Distinct haplotype structure at the innate immune receptor Toll-like receptor 2 across bank vole populations and lineages in Europe

ARTICLE *in* BIOLOGICAL JOURNAL OF THE LINNEAN SOCIETY · JUNE 2015

Impact Factor: 2.54 · DOI: 10.1111/bij.12593

22 AUTHORS, INCLUDING:

Linas Balciauskas
Nature Research Centre
138 PUBLICATIONS **263** CITATIONS
SEE PROFILE

Nico W Van den Brink
Wageningen University
128 PUBLICATIONS **851** CITATIONS
SEE PROFILE

Roman Biek
University of Glasgow
46 PUBLICATIONS **1,212** CITATIONS
SEE PROFILE

The LINNEAN SOCIETY of London BIOLOGICAL Journal of the Linnean Society

Biological Journal of the Linnean Society, 2015, ••, ••–••. With 2 figures.

Distinct haplotype structure at the innate immune receptor Toll-like receptor 2 across bank vole populations and lineages in Europe

JENNIFER MORGER[1], LARS RÅBERG[2], SABINE M. HILLE[3], SANNE HELSEN[4], JAN ŠTEFKA[5], MOHAMMAD M. AL-SABI[6], CHRISTIAN M. O. KAPEL[6], TAPIO MAPPES[7], SANDRA ESSBAUER[8], RAINER G. ULRICH[9], PAOLA BARTOLOMMEI[10], ALESSIO MORTELLITI[11,12], LINAS BALČIAUSKAS[13], NICO W. VAN DEN BRINK[14,15], ALICE RÉMY[16], ANNA BAJER[17], MIHAIL CHEPRAKOV[18], MISA KORVA[19], ANA L. GARCÍA-PÉREZ[20], ROMAN BIEK[21], SUSAN WITHENSHAW[22] and BARBARA TSCHIRREN[1]*

[1]*Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland*
[2]*Department of Biology, Lund University, Sölvegatan 35, 223 62 Lund, Sweden*
[3]*Institute of Wildlife Biology and Game Management, University of Natural Resources and Life Sciences, Gregor Mendel-Strasse 33, 1180 Vienna, Austria*
[4]*Evolutionary Ecology Group, Department of Biology, University of Antwerp, Groenenborgerlaan 171, 2020 Antwerp, Belgium*
[5]*Faculty of Science, Biology Centre ASCR, Institute of Parasitology and University of South Bohemia, Branišovská 31, 37005 České Budějovice, Czech Republic*
[6]*Department of Plant and Environmental Sciences, Faculty of Science, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark*
[7]*Department of Biological and Environmental Science, University of Jyväskylä, PO Box 35, 40014 Jyväskylä, Finland*
[8]*Department Virology and Rickettsiology, Bundeswehr Institute of Microbiology, Neuherbergstrasse 11, 80937 Munich, Germany*
[9]*Friedrich-Loeffler-Institut, Institute for Novel and Emerging Infectious Diseases, Federal Research Institute for Animal Health, Südufer 10, 17493 Greifswald – Insel Riems, Germany*
[10]*Fondazione Ethoikos, Convento dell'Osservanza, Radicondoli, 53030 Siena, Italy*
[11]*Department of Biology and Biotechnology 'Charles Darwin', University of Rome 'La Sapienza', Viale dell'Università 32, 00185 Rome, Italy*
[12]*National Environmental Research Program, Fenner School of Environment and Society, Australian Research Council Centre for Environmental Decisions, The Australian National University, Canberra, ACT, 0200 Australia*
[13]*Nature Research Centre, Akademijos 2, 08412 Vilnius, Lithuania*
[14]*Alterra, Wageningen UR, PO-Box 47, 6700 AA Wageningen, the Netherlands*
[15]*Sub-Department of Toxicology, Wageningen University Wageningen UR, PO-Box 8000, 6700 EA Wageningen, the Netherlands*
[16]*Faculty of Applied Ecology and Agricultural Sciences, Hedmark University College, Anne Evenstadsvei 80, 2480 Koppang, Norway*
[17]*Department of Parasitology, Faculty of Biology, Institute of Zoology, University of Warsaw, 1 Miecznikowa Street, 02-096 Warsaw, Poland*
[18]*Ural Branch, Institute of Plant and Animal Ecology, Russian Academy of Sciences, Str 8 Marta 202, Yekaterinburg, 620144 Russia*
[19]*Faculty of Medicine, Institute of Microbiology and Immunology, Zaloška 4, 1000 Ljubljana, Slovenia*
[20]*Department of Animal Health, NEIKER – Instituto Vasco de Investigación y Desarrollo Agrario, Berreaga 1, 48160 Derio, Bizkaia, Spain*

[21]*College of Medical Veterinary and Life Sciences, Boyd Orr Centre for Population and Ecosystem Health, Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Glasgow G12 8QQ, UK*
[22]*Department of Evolution, Ecology and Behaviour, Institute of Integrative Biology, University of Liverpool, Crown Street, Liverpool L69 7ZB, UK*

Received 10 March 2015; revised 3 April 2015; accepted for publication 3 April 2015

Parasite-mediated selection may contribute to the maintenance of genetic variation at host immune genes over long time scales. To date, the best evidence for the long-term maintenance of immunogenetic variation in natural populations comes from studies on the major histocompatibility complex (*MHC*) genes, whereas evidence for such processes from other immune genes remains scarce. In the present study, we show that, despite pronounced population differentiation and the occurrence of numerous private alleles within populations, the innate immune gene Toll-like receptor 2 (*TLR2*) displays a distinct haplotype structure in 21 bank vole (*Myodes glareolus*) populations across Europe. Haplotypes from all populations grouped in four clearly differentiated clusters, with the three main clusters co-occurring in at least three previously described mitochondrial lineages. This pattern indicates that the distinct *TLR2* haplotype structure may precede the split of the mitochondrial lineages 0.19–0.56 Mya and suggests that haplotype clusters at this innate immune receptor are maintained over prolonged time in wild bank vole populations. *Biological Journal of the Linnean Society*, 2015, **00**, 000–000.

