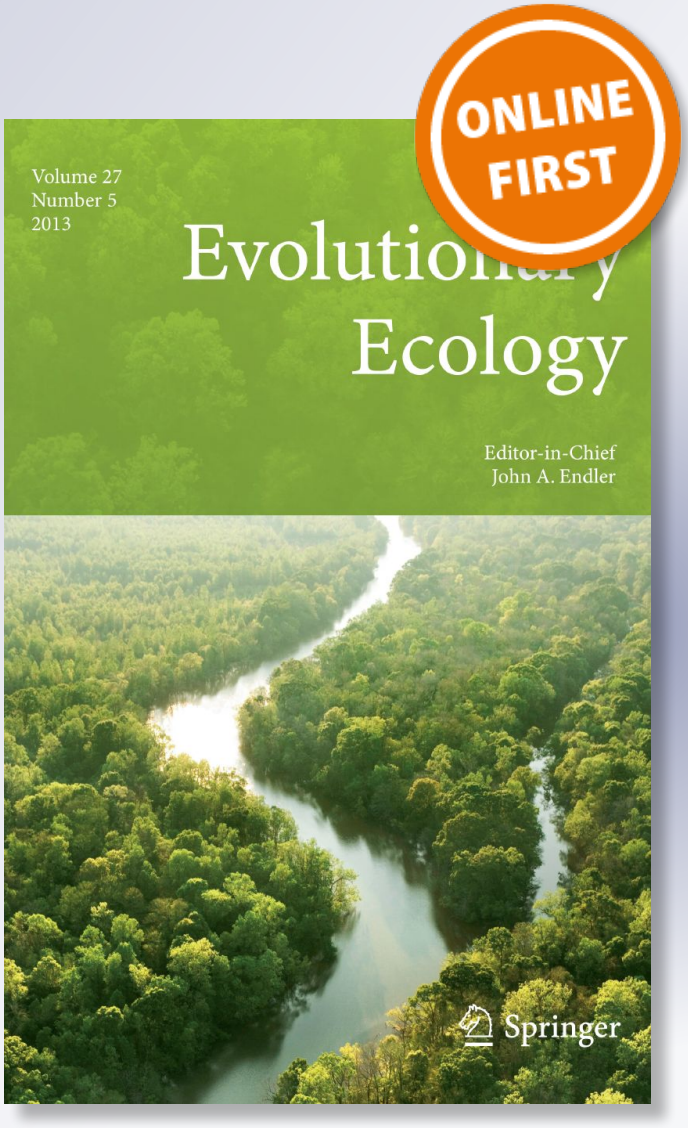
*Effects of parasite and historic driven selection on the diversity and structure of a MHC-II gene in a small mammal species (*Peromyscus leucopus*) undergoing range expansion*

A. André, V. Millien, M. Galan, A. Ribas & J. R. Michaux

Evolutionary Ecology

ISSN 0269-7653

Evol Ecol
DOI 10.1007/s10682-017-9898-z



Springer

Evol Ecol
DOI 10.1007/s10682-017-9898-z

ORIGINAL PAPER

Effects of parasite and historic driven selection on the diversity and structure of a MHC-II gene in a small mammal species (*Peromyscus leucopus*) undergoing range expansion

A. André[1,2] · V. Millien[2] · M. Galan[3] · A. Ribas[4] · J. R. Michaux[1,5]

Received: 21 December 2016 / Accepted: 26 April 2017

Abstract Genetic diversity may decrease from the centre to the margin of a species distribution range due to neutral stochastic processes. Selection may also alter genetic diversity in non-neutral markers, such as genes associated with the immune system. Both neutral processes and selection on the immune system are thus expected to affect the spatial distribution of such markers, but the relative strength of each has been scarcely studied. Here, we compared the diversity of a neutral marker (mitochondrial cytochrome *b*) and a selected marker (DRB gene from the MHC-II), in eastern-North American populations of white-footed mice (*Peromyscus leucopus*), a species known for its role of main reservoir of the Lyme disease. We observed distinct phylogeographic patterns with these two markers, which may be the result of selection pressure acting upon the DRB gene. As predicted by the central marginal hypothesis, we observed a loss of neutral genetic diversity toward the margin of the species distribution. A decrease in diversity was also observed for the DRB gene, likely due to genetic drift and positive selection operated by helminth parasites. Such a loss in genetic diversity at the range margin may slow down the ongoing expansion of *P. leucopus*, by counterbalancing the effect of global warming on the mouse survival at higher latitude.

Electronic supplementary material The online version of this article (doi:10.1007/s10682-017-9898-z) contains supplementary material, which is available to authorized users.

✉ A. André
adrien.andre@ulg.ac.be

[1] Conservation Genetics Laboratory, University of Liège, Boulevard du Rectorat 26, 4000 Liège, Belgium

[2] Redpath Museum, McGill University, 859 Sherbrooke West, Montreal, QC H3A OC4, Canada

[3] INRA, UMR CBGP, (INRA/IRD/Cirad/Montpellier SupAgro), Campus International de Baillarguet CS 30016, 34988 Montferrier-sur-Lez Cedex, France

[4] Section of Parasitology, Department of Biology, Healthcare and the Environment, Faculty of Pharmacy and Food Sciences, University of Barcelona, Av. Joan XXIII, sn, 08028 Barcelona, Spain

[5] CIRAD, UR AGIRs, 34398 Montpellier, France

Published online: 04 May 2017

Keywords Cytochrome-b gene · Major histocompatibility complex diversity · *Peromyscus leucopus* · Helminths

Introduction

Genetic diversity across populations of a species is generally not homogeneously distributed throughout the species range, and distinct processes drive such heterogeneity (Johansson et al. 2006; Eckert et al. 2008; Excoffier et al. 2009; Adams and Hadly 2012; Lau et al. 2014; Chen et al. 2015; Kohyama et al. 2015). The central marginal hypothesis (Prakash et al. 1969; Brussard 1984) states that genetic diversity decreases from the core to the margin of a species distribution range, with effective population size and gene flow decreasing in populations towards the range margin. Peripheral populations also tend to encounter more founder events (Hewitt 1999) and severe population bottlenecks (Nei et al. 1975), further reducing their genetic diversity (Hoffmann and Blows 1994). Climate warming that followed the last glacial maximum also had an impact on the current pattern of genetic diversity within a species range, and populations located today within their former glacial refuges are generally characterised by higher levels of genetic diversity than populations found in post glacial recolonised regions (refuge theory) (Michaux et al. 2003). In their review representing 115 plant and animal species (Eckert et al. 2008) detected the expected decline in diversity in peripheral populations for 64.2% of the studied species. Another review, published more recently by Pironon et al. (2016), found conflicting results as they observed that only 38% (43/114) of the studies conducted after 2008 provided evidences for the expected decline in genetic diversity.

