MOLECULAR ECOLOGY

Molecular Ecology (2012) **21**, 887–906 doi: 10.1111/j.1365-294X.2011.05347.x

Interspecific hybridization increases MHC class II diversity in two sister species of newts

KRYSTYNA NADACHOWSKA-BRZYSKA,*,‡ PIOTR ZIELIŃSKI,* JACEK RADWAN* and WIESŁAW BABIK†

*Institute of Environmental Sciences, Jagiellonian University, Gronostajowa 7, 30-387 Kraków, Poland,

†Institute of Systematics and Evolution of Animals, Polish Academy of Sciences, Sławkowska 17, 31-016 Kraków, Poland

Abstract

Our understanding of the evolutionary mechanisms generating variation within the highly polymorphic major histocompatibility complex (MHC) genes remains incomplete. Assessing MHC variation across multiple populations, of recent and ancient divergence, may facilitate understanding of geographical and temporal aspects of variation. Here, we applied 454 sequencing to perform a large-scale, comprehensive analysis of MHC class II in the closely related, hybridizing newts, *Lissotriton vulgaris* (*Lv*) and *Lissotriton montandoni* (*Lm*). Our study revealed an extensive (299 alleles) geographically structured polymorphism. Populations at the southern margin of the *Lv* distribution, inhabited by old and distinct lineages (southern *Lv*), exhibited moderate MHC variation and strong population structure, indicating little gene flow or extensive local adaptation. *Lissotriton vulgaris* in central Europe and the northern Balkans (northern *Lv*) and almost all *Lm* populations had a high MHC variation. A much higher proportion of MHC alleles was shared between *Lm* and northern *Lv* than between *Lm* and southern *Lv*. Strikingly, the average pairwise F_{ST} between northern *Lv* and *Lm* was significantly lower than between northern and southern *Lv* for MHC, but not for microsatellites. Thus, high MHC variation in *Lm* and northern *Lv* may result from gene flow between species. We hypothesize that the interspecific exchange of MHC genes may be facilitated by frequency-dependent selection. A marginally significant correlation between the MHC and microsatellite allelic richness indicates that demographic factors may have contributed to the present-day pattern of MHC variation, but unequivocal signatures of adaptive evolution in MHC class II sequences emphasize the role of selection on a longer timescale.

Keywords: introgression, *Lissotriton montandoni*, *Lissotriton vulgaris*, MHC variation, positive selection, 454 genotyping

Received 26 May 2011; revision received 19 August 2011, 21 September 2011; accepted 1 October 2011

Introduction

Major histocompatibility complex (MHC) is a gene-dense region present in all jawed vertebrates (Kelley *et al*. 2005). Classical MHC class I and II genes play an essential role in the adaptive immune response; they encode proteins that present antigens derived from pathogens to the immune system. Recognition of the MHC–antigen complex by T cells triggers a highly specific immune response against the pathogen (Janeway *et al*. 2004).

Major histocompatibility complex genes are the most polymorphic genes in vertebrates (Garrigan & Hedrick 2003), and understanding the patterns and causes of their variation has been a major goal of research in the fields of ecology and evolution (Sommer 2005; Milinski

Correspondence: Krystyna Nadachowska-Brzyska, Fax: +46 (0)18-471 6310;
E-mail: krystyna.nadachowska-brzyska@ebc.uu.se
‡Present address: Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, SE-75236 Uppsala, Sweden.

2006; Piertney & Oliver 2006). High variation may be manifested both as a large number of alleles per locus (e.g. human HLA-A, B, DRB, http://www.ebi.ac.uk/imgt/hla/stats.html) and as the presence of a number of recently duplicated, and presumably functionally equivalent, loci (Delarbre *et al.* 1992; Bowen *et al.* 2004; Otting *et al.* 2005; Zagalska-Neubauer *et al.* 2010). High MHC polymorphism, especially in parts that encode for functionally important antigen-binding sites (ABS), is believed to be maintained by two major types of balancing selection: negative frequency-dependent selection (Clarke & Kirby 1966; Snell 1968; Borghans *et al.* 2004) and overdominant selection (heterozygote advantage; Doherty & Zinkernagel 1975; Takahata & Nei 1990; Hughes & Nei 1992). Pressure from pathogens is considered to be the major selective factor maintaining MHC variation in natural populations (Wegner *et al.* 2003; Sommer 2005; Bonneaud *et al.* 2006; Milinski 2006; Schwensow *et al.* 2007, 2010; Deter *et al.* 2008; Kloch *et al.* 2010), but disassortative mate choice can also play a role (Penn 2002).

Natural selection may act on MHC genes over short time periods, but its signatures may be detected in patterns of sequence variation over thousands of generations (Garrigan & Hedrick 2003). On the other hand, MHC variation can be shaped predominantly by demographic processes, especially in small, bottlenecked populations (reviewed in Radwan *et al.* 2010). Thus, both selection and drift appear to affect patterns of MHC variation, and each of the mechanisms may dominate at different timescales (Bernatchez & Landry 2003; Borghans *et al.* 2004; Babik *et al.* 2005b; Piertney & Oliver 2006; Radwan *et al.* 2010). The relative role of natural selection vs. genetic drift in MHC evolution is not yet fully understood, and studies of structured populations on a large geographical scale may provide insight into mechanisms that maintain MHC variation.

Population structure in MHC may result not only from drift but also from divergent selection pressures acting between populations (Ekblom *et al.* 2007; Loiseau *et al.* 2009; Eizaguirre *et al.* 2011). Indeed, there is evidence that selection on MHC alleles differs between populations, and gene flow between populations has been hypothesized to contribute to the maintenance of high MHC polymorphism (Kloch *et al.* 2010). Gene flow may also occur among species (Mallet 2005), and it therefore seems possible that introgression may be another mechanism that increases MHC variation. It follows that frequency-dependent selection acting on MHC may facilitate gene flow, as introgressed alleles will initially be rare, and therefore favoured by selection. Such alleles may rapidly become established in the recipient species (Barton 2001).

Major histocompatibility complex class II architecture and variation have been investigated less in amphibians than in other vertebrate groups (Kelley *et al.* 2005; Piertney & Oliver 2006). Anura data on MHC variation exist for several species (Hauswaldt *et al.* 2007; May & Beebee 2009; Zeisset & Beebee 2009; Kiemnec-Tyburczy *et al.* 2010), and the most extensive information is available for *Xenopus* and *Silurana* (Bos & Waldman 2006; Ohta *et al.* 2006). Among urodele amphibians, data on MHC class II variation are only available for two closely related *Ambystoma* species (*Ambystoma mexicanum* and *Ambystoma tigrinum*; Tournefier *et al.* 1998; Sammut *et al.* 1999; Laurens *et al.* 2001) and two newts, the great crested (*Triturus cristatus*, Babik *et al.* 2009a) and the alpine newt (*Mesotriton alpestris*, Babik *et al.* 2008). Existing data reveal substantial differences in MHC organization and patterns of variation between species, and both selection and drift appear to have left strong signatures on the genome. The interest in amphibian MHC class II is also motivated by its potential link to the conservation of amphibian species. Many amphibian species are highly endangered and are declining worldwide because of several interrelated factors, of which the emergence of infectious pathogens is one of the most severe (Stuart *et al.* 2004; Beebee & Griffiths 2005; Pounds *et al.* 2006; Lips *et al.* 2008).

Here, we investigate MHC variation and population structure in two sister species of newts: the smooth (*Lissotriton vulgaris*) and Carpathian newt (*Lissotriton montandoni*). A comparative study of MHC variation in these species is interesting for several reasons. First, the species differ greatly in the size and characteristics of their distribution ranges. *Lissotriton vulgaris* inhabits large expanses of Eurasia and exhibits an ancient genetic structure, as reflected by multiple subspecies, and multiple glacial refugia have been identified for this species (Babik *et al.* 2005a; Nadachowska & Babik 2009). In contrast, *L. montandoni* is confined to the Carpathian Mountains in central Europe where its refugia were located. Thus, a higher overall MHC variation and its stronger geographical structuring are expected in *L. vulgaris* if MHC variation has been shaped mainly by historical processes and/or local adaptations. Second, the species show a number of behavioural and ecological differences (Babik *et al.* 2003, 2005a and references therein), which may be reflected in patterns of MHC variation. Third, despite these differences, the species hybridize wherever their parapatric ranges meet (Babik *et al.* 2003) and analyses of mtDNA and several nuclear sequence markers revealed substantial, mainly unidirectional, gene flow from *L. vulgaris* to *L. montandoni* (Babik *et al.* 2005a). It is thus interesting to determine the extent to which interspecific gene flow affected MHC class II.

The specific aims of this study were as follows: (i) to assess the number of loci, expression pattern and levels of MHC class II variation; (ii) to test for signatures of historical positive natural selection and to evaluate its role in shaping patterns of MHC II variation; (iii) to characterize population genetic structure in MHC II genes, differentiation between *L. montandoni* and *L. vulgaris* and among *L. vulgaris* subspecies; (iv) to investigate whether interspecific gene flow contributes to MHC variation in these species and (v) to test for a correlation between MHC variation and putatively neutral microsatellite variation.

Materials and methods

Samples

We sampled 26 populations representing all *L. vulgaris* subspecies (13 populations of *L. v. vulgaris*, 5—*L. v. greacus*, 2—*L. v. ampelensis*, 2—*L. v. lanzi*, 1—*L. v. meredionalis*, 1—*L. v. kosswigi*, 2—*L. v. schmidtlerorum)* and nine populations of *L. montandoni* distributed over the entire species range (Fig. 1; Table S1, Supporting information). Our sampling encompassed both putative glacial refugia and postglacial expansion areas and concentrated in the regions where we expected to find the highest diversity, based on the earlier phylogeographic study (Babik *et al.* 2005a).

Development of primers, amplification and 454 sequencing

The second exon of MHC class II genes encodes the ABS sites, is the most polymorphic part of the gene and thus may serve as a proxy for the overall variation. Initially, we obtained exon 2 sequences from *c.* 100 newts from multiple populations using primers TrMHCII1F-TrMHCII4R (Babik *et al.* 2008). Additional vectorette PCR experiments performed according to the protocol of Babik *et al.* (2008) and, with primers reported there, allowed better characterization of sequence diversity at both ends of the exon. Primers used for genotyping were designed on the basis of information gathered in these preliminary steps (Table 1; Fig. S1, Supporting information). A 204-bp fragment (excluding primers) of the MHC II 2nd exon was amplified for each individual in two PCRs, each employing a different primer pair: (i) Fusion_MHCvul_F/Fusion_MHCvul_R1 and (ii) Fus-ion_MHCvul_F/Fusion_MHCvul_R2; these two primer combinations are referred to as R1 and R2 later on. The fusion primers were necessary for sequencing amplicons using 454 Titanium technology. The forward

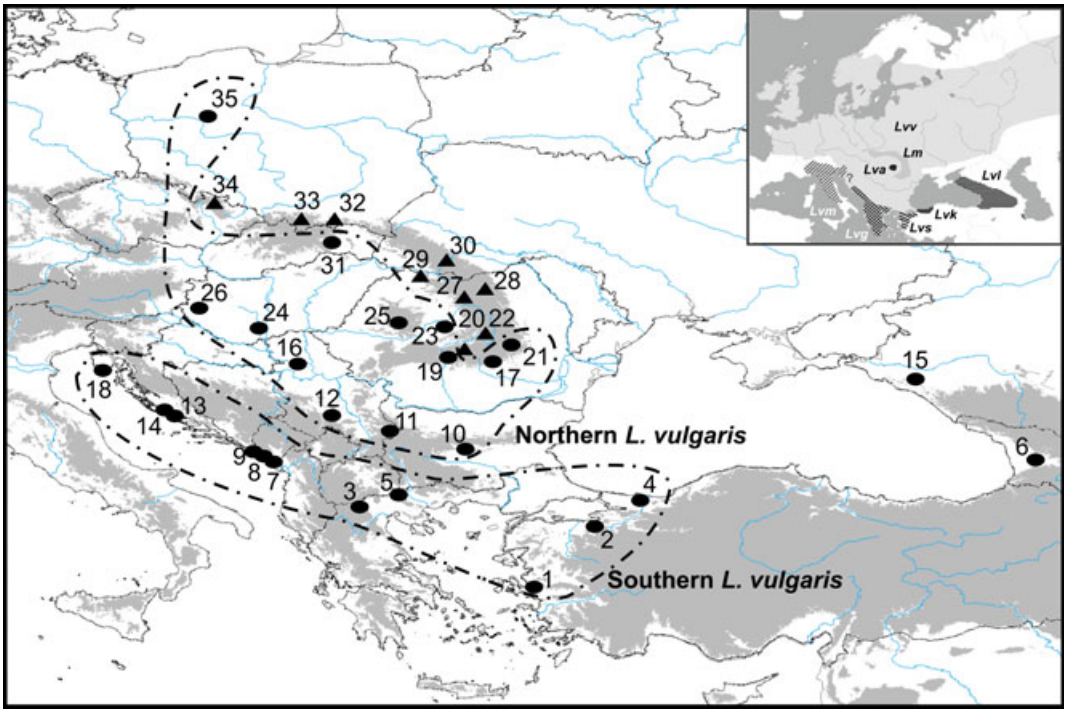

Fig. 1 Sampling localities of *Lissotriton montandoni* (triangles) and *Lissotriton vulgaris* (circles). Inset shows the distribution of species and *L. vulgaris* subspecies (*L. montandoni*—*Lm*, *L. v. vulgaris*—*Lvv*, *L. v. graecus*—*Lvg*, *L. v. ampelensis*—*Lva*, *L. v. lanzi*—*Lvl*, *L. v. meridionalis*—*Lvm*, *L. v. kosswigi*—*Lvk* and *L. v. schmidtlerorum*—*Lvs*). The dashed curves represent the northern *L. vulgaris* and southern *L. vulgaris* groups (see main text for further explanation). The names of populations corresponding to numbers are given in Tables 5 and S1, Supporting information.

