Genetic drift shaped MHC IIB diversity of an endangered anuran species within the Italian glacial refugium

L. Talarico[1] (iD), W. Babik[2], S. Marta[1,3,4] & M. Mattoccia[1]

1 Department of Biology, University of Tor Vergata, Rome, Italy
2 Institute of Environmental Sciences, Jagiellonian University, Kraków, Poland
3 Institute of Ecosystem Studies, National Research Council, Rome, Italy
4 Department of Environmental Sciences and Policy, University of Milan, Milan, Italy

Abstract

Highly polymorphic genes of the major histocompatibility complex (MHC) encode proteins involved in the immune response that protect vertebrates from parasites and pathogen infections. MHC variation is shaped by a complex and poorly understood interplay of selective and demographic forces. Studies of MHC variation provide biologically meaningful insights useful for conservation of threatened taxa such as amphibians, which are globally declining partly due to the impact of emerging infectious diseases (e.g. chytridiomycosis). Here, we characterized MHC class IIB variation in the Apennine yellow-bellied toad (*Bombina pachypus*), an endangered anuran, across its whole distribution range within the Italian peninsula. We compared MHC IIB diversity between refugial (REF) and post-glacial expansion (PGE) areas and tested for the correlation between MHC and neutral variation to quantify the role of drift in shaping MHC diversity. Overall MHC polymorphism was limited compared to other amphibian species. Despite the clear evidence of historical positive selection acting on antigen-binding sites, we found a significant correlation between the diversity of MHC and that of putatively neutral microsatellite markers, which suggests that genetic drift has contributed extensively to shaping MHC variation. MHC diversity was higher in the REF populations and decreased northwards into the PGE area, in accordance with the 'southern richness and northern purity' pattern. Past demographic events likely determined the reduction of MHC variation in the PGE area; in particular, the northern populations appeared to survive for a long time with a depleted MHC variation. We discussed recent population declines of *B. pachypus* in light of our findings.

Introduction

The major histocompatibility complex (MHC), a gene-dense region present in jawed vertebrates, contains many genes involved in the immune response (Kelley, Walter & Trowsdale, 2005). The most emblematic ones are classical MHC genes that encode proteins presenting antigens to T-cells, triggering adaptive immune response (Kumánovics, Takada & Lindahl, 2003). MHC class II proteins recognize and bind peptide antigens derived from extra-cellular pathogens. Each MHC molecule can bind a limited spectrum of peptides, determined by amino acids in the Antigen Binding Sites (ABS) (Piertney & Oliver, 2005). In class II genes, ABS are in the high variable exon 2 (Brown *et al.*, 1993; Tong *et al.*, 2006).

Extensive polymorphism of MHC genes (numerous alleles per locus and/or multiple loci) has adaptive significance, conferring potential resistance to a wide range of pathogens (Piertney & Oliver, 2005). Novel MHC alleles result from positive selection for amino acid replacement in ABS and allelic lineages are often retained over long periods and shared between species, resulting in trans-species polymorphism or TSP (Klein, Sato & Nikolaidis, 2007). Intra-population MHC diversity is believed to be maintained primarily by pathogen-driven balancing selection with additional mechanisms, such as sexual selection, contributing in some cases (Piertney & Oliver, 2005). Recent studies based on experiments in semi-natural settings provided evidence that novel alleles are strongly favored (Phillips *et al.*, 2018). Furthermore, spatially and temporally fluctuating selection may increase inter-population MHC diversity (Spurgin & Richardson, 2010), as found in species inhabiting broad geographical ranges (Ekblom *et al.*, 2007; Loiseau *et al.*, 2009). However, non-adaptive processes also affect MHC variation. Genetic drift has been demonstrated to play a central role in shaping both MHC and neutral variation (Radwan, Biedrzycka & Babik, 2010; Cortázar-Chinarro *et al.*, 2017) and is expected to be particularly important in stenoecious taxa

with low dispersal ability, such as amphibians (Blaustein, Wake & Sousa, 1994). MHC differentiation among populations is probably determined by a complex and so far poorly understood interplay between local adaptation and drift. The distribution of MHC variation may also reflect demographic history of populations. Loss of genetic variation during expansion out of glacial refugia has been observed in many temperate species (Hewitt, 2004). A similar pattern has been found in the few studies comparing MHC variation between glacial refugia (REF) and post-glacial expansion (PGE) areas in amphibians (Babik *et al.*, 2009; Zeisset & Beebee, 2014), providing insight into the role of evolutionary forces and demographic events in shaping MHC diversity. Such results raise some interesting questions about population ability to survive for a long time despite the depleted MHC variation (Radwan *et al.*, 2010) and these issues deserve further investigation, especially in endangered species.

Amphibians are globally threatened and are experiencing declines, the causes of which are not fully understood (Stuart *et al.*, 2004). Emerging infectious diseases contribute to the declines and the impact of novel widespread pathogens (e.g. *Batrachochytrium dendrobatidis*) is well documented (Kilpatrick, Briggs & Daszak, 2010). Some studies found a relationship between MHC polymorphism and susceptibility to chytridiomycosis (Savage & Zamudio, 2011), or linked structural MHC features to *B. dendrobatidis* resistance across a number of amphibian species (Bataille *et al.*, 2015). These findings emphasize the importance of variation in immune genes for conservation (Sommer, 2005). Unfortunately, data on MHC variation in amphibians are generally scarce, in particular studies comparing MHC variation of threatened species between REF and PGE areas.

Here, we investigated the variation of MHC class II across the whole range of the Apennine yellow-bellied toad, *Bombina pachypus* (Bonaparte, 1838). The current range of this species is the result of post-glacial expansions from southern Italian refugia (Canestrelli *et al.*, 2006). *Bombina pachypus* is an endangered amphibian experiencing severe declines across much of its range (Andreone *et al.*, 2009) that spans the Italian peninsula. The causes of these declines are not fully understood, in particular the involvement of chytridiomycosis is not clear (Stagni *et al.*, 2004; Canestrelli, Zampiglia & Nascetti, 2013). In this study, we aimed to: (1) characterize MHC class IIB in *B. pachypus*; (2) quantify variation within and differentiation among populations; (3) compare MHC IIB variation between REF and PGE areas and test for the correlation between MHC and neutral variation.

Materials and methods

Samples and DNA extraction

We collected 283 *B. pachypus* samples from 42 populations across the whole species range. Additionally, we obtained 17 samples of *Bombina variegata* (Linnaeus, 1758), as a closely related reference species, from 10 sites in Italy and one in Albania (Data S1). Genomic DNA was extracted from buccal swabs using the GenElute™ Mammalian Genomic DNA Miniprep Kit (Sigma-Aldrich, St. Louis, MO, USA) following the mammalian tissue protocol.

