Cortázar-Chinarro et al. BMC Evolutionary Biology (2017) 17:189
DOI 10.1186/s12862-017-1022-z

BMC Evolutionary Biology

RESEARCH ARTICLE Open Access

Drift, selection, or migration? Processes affecting genetic differentiation and variation along a latitudinal gradient in an amphibian

Maria Cortázar-Chinarro[1*], Ella Z. Lattenkamp[1,2], Yvonne Meyer-Lucht[1], Emilien Luquet[1,3], Anssi Laurila[1] and Jacob Höglund[1]

Abstract

Background: Past events like fluctuations in population size and post-glacial colonization processes may influence the relative importance of genetic drift, migration and selection when determining the present day patterns of genetic variation. We disentangle how drift, selection and migration shape neutral and adaptive genetic variation in 12 moor frog populations along a 1700 km latitudinal gradient. We studied genetic differentiation and variation at a MHC exon II locus and a set of 18 microsatellites.

Results: Using outlier analyses, we identified the MHC II exon 2 (corresponding to the β-2 domain) locus and one microsatellite locus (RCO8640) to be subject to diversifying selection, while five microsatellite loci showed signals of stabilizing selection among populations. STRUCTURE and DAPC analyses on the neutral microsatellites assigned populations to a northern and a southern cluster, reflecting two different post-glacial colonization routes found in previous studies. Genetic variation overall was lower in the northern cluster. The signature of selection on MHC exon II was weaker in the northern cluster, possibly as a consequence of smaller and more fragmented populations.

Conclusion: Our results show that historical demographic processes combined with selection and drift have led to a complex pattern of differentiation along the gradient where some loci are more divergent among populations than predicted from drift expectations due to diversifying selection, while other loci are more uniform among populations due to stabilizing selection. Importantly, both overall and MHC genetic variation are lower at northern latitudes. Due to lower evolutionary potential, the low genetic variation in northern populations may increase the risk of extinction when confronted with emerging pathogens and climate change.

Keywords: Genetic drift, Natural selection, Major histocompatibility complex, Microsatellites, Outlier tests, *Rana arvalis*

Background

The relative roles of selection and drift shaping genetic diversity among and within natural populations are a contentious issue in evolutionary biology [1, 2]. Generally, the force of genetic drift is inversely related to population size, while the opposite is true for selection [3]. At the molecular level, the most common form of selection is purifying selection (either favouring or disfavouring allelic variants) which leads to loss of genetic variation. However, genetic loci can also be under various forms of balancing selection, which uphold and maintain variation over space and time [4]. As the effect of genetic drift is imminently linked with population size, variation in population size within a species' range is expected to play a central role for the importance of genetic drift. As species are generally expected to achieve their highest abundances at the center of their range, and populations tend to become progressively

* Correspondence: maria.cortazar@ebc.uu.se
[1]Animal Ecology/Department of Ecology and Genetics, Uppsala University, Norbyvägen 18D, 75236 Uppsala, Sweden
Full list of author information is available at the end of the article

© The Author(s). 2017 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

smaller towards the edge of the range (the 'abundant centre' model; [5, 6]), drift is considered as more important in populations closer to the range edge, leading into reduced genetic variation and stronger population structure in these peripheral populations [7, 8].

While the abundant center model has been influential for much of the theory, evidence for populations being more numerous or frequent at the centre of the distribution is, at best, limited [5]. Here, historical events like colonization processes can have a significant influence on patterns of genetic variation [8]. When new populations are repeatedly founded by a few individuals these founder effects will lead to loss of genetic variation [9–11]. In Northern Europe, the climate was considerably colder in the northern hemisphere during the Pleistocene, and most of the species currently present had withdrawn to refugia in Southern Europe [12, 13]. After the retreat of the glaciers, species expanded and recolonized the formerly uninhabitable areas [14, 15]. During the Holocene recolonization genetic variation was lost due to repeated founder events. As a result, less diverse populations are often found at northern latitudes [16–18], and these populations are predicted to genetically differ from the southern ones due to their demographic history.

