class II α and II β , which are physically linked and inherited in a Mendelian fashion. Both genes are ubiquitously expressed and detectable in the brood pouch of male seahorses throughout pregnancy. Genetic variability of the two genes is high, dominated by non-synonymous variation concentrated in their peptide-binding regions. Coding variation outside these regions is negligible, a pattern thought to be driven by intra- and interlocus recombination. Despite the tight physical linkage of MH II α and II β loci, recombination has produced novel composite alleles, increasing functional diversity at sites responsible for antigen recognition.

Conclusions: Antigen recognition by the adaptive immune system of the seahorse is enhanced by high variability at both MH class II α and II β loci. Strong positive selection on sites involved in pathogen recognition, coupled with high levels of intra- and interlocus recombination, produce a patchwork pattern of genetic variation driven by genetic hitchhiking. Studies focusing on variation at individual MH loci may unintentionally overlook an important component of ecologically relevant variation.

Keywords: Gene conversion, Major histocompatibility complex, Natural selection, Recombination, Sexual selection, Syngnathidae

Background

The genes of the major histocompatibility complex (MHC) are an essential component of the adaptive immune system, and among the most variable loci in the vertebrate genome [1]. Classical major histocompatibility (MH) class I and II genes are responsible for the recognition of foreign

antigens [2,3], and the peptide-binding regions (PBR) of these genes represent some of the most striking examples of balancing selection [4]. MH class II molecules are protein dimers produced from the combination of class II α and II β genes, which together form the immunologically-active peptide-binding cleft involved in the recognition of extracellular pathogens [5]. The high variability of the peptide-binding regions of these two genes ensures that hosts can recognize a broad spectrum of pathogens, and these regions are consequently subject to strong positive selection.

Despite the fact that antigen recognition by MH class II is determined by a protein complex including both II α

* Correspondence: twilson@brooklyn.cuny.edu

¹Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

²Department of Biology, Brooklyn College and The Graduate Center, City University of New York, 2900 Bedford Avenue, Brooklyn, New York 11210, USA

Full list of author information is available at the end of the article

and I β molecules, the majority of studies have focused on only a single component of this complex (typically MH I β), neglecting the fact that antigen discrimination is likely to be realized via unique permutations of MH I α and I β alleles in the host. This gene-centric approach to the analysis of MH diversity has been premised on the assumption that allelic variability at class I α loci is negligible relative to that found at class I β genes. While genetic variability of MH class I α is comparatively low in human populations [6], this hypothesis has not been well investigated in other vertebrate taxa.

The majority of vertebrates carry multiple copies of classical MH loci [7], complicating efforts to study intra- and interlocus allelic variability in natural populations. In contrast to the tightly linked structure of major histocompatibility genes in mammals, MH genes are unlinked in teleost fishes, and members of this group show considerable variation in the number and structure of MH class I and II loci [8], providing a powerful comparative model for the study of evolutionary immunology. A recent study comparing patterns of allelic variability of MH I α and I β loci in salmonid fishes with a minimal MH class II system found high levels of genetic variation at both loci, diversity that could serve to maintain the efficiency of pathogen recognition despite the absence of additional gene copies [9]. Unfortunately, given the relative paucity of studies directly investigating allelic variability of both components of the MH II molecule, it remains unclear whether the high genetic diversity found at both loci in salmonids is a lineage-specific phenomenon associated with their minimal MH system, or whether it reflects a broader pattern relevant for understanding MH class II function in other taxa.

Seahorses (*Hippocampus* spp.) are a genus of marine fishes well known for their unique form of reproduction, male pregnancy, in which males provide all parental care after fertilization in highly specialized brooding structures located on their ventral surface [10]. Recent research on the pot-bellied seahorse (*H. abdominalis*) found evidence of a single MH I β locus with high allelic diversity in this species [11], and a highly skewed distribution of non-synonymous variation across the gene [12]. Behavioral studies of *H. abdominalis* indicate that females can discriminate male MH I β genotype using olfactory cues, and use this information when making mate choice decisions [13]. Males, in contrast, show a lack of MH I β olfactory discrimination.

Transcriptome profiling of the seahorse indicates that this species carries a minimal MH class II system, with single expressed copies of both MH I α and β . Intriguingly, while these genes are intact in the seahorse, pipefish (*Syngnathus typhle*) have apparently lost MH class II genes as well as CD4⁺, a critical component of MH II-associated T-helper cells [14]. Given the close

relationship between seahorses and pipefish [15,16], the loss of MH II function in the pipefish appears to be a relatively recent evolutionary event, one which would be expected to have an important impact on adaptive immune function in this group.

In this study, we identify the genomic structure and pattern of inheritance of major histocompatibility class II genes in the seahorse, and use population-level screening to examine the distribution of genetic variation across their peptide-binding regions. We characterize tissue-specific patterns of expression of the two loci, with particular emphasis on the male brood pouch, the reproductive organ associated with male pregnancy in this species. Finally, MH I α and I β data are used to reconstruct the functionally important peptide-binding cleft of the molecule via homology modelling to the known crystallographic structure of the mature protein. Despite the tight physical linkage between MH I α and I β in the seahorse, composite allele profiles reveal high levels of functional diversity at both loci, and a strong signature of intra- and interlocus recombination, insights essential for understanding the structure, function and evolution of the biologically-active protein complex.

Results

The seahorse possesses a minimal MH II system

Full length cDNA and gDNA amplification of MH I α in the seahorse revealed the existence of a single four exon MH I α locus in this species. PCR reactions designed to bridge the full MH class II gene region produced two PCR products in an individual heterozygous for MH I β , both of which were successfully cloned and sequenced using primers distributed across the length of the sequence (Table 1). Analysis of cloned products identified extended single bp and tandem repeats in intronic regions of both MH I α and I β , along with a large 786 bp indel located in the intergenic region (Figure 1). The full length MH class II region of the seahorse is 9,130 - 9,884 bp in length, and contains a single class I α and I β locus (Figure 1). Exonic variation is concentrated in the putative peptide binding regions (PBR) of the two MH loci (Figure 1), and while overall variation in coding regions is low, all exonic SNPs in the sequenced individual are non-synonymous (MH I α Exon 2 (PBR): 11/11 SNPs non-synonymous, MH I α Exon 4: 3/3, MH I β Exon 2 (PBR): 4/4, MH I β Exon 3: 1/1; MH I β Exon 5: 1/1).

MH I α and I β are constitutively expressed, and present in the brood pouch of both pregnant and non-pregnant male seahorses (Figure 2). RT-PCR analysis showed no differences in expression levels of either gene across different tissues or different stages of pregnancy.

Analysis of the MH class I α and I β PBR in five families of known parentage demonstrates that genetic variation at both loci is inherited in a Mendelian fashion (Figure 3).

