

# Ecology and Evolution

Open Access

## Combining molecular evolution and environmental genomics to unravel adaptive processes of MHC class IIB diversity in European minnows (*Phoxinus phoxinus*)

Helene Collin<sup>1,2</sup>, Reto Burri<sup>1,3</sup>, Fabien Comtesse<sup>1</sup> & Luca Fumagalli<sup>1</sup>

<sup>1</sup>Department of Ecology and Evolution, Laboratory for Conservation Biology, University of Lausanne, Biophore, 1015 Lausanne, Switzerland

<sup>2</sup>Institute of Integrative Biology, University of Liverpool, Biosciences Building, Crown St., Liverpool L69 7ZB, U.K.

<sup>3</sup>Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, 75236 Uppsala, Sweden

### Keywords

454 sequencing, AFLP, MHC class IIB, pathogens, *Phoxinus phoxinus*, selection.

### Correspondence

Hélène Collin, Institute of Integrative Biology, Department of Ecology, Evolution and Behaviour, Biosciences Building, University of Liverpool, Crown St., Liverpool L69 7ZB, U.K.  
Tel: 44 (0) 151 695 45 28;  
Fax: 44 (0) 151 795 44 10;  
E-mail: [hcollin25200@gmail.com](mailto:hcollin25200@gmail.com)

### Funding Information

The study was supported by a Swiss National Science Foundation grant (3100A0-109852) to L. F. and a "Société Académique Vaudoise" grant to H. C.

Received: 27 March 2013; Revised: 26 April 2013; Accepted: 29 April 2013

*Ecology and Evolution* 2013; 3(8): 2568–2585

doi: 10.1002/ece3.650

## Introduction

In natural environments, heterogeneity in selection due to habitat fragmentation, climate fluctuations, and pathogen regime variation play a major role in shaping populations' genetic diversity, and eventually can lead to local adaptation (Kawecki and Ebert 2004; Sommer 2005; Barrett et al. 2008; Gandon and Nuismer 2009; Spurgin and Richardson 2010). Pathogen communities can affect various fitness traits linked to immune resistance in hosts (Summers et al. 2003; Simkova et al. 2006; Evans and Neff 2009). In return, hosts develop a wide range of immune responses to overcome pathogen infections (Acevedo-Whitehouse and Cunningham 2006). Host–pathogen interactions are therefore ideal to study local

## Abstract

Host–pathogen interactions are a major evolutionary force promoting local adaptation. Genes of the major histocompatibility complex (MHC) represent unique candidates to investigate evolutionary processes driving local adaptation to parasite communities. The present study aimed at identifying the relative roles of neutral and adaptive processes driving the evolution of MHC class IIB (MHCIIB) genes in natural populations of European minnows (*Phoxinus phoxinus*). To this end, we isolated and genotyped exon 2 of two MHCIIB gene duplicates (DAB1 and DAB3) and 1'665 amplified fragment length polymorphism (AFLP) markers in nine populations, and characterized local bacterial communities by 16S rDNA barcoding using 454 amplicon sequencing. Both MHCIIB loci exhibited signs of historical balancing selection. Whereas genetic differentiation exceeded that of neutral markers at both loci, the populations' genetic diversities were positively correlated with local pathogen diversities only at DAB3. Overall, our results suggest pathogen-mediated local adaptation in European minnows at both MHCIIB loci. While at DAB1 selection appears to favor different alleles among populations, this is only partially the case in DAB3, which appears to be locally adapted to pathogen communities in terms of genetic diversity. These results provide new insights into the importance of host–pathogen interactions in driving local adaptation in the European minnow, and highlight that the importance of adaptive processes driving MHCIIB gene evolution may differ among duplicates within species, presumably as a consequence of alternative selective regimes or different genomic context.

adaptation, as both host and pathogen are subjected to varying selection in time and space, modifying their reciprocal fitness.

Genes of the major histocompatibility complex (MHC) are among the best candidates to study the genetics of the adaptive immune response (e.g., Klein 1987; Kalz and Shykoff 1998; Piertney and Oliver 2006). MHC genes are the most polymorphic genes in vertebrates, and participate in the activation of the adaptive immune response (Klein 1986). While MHC class I is involved in immune response against intracellular pathogens, class II triggers the immune response against extracellular pathogens. MHC polymorphism is believed to be mainly maintained by different mutually nonexclusive forms of pathogen-mediated balancing selection. First, heterozygote advantage

confers a greater resistance to MHC-heterozygote individuals, as they are able to recognize a greater array of pathogen-derived antigens than homozygotes (Doherty and Zinkernagel 1975; Hughes and Nei 1988; Bernatchez and Landry 2003; Garrigan and Hedrick 2003). Second, under rare-allele advantage, rare alleles are selected for as they provide better resistance than common genotypes to which pathogens are adapted (Takahata and Nei 1990; Bernatchez and Landry 2003; Garrigan and Hedrick 2003). Finally, pathogen communities may fluctuate over time and space, leading to temporally and spatially variable selection (Gandon and Nuismer 2009; Spurgin and Richardson 2010; Eizaguirre *et al.* 2012). As a consequence, various subsets of alleles are selected across pathogen fluctuation cycles and thereby high MHC polymorphism is promoted. Also, phases of positive selection may alternate with phases where neutral processes are more important in shaping MHC polymorphism (Klein 1987; Landry and Bernatchez 2001). The null hypothesis that neutral processes (mutation, genetic drift, gene flow) shape MHC diversity can be tested by contrasting patterns of genetic differentiation between MHC genes and neutral markers. Under local adaptation to spatially heterogeneous pathogen pressures, populations are expected to be more genetically differentiated at MHC genes than at neutrally evolving genetic markers. Even though pathogen-mediated selection seem to be a major force driving genetic diversity at MHC genes, neutral processes can decrease genetic differentiation at putatively selected loci (Bernatchez and Landry 2003; Van Oosterhout 2009; Miller *et al.* 2010). For instance, numerous cases reported bottleneck effects on MHC genes due to genetic drift in small and isolated populations (Miller *et al.* 2010; Bollmer *et al.* 2011; Spurgin *et al.* 2011; Strand *et al.* 2012). In these cases, genetic differentiation at MHC can be larger than at neutral loci as neutral genetic structure can be overestimated when population size is very low (Hedrick 1999; Bernatchez and Landry 2003; Van Oosterhout 2009). Also, differentiation at MHC genes can be lower compared to neutral loci in the face of balancing selection (including cases of overdominance and heterozygote advantage) (Schierup *et al.* 2000; Muirhead 2001; Bernatchez and Landry 2003; Blais *et al.* 2007; André *et al.* 2011; Bollmer *et al.* 2011).

MHC class I and IIB genes have been well described in fish, especially in commercially important Salmonids like the Atlantic salmon (*Salmo salar*). This species frequently suffers from infectious diseases, such as the salmon anemia virus, or furunculosis caused by the bacteria *Aeromonas salmonicida*. The polymorphism observed at MHC class I and IIB is maintained through pathogen-mediated selection (Dionne *et al.* 2007, 2009) with particular alleles conferring resistance to furunculosis (Langefors *et al.*

2001; Grimholt *et al.* 2003). Other studies also reported a joint role of selection and neutral processes on the evolution of local adaptation at immune relevant genes in European trout (*Salmo trutta*) or Atlantic salmon (*S. salar*) (Landry and Bernatchez 2001; Dionne *et al.* 2007; Keller *et al.* 2011). In comparison to *S. salar* and other Salmonids, cyprinid fish present a complex MHC architecture, with the presence of multiple duplicated MHC genes (Stet *et al.* 1998; Shum *et al.* 2001; Consuegra *et al.* 2005; Seifertova and Simkova 2011). More particularly, Simkova *et al.* (2006) found a positive relationship between pathogen species richness and MHCIIB diversity in various other European cyprinids.

The European minnow (*Phoxinus phoxinus*) is a cyprinid occurring in most European freshwater systems. Local adaptation in growth and survival of *P. phoxinus* to a specific pathogen, the trematode *Diplostomum phoxini*, has been highlighted in the early 1990's (Ballabeni and Ward 1993; Ballabeni 1994). Little is known about the influence of bacterial pathogens on this species' immunogenetics, especially on patterns of genetic structure and polymorphism at MHC loci. Recently, it has been shown that *P. phoxinus* displayed different body shape and adaptive genetic divergence to various abiotic factors such as lake-stream habitat, landscape topography, trophic dynamics, and geography, highlighting the capacity of this species to locally adapt to its environment (Collin and Fumagalli 2011). The present study aims at identifying whether genes of the immune system (MHCIIB genes) contribute to local adaptation to pathogens in natural populations of European minnows. More specifically, we intended to identify local adaptation by (1) comparing patterns of genetic differentiation at MHCIIB exon 2 and neutral markers, and (2) identifying pathogen-mediated selection by testing the correlation of diversity of potentially pathogenic bacteria with MHCIIB polymorphism. To this end we characterized diversity of MHCIIB exon 2 in European minnows and bacterial 16S rDNA in water samples using 454 amplicon sequencing. We identified two MHCIIB gene duplicates (DAB1 and DAB3) both presenting signs of historical balancing selection, but displaying contrasted patterns of adaptive evolution. Indeed, DAB3 duplicate exhibited evidence for pathogen-mediated selection contrarily to DAB1 duplicate, demonstrating that neutral and adaptive processes' roles differ among MHCIIB duplicates in European minnows.