While such spatial variation in genetic diversity have been reported for neutral markers such as mitochondrial DNA gene sequences (Adams and Hadly 2012) or nuclear microsatellite genotyping (Gassert et al. 2013; Johansson et al. 2006), selection can further alter genetic diversity in non-neutral markers. Of particular interest are genes associated with the immune system, considering that evolutionary processes associated with pathogen resistance can deeply shape the pattern of genetic diversity in such markers (Sommer 2005; Oliver et al. 2009; Turner et al. 2011; Zhang and He 2013). Major histocompatibility complex (MHC) genes play a key role in the immunity in mammals and are among the most polymorphic loci in vertebrates (Klein 1987). Their diversity is maintained by pathogen-driven balancing selection (Spurgin and Richardson 2010), mediated by two main processes. First, under the heterozygous advantage hypothesis (Doherty and Zinkernagel 1975; Hughes and Nei 1989; Hedrick 2012), individuals with a higher allelic richness are able to recognize and fight a wider array of pathogens, which increases their survival rate. Second, under the rare allele advantage hypothesis (Takahata and Nei 1990), rare alleles confer to their host a selective advantage due to the coevolution between hosts and parasites. The most resistant alleles are indeed supposed to be favoured and spread into the population, while parasites tend to evolve to evade the recognition of these most common alleles, favouring the hosts with rare alleles. Both neutral processes and pathogen-driven selection are thus expected to affect the spatial distribution in MHC diversity in mammal species, but the relative strength of each has been scarcely studied. Here, we addressed this question by comparing the DRB gene (MHC-II) diversity in populations of the white-footed mouse (*Peromyscus leucopus*) at its northern range margin in southern Quebec (Canada) and in more central regions of its range in the USA.

Over the last decades, *P. leucopus* has been shifting northward its northern range limit into Southern Quebec at a rate of 10 km per year, in response to global warming (Roy-Dufresne et al. 2013). A recent study of neutral markers in populations of *P. leucopus* from Southern Quebec hypothesized the effect of post-glacial recolonization on the current pattern of genetic diversity in the region (Fiset et al. 2015). A combination of five mitochondrial and nuclear sequences revealed two distinct lineages of *P. leucopus* in Southern Quebec separated by the Saint-Lawrence River, that were assigned to clades identified in the North-Eastern and Central USA (Fiset et al. 2015; Rowe et al. 2006). At a smaller geographic scale, major landscape barriers such as rivers and roads are limiting gene flow and modulating the pattern of ongoing range expansion in *P. leucopus* (Rogic et al. 2013; Marrotte et al. 2014; Leo et al. submitted). The monitoring of *P. leucopus* dispersal is of prime interest because this species is often considered as being the main reservoir for Lyme disease in eastern-North America (Ostfeld 2011). It is indeed the species that has the better success rate at transmitting the *Borrelia* to feeding ticks (Mather 1993). An improved understanding of *P. leucopus* genetic structure and diversity would therefore allow a better evaluation of the Lyme disease expansion risks in Quebec.

Here, we investigated the pattern of DRB and Cyt*b* gene diversity at a larger spatial scale than previous studies (Rogic et al. 2013; Fiset et al. 2015), and tested for a gradient of decreasing genetic diversity from the core region (North USA) to recently colonized regions in southern Quebec, as predicted by the central margin hypothesis. We also investigated the effect of past climate change and post-glacial recolonization by performing a phylogeographic analysis on the DRB and cytochrome B mitochondrial genes. The comparison of the genetic structure in a neutral (Cyt*b*) and a selected (DRB) marker allowed to better estimate the relative effects of past climate change and ongoing selection on the DRB gene in a species undergoing rapid range expansion. To assess the potential for pathogen-driven selection in *P. leucopus* in the most northern part of its range, we quantified helminth diversity and evaluated the relationship between infection rate and DRB diversity pattern in Quebec.

Materials and methods

Field sampling and museum specimens

Field sampling included 18 sites during the summers 2011–2014 across the Monteregie and Estrie regions in Southern Quebec, and located on the northern and southern shores of the Saint-Lawrence river (Table 1). The Richelieu river is located south of the St Lawrence river and runs in a north–south direction. At each site, 160 Sherman traps were placed every 10 meters in 4 grids of 4 × 10 traps. Traps were baited with a mixture of oat and peanut butter late afternoon and checked the following morning for three consecutive nights. A total of 407 *Peromyscus* sp. individuals were collected and identified to the species level using species-specific primers as described in Rogic et al. (2013). All samples were dissected on site and gastrointestinal tracts were stored frozen for subsequent analyses in the laboratory, few months later. All procedures were approved by the Ministère des Ressources Naturelles et de la Faune du Québec (SEG Permit #2011-05-15-014-00-S-F, #2012-07-16-1417-16-17-SF, #2013-07-04-14-16-17-SF and #2014-05-02-1638-05-16-SF) and McGill University Animal Care Committee (AUP#5420). We also obtained tissue samples (collected from 1980 to 2012) for 115 additional individuals from the Smithsonian

Table 1 Sample sizes for the analyses of the DRB and Cyt*b* genes, as well as parasite screening at each of the 27 study sites

Site	Origin	Latitude N	Longitude W	Sample size		
				DRB	Cyt*b*	Parasitology
GL1	USA	44.337508	89.382786	11	16	–
EC1	USA	36.915910	76.190741	18	1	–
EC2	USA	41.470692	71.698295	9	7	–
EC3	USA	41.473861	71.575224	18	6	–
EC4	USA	41.838283	71.539904	9	–	–
EC5	USA	41.892183	71.746949	13	9	–
EC6	USA	41.952682	72.123793	13	–	–
EC7	USA	42.647981	71.185085	15	–	–
EC8	USA	44.049466	71.273830	9	1	–
N1	Quebec	45.330541	74.394675	28	6	19
N2	Quebec	45.811591	73.463493	4	3	–
N3	Quebec	46.299909	73.093551	32	10	21
OR1	Quebec	45.036707	74.458869	30	3	21
OR2	Quebec	45.219070	73.928670	39	–	18
OR3	Quebec	45.241740	73.471424	19	1	19
OR4	Quebec	45.554397	73.336828	21	10	18
OR5	Quebec	45.878411	73.180910	4	5	–
ER1	Quebec	45.062074	73.284884	37	10	20
ER2	Quebec	45.117930	73.212812	15	6	10
ER3	Quebec	45.478300	73.171917	30	–	–
ER4	Quebec	45.521833	73.203283	30	1	–
ER5	Quebec	45.389050	73.203167	16	–	–
ER6	Quebec	45.423350	73.064050	26	–	–
ER7	Quebec	45.448116	72.907767	20	2	–
ER8	Quebec	45.655675	72.748593	14	–	–
ER9	Quebec	45.782053	73.016718	6	–	–
ER10	Quebec	45.864644	72.563806	36	1	19

Northern (N1–N3) and southern (OR 1–5 and ER 1–10) shores of the St Lawrence river, southern Quebec, Canada. GL1: Great Lakes, EC 1–8: East Coast, USA

Institution, National Museum of Natural History, and the Harvard Museum of Natural History collections for 9 additional sites in the USA from the East Coast (Virginia, Connecticut, Rhode Island, Massachusetts, New Hampshire) and the Great Lakes region (Table 1). These samples were all collected using Sherman traps too.

DNA extraction and sequencing

Genomic DNA was extracted from tissue samples (liver, muscle or ear) using the DNeasy™ Tissue Kit (Qiagen®, Hilden) following the manufacturer's protocol from the QIAcube (Qiagen®, Hilden).