Table 1 Primers (5' → 3') used to amplify and sequence major histocompatibility class II genes in *H. abdominalis*

#	Name	Sequence 5'-3'	Location
1	MHIIα-E1F3	TTACTCCGTTGCGGCGACGCC	Exon 1
2	MHIIα-I1R	CGTCTGTACAATACTTCCGTAC	Intron 1
3	MHIIα-I1F	CAGTTACCAGGACAATGAC	Intron 1
4	MHIIα-I1F2	CAAAAAGCGGTGCTTATCGAG	Intron 1
5	MHIIα-I1R2	GCTTTAATTGAGATACAAGTACC	Intron 1
6	MHIIα-I1R3	CACCAACAGTAAAAACCAAG	Intron 1
7	MHIIα-I1F3	CTTGCGGGGAAACGGCGAAG	Intron 1
8	MHIIα-E2F*	GACGTCATCCACAGACATGC	Exon 2
9	MHIIα-E2R2	GATTCTAGTGGCTGATCCTCAATGC	Exon 2
10	MHIIα-E3R*	TGATGGGATGGGAGGAGGATC	Exon 3
11	MHIIα-E3F	AGACCTTCCAACCTGGACTTCAC	Exon 3
12	MHIIα-E3R2	GTGAAGTCCAGTTGAAGGTCTGG	Exon 3
13	MHIIα-I3R	CACGTTATTTATCATCGGCATAG	Intron 3
14	MHIIα-E4F	CATCAAAGGAAACCAATGCAACTG	Exon 4
15	MHIIα-3UTRR	GTGAGGACATAATGTGCCACC	3' UTR
16	MHIIα-3UTRF	GGTGGCGCACATTATGCCTCAC	3' UTR
17	MHIIβ-5UTRR5	GCAATACACTGGGGTTCTCTG	5' UTR
18	MHIIβ-5UTRF2	CAGTGTTATGCGAACATGG	5' UTR
19	MHIIβ-5UTRR6	TATCGGTACTGGCCATCC	5' UTR
20	MHIIβ-5UTRF	GCGACCAAAATCCGTTCCAG	5' UTR
21	MHIIβ-5UTRR4	GGCGTGGTTGTGTGATTTACAC	5' UTR
22	MHIIβ-5UTRF3	GCTTTGTGCAACTAATTGTGC	5' UTR
23	MHIIβ-5UTRR2	GGATCCAGTCTAATCTGAATCCCG	5' UTR
24	MHIIβ-5UTRR	CCAACCTGAGCACAACTTG	5' UTR
25	MHIIβ-UTR5F	CAAGTTTGTGCTCAGTTGG	5' UTR
26	MHIIβ-E1R	AAGGTGAGGAAAAGGAGGC	Exon 1
27	MHIIβ-E1F2*	GCCTCCTTTCTCACCTTC	Exon 1
28	MHIIβ-E2R2	GATGTCAATTCTGGTCACTCGAGTT	Exon 2
29	MHIIβ-E2R*	GAGCGCACTTTCGTAGTCAA	Exon 2

*RT-PCR primers.

Analyses of allelic phase in parents and offspring support the tight physical linkage of MH IIα and β, with no evidence of germline recombination evident in the dataset (Figure 3).

Evidence of balancing selection on the MH IIα PBR

Direct sequencing of the putative MH IIα peptide-binding region in a population sample of 101 individuals identified a high level of functionally important genetic variation, with 18 alleles observed in the 246 bp region, including 25 substitutions at 24 polymorphic nucleotide sites. Nucleotide diversity (π) in the population sample equaled 0.030 ± 0.007 , similar to the level of genetic variation found at MH IIβ of the seahorse ($\pi = 0.031 \pm 0.007$) [11]. 84% of individuals were heterozygous for MH IIα (85 of 101), while 16% were homozygous, values consistent with Hardy-Weinberg expectations (HWE Exact Test: $p = 0.61$).

All 25 observed nucleotide substitutions in the MH IIα PBR were found to be non-synonymous, translating into amino acid substitutions at 22 of the 82 sites of the translated protein sequence, striking evidence of positive selection ($dN/dS: \infty$; $dN = 0.039$; $dS = 0.000$; Z-test for positive selection: $p < 0.001$; Table 2). A site-specific analysis of nucleotide variation including positive selection (M8) fit the data significantly better than a neutral model of molecular evolution (M7) ($LRT = 51.507$, $df = 2$, $p < 0.001$) and revealed significant evidence of positive selection at all 22 variable amino acid sites (Bayes Empirical Bayes Analysis; $p > 99\%$; Figure 4), a pattern consistent with that observed at the physically linked MH IIβ locus, where coding variation is also concentrated at peptide-binding sites [11]. Inferences based on a comparison of the M1a (Nearly Neutral) and M2a (Positive Selection) models in Codeml were identical (data not shown).

Network comparison reveals the existence of composite recombinants

An allelic network was constructed to visualize relationships among the 18 MH IIα alleles and their relative frequencies (Figure 5a). The structure of the MH IIα

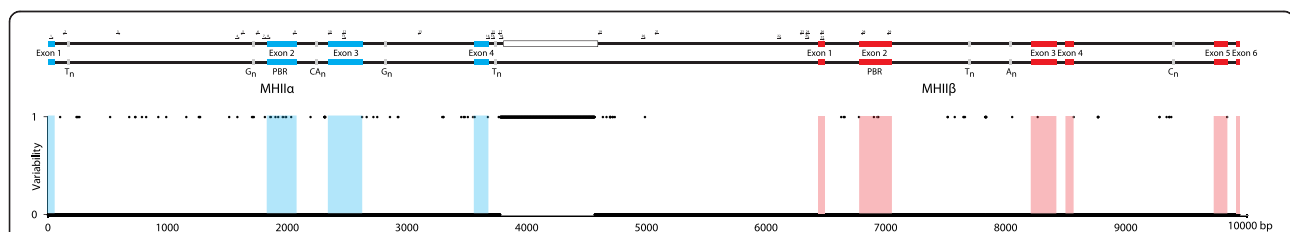


Figure 1 Sequence variability map of the MH class II region of the seahorse, based on the sequencing of alleles from a single heterozygous individual. Intron-exon structures of MH class IIα and IIβ are indicated as well as SNPs distinguishing the two full-length alleles (i.e. variability = 1). Note the presence of regions of repetitive sequence (A_n , C_n , CA_n , G_n and T_n) as well as a 786 bp deletion in the region between the IIα and IIβ loci. Primers used in the amplification and sequencing of this region are indicated (see Table 1 for primer sequences). Full length sequences have been deposited in GenBank (KP259908, KP259909).

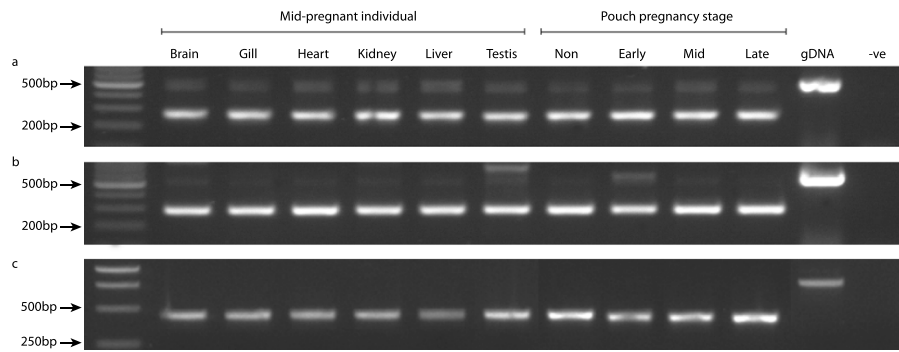


Figure 2 MH class II α and II β genes are ubiquitously expressed in the seahorse. Gel images showing results of RT-PCR on a range of seahorse tissues: (a) MH class II α , (b) MH class II β , (c) Beta-actin (positive control).

network is similar to that observed at MH II β , with a small number of common alleles (DAA-E2*02, 03, 05 and 10), and a larger number of low frequency alleles present in ≤ 10 copies in the population sample (Figure 5).