## Material and Methods

### Sampling and DNA extraction

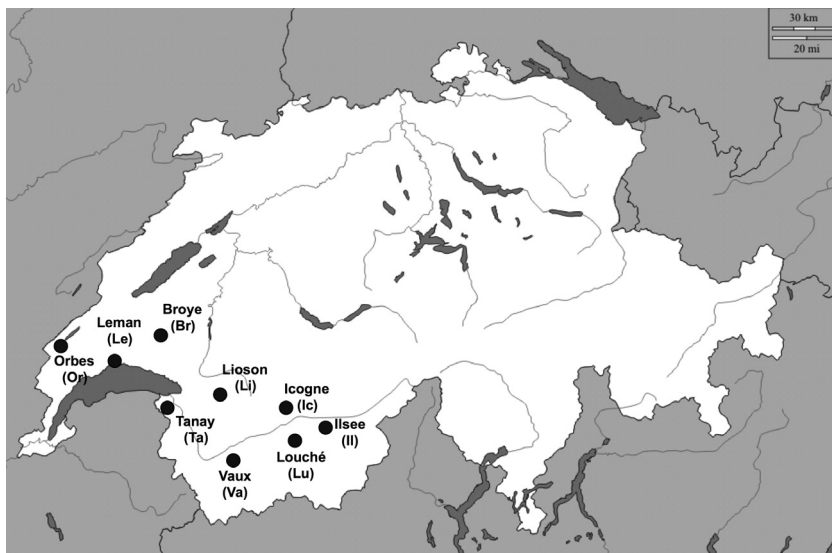
One hundred and seventy-six European minnows were collected in nine populations from the Western Swiss

Alps (Fig. 1) during summer 2008 and 2009, when bacteria communities are the most diverse. Adult fish were killed using AQUI-S® (AQUI-S New Zealand Ltd, Lower Hutt, NZ) and immediately stored in absolute ethanol for further genetic analyses. To assess bacterial diversity for each population, water samples were collected in 1 L sterile bottles and filtered using autoclaved membrane filters (0.2  $\mu\text{m}$ ). Filters were stored in 3 mL of lysis buffer (0.75 mol/L Sucrose, 400 mmol/L NaCl, 20 mmol/L EDTA (pH 8.0), 50 mmol/L Tris-HCl (pH 9.0) at  $-80^{\circ}\text{C}$ . Filters in lysis buffer containing bacterial DNA were digested overnight at  $37^{\circ}\text{C}$  with sodium dodecyl sulfate (1%), proteinase K (60  $\mu\text{g mL}^{-1}$ ), and  $\text{CaCl}_2$  (90 mmol/L). Bacterial DNA was extracted from 500  $\mu\text{L}$  of digested lysate using phenol chloroform (pH 7.5), sodium acetate (15  $\mu\text{mol/L}$ ), isopropanol, and ethanol precipitation procedure. DNA was resuspended in sterile water and concentration was adjusted to 10 ng  $\mu\text{L}^{-1}$ . DNA from fish was extracted using the Gentra Puregene Qiagen kit according to the manufacturer instructions. DNA concentration was adjusted to 50 ng  $\mu\text{L}^{-1}$ .

### MHCIIB isolation and characterization

In order to isolate and characterize MHCIIB exon 2 of *P. phoxinus*, a subset of four individuals from six populations (Ta, Ic, Li, Il, Va, Lu) were used. We first amplified the 3'-end of exon 2, intron 2 and the 5'-end of exon 3 using the primers FishC12S and Fish12R from Ottova *et al.* (2005) (for primer description and sequences, see Fig. S1 and Table S1). Additionally, intron 1 and the 5'-end of exon 2 were amplified using the reverse complement of FishC12S (FishC12S-R), and a newly designed primer in exon 1 (FishEx12F-bis). The latter primer was

designed in conserved regions of exon 1 from *Cyprinus carpio*, *Barbus intermedius*, and *Danio rerio* sequences (Van Erp *et al.* 1996; Kuroda *et al.* 2002; Kruiswijk *et al.* 2004). We then designed species-specific primers amplifying the entire exon 2, which was included in *P. phoxinus* the two duplicates DAB1 and DAB3 (inferred from phylogenetic analyses with other cyprinid MHCIIB exon 2 sequences, see below). PCR for the two duplicates of MHCIIB exon 2 characterization were performed in a 25  $\mu\text{L}$  reaction volume containing 1.5 mmol/L  $\text{MgCl}_2$ ,  $1\times$  PCR buffer (Qiagen), 0.4  $\mu\text{mol/L}$  of each primer, 100  $\mu\text{mol/L}$  of each dNTP, 1 unit of Taq Polymerase (Qiagen), and 50 ng of genomic DNA. Thermal cycling was carried out on a Biometra T Professional Standard Thermocycler (Biometra, Goettingen, Germany), with an initial denaturing step of  $95^{\circ}\text{C}$  for 3 min, followed by 39 cycles of 30 sec of denaturing at  $95^{\circ}\text{C}$ , 1 min annealing at  $52^{\circ}\text{C}$  and 1 min 30 sec of extension at  $72^{\circ}\text{C}$ , and a final extension of 5 min at  $72^{\circ}\text{C}$ . PCR products were cloned into the pGEM T-easy vector (Promega, Madison, WI). Between four to eight positive clones for each duplicate of MHCIIB exon 2 were purified with the Wizard SV PCR clean-up system (Promega) and sequenced using the Big Dye 3.1 Terminator cycle sequencing kit (Applied Biosystems, Carlsbad, CA) on a ABI 3100 PRISM genetic analyzer (Applied Biosystems) using SP6 and T7 standard primers and universal procedures. Sequences from clones were read using BioEdit 7.09 (Hall 1999). In order to ensure that the two duplicates corresponded to functional genes and not pseudogenes, we also amplified *P. phoxinus* cDNA following the protocol mentioned above (Text S1). Start and end of each exon 2 duplicates were deduced by comparison with known cDNA and gDNA sequences of *D. rerio* (Kuroda *et al.* 2002).



**Figure 1.** Map showing the locations of the nine populations sampled in the South Western Swiss Alps. Population abbreviations are indicated in brackets.

## 454 sequencing of MHCIIB exon 2 and 16S bacterial rDNA

In order to amplify the two duplicates of MHCIIB exon 2 and bacterial 16S rDNA in a partial 454 GS FLX Titanium run (Roche, Branford, CT), we pooled 18–20 individuals per population and randomly assigned each population five base pair tagged PCR primers (Table S1, Fig. S1). These primers were then used in locus-specific PCRs to amplify DAB1 and DAB3. The same tagging strategy was used to amplify bacterial 16S rDNA from the nine populations. PCRs for both MHCIIB exon 2 and 16S rDNA were conducted on Biometra T Professional Thermocyclers (Biometra) in a final volume of 25  $\mu\text{L}$ , using 1  $\times$  Q-solution (Qiagen), 1  $\times$  buffer (Qiagen), 1.5 mmol/L MgCl<sub>2</sub> (Qiagen), 0.2 mmol/L dNTPs (Peqlab, Erlanger, Germany), 0.2  $\mu\text{mol/L}$  of each primer, 2 units of Taq (Qiagen), and 20 ng of genomic DNA. Cycling conditions included an initial denaturation step at 95°C for 3 min followed by 35 cycles at 95°C for 30 sec, annealing at 52°C (DAB1 and DAB3 loci) or 58°C (bacterial 16S rDNA) for 1 min, 72°C for 1.5 min, and a final extension step at 72°C for 5 min. PCR product concentrations were estimated simultaneously on agarose gel and with a Nanodrop spectrometer (Nanodrop Technologies, Wilmington, DE). PCR products with similar concentrations were pooled and purified using the MinElute PCR Purification kit (Qiagen). Purified PCR products' concentrations were estimated a second time in order to produce a final pool with equimolar quantities of PCR products (Babik *et al.* 2009; Kloch *et al.* 2010). Final pool concentration was adjusted to 20 ng  $\mu\text{L}^{-1}$  in a final volume of 30  $\mu\text{L}$  and sequenced on a partial 454 GS FLX Titanium run (Microsynth AG, Switzerland).

## Sequence retrieval and filtering

### Bacterial 16S rDNA data

The 16S rDNA multifasta file resulting from the 454 run was sorted according to tag and primer sequences. In order to attribute each bacterial 16S rDNA sequence to a taxonomic unit (genus or species in our case), 16S rDNA sequences were submitted to the Bayesian Ribosomal Database Project (RDP) Classifier (<http://rdp.cme.msu.edu/classifier/classifier.jsp>; Wang *et al.* 2007; Cole *et al.* 2009). Sequences shorter than the expected amplicon size of 250 bp and with a bootstrap support inferior to 50% for taxon assignment were discarded to guarantee sufficient sequence length to reach a high taxonomic resolution and accuracy, respectively. A community data matrix containing the presence–absence of each species per population was built. We identified bacteria potentially pathogenic in fish using the classification established by Austin and Austin (1999).

### MHCIIB data

The primer used to sequence DAB1 and DAB3 using 454 technology amplify 276 bp approximately (Fig. S1). Altogether with primers and tag, read length exceeded the length that 454 run could sequence, resulting in reads whose length was shorter than expected. All sequences were thus truncated to 220 bp in order to maximize the number of sequences for further analyses (Spurgin *et al.* 2011). Sequences from the multifasta DAB1 and DAB3 files containing a complete tag and a match to the forward or reverse primer were retained for further analysis. From these files we identified authentic MHCIIB variants (true variants or TV) according to several criteria, partially following Galan *et al.* (2010). First, we counted for each population the number of unique variants (all sequences differing in at least one base pair or one indel). We then removed variants appearing only once, or in less than two populations in order to eliminate artifactual variants (AV) resulting from polymerase errors and sequencing errors. This step might remove some rare alleles, but avoids including AV in further analyses. As in the four to eight cloned sequences from four test populations we never observed frameshift mutations and because here we are exclusively interested in functional MHCIIB variation, the remaining sequences were aligned using ClustalW (Thompson *et al.* 1994) in Mega 5.0 (Tamura *et al.* 2011) and sequences with indels whose lengths differ from a multiple of three were excluded. Even though PCR products were included in equimolar quantities for sequencing, the number of variants represented only semiquantitative data, and the number of reads per TV might not be representative of the actual TV number in the data set. For further analyses, we thus decided to use a presence–absence matrix of TV instead of allele frequencies.

### AFLP data

In order to compare patterns of genetic differentiation at neutral markers with those observed at MHCIIB, we generated amplified fragment length polymorphism (AFLP) markers using the same protocol as described in Collin and Fumagalli (2011). AFLP were automatically genotyped using Genemapper version 4.0 software (Applied Biosystems) and genotypes were manually cross-validated twice. To avoid genotyping errors and a confounding factor from genotyping plates, 38 individuals from different populations were replicated independently thrice, and all individuals within and between populations were randomized across plates from the extraction step on. All loci that were not 100% reproducible across the 38 individuals were removed from the analysis. For data analysis, popu-

lations were assumed to be under Hardy–Weinberg equilibrium and nonpolymorphic (noninformative) loci were removed, resulting in 1'930 polymorphic loci. As we were interested in comparing neutral genetic divergence with MHCII B divergence between populations, we excluded loci potentially under selection. The latter were inferred using BAYESCAN 1.0 (Foll and Gaggiotti 2008) following a method previously developed specifically for European minnows (Collin and Fumagalli 2011). This software implements two models (including and not including selection, respectively) using reversible jump Markov Chain Monte Carlo sampling. It then estimates the posterior probability for a locus to be under selection, depending on a detection level that we set to “substantial” corresponding to a posterior probability superior to 0.76. The choice of this specific detection level is arbitrary, but has been previously demonstrated to not affect conclusions (Collin and Fumagalli 2011). Also, choosing a “substantial” detection level allows the detection of loci under very weak selection to generate a fully “neutral” data set with presence–absence of 1'665 neutral AFLP loci per population.