MHC amplification, amplicon sequencing and post-processing

We amplified a 157 bp fragment of MHC class IIB exon 2 with primers Bobom_MHCIIexon2-F2 and Bobom_MHCIIexon2-R1 (Hauswaldt *et al.*, 2007). Primers were barcoded with four bp tags and contained a fragment of Illumina Nextera adaptors (Data S2). By combining four forward and four reverse primers, we individually barcoded up to 16 amplicons. PCR reactions in a total volume of 20 μL contained 20–100 ng DNA, 0.375 μM each primer, 1× Phusion Buffer HF, 200 μM each dNTP and 0.4 U of Phusion Hot Start II High-Fidelity DNA polymerase (Thermo Fisher Scientific, Waltham, MA, USA). Amplification was carried out in a 2720 Thermal cycler (Applied Biosystems, Foster City, CA, USA) as follows: initial denaturation at 98°C for 2 min, followed by 30 cycles at 98°C for 10 s, 56°C annealing for 20 s, 72°C for 15 s and a final elongation step at 72°C for 5 min. We prepared 21 batches (Data S3) pooling 5–16 barcoded amplicons approximately equimolarly as assessed by visual examination of band intensities on 1% agarose gels. A commercial company carried out the library preparation (i.e. adding the remaining part of Illumina adapters and indexes by a second PCR step) and the Illumina sequencing (MiSeq 2 × 300 runs). Two additional amplicons without barcode were sequenced as well (Data S3).

Paired-end reads were merged (minimum overlap 20 bp) in FLASH (Magoč & Salzberg, 2011). We discarded merged reads differing from the expected length by > 10 bp and retained the ones with an average base call accuracy greater than 99.9% (Q > 30) using the PRINSEQ online tool (Schmieder & Edwards, 2011). Finally, AmpliCHECK (Sebastian *et al.*, 2016) allowed primer removing, de-multiplexing and quantification of each variant per amplicon.

MHC genotyping pipeline and validation of alleles

AmpliCHECK (Sebastian *et al.*, 2016) was used for initial quality control. Tag-switching (Carlsen *et al.*, 2012) appeared the main source of artifacts in our data, leading to the detection of common variants in most amplicons. Other artifacts (i.e. chimeras and low-frequency variants generated by substitutions from high-frequency variants in the same amplicon) were less common, achieving no more than 5% of the overall maximum per-amplicon frequency (PAF).

Error-correction and genotyping were performed by the adjustable clustering approach implemented in AmpliSAS (Sebastian *et al.*, 2016). We set minimum PAF = 15% for the clustering step and minimum PAF = 5% for filtering, chimera removal option was on and all the other settings were in default mode.

The genotypes obtained by AmpliSAS still included variants that could be explained as the result of extensive tag-switching that likely occurred during the second PCR step, therefore we cross-validated genotypes by Sanger sequencing. All samples were amplified using primers without tags under the previously described conditions, and PCR products were sequenced with an ABI3730xl automated DNA sequencer (Applied Biosystems). The resulting chromatograms were manually checked in

order to validate individual genotypes previously obtained by AmpliSAS. We rejected variants that did not match peaks of the chromatogram in at least two sites. Note that this method was effective because of the low number of amplified MHC loci, the absence of indels and the high number of polymorphic sites (see Results).

MHC IIB characterization

MHC alleles of *B. pachypus* and *B. variegata* were aligned using MUSCLE in MEGA7 (Kumar, Stecher & Tamura, 2016). The overall average pairwise nucleotide distances (Kimura 2-Parameter model), the amino acid distances (Poisson correction model) and the synonymous (d_S) and non-synonymous (d_N) substitution rates (modified Nei-Gojobori method with Jukes-Cantor correction) were computed in MEGA7.

We looked for recombination using several methods implemented in RDP v.4.95 (Martin *et al.*, 2015): SisScan, PhylPro and Bootscan (window = 30, step size = 3); RDP (window = 30); MaxChi and Chimera (variable sites = 10); GENECONV (default settings). The Bonferroni correction was applied for multiple tests ($P < 0.05$). Additionally, the GARD algorithm (Kosakovsky Pond *et al.*, 2006) in the Datamonkey server (http://classic.datamonkey.org/) was used to detect possible recombination breakpoints in two separate alignments (*B. pachypus* and *B. variegata*).

We tested for positive selection using two approaches. Firstly, we evaluated three codon-based evolution models in PAML v.4.8 (Yang, 2007): (1) M0 considers a single ω (dN/dS ratio) value for all codons; (2) M7 assumes purifying selection and neutrality (ω ≤ 1) with the beta distribution approximating ω variation; (3) M8 allows positive selection (i.e. a proportion of sites evolving with ω > 1) with the beta distribution approximating ω variation. The best model was identified using the Akaike information criterion (AIC). Secondly, we performed the one-tailed *Z*-test in MEGA7 to verify if ω > 1 in three alignment partitions: (1) all sites; (2) ABS corresponding to those in human MHC (Brown *et al.*, 1993; Tong *et al.*, 2006); (3) non-ABS. Positively selected codons were detected using six methods: the single-likelihood ancestor counting (SLAC), the fixed effects likelihood (FEL), the random-effects likelihood (REL) analysis, the fast unconstrained Bayesian approximation (FUBAR), and the mixed-effects model of evolution (MEME) (Kosakovsky Pond & Frost, 2005; Murrell *et al.*, 2012) implemented on Datamonkey; PAML was used to perform the Bayes empirical Bayes (BEB) procedure.

As MHC alleles often show complex genealogies, a phylogenetic network was used to examine the relationship among variants and to visualize conflicting signals. The network was built in SplitsTree 4.14 (Huson & Bryant, 2006) using the neighbour-net algorithm (*P*-distance method) and including all sequences available in GenBank for the genus *Bombina*.

MHC variation

As it was not possible to assign MHC alleles to loci, variation within populations was measured as the number of alleles (MHC_{AN}) and allelic richness (MHC_{AR}). MHC_{AR} was calculated in samples with more than four individuals as the average MHC_{AN} present in 1000 random samples of five individuals drawn from the same population. We also recorded the number of private alleles per population (MHC_{PA}). Pie charts displaying MHC allele frequencies were mapped and edited in QGIS 2.18 (Quantum GIS Development Team, http://qgis.osgeo.org).