An analysis of genetic linkage between MH II α and II β alleles in the population sample supported the results of the parentage analysis, and indicated significant linkage disequilibrium between the two loci ($\chi^2 = 469.70$; $p < 0.001$). With 18 alleles detected at MH II α and 17 alleles at MH II β , a minimum of 18 composite alleles and maximum of $18 * 17 = 306$ alleles could be realized in the

population sample. The observation of 36 composite MH II α /II β alleles in the population dataset (Figure 5c) reflects a significant history of recombination in this region.

Evidence of intra- and interlocus recombination

A statistical analysis of recombination at MH II α failed to detect significant evidence of intralocus recombination ($p = 0.077$). An analysis of the phased MH II α /II β dataset, in contrast, found a strong signature of recombination ($p = 0.0002$), with 8 MH II α /II β recombinants and 2 of the 3 MH II β intralocus recombinant genotypes

Individual	Family A		Family B		Family C		Family D		Family E	
	MH II α	MH II β	MH II α	MH II β	MH II α	MH II β	MH II α	MH II β	MH II α	MH II β
Father	2	4	2	6	3	13	2	4	3	2
	3	3	3	2	8	5	2	9	2	6
Mother	2	4	3	1	2	8	2	6	3	1
	2	15	2	8	3	2	7	16	6	11
Juvenile 1	2	4	2	6	3	13	2	4	2	6
	2	4	3	1	3	2	2	6	6	11
Juvenile 2	3	3	2	6	8	5	2	4	3	2
	2	15	2	8	3	2	7	16	6	11
Juvenile 3	2	4	2	6	8	5	2	9	2	6
	2	15	3	1	3	2	7	16	3	1
Juvenile 4	2	4	2	6	3	13	2	4	3	2
	2	15	3	1	2	8	2	6	6	11
Juvenile 5	3	3	2	6	8	5	2	9	2	6
	2	15	2	8	2	8	7	16	3	1
Juvenile 6	3	3	3	2	3	13	2	9	3	2
	2	15	3	1	2	8	2	6	3	1
Juvenile 7	3	3	2	6	3	13	2	9	3	2
	2	4	2	8	2	8	2	6	3	1
Juvenile 8	2	4	2	6	3	13	2	4	3	2
	2	4	2	8	2	8	7	16	6	11
Juvenile 9	3	3	3	2	8	5				
	2	15	3	1	2	8				
Juvenile 10	3	3								
	2	15								
Juvenile 11	3	3								
	2	4								
Juvenile 12	3	3								
	2	15								
Juvenile 13	2	4								
	2	15								

Figure 3 Mendelian inheritance of MH class II alleles in five families of known parentage. Parental alleles are color-coded for ease of discrimination in offspring (Paternal alleles: light/dark blue; Maternal alleles: light/dark red). Blockwise inheritance of parental MH alleles provides direct evidence of tight physical linkage.

Table 2 Synonymous and non-synonymous substitution rates for exon 2 alleles of the seahorse MH class IIa gene

Locus	Length (bp)	Samples	Alleles	dN	dS	dN/dS
Exon 2	246	101	18	0.039	0.000	∞^{**}
Exon 2, PBS	60	101	17	0.110	0.000	∞^*
Exon 2, non-PBS	186	101	11	0.018	0.000	∞^*

Probabilities (* <0.05 , ** <0.001 , ns = not significant) are derived from a Z-test (H1 = positive selection). Peptide-binding sites (PBS) refer to the human sites, identified by crystallographic analysis [3] (Figure 4).

identified in an earlier analysis of MH II β PBR variation [11] (Table 3).

Protein structure

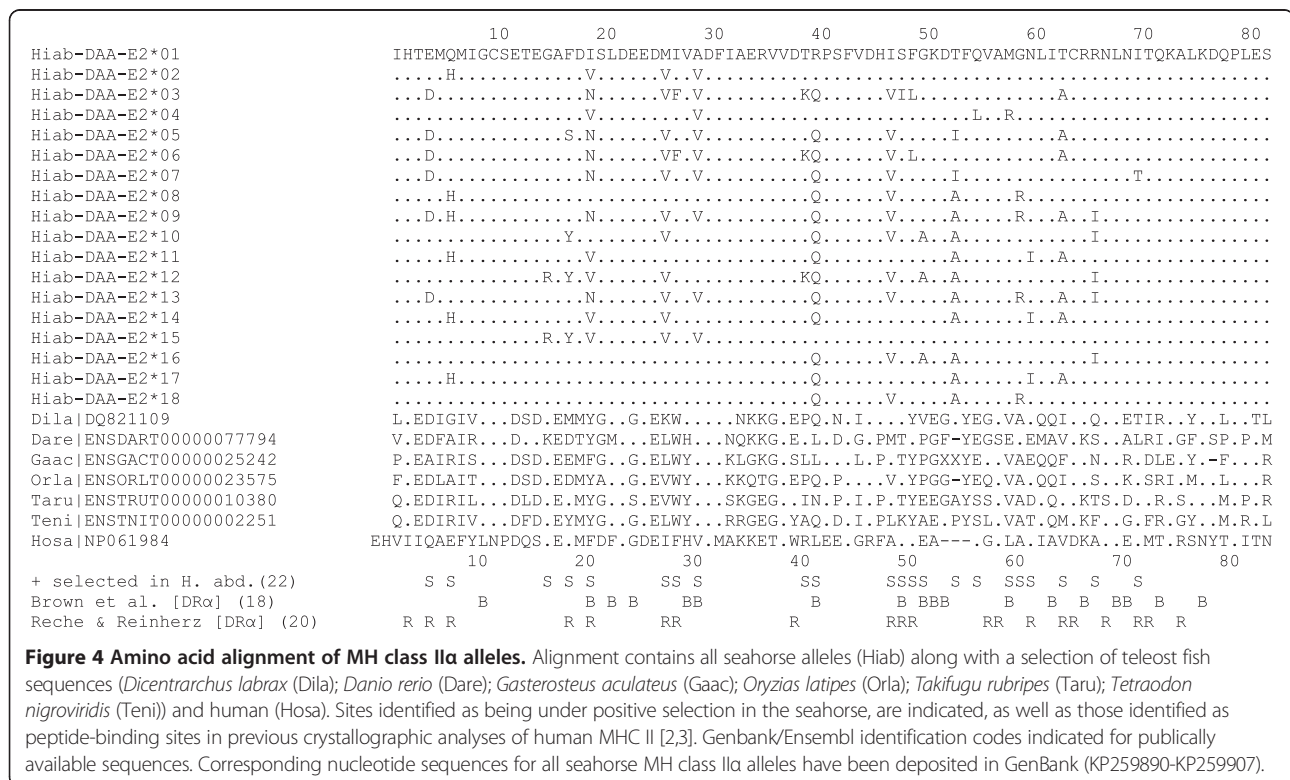
The inferred quaternary structure of the seahorse MH class II molecule closely resembles that of the mouse model, with a clearly defined peptide binding groove formed by the interaction of MH class II α and II β loci (Figure 6). While the majority of positively selected sites in the seahorse are located in this peptide-binding groove, several selected sites at both the II α (LYS.39 α and GLN.40 α) and II β loci (LEU.18 β and ASN.61 β) are located away from this groove on the surface of the mature peptide.

Discussion

Adaptive immune diversity at the minimal MH class II system of the seahorse is enhanced by high levels of genetic variation at both MH II α and II β , highlighting the importance of considering both loci in order to understand

MH class II function in this system. Intra- and interlocus recombination act to further augment genotypic variation, generating novel genotypes despite the tight physical linkage of MH class II loci in the seahorse.