## Statistical analyses

### Bacteria diversity

In order to explore nonpathogenic and pathogenic bacterial composition between sampling locations, we conducted a permutational multivariate analysis of variance based on dissimilarities, that is, beta nonpathogenic and pathogenic bacteria diversity (ADONIS), using populations as a factor. Betadiversity was calculated using the nonpathogenic bacteria and pathogens' abundance matrices (excluding two populations, Br and Ta, as no pathogenic bacteria were recorded in these two locations). We finally conducted a canonical correspondence analysis (CCA) in order to analyze the correlation between pathogen composition and various environmental factors characterizing each sampling location and which have been shown to influence bacteria distribution (Yannarell *et al.* 2003; Lindström *et al.* 2005; Yannarell and Triplett 2005; Newton *et al.* 2007; Shade *et al.* 2008; Lindström and Langenheder 2012) (dissolved oxygen, altitude, landscape slope, soil pH, mean number of annual frost days, primary productivity; see Collin and Fumagalli 2011 for a full description of environmental variables). All analyses were conducted in R using the “vegan” package (Oksanen *et al.* 2008).

### MHCII B diversity

We were able to amplify 55 codons (165 bp) for DAB1 and 58 codons (174 bp) for DAB3, in which peptide-

binding region (PBR) represented the most polymorphic codons as they bind pathogenic antigens (Hughes and Yeager 1998). We inferred the locations of PBR and non-PBR regions from human MHCII B molecular structure (Brown *et al.* 1993). Two types of indices were estimated for MHCII B loci. First, amino acid diversity at every codon position was estimated separately for DAB1 and DAB3 using DIVAA (Rodi *et al.* 2004). Then DAB1 and DAB3 amino acid polymorphism was measured as the percentage of polymorphic codons for PBR, non-PBR and all codons, respectively. We tested for differences between DAB1 and DAB3 amino acid polymorphism at PBR, non-PBR, and all codons using a t-test. The relative rate of synonymous ( $d_S$ ) and nonsynonymous ( $d_N$ ) substitutions was calculated according to Nei and Gojobori (1986) using MEGA 5.0 (Tamura *et al.* 2011). MEGA 5.0 was also used to perform a Z-test of selection at all amino acid positions, PBR and non-PBR sites, under the null hypothesis that  $d_N = d_S$ . Additionally, we tested for positive selection using a maximum likelihood method implemented in CODEML (PAML v4.3 package; Yang 2007). Phylogenetic tree topologies were reconstructed using MrBayes 3.1 (Ronquist and Huelsenbeck 2003) for DAB1 and DAB3 duplicates separately using *Squalius cephalus* as outgroup (Genbank accession number HQ595117 for DAB1 and HQ595146 for DAB3). We then used CODEML to implement four site models: two neutral models (M1a and M7) and two models allowing for positive selection (M2a and M8) (Yang *et al.* 2000). The models were compared using likelihood ratio tests statistics (LRT statistics) calculated as two times the difference between the Log-likelihoods of the models allowing for selection and the neutral models. LRT statistics were compared to a chi-square distribution with 2 degrees of freedom. Positively selected codons were identified using the Bayes empirical Bayes (BEB) approach described in Yang *et al.* (2005).

### MHCII B and bacteria diversity

In order to provide evidence for pathogen-mediated selection acting on particular MHC sites, we tested the null hypothesis of no difference in the correlation between MHCII B polymorphism and bacterial pathogen diversity (Dionne *et al.* 2007). We calculated species richnesses of pathogenic bacteria, nonpathogenic bacteria, and overall bacterial species richness per population using the R library “vegan” (Oksanen *et al.* 2008). We then analyzed the relationship between the percentage of amino acid polymorphism at PBR, non-PBR codons, and bacterial species richness (pathogenic, nonpathogenic, overall) using linear models. Significance was tested using analysis of variance (ANOVA).

## Genetic differentiation at MHCIIB and neutral AFLPs

As our study includes population-based and not individual-based data, we used dissimilarity indices, that is, Euclidean distances between pairs of populations as an estimate for population differentiation. First, Euclidean distance matrices were estimated based on (1) TV presence–absence in each population for DAB1 and DAB3 and (2) AFLP band presence–absence in each population. DAB1, DAB3, and AFLP distance matrices were then standardized to a 0/1 scale using the following formula:  $(d_x - d_{\min}) / (d_{\max} - d_{\min})$ , where  $d_x$  represents the pairwise Euclidean distance between two populations,  $d_{\max}$  the maximum Euclidean distance between two populations, and  $d_{\min}$  the minimum Euclidean distance between two populations identified in the matrices. We first tested for isolation by geographical distance (IBD) with Mantel tests both for neutral AFLP and MHCIIB (separately for DAB1 and DAB3). In order to estimate if MHCIIB differentiation exceeds neutral differentiation along geographic clines, we tested whether isolation by distance (IBD) patterns in MHCIIB persisted when accounting for neutral population differentiation using a partial Mantel test. We tested for correlation between differentiation at AFLP and at MHCIIB loci (DAB1 and DAB3 separately) using a Mantel test in order to highlight a potential effect of divergent selection which would increase levels of genetic differentiation at immune loci compared to neutral markers. We also aimed at obtaining the full neutral distribution of IBD slopes, in order to directly compare this distribution to IBD slope values observed at the MHC, hence providing a cross-test of IBD not only using neutral population-based data but individual neutral data. We estimated individual-based AFLP  $F_{ST}$  matrices between pairs of populations using a subset of randomly chosen AFLP markers among the 1665 available. The regression slopes between  $F_{ST}$  matrices and geographical distance matrices were obtained from Mantel tests. We then compared the distribution of neutral regression slopes (minimum and maximum values of  $R^2$  obtained with AFLP markers) with values obtained from MHC loci. All Mantel and partial Mantel tests were implemented with 1000 permutations in each case using the “vegan” library in R (Oksanen et al. 2008).

## Results

### MHCIIB characterization and diversity

All MHCIIB sequences obtained by cloning four individuals of six *P. phoxinus* populations blasted against previously published MHCIIB sequences of other cyprinids

(including *Abramis brama*, *Barbus bocagei*, *B. intermedius*, *Brachydanio rerio*, *C. carpio*, *Ladigesocypris ghigii*, *Rhodeus ocellatus*, and *S. cephalus*) and salmonids (*S. salar*), with sequence similarities ranging from 79% to 92%. As reported also from other species (Ottova et al. 2005; Seifertova and Simkova 2011), we identified two functional duplicates, corresponding to DAB1 and DAB3. Functionality of the genes was supported by successful cDNA amplification (Text S1), and orthology with DAB1 and DAB3 of other cyprinids was confirmed by two well-separated paralog sequence clusters in a phylogenetic analysis (Appendix 1), with alleles from different species found in each duplicate cluster. We thus further studied two MHCIIB paralogs, with large differences in intron 1 and 2 lengths between the two groups of sequences strengthening the evidence for at least two gene copies (data not shown). A number of clues such as (1) successful amplification of cDNA (2) the absence of stop codon or frameshift mutations (3) the presence of many features characteristic of functional MHCIIB genes (N-linked glycosylation sites and the two conserved cysteine residues) are consistent with a functional antigen binding protein.

454 sequencing produced 7'591 DAB1 and 7'877 DAB3 reads (Text S2 and S3). Average, minimum and maximum read lengths and quality scores frequencies are reported in Appendix 2 and Fig. S2. The first step consisting in the identification of unique TV and AV resulted in 3'456 retained sequences for DAB1 and 3'811 for DAB3. After removing sequences appearing only once or less in the two populations, 162 variants for DAB1 and 177 for DAB3 were left. Sequences were aligned and partial intronic sequences were removed. After alignment and removal of sequences with frameshift indels, 19 TV were found for DAB1 (956 sequences in total, 13% of the initial reads number) and 30 TV for DAB3 (1'046 sequences in total, 13% of the initial reads number).

For both DAB1 and DAB3, amino acid diversity was highest at the 10 (DAB1) and 11 (DAB3) codons matching human PBR codons (Appendix 3). Mean DAB1 and DAB3 amino acid diversity at PBR codons were 0.117 and 0.139, respectively, and were significantly lower at non-PBR sites (0.009 and 0.059 for DAB1 and DAB3, respective  $t$ -tests:  $df = 8.832$ ,  $t = -3.685$ ,  $P = 0.005$ ;  $df = 10.438$ ,  $t = -4.219$ ,  $P = 0.002$ ). In concordance with these results, the  $Z$ -test of selection showed evidence for diversifying selection on PBR codons as the rate of non-synonymous substitution ( $d_S$ ) over synonymous substitution ( $d_N$ ) was significantly higher than expected under neutral expectations (Table 1). We tested for differences in amino acid polymorphism between DAB1 and DAB3 duplicates. The  $t$ -tests revealed that DAB3 presented greater amino acid polymorphism only at PBR codons ( $t = 2.402$ ,  $df = 15.887$ ,  $P = 0.014$ ). The maximum likelihood method

**Table 1.** Ratio of synonymous ( $d_s$ ) to nonsynonymous ( $d_n$ ) substitutions for the two duplicates of MHC class IIB (MHCIIB) exon 2 (DAB1 and DAB3).

Region	DAB1			DAB3		
	Codon	$d_n/d_s$	$P$	Codon	$d_n/d_s$	$P$
Non-PBR	45	0.664	0.508	47	0.548	0.585
PBR	<b>10</b>	<b>4.320</b>	<b>&lt;0.001</b>	<b>11</b>	<b>7.038</b>	<b>&lt;0.001</b>
Total	<b>55</b>	<b>2.131</b>	<b>0.035</b>	58	1.240	0.217

Bold font indicates significant departure from neutrality test at  $P < 0.05$ .

implemented in CODEML revealed that models allowing for positive selection (M2a and M8) better explained DAB1 and DAB3 sequences evolution than neutral models (M1a and M8) (Table 2). Models M2a and M8 identified 13 and 14 codons, respectively, as having evolved under positive selection in DAB1, with, respectively, seven and eight of these corresponding to human PBR sites. For DAB3, five codons were inferred having evolved under positive selection, all of them corresponding to PBR sites.

## Genetic differentiation and diversity at MHCIIB and neutral AFLPs

### Isolation by distance

Isolation by distance was observed at neutral loci and DAB1 (Fig. 2; Mantel test respectively,  $R = 0.330$ ,  $P = 0.037$  and  $R = 0.403$ ,  $P = 0.025$ ). DAB3 showed a tendency of IBD (Mantel test,  $R = 0.242$ ,  $P = 0.086$ ). The nonsignificance of this correlation was driven by three outlier population pairs (Ic-Or, Il-Or, Lu-Or) for which low levels of genetic differentiation was not explained by the geographic distance separating them. The three outlier

population pairs are highlighted by black circles in Fig. 2C. Removing them resulted in significant IBD (Mantel test,  $R = 0.515$ ,  $P = 0.004$ ). The pattern of IBD at DAB1 persisted when controlling for neutral genetic differentiation ( $R = 0.297$ ,  $P = 0.042$ ), and stayed nonsignificant for DAB3 ( $R = 0.188$ ,  $P = 0.140$ ). Again, when removing the three outlier pairs of populations and controlling for neutral differentiation, significant IBD was found also at DAB3 ( $R = 0.459$ ,  $P = 0.014$ ).