Disentangling whether current patterns of genetic variation along geographical gradients are brought about by central-marginal or colonization processes is not simple, because, due to the location of glacial refugia, the two patterns are geographically correlated [8]. Moreover, these processes can be complicated by historical processes when populations following different migration routes meet in a secondary contact zone [8] and confounded by the fact that divergent selection in different parts of the species range may also cause population divergence [14]. However, such information is pivotal for understanding the genetic processes in contemporary populations.

In population and conservation genetics, spatial population structure and gene flow are usually assessed with neutral markers, assuming that this variation also reflects adaptive potential [19] However, the use of only neutral markers has important limitations and does not provide a complete picture on genetic variation and the evolutionary potential of wildlife populations [20–23]. In contrast to neutral markers, loci of the Major Histocompatability Complex (MHC) comprise well understood coding genes which show a high level of variation [24–28]. The MHC is a multigene family and the most polymorphic functional set of loci in vertebrate genomes described so far [29–33]. It codes for proteins that are part of the adaptive immune system and associated with disease and parasite resistance [4, 34–36]. Previous studies have shown a direct association between specific MHC alleles and pathogen resistance (e.g. [24, 37–40]). However, other studies have claimed a lack of association between disease resistance and MHC genetic variation [41, 42].

Amphibians are globally subjected to severe declines and currently the most threatened vertebrate class [43]. Along with the severe impact of habitat fragmentation, emerging infectious diseases are accounting for the global decline such as the chytridiomycosis caused by the chytrid fungus *Batrachochytrium dendrobatidis* (*Bd*), which has caused a large number of local and global amphibian extinctions [44, 45]. Several studies have emphasized the importance of MHC variation for disease resistance in amphibians [46]. For example, transcriptomic studies show that *Bd* infection activates immune pathways in many species [47, 48], and MHC heterozygous individuals or individuals with unique MHC alleles have been shown to be resistant to *Bd* [38, 39, 49] and other pathogens [50, 51]. The study of MHC variation is especially important in amphibian populations at northern latitudes, where genetic variation is often lower due to postglacial colonization processes. Indeed earlier studies in two amphibian species (*Triturus cristatus*: [52]; *Epidalea calamita*: [53, 54]) have found low MHC variation in high-latitude populations, potentially rendering these populations especially vulnerable to disease. With these exceptions, little is known about how adaptive genetic variation in general, and MHC variation in particular, varies with latitude in amphibians.

Studies on life-history variation of ranid frogs in northern Europe have revealed extensive adaptive genetic divergence in larval development rates, both at local scale (*Rana temporaria*; *Rana arvalis;* [55–57] and over wider latitudinal gradients (*Rana temporaria; Pelophylax lessonae* [58, 59]. Here the thyroid hormone pathway has an important dual function repressing thyroid-induced gene expression in premetamorphic tadpoles and, on the other hand, activating thyroid-induced genes to initiate metamorphosis (*Xenopus laevis;* [60, 61]). The microsatellite locus RCO8640 located inside the up-regulated transcription factor gene C/EBP is involved in the thyroid hormone pathway (*Lithobates catesbiana*; [62, 63], and has been linked with adaptive divergence in larval developmental rate (*Rana arvalis*; [56, 57]) providing an additional marker of adaptive genetic variation.

In this study, we assessed genetic variation in 12 moorfrog *Rana arvalis* populations along a 1700 km latitudinal gradient from northern Germany to northern Sweden. Sampling along this gradient allowed us to study variation in the single MHC class II exon 2 locus and 18 microsatellites in order to: 1) trace the postglacial history of the moor frog along the gradient, 2) study the level of population genetic diversity along the gradient in order to elucidate adaptability to more local environments, and 3) investigate possible different effects of selection along the gradient. Our ultimate goal is to

understand how these processes combine to create a complex pattern of within and among site genetic diversity of *R. arvalis* along the latitudinal gradient.