Table 1 454 Primers used for amplification of MHC class II 2nd exon in *Lissotriton vulgaris* and *Lissotriton montandoni*

Name	Sequence 5′ → 3′
Specific primers	
MHCvul_F	GAGTGYCWSTWSSTBAACG
MHCvul_R1	CTCACRCYTCCGSTGCTCCATG
MHCvul_R2	CTCACGCCTCCGKTKGTACAGG
454 adapters	
Adapter A	CGTATCGCCTCCCTCGCGCCATCAG
Adapter B	CTATGCGCCTTGCCAGCCCGCTCAG

MHC, major histocompatibility complex.

fusion primer consisted of 454 Titanium amplicon adaptor A, followed by a 6-bp sequence tag, and MHCvul_F sequence. The reverse fusion primers consisted of 454 Titanium amplicon adaptor B and MHC II-specific primer (MHCvul_R1 or MHCvul_R2, Fig. S1, Supporting information). The 6-bp tags were used to assign 454 sequencing reads to individuals. We used 96 tags, differing from each other in at least three positions, which minimizes misassignment of alleles, i.e. even if the tag in a sequencing read contains an error, this error is unlikely to convert it to the sequence of another tag used in the experiment; thus, the read will simply be left unassigned. The R1 primer pair amplified the majority of alleles and yielded PCR product for all individuals. The R2 pair amplified just a small subset of alleles, and PCRs were successful for 38% of analysed individuals, reflecting the presence of R2 amplifiable alleles. Each PCR was performed in 15 μL and contained 7.5 μL of HotStar PCR Master Mix (Qiagen) and 3 μM (R1) or 1.5 μM (R2) of each primer. The following PCR protocols were used: 95 °C/15 min, 42× (94 °C/30 s, 52 °C/30 s, 72 °C/90 s), 72 °C/10 min for R1 and 95 °C/15 min, 40× (94 °C/30 s, 55 °C/30 s, 72 °C/90 s), 72 °C/10 min for R2.

The number of PCR cycles was high because of the low amplification efficiency of primers. Normally, careful primer design and optimization of PCR conditions ensure high efficiency of amplification. However, in cases of high sequence variation, such as observed in this study, this may not be possible. One can follow two alternative routes to study such extraordinarily polymorphic systems by PCR. First, which we have followed, is to design primers at the most conservative fragments of the sequence even though the primers may not be optimal for PCR efficiency. The advantages of this approach are a low number of primer sets required for amplification of all target variants (two sets in our case) and low probability of missing true alleles owing to mutations in primer-binding sites. On the other hand, such primers, if not optimal for PCR, require a high number of PCR cycles, resulting in higher probability of PCR artefacts (Kanagawa 2003). Alternatively, in the second approach, one may try to design a set of high-quality primer pairs in a more polymorphic region. In our opinion, this approach has serious disadvantages: (i) expected unequal amplification by various primer pairs precludes the use of the multiplex PCR, thus increasing the overall cost and logistic burden in setting up multiple PCRs and their downstream processing, (ii) variation in the primer-binding sites not recognized at the stage of primer design increases the risk of missing some alleles.

We pooled approximately equimolar quantities of PCR products (assessed by visual examination of band intensities on agarose gels) obtained with up to 96 different tags. The pools were purified using the MinElute PCR Purification Kit (Qiagen). Purified pools from two PCR (R1 and R2) containing the same individuals were combined according to the required coverage and sequenced as a part of two 454 Titanium runs (each run contained also MHC amplicons from other species) at the Functional Genomics Center Uni/ETH in Zurich. Each run was divided into eight sections, which allowed the use of the same tagged primers multiple times. Seventy-seven individuals were amplified twice with R1, and both amplicons were sequenced. The replicates were used to assess the genotyping error. To extract, analyse and visualize sequencing data for each amplicon, we used the jMHC software (Stuglik *et al.* 2011) and the output from jMHC was further analysed in Excel and Bioedit (Hall 1999).

Analysis of expression

The preliminary analyses of expression patterns were conducted on five *L. vulgaris* larvae preserved in the RNAlater reagent (Sigma). Gills were homogenized, and RNA was extracted using the RNeasy Mini kit (Qiagen) according to the manufacturer's protocol including DNase digestion step. Messenger RNA (mRNA) was reverse transcribed using Omniscript Reverse Transcriptase kit (Qiagen) and Oligo(dT)$_{12–18}$ primer (Invitrogen). MHC II exon 2 was amplified from cDNA and 454 sequenced (together with MHC alleles amplified from genomic DNA of these individuals) as described above.

Distinguishing true alleles from artefacts and MHC genotyping

Preliminary analysis of R1 amplicons, based on a careful examination of the allele sequences and read counts for a subset of individuals, suggested the range of 2–6 alleles per individual. Assuming that the maximum number of alleles per individual was 6, we then calcu-

lated the minimum coverage for which the probability of obtaining at least two copies of each allele would exceed 0.95 using the software of Galan *et al.* (2010). The 0.95 threshold was reached at the coverage of 39; thus, we accepted 40 reads as the coverage sufficient for genotyping of R1 amplicons.

Distinguishing true alleles from artefacts generated at various stages of the analysis, and the determination of sufficient (i.e. enabling identification of all alleles present in an amplicon) coverage are required for reliable genotyping. PCR may produce two kinds of artefacts: point mutations (owing to DNA polymerase errors) and chimeras—in vitro recombinants. Four hundred and fifty-four sequencing produces additional errors: substitutions and small indels (especially in homopolymer runs, Babik *et al.* 2009b; Galan *et al.* 2010). Point mutations introduced by PCR and/or sequencing reactions are usually easy to detect: the artefacts differ just slightly (e.g. one substitution) from the true alleles, and substitutions are expected to occur approximately randomly across the sequence, keeping probability of multiple identical substitution errors low (Galan *et al.* 2010). Sequences with small indels are much more common (Margulies *et al.* 2005; Huse *et al.* 2007) and are nonrandomly distributed along the sequence. Therefore, it may be challenging to distinguish them from true alleles in noncoding sequences. In the case of coding genes, such sequences may be simply discarded because even if real, they can be considered loss-of-function frameshifts. The most difficult artefacts to address are PCR chimeras produced by in vitro recombination between true alleles, as these are common artefacts of all existing methods of MHC genotyping (Longeri *et al.* 2002; Lenz & Becker 2008; Galan *et al.* 2010) and may occur in a relatively high number of reads. However, because chimeras always co-occur with parental alleles, they should be easily identified by examining all sequence variants present in an amplicon (Zagalska-Neubauer *et al.* 2010).

Taking into account these considerations, we followed procedures developed by Zagalska-Neubauer *et al.* (2010) to filter out artefacts. In the first step, we excluded all variants containing indels (mostly 1 or 2 bp) causing frameshifts or present in just one copy at the level of the full data set. In the second step, considering all R1 amplicons with sufficient coverage, we calculated the maximum per-amplicon frequency (MPAF) of each sequence variant and sorted the variants according to their MPAF. Then, for each variant with a MPAF of at least 3%, we checked whether it could be explained as an artefact (derived by 1-bp substitution or recombination from other, more abundant, variants present in the amplicon). Nearly all variants with MPAF below 5% could be explained in this way and were considered artefacts, whereas all variants above 9% could not be explained in this way and were therefore considered to be putative true alleles. Both putative true alleles and artefacts were present in the MPAF range of 5–9% and were considered on a per-individual basis. Thus, during genotyping, for each amplicon meeting the minimum coverage threshold, we recorded all putative true alleles present in at least 9% of the amplicon's reads. Of variants occurring in 5–9% of reads, which had sequences identical to true alleles, we excluded these, which could have been derived as recombinants or single base pair substitutions from other, more abundant alleles detected in the amplicon. This step was necessary because some artefacts may have sequences identical to putative true alleles and are thus indistinguishable by sequence alone.

The coverage for R2 amplicons was lower than for R1 amplicons (Table 2) because we expected only one or two R2 alleles per individual. Indeed, the preliminary screening of the 454 sequencing results suggested the presence of up to two R2 alleles per amplicon. The minimum coverage for which the probability of obtaining at least two copies of each allele would exceed 0.95 was 9; therefore, nine was used as the genotyping threshold. Sequence variants were considered to be true alleles if they were present in at least two copies in two independent PCR and could have not been explained as artefacts derived from more abundant alleles present in the amplicon. Singletons were considered true alleles if present in at least three copies.

The standard in MHC studies is the two-PCR criterion; to be confirmed, an allele must be obtained from two independent PCRs to guard against PCR artefacts (Babik *et al.* 2009b). The current study, however, included multiple populations expected to differ in their MHC allelic composition, and therefore, many true alleles were probably observed only once. Thus, the strict application of the two-PCR criterion would underestimate diversity. In contrast, the procedure we applied to distinguish true alleles from artefacts accounted for the known sources of artefacts; the problem of mosaic sequences, created by bacterial repair of heteroduplexes during traditional cloning (Longeri *et al.* 2002), does not apply to 454 sequencing. We are therefore convinced that alleles detected only once (in a single amplicon, singletons) were, in most

Table 2 Statistics of 454 Titanium runs

Pools	R1	R2
Total number of reads	129 781	13 046
Variants with only one read	25 659	2413
Erroneous tags	809	103
Average per amplicon coverage	88.31	23.66

cases, true alleles. Nevertheless, to evaluate the effect of singletons, we created two data sets, one including and one excluding singletons (subsequently called data set plusS and data set minusS, respectively), and performed all analyses on both data sets. The results were qualitatively identical and quantitatively very similar; therefore, we only present results for data set plusS in the main text (results for data set minusS are available in Supporting information).

To assess the rate of genotyping error, we obtained replicate (amplified and sequenced independently in separate 454 pools) genotypes of R1 alleles for 77 individuals. Based on consensus genotypes (comparing genotypes of the same individuals obtained from different PCRs), we calculated the fraction of alleles that were not identified in one replicate (the actual genotyping error rate). To compare this actual error rate to the theoretical expectation, we also computed, from the multinomial distribution, the probability of not detecting a particular allele (obtaining 0 reads) when genotyping individuals possessing the specified total number of alleles.