### Patterns of genetic differentiation at MHCIIB and neutral AFLP

Patterns of differentiation at neutral and MHCIIB duplicates were significantly correlated for DAB1, but not for DAB3 (Fig. 3; Mantel test; DAB1,  $R = 0.469$ ,  $P = 0.006$  and DAB3,  $R = 0.207$ ,  $P = 0.137$ ). Again, removing the three outlier pairs of populations resulted in significantly correlated genetic differentiation at AFLP markers and DAB3 (Mantel test;  $R = 0.313$ ,  $P = 0.062$ ). The neutral distribution of regression slopes ranged from  $-0.263$  to  $0.839$ , which includes values observed at MHC loci (0.445 and 0.242 for DAB1 and DAB3, respectively).

### Bacterial and MHCIIB diversity

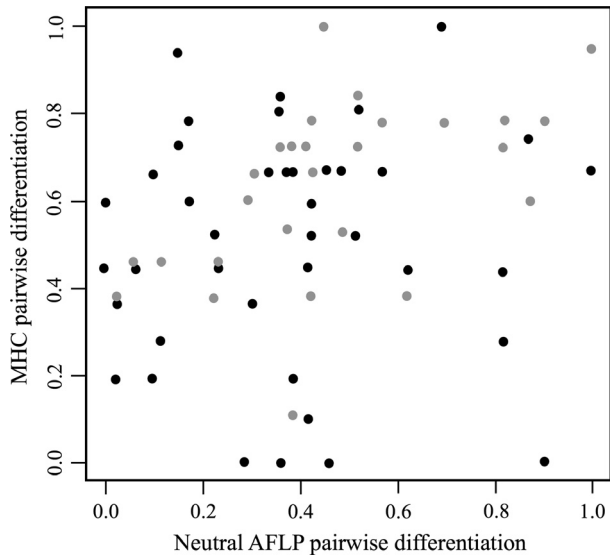
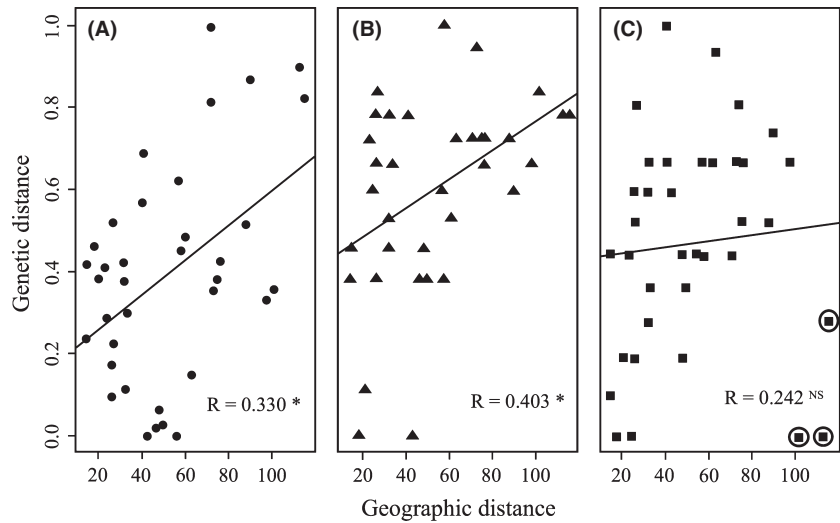
To investigate how pathogenic bacteria affect *P. phoxinus* populations' MHCIIB diversity at a local scale during summer, we first characterized pathogenic and nonpathogenic bacterial communities for each population using 454 amplicon sequencing of 16S rDNA. Classification in the Ribosomal Database Project retained sequences with a length  $>200$  bp resulting in a total of 374 sequences of 16S bacterial rDNA out of 548 reads (see Appendix 2 for

**Table 2.** Evidence of positive selection on MHC class IIB (MHCIIB) exon 2 duplicates.

Duplicate	Model	Ln L	Parameter estimates	Positively selected sites	LRT statistic
DAB1	M1a	-807.21	$\rho = 0.548$ ( $p_1 = 0.452$ ), $\omega = 0$ , $\omega_1 = 1$	Not allowed	
	M2a	-775.92	$\rho_0 = 0.501$ , $\rho_1 = 0.144$ ( $p_2 = 0.355$ ), $\omega = 0$ , $\omega_1 = 1$ , $\omega_2 = 9.121$	6, 8, 10, 14, 22, 24, 25, 35, 36, 44, 45, 51, 54	62.584 ( $P < 0.001$ )
	M7	-807.43	$\rho = 0.005$ , $q = 0.005$	Not allowed	
	M8	-775.93	$\rho_0 = 0.643$ ( $p_1 = 0.357$ ) $\rho = 0.005$ , $q = 0.020$ , $\omega = 8.965$	6, 8, 10, 14, 22, 24, 25, 28, 35, 36, 44, 45, 50, 51, 54	62.991 ( $P < 0.001$ )
DAB3	M1a	-1015.73	$\rho = 0.663$ ( $p_1 = 0.337$ ), $\omega = 0.081$ , $\omega_1 = 1$	Not allowed	
	M2a	-971.28	$\rho_0 = 0.454$ , $\rho_1 = 0.457$ ( $p_2 = 0.089$ ), $\omega = 0.065$ , $\omega_1 = 1$ , $\omega_2 = 9.402$	8, 10, 24, 26, 28	88.906 ( $P < 0.001$ )
	M7	-1019.09	$\rho = 0.159$ , $q = 0.221$	Not allowed	
	M8	-971.60	$\rho_0 = 0.911$ ( $p_1 = 0.089$ ) $\rho = 0.023$ , $q = 0.020$ , $\omega = 9.535$	8, 10, 24, 26, 28	94.963 ( $P < 0.001$ )

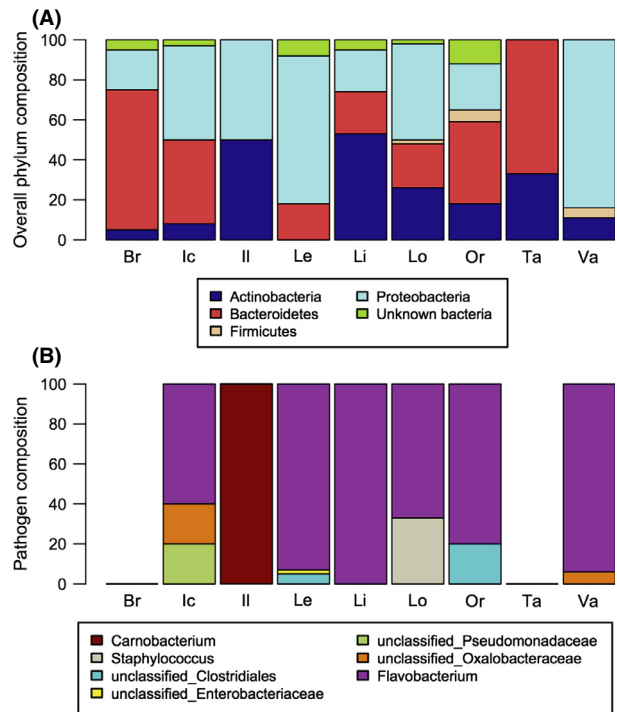
Sites inferred as being under positive selection with a posterior probability  $>0.99$  are also reported as numbers corresponding to their position in Fig. S4. LRT statistics were used to compare (1) M1a and M2a models and (2) M7 and M8 models to a chi-square distribution ( $df = 2$ ). Log-likelihood and parameter estimates calculated by CODEML are presented.

**Figure 2.** Isolation by distance patterns for genetic differentiation between populations at neutral AFLP (A), DAB1 (B), DAB3 (C). Lines represent the least square regression between genetic distance and geographical distance. R statistics from Mantel tests and their significance (\* significant at 95% level, <sup>NS</sup> non-significant at the 95% level) are indicated. The three outlier pairs of populations breaking down the isolation by distance (IBD) pattern at MHC class IIB (MHCIIB) DAB3 are circled in black (C).



**Figure 3.** Correlation between the extent of genetic differentiation at neutral AFLP loci and MHC class IIB (MHCIIB) loci (DAB1: gray closed dots, Mantel test:  $R = 0.469$ ,  $P = 0.006$ ; DAB3: black closed dots Mantel test:  $R = 0.207$ ,  $P = 0.137$ ).

read details). Based on Bergey’s nomenclature used by the classifier taxonomic assignment tool implemented in RDP, 41 different species were identified in the nine sampling sites analyzed, and 40 other species were not determined and thus treated as being different species in the data set. Seven potentially pathogenic species were identified following Austin and Austin (1999): *Lactobacillus* sp. (Carnobacterium), *Staphylococcus* sp., *Flavobacterium* sp., unclassified *Clostridiales*, unclassified *Enterobacteriaceae*, unclassified *Pseudomonadaceae*, and unclassified *Oxalobacteraceae*. We therefore partitioned the bacteria into

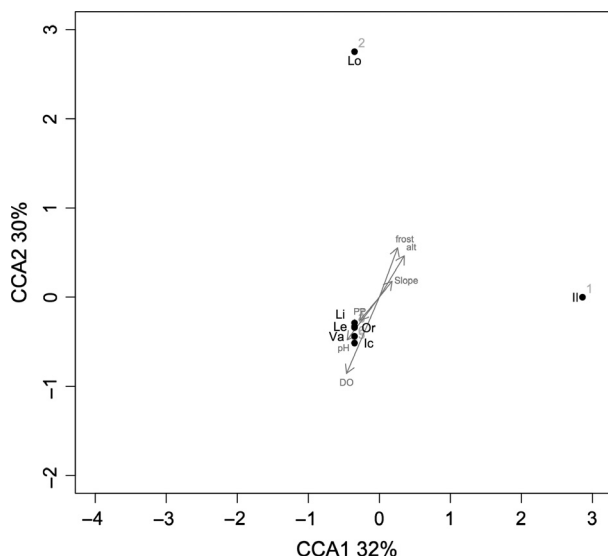


**Figure 4.** Barplot of 16S rDNA sequences relative abundance of phyla (A) and pathogen species (B) for sampling location (full population names and abbreviation are detailed in Figure 1).