Microsatellite variation and genetic structure

Putatively neutral variation was assessed with nine polymorphic microsatellite loci (hereafter STR). We adjusted a previously generated dataset (Ciambotta, 2015) to obtain STR genotypes of 283 *B. pachypus* used in the present study (see Data S5 for details). To allow comparison with MHC diversity, STR_{AN} and STR_{PA} were computed in GenAlEx 6.5 (Peakall & Smouse, 2012). We calculated STR_{AR} by rarefaction (minimum number of sampled alleles = 10) in populations of at least five individuals with the PopGenReport R package (Adamack & Gruber, 2014). We used STRUCTURE v.2.3.4 (Pritchard, Stephens & Donnelly, 2000) to identify genetic clusters based on STR data, as described in Data S6.

Relationship between MHC and STR variation

A linear model (LM) tested the association between MHC variation (the response variable) and STR diversity using STR cluster membership as the second predictive variable. The model included both additive main effects and their interaction. Two versions of the analysis were performed: the first one (AR model) employed STR_{AR} and MHC_{AR}, while the second one (AN model) used STR_{AN} and MHC_{AN} as measures of genetic diversity. In the AN model, we introduced population weights (i.e. the number of individuals divided by the average population size) to take into account differences in sample sizes. The data were analyzed using the lm function in R (R Development Core Team, 2018).

Results

Characterization of MHC IIB alleles

We sequenced fragments of MHC class II B exon 2 in 283 *B. pachypus* and 17 *B. variegata* individuals from 42 and 11 populations, respectively (Data S4). Individuals showed 1–4 alleles (mean: 2.06; SE: 0.71) indicating at least two loci co-amplified.

Overall, we found 46 MHC IIB alleles (GenBank accessions: MH220884-MH220929), 157 bp each, which translated into 44 different protein sequences (52 amino acids long); no stop codons were detected (Fig. 1). Twice as many alleles were found in *B. pachypus* as in *B. variegata* (32 vs. 15), but many more *B. pachypus* individuals were sampled (Table 1); only one allele (Bopa-DXB*04) was shared between the two species. Unfortunately, neither sequence similarity nor patterns of allele co-occurrence allowed assignment of alleles to loci. However, in several *B. pachypus* populations from the central

and northern regions, individuals with two alleles predominated (usually DXB*02 with *03, and DXB*04 with *05), suggesting that alleles in each pair were from different loci. The overall divergence between alleles was high (Table 1). Out of the eight methods used to test for recombination, only SisScan analysis found a single significant recombination event (results not showed), indicating negligible evidence of recombination.

The non-synonymous divergence exceeded synonymous divergence in the entire alignment ($P < 0.05$). In particular, d_N was five times greater than d_S in the putative ABS ($P < 0.01$), while the rates were almost the same in non-ABS sites (Table 2). Consistently, the M8 model, allowing for positive selection, fitted our data best ($\Delta AIC = 92.443$; Data S7). We identified 14 codons as putative targets of positive selection. Most of them are likely involved in peptide binding as 10 out of 14 (71.4%) agreed with human ABS (P-value < 0.001, Fisher's Exact Test) (Fig. 1).

The neighbour network including four *Bombina* species did not show clearly defined relationships between alleles (Fig. 2). However, all *B. pachypus* alleles appeared to group together, except for 3F and Bopa-DXB*04 (the only variant shared with *B. variegata*), suggesting a limited TSP, while *B. variegata* formed several mixed clusters with *B. bombina* and *B. orientalis* alleles. The three alleles most common and widespread in *B. pachypus* (i.e. Bopa-DXB*02, *03 and *04) were highly divergent from each other (Figs 2 and 4).

MHC diversity within populations and geographical variation

The average number of MHC alleles per population (MHC$_{AN}$), calculated for all 42 *B. pachypus* populations (average sample size = 6.74 ± 4.48 SE) was 3.55 (± 1.65 SE). The average MHC$_{AR}$ calculated for the sample size of five individuals (24 populations) was 3.33 (± 1.40 SE). We observed a higher MHC diversity (i.e. AN, AR and PA) within the southern regions, particularly in populations 34, 38, 39 and 40 (Fig. 3 and Data S8).

STRUCTURE analysis based on nine STR loci found two genetic clusters: the central-northern (C-N cluster, 32 populations, 203 individuals) and the southern populations (S cluster, nine populations, 70 individuals) (Data S6). The two clusters roughly corresponded to the PGE and REF areas respectively, in agreement with previous studies (Canestrelli *et al.*, 2006; Ciambotta, 2015).

We found 22 and 17 MHC alleles in S and C–N clusters, respectively (Fig. 4). Within the C–N cluster two sub-groups, each relatively homogeneous with respect to MHC IIB, were observed in the northern part, alleles DXB*02 and *03 were widespread (these alleles, the most divergent within the species, represent most likely separate loci) (Fig. 3), while in central Italy DXB*04 was very common, usually co-occurring with *05 or *06. MHC diversity appeared to decrease northwards (Figs 3 and 4).

Correlation between MHC and STR variation

The association between MHC and STR variation was tested including the cluster membership as a categorical explanatory variable; the single admixed locality (pop. 31) was excluded

from this analysis. There was no significant interaction between clusters and STR diversity so this interaction was dropped from the final model. MHC diversity was positively correlated with STR variation expressed as both AR and AN, and the effect of cluster was not significant (Table 3). After removing cluster from the model, STR variation explained 26% (AR) or 42% (AN) of variation in MHC diversity between populations (Fig. 5).

Discussion

MHC IIB in *B. pachypus* and *B. variegata* is duplicated, with at least two loci simultaneously amplified, consistent with results reported for *B. bombina* (Hauswaldt *et al.*, 2007) and several species of *Bufo* (Zeisset & Beebee, 2013) and *Rana* (Kiemnec-Tyburczy *et al.*, 2010). We probably characterized most of the current MHC class II variability detectable using internal exon primers in *B. pachypus*, as we sampled a large portion of the species range. The allelic diversity is similar to that of *Bufo calamita* (38 DAB alleles in 325 individuals in a biogeographical scale study; Zeisset & Beebee, 2014), but it is considerably lower than in several non-European anurans, such as *Rana omeimontis* (130 alleles from 215 individuals in at least three loci; Li *et al.*, 2016) and *Leiopelma hochstetteri* (74 alleles isolated from 121 individuals in a single locus; Lillie *et al.*, 2015). Sequence polymorphism is similar to that of *B. calamita* (Zeisset & Beebee, 2014) and *Rana temporaria* (Zeisset & Beebee, 2009). The estimates of MHC variation are not fully comparable between these studies as they used fragments of exon 2 of various lengths. However, because MHC alleles are often rather divergent and differences occur throughout the exon length, the differences in the assayed fragment length should have only a minor effect on estimates of polymorphism. *B. pachypus* shows a lower MHC diversity than *B. variegata*, and they exhibit a limited TSP although the two species are closely related with an estimated divergence time of 2–3 Ma. Curiously, we observed TSP in the other congeneric species despite the fact they are more distantly related (Pabijan *et al.*, 2013). Our data suggest, positive selection affected the evolution of MHC IIB exon 2 in *Bombina*, as indicated by an excess of non-synonymous divergence between alleles.