Despite high levels of functionally important genetic diversity and clear evidence of positive selection in the PBR of MH II α , this region shows a complete absence of synonymous substitutions. The significant excess of non-synonymous substitutions relative to synonymous variation is also characteristic of the PBR of MH II β [11], and low levels of genetic variation are found in the other coding regions of the two genes [Figure 1; 12]. While the high number of non-synonymous variants in the PBR is expected given its role in antigen recognition, the low levels of synonymous substitutions suggest that selective sweeps driven by positively selected sites may act to homogenize neutral variation through a process of genetic hitchhiking [12]. New mutations at sites involved in antigen detection that provide a selective advantage in host-pathogen interactions are expected to rapidly increase in frequency [17],



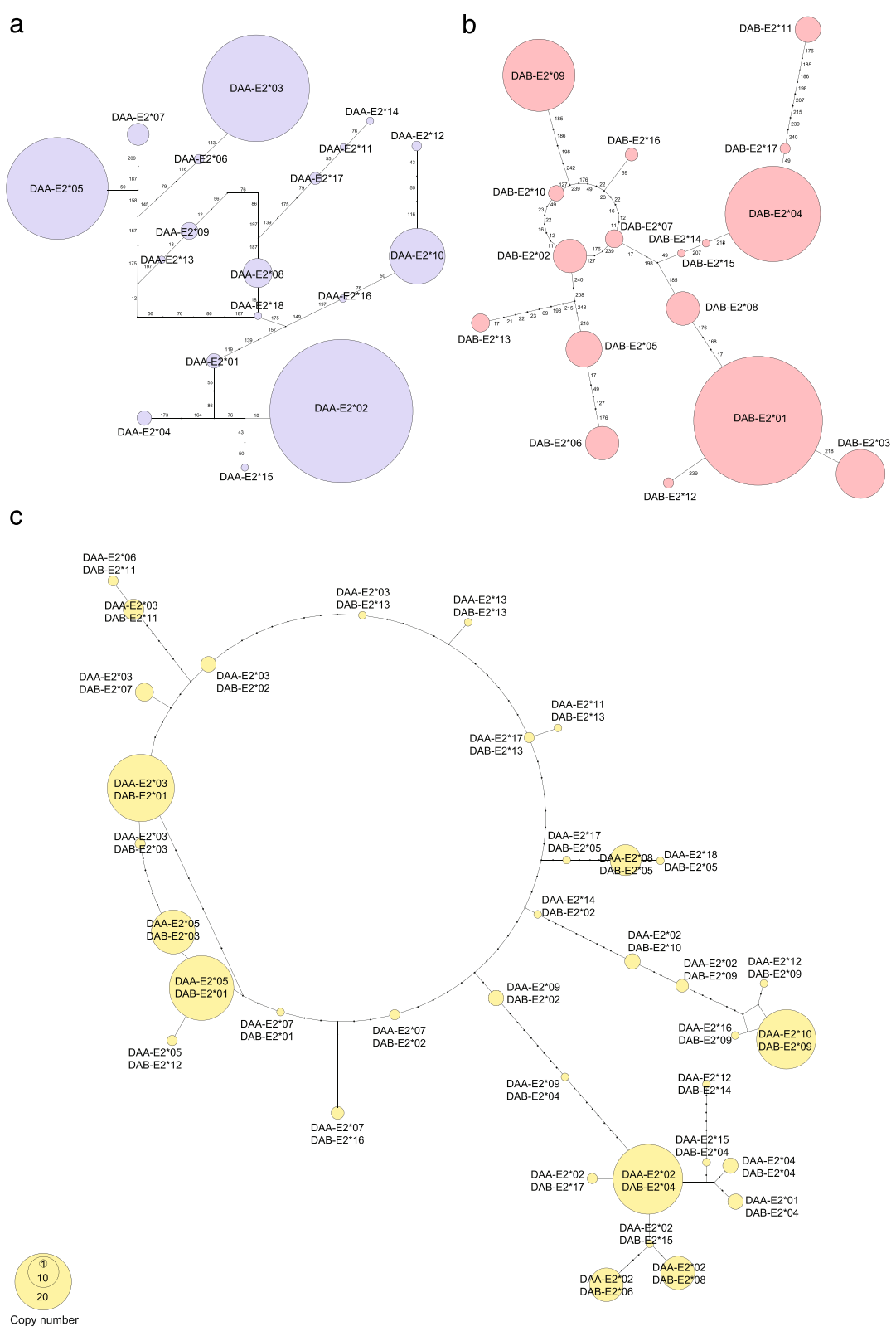


Figure 5 Allelic networks of MH class II variation (n = 101 individuals). Figure shows allelic variability in the pattern of non-synonymous substitutions across exon 2 of (a) MH class II α and (b) MH class II β , the immunologically-active region of the two genes; (c) Composite MH network constructed from phased genotype data. Note the presence of reticulation in the MH II β network, evidence of intralocus recombination [11]. Reticulation is also evident in the composite allele network.

simultaneously increasing the frequency of linked neutral variants. Subsequent short-tract gene conversion and recombination may enhance functional diversity at MH peptide binding sites, without increasing neutral genetic variation. Over time, the hitchhiking of linked genetic variants with positively selected sites may erode local neutral variation, and impact functionally important genetic variation at neighboring loci. The combination of positive selection and genetic hitchhiking is likely responsible for the highly skewed dN/dS ratios observed here.

Amino acids experiencing positive selection at both MH II α and II β are preferentially located at sites predicted to be directly involved in antigen binding (Figure 4), as inferred from homology modelling to the experimentally determined crystallographic structures of human and mouse MHC [3,18]. The presence of several positively selected sites outside this region (Figure 6) suggests potential differences in the quaternary structure of the mature protein *in vivo* and/or additional selective constraints acting away from the main peptide binding groove of the seahorse.

Despite the tight physical linkage of MH class II loci in the seahorse, both intra- and interlocus recombinants were detected in the population sample, a finding which likely reflects the selective advantages of novel allelic variants generated by recombination. While empirical recombination rates are not available for the seahorse, a rough approximation based on the average recombination rate in the human genome (ca. 1 cM/Mb) would suggest that a crossover event in the 5,000 bp interval separating the PBRs of MH II α and II β would be expected every 20,000 meioses. Recombination is likely to be even more common in the MH region, as hotspots have been identified in the MH class II regions of humans [19] and mice [20] that have been shown to provide an

important source of allelic variation at these loci. Sperm profiling indicates that gene conversion, the unidirectional exchange of short sequence tracts without crossover, may make up as much as 90% of all recombination events at MH class II loci [21], and serve a particularly important role in generating new PBR variants from standing genetic variation. Once new variants are produced, they may rapidly spread through the population if they offer selective advantages relative to existing allelic variation [17].

While analyses of linkage disequilibrium in population samples provide evidence of breakpoints of historical recombination events, the estimation of recombination rates requires the identification of germline recombinants from pedigree data and/or the direct typing of gametes. No MH II α /II β recombinants were identified in the parentage analysis carried out as part of this study, but the modest size of this dataset had limited power to identify meiotic recombination. Although seahorses have reduced testes and produce low amounts of sperm [22], sperm typing could provide a high throughput method to estimate meiotic recombination rates, and to explore the relative importance of gene conversion and classical recombination in structuring patterns of genetic variation across the major histocompatibility region of the seahorse. An understanding of baseline recombination rates would also facilitate more detailed analyses of the potential selective advantages of novel recombinant genotypes.

Studies of MH class II genetic variability in non-model organisms have been heavily biased towards MH II β , an approach which has been informed by the low variability of MH II α relative to MH II β in humans. Large scale population genotyping of the 5 classical MH II α and II β genes in more than 20,000 individuals has identified >2,750 MH II β alleles and only 116 MH II α variants, a greater than 20 \times excess of MH II β allelic diversity relative to MH II α [6].