three different data sets: “pathogenic bacteria,” “non-pathogenic bacteria”, and “all bacteria.” The species richness index calculated for these three data sets ranged from zero to three for pathogens, from two to 36 for nonpathogens, and from three to 39 for all bacteria. We examined the overall bacteria phylum composition for each sampling location (Fig. 4A). We found four main



phyla across the nine sampling locations screened: *Actinobacteria*, *Bacteroidetes*, *Firmicutes*, and *Proteobacteria*. A small proportion of unknown bacteria were recorded. Il and Va locations were characterized by the absence of the *Bacteroidetes* phylum and Ta by the absence of *Proteobacteria*. *Firmicutes* was the less abundant phylum in all sampling locations. No pathogenic bacteria were recorded at Br and Ta locations (Fig. 4B). Il location was characterized by *Carnobacterium* (*Firmicute*), while in other locations *Flavobacterium* pathogens (*Bacteroidetes*) predominated (Fig. 4B). ADONIS analysis revealed no significant difference in pathogenic and nonpathogenic bacteria composition between locations. CCA highlighted that 62% of the variance observed in the multivariate data set (pathogen diversity and environmental variables for each location) were explained by CCA1 and CCA2 axes (Fig. 5). As mentioned previously, Il and Lo locations were differentiated from the others as their pathogen distribution (1) was rather different, dominated by *Carnobacterium* and *Staphylococcus*, respectively, and (2) was positively associated to higher altitudinal sites and topography. Pathogen



**Figure 5.** Canonical correspondence analysis correlation biplot showing the relationship between pathogen abundance and composition per sampling location (for full names, see Fig. 1) and environmental variables (DO, dissolved oxygen; Slope, landscape slope; pH, soil pH; frost, mean number of annual days of frost; PP, primary productivity). Explanatory environmental variables are represented by gray arrows whose length indicated the strength of correlation between variables and pathogen's ordination score. Angles between variables reflect their degree of correlation. Gray numbers represent each pathogen species (1, *Carnobacterium*; 2, *Staphylococcus*; 3, unclassified *Clostridiales*; 4, unclassified *Enterobacteriaceae*; 5, unclassified *Pseudomonaceae*; 6, unclassified *Oxalabacteraceae*; 7, *Flavobacterium*).

compositions at other sampling locations were characteristic of highly productive (eutrophic) sites at lower elevation.

We used linear models and ANOVAs to explore the relationship between amino acid polymorphism at MHCIIB and species richness of pathogenic bacteria, non-pathogenic bacteria, and all bacteria. The percentage of polymorphic amino acid sites for the two MHCIIB duplicates ranged from 80% to 100% for DAB1, 82% to 100% for DAB3 at PBR codons, and at non-PBR codons from 20% to 27% for DAB1 and 19% to 36% for DAB3. Only the percentage of polymorphic amino acid sites within PBR codons at DAB3 showed significant positive association with pathogen species richness (ANOVA:  $df = 7$ ,  $F = 12.015$ ,  $P = 0.010$ ). ANOVA results of linear regression models are summarized in Table 3.

### Discussion

The present study aimed at examining whether the cyprinid fish *P. phoxinus* exhibited MHCIIB-related local adaptation to potentially pathogenic bacterial communities to which fish populations are exposed to. Our results suggest pathogen-mediated local adaptation in European minnows at two MHCIIB loci (DAB1 and DAB3), though they appeared to be driven by different aspects of pathogen-mediated selection. Patterns of genetic differentiation at DAB1 suggest that selection favored different alleles in each population, resulting presumably in better immune defense against local pathogens. This pattern was also found for DAB3, but only partially when removing the three pairwise comparisons with largest geographic distance. Contrary to its paralog, genetic diversity at DAB3 was correlated with local pathogen diversities. This suggests that although locally adapted alleles were favored by selection also at DAB3, selection for locally adapted genetic diversity overpowered the selection for locally adapted alleles. The interplay between these different selection processes may explain the breakdown of the overall pattern of IBD at DAB3.

**Table 3.** General linear model results regarding the effects of bacterial richness (nonpathogen and pathogen) on the percentage of polymorphic amino acid sites for the two MHCIIB duplicates (DAB1 and DAB3).

	PBR	Non-PBR	PBR	Non-PBR
All bacteria	$F = 0.536$	$F = 0.355$	$F = 0.440$	$F = 3.488$
( $df = 7$ )	$P = 0.488$	$P = 0.570$	$P = 0.528$	$P = 0.104$
Nonpathogen	$F = 0.740$	$F = 0.460$	$F = 0.244$	$F = 4.142$
( $df = 7$ )	$P = 0.418$	$P = 0.521$	$P = 0.637$	$P = 0.081$
Pathogen	$F = 0.532$	$F = 0.148$	<b><math>F = 12.015</math></b>	$F = 0.015$
( $df = 7$ )	$P = 0.489$	$P = 0.712$	<b><math>P = 0.010</math></b>	$P = 0.905$

Bold font indicates a significant effect at  $P < 0.05$ .

### Bacterial community composition

In the present study, pyrosequencing provided a fast way to characterize bacterial communities with a greater sequencing depth than any other method (Sogin et al. 2006). We found common phyla of freshwater lakes, such as *Bacteroidetes*, *Actinobacteria*, and *Proteobacteria* (Newton et al. 2011). We found seven potentially pathogenic bacteria species that were inferred from multiple previous studies as driving factors of fish diseases (Austin and Austin 1999; Isik et al. 1999; Clements et al. 2007, 2009; Moyer and Hunnicutt 2007; Embar-Gopinath et al. 2008; Loch et al. 2008; Sekar et al. 2008; Von Siebenthal et al. 2008; Prasad and Kumar 2010; Newton et al. 2011; Mouchet et al. 2012). Their composition didn't significantly differ between locations, but was influenced by various environmental factors such as altitude, landscape topography, and productivity. These factors have been previously reported to play a role in bacteria species distribution and abundance at local and regional scales (Yannarell et al. 2003; Lindström et al. 2005; Yannarell and Triplett 2005; Newton et al., 2007; Shade et al. 2008; Lindström and Langenheder 2012). No pathogenic bacteria were recorded at two locations (Br and Ta), suggesting that other types of pathogens may be present and play a role in shaping MHC genetic diversity in these populations. Even though ADONIS analysis showed little differences in pathogen composition, the CCA analysis revealed a high heterogeneity in pathogen distribution between sites, especially for Il and Lo. Il was dominated by *Carnobacterium*, which was reported as a major fish pathogen associated with warm water streptococcosis (Michel et al. 2003; Mata et al. 2004; Leisner et al. 2007). *Flavobacterium*, *Pseudomonas*, *Enterobacteriaceae*, or *Clostridium* are known to be pathogenic bacteria commonly colonizing digestive tract of fish (Sugita et al. 1985; Nayak 2010). *Enterobacteriaceae* recorded in one population, Le, can be a major disease agent in fish and is probably of anthropogenic origin (Newton et al. 2011) and Le is a human managed freshwater location. ADONIS and CCA presented contrasted results suggesting that pathogen richness and distribution differ but not their composition. Also, the environmental factors identified by the CCA to have potential influence on pathogen distribution were also found to play a role in morphological adaptation to various habitats in the European minnow (Collin and Fumagalli 2011). These concordant lines of evidence suggest that European minnows are locally adapted to both abiotic and biotic factors in their respective environment.

### Pathogen-mediated selection on MHCIIB loci

In the present study, we found that DAB1 amino acid polymorphism at PBR sites was not associated to pathogen

richness, but patterns of differentiation at neutral loci and DAB1 were significantly correlated. These results could suggest an effect of migration and drift on DAB1 diversity. However, contrary to expectations where neutral processes alone were explaining genetic structure at DAB1, the pattern of isolation by distance at DAB1 remained significant even when controlling for neutral genetic structure. These results suggest (1) a strong spatial structure between populations and (2) a role for diversifying selection, with selection on DAB1 favoring different, locally adapted alleles in every population. DAB1 genetic diversity seems thus to be the result of diversifying selection promoting different locally adapted alleles frequencies in different populations. Similar results were observed on MHCIIB in recent studies on great snipes (*Gallinago media*; Ekblom et al. 2007) and MHC class I in house sparrow (*Passer domesticus*; Loiseau et al. 2009).

Contrarily to DAB1, a significant positive relationship between diversity of potentially pathogenic bacteria and DAB3 amino acid polymorphism was found. Also, patterns of genetic differentiation at DAB3 and neutral markers were not correlated, implying that selection shaped DAB3 diversity. The latter results suggest a role for contemporary pathogen-mediated selection in promoting local adaptation in terms of genetic diversity of *P. phoxinus* at DAB3. Pathogen-mediated selection linked to habitat heterogeneity has been proposed as a major mechanism shaping MHC variation (Blais et al. 2007; Ekblom et al. 2007; Alcaide et al. 2010). In fish, there are numerous empirical examples suggesting that pathogen-mediated selection is acting to promote MHC diversity through three main mechanisms (reviewed in Wegner 2008 and Eizaguirre and Lenz 2010): (1) heterozygote advantage (Rakus et al. 2008; Evans and Neff 2009; Kekäläinen et al. 2009), (2) frequency-dependent selection (Langefors et al. 2001; Croisetière et al. 2008; Dionne et al. 2009; Eizaguirre et al. 2009) and/or (3) habitat heterogeneity (Dionne et al. 2007, 2009; Eizaguirre et al. 2011, 2012). MHC differentiation observed at DAB3 is likely to be a result of habitat heterogeneity and local adaptation to the local pathogen diversities. Several lines of evidence suggest that pathogen-mediated selection promotes and maintains genetic diversity at DAB3. First, IBD became only apparent for DAB3 when excluding three outlier pairs of populations and was absent otherwise. This highlights a limited role for distance-dependent gene flow in shaping the distribution of DAB3 alleles. Significant IBD after the exclusion of the three outlier pairs of populations suggests that selection (local adaptation) at DAB3 has driven allele frequency evolution in a number of populations, as for DAB1. As the outliers show low genetic differentiation at large distances, selection appears to have favored similar alleles in isolated populations. We

found a significant relationship between pathogen richness and DAB3 amino acid polymorphism at PBR sites. Overall, these results suggest that the genetic diversity observed at DAB3 may be adaptive at a local scale, and that bacterial pathogen diversity represents a main selective agent. Though, we cannot exclude that only balancing selection may act on the three outlier pairs of populations at DAB3 loci. Indeed, as mentioned earlier, balancing selection usually produces pattern low of genetic differentiation at MHC loci compared to neutral markers (Hedrick *et al.* 2001; Bernatchez and Landry 2003; Seddon and Ellegren 2004; Van Oosterhout 2009; Evans *et al.* 2010; Miller *et al.* 2010). For example, Van Oosterhout *et al.* (2006) found lower levels of genetic differentiation at MHC loci than at neutral markers in the *Poeciliid* fish *Poecilia reticulata*. They explain such a pattern by MHC loci being under balancing selection, increasing the effective migration rate for these genes compared to what is expected at neutral markers. Finally, the pattern of genetic differentiation observed between populations of European minnows and the presence of outlier pairs of populations could suggest an effect of fish stocking. We can rule out this hypothesis with high confidence, as the levels of genetic differentiation for outlier pairs of population are higher at AFLP loci and DAB1 than they are at DAB3 (gene flow from stocking would lead to genome-wide low differentiation).