ADDITIONAL KEYWORDS: balancing selection – diversity – gene evolution – immunogenetics – maintenance of genetic variation – *Myodes glareolus* – parasite resistance – phylogeography – rodents – *TLR2*.

INTRODUCTION

Understanding how genetic variation is maintained in natural populations is a fundamental quest in evolutionary biology. Immune genes frequently display particularly high diversity, presumably as a result of parasite-mediated selection (Haldane, 1949). There are different, non-mutually exclusive processes by which parasites (i.e. parasitic and pathogenic organisms and agents) may contribute to the maintenance of genetic variation at host defence genes. First, parasites may specialize in the most common host genotype, leading to negative frequency-dependent selection (Takahata & Nei, 1990; Woolhouse *et al.*, 2002). Second, heterozygous individuals may have a selective advantage over homozygous individuals (i.e. overdominance); for example, as a result of enhanced parasite recognition ability or reduced immunopathology (Doherty & Zinkernagel, 1975; Apanius *et al.*, 1997; Khor *et al.*, 2007). Finally, variation in the composition of the parasite community in time and/or space may lead to fluctuating selection, and hence contribute to the maintenance of genetic variation in host populations (Hedrick, 2002).

To date, the best evidence for the maintenance of immunogenetic variation over long time scales in natural populations comes from studies on the major histocompatibility complex (*MHC*) genes (Hedrick, 1998; Bernatchez & Landry, 2003; Piertney & Oliver, 2006). In humans, a number of other immune loci with signatures of long-term balancing selection have been identified (Bubb *et al.*, 2006; Andrés *et al.*, 2009). In other species, however, evidence for such patterns outside the MHC complex remains scarce (Ferguson *et al.*, 2012).

Toll-like receptors (TLRs) are pattern-recognition receptors that play a key role in the recognition of intra- and extracellular pathogens. They recognize essential structures of pathogens (pathogen-associated molecular patterns; PAMPs) and initiate innate and adaptive immune responses (Medzhitov, 2001; Akira, Uematsu & Takeuchi, 2006). Toll-like receptor 2 (TLR2), for example, plays an important role in the recognition of lipopeptides, crucial components of the cell membranes of Gram-positive bacteria (Jin *et al.*, 2007). Genetic polymorphisms at *TLRs* have been associated with variation in infectious disease susceptibility in humans and laboratory animals (Texereau *et al.*, 2005; Netea, Wijmenga & O'Neill, 2012) and there is accumulating evidence that they play an important role in mediating host–parasite interactions and disease susceptibility in wildlife (Jackson *et al.*, 2009; Turner *et al.*, 2011; Tschirren *et al.*, 2013; Fornůsková *et al.*, 2014; Morger *et al.*, 2014).

It has previously been reported that in a Swedish bank vole *Myodes glareolus* (Schreber, 1780; Rodentia, Cricetidae) population, *TLR2* displays a high diversity, with haplotypes grouping in three dis-

*Corresponding author: E-mail: barbara.tschirren@ieu.uzh.ch

tinct clusters (Tschirren *et al.*, 2013). Furthermore, voles carrying haplotypes of one particular cluster (c_2) had a reduced *Borrelia afzelii* infection rate, and an analysis of haplotype frequencies and tree topology strongly suggested that these 'protective' alleles had evolved under positive selection (Tschirren *et al.*, 2013). One interpretation of these patterns is that the high diversity at *TLR2* is the result of parasite-mediated selection. However, the distinct *TLR2* haplotype structure could also be a result of neutral processes, such as drift in isolated populations followed by population admixture. Indeed, eight well-defined mitochondrial bank vole lineages have been described in Europe (Deffontaine *et al.*, 2005, 2009; Kotlík *et al.*, 2006) and there is evidence that in Southern Fennoscandia, re-colonization after the last glaciation (approximately 11 000–10 000 years BP; Jaarola, Tegelström & Fredga, 1999) occurred from separate glacial refugia, creating contact zones between bank vole lineages that have evolved genetic differences while they were geographically isolated (Hewitt, 1999; Jaarola *et al.*, 1999).

In the present study, we sequenced *TLR2* from 21 bank vole populations across Europe, belonging to five previously described mitochondrial lineages, aiming to test whether the different *TLR2* clusters are associated with different bank vole lineages. If, on the other hand, the co-occurrence of the distinct *TLR2* clusters is a general feature of this innate immune gene and is observed in several lineages and populations on a large geographical scale, this would provide indirect evidence for the long-term maintenance of *TLR2* diversity.

MATERIAL AND METHODS

Study species and sampling sites

The bank vole is a common rodent in Europe. Its distribution ranges from Fennoscandia and the British Isles to northern Spain, Italy, and Russia (Corbet & Harris, 1991). For the present study, we caught 385 bank voles at 21 locations in 18 countries for tissue or hair collection (mean per location of 18.3 animals; range 10–21) (Table 1; see also Supporting information, Table S1). Trapping and tissue or hair sampling were performed in accordance with the regulations of the respective countries and performed under licenses provided by the local animal welfare committees (see Supporting information, Table S1). Distances between locations ranged from 69 to 4454 km.

TLR2 and cytochrome b (*CYTB*) genotyping

Total genomic DNA was extracted from the samples using a standard extraction method (Laird *et al.*, 1991). We sequenced a 1149-bp fragment of *TLR2* for all individuals (N = 385) (Table 1) in acordance with previously reported protocols (Tschirren *et al.*, 2012). The amplified fragment contains the functionally relevant sites involved in pathogen recognition and TLR heterodimerization (Jin *et al.*, 2007) (see Supporting information, Fig. S1), and we previously demonstrated molecular signatures of positive selection during the evolutionary history of rodents within this gene region (Tschirren, Råberg & Westerdahl, 2011). *TLR2* amplifications were performed in a total volume of 10 μL containing 0.2 μL of JumpStart Taq DNA Polmyerase (Sigma-Aldrich, Buchs, Switzerland), 300 M of the primers MglTLR2F: CAT CCATCACCTGACCCTTC and MglTLR2R: CCAGTA GGAATCCTGCTCG and 25 ng of genomic DNA.