Major histocompatibility complex DRB gene

Amplification of the DRB exon 2 gene was performed on 522 *P. leucopus* individuals (Table 1). Two batches of samples were sequenced at two different times using a 454 GS-FLX pyrosequencer (Roche) and an Illumina MiSeq sequencer. To check for congruency of the results, eight samples were processed using both sequencing methods.

454 GS-FLX sequencing

The second exon of the Mhc-Drb class II encoding the ligand-binding domain of the protein was amplified and sequenced in 213 individuals from 16 localities. We used a modified version of the primers JS1 (5′-GAGTGTCATTTCTACAACGGGAC-3′) and JS2 (5′-GATCCCGTAGTTGTGTYTGCA-3′), which amplify a 172-bp fragment (excluding primers) of exon 2 from the DRB gene in several mammal species, with the addition of individual-specific MIDS (multiplex identifiers) and adaptors required for emPCR and 454 sequencing (Galan et al. 2010). PCRs were performed following the procedure detailed in Galan et al. (2010). The SESAME software (Sequence Sorter and AMplicon Explorer) (Meglecz et al. 2011) was used to sort sequences, identify and discard artefactual variants, and generate the haplotypes and individual genotypes. Gene duplication is common in MHC genes, and, in theory, an individual can have more than two alleles. However, this was ruled out for several rodent species using the same PCR primers (Galan et al. 2010). Furthermore, Peromyscus individuals used here had a maximum of two alleles amplified by these PCR primers.

Illumina MiSeq sequencing

For the second batch of samples (317 individuals from 21 localities), the same gene specific primers JS1 and JS2 were used but with the Illumina specific adapters in 5′ position (Forward: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG; Reverse GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG). All PCRs were performed in a reaction volume of 25 µl, each containing 5 µl of fidelity buffer, 0.75 µl of dNTP, 0.75 µl of each primer, 0.25 µl of KAPA HiFi Hotstart Dna polymerase (KAPA Biosystems), 12.5 µl of pure grade water (GE Healthcare—Hyclone) and 5 µl of DNA. Thermocycling was carried out on a Mastercycler Gradient (Eppendorf) with an initial denaturation step of 5 min à 95° followed by 30 cycles of 30 s denaturation at 98 °C, 30 s annealing at 60° and 30 s elongation at 72° finished with a 10-min elongation at 72° and stopped by cooling down at 4 °C. Twenty-five µl of PCR products were purified using 20 µl of Agencourt AMPure XP beads and following the manufacturer protocol with the following change: purified products were eluted in a final volume of 25 µl of 10 nM Tris pH 8.5 instead of 50 µl. A second PCR was conducted using Nextera Index Kit to attach the dual indices and the Illumina sequencing adapters. Twenty-four indices were distributed horizontally and 16 indices were distributed vertically in order to individually tag 384 samples. These PCR reactions were performed in a reaction volume of 50 µl, containing 5 µl of each index primer, 10 µl of Fidelity buffer, 1.5 µl of dNTP, 23 µl of pure water (GE Healthcare—Hyclone), 0.5 µl of KAPA HiFi Hotstart Dna polymerase (KAPA Biosystems) and 5 µl of purified DNA product. Thermocycling was carried out on the same thermocycler with an initial denaturation step of 3 min à 95° followed by 8 cycles of 30 s denaturation at 95 °C, 30 s annealing at 55 °C and 30 s elongation at 72 °C finished with a 5-min elongation at

72 °C and stopped by cooling down at 4 °C. A second purification was conducted using 45 µl of PCR2 product and 50 µl of Agencourt AMPure XP beads following the manufacturer protocol. Purified products were quantified using Quant-iT^{TM} PicoGreen$^{®}$ dsDNA Assay Kit according to the manufacturer protocol and a fluorimeter (FilterMax F3, Molecular Devices). Following this step quantified product have been pooled in equimolarity before being sent to the GIGA Genomics platform (ULg) for sequencing using an Illumina MiSeq V2 benchtop sequencer. Raw sequences were cleaned using a script consisting in a mix of Fastx-toolkit (http://hannonlab.cshl.edu/fastx_toolkit) and USEARCH (Edgar 2010) functions. Briefly, the paired-end reads presenting an overlap of at least 10 bases long with a maximum difference of 8% were joined. Primers were removed, each sequence were filtered so that only the sequences with at least 95% of the bases presenting a quality superior to Q30 were kept. We used once again the SESAME software to generate the individual genotypes.

Mitochondrial cytochrome b gene

In order to compare the genetic structure obtained with the selected MHC genes with those based on neutral markers, we selected a subset of 98 *P. leucopus* individuals distributed throughout the studied area and corresponding to putative refuge areas as well as to newly colonised regions. Indeed, such sampling size corresponds to those generally used for phylogeographic studies performed at this geographic scale (Michaux et al. 2005; Mouton et al. 2016). We sequenced the mitochondrial cytochrome *b* for all these samples (Table 1). Indeed, this gene is generally considered as a neutral marker, upon which selection should be weak or absent. Polymerase Chain Reaction (PCR) was performed using universal primers H6/L7 described by Kocher et al. (1989) in a reaction volume of 25 µl, each containing 10 µl of Multiplex PCR kit (Qiagen Inc., Hilden, Germany), 2.5 µl of each primer (initial concentration: 10 µM), 8 µl of pure grade water (GE Healthcare—Hyclone) and 2 µl of DNA. Thermocycling was carried out on a Mastercycler Gradient (Eppendorf) with the following cycling conditions: one activation step at 94 °C for 4 min followed by 40 cycles of denaturation at 94 °C for 30 s, annealing at 50 °C for 30 s, elongation at 72 °C for 1 min 30 s, and a final extension at 72 °C for 10 min. PCR products were then sent to the MACROGEN society for Sanger sequencing.

Parasite screening

Guts from a subset of 165 mice belonging to populations sampled on the northern and southern shores of the St Lawrence river in southern Quebec were screened for helminth parasites using a binocular microscope. These samples are representative of the entire range in Québec and sampling size corresponds to those generally used in parasitology studies (Froeschke and Sommer 2005; Harf and Sommer 2005; Schad et al. 2005). Unfortunately, guts were not available for the American samples. All helminths were identified by their morphology and morphometry according to the literature. Parasites from the same species were grouped together and counted. The infection rate was computed as the percentage of hosts infected by at least one parasite.

Genetic analyses

Mitochondrial Cyt*b* sequences were aligned using the Sequence Scanner software (V1.0; Applied-Biosystems^{TM}). Ambiguous base pairs were removed at the beginning and the end of the sequences. Mega6 version 6.06 (Tamura et al. 2013) was used to align the resulting sequences and find the best nucleotide substitution model. We used Mega6 and Mr Bayes to draw phylogenetic trees based on maximum likelihood (1000 bootstrap resampling) and Bayesian inferences. Two cytochrome *b* sequences of *P. maniculatus* sequences from NCBI (accession numbers DQ385642 and DQ385706) were used as outgroups.