Gomez [9] carried out the first integrative analysis of MH II α and II β variation in a non-model species, simultaneously characterizing genetic variation at both loci of the minimal MH system of salmonid fishes. An analysis of genetic variation in population samples of three salmonid species revealed similar levels of variation at both loci (MH II α /II β alleles - *Oncorhynchus mykiss* (n = 40): 10/11; *Oncorhynchus kisutch* (n = 75): 4/8; *Salmo salar* (n = 27): 9/9), a result which led the authors to suggest that high genetic variability at MH II α could allow species with a minimal MH system to mount a successful immune response without the additional genetic variation associated with multiple gene copies [9].

Recent work suggests that the high genetic variability of MH II α may not be restricted to species with a minimal MH system such as salmonids or seahorses,

Table 3 Inferred recombinant genotypes from analysis of the phased MH class II α /II β PBR dataset

Genotype	Recombination breakpoint	Mutation savings	p-val
DAA-E2*9/DAB-E2*4	198-4896	9	0.0001
DAA-E2*3/DAB-E2*13	180-4896	9	0.0002
DAA-E2*12/DAB-E2*14	188-4890	7	0.0002
DAA-E2*13/DAB-E2*13	198-4890	6	0.0001
DAA-E2*7/DAB-E2*2	210-5007	6	0.0016
DAA-E2*2/DAB-E2*6	5058-5079	6	0.0004
DAA-E2*9/DAB-E2*2	198-4896	6	0.0001
DAA-E2*2/DAB-E2*9	193-5007	5	0.0021
DAA-E2*2/DAB-E2*10	4930-5007	5	0.0001
DAA-E2*14/DAB-E2*2	180-4896	5	0.0053

Note: Recombination breakpoints for interlocus recombinants include 4630 bp separating the sequenced peptide binding regions of MHII α and II β (see Figure 1).

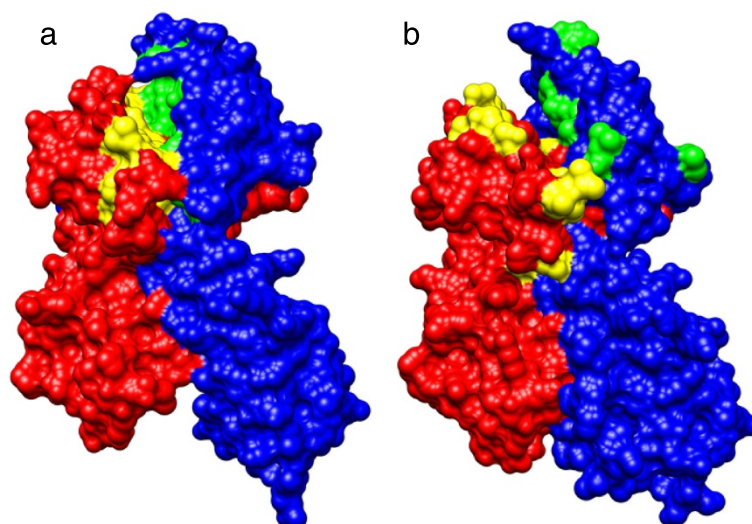


Figure 6 Positively selected sites of the seahorse MH class II are concentrated in the inferred peptide binding groove of the mature protein. Quaternary structure models of the major histocompatibility class II molecule for **(a)** *Mus musculus* and **(b)** *Hippocampus abdominalis*, based on crystallographic reconstructions for *Mus* (pdb1es0). Surface structures of extracellular components of MH II α (red) and MH II β (blue) are shown, as are **(a)** peptide-binding sites, and **(b)** sites inferred to be under positive selection in the seahorse (yellow and green for MH II α and II β respectively).

but may be evolutionarily widespread. A recent population survey ($n = 367$ wild-caught individuals) of single copies of MH II α (H-2Aa) and II β (H-2-Eb) in the two locus system of the house mouse (*Mus musculus*) recovered 27 MH II α and 17 MH II β alleles, an excess of MH II α allelic variability relative to MH II β [23]. Similarly, a large genotyping study of two MH loci in Japanese domestic cattle ($n = 650$) detected 15 and 26 alleles at MH II α (DQA1) and MH II β (DRB3) loci, respectively [24]. Taken together, the results of these recent studies suggest that high genetic diversity at MH II α is likely to be widespread, and is not a phenomenon restricted to species with a minimal MH class II system. As such, studies which continue to focus exclusively on variation at MH II β may be missing an important component of ecologically relevant variation, providing an incomplete picture of how selective pressures influence genetic diversity of the biologically active protein dimer.

Unfortunately, while it is relatively straightforward to characterize MH diversity using traditional Sanger sequencing in species carrying a minimal MH system, researchers working on more complex MH systems have been unable to accurately screen MH allelic diversity and classify variation to specific loci, an essential prerequisite for more detailed investigations of selection. Recent innovations in next-generation sequencing [25], coupled with rapidly maturing statistical approaches for processing population data for multigene families such as the MHC [26,27], offer an exciting new avenue for high-throughput systems-based analyses of immunological variation. While

such studies are still in their infancy, they show great promise for population-level analyses of MH loci e.g. [28-30]. Thus far, studies have followed the design of earlier research in focusing on MH II β , but combined analyses of MH II α and II β should be straightforward once these new technologies have fully matured.

Conclusions

Antigen recognition by the adaptive immune system of the seahorse is enhanced by high levels of variability at both MH class II α and II β loci. Strong positive selection on sites involved in pathogen recognition, coupled with high levels of intra- and interlocus recombination, generate a patchwork distribution of genetic variation essential for understanding the functional diversity of the mature protein.

While studies of MH diversity in non-model organisms are now widespread, the majority of investigations have focused exclusively on MH class II β , an approach that has been justified by the relatively low levels of variation observed at MH class II α loci in humans. Our results suggest that humans may be atypical in this regard, a conclusion supported by recent comparative analyses of MH class II diversity in both mammals and fish, which indicate that genetic diversity at MH class II α loci may often meet or exceed that observed at class II β . Given the high rates of intra- and interlocus recombination found at MH loci, studies aiming to link patterns of MH class II diversity to survival and reproduction should consider the structure and function of the immunologically active protein molecule. In failing to do so,

researchers may be unintentionally overlooking an important component of ecologically relevant genetic variation.

Methods

Sample collection

Adult seahorses (*H. abdominalis*) were purchased from a commercial breeding facility (Seahorse Australia, Beauty Point, Tasmania) and held at the University of Zurich under an animal care and experimentation permit from the Veterinäramt Zürich (Permit 103/2008). In addition, fin clips from 5 wild-caught seahorses from Sydney, Australia (2 individuals collected in 2003) and Tasmania (3 individuals collected from 3 populations in 2003 and 2004) [31] were included in analyses of MH class II genetic diversity.

Gene discovery via high-throughput transcriptome profiling

A full plate of 454 sequencing of transcriptome libraries prepared from pouch and reference (brain, gills, liver, heart, kidney and testes) tissues from a single pregnant and non-pregnant seahorse using the GS FLX Titanium Chemistry (Roche) recovered six partial transcripts of the MH II α locus of the seahorse, which were assembled into a single 541 bp contig spanning the 5' UTR, complete exons 1 and 2 and partial exon 3. Further details on this transcriptome screen are available elsewhere [11]. A TBLASTX search of the nucleotide collection of GenBank identified MH class II α of *Dicentrarchus labrax* as the top hit (DQ821109.1: e-value = 3e-47), followed by class II α genes from other teleost species (*Morone saxatilis* L35062.1: e-value = 8e-42, *Larimichthys crocea* EF681861.1: e-value = 5e-41, *Miichthys miiuy* GU936787.1: e-value = 2e-40, and *Epinephelus coioides* GU992883.1: e-value = 3e-40).