To assign the different populations to the previously described mitochondrial bank vole lineages (Deffontaine *et al.*, 2005; Kotlík *et al.*, 2006), we sequenced a 1101-bp fragment of the mitochondrial *cytb* gene for three to five individuals per population *sensu* Kotlík *et al.* (2006).

Amplifications were performed in a total volume of 10 μL containing 0.2 μL of JumpStart Taq DNA Polymerase (Sigma-Aldrich), 300 M of each primer

Table 1. Toll-like receptor 2 (*TLR2*) gene diversity

Population	N	h	h_{priv}	A	p (ns/s)
Total	385	86	38	51	75 (34/41)
ESP	13	6	4	5	18 (9/9)
ITA1	19	9	2	4	23 (6/17)
ITA2	21	9	4	6	21 (8/13)
AUT	17	8	1	4	25 (9/16)
BEL	19	8	2	6	25 (12/13)
CZE	19	10	2	6	10 (5/5)
ENG	17	3	0	2	3 (1/2)
GER1	21	11	3	6	21 (11/10)
NED	10	5	1	3	6 (2/4)
SLO	20	11	3	8	29 (12/17)
SUI	19	6	1	5	24 (11/13)
DEN	19	5	0	3	15 (7/8)
FIN	20	9	2	7	28 (12/16)
GER2	18	7	2	5	25 (11/14)
POL	12	6	0	3	16 (7/9)
RUS	21	12	4	9	25 (13/12)
UKR	20	8	2	6	8 (6/2)
LTU	20	8	1	5	19 (9/10)
NOR	21	3	1	2	2 (1/1)
SCO	19	2	0	2	2 (1/1)
SWE	20	10	3	8	28 (13/15)

N, number of analyzed individuals; h, number of *TLR2* haplotypes; h_{priv}, number of private haplotypes; A, number of amino acid variants; p (ns/s), polymorphic sites (nonsynonymous substitutions/synonymous substitutions) in the study populations.

and 25 ng of genomic DNA. The polymerase chain reaction (PCR) protocol for *TLR2* amplification consisted of an initial denaturation step at 94 °C for 1 min, followed by 35 cycles of denaturation at 94 °C for 30 s, annealing at 57 °C for 30 s, and extension at 72 °C for 90 s, with a final elongation step at 72 °C for 10 min. For *cytb*, we used 40 cycles and an annealing temperature of 52 °C. PCR products were purified and sequenced in both directions on an ABI Prism 3730 capillary sequencer (Applied Biosystems, Waltham, MA, USA) using Big Dye terminator, version 3.1 (Applied Biosystems). All *TLR2* and *cytb* raw sequences were trimmed, processed and aligned in GENEIOUS, version 5.6.5 (Drummond *et al.*, 2009). Putative polymorphisms identified by the software were manually verified. *TLR2* haplotypes were reconstructed for each population separately in PHASE, version 2.1 (Stephens, Smith & Donnelly, 2001) using the default settings of a thinning interval of 1, 100 burn-in iterations and 100 main iterations, and allowing for recombination (see Supporting information, Table S2). If a haplotype was observed only once in the dataset, we repeated the PCR and sequencing reactions to confirm the sequence and avoid overestimation of diversity as a result of amplification or sequencing errors. *TLR2* and *cytb* sequences were submitted to GenBank (*TLR2* accession numbers: see Supporting information, Table S3; *cytb* accession numbers: KJ612463–KJ612512).

Statistical analysis

We quantified genetic diversity at *TLR2* in the 21 bank vole populations by calculating the number of haplotypes, the number of synonymous and nonsynonymous substitutions, and the number of amino acid variants using DNASP, version 5.10.01 (Librado & Rozas, 2009). To estimate population differentiation at *TLR2*, we calculated pairwise F_{ST} (Weir & Cockerham, 1984) in ARLEQUIN, version 3.5.1.3 (Excoffier & Lischer, 2010). Sequential Bonferroni corrections were applied to account for multiple testing. To test whether patterns of haplotype frequencies deviated from neutral expectations, we calculated Fay & Wu's *H* (Fay & Wu, 2000). The empirical distribution of the test statistics was generated using neutral coalescent simulations in DNASP (Librado & Rozas, 2009) based on the observed number of segregating sites, 20 000 replicates, and allowing for a recombination rate of 0.001. A *Mus musculus TLR2* sequence (NM_011905.3) was used as an outgroup.