The slight contrasting pattern of differentiation between DAB1 and DAB3 suggests that the evolution of these two duplicates is driven by different selective agents (Landry and Bernatchez 2001) and/or different pathogen-mediated selection intensity (Edwards and Hedrick 1998; Martinsohn *et al.* 1999; Penn and Potts 1999; Van Oosterhout 2009). Contrasting patterns of evolution of the MHCIIB duplicates similar to the ones identified here in *P. phoxinus* have already been demonstrated in other cyprinid species (Seifertova and Simkova 2011). For example, chubs (*Squalius cephalus*) display a similar MHCIIB architecture with each two of the duplicates present in European minnows highlighted in the present study. In *S. cephalus* DAB3 also presents a greater diversity than DAB1. Although these data support the results found in the present work, several studies on other cyprinid fish reported the inverse pattern (Ottova *et al.* 2005, 2007).

### Combining molecular evolution and environmental genomics

Our study used a combination of molecular evolution and environmental genomics to understand how pathogenic bacteria communities and their environmental characteristics influence MHCIIB genetic diversity in nine populations of European minnow. We also used genome-wide neutral

AFLP markers to compare levels of genetic differentiation at both neutral and selected markers. This study attempted to (1) characterize MHCIIB in *P. phoxinus* and (2) understand how pathogen-mediated selection shapes diversity at two MHCIIB copies. Despite providing important insights into these questions, our study also shows the technical limitations of this approach.

First, the bacteria diversity survey was based on water samples collected from a single water sampling event. Even though this may represent some limitation, this approach gave some important information about pathogen distribution and abundance that can be a starting point for further studies on European minnows or other species. Indeed, most studies aiming at detecting pathogen-mediated selection, often do not report diversity, spatial distribution, or abundance of pathogens (Wegner 2008). Even though in the present study potentially pathogenic bacteria were recovered from water samples and not from fish, numerous studies on bacteria associated with fish demonstrated that gut and gill microflora (including potential pathogens) is colonized by free living aquatic bacteria and/or ones associated to food (Nayak 2010; Mouchet *et al.* 2012). Further work on the topic should also consider temporal distribution of pathogens, as time scale can be of particular importance in shaping MHC diversity.

We demonstrated that both copies presented historical evidence for positive selection, but only one copy (DAB3) presented evidence for contemporary pathogen-mediated selection. DAB1 may also be subjected to this type of selection, but as fish and water samples have been sampled during the summer, we can hypothesize that this copy may be involved in the recognition of pathogens that were not sampled at that time of the year, or other pathogen taxa.

Our next-generation sequencing of MHCIIB copies presents several limits as compared to more classical approaches. The major limitation of our approach was the population-level tagging rather than individual-based sequencing. Even though it provided a rapid and cheap way to sequence a large number of individuals, the frequency of the identified variants could not be reliably inferred. Several hypotheses related to pathogen-mediated selection could not be tested such as a potential heterozygote advantage or frequency dependent selection. Nevertheless, several recent publications successfully used a population-based 454 sequencing approach similar to the one presented in our study to assess population-level MHC variation (Spurgin *et al.* 2011; Radwan *et al.* 2012; Sepil *et al.* 2012). In addition, we think that the technological progress in tagging possibilities and the longer read length now available with next-generation sequencing will soon unravel the limitations mentioned above.

## Conclusion

Duplication events and the fate of duplicates in the face of both neutral processes and selection are still sparsely understood. In this context, the present study demonstrates how host–pathogen interactions can be the driver of genetic diversity in immune-related genes, and how alternative selective regimes may influence the processes driving the evolution of gene duplicates within the same species. Further studies with higher technical resolution examining the duplication history of MHC genes in cyprinid fish in Europe and the origin of duplicates' diversity should provide new insights into the relative role of historical and contemporary selection, and neutral processes on host–pathogen related local adaptation.

## Acknowledgments

We thank Lee Marowski for providing DIVAA, T. Parchman and P. Taberlet for help with analyses, S. Antoniazza and S. A. Pavey for comments on the manuscript, P. Nosil for discussion. We also thank the “Services de la Faune” from the cantons of Valais and Vaud, L. Anex, Y. Crettenand, and J. M. Trolliet for help with sampling. This work was conducted in compliance with the Swiss laws on animal experimentation under authorization number 2049 delivered by the Swiss Veterinary Services. The study was supported by a Swiss National Science Foundation grant (3100A0-109852) to L. F. and a “Société Académique Vaudoise” grant to H. C.

## Conflict of Interest

None declared.

## References

- Acevedo-Whitehouse, K., and A. A. Cunningham. 2006. Is MHC enough for understanding wildlife immunogenetics?. *Trends Ecol. Evol.* 21:433–438.
- Alcaide, M., J. A. Lemus, G. Blanco, J. L. Tella, D. Serrano, and J. J. Negro. 2010. MHC diversity and differential exposure to pathogens in kestrels (Aves: *Falconidae*). *Mol. Ecol.* 19:691–705.
- André, C., L. C. Larsson, L. Laikre, D. Bekkevold, J. Brigham, G. R. Carvalho, et al. 2011. Detecting population structure in a high gene-flow species, Atlantic herring (*Clupea harengus*): direct, simultaneous evaluation of neutral vs putatively selected loci. *Heredity* 106:270–280.
- Austin, B., and D. A. Austin. 1999. Bacterial fish pathogens: disease in farmed and wild fish. Springer, Chichester, U.K.
- Babik, W., P. Taberlet, M. J. Ejsmond, and J. Radwan. 2009. New generation sequencers as a tool for genotyping highly polymorphic multilocus MHC system. *Mol. Ecol. Resour.* 9:713–719.
- Ballabeni, P. 1994. Experimental differences in mortality patterns between European minnows, *Phoxinus phoxinus*, infected with sympatric or allopatric trematodes *Diplostomum phoxini*. *J. Fish Biol.* 45:257–267.
- Ballabeni, P., and P. I. Ward. 1993. Local adaptation of the trematode *Diplostomum phoxini* to the European minnow *Phoxinus phoxinus*, its second intermediate host. *Funct. Ecol.* 7:84–90.
- Barrett, L. G., P. H. Thrall, J. J. Burdon, and C. C. Linde. 2008. Life history determines genetic structure and evolutionary potential of host–parasite interactions. *Trends Ecol. Evol.* 23:678–685.
- Bernatchez, L., and C. Landry. 2003. MHC studies in non model vertebrates: what have we learned about natural selection in 15 years? *J. Evol. Biol.* 16:363–377.
- Blais, J., C. Rico, C. van Oosterhout, J. Cable, G. F. Turner, and L. Bernatchez. 2007. MHC adaptive divergence between closely related and sympatric african cichlids. *PLoS ONE* 2: e734.
- Bollmer, J. L., J. M. Hull, H. B. Ernest, J. H. Sarasola, and P. G. Parker. 2011. Reduced MHC and neutral variation in the Galápagos hawk, an island endemic. *BMC Evol. Biol.* 11:143.
- Brown, J. H., T. S. Jadetzky, J. C. Gorga, L. J. Stern, R. G. Urban, J. L. Strominger, et al. 1993. Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* 364:33–39.
- Clements, K. D., I. B. Y. Pasch, D. Moran, and S. J. Turner. 2007. Clostridia dominate 16S rRNA gene libraries prepared from the hindgut of temperate marine herbivorous fishes. *Mar. Biol.* 150:1431–1440.
- Clements, K. D., D. Raubenheimer, and J. H. Choat. 2009. Nutritional ecology of marine herbivorous fishes: ten years on. *Funct. Ecol.* 23:79–92.
- Cole, J. R., Q. Wang, E. Cardenas, J. Fish, B. Chai, R. J. Farris, et al. 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* 37:141–145.
- Collin, H., and L. Fumagalli. 2011. Evidence for morphological and adaptive genetic divergence between lake and stream habitats in European minnows (*Phoxinus phoxinus*, Cyprinidae). *Mol. Ecol.* 20:4490–4502.
- Consuegra, S., H. J. Megens, K. Leon, R. J. M. Stet, and W. C. Jordan. 2005. Patterns of variability at the major histocompatibility class II alpha locus in Atlantic salmon contrast with those at the class I locus. *Immunogenetics* 57:16–24.
- Croisatière, S., P. D. Tarte, L. Bernatchez, and P. Belhumeur. 2008. Identification of MHC class II beta resistance/susceptibility alleles to *Aeromonas salmonicida* in brook charr (*Salvelinus fontinalis*). *Mol. Immunol.* 45: 3107–3116.