Microsatellite analysis

Six microsatellite loci were used to estimate the amount of neutral genetic variation in populations. Four loci were previously described (*Tv3CA9*, Johanet *et al.* 2009; *Lm_521*, *Lm_749*, *Lm_870*, Nadachowska *et al.* 2010), and two loci were mined from 454 transcriptome sequences (*Lm_346, Lm_AHNC3*). Five loci was amplified in two multiplexes (MPX1: *Lm_521* (forward primer FAM labelled), *Lm_AHNC3* (NED); MPX2: *Tv3CA9* (HEX), *Lm_749* (PET), *Lm_870* (NED)), and locus *Lm_346* was amplified separately. Multiplex reactions were performed in 8 μL and contained 4 μL of Multiplex PCR Master Mix (Qiagen), 0.2–0.4 μM of each primer and 30–100 ng of genomic DNA. The following cycling scheme was used: 95 °C/15 min, followed by 30 cycles of 30 s/94 °C, 90 s/54 °C (MPX1) or 56 °C (MPX2) and 90 s/72 °C and a final extension step of 10 min at 72 °C. Locus *Lm_346* was amplified in 5 μL PCR containing 2.5 μL of Multiplex PCR Master Mix and 0.2 μM of each primer. The following cycling scheme was used: 95 °C/15 min, followed by 35 cycles of 30 s/94 °C, 90 s/56 °C and 90 s/72 °C and a final extension of 10 min at 72 °C. The PCR products were electrophoresed on an ABI 3130xl with GeneScan 500 LIZ size standard, and GeneMapper was used for genotyping.

Statistical, population genetics and phylogenetic analysis

Allelic richness, i.e. the per-population number of MHC class II alleles corrected for sample size, was estimated for all population samples with $N \geq 11$ (30 populations in total) through randomization tests in PopTools (Hood 2010). We chose this method because assigning alleles to loci was not possible, so no standard population genetic packages could be used for computations.

DT-ModSel (Minin *et al.* 2003) was used to select the model of sequence evolution. Relationships among all *L. vulgaris* and *L. montandoni* MHC class II alleles were reconstructed using the neighbour-joining method under the Tamura–Nei model of sequence evolution. We used *A. tigrinum* sequence (GenBank DQ125478) as out-group, and the robustness of the tree was assessed with 1000 bootstrap replicates. To place the *L. vulgaris/L. montandoni* MHC variation in the broader context of other newts, we chose 22 *L. vulgaris/L. montandoni* alleles representing the entire spectrum of sequence diversity and reconstructed relationships among them and alleles from other newt species: *T. cristatus* (four alleles from main clades, including one pseudogene sequence, Babik *et al.* 2009a) and *M. alpestris* (four alleles from main clades, including one pseudogene sequence, Babik *et al.* 2008). This tree was reconstructed using *A. tigrinum* allele as an out-group under the Bayesian approach with MrBayes 3.1 (Ronquist & Huelsenbeck 2003) under the general time reversible model of sequence evolution with rate variation (GTR + Γ). Priors were set to default values. We ran two independent Metropolis-coupled Markov chain Monte Carlo analyses with three heated chains ($T = 0.20$), sampled every 10^3 generations and performed the simulation for 2×10^6 generations, which was sufficient for achieving convergence as evidenced by the average standard deviation of split frequencies. Log-likelihood values were plotted against generation time, which ensured that no trees were retained prior to the run reaching stationarity and the first 500 trees were discarded as burn-in. To calculate the posterior probability (PP) of each bipartition, the majority-rule consensus tree was computed from the 3000 sampled trees.

We tested for signatures of positive selection in MHC class II sequences with one-tailed *Z*-test in MEGA 4 (Tamura *et al.* 2007) and by comparing the likelihoods of codon-based models of sequence evolution in PAML 4 (Yang 2007). The *Z*-test was performed for all codons, ABS codons and non-ABS codons. The putative locations of ABS codons were taken from (Tong *et al.* 2006). In the codon-based approached, we evaluated three models: M0, one dN/dS ratio; M7, nearly neutral model ($dN/dS \leq 1$) with beta distribution approximating dN/dS variation; M8, positive selection, model with a proportion of sites evolving with $dN/dS > 1$. The best fitting model was chosen by Akaike information criterion (AIC), and codons under positive natural selection

were identified through Bayes empirical Bayes procedure (BEB, Yang *et al.* 2005).

We checked for signatures of recombination using two methods GeneConv (Padidam *et al.* 1999) and MaxChi2 (Smith 1992), both implemented in RDP 3 (Martin *et al.* 2010). The methods were shown to perform very well in an assessment of 14 recombination detection methods (Posada 2002). Because of high number of putative alleles in both of our data sets, recombination tests using all sequences would be extremely conservative owing to repetitive statistical testing and correction for multiple tests. Therefore, we tested smaller data sets: the representative *L. vulgaris* and *L. montandoni* sequences used to reconstruct phylogenetic relationships of MHC class II among four newts species and 20 subsets of 25 randomly chosen sequences (10 of them were sampled from plusS and 10 from minusS data set). We applied Bonferroni correction for multiple tests.

To assess genetic structure of *L. vulgaris* and *L. montandoni* populations in MHC class II, we applied several approaches using binary-encoded data, with each allele considered a separate dominant locus, and its presence/absence coded 1/0 (all analyses were performed on both data sets, plusS and minusS). First, analysis of molecular variance (AMOVA) was used to partition MHC variation into hierarchical levels. Computations were performed in Arlequin (Excoffier *et al.* 2005). Significance of AMOVA components were tested with 10 000 permutations. Second, also with Arlequin, we calculated F_{ST} between all pairs of populations and tested their statistical significance with 10 000 permutations; multidimensional scaling (MDS) was used for the visualization of the F_{ST} matrix. Third, to determine the most likely number of MHC genetic clusters present in the studied newts, we used an individual-based Bayesian clustering method implemented in STRUCTURE 2.3.3 (Pritchard *et al.* 2000; Falush *et al.* 2007; Hubisz *et al.* 2009). Structure uses multilocus genotype data to cluster individuals, usually not taking into account the population of origin, into K genetic clusters, each characterized by a set of allele frequencies. The individuals are probabilistically assigned to clusters in such a way that the deviations from Hardy–Weinberg and linkage equilibrium are minimized. The method has been modified to deal with dominant markers like AFLP (Falush *et al.* 2007). We ran the program under a model with correlated allele frequency and allowed admixture. We examined K values from 1 to 13. For each K value, we performed five runs of 4×10^5 burn-in steps followed by 4×10^6 postburn MCMC iterations. To infer most likely number of clusters in the data, we calculated ΔK, a measure of second-order rate of change in the likelihood of K (Evanno *et al.* 2005).

For microsatellites, pairwise F_{ST} between populations were computed from allele frequencies adjusted for the presence of null alleles using the ENA correction in FreeNA (Chapuis & Estoup 2007), and MDS was used for the visualization of the F_{ST} matrix. Microsatellite allelic richness for each locus and population was calculated in ADZE (Szpiech *et al.* 2008). Because the presence of null alleles was inferred for multiple loci and populations, we applied the following procedure to minimize the effect of null alleles on the estimates of microsatellite allelic richness. We estimated frequencies of null alleles (Pnull) with the EM algorithm in FreeNA (Chapuis & Estoup 2007) and divided allelic richness estimates by 1—Pnull. Null allele-corrected estimates of allelic richness are reported throughout the article. The procedure assumes that the distribution of allele frequencies for alleles collectively labelled 'null' is the same as for observed alleles, which seems reasonable (Chapuis & Estoup 2007). The correlation between MHC class II allelic richness and microsatellite variation (mean allelic richness) was estimated for three data sets, which was necessary because some microsatellite loci did not amplify in all populations. The first data set included all six loci and 24 populations (further on called six loci data set), the second excluded locus Lm_521 and consisted of 27 populations (Seli [3], Kentriko [5] and Benkovac [13] excluded, minus Lm_521 data set) and the third one excluded locus Lm_AHNC3 and also comprised 27 populations (Karagol [1], Subasi [2] and Salakovac [18] excluded, minus Lm_AHNC3 data set). The six loci data set was analysed in Structure under the admixture model with correlated allele frequencies. For direct comparison of clustering based on MHC and microsatellites, we also ran Structure on trimmed MHC data including only the populations present in microsatellite data set (24 populations).

Results

MHC II genotyping

Statistics regarding the number of reads and individual coverages are provided in Table 2. The coverage threshold of 40 reads suitable for genotyping of R1 amplicons was achieved for 526 individuals. The average coverage for genotyped individuals was 117 (SD = 88).

Replicate genotypes, for which both replicates achieved the coverage threshold, were obtained for 30 individuals. Replicated individuals had a total of 87 alleles (sum of individual counts), of which five were detected in only one of the individual's replicates, setting the genotyping error at 5.7%. Discrepancies were found for individuals with 3 or 4 alleles; thus, from the multinomial distribution, we estimated probabilities of

obtaining 0 reads of a single allele for 3- and 4-allele genotypes. Both probabilities were $<10^{-4}$. Therefore, the substantial difference between the probability of not detecting alleles during genotyping and the actual genotyping error may indicate unequal amplification of some alleles and/or the duplication of some loci.

MHC II diversity

Two data sets, one including singletons (alleles detected in only one individual) and the other excluding them (subsequently referred to as plusS and minusS, respectively), were analysed. Because the results of all analyses were concordant for both data sets, we present here only the plusS data set; results for the minusS can be found in Supporting information (Figs S2–S5, Tables S2–S6, Supporting information). We found 299 MHC II exon 2 alleles in the plusS data set, 237 occurred in *L. vulgaris* and 101 in *L. montandoni*. The number of alleles per individual ranged from one to six (median = 3), indicating the presence of a minimum of three MHC loci and suggesting variation among haplotypes in the number of loci (Fig. S5, Supporting information). There were no alleles present in all or most individuals, which would be expected if monomorphic or only slightly polymorphic loci were present. There were single-codon deletions in 28 alleles (Appendix S2, Supporting information). To test whether the obtained number of alleles per individual was affected by sequencing depth, we checked the correlation between number of alleles and amplicon coverage. The correlation was not significant ($r^2 = 0.0006$, $P = 0.571$), indicating coverage sufficient for reliable genotyping.

The majority of alleles were amplified by Fusion_MHCvul_F/Fusion_MHCvul_R1 primer pair (268, R1 alleles). Only 28 alleles (9% of all, R2 alleles) were amplified exclusively with Fusion_MHCvul_F—Fusion_MHCvul_R2 primer combination. Fifty R1 alleles were amplified by both sets of primers, constituting additional evidence for their classification as true alleles. R2 alleles were present in 38% of individuals; most of them (80%) had only a single R2 allele, while others had two alleles. Of 38 individuals with two R2 alleles, 12 were from the Benkovac [13] population. We found 198 unique alleles in the minusS data set: 161 in *L. vulgaris* and 76 in *L. montandoni*. Despite the small number of MHC II loci, assignment of alleles to loci was not possible, as phylogenetic analysis (see below) did not reveal well-supported clades (Fig. 2).

Expression pattern

For five individuals, we amplified and sequenced MHC alleles from both cDNA and genomic DNA. Three had coverage sufficient for genotyping (Table S7, Supporting information). We combined the data with information on preliminary expression pattern obtained from previous experiments conducted on *L. vulgaris* and *L. montandoni* cDNA (W. Babik, unpublished) and identified 10 expressed (*Limv-DNB*004*, *Limv-DNB*007*, *Limv-DNB*008*, *Limv-DNB*046*, *Limv-DNB*056*, *Limv-DNB*073*, *Limv-DNB*124*, *Limv-DNB*132*, *Limv-DNB*140* and *Limv-DNB*197*) and three nonexpressed alleles (*Limv-DNB*001*, *Limv-DNB*019* and *Limv-DNB*068*). Expressed alleles did not form a distinct group on a phylogenetic tree; in some cases, the expressed and nonexpressed alleles had very similar sequences (Fig. 2). Thus, distinguishing functional from nonfunctional MHC II alleles was not possible with the limited cDNA information available.

Signatures of natural selection and recombination

The average rates of nonsynonymous substitutions per nonsynonymous site and synonymous substitutions per synonymous site are provided in Table 3. The overall nucleotide divergence in the ABS was 5.6 times higher than in non-ABS sites. The dN/dS ratio was significantly elevated at ABS (one-tailed test, $Z = 2.232$, $P = 0.014$), indicating the role of historical positive natural selection in shaping MHC sequence variation. Interestingly, the dN value was very high (0.7), suggesting a saturation of nonsynonymous sites. In codon-based tests of natural selection, the M8 model of codon evolution, which assumes that a fraction of sites is under positive selection, produced a much better fit to the data than remaining models (Table 4). The Bayes empirical Bayes procedure identified two ABS codons evolving under positive selection: 8 and 66 (both *PP*s > 99%).