We then reconstructed a *TLR2* haplotype genealogy in TCS, version 1.21 (Clement, Posada & Crandall, 2000) by taking synonymous and nonsynonymous substitutions into account. We used the resulting genealogy to assign haplotypes to *TLR2* haplogroups (or clusters, defined as groups of haplotypes that are separated from other groups of haplotypes by at least seven mutations) *sensu* Tschirren *et al.* (2013). The three 'intermediate' haplotypes in the Italian populations were separated by more than seven mutations from the other previously described clusters (c_1, c_2, c_3) (see Results) and were therefore treated as a forth cluster (c_4) in the analyses. We tested for deviations of *TLR2* cluster frequencies from Hardy–Weinberg equilibrium within populations in ARLEQUIN, version 3.5.1.2. (Excoffier & Lischer, 2010). To determine which of the previously described European bank vole lineages (Deffontaine *et al.*, 2005; Kotlík *et al.*, 2006) the different study populations belonged to, we compared the *cytb* sequences with 14 randomly chosen *cytb* sequences from the seven lineages obtained from GenBank (Basque: EF408066, EF408062; Spanish: EF408061, EF408060; Italian: AJ639664, AJ639692; Western: DQ472319, DQ472333; Basal: DQ472327, DQ472344; Eastern: DQ472347, DQ472346; Carpathian: DQ472253, DQ472339). We excluded the eighth lineage (Ural) because of its large genetic distance from all other lineages and because preliminary analyses showed that we had no representative of this lineage in our sample. A Neighbour-joining tree (maximum composite likelihood, 1000 bootstrap iterations) with the Northern red-backed vole, *Myodes rutilus* (Pallas, 1779; Rodentia, Cricetidae) (JF714850), as an outgroup was created in MEGA, version 6 (Tamura *et al.*, 2013) and used to assign the *cytb* sequences to one of the seven mitochondrial lineages. To further confirm the robustness of the assignment of the study populations to the different lineages, we reconstructed a *cytb* genealogy in TCS, version 1.21 (Clement *et al.*, 2000) by taking synonymous and nonsynonymous substitutions into account. Finally, we used a simulated annealing approach to identify groups that are maximally differentiated from each other without taking geographical constrains into account in SAMOVA, version 2.0 (Dupanloup, Schneider & Excoffier, 2002). This lineage classification was then included in an analysis of molecular variance (AMOVA) in ARLEQUIN, version 3.5.1.3 (Excoffier & Lischer, 2010), which partitions the total *TLR2* cluster variation into covariance fractions, resulting from the variance among mitochondrial lineages, among populations within lineage, and within populations. The number of permutations for this analysis was set to 1000.

RESULTS

Mitochondrial lineages

The different bank vole populations could be assigned to one of five previously described mito-

chondrial lineages (Spanish, Italian, Western, Eastern, and Carpathian) (Deffontaine *et al.*, 2005; Kotlík *et al.*, 2006) based on *cytb* sequencing (Fig. 1; see also Supporting information, Figs S2, S3).

Groups that were maximally genetically differentiated from each other were identified as (SAMOVA; $K = 5$): Group 1 = ESP; Group 2 = ITA1, ITA2; Group 3 = NOR, SCO, SWE, LTU; Group 4 = RUS, UKR, DEN, FIN, POL, GER2; Group 5 = AUT, SUI, ENG, SLO, BEL, NED, GER1, CZE, which is identical to the lineage assignment of the populations based on the phylogenetic tree (see Supporting information, Fig. S2).

TLR2 HAPLOTYPE DIVERSITY AND DIFFERENTIATION

Genetic diversity at *TLR2* was high across bank vole populations, with 86 haplotypes and 51 amino acid variants (Table 1). We observed a large number of private alleles within populations ($N = 67$) (Table 1), even among populations in close proximity; for example, the two Italian populations ITA1 and ITA2 (see Supporting information, Fig. S4). Furthermore, we observed significant *TLR2* population differentiation (F_{ST}) for most pairwise population comparisons (see Supporting information, Table S4). An excess of high-frequency derived haplotypes was observed in most populations, which is indicative of positive selection (see Supporting information, Table S5).

TLR2 CLUSTERS

TLR2 haplotypes from all European bank vole populations grouped in the same, distinct clusters (c_1, c_2, c_3, c_4) (Fig. 2; see also Supporting information, Fig. S5, Table S3). Despite the high *TLR2* diversity, a large number of private alleles, and strong population differentiation (see above), haplotypes of the