- Dionne, M., K. M. Miller, J. J. Dodson, F. Caron, and L. Bernatchez. 2007. Clinal variation in MHC diversity with temperature: evidence for the role of host-pathogen interaction on local adaptation in Atlantic salmon. *Evolution* 61:2154–2164.
- Dionne, M., K. M. Miller, J. J. Dodson, and L. Bernatchez. 2009. MHC standing genetic variation and pathogen resistance in wild Atlantic salmon. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364:1555–1565.
- Doherty, P. C., and R. M. Zinkernagel. 1975. Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature* 256:50–52.
- Edwards, S. V., and P. W. Hedrick. 1998. Evolution and ecology of MHC molecules: from genomics to sexual selection. *Trends Ecol. Evol.* 13:305–311.
- Eizaguirre, C., and T. L. Lenz. 2010. Major histocompatibility complex polymorphism: dynamics and consequences of parasite-mediated local adaptation in fishes. *J. Fish Biol.* 77:2023–2047.
- Eizaguirre, C., S. E. Yeates, T. L. Lenz, M. Kalbe, and M. Milinski. 2009. MHC-based mate choice combines good genes and maintenance of MHC polymorphism. *Mol. Ecol.* 18:3316–3329.
- Eizaguirre, C., T. L. Lenz, R. D. Sommerfeld, C. Harrod, C. Kalbe, and M. Milinski. 2011. Parasite diversity, patterns of MHC II variation and olfactory based mate choice in diverging three-spined stickleback ecotypes. *Evol. Ecol.* 25:605–622.
- Eizaguirre, C., T. L. Lenz, M. Kalbe, and M. Millinski. 2012. Divergent selection on locally adapted major histocompatibility complex immune genes experimentally proven in the field. *Ecol. Lett.* 15:723–731.
- Eklblom, R., S. A. Sæther, P. Jacobsson, P. Fiske, T. Sahlman, M. Grahn, et al. 2007. Spatial pattern of MHC class II variation in the great snipe (*Gallinago media*). *Mol. Ecol.* 16:1439–1451.
- Embar-Gopinath, S., J. P. Bowman, J. Carson, P. B. B. Crosbie, and B. F. Nowak. 2008. A culture-dependent 16S rRNA gene-based approach to identify gill bacteria associated with amoebic gill disease in Atlantic salmon. *Bull. Eur. Assn. Fish P.* 28:27–34.
- Evans, M. L., and B. D. Neff. 2009. Major histocompatibility complex heterozygote advantage and widespread bacterial infections in populations of Chinook salmon (*Oncorhynchus tshawytscha*). *Mol. Ecol.* 18:4716–4729.
- Evans, M. L., B. D. Neff, and D. D. Heath. 2010. MHC genetic structure and divergence across populations of Chinook salmon (*Oncorhynchus tshawytscha*). *Heredity* 104:449–459.
- Foll, M., and O. Gagiotti. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* 180:977–993.
- Galan, M., E. Guivier, G. Caraux, N. Charbonnel, and J.-F. Cosson. 2010. A 454 multiplex sequencing method for rapid and reliable genotyping of highly polymorphic genes in large-scale studies. *BMC Genomics* 11:296.
- Gandon, S., and S. L. Nuismer. 2009. Interactions between genetic drift, gene flow, and selection mosaics drive parasite local adaptation. *Am. Nat.* 173:212–224.
- Garrigan, D., and P. W. Hedrick. 2003. Detecting adaptive molecular polymorphism: lessons from the MHC. *Evolution* 57:1707–1722.
- Grimholt, U., S. Larsen, R. Nordmo, P. Midtlyng, S. Kjoeglum, A. Storset, et al. 2003. MHC polymorphism and disease resistance in Atlantic salmon (*Salmo salar*): facing pathogens with single expressed major histocompatibility class I and class II loci. *Immunogenetics* 55:210–219.
- Hall, T. A. 1999. BioEdit: a user-friendly biological alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41:95–98.
- Hedrick, P. W. 1999. Balancing selection and MHC. *Genetica* 104:207–214.
- Hedrick, P. W., G. A. Gutierrez-Espeleta, and R. N. Lee. 2001. Founder effect in an island population of bighorn sheep. *Mol. Ecol.* 10:851–857.
- Hughes, A. L., and M. Nei. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 335:167–170.
- Hughes, A. L., and M. Yeager. 1998. Natural selection at the major histocompatibility complex loci of vertebrates. *Annu. Rev. Genet.* 32:415–435.
- Isik, K., J. Chun, Y. C. Hah, and M. Goodfellow. 1999. *Nocardia salmonicida* nom. Rev., a fish pathogen. *Int. J. Syst. Bacteriol.* 49:833–837.
- Kalz, O., and J. A. Shykoff. 1998. Local adaptation in host-parasite systems. *Heredity* 81:361–370.
- Kawecki, T. J., and D. Ebert. 2004. Conceptual issues in local adaptation. *Ecol. Lett.* 7:1225–1241.
- Kekäläinen, J., J. Albert Vallunen, C. R. Primmer, J. Räättä, and J. Taskinen. 2009. Signals of major histocompatibility complex overdominance in a wild salmonid population. *Proc. Biol. Sci.* 276:3133–3140.
- Keller, I., A. Taverna, and O. Seehausen. 2011. Evidence of neutral and adaptive genetic divergence between European trout populations sampled along altitudinal gradients. *Mol. Ecol.* 20:1888–1904.
- Klein, J. 1986. Natural history of the major histocompatibility complex. John Wiley, New York, NY.
- Klein, J. 1987. Origin of major histocompatibility complex polymorphism: the trans-species hypothesis. *Hum. Immunol.* 19:155–162.
- Kloch, A., W. Babik, A. Bajer, E. Sinski, and J. Radwan. 2010. Effects of an MHC-DRB genotype and allele number on the load of gut parasites in the bank vole *Myodes glareolus*. *Mol. Ecol.* 19:255–265.
- Kruiswijk, C. P., T. Hermsen, K. Fujiki, B. Dixon, H. F. J. Savelkoul, and R. J. M. Stet. 2004. Analysis of genomic and

- expressed major histocompatibility class Ia and class II genes in a hexaploid Lake Tana African “large” barb individual (*Barbus intermedius*). *Immunogenetics* 55:770–781.
- Kuroda, N., F. Figueroa, C. O’huigin, and J. Klein. 2002. Evidence that the separation of Mhc class II from class I loci in the zebrafish, *Danio rerio*, occurred by translocation. *Immunogenetics* 54:418–430.
- Landry, C., and L. Bernatchez. 2001. Comparative analysis of population structure across environments and geographical scales at major histocompatibility complex and microsatellite loci in Atlantic salmon (*Salmo salar*). *Mol. Ecol.* 10:2525–2539.
- Lanfords, A. H., J. Lohm, M. Grahn, O. Andersen, and T. Schantz. 2001. Association between major histocompatibility complex class IIB alleles and resistance to *Aeromonas salmonicida* in Atlantic salmon. *Proc. Biol. Sci.* 268:479–485.
- Leisner, J. J., B. G. Laursen, H. Prévost, D. Drider, and P. Dalgaard. 2007. *Carnobacterium*: positive and negative effects in the environment and in foods. *FEMS Microbiol. Rev.* 31:592–613.
- Lindström, E. S., and S. Langenheder. 2012. Local and regional factors influencing bacterial community assembly. *Environ. Microbiol. Rep.* 4:1–9.
- Lindström, E. S., M. P. Kamst-Van Agterveld, and G. Zwart. 2005. Distribution of typical freshwater bacterial groups is associated with pH, temperature and lake water retention time. *Appl. Environ. Microbiol.* 71:8201–8206.
- Loch, T. P., W. Xu, S. M. Fitzgerald, and M. Faisal. 2008. Isolation of a *Carnobacterium* maltaromaticum-like bacterium from systemically infected lake whitefish (*Coregonus clupeaformis*). *FEMS Microbiol. Lett.* 288: 76–84.
- Loiseau, C., M. Richard, S. Garnier, O. Chastel, R. Julliard, R. Zoorob, et al. 2009. Diversifying selection on MHC class I in the house sparrow (*Passer domesticus*). *Mol. Ecol.* 18:1331–1340.
- Martinsohn, J. T., A. B. Sousa, L. A. Guethlein, and J. C. Howard. 1999. The gene conversion hypothesis of MHC evolution: a review. *Immunogenetics* 50:168–200.
- Mata, A. I., A. Gibello, A. Casamyor, M. M. Blanco, L. Domingez, and J. F. Fernandez-Garayzabal. 2004. Multiplex PCR assay for detection of bacterial pathogens associated with warm-water streptococcosis in fish. *Appl. Environ. Microbiol.* 70:3183–3187.
- Michel, C., B. Kerouault, and C. Martin. 2003. Chloramphenicol and florfenicol susceptibility of fish-pathogenic bacteria isolated in France: comparison of minimum inhibitory concentration, using recommended provisory standards for fish bacteria. *J. Appl. Microbiol.* 95:1008–1015.
- Miller, H. C., F. Allendorf, and C. H. Daugherty. 2010. Genetic diversity and differentiation at MHC genes in island populations of tuatara (*Sphneodon spp.*). *Mol. Ecol.* 19:3894–3908.
- Mouchet, M. A., C. Bouvier, T. Bouvier, M. Troussellier, A. Escalas, and D. Mouillot. 2012. Genetic difference but functional similarity among fish gut bacterial communities through molecular and biochemical fingerprints. *FEMS Microbiol. Ecol.* 79:568–580.
- Moyer, T. R., and D. W. Hunnicutt. 2007. Susceptibility of zebra fish *Danio rerio* to infection by *Flavobacterium columnare* and *F. johnsoniae*. *Dis. Aquat. Organ.* 76: 39–44.
- Muirhead, C. A. 2001. Consequences of population structure on genes under balancing selection. *Evolution* 55:1532–1541.
- Nayak, S. K. 2010. Role of gastrointestinal microbiota in fish. *Aquacult. Res.* 41:1553–1573.
- Nei, M., and T. Gojobori. 1986. Simple method for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* 3:418–426.
- Newton, R. J., E. S. Jones, M. R. Helmus, and K. D. McMahon. 2007. Phylogenetic ecology of freshwater Actinobacteria acl lineage. *App. Env. Microbiol.* 73: 7169–7176.
- Newton, R. J., S. E. Jones, A. Eiler, K. D. McMahon, and S. Bertilsson. 2011. A guide to the natural history of freshwater lake bacteria. *Microbiol. Mol. Biol. Rev.* 75: 14–49.
- Oksanen, J., R. Kindt, P. Legendre, B. O’Hara, G. L. Simpson, P. Solymos, et al. 2008. Vegan: community ecology package. R package version 1.15-0. Available at <http://vegan.r-forge.r-project.org> (accessed January 14, 2012).
- Ottova, E., S. Andrea, J.-F. Martin, J. G. De Belloq, M. Geinar, J.-F. Allienne, et al. 2005. Evolution and trans-species polymorphism of MHC class IIB genes in cyprinid fish. *Fish Shellfish Immun.* 18:199–222.
- Ottova, E., A. Simkova, and S. Morand. 2007. The role of major histocompatibility complex diversity in vigour of fish males (*Abramis brama* L.) and parasite selection. *Biol. J. Linn. Soc.* 90:525–538.
- Penn, D., and W. K. Potts. 1999. The evolution of MHC-disassortative mating preferences. *Am. Nat.* 153: 145–164.
- Piertney, S. B., and M. K. Oliver. 2006. The evolutionary ecology of the major histocompatibility complex. *Heredity* 96:7–21.
- Prasad, Y., and A. D. Kumar. 2010. Isolation and efficacy evaluation of virulent bacteriophages specific to fish pathogenic bacterium, *Flavobacterium columnare*. *J. Appl. Anim. Res.* 38: 169–172.
- Radwan, J., M. Zagalska-Neubauer, M. Cichon, J. Sendek, K. Kulma, L. Gustafsson, et al. 2012. MHC diversity, malaria and lifetime reproductive success in collared flycatchers. *Mol. Ecol.* 21:2469–2479.
- Rakus, K. L., G. F. Wiegertjes, P. Jurecka, P. D. Walker, A. Pilarczyk, and I. Irnazarow. 2008. Major histocompatibility (MH) class II B gene polymorphism influences disease