Recombination tests detected from 0 to 3 recombination events in all analysed data subsets. At least one recombination event was observed in 20 of 42 tests, suggesting that recombination occurs frequently in the 2nd exon of MHC II in the studied species.

Relationships between newt MHC class II sequences

Phylogenetic analysis of representative MHC II alleles from four newt species (*L. vulgaris*, *L. montandoni*, *M. alpestris* and *T. cristatus*) showed that some *L. montandoni* and *L. vulgaris* alleles are more similar to *M. alpestris* or *T. cristatus* alleles than to alleles identified in their own species; this is a signature of trans-species polymorphism commonly observed in MHC genes (Fig. 3). Two alleles were highly similar to *T. cristatus* (*Limv-DNB*040*) and *M. alpestris* (*Limv-DNB*134*) pseudogenes.

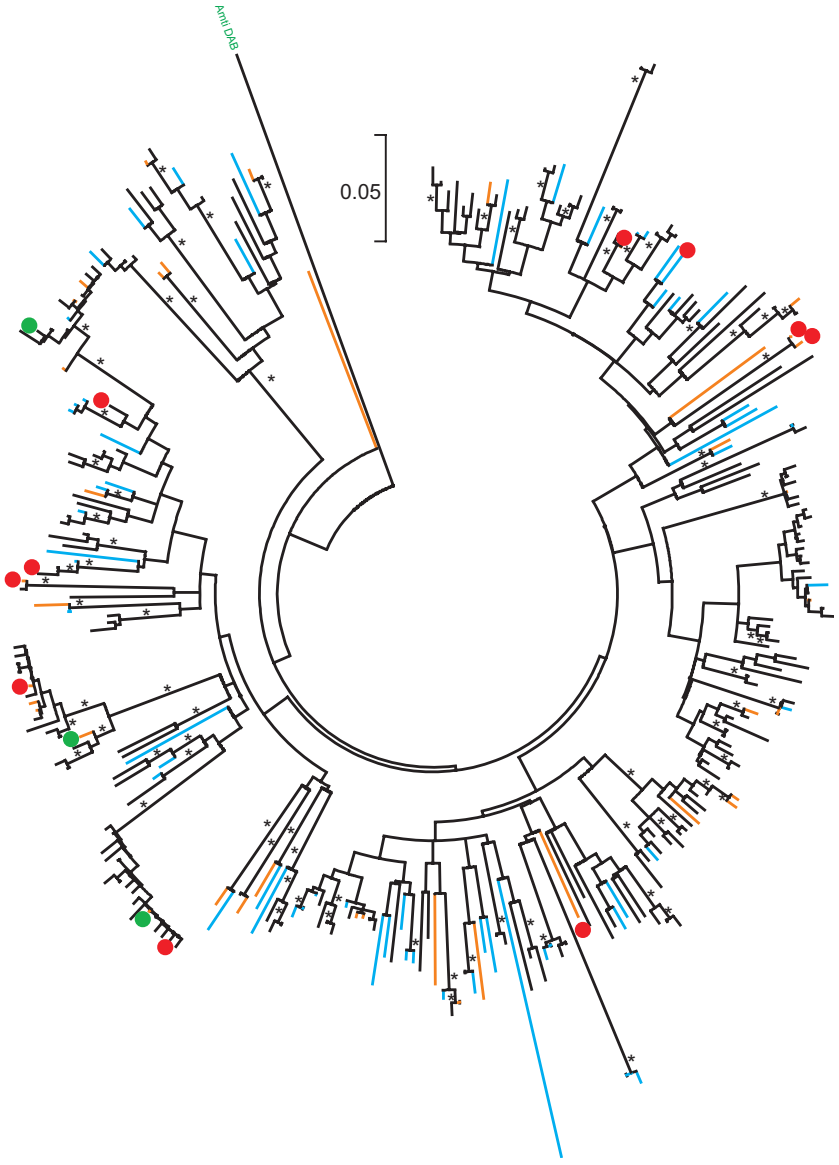

Fig. 2 The neighbour-joining tree showing relationships of *Lissotriton vulgaris* and *Lissotriton montandoni* major histocompatibility complex class II alleles. Alleles of *L. montandoni* are shown in blue and shared alleles in orange. Red dots—alleles with confirmed expression and green dots—non-expressed alleles. Bootstrap values ≥70% are indicated by asterisks. The tree was rooted with an *Ambystoma tigrinum* sequence.

Table 3 The average rates of nonsynonymous substitutions per nonsynonymous site (dN) and synonymous substitutions per synonymous sites (dS) with standard errors obtained through 1000 bootstrap replicates in parentheses, and the results of the *Z*-test of neutrality

Sites	dN (SE)	dS (SE)	*Z*	*P*
ABS	0.70 (0.119)	0.34 (0.100)	2.232	0.014*
Non-ABS	0.13 (0.023)	0.15 (0.003)	−0.560	1.000
All	0.19 (0.029)	0.16 (0.034)	0.573	0.284

ABS, antigen-binding sites.
*$P < 0.05$.

Table 4 Evaluation of the goodness of fit for different models of codon evolution and estimated parameter values

Model	ln L	Δ AIC	Parameters
M0	−8882.8	1395.9	ω = 0.307
M7	−8191.2	14.6	
M8	−8181.9	Best	p_0 = 0.952, p_1 = 0.048, ω_1 = 1.549

ω, dN/dS; nearly neutral with beta, for all sites ω ≤ 1 and the beta distribution approximates ω variation; positive selection, a proportion of sites evolves with ω > 1; p_0, proportion of sites with ω ≤ 1; p_1, proportion of positively selected sites (ω > 1); ω_1, estimated value of ω for sites under positive selection.

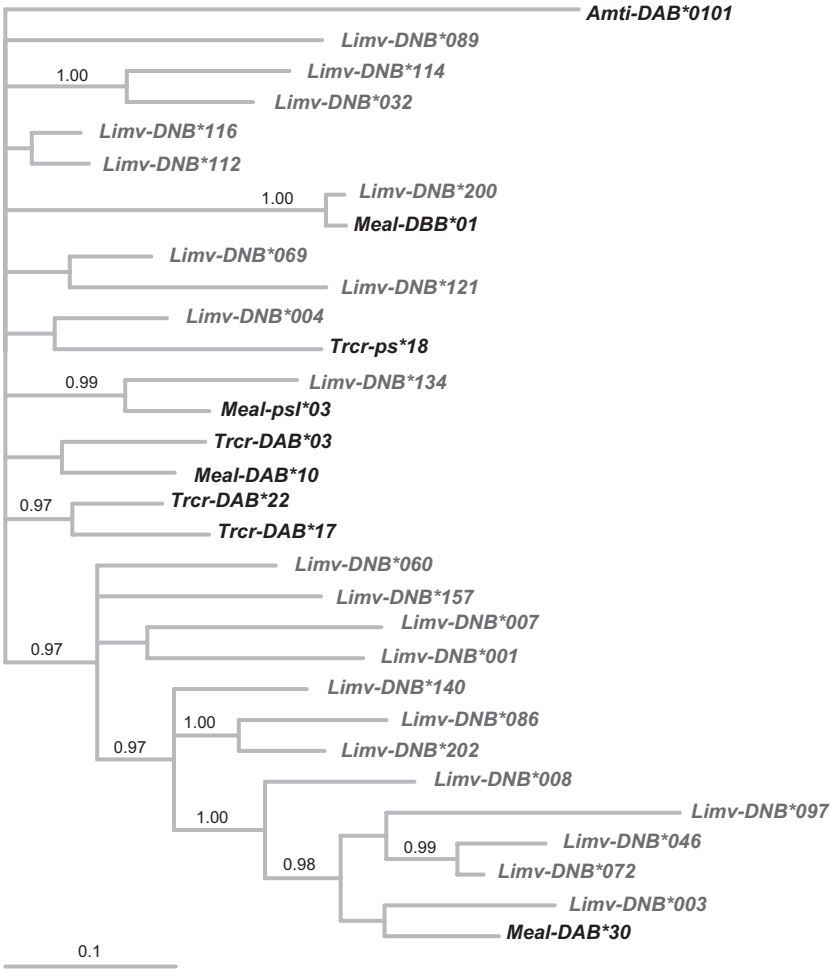

Fig. 3 The Bayesian phylogenetic tree of 22 representative alleles of *Lissotriton montandoni* and *Lissotriton vulgaris* (grey) and major histocompatibility complex class II alleles from other newt species (black): *Triturus cristatus* (four alleles from main clades, including one pseudogene sequence) and *Mesotriton alpestris* (four alleles from main clades, including one pseudogene sequence). The tree was rooted using *Ambystoma tigrinum* sequence. Posterior probabilities ≥95% are shown above branches.

MHC II variation within populations

The number of alleles, allelic richness and the number of private alleles for each population are provided in Table 5. The correlation between allelic richness and the average per individual coverage in populations was not significant ($r^2 = 0.0026$, $P = 0.788$). Thus, the estimates of allelic richness were not affected by the differences in coverage among populations. Overall, we found high MHC variation within populations, but differences between populations were also substantial. Considering both species together, allelic richness was highest in *L. vulgaris* population Sighişoara [23] (25.1) and lowest in *L. vulgaris* population Bátászek [24] (7.1, Table 5, Fig. 4). *Lissotriton montandoni* showed higher allelic richness than *L. vulgaris* (Mann–Whitney test, excluding isolated *L. montandoni* Jeseniki [34] population—$P < 0.05$, including Jeseniki—$P = 0.095$). However, both species showed a similar allelic richness in Romania and adjacent Ukraine, where several populations of each species were sampled: 19.7 for *L. montandoni* and 20.1 for *L. vulgaris* ($P = 0.323$, Mann–Whitney test). Among *L. montandoni* populations, Dzembronia [30] from the Ukrainian Carpathians had the highest variation ($R = 23.0$, Table 5, Fig. 4). Other populations from the Carpathian Mountains also showed high allelic richness. The isolated population from the Jeseniki Mountains [34] in the eastern Sudetes showed the lowest diversity in *L. montandoni* ($R = 9.2$). The highest allelic richness in *L. vulgaris* was detected in populations from Romania as well as northern and central Serbia (for most of them $R > 20$). Less diverse populations ($7 < R < 10$) inhabit the southern part of the species distribution (Greece, Turkey, Montenegro, Croatia). Comparing both *L. vulgaris* groups (southern and northern), significantly higher allelic richness was observed in the north (Mann–Whitney test, $P = 0.002$). The number of private alleles in particular populations ranged from 0 to 16 (Table 5, Fig. 4), and the average proportion of private alleles was higher in *L. vulgaris* (0.345) than in *L. montandoni* (0.172) populations (Mann–Whitney test $P = 0.044$).