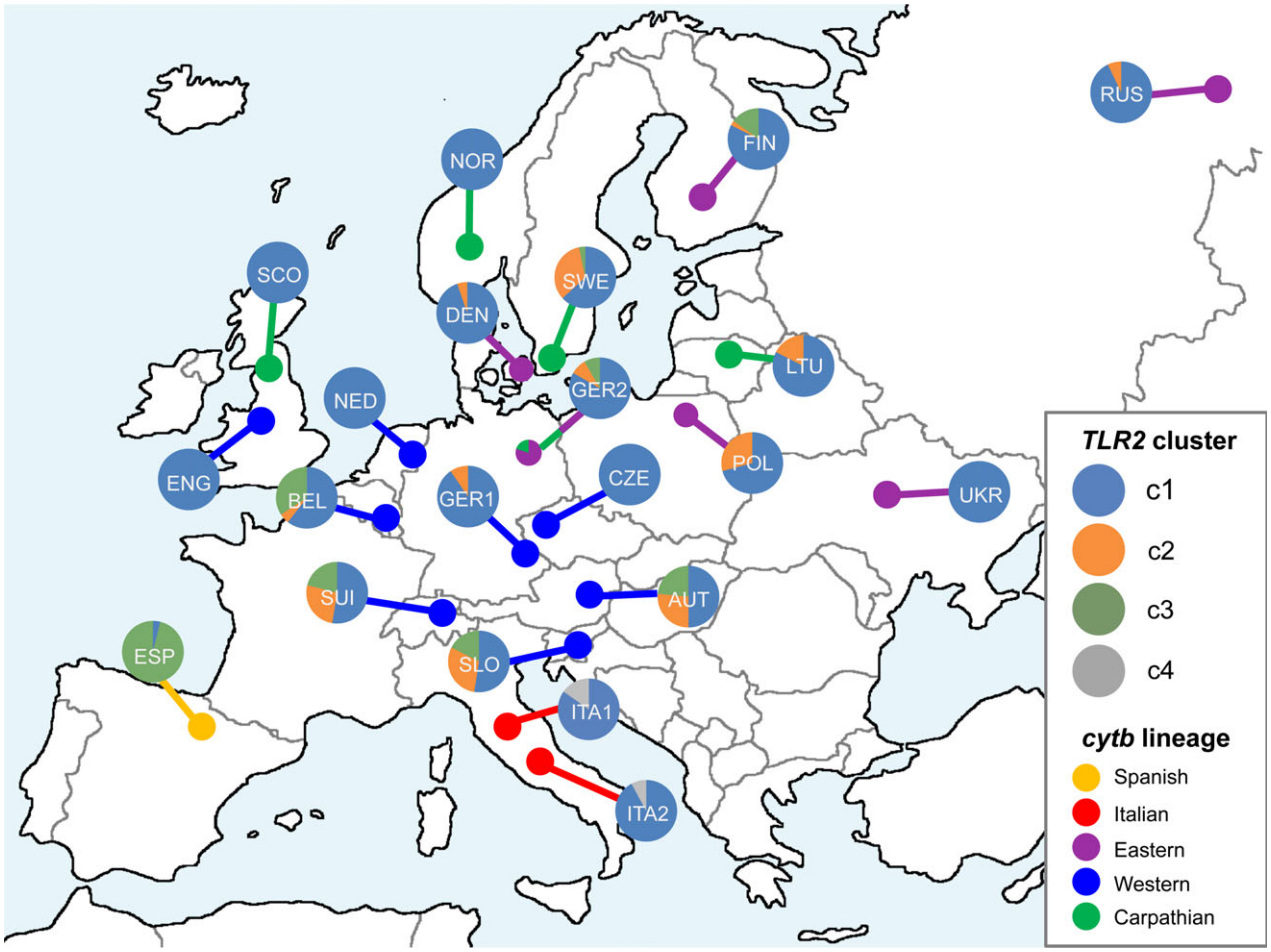

Figure 1. Toll-like receptor 2 (*TLR2*) haplotype cluster frequencies and mitochondrial lineage of the sampled bank vole populations. Population IDs are given within the corresponding *TLR2* cluster pie. $N = 385$ animals; $N = 20$–42 alleles per population.

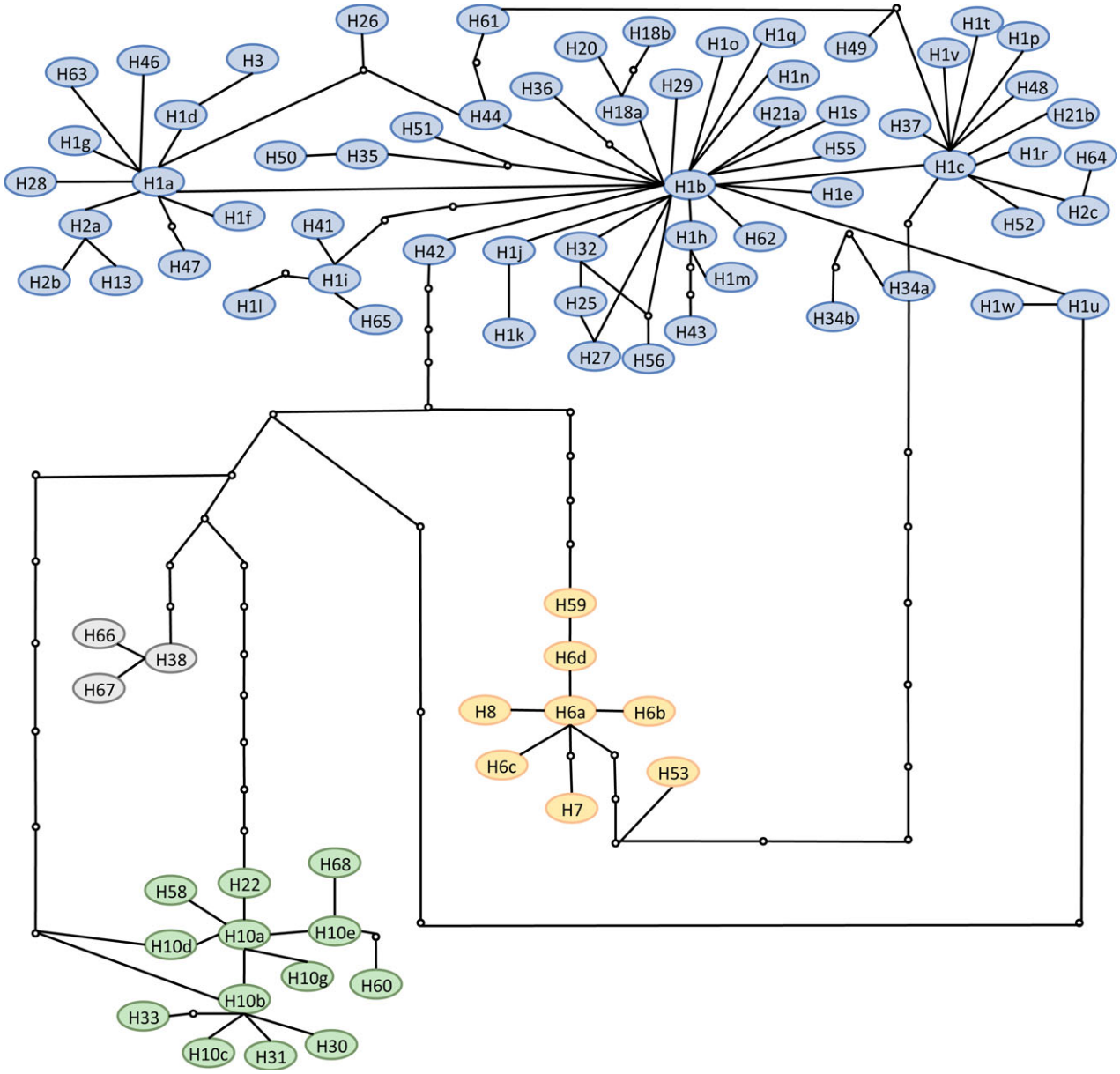

Figure 2. Toll-like receptor 2 (*TLR2*) haplotype network. Haplotypes that differ at the amino acid level are labelled with different numbers (e.g. H1, H2, and H3). Haplotypes that do not differ at the amino acid level are labelled with different letters (e.g. H1a, H1b, and H1c). Haplotypes are coloured according to clusters: c_1 (blue), c_2 (orange), c_3 (green), c_4 (grey). Each line indicates a change at one nucleotide.

three main clusters ($c_{1–3}$) were found in the Western, Eastern, and Carpathian lineages, and two of the three main clusters (c_1, c_3) were observed in the Spanish lineage (Fig. 1). Only haplotypes of one previously described cluster (c_1) were present in the Italian lineage. In addition, a fourth haplotype cluster, restricted to the Italian populations, was observed (c_4) (Figs 1, 2).