Full-length cDNA sequencing

Total RNA was extracted from brain, gill, kidney, liver, pouch and testes tissues of two pregnant seahorses with RNeasy extraction columns (Qiagen). RNA extractions were subsequently DNase treated, standardized to a common concentration of 85 ng, and pooled for library construction. 5' and 3' RACE libraries were prepared from 1 μ g of total RNA using a SMARTer RACE cDNA amplification kit (Clontech).

The full-length cDNA sequence of the MH class II α gene of the seahorse was obtained using separate 5' and 3' RACE reactions primed with gene-specific primers (MHII α E2F and MHII α E3R; Table 1), in 25 μ L PCR reactions using 1.5 μ L of 5'/3' RACE-ready cDNA. Both reactions produced single products from the 5' and 3' ends of the class II α gene, which were PCR-purified with Montage PCR μ ₉₆ Filter Plates (Millipore) and eluted in 20 μ L ddH₂O in preparation for sequencing.

Sequencing reactions were carried out in 10 μ L volumes consisting of 2–4 μ L purified PCR product, 0.2 μ M primer, and 1 μ L Big Dye v3.1 Terminator Cycle Sequencing mixture (Applied Biosystems). Cycling conditions consisted of 30 cycles of 10 sec at 96°C, 5 sec at 50°C and 4 min at 60°C. Ethanol-purified products were sequenced on an ABI 3730 automated sequencer (Applied Biosystems).

Full-length genomic sequencing of the Major Histocompatibility class II gene region

To elucidate the structure and distribution of variation across the MH region of the seahorse, genomic DNA of a single non-pregnant individual was extracted from muscle tissue (DNeasy, Qiagen). The quality of extracted DNA was assessed on a 1.5% agarose gel and spectrophotometrically quantified using a Nanodrop 2000 (Thermo Scientific). The gDNA sequence of the MH class II β for this individual has been previously published [12].

The full-length gDNA sequence of the seahorse MH class II gene region was determined using long-range PCR with gene-specific primers designed from the cDNA sequence of MH class II α and the 5' UTR region of the class II β locus (MHII α -E1F3/MHII β -5UTRR4; Table 1).

Long range PCR was performed in a 25 μ L volume containing 3U LongAmp Taq (New England Biolabs), 1 \times LongAmp reaction buffer, 0.4 mM dNTPs, 0.2 μ M primers and 250 ng DNA. PCR amplification involved a 2 min denaturation step at 92°C, followed by 30 cycles of 92°C (20 s), 65°C (20 s) and 65°C (10 min), and a final extension step for 10 min at 65°C. The PCR reaction was filter-purified in preparation for cloning.

4 μ L of purified PCR product was cloned into a TOPO TA cloning vector (Invitrogen) following the manufacturers' recommendations. Following overnight culture of transformed chemically competent *E. coli* at 37°C, 5 positive colonies were picked and grown for 16 h in liquid culture on a 200 rpm horizontal shaker at 37°C. Liquid cultures were purified for downstream sequencing using a QIAprep Spin Miniprep kit (Qiagen).

Screening of plasmid DNA with primers for the hyper-variable peptide binding region of the seahorse MH class II α peptide binding region revealed the presence of two alleles, both of which were sequenced to completion using a nested sequencing strategy involving primers distributed across the full length of the amplified region (Table 1) using the protocols outlined above. DNA sequencing revealed a 786 bp deletion between MH class II α and II β in one of the two full-length alleles (Figure 1).

The intervening non-coding sequence between the class II α and II β loci was PCR-amplified and sequenced from genomic DNA using two sets of PCR reactions, one using MHII α -E4F/MHII β -5UTRR2 (60°C anneal) and a second using MHII β -5UTRF2/MHII β -E2R2 (55°C

anneal), both of which spanned the deletion region, allowing the determination of allelic phase of sequences from $II\alpha$ and $II\beta$ loci.

RT-PCR screening of tissue-specific expression

Samples were obtained from captive *H. abdominalis* individuals (Seahorse Australia, Beauty Point, Australia) preserved in RNAlater (Sigma-Aldrich) and then stored at -80°C . Four reproductively active adult males were screened for MH II gene activity, with the stage of pregnancy estimated using a recently published developmental key for syngnathid fishes [32]. Total RNA was extracted from a panel of tissues (brain, gill, heart, kidney, liver, pouch, testis from one mid-pregnant animal; and pouch from one non-pregnant, one early pregnant, and one late pregnant individual) using an RNeasy Mini Kit with QiaShredder (Qiagen) and DNase I (Invitrogen) digestion. First-strand cDNA was synthesized with SuperScript III First-Strand Synthesis System for RT-PCR (Invitrogen) using random hexamer priming and 200 ng of RNA.

MH class II PCRs were carried out using intron-spanning primers indicated in Table 1. Beta-actin (ACTB) was used as the positive control to ensure uniform amplification for each tissue, and was amplified using ACTB-E2F - GTCATGGTCGGCATGGGAC and ACTB-E3R - AGGTAGTCTGTGAGGTCTCG. PCR reactions for ACTB were performed in 20 μl volumes containing 0.5U Taq (New England Biolabs), 1 \times NEB reaction buffer, 0.75 μM MgCl_2 , 0.2 mM dNTPs, 0.2 μM primers, and 1.0 μl DNA, with the following PCR cycling conditions: 95°C for 1:30, then 30 cycles of 95°C for 0:30, 53°C for 0:30, 68°C for 1:00, followed by 68°C for 5:00. MH amplifications were performed in 25 μl volumes containing 1U Taq (NEB), 1 \times NEB reaction buffer, 1.0 μM MgCl_2 , 0.4 mM dNTPs, 0.2 μM primers and 2.5 μl DNA, with the following cycling parameters: 92°C for 0:10, then 40 cycles of 92°C for 0:10, 55°C for 0:30, 68°C for 2:00 (for MH $II\alpha$) or 92°C for 5:00, then 40 cycles of 92°C for 0:30, 62°C for 0:30, 68°C for 4:00, followed by 68°C for 15:00 (for MH $II\beta$).

PCR products were subjected to electrophoresis at 100 V for 20 min in 1.5% agarose gels stained with ethidium bromide, and visualized using an Alphamager gel documentation system (Alpha Innotech).

MH $II\alpha$ inheritance and MH II linkage analysis

MH class $II\alpha$ Exon 2, containing the immunologically active peptide-binding region of the gene, was PCR-amplified and sequenced in a sample of 47 F1 individuals from 5 families ($n = 8\text{--}13$ per family) which had previously been characterized for patterns of genetic variation at the MH $II\beta$ peptide-binding region [11]. A comparison of parent-offspring genotype profiles allowed the inference

of the mode of MH $II\alpha$ inheritance and a means to test for linkage of $II\alpha$ and $II\beta$ loci in this species.