Table 5 Summary of MHC class II and microsatellite variation in populations

No.	Population	Country	N	Species	Group	Na	R	Nap	R_{6loci}	$R_{minus\_Lm\_521}$	$R_{minus\_Lm\_AHNC3}$
1	Karagol	Turkey	19	*Lvsch*	vulS	12	9.9	11	–	–	4
2	Subasi	Turkey	15	*Lvsch*	vulS	10	9.4	6	–	–	4.7
3	Seli	Greece	13	*Lvg*	vulS	8	7.5	4	–	3	–
4	Demirbey	Turkey	14	*Lvk*	vulS	12	10.7	9	5.5	4.8	6.2
5	Kentriko	Greece	12	*Lvg*	vulS	9	8.8	2	–	3.9	–
6	Bakuriani	Georgia	5	*Lvl*		2		2	–	–	–
7	Donji Stoj	Montenegro	15	*Lvg*	vulS	14	12.9	4	5.7	5	6.3
8	Kostanjica	Montenegro	20	*Lvg*	vulS	10	8.5	3	3.7	3.5	3.9
9	Tivat	Montenegro	16	*Lvg*	vulS	10	8.6	1	3.3	3.5	3.5
10	Alexandrowo	Bulgaria	9	*Lvv*	vulN	10		3	–	–	–
11	Vlasi	Serbia	17	*Lvv*	vulN	11	10	1	5.2	4.4	5.5
12	Gornja Sabanta	Serbia	18	*Lvv*	vulN	29	22	16	7.3	6.9	8
13	Benkovac	Croatia	18	*Lvv*	vulS	13	9.9	1	–	3.3	–
14	Donji Zemunik	Croatia	8	*Lvv*	vulS	3		0	–	–	–
15	Goryachiy Klyuch	Russia	1/0	*Lvl*		4		4	–	–	–
16	Glušci	Serbia	20	*Lvv*	vulN	34	22.6	15	7.3	7.1	7.9
17	Plopu	Romania	17	*Lvv*	vulN	29	22.5	11	5.9	5.5	6.5
18	Salakovac	Croatia	14	*Lvm*	vulS	9	8.4	6	–	–	6.5
19	Vilanesti	Romania	19	*Lvv*	vulN	26	18.2	11	6.2	5.8	6.8
20	Voina	Romania	13	*Lm*	vulN	18	16.7	1	3.9	3.5	3.7
21	Budeni	Romania	15	*Lvv*	vulN	14	12.9	2	4.4	4.2	4.9
22	Pasul Musat	Romania	18	*Lm*	mon	20	16.7	0	5.2	4.6	5.6
23	Sighişoara	Romania	13	*Lva*	vulN	28	25.1	10	6.4	5.9	7.1
24	Bátászek	Hungary	15	*Lvv*	vulN	8	7.1	1	6.8	6.5	7.5
25	Cărpiniş	Romania	22	*Lva*	vulN	29	21.5	12	6.6	6	7.1
26	Zalaháshágy	Hungary	26/25	*Lvv*	vulN	27	17.6	12	6.6	6.2	7.6
27	Secu	Romania	11	*Lm*	mon	21	21	2	4.9	4	5.4
28	Petru Voda	Romania	20	*Lm*	mon	27	19.5	3	4.9	4.4	5.3
29	Pasul Gutai	Romania	13	*Lm*	mon	23	21.2	10	6.4	5.6	7.2
30	Dzembronia	Ukraine	21	*Lm*	mon	35	23	14	5.1	4.5	5.6
31	Lipníky	Slovakia	9	*Lvv*	vulN	8		1	–	–	–
32	Krempna	Poland	13	*Lm*	mon	22	20.1	6	4	3.5	4.3
33	Łopuszna	Poland	17	*Lm*	mon	16	13.9	0	4	3.6	4.2
34	Jeseniki	Czech Republic	17	*Lm*	mon	11	9.2	2	3.8	3.6	4.3
35	Radzyny	Poland	14	*Lvv*	vulN	12	10.1	1	5.4	4.5	6.1
	Lissotriton montandoni		**143**	***Lm***		**21**	**17.9**	**4**			
	Lissotriton vulgaris		**383/381**	***Lvv***		**15**	**13.5**	**6**			
	Total number of alleles					**299**					

MHC, major histocompatibility complex.
Sample size (N, if sample sizes differed for the plusS and minusS data sets, both are given separated by slash), number of alleles (Na), allelic richness (R), number of private alleles (Nap), mean allelic richness for microsatellite loci in three data sets: six loci (R_{6loci}), minus_Lm_521 ($R_{minus\_Lm\_521}$) and minus_Lm_AHNC3 ($R_{minus\_Lm\_AHNC3}$). The populations are assigned to *Lissotriton montandoni* (mon), southern (vulS) and northern *L. vulgaris* group (vulN). Average values for all summary statistics for *L. montandoni* and *L. vulgaris* are given in bold.

Structuring of MHC II variation among populations and species

Two aspects of MHC class II structuring were worth exploring: the distribution of variation between species and the geographical structuring within species. The species shared 39 alleles, significantly fewer than expected from the random assignment of alleles to species (randomization test, $P < 0.001$). The nonrandom distribution of alleles between species also occurred in Romania and Ukraine, where multiple, geographically adjacent populations of both species were sampled (randomization test, $P < 0.001$). In contrast, almost all *L. montandoni* individuals had one or more alleles shared between species, indicating that some genetic exchange between species may have occurred. To test this hypothesis, it was useful to divide *L. vulgaris* populations into two groups: northern, which, due to geographical proximity, might have recently (e.g. since the last glacial period) exchanged genes with *L. montandoni*,

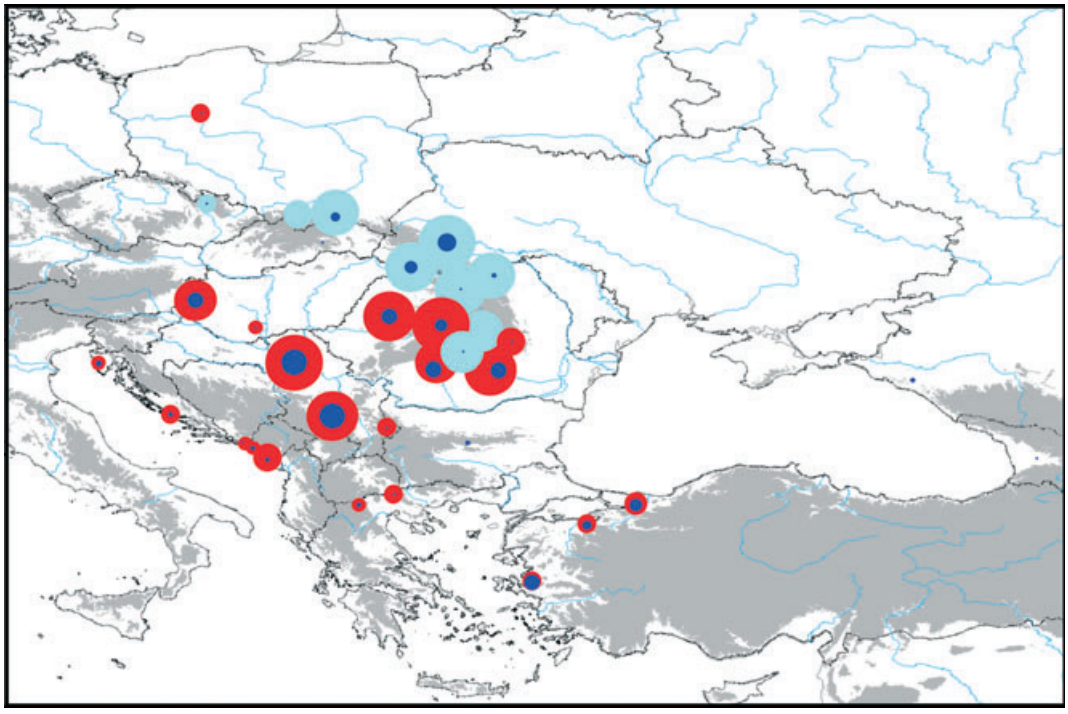

Fig. 4 Allelic richness and number of private alleles (dark blue, range 0–16) in populations of *Lissotriton vulgaris* (red, range 7.1–25.1) and *Lissotriton montandoni* (light blue, range 9.2–23.0).

Table 6 The results of AMOVAS, in which northern *Lissotriton vulgaris*, southern *L. vulgaris* and *Lissotriton montandoni* were used as the highest level of hierarchical grouping of populations, and two groups were analysed at a time

	Source of variation		
Groups	Among groups	Among populations	Within populations
Northern *L. vulgaris* and *L. montandoni*	0.89^{NS}	18.06***	81.04***
Southern *L. vulgaris* and *L. montandoni*	4.61***	25.79***	69.60***
Northern *L. vulgaris* and southern *L. vulgaris*	3.95***	29.26***	66.79***

Percentages of variation accounted for by the respective hierarchical levels are given. NS, not significant; ***$P < 0.001$.

and southern, for which recent gene exchange with *L. montandoni* is unlikely (Fig. 1). We found that a higher fraction of alleles was shared between *L. montandoni* and northern *L. vulgaris* (39/172) than between *L. montandoni* and southern *L. vulgaris* (3/75) (Fisher's exact test, $P = 2 \times 10^{-4}$). The average pairwise F_{ST} between northern *L. vulgaris* and *L. montandoni* (0.187 ± (SE) 0.0080) was significantly lower than between northern and southern *L. vulgaris* populations (0.357 ± 0.0088, randomization test, $P < 0.001$); it was also lower than between southern *L. vulgaris* populations and *L. montandoni* (0.296 ± 0.0073, $P < 0.001$). These observations support the hypothesis of interspecific MHC class II gene flow between *L. vulgaris* and *L. montandoni*. Further support is provided by the AMOVAS (Table 6). When northern, southern *L. vulgaris* and *L. montandoni* were used as the highest level of hierarchical grouping of populations, with two groups of populations analysed at a time, the between-group component was not significant ($P = 0.07$) in AMOVA for northern *L. vulgaris* and *L. montandoni*, but it was highly significant for southern *L. vulgaris* and *L. montandoni* (4.61% of variation explained, $P < 0.001$) and for northern and southern *L. vulgaris* (3.95%, $P < 0.001$) AMOVAS.

Structure analysis revealed ten clusters, with geographical distributions not corresponding to these of either subspecies or species (Fig. 5). Mixed ancestry was inferred for individuals in many *L. montandoni* and northern *L. vulgaris* populations. In contrast, most populations from the southern part of *L. vulgaris* were homogeneous, i.e. individuals were classified to a single cluster only.

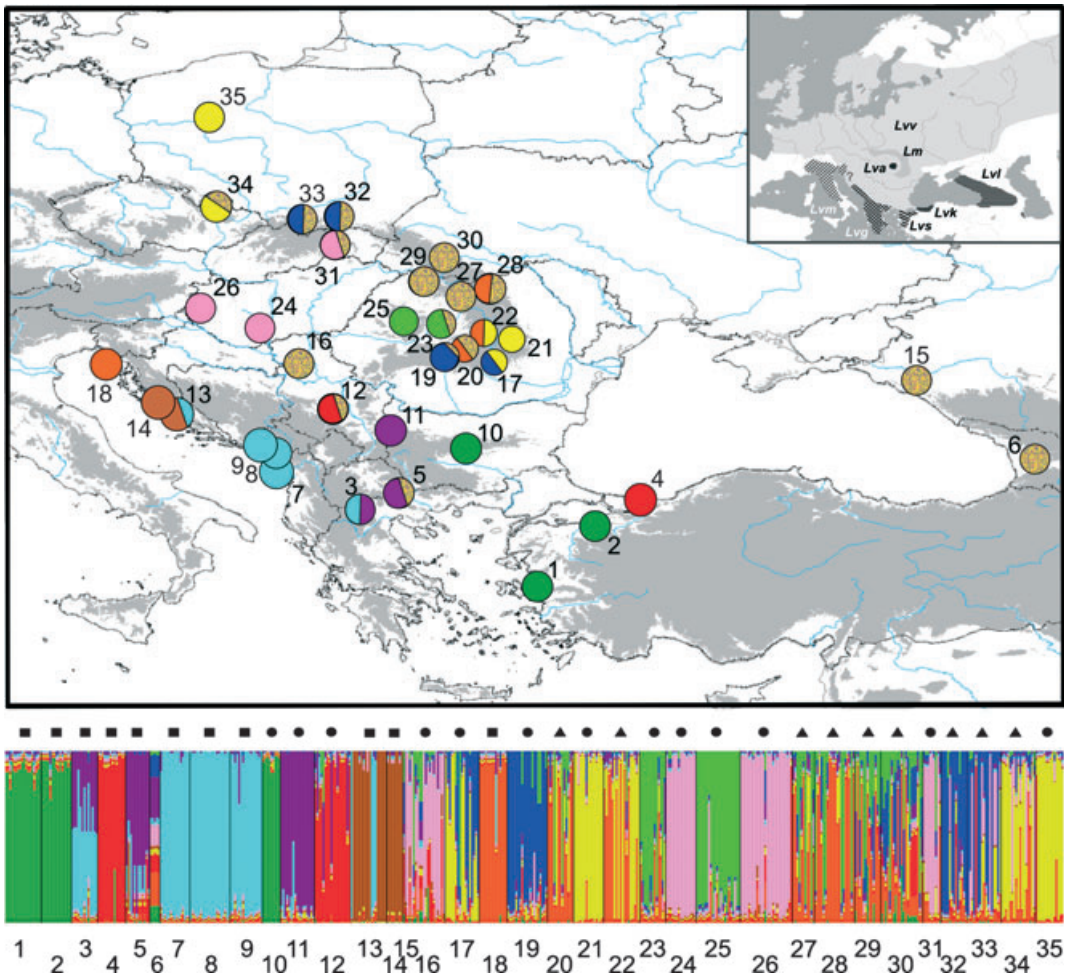

Fig. 5 Genetic structure of *Lissotriton vulgaris* and *Lissotriton montandoni* populations in major histocompatibility complex II inferred by Structure analysis for $K = 10$. Assignment of populations to particular clusters is shown on the map. The circles, squares and triangles correspond to northern *L. vulgaris*, southern *L. vulgaris* and *L. montandoni* populations, respectively.