Results

Microsatellites – Null alleles and detection of loci under selection

Out of 28 initially tested microsatellite loci, three (Rtempμ9, Rtempμ7, Rt2Ca9) were monomorphic, three (WRA1–22, WRA1–28, WRA6–8) failed to amplify and four (RtμJ, RNTYR2, RNTYR2, RtCa11) produced ambiguous scoring patterns and were therefore excluded from subsequent analyses (Additional file 1: Table S1). The remaining 18 microsatellites were polymorphic in all populations with 6.3 alleles per locus on average. Three out of the remaining 18 loci (EU_11, RECALQ and RtCa41) were found to contain null alleles and were removed from further analyses. Five loci (RRDD590, Rtca18, RtuP, RtCa25, and WRA_160) were identified to be under stabilizing selection with F_{ST} lower than 0.03, nine loci were identified as neutral (R1atCa17, Rtempμ4, Rtempμ5, RCIDII, EU_06, EU_12, EU_15, EU_19, EU_24), whereas the locus RCO8640 was under diversifying selection with high F_{ST} (0.48; Additional file 1: Table S1 and Additional file 2: Table S2). All loci identified as outliers by Lositan and/or BayeScan were deemed as "under selection" and excluded from the neutral expectation analyses (Additional file 2: Table S2). All loci that were under diversifying selection, stabilizing selection, identified either by Lositan or Bayescan, or neutral expectations when the MHC II exon 2 was included in the analyses are summarized in Additional file 3: Table S3. The MHC II exon 2 and RCO8640 were under diversifying selection in the full gradient and in the southern cluster but not in the northern cluster (Fig. 1, Additional file 3: Table S3).

MHC II exon 2 diversity: Miseq run summary

We obtained a total of 20 million of MiSeq sequence reads in two separate runs, an average of 2.5–2.7 million reads per pool with intact primers and attached barcode information that could be assigned to a total of 361 amplicons from 207 individuals. In total 33 samples out of 240 failed due to PCR amplification problems. The average number of reads per amplicon was 13,404, ranging from 300 to 53,006; five amplicons with <300 reads were discarded. We amplified and sequenced 111 of the 207 genotyped individuals in two or three replicates, which corresponds to a replication rate of 53.6%, including every amplicon that revealed a unique MHC allele. Replicates were randomly assigned across different pools to avoid false allele identification. All replicates revealed the identical genotype as the respective original sample, leading to a genotyping reproducibility of 100%. Among the 207 individuals, we assigned 57 valid MHC II exon 2 alleles with a length of 272 bp and 27 polymorphic nucleotide positions. None of these sequences showed insertions, deletions or stop codons, therefore we assume that these are true alleles. All the 57 valid MHC II exon 2 allele sequences were translated into unique amino acid alleles. We checked for signs of recombination by using Omegamap [64] and did not find any signal of recombination in MHC class II exon 2. By following the DOC method [65], we detected a single MHC class II locus in 206 individuals. One sample (A10; Germany), however, revealed a second MHC class II locus in apparently lower read numbers in two of the three replicates, pointing to the possible existence of a very rare MHC class II duplication. We conclude that we are working with a single MHC class II locus in our data set.

Genetic structure among populations

The neutral microsatellites markers showed a global F_{ST} of 0.19 (95% C.I. = 0.14–0.26). In the northern cluster (see below) F_{ST} for neutral microsatellites was 0.12 (95% C.I. = 0.05–0.16) and in the southern cluster 0.09 (95% C.I. = 0.00–0.15). Population differentiation increased with geographic distance, showing clear and significant IBD (R^2 = 0.54; p = 0.003; Additional file 4: Figure S1). Using the STRUCTURE admixture model, the most