DRB sequences were aligned using Mega version 6.06 (Tamura et al. 2013). The STRUCTURE software version 2.3.4 (Pritchard et al. 2000) was used to identify the number of putative genetic clusters (K) within our 27 sampling sites. A range of $K = 1$ to $K = 10$ with 10 replicates for each K was assessed with a length of burnin period of 100,000 and a number of MCMC reps after burnin of 1,000,000. The number of genetic clusters was inferred using delta K (Evanno et al. 2005) in STRUCTURE HARVESTER (Earl and vonHoldt 2012), and individuals were assigned to a cluster if their probability of membership was higher than 0.6. This analysis was performed again separately on the two subgroups identified by the first Structure analysis to putatively detect some substructures. Rates of nonsynonymous (dN) and synonymous base pair substitutions (dS) were calculated with Mega version 6.06, using the model originally described by Nei and Gojobori (1986) with the Jukes and Cantor (1969) correction for multiple substitutions (Musolf et al. 2004; Froeschke and Sommer 2005). These rates were computed separately for the antigen binding sites (ABS) and for the non-antigen binding sites (non-ABS) as defined by Brown et al. (1993). The difference between dN and dS was assessed using a z-test calculated following the formula: $Z = (dN - dS)/SQRT(Var(dS) + Var(dN))$ (Tamura et al. 2013). To determine whether there is a relationship between heterozygosity and infection, we performed Chi square tests on individuals representing the entire range in Quebec.

Genetic diversity was computed separately for the Cyt*b* and DRB genes for four groups of populations corresponding to the range centre and to the range margin from the two lineages detected in the Cyt*b* phylogeny: GL (great lakes region population), NSSLR (populations from the North Shore of the Saint-Lawrence River), EC (populations from the American East Coast) and SSSLR (populations from the South Shore of the Saint-Lawrence River). For this analysis, the EC group was limited to populations EC2 to EC5. Indeed, mice from populations EC1 to EC8 have been captured over a 32 years period throughout a geographical range of 900 km. By limiting this group to EC2 to EC5, we decreased the time period to 2 years and the geographical gradient to 50 km (4 years and 175 km for SSSLR). By doing so, we avoid comparing diversity data that would have been artificially increased owing to large sampling differences. DRB and Cyt*b* genetic diversities were measured using the haplotype (h) diversity, computed using the DNASP software (Librado and Rozas 2009). Differences in diversity between core and marginal populations were assessed separately for the two lineages defined by Structure and for the two genes using unpaired Student t-tests.

Results

Cytochrome *b* genetic structure

The Cyt*b* gene from a subset of 98 specimens from 18 localities distributed throughout the studied area was successfully sequenced and 37 different haplotypes were detected in our samples (Table S1). Both the Bayesian Information and the corrected Akaike Information Criterions, selected the Hasegawa-Kishino-Yano plus Gamma model as the best model for nucleotide substitution. We obtained the same global structure using Maximum-likelihood and Bayesian inference methods, with two distinct genetic lineages, with a high level of robustness (ML posterior probability/BI bootstrap value for the two clades: 63/0.75 and 90/1) (Fig. 1). The first lineage grouped populations from the NSSLR regions with the Great lakes population; the second one grouped the SSSLR and the EC populations together (Fig. 2). No further structure was detected within each of these two lineages.

DRB genetic structure

The eight samples processed using both Illumina and 454 based protocols returned the exact same results, confirming the congruency between the two methods. The DRB gene exon 2 was successfully sequenced for 522 specimens in which we detected a total of 81 different alleles labelled following the nomenclature of Klein et al. (1990) from Pele-DRB*001 to Pele-DRB*081 (Table S1). No more than two alleles were found in any individual, suggesting that no duplication event occurred. The most common allele was found at a frequency of 9.5%.

STRUCTURE analysis identified two main clusters among our 27 populations. The first cluster was formed by the NSSLR, GL and EC populations (hereafter DRB-1 cluster), while the second one grouped all 15 SSSLR populations together (DRB-2 cluster) (Fig. 3). Sub-structuring was present within the DRB-1 cluster with the presence of two sub-clusters. The first one (DRB-1a) aggregated the GL and EC populations, while the second cluster (DRB-1b) grouped the three NSSLR populations. Likewise, Structure identified three sub-clusters within the DRB-2 cluster. The first one (DRB-2a) aggregated all populations west of the Richelieu River while both the second and the third clusters (DRB-2b and 2c) comprised populations from the Eastern side of this river (Fig. 3).

Helminth diversity

We detected 9 different helminth species in the 165 *P. leucopus* samples from the 9 analysed localities in southern Quebec (four species of trematode, one nematode cyst, one cestode, one Acantocephala, one *Syphacia* sp., and one *Aspiculuris* sp.). The most common parasite was the nematode *Syphacia* sp., which was present in nearly 20% of the screened mice. The remaining species were present in 1–10% of the individuals. In total, 37% of the screened mice were positive for at least one species and 10% of the sampled individuals presented a co-infection with more than one species. The infection rate in each population ranged from 15.8% in ER 10 to 63.2% in N1.