Standard PCR was performed in 25 μL volumes containing 1U Taq (NEB), 1 \times NEB reaction buffer, 1.0 μM MgCl_2 , 0.4 mM dNTPs, 0.2 μM of either MH $II\alpha$ -E2F/MH $II\alpha$ -E3R or MH $II\alpha$ -I1F3/MH $II\alpha$ -E3R and 25–250 ng DNA. PCR amplifications involved a 10 s denaturation step at 92°C , followed by 40 cycles of 92°C (10 s), 55°C (30 s) and 68°C (2 min). All individuals were PCR-amplified and sequenced using both sets of primer pairs. PCR amplifications of both PCR purification and sequencing followed that outlined for the cDNA experiment above, producing the full length sequence of the 249 bp exon. After trimming 2 bp from the 5' end of the sequence alignment and 1 bp from the 3' terminus to exclude incomplete amino acids, the analyzed exon 2 dataset included 246 bp/82 amino acids.

Characterization of the MH $II\alpha$ peptide-binding region (PBR)

Exon 2 of MH class $II\alpha$ was also sequenced in a population of 101 seahorse individuals for which the peptide-binding region of the MH class $II\beta$ locus had previously been characterized [11] to obtain an estimate of population-level variability of this region. PCR amplification and sequencing conditions were identical to those outlined above.

Sequence processing

All PCR reactions were sequenced in both directions, aligned using ClustalW [33] and visualized in BioEdit v.7.0.9 [34]. Heterozygous sites were coded using IUPAC nomenclature for degenerate positions, and allelic sequences were inferred using the default settings of PHASE V2.1.1 [35]. Individuals for which allelic phase could not be reliably inferred by statistical inference (Phase probabilities ≥ 0.95) were re-amplified and cloned (MH $II\alpha$: 1 individual, MH $II\beta$: 1 individual, MH $II\alpha$ /MH $II\beta$: 4 individuals). Four to five colonies were sequenced from each cloned individual, allowing the direct determination of individual alleles. All private alleles were separately re-amplified and sequenced to verify their identity (MH $II\alpha$: 6 individuals, MH $II\beta$: 2 individuals, MH $II\alpha$ /MH $II\beta$: 13 individuals).

Analyses of sequence polymorphism/linkage

Nucleotide diversity (π) at MH $II\alpha$ was estimated under the maximum composite likelihood model implemented in Mega v6.0 [36], with standard error estimates derived from 500 bootstrap replicates. Exact tests of Hardy-Weinberg equilibrium (1,000,000 step Markov Chain, 100,000 dememorization steps) were performed in Arlequin v3.5.1.2 [37].

Analysis of gametic phase of MH $II\alpha$ and $II\beta$ genotypes was performed using the Bayesian ELB approach [38] implemented in Arlequin v.3.5.12. Pairwise linkage

analysis of unphased MH II α and II β data was also carried out using Arlequin (20,000 permutations, 5 EM replicates).

Site-specific tests of positive selection

Characterization of synonymous and non-synonymous substitutions across the peptide binding region of MH II α was performed in Mega v6.0 [36] using the Nei-Gojobori method with Jukes-Cantor distances. A one-tailed Z-test of positive selection (500 bootstrap replicates) tested the null hypothesis of neutral evolution for putative peptide binding sites, non-binding sites, and the full peptide binding region.

A neighbor-joining tree was constructed from MH II α alleles using the maximum composite likelihood method implemented in Mega v6.0 [36]. This tree served as a starting tree for a site-specific analysis of positive selection in Codeml v4.8 [39], which compared the fit of a neutral evolution model with recombination (M7) with one allowing for positive selection (M8), using a likelihood-ratio test (LRT). Sites experiencing positive selection were identified following a Bayes Empirical Bayes analysis (posterior probability ≥ 0.95) [40].

Network construction

An allelic network was constructed to visualize genetic relationships among alleles of the MH II α PBR using TCS v.1.21 [41], and prepared for publication using yED v3.12.2 [42]. A second network was constructed using phased MH II α and II β data, in order to visualize the frequency and distribution of MH class II composite genotypes.

Recombination

The presence of recombination in the seahorse MH class II region was investigated using RECCO v.0.93 [43] (10,000 permutations), using a minimum mutation savings criterion of 5 to identify recombinants. Recombination analyses were carried out independently for the MH II α PBR dataset, and for a concatenated alignment of phased MH II α /II β data, allowing the identification of intra- and interlocus recombination. Inferred recombination breakpoints for interlocus recombinants included an offset of 4,630 bp of unknown sequence separating the PBRs of the two genes (Figure 1), reflecting uncertainty in the location of breakpoints in the unsequenced region between the two PBRs.

Protein structure

The quaternary structure of the MH class II complex of the seahorse was reconstructed via homology modeling of the full-length MH class II α and II β loci to the previously determined crystallographic structure of the mouse MH class II molecule, using Protinfo PPC [44]. Inferred

protein surface models of target and database sequences were annotated and visualized in Chimera v1.6.2 [45].

Protinfo PPC returned five significant hits (structure confidence: 41-55%), all of which matched PDB models for the extracellular domain of the MH class II complex of *Mus musculus*. One of the three top hits (PDB ID: 1ES0: structure confidence 55%) was selected as a model for the seahorse MH class II complex. The expression vector, peptide and linker of the mouse structure were omitted from the modeled data, as well as 8 aa of MH II β not resolved in the original model, resulting in a total of 182 aa and 180 aa for the MH class II α and II β loci, respectively. Known peptide-binding sites for the human MH class II molecule [3] were annotated on the mouse model, along with sites under positive selection in the seahorse.

Availability of supporting data section

Sequence data generated for this project have been deposited in GenBank (Accession #: KP259890-KP259909).

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

ABW designed the study, carried out the laboratory and statistical analyses and drafted the manuscript. CMW and AB contributed to the laboratory analyses and participated in the preparation of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Annemarie Honegger for discussions concerning homology modelling of the seahorse MHC. This research was supported with funding from the Swiss National Science Foundation, the University of Zurich and City University of New York.

Author details

¹Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland. ²Department of Biology, Brooklyn College and The Graduate Center, City University of New York, 2900 Bedford Avenue, Brooklyn, New York 11210, USA. ³School of Biological Sciences, University of Sydney, Heydon-Laurence Building A08, Sydney, NSW 2006, Australia. ⁴Institute of Medical Molecular Genetics, University of Zurich, Wagistrasse 12, 8952 Schlieren, Switzerland. ⁵Department of Biology, Brooklyn College, 2900 Bedford Avenue, Brooklyn, New York 11210, USA.

Received: 14 August 2014 Accepted: 12 December 2014

Published online: 20 December 2014

References

- Kelley J, Walter L, Trowsdale J: Comparative genomics of major histocompatibility complexes. *Immunogenetics* 2005, **56**:683–695.
- Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL, Wiley DC: Three-dimensional structure of the human class-II histocompatibility antigen HLA-DR1. *Nature* 1993, **364**:33–39.
- Reche PA, Reinherz EL: Sequence variability analysis of human class I and class II MHC molecules: functional and structural correlates of amino acid polymorphisms. *J Mol Biol* 2003, **331**:623–641.
- Hughes AL: Looking for Darwin in all the wrong places: the misguided quest for positive selection at the nucleotide sequence level. *Heredity* 2007, **99**:364–373.
- Yeager M, Hughes AL: Evolution of the mammalian MHC: natural selection, recombination, and convergent evolution. *Immunol Rev* 1999, **167**:45–58.