- resistance of common carp (*Cyprinus carpio* L.). *Aquaculture* 288:44–50.
- Rodi, D. J., S. Mandava, and L. Makowski. 2004. DIVAA: analysis of amino acid diversity in multiple aligned protein sequences. *Bioinformatics* 20:3481–3489.
- Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.
- Schierup, M. H., X. Vekemans, and D. Charlesworth. 2000. The effect of subdivision on variation at multi-allelic loci under balancing selection. *Genet. Res.* 76:51–62.
- Seddon, J. M., and H. Ellegren. 2004. A temporal analysis shows MHC loci in the Scandinavian wolf population are consistent with neutral evolution. *Proc. Biol. Sci.* 271:2283–2291.
- Seifertova, M., and A. Simkova. 2011. Structure, diversity and evolutionary patterns of expressed MHC class IIB genes in chub (*Squalius cephalus*), a cyprinid fish species from Europe. *Immunogenetics* 63:167–181.
- Sekar, V. T., T. C. Santiago, K. K. Vijayan, S. V. Alavandi, V. S. Raj, J. J. Rajan, et al. 2008. Involvement of *Enterobacter cloacae* in the mortality of the fish, *Mugil cephalus*. *Lett. Appl. Microbiol.* 46:667–672.
- Sepil, I., H. K. Moghadam, E. Huchard, and B. Sheldon. 2012. Characterization and 454 pyrosequencing of Major Histocompatibility Complex class I genes in the great tit reveal complexity in a passerine system. *BMC Evol. Biol.* 12:68.
- Shade, A., S. E. Jones, and K. D. McMahon. 2008. The influence of habitat heterogeneity on freshwater bacterial community composition and dynamics. *Environ. Microbiol.* 10:1057–1067.
- Shum, B. P., L. Guethlein, L. R. Flodin, M. D. Adkison, R. P. Hedrick, R. B. Nehring, et al. 2001. Modes of Salmonid MHC class I and II evolution differ from the primate paradigm. *J. Immunol.* 166:3297–3308.
- Simkova, A., E. Ottova, and S. Morand. 2006. MHC variability, life-traits and parasite diversity of European cyprinid fish. *Evol. Ecol.* 20:465–477.
- Sogin, M. L., H. G. Morrison, J. A. Huber, D. M. Welch, S. M. Huse, and P. R. Real. 2006. Microbial diversity in the deep sea and the underexplored “rare biosphere”. *Proc. Natl. Acad. Sci. USA* 103:12115–12120.
- Sommer, S. 2005. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Front. Zool.* 2:1–18.
- Spurgin, L. G., and D. S. Richardson. 2010. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proc. Biol. Sci.* 277:979–988.
- Spurgin, L. G., C. Van Oosterhout, J. C. Illera, S. Bridgett, K. Gharbi, B. C. Emerson, et al. 2011. Gene conversion rapidly generates major histocompatibility complex in recently founded bird populations. *Mol. Ecol.* 20:5213–5225.
- Stet, R. J. M., C. P. Kruiswijk, J. P. J. Saeij, and G. F. Wiegertjes. 1998. Major histocompatibility genes in cyprinid fishes: theory and practice. *Immunol. Rev.* 166:301–316.
- Strand, T. M., G. Segelbacher, M. Quintela, L. Xiao, T. Axelsson, and J. Höglund. 2012. Can balancing selection on MHC counteract genetic drift in small fragmented populations of black grouse. *Ecol. Evol.* 2:341–353.
- Sugita, H., K. Tokuyama, and Y. Deguchi. 1985. The intestinal microflora of carp *Cyprinus carpio*, grass carp *Ctenopharyngodon idella* and tilapia *Sarotherodon niloticus*. *Bull. Jap. Soc. Sci. Fish.* 50:1325–1329.
- Summers, K., S. McKeon, J. Sellars, M. Keusenkothen, J. Morris, D. Gloeckner, et al. 2003. Parasitic exploitation as an engine of diversity. *Biol. Rev.* 78:639–675.
- Takahata, N., and M. Nei. 1990. Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics* 124:967–978.
- Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar. 2011. MEGA5: molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol. Biol. Evol.* 28:2731–2739.
- Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:4673–4680.
- Van Erp, S. H., E. Egberts, and R. J. M. Stet. 1996. Characterization of major histocompatibility complex class II A and B genes in gynogenetic carp clone. *Immunogenetics* 44:192–202.
- Van Oosterhout, C. 2009. A new theory of MHC evolution: beyond selection on the immune genes. *Proc. Biol. Sci.* 276:657–665.
- Van Oosterhout, C., D. A. Joyce, and S. M. Cummings. 2006. Evolution of MHC class IIB in the genome of wild and ornamental guppies, *Poecilia reticulata*. *Heredity* 97:111–118.
- Von Siebenthal, B. A., A. Jacob, and C. Wedekind. 2008. Tolerance of whitefish embryos to *Pseudomonas fluorescens* linked to genetic and maternal effects, and reduced by previous exposure. *Fish Shellfish Immun.* 26:531–535.
- Wang, Q., G. M. Garrity, J. M. Tiedje, and J. R. Cole. 2007. Naïve Bayesian Classifier for Rapid assignment of rRNA Sequences into the New Bacterial taxonomy. *Appl. Environ. Microbiol.* 73:5261–5267.
- Wegner, K. M. 2008. Historical and contemporary selection of teleost MHC genes: did we leave the past behind? *J. Fish Biol.* 73:2110–2132.
- Yang, Z. 2007. PAML 4: a program package for phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24:1586–1591.

- Yang, Z., R. Nielsen, N. Goldman, and A.-M. K. Pedersen. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155:431–449.
- Yang, Z., W. S. W. Wong, and R. Nielsen. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* 22:1107–1118.
- Yannarell, A. C., and E. W. Triplett. 2005. Geographic and environmental sources of variation in lake bacterial community composition. *Appl. Environ. Microbiol.* 71:227–239.
- Yannarell, A. C., A. D. Kent, G. H. Lauster, T. K. Kratz, and E. W. Triplett. 2003. Temporal patterns in bacterial communities in three temporal lakes of different trophic levels. *Microb. Ecol.* 46:391–405.

## Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Figure S1.** Schematic illustration of the *Phoxinus phoxinus* MHCIIB duplicates DAB 1 (A) and DAB 3 (B). Shaded boxes represent exons. The positions of the primers used in this study are indicated by arrows.

**Figure S2.** Details on sequence reads and quality scores from the 454 run. (A) Histogram showing sequence read quality score frequencies for all reads, bacteria 16S rDNA reads, DAB1 and DAB3 reads. Red line indicates the average quality score per class of reads (B) Histogram showing sequence read length frequencies for all reads, bacteria 16S rDNA reads, DAB1 and DAB3 reads. Red line indicates the average quality score per class of reads (C) Correlation between reads qualities and reads length for each class of reads. Red lines represent the least square regression between read quality and length. R statistics from correlation Pearson tests and their significance (\*\*\*) significant at 99% level, ns nonsignificant at the 99% level) are indicated.

**Table S1.** List of primers used to (A) characterize MHC Class II B in *Phoxinus phoxinus*; (B) amplify bacterial 16S rDNA and genotype MHC Class II B in *P. phoxinus*.

**Text S1.** cDNA sequence amplified from *Phoxinus phoxinus* with FishEx12F-bis and FishC12S-R primers.

**Text S2.** MHCIIB partial Exon 2 DAB1 and DAB3 sequences amplified using 454 sequencing.

**Text S3.** 16s rDNA sequences amplified using universal bacteria primers and 454 sequencing.

## Appendix 1:

Description of the 454 sequencing run details, before read sorting and classification of the two MHC classIIb exon 2 and 16S rDNA sequences.

Amplicon	Number of reads	Mean read length	Minimum read length	Maximum read length	Mean quality score	Minimum quality score	Maximum quality score
16S rDNA	548	245	27	655	27	16	40
DAB1	7591	275	15	685	28	11	40
DAB3	7877	274	18	768	28	11	40

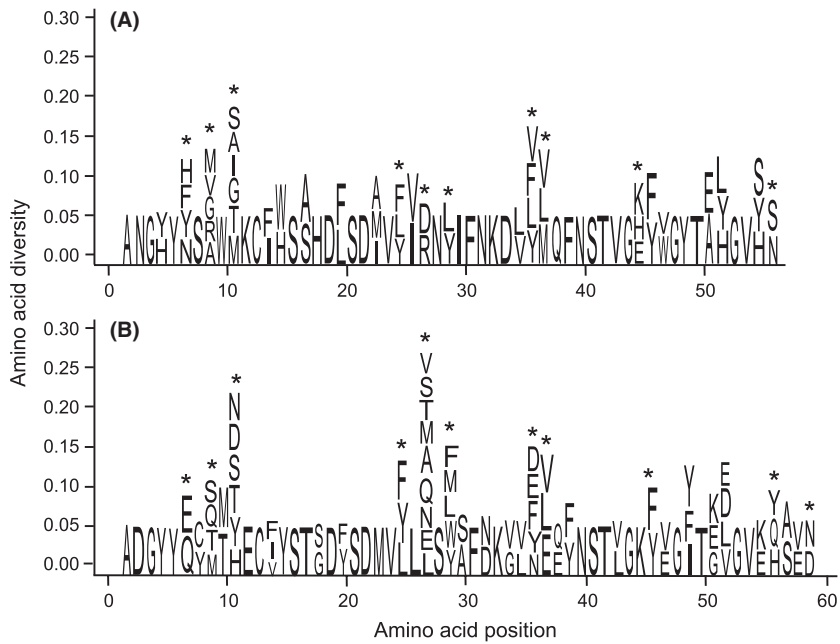


**Appendix 2:**



Unrooted neighbor-joining tree of *Phoxinus phoxinus* DAB exon 2 alleles (Jukes Cantor nucleotides distances) based on a 135 bp and 174 bp for DAB1 and DAB3, respectively. DAB alleles from additional cyprinids (*Abramis brama*, *Squalius cephalus*, *Cyprinus carpio*, *Barbus bocagei*, *Brachydanio rerio*, *Barbus intermedius*, *Rhodeus ocellatus*, and *Ladigesocypris ghigii*; Genbank accession number beside allele names) and one salmonid (*Salmo salar*) are included. “HAP” refers to true variants of MHCIIB for DAB1 (pink branches) and DAB3 (green branches) duplicates obtained for *Phoxinus phoxinus* in the present study. Bootstrap values >50% (1000 replicates) are shown.

### Appendix 3:



Sequence logo showing amino acid diversity for each amino acid position of the two MHC class IIB duplicates examined: DAB1 (A) and DAB3 (B). Human peptide-binding region (PBR) sites (Brown *et al.* 1993) are indicated with an asterisk. The amino acid positions refer to the beginning of MHC class IIB (MHCIIb) exon 2.