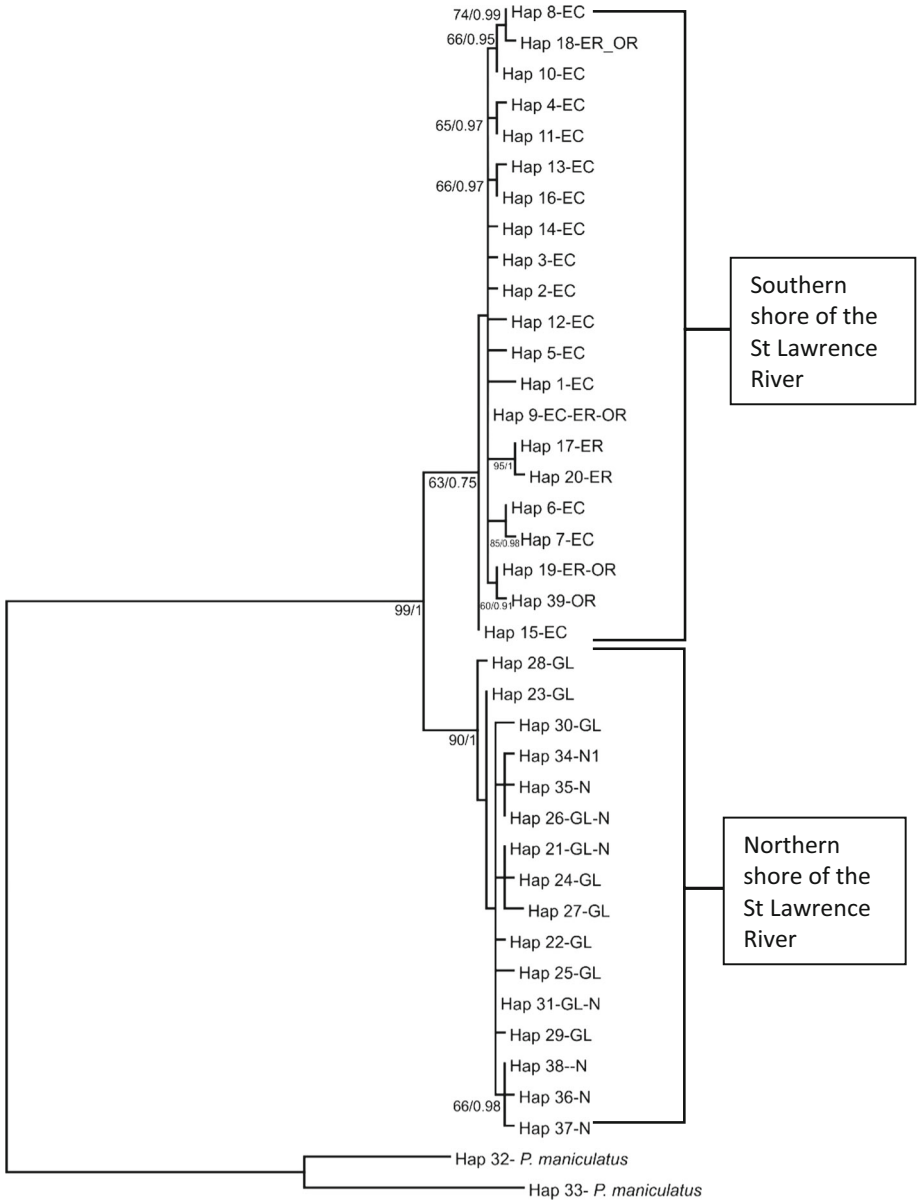


Fig. 1 Phylogenetic relationships based on Maximum likelihood (ML) and Bayesian inference (BI) among mitochondrial gene cytochrome b sequences. Values at branch nodes correspond to bootstrap support (%, 1000 pseudoreplicates) obtained with the ML analyses and to posterior probabilities obtained with the BI analyses. Haplotypes are named by a number followed by the site where they are present. Abbreviations for sites are as in Table 1

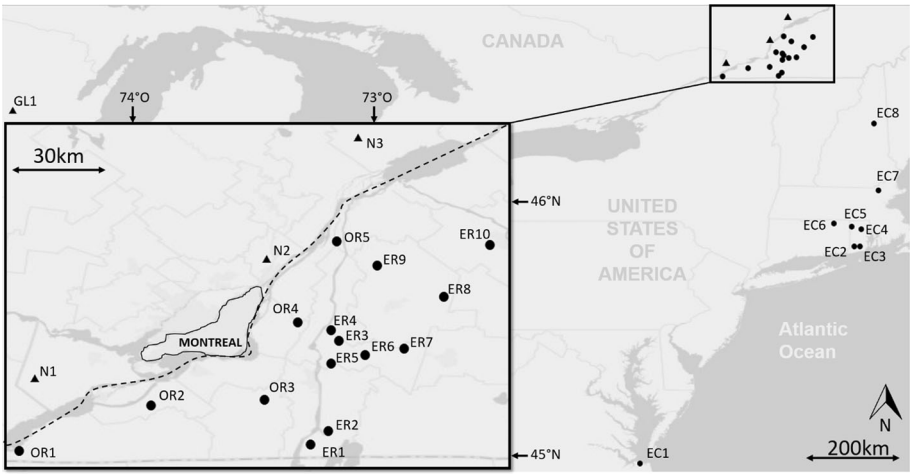


Fig. 2 Pattern of genetic diversity in the Cyt*b* sequence, characterized by two distinct lineages separating sites on the northern shore of the St Lawrence river (*triangles*) from sites on the southern shore (*circles*). The Saint-Lawrence river is represented by the dashed line. Abbreviations for sites are as in Table 1

Signature of selection acting on DRB

The 13 amino acids in the ABS (antigen binding sites) region were all polymorphic, whereas only 13 amino-acids out of 42 (31%) were variable in the non-ABS region. The rate of non-synonymous substitutions (*dN*) was 2.82 times higher than the rate of synonymous substitutions (*dS*) in the ABS region ($z = 5.307$; $p < 0.001$). In the non-ABS region, the rate of non-synonymous substitutions (*dN*) did not differ from the rate of synonymous substitutions (*dS*) ($z = 1.33$; $p = 0.186$; Table 2).

The proportion of individuals infected with helminths did not differ between individuals with a DRB homozygote genotype and individuals with a DRB heterozygote genotype (36.6%; N = 41 and 35.5%; N = 124, respectively; $\chi^2 = 0.0163$; $p = 0.90$). The 104 individuals carrying one or two of the five most common alleles had a mean infection rate of 26.9%. This is significantly lower than the infection rate from the 100 individuals carrying one or two of the 23 rarest alleles (mean infection rate of 44%; $\chi^2 = 6.5$, $p = 0.01$).

Within the northern lineage, haplotype diversity for the Cyt*b* and DRB was larger in the core population GL than in the marginal populations NSSLR ($t = 6.43$; $p < 0.0001$ and $t = 14.54$; $p < 0.0001$, respectively; Table 3). A similar pattern was apparent for the southern lineage when comparing haplotype diversity between the core populations EC and the marginal populations SSSLR ($t = 13.77$; $p < 0.0001$ and $t = 29.14$; $p < 0.0001$ for the Cyt*b* and DRB, respectively; Table 3).

Discussion

Here, we investigated the genetic structure of north-eastern American populations of *P. leucopus* using two genetic markers: one considered as neutral, the mitochondrial cytochrome *b;* the other considered as under selection, the Major Histocompatibility Complex

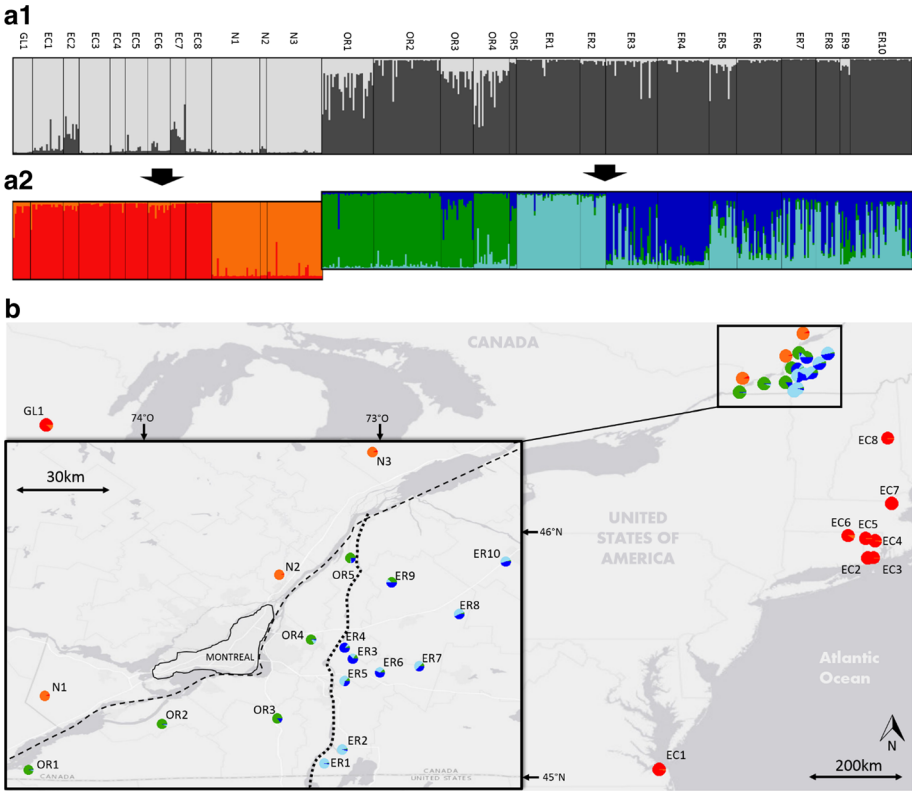


Fig. 3 Patterns of genetic diversity in the DRB sequences. **a1** Structure analysis on all populations showing two distinct clusters; *light gray*: populations from the NSSLR, GL and EC regions, *dark grey*: populations from the SSSLR region. **a2** Further sub-structure is observed within these two intial clusters. **b** Geographical distribution of the 5 genetic sub-clusters. The Saint-Lawrence and Richelieu rivers are represented by the *dashed and dotted lines*, respectively