AMOVA revealed that mitochondrial lineage explained 18% of variation in *TLR2* clusters, whereas population within lineage explained 14% of variation in *TLR2* clusters. Most variation in *TLR2* clusters was observed within populations (68%) (Table 2). In comparison, lineage explained 55% of *cytb* variation (see Supporting information, Table S6). When restricting the AMOVA analysis to the Western, Eastern and Carpathian lineages, the proportion of variation in *TLR2* explained by lineage dropped to 0.8% (Table 2), showing that the lineage effect in the overall analysis was a result of differ-

Table 2. Analysis of molecular variance of Toll-like receptor 2 (*TLR2*) haplotype clusters in bank voles

Level of comparison	d.f.	Sum of squares	Variance component	% Variation	*P*
All lineages					
Among lineages	4	21.95	0.032	17.9	0.024
Among populations, within lineages	16	16.47	0.025	13.7	< 0.001
Within populations	749	91.80	0.123	68.4	< 0.001
Western, Eastern, and Carpathian lineages					
Among lineages	2	2.76	0.001	0.8	0.299
Among populations, within lineages	15	16.05	0.026	16.6	< 0.001
Within populations	646	82.49	0.128	82.6	< 0.001

ences in the *TLR2* composition of the Italian and Spanish clades. There was no evidence that *TLR2* cluster frequencies deviated from Hardy–Weinberg equilibrium within populations (all nonsignificant after Bonferroni correction).

DISCUSSION

Molecular genetic analyses revealed a high *TLR2* diversity in 21 bank vole populations across Europe. Although the haplotype composition differed markedly among populations, and a large number of private alleles was observed within populations, *TLR2* haplotypes from all populations grouped in the same distinct clusters.

Haplotypes of the three main clusters were observed in the Western, Eastern, and Carpathian mitochondrial lineages, and it is possible that we underestimated *TLR2* diversity in the two other lineages (i.e. the Italian lineage where c_1 and c_4 occurred, and the Spanish lineage where c_1 and c_3 occurred) because fewer samples were analyzed for these groups. A molecular analysis of variance confirmed that most *TLR2* cluster variation was present within populations, rather than among populations or mitochondrial lineages. Furthermore, the variance explained by lineage was a result of the very high frequency of c_3 in the Spanish lineage, as well as the occurrence of a fourth cluster that was restricted to the Italian lineage. When excluding the Italian and Spanish clades, the variation explained by lineage dropped to 0.8% and was no longer significant. It shows that, at least among the Eastern, Western and Carpathian clades, there is no association between *TLR2* clusters and mitochondrial lineages.

Interestingly, the finding that *TLR2* clusters are shared among lineages and populations across Europe is similar to patterns observed at the bank vole MHC class II (*Dqa*-exon 2) where no genetic structure was observed at the continental scale and the mitochondrial lineage explained < 2% of *Dqa*-exon 2 variation (Malé *et al.*, 2012). One interesting hypothesis is that parasites might impose similar selective pressures on the two immune genes, thereby maintaining genetic variation in time and space.

Our results suggest that the distinct *TLR2* clusters may precede the split of the bank vole lineages, which is estimated to have occurred during the Upper Pleistocene between 0.19 and 0.56 Mya (Deffontaine *et al.*, 2009) and that, subsequently, clusters have been maintained within lineages. Alternatively, *TLR2* clusters may have arisen several times independently by convergent evolution after the last glaciation. This, however, appears unlikely given the large number of populations in which all three clusters were observed. Importantly, both scenarios suggest that the co-occurrence of the three main *TLR2* clusters may be favoured by natural selection. Haplotypes within clusters, on the other hand, do not appear to be maintained across populations and lineages, as suggested by the strong population differentiation and the large number of private alleles within populations. This within-cluster differentiation among population is likely the result of population isolation and drift, although we can currently not exclude other evolutionary forces.

Only three haplotypes in the southernmost populations ITA1 and ITA2 did not group with one of the three previously described *TLR2* clusters. It is possible that these haplotypes represent a forth cluster that was lost in other populations during glaciation but persisted on the Italian peninsula, which remained largely free of ice (Hewitt, 1999). Alternatively, the occurrence of these 'intermediate' haplotypes could be an indication for relaxed selection in southern Europe. Indeed, a strong association between *TLR2* genotype and *B. afzelii* resistance has recently been observed in bank voles (Tschirren *et al.*, 2013) and, although *Borrelia* sp. are common tick-transmitted pathogens in most parts of Europe (Lindgren & Jaenson, 2006), they are absent in cen-

tral Italy where populations ITA1 and ITA2 are situated (Cimmino *et al.*, 1992). Reduced parasite-mediated selection and/or bottlenecks during post-glacial recolonization could also explain the low *TLR2* diversity observed in populations at the northern periphery of the species range (i.e. Norway and the British Isles).

Interestingly, in Spain, the *Borrelia* genospecies 'R57' is highly prevalent (but absent in other parts of Europe), as is the otherwise rare *TLR2* cluster c_3 (Gil *et al.*, 2005; Barandika *et al.*, 2007). Testing for interactive effects between different *Borrelia* species and *TLR2* clusters will thus be a fruitful next step to better understand why the distinct *TLR2* clusters are maintained within bank vole lineages and populations, and why some *TLR2* clusters are common in some populations but rare in others. Furthermore, TLR2 is involved in the recognition of other pathogens (Lien *et al.*, 1999; Medzhitov, 2001), which constitute multiple selective forces on TLR2 and may contribute to the observed patterns of *TLR2* diversity and differentiation across Europe.