MHC variation is generally low within *B. pachypus* populations and it correlates positively with the neutral variation. The spatial distribution of MHC diversity follows the 'southern richness and northern purity' pattern detected previously for this species (Canestrelli *et al.*, 2006) and many taxa inhabiting regions affected by the Pleistocene climatic fluctuations (Hewitt, 2004). This suggests that drift shapes MHC diversity over the entire range of *B. pachypus*, and the current distribution of MHC variability appears to reflect predominately historical rather than selective processes, despite strong signatures of historical positive and balancing selection. Interestingly, a similar pattern was detected across *Triturus cristatus* (Babik *et al.*, 2009) and *B. calamita* ranges (Zeisset & Beebee, 2014): a loss of both MHC and microsatellites diversity, ascribed to the effects of genetic drift, was observed in the PGE areas. Conversely, in other amphibians, such as *R. omeimontis* (Li *et al.*,
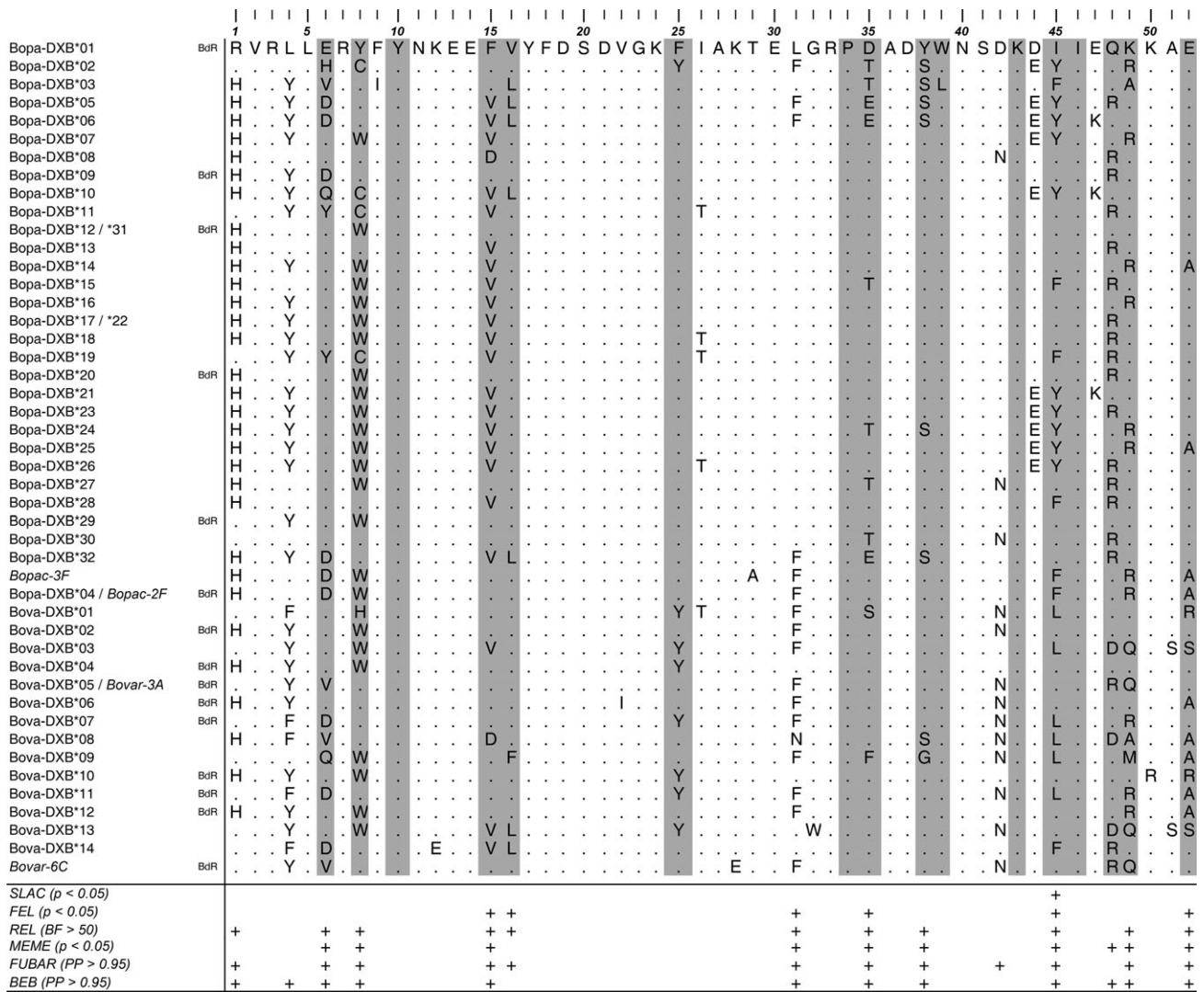
Figure 1 Amino acid alignment of MHC IIB alleles detected in this study and obtained from literature data (in *italic*; GenBank accessions: EF210744 - EF210747) in two *Bombina* species: *B. pachypus* (Bopa) and *B. variegata* (Bova) (note that allele DXB-Bopa*04 is shared between the two species). Allele codes separated by a slash (/) indicate DNA sequences translating into an identical amino acid sequence. Positions shaded in gray identify putative ABS in human HLA (Brown *et al.*, 1993; Tong *et al.*, 2006). Plus sign (+) marks codons under positive selection as inferred by each test (*P* = *P*-value; PP = posterior probability; BF = Bayes factor). Acronym BdR indicates variants resistant to *Batrachochytrium dendrobatidis* infections, according to the protective conformation of P9 binding pocket of MHC II molecules (Bataille *et al.*, 2015).

Table 1 Summary of MHC allelic polymorphism in *B. pachypus* and *B. variegata*: overall nucleotide (Kimura 2-Parameters) and amino acid (Poisson correction) average pairwise distance, with standard errors (SE, 1000 bootstrap replicates) given in parentheses

Group	# individuals	# alleles	K2P nt-distance ($\pm$ SE)	Poisson aa-distance ($\pm$ SE)
B. pachypus	284	33	0.073 (0.012)	0.142 (0.033)
B. variegata	18	16	0.096 (0.016)	0.189 (0.042)
Both	302	48	0.087 (0.013)	0.173 (0.035)

Data include alleles previously detected by Hauswaldt *et al.* (2007) in a single specimen for each species.

2016) and *L. hochstetteri* (Lillie *et al.*, 2015), diversifying selection is the main force determining MHC variation and leading to high interpopulation differentiation.