Interestingly, *L. montandoni* populations often exhibited substantial admixture from genetic clusters found predominantly in *L. vulgaris*, whereas the opposite has not been observed (Fig. 5). Such pattern suggests unidirectional MHC gene flow from *L. vulgaris* to *L. montandoni*.

Pairwise F_{ST} showed a pattern of interpopulation differentiation consistent with Structure analysis (Table S8, Supporting information). F_{ST} ranged from 0.04 (Donji Stoj [7]—Kostanjica [8]) to 0.78 (Bakuriani [6]—Donji Zemunik [14]). The most divergent populations were found in southern Europe, as evident from a MDS of the pairwise F_{ST} matrix (Fig. 6A).

Microsatellite variation and comparison of MHC and neutral variation

The total number of distinct alleles at microsatellite loci ranged from 11 (Lm_AHNC3) to 180 (Lm_521). Microsatellite allelic richness for each population is provided in Table 5, and the number of alleles per locus and the estimated frequency of null alleles (Pnull) are in Table S9 (Supporting information). The pattern of higher allelic richness in *L. vulgaris* held for all data sets, but significance was reached only in the six loci data set (Mann–Whitney test, $P = 0.0136$) compared to the minus Lm_521 ($P = 0.0708$) and minus Lm_AHNC3 ($P = 0.0694$) data sets. We observed a significant but moderate correlation between population-level allelic richness in MHC and microsatellites (six loci data set: $r^2 = 0.21$, $P = 0.023$; minus Lm_521 data set: $r^2 = 0.26$, $P = 0.006$; minus Lm_AHNC3 data set: $r^2 = 0.20$, $P = 0.021$). After the correction for multiple tests, only the correlation for the minus Lm_521 data set remained significant ($P < 0.02$). The average pairwise F_{ST} between northern *L. vulgaris* and *L. montandoni* (0.160 ± 0.0051) was not significantly different than that between northern and southern

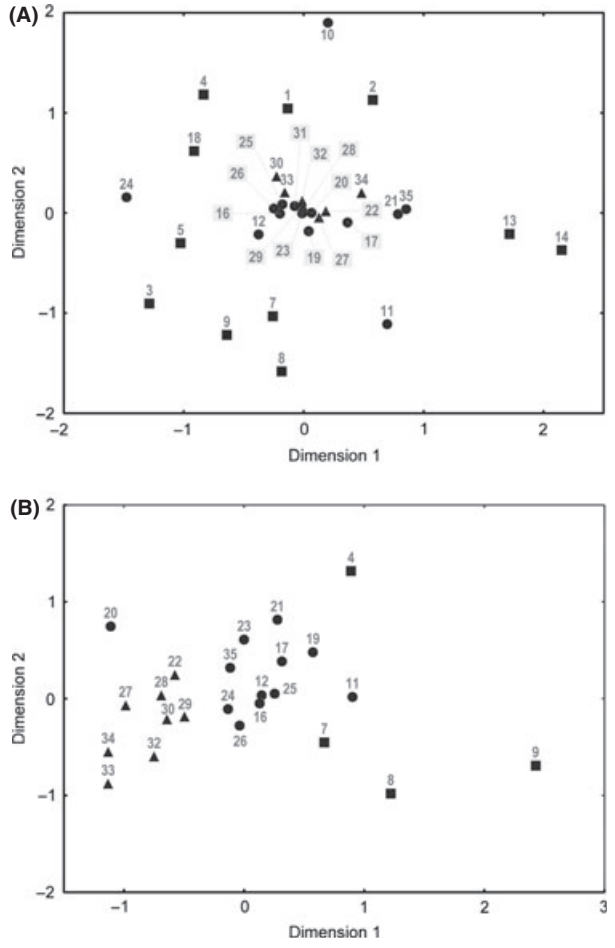

Fig. 6 Nonmetric two-dimensional scaling of the pairwise F_{ST} matrix computed for binary-encoded major histocompatibility complex class II data, stress 0.233 (A) and for microsatellites (six loci data set, 24 populations), stress 0.145 (B). The names of populations corresponding to numbers are given in Tables 5 and S1, Supporting information. The circles, squares and triangles correspond to northern *Lissotriton vulgaris*, southern *L. vulgaris* and *Lissotriton montandoni* populations, respectively.

L. vulgaris populations (0.200 ± 0.0095, randomization test, $P = 0.193$) but was significantly lower than between southern *L. vulgaris* and *L. montandoni* (0.276 ± 0.0116, randomization test, $P < 0.001$). Qualitatively identical results were obtained with microsatellites encoded as binary data, which facilitates direct comparison with binary-encoded MHC (data not shown).

Structure analysis performed for microsatellites in the six loci data set (24 populations) revealed little admixture between *L. montandoni* and *L. vulgaris*, contrary to the pattern observed in MHC (Fig. 7). Also, MDS of the pairwise F_{ST} matrix confirmed distinctiveness of *L. montandoni* from *L. vulgaris* in microsatellites (Fig. 6B).

Discussion

Our study is a rare example of research assessing MHC variation across multiple populations of closely related species and of both recent and ancient divergence, distributed over a large geographical area. This design takes into account the geographical and temporal aspects of MHC variation, provides information about the actual species-wide extent of MHC variation and complements more localized studies that typically investigate evolutionary mechanisms shaping MHC variation at the microscale and/or document associations between MHC variation and pathogens (reviewed in Bernatchez & Landry 2003; Sommer 2005; Piertney & Oliver 2006). In addition, the study demonstrates the utility of high-throughput sequencing not only for complex, multilocus systems with a large number of alleles per individual, as has already been shown (Kloch *et al.* 2010; Zagalska-Neubauer *et al.* 2010), but also for systems with more limited within-individual diversity but with large population-scale variation. The most attractive aspect of using the high-throughput sequencing here is the combination of genotyping and characterization of new alleles in a single step, which saves time and effort. Below, we discuss the organization of MHC class II genes, the variation within and between species and the evolutionary factors that may have contributed to the observed genetic and geographical patterns.

The organization of MHC II in *L. vulgaris* and *L. montandoni* shares important similarities with other newts investigated so far, the Alpine and great crested (Babik *et al.* 2008, 2009a), yet it also exhibits remarkable differences. All newt species have multiple MHC class II loci, some of which have undergone pseudogenization. The presence of MHC II pseudogenes is a universal feature of MHC class II in vertebrates and is believed to result from frequent duplications followed by a subsequent divergence and loss of function (Nei & Rooney 2005). Despite limited within-individual variation and, consequently, few loci, distinguishing functional alleles from pseudogenes proved much more difficult in *L. vulgaris* and *L. montandoni* than in other newt species. Likewise, assignment of alleles to particular loci was not possible. These difficulties resulted from: the lack of signatures of pseudogenization in MHC sequences (e.g. stop codons or substitutions in conserved amino acid positions), the lack of clustering of expressed or nonexpresed alleles in the phylogenetic tree and, in some cases, a high sequence similarity of expressed and nonexpressed alleles. Whereas assignment of alleles to loci may sometimes be achieved by examining intron sequences (Ono *et al.* 1993; Canal *et al.* 2010), obtaining the complete picture of MHC II expression pattern in *Lissotriton* newts would be much

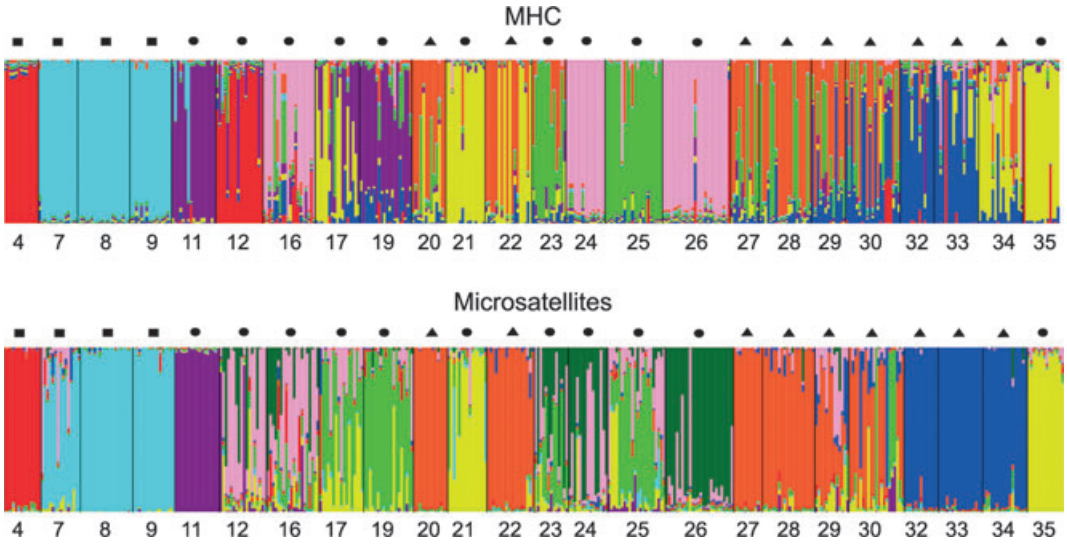

Fig. 7 Comparison of genetic structure of *Lissotriton vulgaris* and *Lissotriton montandoni* populations in major histocompatibility complex (MHC) II and microsatellites inferred by Structure analysis for $K = 8$ (MHC) and $K = 9$ (microsatellites). The names of populations corresponding to numbers are given in Tables 5 and S1, Supporting information. The circles, squares and triangles correspond to northern *L. vulgaris*, southern *L. vulgaris* and *L. montandoni* populations, respectively.

more challenging. It could be argued that examining a larger sample of cDNA amplicons may help clarify the expression pattern, but the majority of individuals would need to be screened, which is not possible because of conservation concerns. Moreover, if certain alleles are expressed on some haplotypes but not on others, determining the MHC class II expression pattern would presumably require examination of cDNA from all individuals (or at least all genotypes). Finally, we cannot exclude the possibility that some alleles we consider nonexpressed are actually derived from nonclassical MHC genes, expressed at low level or only in certain tissues.

The observed range of the number of alleles per individual indicates either differences among MHC haplotypes in the number of loci or the presence of identical alleles in more than one locus. Both are common features of MHC architecture and result from its extraordinary dynamics: frequent duplications and recombination between loci (e.g. Bowen *et al.* 2004; Ellis *et al.* 2005; Otting *et al.* 2005; Bontrop 2006). The lack of alleles or clusters of similar alleles present in most individuals indicates that all loci in *Lissotriton* newts are at least moderately polymorphic. This is in contrast to the alpine newt, in which one of the loci displayed a very low variation and probably represents nonclassical MHC II (Babik *et al.* 2008).