Besides interactions between host genotype and parasite strains or species (i.e. host G × parasite G interactions), other processes could contribute to the maintenance of *TLR2* clusters within populations. For example, overdominance appears to play a role in the maintenance of diversity at the MHC (Doherty & Zinkernagel, 1975; Apanius *et al.*, 1997; Oliver, Telfer & Piertney, 2009) and has also been suggested to contribute to the maintenance of genetic variation at TIRAP, an adaptor in the TLR signalling pathway (Khor *et al.*, 2007). However, we found no evidence for an excess of *TLR2* heterozygotes in our study populations. Alternatively, negative-frequency dependent selection might underlie the maintenance of *TLR2* clusters within populations (Tanaka & Nei, 1989; Woolhouse *et al.*, 2002).

In conclusion, the present study shows that *TLR2* haplotypes group in distinct clusters in bank vole populations across Europe. Because haplotypes of the three main *TLR2* clusters were found in at least three mitochondrial lineages, our results suggest that selection maintains genetic supertypes at this innate immune receptor in time and space.

ACKNOWLEDGEMENTS

We thank Robert J. Baker from the Texas Tech University for providing Ukrainian bank vole samples; Irene Keller and three anonymous reviewers for their comments on the manuscript; and the numerous helpers who assisted with bank vole trapping and DNA extraction in the different study populations. The present study was supported financially by the University of Zurich Research Priority Program 'Evolution in Action: From Genomes to Ecosystems' and the Swiss National Science Foundation (PP00P3_15745) (to BT).

REFERENCES

Akira S, Uematsu S, Takeuchi O. 2006. Pathogen recognition and innate immunity. *Cell* **124:** 783–801.

Andrés AM, Hubisz MJ, Indap A, Torgerson DG, Degenhardt JD, Boyko AR, Gutenkunst RN, White TJ, Green ED, Bustamante CD, Clark AG, Nielsen R. 2009. Targets of balancing selection in the human genome. *Molecular Biology and Evolution* **26:** 2755–2764.

Apanius V, Penn D, Slev PR, Ruff LR, Potts WK. 1997. The nature of selection on the major histocompatibility complex. *Critical Reviews in Immunology* **17:** 179–224.

Barandika JF, Hurtado A, Garcia-Esteban C, Gil H, Escudero R, Barral M, Jado I, Juste RA, Anda P, Garcia-Perez AL. 2007. Tick-borne zoonotic bacteria in wild and domestic small mammals in northern Spain. *Applied and Environmental Microbiology* **73:** 6166–6171.

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.

Bubb KL, Bovee D, Buckley D, Haugen E, Kibukawa M, Paddock M, Palmieri A, Subramanian S, Zhou Y, Kaul R, Green P, Olson MV. 2006. Scan of human genome reveals no new loci under ancient balancing selection. *Genetics* **173:** 2165–2177.

Cimmino MA, Fumarola D, Sambri V, Accardo S. 1992. The epidemiology of Lyme borreliosis in Italy. *Microbiologica* **15:** 419–424.

Clement M, Posada D, Crandall K. 2000. TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9:** 1657–1660.

Corbet GB, Harris S. 1991. *The handbook of British mammals*. Oxford: Blackwell Scientific Publications.

Deffontaine V, Libois R, Kotlík P, Sommer R, Nieberding C, Paradis E, Searle JB, Michaux J. 2005. Beyond the Mediterranean peninsulas: evidence of central European glacial refugia for a temperate forest mammal species, the bank vole (*Clethrionomys glareolus*). *Molecular Ecology* **14:** 1727–1739.

Deffontaine V, Ledevin R, Fontaine MC, Quere JP, Renaud S, Libois R, Michaux JR. 2009. A relict bank vole lineage highlights the biogeographic history of the Pyrenean region in Europe. *Molecular Ecology* **18:** 2489–2502.

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

Drummond AJ, Ashton B, Cheung M, Heled J, Kearse M, Moir R, Stones-Havas S, Thierer T, Wilson A. 2009. *Geneious*, Version 5.6.5. Available at: http://www.geneious.com

Dupanloup I, Schneider S, Excoffier L. 2002. A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology* **11:** 2571–2581.

Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10:** 564–567.

Fay JC, Wu CI. 2000. Hitchhiking under positive Darwinian selection. *Genetics* **155:** 1405–1413.

Ferguson W, Dvora S, Fikes RW, Stone AC, Boissinot S. 2012. Long-term balancing selection at the antiviral gene OAS1 in Central African chimpanzees. *Molecular Biology and Evolution* **29:** 1093–1103.

Fornůsková A, Bryja J, Vinkler M, Macholán M, Piálek J. 2014. Contrasting patterns of polymorphism and selection in bacterial-sensing toll-like receptor 4 in two house mouse subspecies. *Ecology and Evolution* **4:** 2931–2944.

Gil H, Barral M, Escudero R, Garcia-Perez AL, Anda P. 2005. Identification of a new *Borrelia* species among small mammals in areas of northern Spain where Lyme disease is endemic. *Applied and Environmental Microbiology* **71:** 1336–1345.

Haldane JBS. 1949. Disease and evolution. *La Ricerca Scientifica* **19:** 68–76.

Hedrick PW. 1998. Balancing selection and MHC. *Genetica* **104:** 207–214.

Hedrick PW. 2002. Pathogen resistance and genetic variation at MHC loci. *Evolution* **56:** 1902–1908.

Hewitt GM. 1999. Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* **68:** 87–112.

Jaarola M, Tegelström H, Fredga K. 1999. Colonization history in Fennoscandian rodents. *Biological Journal of the Linnean Society* **68:** 113–127.

Jackson JA, Friberg IM, Bolch L, Lowe A, Ralli C, Harris PD, Behnke JM, Bradley JE. 2009. Immunomodulatory parasites and toll-like receptor-mediated tumour necrosis factor alpha responsiveness in wild mammals. *BMC Biology* **7:** doi:10.1186/1741-7007-1187-1116.