Table 2 Estimated Rate (±SE) of non-synonymous (dN) and synonymous (dS) substitutions for Antigen-Binding Sites (ABS) and Non–Antigen-Binding-Sites (non-ABS) for the DRB gene; N: number of codons, *p*: probability that dN and dS are different using a z-test, where z = (dN − dS)/SQRT(Var(dS) + Var(dN))

Position	N	dN	dS	dN/dS	*p*
ABS	13	0.482 ± 0.089	0.171 ± 0.088	2.82	0.001
Non-ABS	43	0.1 ± 0.027	0.087 ± 0.026	1.15	0.186
ALL	56	0.173 ± 0.032	0.105 ± 0.027	1.65	0.001

DRB exon2 genes. We found that the spatial genetic structure was not congruent in these two genes, which may be the result of parasite-driven selection acting upon the DRB gene. In both genes, however, we detected a decrease in genetic diversity from the most central populations towards the species distribution margin.

Table 3 Haplotype (h) diversity (±SD) within the two lineages for the Cyt*b* and the DRB genes; N: sample size; N(h): haplotype number

Lineage	Population	Cyt B			DRB		
		N	N(h)	h	N	N(h)	h
Northern	GL	16	12	0.942 ± 0.048	11	15	0.961 ± 0.024
	NSSLR	19	9	0.848 ± 0.059	64	20	0.9 ± 0.01
Southern	EC	24	15	0.891 ± 0.057	49	38	0.957 ± 0.01
	SSSLR	35	6	0.637 ± 0.077	343	37	0.937 ± 0.003

Population genetic structure

The mitochondrial cytochrome *b* sequences revealed two distinct genetic lineages of *P. leucopus* in our study area. The first lineage grouped all individuals from the northern shore of the St Lawrence river in Quebec with the population from the Great Lakes region. The second lineage comprised all populations from the southern shore of the St Lawrence river in Quebec and populations from the East Coast region.

This result confirms at a highly larger geographic scale, the pattern previously found using a combination of five mitochondrial and nuclear sequences, including the Cyt*b* from populations from southern Quebec, where the St-Lawrence River acted as a barrier between two distinct genetic lineages (Fiset et al. 2015). Three major clusters for *P. leucopus* were also evidenced across the United States based on mitochondrial DNA markers (Rowe et al. 2006): A North-Eastern, a Central Western and a Western lineage. Our larger sampling of cytochrome b sequences, including now populations from south of Quebec, provides new evidence for the existence of two distinct glacial refugia in the Northern USA probably associated with two distinct post-glacial recolonization routes into Southern Quebec, as hypothesized in Fiset et al. (2015). Under this scenario, *P. leucopus* populations currently occurring on the northern shore of the St Lawrence river would be the result of a West-East recolonization originating from refuge populations in the Great Lakes region, while populations occurring in southern Quebec, south of the St Lawrence river would have originated from the East Coast region of the United States.

Our results obtained with the DRB marker differ markedly from those obtained with the Cyt*b*. With the DRB gene, we still detected two distinct clusters, but surprisingly populations from the East Coast appeared closer to populations from the northern shore of the St Lawrence river and the population from the Great Lakes region, than to populations from the southern shore of the St Lawrence river. Further sub-structure was detected among the populations south of the St Lawrence river, with an effect of the Richelieu river, as previously shown using neutral microsatellite markers (Rogic et al. 2013; Marrotte et al. 2014) and based on the morphological variation in skull shape (Ledevin and Millien 2013).

The discrepancy between the two genetic markers we studied may be due to differing selection processes acting upon the two genes. The mitochondrial cytochrome *b* gene is generally considered as a neutral marker, upon which selection should be weak or absent. This gene is thus expected to reflect the phylogeographical structure of the species, rather than local adaptation. By contrast, the DRB gene is generally considered to be under selection, particularly for functions related to the immune system and resistance to pathogens and parasites. The geographic structure obtained with this last marker may

therefore reflect different selection pressures operated by pathogens, and in the case of the DRB gene, particularly by macroparasites such as helminths.

Selection process on the DRB gene

At the sequence level, the rate of nonsynonymous substitutions (dN) was 2.82 times higher than the rate of synonymous substitutions (dS) in the ABS regions, while these substitution rates did not differ in the non-ABS region. This result provides evidence for a general positive selection having acted on the DRB gene over historical scales and at a large geographic scale. This would partly explain the discrepancy observed between Cyt*b* and DRB phylogenies. However, other processes would also have happened at a more contemporary scale and would have increased selection pressures on the DRB marker. For example, many studies evidenced, signatures of contemporary selection acting on immune system genes. The high degree of diversity in MHC genes has often been explained by balancing selection processes acted by parasites and pathogens (Quinnell et al. 2003; Harf and Sommer 2005; Schad et al. 2005; Zhang and He 2013). This mode of selection usually results from two main mechanisms, the heterozygous advantage and the rare allele advantage (or negative frequency dependent selection). Under the heterozygous advantage hypothesis, heterozygous animals should be able to recognize a broader range of pathogens, presenting thereby lower infection rates. Our data however did not support this hypothesis with respect to helminth resistance, as infection rate was identical between homozygous and heterozygous individuals. Although this hypothesis has received empirical support for MHC genes (Penn et al. 2002; Froeschke and Sommer 2005; Worley et al. 2010) and for other genes such as the cystic fibrosis transmembrane regulator (CFTR) (Schroeder et al. 1995; Common et al. 2004), our results are similar to those obtained by Meyer-Lucht and Sommer (2005) and by Schad et al. (2005) who did not detect any heterozygous advantage in their study on yellow-necked mice (*Apodemus flavicollis)* and Malagasy mouse lemurs (*Microcebus murinus*) and their respective nematodes. Another mechanism for balancing selection is the rare allele advantage hypothesis. It assumes that the evolution of new parasite and pathogen phenotypes reduces the fitness of common host genotypes, providing a selective advantage to rare genotypes. In other words, alleles that provide better immunity against parasites increase in frequency within a population (Parham and Ohta 1996), until parasites get adapted to those specific alleles, which may become a disadvantage for the host. Our data did not support this hypothesis either, as we found that common alleles were associated with lower infection rates than rare alleles, suggesting a positive selection. Again, this result is quite unusual, as most similar studies found evidence of negative frequency dependent selection (Quinnell et al. 2003; Harf and Sommer 2005; Schad et al. 2005; Zhang and He 2013). This difference may be due to the source of data we used, considering that the populations we sampled for parasites were all from the northern edge of the distribution range of *P. leucopus*, which is known to have only recently colonized this region (Roy-Dufresne et al. 2013). While parasites move along during range shift of their hosts, a time lag may be observed between parasites and their hosts (Phillips et al. 2010). Thus, even though parasites were detected in our sampled populations, they may not have adapted yet to the most common alleles, a process expected to occur locally with a delay, following the establishment of populations of hosts in newly colonized areas.