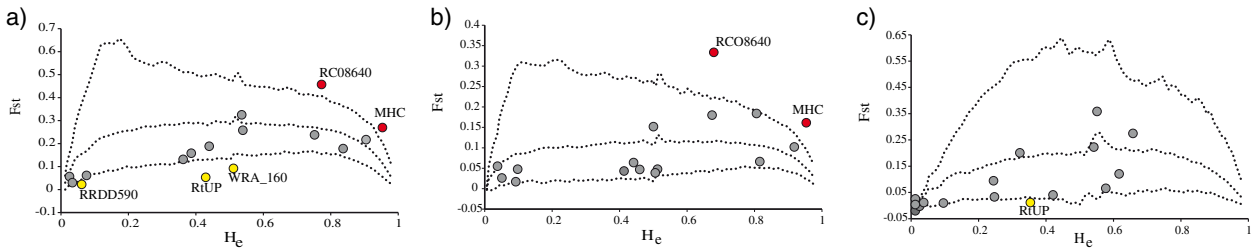

Fig. 1 F_{ST} vs expected heterozygosity for each 15 microsatellite and the MHC II exon 2 locus. Black dashed lines show the upper and lower 99% confidence intervals with 10,000 simulations from a stepwise mutation model (SMM), loci under neutrality expectations are colored in grey, loci under differential selection are colored in yellow and loci under diversifying selection are colored in red. Figure **a** represent the plot for the entire gradient, figure **b** the southern cluster and figure **c** northern clusters, respectively

likely number of genetic clusters was two (K = 2), revealing a northern (Umeå and Luleå) and a southern cluster (Uppsala, Skåne and Germany). Further STRUCTURE runs suggested population substructure (Additional file 5: Figure S2). The Uppsala and Germany populations seemed to be more similar to each other than to the Skåne populations when K = 3 was modelled. Four different clusters were observed when K = 4 was modelled while the same four clusters were observed with K = 5 since the northern populations (Luleå and Umeå) were not differentiated. A similar result was also found in the DAPC analyses which revealed two well defined clusters (a southern and a northern) with further substructure in concordance with the STRUCTURE results (Additional file 6: Figure S3). We did not find any evidence of gene flow between any of populations in the entire gradient using divMigrate [66]. This was the case even among local populations within the regions.

The global F_{ST} for MHC II exon 2 in the entire gradient was 0.36 with the German populations included (95% C.I. = 0.32–0.40) and 0.42 (95% C.I. = 0.23–0.47) when excluded. We found significant IBD (R^2 = 0.53; p = 0.031; Additional file 4: Figure S1). F_{ST} was 0.14 (95% C.I. = 0.07–0.18) and 0.28 with the German populations included (95% C.I. = 0.21–0.33) within the northern and southern cluster, respectively and 0.35 in the south when German population were excluded (95% C.I. = 0.24–0.42). The populations were clearly differentiated and genetically structured (Additional file 7: Table S4). Allele frequency pie-charts showed a structured pattern from northern to southern populations forming two clearly different clusters - a northern (Luleå and Umeå) and a southern (Uppsala, Skåne and Germany) cluster - for the MHC gene (Fig. 2; Additional file 8: Figure S4). F_{ST} between these two clusters was 0.42 (p < 0.001). Surprisingly, we found several alleles at very low frequencies at intermediate latitudes (Uppsala) and only five alleles were shared between Umeå in the north and the southern populations (Fig. 2; Additional file 8: Figure S4 and Additional file 9: Table S5).

The outlier microsatellite locus RCO8640 showed strong differentiation along the gradient with an overall F_{ST} of 0.45 (95% C.I. = 0.35–0.55). F_{ST} was 0.33 (95% C.I. = 0.20–0.47) and 0.32 (95% C.I. = 0.23–0.45) within the northern and southern cluster, respectively. Four different alleles were found at this locus. In Luleå, the northernmost populations in the gradient, only one allele was present, and in Umeå this was the most common allele (Fig. 2; Additional file 8: Figure S4).