Both species exhibited substantial MHC variation over their distribution area. Diversity was manifested by a large overall number of alleles and generally high within-population variation. MHC allelic richness differed substantially between *L. vulgaris* and *L. montandoni* and over the geographical range of *L. vulgaris*. Almost all populations of *L. montandoni* had high allelic richness, which appears surprising at first sight. In the light of the limited distribution and glacial history (apparently small refugia in the Carpathians), we would expect lower MHC variation in this species compared to *L. vulgaris*, as observed in microsatellites. In general, substantial interspecific differences in MHC variation may be an effect of dissimilar selective pressures stemming from general ecological differences between the two species. Such differences were described for several species (e.g. Ekblom *et al.* 2007; Matthews *et al.* 2010; Eizaguirre *et al.* 2011). However, because MHC II allelic richness was similar for both species in Romania, it appears that differences in within-population variation, rather than being a species-specific feature, may reflect the diverse histories of *L. vulgaris* populations in various geographical regions. A high diversity of mtDNA was observed in *L. montandoni* as a consequence of well-documented introgression from several *L. vulgaris* mtDNA lineages (Babik *et al.* 2005a). Our analyses indicate that high MHC II variation observed in the Carpathian region in both species and in central Europe and northern Balkans in *L. vulgaris* may also result from interspecific hybridization and gene flow between the two species. We found that *L. vulgaris* in the northern, but not in the southern part of the range, shares many alleles with *L. montandoni*. If identical alleles were retained across the speciation event (trans-species polymorphism), we would not expect the number of shared alleles to be associated with an opportunity for hybridization. Even more

strikingly, the average pairwise F_{ST} between northern *L. vulgaris* and *L. montandoni* was significantly lower than between northern and southern *L. vulgaris* populations for MHC, and the between-group component of variation was not significant for the AMOVA comprising northern *L. vulgaris* and *L. montandoni*. Unequivocal distinguishing between uni- (as observed for mtDNA) and bi-directional MHC introgression is difficult with the available data; however, the Structure analyses suggest that most interspecific MHC gene flow has occurred from *L. vulgaris* to *L. montandoni*. The directionality of MHC introgression could be tested directly in the fine-scale genetic analysis of hybrid zones. Apart from hybridization, similar selective pressures acting on MHC in both species in the northern region might increase the probability that alleles are shared among species owing to regional retention of ancestral polymorphism. However, given multiple ecological and behavioural differences between the species (Babik *et al.* 2003, 2005a and references therein), we consider this explanation unlikely.

Major histocompatibility complex alleles, if advantageous through, for example, the mechanism of negative frequency-dependent selection (Snell 1968; Borghans *et al.* 2004), would introgress much more quickly than neutral variation (Schierup *et al.* 2000; Barton 2001; Muirhead 2001). Indeed, the pattern of variation between regions for the microsatellites differed from that of MHC, with average pairwise F_{ST} between northern *L. vulgaris* and *L. montandoni* of similar magnitude and not significantly different from that between northern and southern *L. vulgaris* populations for MHC.

The evidence for more extensive (although not free) MHC gene flow among populations in the north in general (intra- and interspecific) was provided by Structure analysis and MDS of the F_{ST}. Mixed ancestry was commonly inferred for individuals from populations in question, and the overall genetic differentiation among populations, although mostly significant, was moderate.

In contrast to the northern Balkans and central Europe, populations inhabiting the southern fringe of the *L. vulgaris* distribution exhibited lower MHC class II variation, which, in principle, may have been caused by the long-term isolation of these groups and the loss of variation through drift. The region is inhabited by multiple morphologically distinct subspecies of *L. vulgaris*, and several old, divergent mtDNA lineages with localized distributions have been described in this area (Babik *et al.* 2005a). Relative isolation and distinctiveness of some of the southern groups have also been suggested by MHC II analysis in Structure and MDS of the F_{ST} matrix. On the other hand, Structure groups agree poorly with the distribution of subspecies and mtDNA lineages (Fig. 5; Babik *et al.* 2005a), and populations of *L. vulgaris* in at least some southern regions have been large and historically stable as indicated by our previous studies in Turkey (Nadachowska & Babik 2009). Thus, it is rather unlikely that the populations have been subjected to prolonged and strong drift and the evident signatures of positive selection emphasize the importance of adaptive factors in shaping the patterns of MHC variation in *Lissotriton* newts. The role of demographic factors cannot be ruled out entirely, however, because of a significant correlation between allelic richness in MHC and microsatellites. Either the effect of drift, although present, is of a relatively minor importance or microsatellites are not very informative when studying highly divergent populations, which increases the risk of encountering null alleles (Chapuis & Estoup 2007), as was also evident in our study. Nevertheless, we do believe that the adopted correction alleviated the problem of null alleles, and the correlation between MHC and microsatellite variation is indeed moderate.

In addition to differences in the extent of gene flow or the apparently moderate effect of drift, other factors may also contribute to the observed differences in MHC II variation over the geographical range of *L. vulgaris*. Stronger genetic structure in the south may reflect the existence of local adaptations as indicated by a relatively high proportion of private alleles in many populations. Finally, we cannot exclude the possibility that the genomic architecture of MHC class II may differ among geographical regions, and in the southern, more divergent populations of *L. vulgaris* haplotypes with a lower number of loci occur in higher frequencies than elsewhere. More extensive genomic data would be necessary to test this hypothesis.

Regardless of the reason for the observed regional differences in MHC variation, the geographical distribution of MHC II variation reflected the distribution of newts' glacial refugia identified previously by mtDNA analysis (Babik *et al.* 2005a). Carpathians and the Carpathian Basin, which served as the main sources of populations that recolonized northern areas after glaciations (Jaarola & Searle 2002; Babik *et al.* 2004; Sommer & Nadachowski 2006; Ursenbacher *et al.* 2006; Hofman *et al.* 2007), displayed the highest variation of MHC class II, and this pattern is consistent with high MHC variation observed for the great crested newt in this area (Babik *et al.* 2009a).

Conclusions

The large-scale, comprehensive analysis of MHC class II variation in the closely related newts *L. vulgaris* and *L. montandoni* revealed an extensive, geographically structured polymorphism. Populations at the southern margin of the *L. vulgaris* distribution, inhabited by old

and distinct lineages, exhibited moderate MHC variation and little gene flow or extensive local adaptation. The most variable populations of *L. vulgaris* were found in central Europe and the northern Balkans, areas also harbouring multiple, highly variable mtDNA lineages. Thus, gene flow between genetically differentiated populations may have contributed to the high MHC variation there. Strikingly, *L. montandoni*, a species with limited geographical distribution, as it is confined to the Carpathian Mountains, has high MHC variation. Our results indicate that this substantial polymorphism may be attributed to gene flow from the neighbouring populations of *L. vulgaris*, which are also highly variable. A moderate but significant correlation between the MHC and microsatellite allelic richness indicates that demographic factors may have, to some extent, contributed to the present-day pattern of MHC variation. However, unequivocal signatures of adaptive evolution in MHC class II sequences emphasize the role of selection in the longer timescale.

Acknowledgements

We thank Dan Cogălniceanu, Severus Daniel Covaciu-Marcov, Piotr Grodek and Marcin Liana for their help in samples collection. We thank Marzanna Kuenzli and Weihong Qui for running 454 sequencing and Michał Stuglik and Filip Kapustka for help in bioinformatic analysis. The study was funded by the Foundation for Polish Science, professor subsidy 9/2008 to J.R., N N304 074235 to K.N-B. and the Jagiellonian University (DS/WBINOZ/INOŚ/762/10).

References

Babik W, Szymura JM, Rafinski J (2003) Nuclear markers, mitochondrial DNA and male secondary sexual traits variation in a newt hybrid zone (*Triturus vulgaris* × *T. montandoni*). *Molecular Ecology*, **12**, 1913–1930.

Babik W, Branicki W, Sandera M *et al.* (2004) Mitochondrial phylogeography of the moor frog, *Rana arvalis*. *Molecular Ecology*, **13**, 1469–1480.

Babik W, Branicki W, Crnobrnja-Isailovic J *et al.* (2005a) Phylogeography of two European newt species – discordance between mtDNA and morphology. *Molecular Ecology*, **14**, 2475–2491.

Babik W, Durka W, Radwan J (2005b) Sequence diversity of the MHC DRB gene in the Eurasian beaver (*Castor fiber*). *Molecular Ecology*, **14**, 4249–4257.

Babik W, Pabijan M, Radwan J (2008) Contrasting patterns of variation in MHC loci in the Alpine newt. *Molecular Ecology*, **17**, 2339–2355.

Babik W, Pabijan M, Arntzen JW *et al.* (2009a) Long-term survival of a urodele amphibian despite depleted major histocompatibility complex variation. *Molecular Ecology*, **18**, 769–781.

Babik W, Taberlet P, Ejsmond MJ, Radwan J (2009b) New generation sequencers as a tool for genotyping of highly polymorphic multilocus MHC system. *Molecular Ecology Resources*, **9**, 713–719.

Barton NH (2001) The role of hybridization in evolution. *Molecular Ecology*, **10**, 551–568.

Beebee TJC, Griffiths RA (2005) The amphibian decline crisis: a watershed for conservation biology? *Biological Conservation*, **125**, 271–285.

Bernatchez L, Landry C (2003) MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *Journal of Evolutionary Biology*, **16**, 363–377.

Bonneaud C, Perez-Tris J, Federici P, Chastel O, Sorci G (2006) Major histocompatibility alleles associated with local resistance to malaria in a passerine. *Evolution*, **60**, 383–389.

Bontrop RE (2006) Comparative genetics of MHC polymorphisms in different primate species: duplications and deletions. *Human Immunology*, **67**, 388–397.

Borghans JAM, Beltman JB, De Boer RJ (2004) MHC polymorphism under host-pathogen coevolution. *Immunogenetics*, **55**, 732–739.

Bos DH, Waldman B (2006) Polymorphism, natural selection, and structural modeling of class Ia MHC in the African clawed frog (*Xenopus laevis*). *Immunogenetics*, **58**, 433–442.

Bowen L, Aldridge BM, Gulland F *et al.* (2004) Class II multiformity generated by variable MHC-DRB region configurations in the California sea lion (*Zalophus californianus*). *Immunogenetics*, **56**, 12–27.

Canal D, Alcaide M, Anmarkrud JA, Potti J (2010) Towards the simplification of MHC typing protocols: targeting classical MHC class II genes in a passerine, the pied flycatcher *Ficedula hypoleuca*. *BMC Research Notes*, **3**, 236.

Chapuis MP, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution*, **24**, 621–631.

Clarke B, Kirby DRS (1966) Maintenance of histocompatibility polymorphisms. *Nature*, **211**, 999.

Delarbre C, Jaulin C, Kourilsky P, Gachelin G (1992) Evolution of the major histocompatibility complex – a 100-fold amplification of Mhc class-I genes in the African pygmy mouse *Nannomys setulosus*. *Immunogenetics*, **37**, 29–38.

Deter J, Bryja J, Chaval Y *et al.* (2008) Association between the DQA MHC class II gene and Puumala virus infection in *Myodes glareolus*, the bank vole. *Infection Genetics and Evolution*, **8**, 450–458.

Doherty PC, Zinkernagel RM (1975) Enhanced immunological surveillance in mice heterozygous at H-2 gene complex. *Nature*, **256**, 50–52.

Eizaguirre C, Lenz TL, Sommerfeld RD *et al.* (2011) Parasite diversity, patterns of MHC II variation and olfactory based mate choice in diverging three-spined stickleback ecotypes. *Evolutionary Ecology*, **25**, 605–622.

Ekblom R, Saether SA, Jacobsson P *et al.* (2007) Spatial pattern of MHC class II variation in the great snipe (*Gallinago media*). *Molecular Ecology*, **16**, 1439–1451.

Ellis SA, Morrison WI, MacHugh ND *et al.* (2005) Serological and molecular diversity in the cattle MHC class I region. *Immunogenetics*, **57**, 601–606.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.

Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes*, **7**, 574–578.

Galan M, Guivier E, Caraux G, Charbonnel N, Cosson JF (2010) A 454 multiplex sequencing method for rapid and reliable genotyping of highly polymorphic genes in large-scale studies. *BMC Genomics*, **11**, 296.

Garrigan D, Hedrick PW (2003) Perspective: detecting adaptive molecular polymorphism: lessons from the MHC. *Evolution*, **57**, 1707–1722.

Hall TA (1999) Bioedit: an user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposia Series*, **41**, 95–98.

Hauswaldt JS, 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.

Hofman S, Spolsky C, Uzzell T *et al.* (2007) Phylogeography of the fire-bellied toads Bombina: independent Pleistocene histories inferred from mitochondrial genomes. *Molecular Ecology*, **16**, 2301–2316.