6. Robinson J, Mistry K, McWilliam H, Lopez R, Parham P, Marsh SGE: **The IMGT/HLA database.** *Nucl Acids Res* 2011, **39**(Database):D1171–D1176.
7. Flajnik MF, Kasahara M: **Comparative genomics of the MHC: glimpses into the evolution of the adaptive immune system.** *Immunity* 2001, **15**:351–362.
8. Wegner KM: **Historical and contemporary selection of teleost MHC genes: did we leave the past behind?** *J Fish Biol* 2008, **73**:2110–2132.
9. Gomez D, Conejeros P, Marshall SH, Consuegra S: **MHC evolution in three salmonid species: a comparison between class II alpha and beta genes.** *Immunogenetics* 2010, **62**:531–542.
10. Stolting KN, Wilson AB: **Male pregnancy in seahorses and pipefish: beyond the mammalian model.** *BioEssays* 2007, **29**:884–896.
11. Bahr A, Wilson AB: **The impact of sex-role reversal on the diversity of the major histocompatibility complex: insights from the seahorse (*Hippocampus abdominalis*).** *BMC Evol Biol* 2011, **11**:121.
12. Bahr A, Wilson AB: **The evolution of MHC diversity: evidence of intralocus gene conversion and recombination in a single-locus system.** *Gene* 2012, **497**:52–57.
13. Bahr A, Sommer S, Mattle B, Wilson AB: **Mutual mate choice in the potbellied seahorse (*Hippocampus abdominalis*).** *Behav Ecol* 2012, **23**:869–878.
14. Haase D, Roth O, Kalbe M, Schmiedeskamp G, Scharsack JP, Rosenstiel P, Reusch TBH: **Absence of major histocompatibility complex class II mediated immunity in pipefish, *Syngnathus typhle*: evidence from deep transcriptome sequencing.** *Biol Lett* 2013, **9**:20130044–20130044.
15. Wilson AB, Orr JW: **The evolutionary origins of Syngnathidae: pipefishes and seahorses.** *J Fish Biol* 2011, **78**:1603–1623.
16. Wilson AB, Ahnesjö I, Vincent ACJ, Meyer A: **The dynamics of male brooding, mating patterns, and sex roles in pipefishes and seahorses (family Syngnathidae).** *Evolution* 2003, **57**:1374–1386.
17. Ejsmond MJ, Radwan J, Wilson AB: **Sexual selection and the evolutionary dynamics of the major histocompatibility complex.** *Proc R Soc Biol* 2014, **281**:20141662.
18. McBeth C, Seamons A, Pizarro JC, Fleishman SJ, Baker D, Kortemme T, Gorman JM, Strong RK: **A new twist in TCR diversity revealed by a forbidden $\alpha\beta$ TCR.** *J Mol Biol* 2008, **375**:1306–1319.
19. Jeffreys AJ, Kauppi L, Neumann R: **Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex.** *Nat Genet* 2001, **29**:217–222.
20. Yauk CL, Bois PR, Jeffreys AJ: **High-resolution sperm typing of meiotic recombination in the mouse MHC E β gene.** *EMBO J* 2003, **22**:1389–1397.
21. Jeffreys AJ, May CA: **Intense and highly localized gene conversion activity in human meiotic crossover hot spots.** *Nat Genet* 2004, **36**:151–156.
22. Van Look KJW, Dzyuba B, Cliffe A, Koldewey HJ, Holt WW: **Dimorphic sperm and the unlikely route to fertilisation in the yellow seahorse.** *J Exp Biol* 2007, **210**:432–437.
23. Čížková D, de Belloq JG, Baird SJE, Piálek J, Bryja J: **Genetic structure and contrasting selection pattern at two major histocompatibility complex genes in wild house mouse populations.** *Heredity* 2011, **106**:727–740.
24. Miyasaka T, Takeshima S, Matsumoto Y, Kobayashi N, Matsuhashi T, Miyazaki Y, Tanabe Y, Ishibashi K, Sentsui H, Aida Y: **The diversity of bovine MHC class II DRB3 and DQA1 alleles in different herds of Japanese Black and Holstein cattle in Japan.** *Gene* 2011, **472**:42–49.
25. Babik W: **Methods for MHC genotyping in non-model vertebrates.** *Mol Ecol Resour* 2010, **10**:237–251.
26. Sommer S, Courtiol A, Mazzoni CJ: **MHC genotyping of non-model organisms using next-generation sequencing: a new methodology to deal with artefacts and allelic dropout.** *BMC Genomics* 2013, **14**:542.
27. Lighten J, van Oosterhout C, Bentzen P: **Critical review of NGS analyses for de novo genotyping multigene families.** *Mol Ecol* 2014:in press.
28. Huchard E, Albrecht C, Schliehe-Diecks S, Baniel A, Roos C, Kappeler PM, Peter PMK, Brameier M: **Large-scale MHC class II genotyping of a wild lemur population by next generation sequencing.** *Immunogenetics* 2012, **64**:895–913.
29. Pavey SA, Sevellec M, Adam W, Normandeau E, Lamaze FC, Gagnaire P-A, Filteau M, Hebert FO, Maaroufi H, Bernatchez L: **Nonparallelism in MHCII β diversity accompanies nonparallelism in pathogen infection of lake whitefish (*Coregonus clupeaformis*) species pairs as revealed by next-generation sequencing.** *Mol Ecol* 2013, **22**:3833–3849.
30. Radwan J, Zagalska-Neubauer M, Cichoń M, Sendecka J, Kulma K, Gustafsson L, Babik W: **MHC diversity, malaria and lifetime reproductive success in collared flycatchers.** *Mol Ecol* 2012, **21**:2469–2479.
31. Wilson AB, Martin-Smith KM: **Genetic monogamy despite social promiscuity in the pot-bellied seahorse (*Hippocampus abdominalis*).** *Mol Ecol* 2007, **16**:2345–2352.
32. Sommer S, Whittington CM, Wilson AB: **Standardised classification of pre-release development in male-brooding pipefish, seahorses, and seadragons (Family Syngnathidae).** *BMC Dev Biol* 2012, **12**:39.
33. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucl Acids Res* 1994, **22**:4673–4680.
34. Hall TA: **BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.** *Nucl Acids Symp* 1999, **41**:95–98.
35. Stephens M, Donnelly P: **A comparison of Bayesian methods for haplotype reconstruction from population genotype data.** *Am J Hum Genet* 2003, **73**:1162–1169.
36. Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S: **MEGA6: molecular evolutionary genetics analysis version 6.0.** *Mol Biol Evol* 2013, **30**:2725–2729.
37. Excoffier L, Lischer HEL: **Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows.** *Mol Ecol Resour* 2010, **10**:564–567.
38. Excoffier L, Laval G, Balding D: **Gametic phase estimation over large genomic regions using an adaptive window approach.** *Hum Genomics* 2003, **1**:7–19.
39. Yang Z: **PAML 4: phylogenetic analysis by maximum likelihood.** *Mol Biol Evol* 2007, **24**:1586–1591.
40. Yang Z, Wong WSW, Nielsen R: **Bayes empirical bayes inference of amino acid sites under positive selection.** *Mol Biol Evol* 2005, **22**:1107–1118.
41. Clement M, Posada D, Crandall KA: **TCS: a computer program to estimate gene genealogies.** *Mol Ecol* 2000, **9**:1657–1659.
42. yWorks: **yED Graph Editor.** Tuebingen, Germany: yWorks; 2009.
43. Maydt J, Lengauer T: **Recco: recombination analysis using cost optimization.** *Bioinformatics* 2006, **22**:1064–1071.
44. Kittichotirat W, Guerquin M, Bumgarner RE, Samudrala R: **ProInfo PPC: a web server for atomic level prediction of protein complexes.** *Nucl Acids Res* 2009, **37**(Web Server):W519–W525.
45. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE: **UCSF Chimera—a visualization system for exploratory research and analysis.** *J Comp Chem* 2004, **25**:1605–1612.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