Genetic richness in southern Italy is believed to fit the 'refugia within refugia' model (Gómez & Lunt, 2007) as confirmed in a number of amphibians, such as *Salamandrina terdigitata*

Table 2 The synonymous (d_S) and non-synonymous (d_N) nucleotide substitution rates (Nei-Gojobori method with Jukes-Cantor correction), and results of positive selection test (one-tailed Z-test) in different partitions of exon 2 in *B. pachypus* and *B. variegata* MHC IIB alleles

Partition	d_N ($\pm$SE)	d_S ($\pm$SE)	Z	P
All sites	0.10 (0.02)	0.05 (0.02)	1.69	**0.047**
ABS	0.25 (0.07)	0.05 (0.03)	2.767	**0.003**
non-ABS	0.04 (0.02)	0.05 (0.03)	−0.385	1

Standard errors (SE) are showed in parentheses; significant *P*-values are in bold.

(Mattoccia *et al.*, 2011) and *Rana italica* (Canestrelli, Cimmaruta & Nascetti, 2008). In reference to *B. pachypus*, Canestrelli *et al.* (2006) argued that both allopatric differentiation and demographic stability occurred in the deep southern part of the range (i.e. Calabria region) during glaciations, leading to strong genetic structuring in the above-mentioned refugial area. Accordingly, we found a similar pattern with respect to adaptive (MHC) variability, although our data suggest that the REF area could be more extensive than previously supposed. We found relatively high MHC diversity in the REF populations, and a certain number of MHC alleles are specific to restricted
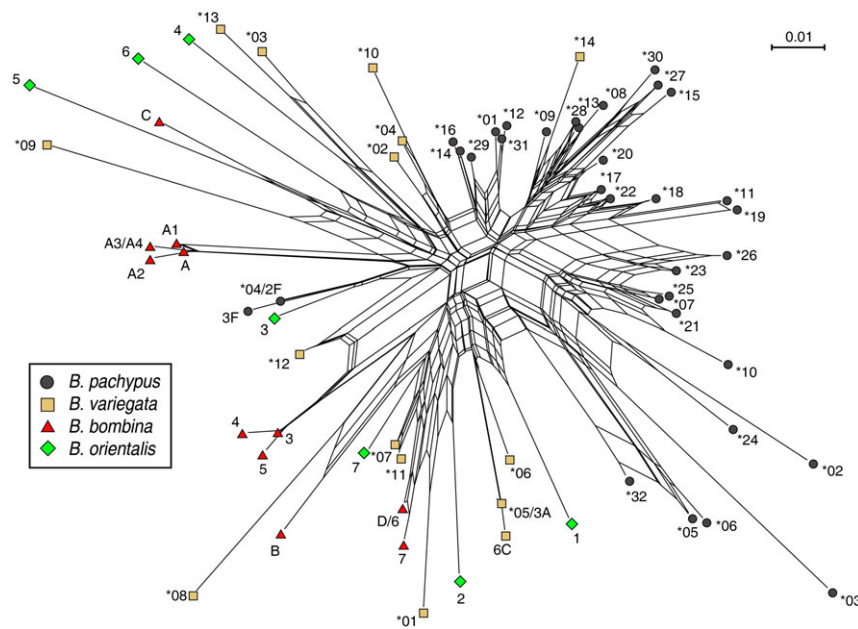
Figure 2 Phylogenetic relationship of MHC IIB alleles in *Bombina* based on neighbour-net network (uncorrected p-distance; 156 bp of exon 2). Allele IDs marked with asterisks (*) refer to sequences detected in this paper (Fig. 1). Unmarked allele IDs refer to sequences previously published in GenBank for: *B. orientalis* (accessions: KJ679325-KJ679331), *B. bombina* (accessions: EF210736-EF210743, EF210770-EF210774), *B. variegata* (accessions: EF210744-EF210745) and *B. pachypus* (accessions: EF210746-EF210747).
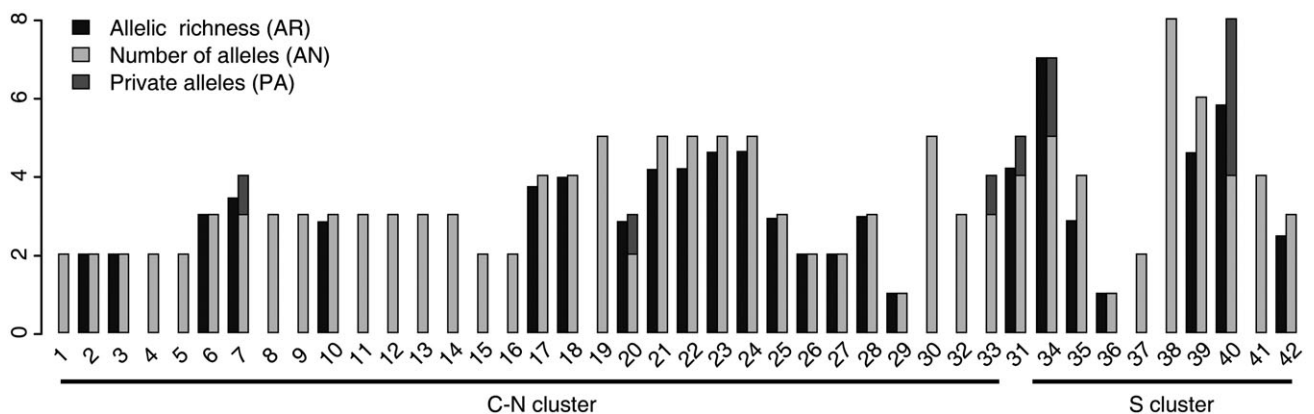
Figure 3 MHC IIB diversity across 42 populations of *B. pachypus* with cluster membership (C-N, Central-Northern, and S, southern). Populations are arranged approximately from northern (left) to southern (right) Italy.
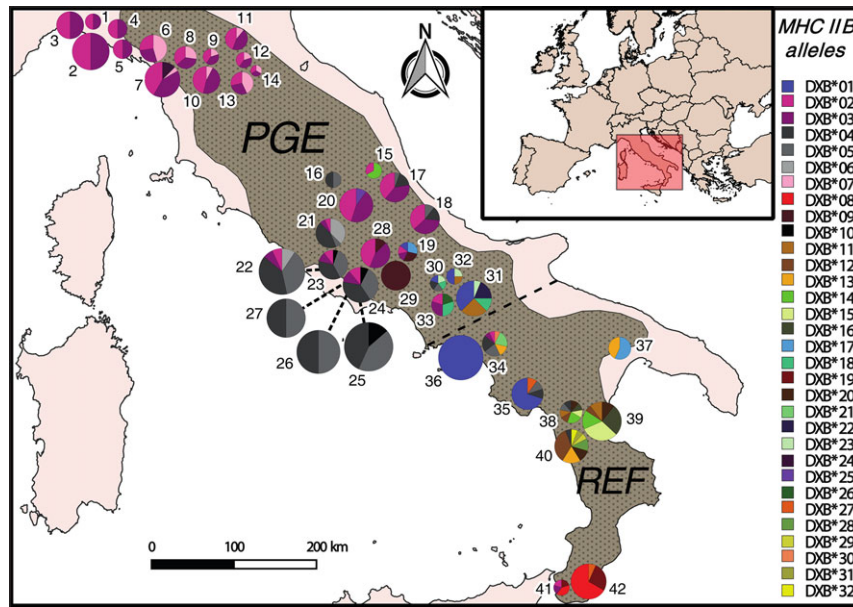
Figure 4 Frequencies of MHC IIB alleles in 42 *B. pachypus* populations. Pie charts, proportional to sample sizes, display allele frequencies per population (i.e. the number of individuals having the allele on the total number of alleles). The thick dashed line indicates the approximate geographic border between PGE and REF areas. *Bombina pachypus* distribution area is showed in the background (downloaded from IUCN, http://www.iucnredlist.org).