When comparing differentiation in loci under diversifying selection (MHC and RCO8640) and neutral microsatellites we found that the Restricted Major Axis (RMA) regression slope for MHC differentiation (F'_{ST}) against neutral differentiation (G'_{ST}) tended to be higher in the southern (slope = 1.89 (95% C.I.: 0.93–3.55) compared to the northern cluster (slope = 0.76 (95% C.I.: 0.35–1.59, Fig. 3) possibly indicating stronger differentiation on MHC in the south. The corresponding RMA slopes for RCO8640 were 2.19 (C.I.: 1.68–2.85) in the southern and 1.41 (C.I.: 0.52–4.00) in the northern cluster (Fig. 3), again suggesting a trend of stronger differentiation in the south although the confidence intervals of the RMA slopes were widely overlapping.

Genetic diversity within populations

For the nine neutral microsatellites, H_E within populations ranged from 0.08 to 0.32 (overall H_E = 0.28) and

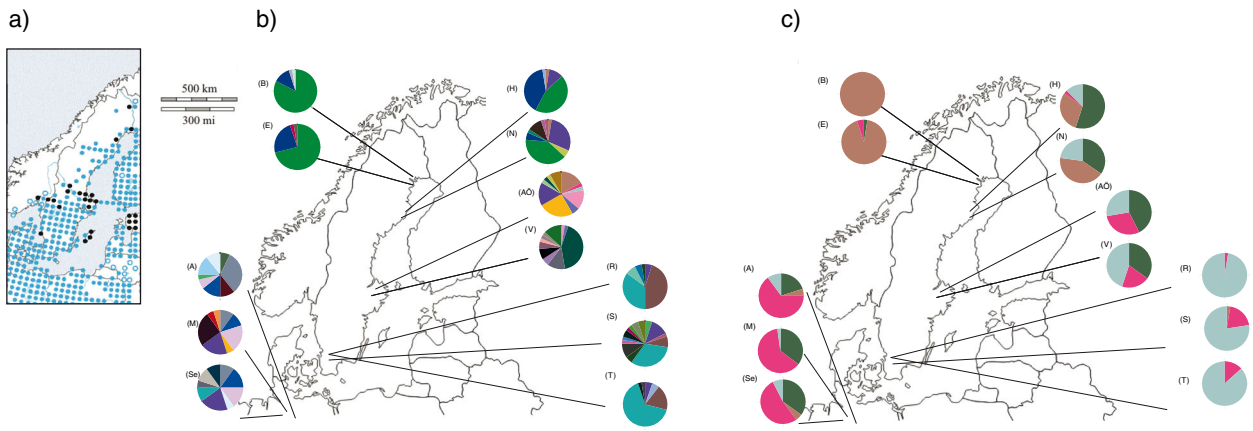

Fig. 2 a Approximate distribution of *R. arvalis* in Europe. The map is based on information in Gasc et al. [116] (*blue dots* and *open circles*), and updated with information from the Swedish Species Information Centre (http://artfaktaartdatabankense/taxon/208250), Sillero et al. [84], and own observations (*black dots*). **b** Allelic distribution of MHC Class II alleles and **c** RCO8640 in 12 *R arvalis* populations (B: Besbyn (Luleå); F: Ernäs (Luleå); N: Nydalasjön (Umeå); H: Holmsjön (Umeå); ÖA: Österbybruk (Uppsala); V: Valsbrunna (Uppsala); R: Räften (Skåne); S: Sjöhusen (Skåne); T: Tvedöra (Skåne), M: Mardorf (Germany), Se: Seebeckwiesen (Germany). A: Altwarmbüchen (Germany)). Colour coding scheme for MHC alleles is given in the (Additional file 5: Fig. S2)