Jin MS, Kim SE, Heo JY, Lee ME, Kim HM, Paik SG, Lee HY, Lee JO. 2007. Crystal structure of the TLR1–TLR2 heterodimer induced by binding of a tri-acylated lipopeptide. *Cell* **130:** 1071–1082.

Khor CC, Chapman SJ, Vannberg FO, Dunne A, Murphy C, Ling EY, Frodsham AJ, Walley AJ, Kyrieleis O, Khan A, Aucan C, Segal S, Moore CE, Knox K, Campbell SJ, Lienhardt C, Scott A, Aaby P, Sow OY, Grignani RT, Sillah J, Sirugo G, Peshu N, Williams TN, Maitland K, Davies RJO, Kwiatkowski DP, Day NP, Yala D, Crook DW, Marsh K, Berkley JA, O'Neill LAJ, Hill AVS. 2007. A Mal functional variant is associated with protection against invasive pneumococcal disease, bacteremia, malaria and tuberculosis. *Nature Genetics* **39:** 523–528.

Kotlík P, Deffontaine V, Mascheretti S, Zima J, Michaux JR, Searle JB. 2006. A northern glacial refugium for bank voles (*Clethrionomys glareolus*). *Proceedings of the National Academy of Sciences of the of the United States of America* **103:** 14860–14864.

Laird PW, Zijderveld A, Linders K, Rudnicki MA, Jaenisch R, Berns A. 1991. Simplified mammalian DNA isolation procedure. *Nucleid Acids Research* **19:** 4293.

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

Lien E, Sellati TJ, Yoshimura A, Flo TH, Rawadi G, Finberg RW, Carroll JD, Espevik T, Ingalls RR, Radolf JD, Golenbock DT. 1999. Toll-like receptor 2 functions as a pattern recognition receptor for diverse bacterial products. *Journal of Biological Chemistry* **274:** 33419–33425.

Lindgren E, Jaenson TGT. 2006. *Lyme borreliosis in Europe: influences of climate and climate change, epidemiology, ecology and adaptation measures*. Copenhagen: World Health Organization Regional Office for Europe, 1–35.

Malé PJG, Martin JF, Galan M, Deffontaine V, Bryja J, Cosson JF, Michaux J, Charbonnel N. 2012. Discongruence of *Mhc* and cytochrome *b* phylogeographical patterns in *Myodes glareolus* (Rodentia: Cricetidae). *Biological Journal of the Linnean Society* **105:** 881–899.

Medzhitov R. 2001. Toll-like receptors and innate immunity. *Nature Reviews Immunology* **1:** 135–145.

Morger J, Bajnok J, Boyce K, Craig PS, Rogan MT, Lun ZR, Hide G, Tschirren B. 2014. Naturally occurring Toll-like receptor 11 (*TLR11*) and Toll-like receptor 12 (*TLR12*) polymorphisms are not associated with *Toxoplasma gondii* infection in wild wood mice. *Infection Genetics and Evolution* **26:** 180–184.

Netea MG, Wijmenga C, O'Neill LAJ. 2012. Genetic variation in Toll-like receptors and disease susceptibility. *Nature Immunology* **13:** 535–542.

Oliver MK, Telfer S, Piertney SB. 2009. Major histocompatibility complex (MHC) heterozygote superiority to natural multi-parasite infections in the water vole (*Arvicola terrestris*). *Proceedings of the Royal Society of London Series B, Biological Sciences* **276:** 1119–1128.

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

Stephens M, Smith NJ, Donnelly P. 2001. A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics* **68:** 978–989.

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. *Molecular Biology and Evolution* **30:** 2725–2729.

Tanaka T, Nei M. 1989. Positive Darwinian selection observed at the variable-region genes of immunoglobulins. *Molecular Biology and Evolution* **6:** 447–459.

Texereau J, Chiche JD, Taylor W, Choukroun G, Comba B, Mira JP. 2005. The importance of Toll-like receptor 2 polymorphisms in severe infections. *Clinical Infectious Diseases* **41:** S408–S415.

Tschirren B, Råberg L, Westerdahl H. 2011. Signatures of selection acting on the innate immunity gene Toll-like

receptor 2 (*TLR2*) during the evolutionary history of rodents. *Journal of Evolutionary Biology* **24:** 1232–1240.

Tschirren B, Andersson M, Scherman K, Westerdahl H, Råberg L. 2012. Contrasting patterns of diversity and population differentiation at the innate immunity gene Toll-like receptor 2 (*TLR2*) across populations of two sympatric rodent species. *Evolution* **66:** 720–731.

Tschirren B, Andersson A, Scherman K, Westerdahl H, Mittl P, Råberg L. 2013. Polymorphisms at the innate immune receptor *TLR2* are associated with *Borrelia* infection in a wild rodent population. *Proceedings of the Royal Society of London Series B, Biological Sciences* **280:** doi:10.1098/rspb.2013.0364.

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 Genetics* **7:** e1002343. doi:1002310.1001371/journal.pgen.1002343.

Weir BS, Cockerham CC. 1984. Estimating F-statistic for the analysis of population-structure. *Evolution* **38:** 1358–1370.

Woolhouse MEJ, Webster JP, Domingo E, Charlesworth B, Levin BR. 2002. Biological and biomedical implications of the co-evolution of pathogens and their hosts. *Nature Genetics* **32:** 569–577.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Figure S1. Schematic representation of the functional regions of *TLR2*.
Figure S2. Phylogenetic tree based on *cytb* sequences.
Figure S3. *cytb* genealogy.
Figure S4. *TLR2* haplotype frequencies within populations.
Figure S5. Amino acid polymorphisms in the four bank vole *TLR2* clusters.
Table S1. Sampling locations.
Table S2. Inferring *TLR* haplotypes using PHASE.
Table S3. Origin of *TLR2* haplotypes.
Table S4. *TLR2* population divergence.
Table S5. Fay & Wu's neutrality test.
Table S6. Analysis of molecular variance of *cytb*.