Table 3 Results of the linear models with MHC diversity as the response variable, and STR diversity and cluster membership as predictors

	Additive AN model			Additive AR model		
	Estimate (± SE)	t	P	Estimate (± SE)	t	P
Intercept	−0.213 (0.735)	−0.289	0.774	0.150 (1.300)	0.116	0.909
STR diversity	1.829 (0.367)	4.982	**<0.001**	1.673 (0.724)	2.311	**0.032**
S cluster	−0.723 (0.603)	−1.199	0.238	−0.542 (0.872)	−0.621	0.541

Diversity was measured as number of alleles in the Additive AN model ($F_{(2,38)}$ = 15.12, $P < 0.001$; adjusted R^2 = 0.41) or allelic richness in the Additive AR model ($F_{(2,20)}$ = 3.702, $P = 0.043$; adjusted R^2 = 0.20). Estimates of C–N cluster intercept (intercept), slope (STR diversity) and S cluster intercept (S cluster) are given for each model with standard errors (SE) in parentheses. Significant P-values (t-test) are in bold.
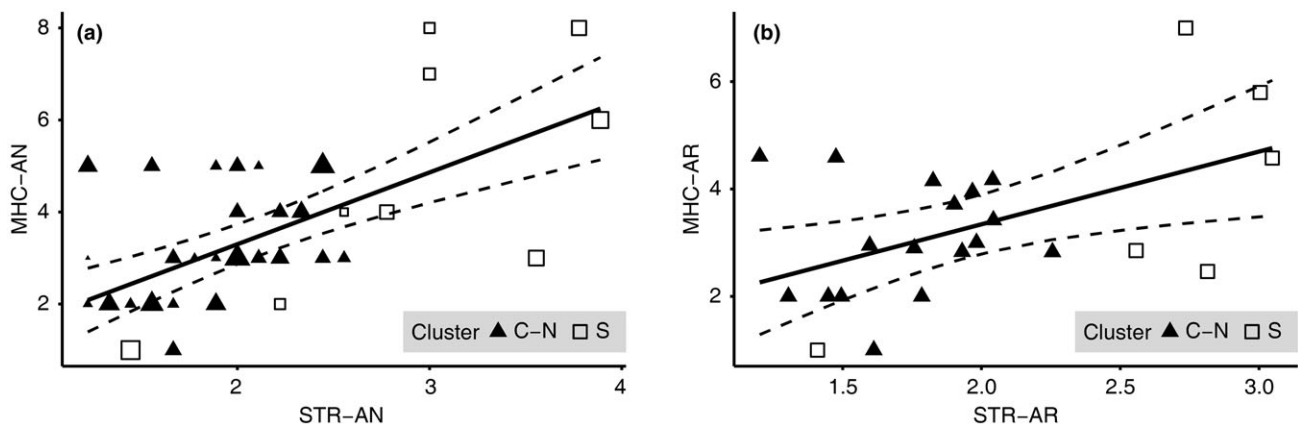
Figure 5 Association between MHC and STR diversity: (a) AN, number of alleles ($F_{(1,39)}$ = 28.48, $P < 0.001$), R^2 = 0.42; (b) AR, allelic richness ($F_{(1,21)}$ = 7.229, $P = 0.014$), R^2 = 0.26. The dashed curves indicate 95% confidence intervals. Symbols in Fig. 5a are proportional to population sizes.

areas (e.g. *01, *12, *15, *16, *19 and *20) or private for a single population. This pattern may reflect the effect of genetic drift and limited gene flow, but it could also mirror patchy local adaptation driven by spatial and temporal fluctuating selection (Spurgin & Richardson, 2010). As argued by Zeisset & Beebee (2014), local adaptation in the southern regions could be favored by a stronger parasite pressure along a latitudinal gradient (i.e. parasite diversity decreases from equator toward the poles; Guernier, Hochberg & Guégan, 2004), but further studies on distribution of pathogens in the Italian peninsula are required to test such hypothesis.

Both MHC and neutral variation are low within PGE populations, and decline northward, consistent with post-glacial expansions from southern refugia (Hewitt, 2004). During recolonization, multiple founder events and bottlenecks likely reduced MHC diversity as suggested by the presence in PGE area of two distinct population groups, each with very low MHC diversity. Although a selective sweep scenario cannot be ruled out entirely, the proposed scenario appears more parsimonious due to the observed correlation between STR and MHC variation. Interestingly, we found several alleles with a local distribution within the PGE areas (e.g. *06, *07, *09, *10, *18, *23). As sampling deficiency in REF populations appears unlikely, though possible, these alleles may have been lost from REF. Alternatively, these alleles could have originated from micro-recombination events (Richman *et al.*, 2003) involving southern alleles, which recombination methods may fail to detect due to the short length of involved fragments (as suggested by a visual inspection).

Loss of MHC diversity was observed in recently bottlenecked populations of different taxa, while the long-time persistence of limited MHC variation is not so common, and implications for population viability are poorly understood (Radwan *et al.*, 2010). The possible long-term adverse effects of a depleted MHC variation in most PGE populations may have been mitigated to some extent by the retention of a haplotype carrying two loci with highly differentiated alleles (Bopa-DXB*02 and *03). Animals carrying this haplotype could possibly bind a broader range of potential antigens increasing fitness as predicted by the divergent allele advantage hypothesis (Lenz, 2011). Alternatively, MHC depletion may not constitute a risk to population viability even in a relatively long time perspective, as in the case of *T. cristatus* (Babik *et al.*, 2009; Hoglund *et al.*, 2015).