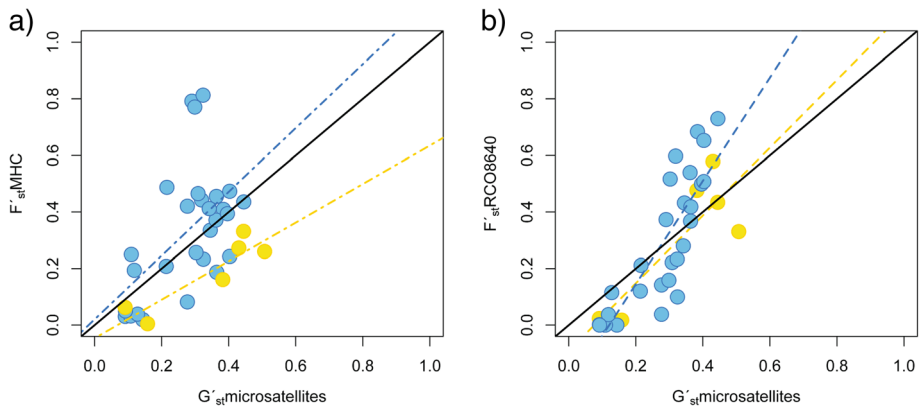

Fig. 3 Standardized F'_{ST} pairwise comparisons **a**) for MHC class II and **b**) RCO860 microsatellite marker **b**) in relation to Standardized G'_{ST} pairwise values for neutral microsatellites. The northern cluster is represented by *yellow circles* and the southern cluster is represented by *blue circles*

allelic richness from 1.45 to 3.0 (overall AR = 2.22). Observed and expected heterozygosity were not significantly different, as would be expected of selectively neutral loci. We found higher H_E and AR values in the southern (i.e., Uppsala and southwards) than in the northern regions (Umeå and Luleå; Table 2). We did not find any significant relationships between latitude and H_E or latitude and AR, along either part of the gradient (Fig. 4). The maximum number of alleles was found in the Uppsala region (46) followed by Germany (45).

Across all 207 individuals we found 57 alleles at the MHC class II β locus (Fig. 2; Additional file 8: Figure S4). The number of alleles per population varied substantially among regions. Levels of expected heterozygosity within populations for the MHC locus ranged from 0.31 to 0.91 (overall H_E = 0.71) and allelic richness ranged from 2.98 to 10.09 (mean AR = 6.45; Table 1). Levels of observed heterozygosity within populations ranged from 0.30 to 0.90 (overall H_O = 0.56; Table 1). Observed and expected heterozygosity were significantly deviant in five out of 12 populations (HWE equilibrium tests; $p < 0.001$). Especially, the German populations (A, M and Se) showed lower H_O values than expected (H_O = 0.35, H_E = 0.85; H_O = 0.30, H_E = 0.87 and H_O = 0.50, H_E = 0.91, respectively) compared to the rest of populations (Table 1). These low values of observed heterozygosity might be the result of an amplification problem of some alleles in these populations, leading to an artificial increase in homozygotes. The northern regions (Umeå and Luleå) showed lower diversity than the southern (all remaining) populations in terms of H_E and AR (Fig. 4; Wilcoxon test AR; $p < 0.001$ HE; $p < 0.001$). There was no significant relationship between latitude and H_E and AR in the southern part of the gradient. However, along the northern part of the gradient H_E was lower in the northernmost region (Luleå, Fig. 4 and Table 1). The highest number of alleles (46) was again found within the Uppsala region. Within the entire