In summary, while we found a clear signal for historical positive selection on the DRB gene using the rates of nonsynonymous and synonymous substitutions, a more contemporary process, the positive selection, may also explain the discrepancy between the results we obtained with the Cyt*b* and the DRB.

Genetic diversity and range expansion

The central-marginal hypothesis states that for a given species, population genetic diversity decreases from the centre to the edge of their distribution range. Even though this hypothesis does not constitute a general rule, as suggested by Eckert et al. (2008) and by Pironon et al. (2016), this pattern was detected in several species including *Podarcis muralis* (Gassert et al. 2013), *Rana temporaria* (Johansson et al. 2006) or *Gypsophila fastigiata* (Lönn and Prentice 2002). In line with these results, the genetic diversity in both the Cyt*b* and DRB sequences for *P. leucopus* showed the expected decrease within both genetic lineages: from the Great Lakes to the northern shore of the St Lawrence, and from the US East Coast to the southern Quebec region south of the St Lawrence river. In eastern-North America, the species is known to expand its range at a rate of 10–15 km per year (Roy-Dufresne et al. 2013; Myers et al. 2009). Our data confirms that this expansion is associated with a loss of neutral as well as selected genetic diversity and may be explained by neutral processes such as genetic drift as observed in (Zeisset and Beebee 2014). The detected loss of DRB genetic diversity may furthermore be explained by an additional process, the positive selection that we observed on the margin of the species distribution. Indeed, in our data, common alleles were correlated to lower infection rate when compared to rare alleles, which may grant to their host a selective advantage, favouring the dominance of these common alleles in the next generations, hence decreasing the DRB genetic diversity.

Such a loss in genetic diversity at the range margin may potentially slow down the ongoing expansion of *P. leucopus*, by counterbalancing the positive effect of global warming on the mouse survival at higher latitude. However, field observations over the last decade, recording the presence of white footed mice at higher latitude year after year (Roy-Dufresne et al. 2013), tend to prove that global warming remain the major factor driving the expansion of the species. This warrants further research, as the rate and pattern of emergence of Lyme disease in Southern Quebec is dependent on the expansion rate of *P. leucopus* in the region (Simon et al. 2014).

Data accessibility

The Cyt*b* sequences have been deposited in Genbank under the Accession Numbers KX784130 to KX784166. Both Cyt*b* and DRB sequences can be found in the supplementary table (Table S1).

Acknowledgements We thank the Smithsonian National Museum of Natural History and the Harvard Museum of Natural History for the loan of some tissue samples. We thank S. Leo, S. Turney, field assistants and land owners. Computational resources were provided by the CBGP HPC computational platform, and by the Consortium des Équipements de Calcul Intensif (CÉCI) funded by the Fonds de la Recherche Scientifique de Belgique (F.R.S.-FNRS) under Grant No. 2.5020.11. This work was supported by Belgian FRS-FNRS (Fonds de la Recherche Scientifique) fellowship to AA and to JRM and by financial grants from the Belgian FRS-FNRS (''credits pour brefs séjours à l'étranger'') to AA and JRM, and from the "Patrimoine de l'université de Liège" to AA. VM is supported by a NSERC DG Grant (RGPIN/341918-2012).

References

Adams RI, Hadly EA (2012) Genetic diversity within vertebrate species is greater at lower latitudes. Evol Ecol 27:133–143

Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL, Wiley DC (1993) Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. Nature 364(6432):33–39
Brussard PF (1984) Geographic patterns and environmental gradients: the central-marginal model in Drosophila revisited. Annu Rev Ecol Syst 15:25–64
Chen W, Bei Y, Li H (2015) Genetic variation of the major histocompatibility complex (MHC class II B gene) in the threatened Hume's pheasant, *Syrmaticus humiae*. PLoS ONE 10:e0116499
Common J, Di W, Davies D, Kelsell D (2004) Further evidence for heterozygote advantage of GJB2 deafness mutations: a link with cell survival. J Med Genet 41:573–575
Doherty PC, Zinkernagel RM (1975) Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. Nature 256:50–52
Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361
Eckert CG, Samis KE, Lougheed SC (2008) Genetic variation across species' geographical ranges: the central-marginal hypothesis and beyond. Mol Ecol 17:1170–1188
Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461
Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620
Excoffier L, Foll M, Petit RJ (2009) Genetic consequences of range expansions. Annu Rev Ecol Evol Syst 40:481–501
Fiset J, Tessier N, Millien V, Lapointe FJ (2015) Phylogeographic structure of the white-footed mouse and the deer mouse, two lyme disease reservoir hosts in Quebec. PLoS ONE 10:e0144112
Froeschke G, Sommer S (2005) MHC class II DRB variability and parasite load in the striped mouse (*Rhabdomys pumilio*) in the Southern Kalahari. Mol Biol Evol 22:1254–1259
Gaitan J, Millien V (2016) Stress level, parasite load, and movement pattern in a small mammal reservoir host for Lyme disease. Can J Zool 94:565–573
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 Genom 11(1):296
Gassert F, Schulte U, Husemann M, Ulrich W, Rödder D, Hochkirch A, Engel E, Meyer J, Habel JC, Parmakelis A (2013) From southern refugia to the northern range margin: genetic population structure of the common wall lizard, Podarcis muralis. J Biogeogr 40(8):1475–1489
Harf R, Sommer S (2005) Association between major histocompatibility complex class II DRB alleles and parasite load in the hairy-footed gerbil, *Gerbillurus paeba*, in the southern Kalahari. Mol Ecol 14:85–91
Hedrick PW (2012) What is the evidence for heterozygote advantage selection? Trends Ecol Evol 27:698–704
Hewitt GM (1999) Post-glacial re-colonization of European biota. Biol J Linn Soc 68:87–112
Hoffmann AA, Blows MW (1994) Species borders: ecological and evolutionary perspectives. Trends Ecol Evol 9:223–227
Hughes AL, Nei M (1989) Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. Proc Natl Acad Sci USA 86:958–962
Johansson M, Primmer CR, Merila J (2006) History vs. current demography: explaining the genetic population structure of the common frog (Rana temporaria). Mol Ecol 15:975–983
Jukes TH, Cantor CR (1969) Evolution of protein molecules. Academy Press, New York
Klein J (1987) Origin of major histocompatibility complex polymorphism: the trans-species hypothesis. Hum Immunol 19:155–162
Klein J, Bontrop RE, Dawkins RL, Erlich HA, Gyllensten UB, Heise ER, Jones PP, Parham P, Wakeland EK, Watkins DI (1990) Nomenclature for the major histocompatibility complexes of different species: a proposal. Immunogenetics 31:217–219
Kocher TD, Thomas WK, Meyer A, Edwards SV, Paabo S, Villablanca FX, Wilson AC (1989) Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc Natl Acad Sci USA 86:6196–6200
Kohyama TI, Omote K, Nishida C, Takenaka T, Saito K, Fujimoto S, Masuda R (2015) Spatial and temporal variation at major histocompatibility complex class IIB genes in the endangered Blakiston's fish owl. Zool Lett 1:13
Lau Q, Jaratlerdsiri W, Griffith JE, Gongora J, Higgins DP (2014) MHC class II diversity of koala (*Phascolarctos cinereus*) populations across their range. Heredity 113:287–296
Ledevin R, Millien V (2013) Congruent morphological and genetic differentiation as a signature of range expansion in a fragmented landscape. Ecol Evol 3:4172–4182