Four amino acid positions known to affect the properties of the MHC II P9 binding pocket have been proposed to be involved in resistance to chytridiomycosis also in genus *Bombina* (for details see Bataille *et al.*, 2015). The putative protective conformation of P9 pocket is surprisingly not widely conserved in *B. pachypus* alleles (being present only in eight out of 33; Fig. 1) and only Bopa-DXB*04 and *01 are quite widespread across populations. On the other hand, a larger fraction of *B. variegata* variants (10 out of 16, Fig. 1) appear to be *B. dendrobatidis* protective, possibly giving *B. variegata* a stronger resistance to the emerging chytridiomycosis. Additionally, just 52.3% of *B. pachypus* populations shows at least one protective allele, and populations in the north and deep south lack any protective variants (see Data S9), although *B. dendrobatidis* was recorded in those areas in recent and/or

historical samples (Canestrelli *et al.*, 2013). Declines of *B. pachypus* populations are observed particularly in the central-northern distribution range (Andreone *et al.*, 2009), but the causes are not understood and it is difficult to assess the role of chytridiomycosis. Our data suggest that the recent and severe declines should not be directly correlated with MHC loss of diversity that occurred long ago, but we cannot reject the involvement of recent emerging diseases especially in populations with reduced MHC variation.

This paper provides further insight into the role of genetic drift in shaping pattern of MHC diversity, focusing on the Italian glacial refugium, and we found evidence for a long-term survival of populations with depleted MHC variation. To the best of our knowledge, this study reports the first MHC data about an Italian amphibian and we offer useful information for future conservation strategies of the endangered *B. pachypus*. Both MHC and genomic neutral variation should be used to define Management Units. Moreover, as the MHC polymorphism is significantly reduced in *B. pachypus*, the retention of as much as possible MHC variability is supposed likely to be essential and it could not be achieved by protecting populations in REF areas only.

Acknowledgements

Authors' contributions

LT, WB and MM conceived and designed the study. LT performed the lab experiments. LT, WB and SM analyzed the data. LT and WB wrote the paper. MM and SM revised the manuscript.

References

Adamack, A.T. & Gruber, B. (2014). PopGenReport: simplifying basic population genetic analyses in R. *Methods Ecol. Evol.* **5**, 384–387.

Andreone, F., Corti, C., Sindaco, R., Romano, A., Giachi, F., Vanni, S. and Delfino, G. 2009. *Bombina pachypus*. The IUCN Red List of Threatened Species 2009: e.T54450A86629977. https://doi.org/10.2305/iucn.uk.2009.rlts.t54450a11147957. Downloaded on 05th April 2018.

Babik, W., Pabijan, M., Arntzen, J.W., Cogalniceanu, D., Durka, W. & Radwan, J. (2009). Long-term survival of a urodele

amphibian despite depleted major histocompatibility complex variation. *Mol. Ecol.* **18**, 769–781.

Bataille, A., Cashins, S.D., Grogan, L., Skerratt, L.F., Hunter, D., McFadden, M., Scheele, B., Brannelly, L.A., Macris, A., Harlow, P.S., Bell, S., Berger, L. & Waldman, B. (2015). Susceptibility of amphibians to chytridiomycosis is associated with MHC class II conformation. *Proc. R. Soc. B Biol. Sci.* **282**, 20143127–20143127.

Blaustein, A.R., Wake, D.B. & Sousa, W.P. (1994). Amphibian declines: judging stability, persistence, and susceptibility of populations to local and global extinctions. *Conserv. Biol.* **8**, 60–71.

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

Canestrelli, D., Cimmaruta, R., Costantini, V. & Nascetti, G. (2006). Genetic diversity and phylogeography of the Apennine yellow-bellied toad *Bombina pachypus*, with implications for conservation. *Mol. Ecol.* **15**, 3741–3754.

Canestrelli, D., Cimmaruta, R. & Nascetti, G. (2008). Population genetic structure and diversity of the Apennine endemic stream frog, *Rana italica* - Insights on the Pleistocene evolutionary history of the Italian peninsular biota. *Mol. Ecol.* **17**, 3856–3872.

Canestrelli, D., Zampiglia, M. & Nascetti, G. (2013). Widespread occurrence of *Batrachochytrium dendrobatidis* in contemporary and historical samples of the endangered *Bombina pachypus* along the Italian Peninsula. *PLoS ONE* **8**, e63349.

Carlsen, T., Aas, A.B., Lindner, D., Vrålstad, T., Schumacher, T. & Kauserud, H. (2012). Don't make a mista(g)ke: is tag switching an overlooked source of error in amplicon pyrosequencing studies? *Fungal Ecol.* **5**, 747–749.

Ciambotta, M. (2015). *Genetic structure and population dynamics of Bombina pachypus (Amphibia: Anura)*. PhD thesis, University of Rome Tor Vergata.

Cortázar-Chinarro, M., Lattenkamp, E.Z., Meyer-Lucht, Y., Luquet, E., Laurila, A. & Höglund, J. (2017). Drift, selection, or migration? Processes affecting genetic differentiation and variation along a latitudinal gradient in an amphibian. *BMC Evol. Biol.* **17**, 189.

Ekblom, R., Saether, S.A., Jacobsson, P., Fiske, P., Sahlman, T., Grahn, M., Kålås, J.A. & Höglund, J. (2007). Spatial pattern of MHC class II variation in the great snipe (*Gallinago media*). *Mol. Ecol.* **16**, 1439–1451.

Gómez, A. & Lunt, D.H. (2007). Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. In *Phylogeography of Southern European Refugia: 155–188*. Weiss, S. & Ferrand, N. (Eds). Dordrecht: Springer, Netherlands.

Guernier, V., Hochberg, M.E. & Guégan, J.-F. (2004). Ecology drives the worldwide distribution of human diseases. *PLoS Biol.* **2**, e141.

Hauswaldt, J.S., Stuckas, H., Pfautsch, S. & Tiedemann, R. (2007). Molecular characterization of MHC class II in a nonmodel anuran species, the fire-bellied toad *Bombina bombina*. *Immunogenetics* **59**, 479–491.

Hewitt, G.M. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philos. Trans. R. Soc. B Biol. Sci.* **359**, 183–195.

Hoglund, J., Wengstrom, A., Rogeli, B. & Meyer-Lucht, Y. (2015). Low MHC variation in isolated island populations of the Natterjack toad (*Bufo calamita*). *Conserv. Genet.* **16**, 1007–1010.

Huson, D.H. & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267.