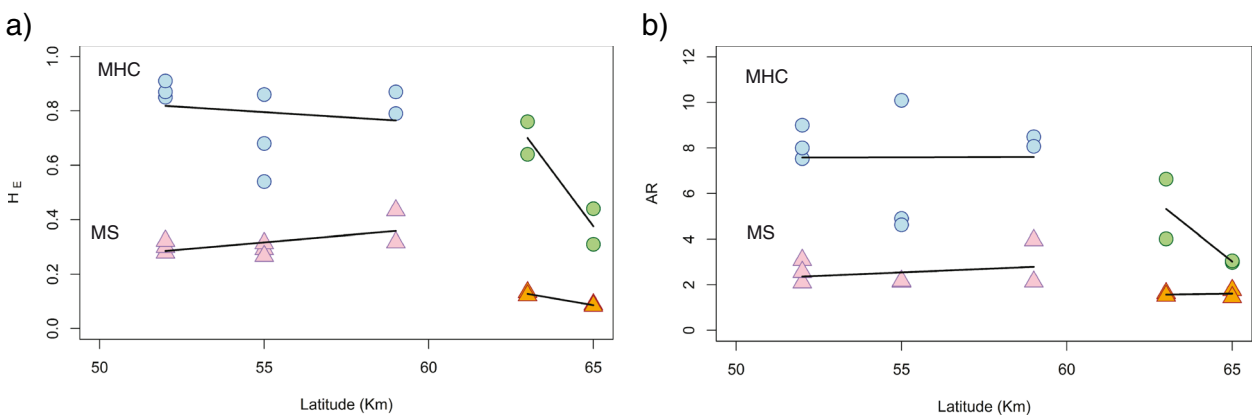

Fig. 4 MHC genetic variation (*blue circles* (southern cluster). and *green circles* (northern cluster). Microsatellite variation is given in *pink tringles* (southern cluster) and *orange triangles* (northern cluster). The linear regression is represented by a *black line*. **a**) H_E = Expected heterozygosity and **b**) AR = allelic richness

Table 1 Genetic variation at the MHC II exon 2 locus in the populations

Locality	Sampling area	code	n	NA	As	PA	H_o	H_e	AR
Altwarmbüchen		A	14	8		3	0.35*	0.85	7.53
Mardorf	Germany	M	10	8	17	3	0.30*	0.87	8.00
Seebeckwiesen		Se	10	9		2	0.50*	0.91	9.00
Sjohusen		R	20	14		2	0.80	0.68	4.90
Tvödöra	Skåne	S	20	6	18	7	0.90	0.86	10.09
Räften		T	19	6		1	0.42*	0.54	4.62
Österbybruk		AÖ	18	11		5	0.83	0.87	8.49
Valsbrunna	Uppsala	V	19	11	19	7	0.78	0.79	8.07
Holmsjön		H	19	5		0	0.57	0.64	4.01
Nydalasjön	Umeå	Ny	19	9	10	3	0.57*	0.76	6.63
Besbyn		B	20	4		2	0.30	0.31	2.98
Ernäs	Luleå	E	19	4	6	2	0.42	0.44	3.05
Total			207			37	0.56	0.71	6.45

The populations are ordered from South to North. n = number of individuals; NA = alleles within a population; As = alleles within a sampling area; PA = private alleles; H_o = observed heterozygosity, H_e = expected heterozygosity; AR = allelic richness. The HO that deviate significantly from H-W expectations are marked with a *

gradient, 37 out of 57 alleles were private to a single population (Table 1). The number of private alleles within a sampling area was highest in A (Germany, PA = 12) and ÖA (Uppsala, PA = 15) populations (Additional file 9: Table S5). Only one allele (Raar_DAB*15) was widespread and observed in 9 populations from different regions, while two alleles (Raar_DAB*20 and Raar_DAB*21) were only shared between the 4 populations in northern Sweden (Umeå and Luleå regions; Fig. 2; Additional file 8: Figure S4; Additional file 9: Table S5).

Discussion

We studied the MHC and microsatellite genetic variation in *R. arvalis* populations along a 1700 km latitudinal gradient and assessed the relative contributions of drift, selection and migration/colonization to understand the postglacial colonization history and the evolutionary forces acting on the adaptive potential and genetic variation of the populations. Four main results can be derived from our analyses. First, the postglacial migration history has resulted in two major clusters currently present in northern Germany and the Scandinavian Peninsula: a northern and a southern. Second, within population genetic variation is higher in the southern as compared to the northern cluster for all the studied genetic markers. Third, there are indications that selection is likely weaker and drift stronger in the northern cluster. Fourth, these forces combined have led to a complex pattern of differentiation along the gradient where some loci are more divergent among populations than predicted from drift expectations due to diversifying selection, while other loci are more uniform among populations due to stabilizing selection. We will discuss each of these conclusions in detail below.