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452

Lönn M, Prentice HC (2002) Gene diversity and demographic turnover in central and peripheral populations of the perennial herb *Gypsophila fastigiata*. Oikos 99:489–498

Marrotte RR, Gonzalez A, Millien V (2014) Landscape resistance and habitat combine to provide an optimal model of genetic structure and connectivity at the range margin of a small mammal. Mol Ecol 23:3983–3998

Mather TN (1993) The dynamics of spirochete transmission between ticks and vertebrates. Ecology and environmental management of Lyme disease. Rutgers University Press, New Brunswick, pp 43–62

Meglecz E, Piry S, Desmarais E, Galan M, Gilles A, Guivier E, Pech N, Martin JF (2011) SESAME (SEquence Sorter & AMplicon Explorer): genotyping based on high-throughput multiplex amplicon sequencing. Bioinformatics 27:277–278

Meyer-Lucht Y, Sommer S (2005) MHC diversity and the association to nematode parasitism in the yellow-necked mouse (*Apodemus flavicollis*). Mol Ecol 14:2233–2243

Michaux JR, Magnanou E, Paradis E, Nieberding C, Libois R (2003) Mitochondrial phylogeography of the woodmouse (*Apodemus sylvaticus*) in the Western Palearctic region. Mol Ecol 12:685–697

Michaux JR, Libois R, Filippucci MG (2005) So close and so different:comparative phylogeography of two small mammal species, theYellownecked fieldmouse (*Apodemus flavicollis*) and the Wood-mouse(*Apodemus sylvaticus*), in the Western Palearctic region. Heredity 94:52–63

Mouton A, Mortelliti A, Grill A, Sara M, Kryštufek B, Juškaitis R, Latinne A, Amori G, Randi E, Büchner S, Schulz B, Ehlers S, Lang J, Adamik P, Verbeylen G, Dorenbosch M, Trout R, Elmeros M, Aloise G, Mazzoti S, Matur F, Poitevin F, Michaux JR (2017) Evolutionary history and species delimitations: a case study of the hazel dormouse, Muscardinus avellanarius. Conserv Genet 18(1):181–196

Musolf K, Meyer-Lucht Y, Sommer S (2004) Evolution of MHC-DRB class II polymorphism in the genus Apodemus and a comparison of DRB sequences within the family Muridae (Mammalia: Rodentia). Immunogenetics 56:420–426

Myers P, Lundrigan BL, Hoffman SMG, Haraminac AP, Seto SH (2009) Climate-induced changes in the small mammal communities of the Northern Great Lakes Region. Glob Change Biol 15:1434–1454

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426

Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. Evolution 29:1–10

Oliver MK, Lambin X, Cornulier T, Piertney SB (2009) Spatio-temporal variation in the strength and mode of selection acting on major histocompatibility complex diversity in water vole (*Arvicola terrestris*) metapopulations. Mol Ecol 18:80–92

Ostfeld RS (2011) Lyme disease: the ecology of a complex system. Oxford University Press, Oxford, p 216

Parham P, Ohta T (1996) Population biology of antigen presentation by MHC class I molecules. Science 272:67–74

Penn DJ, Damjanovich K, Potts WK (2002) MHC heterozygosity confers a selective advantage against multiple-strain infections. Proc Natl Acad Sci USA 99:11260–11264

Phillips BL, Kelehear C, Pizzatto L, Brown GP, Barton D, Shine R (2010) Parasites and pathogens lag behind their host during periods of host range advance. Ecology 91:872–881

Pironon S, Papuga G, Villellas J, Angert AL, García MB, Thompson JD (2016) Geographic variation in genetic and demographic performance: new insights from an old biogeographical paradigm. Biol Rev. doi:10.1111/brv.12313

Prakash S, Lewontin RC, Hubby JL (1969) A molecular approach to the study of genic heterozygosity in natural populations. IV. Patterns of genic variation in central, marginal and isolated populations of *Drosophila pseudoobscura*. Genetics 61:841–858

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155(2):945–959

Quinnell RJ, Kennedy LJ, Barnes A, Courtenay O, Dye C, Garcez LM, Shaw MA, Carter SD, Thomson W, Ollier WE (2003) Susceptibility to visceral leishmaniasis in the domestic dog is associated with MHC class II polymorphism. Immunogenetics 55:23–28

Rogic A, Tessier N, Legendre P, Lapointe FJ, Millien V (2013) Genetic structure of the white-footed mouse in the context of the emergence of Lyme disease in southern Quebec. Ecol Evol 3:2075–2088

Rowe KC, Heske EJ, Paige KN (2006) Comparative phylogeography of eastern chipmunks and white-footed mice in relation to the individualistic nature of species. Mol Ecol 15:4003–4020

Roy-Dufresne E, Logan T, Simon JA, Chmura GL, Millien V (2013) Poleward expansion of the white-footed mouse *Peromyscus leucopus* under climate change: implications for the spread of Lyme disease. PLoS ONE 8:e80724

Schad J, Ganzhorn JU, Sommer S (2005) Parasite burden and constitution of major histocompatibility complex in the Malagasy mouse lemur, *Microcebus murinus*. Evolution 59:439–450

Schroeder SA, Gaughan DM, Swift M (1995) Protection against bronchial asthma by CFTR [Delta]F508 mutation: a heterozygote advantage in cystic fibrosis. Nat Med 1:703–705

Simon JA, Marrotte RR, Desrosiers N, Fiset J, Gaitan J, Gonzalez A, Koffi JK, Lapointe F-J, Leighton PA, Lindsay LR, Logan T, Milord F, Ogden NH, Rogic A, Roy-Dufresne E, Suter D, Tessier N, Millien V (2014) Climate change and habitat fragmentation drive the occurrence of *Borrelia burgdorferi*, the agent of Lyme disease, at the northeastern limit of its distribution. Evol Appl 7:750–764

Sommer S (2005) The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Front Zool 2:16. doi:10.1186/1742-9994-2-16

Spurgin LG, Richardson DS (2010) How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. Proc R Soc Lond B Biol Sci 277(1684):979–988

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, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729

Turner AK, Begon M, Jackson JA, Bradley JE, Paterson S (2011) Genetic diversity in cytokines associated with immune variation and resistance to multiple pathogens in a natural rodent population. PLoS Genet 7:e1002343

Worley K, Collet J, Spurgin LG, Cornwallis C, Pizzari T, Richardson DS (2010) MHC heterozygosity and survival in red junglefowl. Mol Ecol 19:3064–3075

Zeisset I, Beebee TJ (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

Zhang M, He H (2013) Parasite-mediated selection of major histocompatibility complex variability in wild brandt's voles (*Lasiopodomys brandtii*) from Inner Mongolia, China. BMC Evol Biol 13:149