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

Kiemnec-Tyburczy, K.M., Richmond, J.Q., Savage, A.E. & Zamudio, K.R. (2010). Selection, trans-species polymorphism, and locus identification of major histocompatibility complex class IIβ alleles of New World ranid frogs. *Immunogenetics* **62**, 741–751.

Kilpatrick, A.M., Briggs, C.J. & Daszak, P. (2010). The ecology and impact of chytridiomycosis: an emerging disease of amphibians. *Trends Ecol. Evol.* **25**, 109–118.

Klein, J., Sato, A. & Nikolaidis, N. (2007). MHC, TSP, and the origin of species: from immunogenetics to evolutionary genetics. *Annu. Rev. Genet.* **41**, 281–304.

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

Kosakovsky Pond, S.L., Posada, D., Gravenor, M.B., Woelk, C.H. & Frost, S.D.W. (2006). Automated phylogenetic detection of recombination using a genetic algorithm. *Mol. Biol. Evol.* **23**, 1891–1901.

Kumánovics, A., Takada, T. & Lindahl, K.F. (2003). Genomic organization of the mammalian MHC. *Annu. Rev. Immunol.* **21**, 629–657.

Kumar, S., Stecher, G. & Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874.

Lenz, T.L. (2011). Computational prediction of MHC II-antigen binding supports divergent allele advantage and explains trans-species polymorphism. *Evolution* **65**, 2380–2390.

Li, J., Shen, H., Wang, H., Zhao, M., Luo, Z. & Wu, H. (2016). Diversifying selection is the dominant factor affecting the geographical variation of MHC class II genes in the Omei tree frog. *J. Zool. (Lond.)* **300**, 197–204.

Lillie, M., Grueber, C.E., Sutton, J.T., Howitt, R., Bishop, P.J., Gleeson, D. & Belov, K. (2015). Selection on MHC class II supertypes in the New Zealand endemic Hochstetter's frog. *BMC Evol. Biol.* **15**, 1–11.

Loiseau, C., Richard, M., Garnier, S., Chastel, O., Jullard, R., Zoorob, R. & Sorci, G. (2009). Diversifying selection on MHC class I in the house sparrow (*Passer domesticus*). *Mol. Ecol.* **18**, 1331–1340.

Magoč, T. & Salzberg, S.L. (2011). FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963.

Martin, D.P., Murrell, B., Golden, M., Khoosal, A. & Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* **1**, 1–5.

Mattoccia, M., Marta, S., Romano, A. & Sbordoni, V. (2011). Phylogeography of an Italian endemic salamander (genus *Salamandrina*): glacial refugia, postglacial expansions, and secondary contact. *Biol. J. Linn. Soc.* **104**, 903–922.

Murrell, B., Wertheim, J.O., Moola, S., Weighill, T., Scheffler, K. & Kosakovsky Pond, S.L. (2012). Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* **8**, e1002764.

Pabijan, M., Wandycz, A., Hofman, S., We, K., Piwczyn, M. & Szymura, J.M. (2013). Complete mitochondrial genomes resolve phylogenetic relationships within Bombina (Anura : Bombinatoridae). *Mol. Phylogenet. Evol.* **69**, 63–74.

Peakall, R. & Smouse, P.E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* **28**, 2537–2539.

Phillips, K.P., Cable, J., Mohammed, R.S., Herdegen-radwan, M., Raubic, J., Przesmyckaa, K.J., van Oosterhoutf, C. & Radwan, J. (2018). Immunogenetic novelty confers a selective advantage in host-pathogen coevolution. *Proc. Natl. Acad. Sci. USA* **115**, 1552–1557.

Piertney, S.B. & Oliver, M.K. (2005). The evolutionary ecology of the major histocompatibility complex. *Heredity* **96**, 7–21.

Pritchard, J.K., Stephens, M. & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959.

Radwan, J., Biedrzycka, A. & Babik, W. (2010). Does reduced MHC diversity decrease viability of vertebrate populations? *Biol. Conserv.* **143**, 537–544.

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

Savage, A.E. & Zamudio, K.R. (2011). MHC genotypes associate with resistance to a frog-killing fungus. *Proc. Natl. Acad. Sci. USA* **108**, 16705–16710.

Schmieder, R. & Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**, 863–864.

Sebastian, A., Herdegen, M., Migalska, M. & Radwan, J. (2016). Amplisas: a web server for multilocus genotyping using next-generation amplicon sequencing data. *Mol. Ecol. Resour.* **16**, 498–510.

Sommer, S. (2005). The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Front. Zool.* **2**, 16.

Spurgin, L.G. & Richardson, D.S. (2010). How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proc. Biol. Sci.* **277**, 979–988.

Stagni, G., Dallolio, R., Fusini, U., Mazzotti, S., Scoccianti, C. & Serra, A. (2004). Declining populations of Apennine yellow-bellied toad *Bombina pachypus* in the northern Apennines (Italy): is *Batrachochytrium dendrobatidis* the main cause? *Ital. J. Zool.* **71**, 151–154.

Stuart, S.N., Chanson, J.S., Cox, N.A., Young, B.E., Rodrigues, A.S.L., Fischman, D.L. & Waller, R.W. (2004). Status and trends of amphibian declines and extinctions worldwide. *Science* **306**, 1783–1786.

Tong, J.C., Bramson, J., Kanduc, D., Chow, S., Sinha, A.A. & Ranganathan, S. (2006). Modeling the bound conformation of Pemphigus Vulgaris-associated peptides to MHC Class II DR and DQ alleles. *Immunome Res.* **2**, 1.

Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591.

Zeisset, I. & Beebee, T.J.C. (2009). Molecular characterization of major histocompatibility complex class II alleles in the common frog, *Rana temporaria*. *Mol. Ecol. Resour.* **9**, 738–745.

Zeisset, I. & Beebee, T.J.C. (2013). *Bufo* MHC class II loci with conserved introns flanking exon 2: cross-species amplification with common primers. *Conserv. Genet. Resour.* **5**, 211–213.

Zeisset, I. & Beebee, T.J.C. (2014). Drift rather than selection dominates MHC class II allelic diversity patterns at the biogeographical range scale in natterjack toads *Bufo calamita*. *PLoS One* **9**, e100176.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article:

Table S1. List of *B. pachypus* and *B. variegata* sampling sites (POP ID) with the location, the approximated geographic coordinates (datum WGS84) and the number of collected samples (SIZE).

Data S2. Modified primers used for PCR amplification of MHC exon 2.

Data S3. Individual barcode and batch membership for 300 *Bombina* samples.

Data S4. MHC genotypes and amplicon summary.

Data S5. STR dataset.

Data S6. Population structure analysis.

Data S7. PAML models result.

Data S8. Genetic diversity.

Data S9. Distribution map of MHC variants protective against chytridiomycosis.