Hood GM (2010) PopTools version 3.2. Available on the internet URL http://www.poptools.org.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, **9**, 1322–1332.

Hughes AL, Nei M (1992) Maintenance of MHC polymorphism. *Nature*, **355**, 402–403.

Huse SM, Huber JA, Morrison HG, Sogin ML, Mark WelchD (2007) Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biology*, **8**, R143.

Jaarola M, Searle JB (2002) Phylogeography of field voles (*Microtus agrestis*) in Eurasia inferred from mitochondrial DNA sequences. *Molecular Ecology*, **11**, 2613–2621.

Janeway CA, Travers P, Walport D, Shlomchik MJ (2004) *Immunobiology: The Immune System in Health and Disease*. Garland Publishing, New York.

Johanet A, Picard D, Garner TWJ *et al.* (2009) Characterization of microsatellite loci in two closely related *Lissotriton* newt species. *Conservation Genetics*, **10**, 1903–1906.

Kanagawa T (2003) Bias and artifacts in multitemplate polymerase chain reactions (PCR). *Journal of Bioscience and Bioengineering*, **96**, 317–323.

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

Kiemnec-Tyburczy KM, Richmond JQ, Savage AE, Zamudio KR (2010) Selection, trans-species polymorphism, and locus identification of major histocompatibility complex class II beta alleles of New World ranid frogs. *Immunogenetics*, **62**, 741–751.

Kloch A, Babik W, Bajer A, Sinski E, Radwan J (2010) Effects of an MHC-DRB genotype and allele number on the load of gut parasites in the bank vole *Myodes glareolus*. *Molecular Ecology*, **19**, 255–265.

Laurens V, Chapusot C, Ordonez MD *et al.* (2001) Axolotl MHC class II beta chain: predominance of one allele and alternative splicing of the beta 1 domain. *European Journal of Immunology*, **31**, 506–515.

Lenz TL, Becker S (2008) Simple approach to reduce PCR artefact formation leads to reliable genotyping of MHC and other highly polymorphic loci – Implications for evolutionary analysis. *Gene*, **427**, 117–123.

Lips KR, Diffendorfer J, Mendelson JR, Sears MW (2008) Riding the wave: reconciling the roles of disease and climate change in amphibian declines. *Plos Biology*, **6**, 441–454.

Loiseau C, Richard M, Garnier S *et al.* (2009) Diversifying selection on MHC class I in the house sparrow (*Passer domesticus*). *Molecular Ecology*, **18**, 1331–1340.

Longeri M, Zanotti M, Damiani G (2002) Recombinant DRB sequences produced by mismatch repair of heteroduplexes during cloning in *Escherichia coli*. *European Journal of Immunogenetics*, **29**, 517–523.

Mallet J (2005) Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, **20**, 229–237.

Margulies M, Egholm M, Altman WE *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, **437**, 376–380.

Martin DP, Lemey P, Lott M *et al.* (2010) RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics*, **26**, 2462–2463.

Matthews B, Harmon LJ, M'Gonigle L, Marchinko KB, Schaschl H (2010) Sympatric and allopatric divergence of MHC genes in threespine stickleback. *PLoS ONE*, **5**, e10948.

May S, Beebee TJC (2009) Characterisation of major histocompatibility complex class II alleles in the natterjack toad, *Bufo calamita*. *Conservation Genetics Resourses*, **1**, 415–417.

Milinski M (2006) The major histocompatibility complex, sexual selection, and mate choice. *Annual Review of Ecology Evolution and Systematics*, **37**, 159–186.

Minin V, Abdo Z, Joyce P, Sullivan J (2003) Performance-based selection of likelihood models for phylogeny estimation. *Systematic Biology*, **52**, 674–683.

Muirhead CA (2001) Consequences of population structure on genes under balancing selection. *Evolution*, **55**, 1532–1541.

Nadachowska K, Babik W (2009) Divergence in the face of gene flow: the case of two newts (Amphibia: Salamandridae). *Molecular Biology and Evolution*, **26**, 829–841.

Nadachowska K, Flis I, Babik W (2010) Characterization of microsatellite loci in the Carpathian newt (*Lissotriton montandoni*). *Herpetological Journal*, **20**, 107–110.

Nei M, Rooney AP (2005) Concerted and birth-and-death evolution of multigene families. *Annual Review of Genetics*, **39**, 121–152.

Ohta Y, Goetz W, Hossain MZ, Nonaka M, Flajnik MF (2006) Ancestral organization of the MHC revealed in the amphibian *Xenopus*. *Journal of Immunology*, **176**, 3674–3685.

Ono H, Ohuigin C, Vincek V, Klein J (1993) Exon-intron organization of fish major histocompatibility complex class-Ii B-genes. *Immunogenetics*, **38**, 223–234.

Otting N, Heijmans CMC, de Groot NG, Doxiadis GGM, Bontrop RE (2005) Unparalleled complexity of the MHC class 1 region in rhesus macaques. *Genes and Immunity*, **6**, S6.

Padidam M, Sawyer S, Fauquet CM (1999) Possible emergence of new geminiviruses by frequent recombination. *Virology*, **265**, 218–225.

Penn DJ (2002) The scent of genetic compatibility: sexual selection and the major histocompatibility complex. *Ethology*, **108**, 1–21.

Piertney SB, Oliver MK (2006) The evolutionary ecology of the major histocompatibility complex. *Heredity*, **96**, 7–21.

Posada D (2002) Evaluation of methods for detecting recombination from DNA sequences: empirical data. *Molecular Biology and Evolution*, **19**, 708–717.

Pounds JA, Bustamante MR, Coloma LA *et al.* (2006) Widespread amphibian extinctions from epidemic disease driven by global warming. *Nature*, **439**, 161–167.

Pritchard JK, 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. *Biological Conservation*, **143**, 537–544.

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.

Sammut B, Du Pasquier L, Ducoroy P *et al.* (1999) Axolotl MHC architecture and polymorphism. *European Journal of Immunology*, **29**, 2897–2907.

Schierup MH, Vekemans X, Charlesworth D (2000) The effect of subdivision on variation at multi-allelic loci under balancing selection. *Genetical Research*, **76**, 51–62.

Schwensow N, Fietz J, Dausmann KH, Sommer S (2007) Neutral versus adaptive genetic variation in parasite resistance: importance of major histocompatibility complex supertypes in a free-ranging primate. *Heredity*, **99**, 265–277.

Schwensow N, Dausmann K, Eberle M, Fietz J, Sommer S (2010) Functional associations of similar MHC alleles and shared parasite species in two sympatric lemurs. *Infection Genetics and Evolution*, **10**, 662–668.

Smith JM (1992) Analyzing the mosaic structure of genes. *Journal of Molecular Evolution*, **34**, 126–129.

Snell GD (1968) The H-2 locus of the mouse: observations and speculations concerning its comparative genetics and its polymorphism. *Folia Biologica (Prague)*, **14**, 335–358.

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

Sommer RS, Nadachowski A (2006) Glacial refugia of mammals in Europe: evidence from fossil records. *Mammal Review*, **36**, 251–265.

Stuart SN, Chanson JS, Cox NA *et al.* (2004) Status and trends of amphibian declines and extinctions worldwide. *Science*, **306**, 1783–1786.

Stuglik M, Radwan J, Babik W (2011) jMHC: software assistant for multilocus genotyping of gene families using next-generation amplicon sequencing. *Molecular Ecology Resources*, **11**, 739–742.

Szpiech ZA, Jakobsson M, Rosenberg NA (2008) ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics*, **24**, 2498–2504.

Takahata N, Nei M (1990) Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of Major Histocompatibility Complex loci. *Genetics*, **124**, 967–978.

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, **24**, 1596–1599.

Tong JC, Bramson J, Kanduc D *et al.* (2006) Modeling the bound conformation of *Pemphigus vulgaris*-associated peptides to MHC class II DR and DQ alleles. *Immunome Research*, **2**, 1.

Tournefier A, Laurens V, Chapusot C *et al.* (1998) Structure of MHC class I and class II cDNAs and possible immunodeficiency linked to class II expression in the Mexican axolotl. *Immunological Reviews*, **166**, 259–277.

Ursenbacher S, Carlsson M, Helfer V, Tegelstrom H, Fumagalli L (2006) Phylogeography and Pleistocene refugia of the adder (*Vipera berus*) as inferred from mitochondrial DNA sequence data. *Molecular Ecology*, **15**, 3425–3437.

Wegner KM, Kalbe M, Kurtz J, Reusch TBH, Milinski M (2003) Parasite selection for immunogenetic optimality. *Science*, **301**, 1343–1343.

Yang ZH (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**, 1586–1591.

Yang ZH, Wong WSW, Nielsen R (2005) Bayes empirical Bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution*, **22**, 1107–1118.

Zagalska-Neubauer M, Babik W, Stuglik M *et al.* (2010) 454 sequencing reveals extreme complexity of the class II Major Histocompatibility Complex in the collared flycatcher. *BMC Evolutionary Biology*, **10**, 395.

Zeisset I, Beebee TJC (2009) Molecular characterization of major histocompatibility complex class II alleles in the common frog, *Rana temporaria*. *Molecular Ecology Resources*, **9**, 738–745.

The study formed a part of the PhD thesis of K.N-B. She is a Postdoc at the Uppsala University studying speciation genomics in birds; P.Z., a PhD student at Jagiellonian University, is interested in the patterns of historical and contemporary gene flow and intra-specific differentiation between diverging populations; J.R. is a Professor at the Institute of Nature Conservation, Polish Academy of Sciences and at the Institute of Environmental Sciences, Jagiellonian University. His interests focus on sexual selection, its evolutionary-genetic background and conservation; W.B. uses molecular tools to study historical processes at and below the species level.

Data accessibility

Alignment of MHC II alleles is available as Supporting information. The sequences were also deposited in GenBank (accession numbers JN565304–JN565602). The spreadsheet with GenBank accessions numbers and corresponding names of MHC II alleles as well as MHC and microsatellite genotypes for all examined individuals is also available as Appendix S1 (Supporting information).

Supporting information

Additional supporting information may be found in the online version of this article.

Appendix S1 GenBank accessions numbers, MHC II and microsatellite genotypes.

Appendix S2 Alignment of MHC II alleles.

Table S1 Sampling localities.

Table S2 The average rates of nonsynonymous substitutions per nonsynonymous site (dN) and synonymous substitutions per synonymous sites (dS) with standard errors obtained through 1000 bootstrap replicates in parentheses, and the results of the *Z*-test of neutrality.

Table S3 Evaluation of the goodness of fit for different models of codon evolution and estimated parameter values.

Table S4 Summary of MHC class II variation in populations.

Table S5 The results of AMOVAS for the minusS dataset, in which northern *L. vulgaris*, southern *L. vulgaris* and *L. montandoni* were used as the highest level of hierarchical grouping of populations, and two groups were analyzed at a time.

Table S6 Parwise F_{ST} values for the dataset minusS, computed for binary encoded data.

Table S7 Comparison of cDNA and gDNA MHC alleles for three individuals with coverage sufficient for genotyping.

Table S8 Parwise F_{ST} values for the dataset plusS, computed for binary encoded data.

Table S9 Number of microsatellite alleles (Na) and null allele frequencies (Pnull) for each population.

Fig. S1 Schematic representation of the location of primers used to amplify 2nd exon of MHC class II (A) and the structure of amplicons generated with fusion primers (B).

Fig. S2 Allelic richness and number of private alleles (dark blue, range 0–8) in populations of *L. vulgaris* (red, range 5.8–19.3) and *L. montandoni* (light blue, range 7.8–21.0) for data set minusS.

Fig. S3 Genetic structure of *L. vulgaris* and *L. montandoni* populations in MHC II inferred by Structure analysis for $K = 10$ (data set minusS). Assignment of populations to particular clusters is shown on the map.

Fig. S4 Nonmetric two-dimensional scaling of the pairwise F_{ST} matrix computed for binary-encoded MHC class II data; stress = 0.203 (data set minusS). The names of populations corresponding to numbers are given in Tables 5 and S1, Supporting information.

Fig. S5 Distribution of the number of alleles per individual in *L. vulgaris* and *L. montandoni* in data set plusS (A, B) and in data set minusS (C, D).

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.