Postglacial colonization history

Previous phylogeographical studies of *R. arvalis* suggest two routes of postglacial recolonization from southeastern Europe to Scandinavia [67, 68]. Accordingly, *R. arvalis* expanded from refugia in the Balkans and southern Russia, and used two different postglacial recolonization routes to colonize the Scandinavian Peninsula; one crossing from western central Europe to Sweden from the south-west via the postglacial land bridge that connected Denmark and Sweden, and the other using a route on the eastern side of the Baltic Sea via Finland to northern Sweden [67, 68]. Under this scenario the two lineages met in northern Sweden forming a contact zone, which exact location remains to be formally identified. Our results are in accordance with the previous studies. However, our study suggests that the contact zone lies further south (i.e. between Uppsala and Umeå) than the previous, primarily mt-DNA based, analysis [68], which placed the contact zone between Umeå and Luleå.

Within region diversity

Our results show lower genetic variation for both neutral microsatellites and the MHC II exon 2 locus in populations from the north of Sweden as compared to southern populations. For the MHC II exon 2 locus this difference is especially pronounced in the northernmost sampling region (Luleå), where we only found six of the 57 alleles present along the gradient. These results are in concordance with studies on different taxa, which have

found lower genetic variability in northern Europe as compared to central European populations [13, 16], and this is usually interpreted as a consequence of northern Sweden being the last area in Europe to be recolonized after the last glaciation events. This is also what is predicted by the central marginal hypothesis of genetic variation [7, 8]. However, in the case of *R. arvalis*, our analyses showed that the observed clinal variation is not simply due to a gradual decline along the colonization gradient as the two lineages colonizing the Scandinavian Peninsula meet at intermediate latitudes somewhere between Uppsala and Umeå. This leads to a prediction that these two localities are the least diverse along each of the respective colonization routes. However, the northernmost locality, Luleå, is the one showing the lowest within site diversity.

The depletion in genetic variation observed in the north suggests lower adaptive potential in response to climate change in northern populations [69]. The reduced variation in MHC exon II in the northern cluster is in line with the earlier results on two amphibian species [52–54]. Although these results suggest long-term survival of populations with very low MHC variation in postglacial expansion areas for hundreds of generations, the low number of MHC II alleles in the northern population (Luleå) can be a disadvantage in the future if the populations are confronted with novel or emerging diseases [70, 71].

The high levels of expected heterozygosity, the large number of rare alleles - numbers exceeding or being in line with what has previously been reported for refugial populations in unglaciated areas of Europe [53] – as well as the heterozygosity excess for the MHC exon II in some southern populations might be explained by two hypotheses on pathogen mediate-selection mechanisms (PMS). First, *the heterozygote advantage hypothesis* assumes that heterozygous individuals are favored because they can recognize a broader range of pathogens [72, 73]. However, published evidence confirming a MHC-specific heterozygote advantage is limited [38, 74, 75]. Second, the *rare allele advantage hypothesis* assumes that uncommon alleles in the populations are likely to offer more protection than common alleles and thus confer a selective advantage [76]. In our data, we found rare alleles in almost all the populations over the entire gradient. These rare alleles could be a potential source for defense against pathogens in these populations. With our data we cannot distinguish among these two hypotheses and they are not mutually exclusive [35, 38, 39, 77, 78]. Further investigation regarding allele frequency distributions and parasite infection are needed to understand which mechanisms are responsible for maintaining genetic variation in relation to parasite resistance.

Selection/drift patterns along the gradient

R. arvalis populations showed clear structure and IBD in both adaptive and neutral markers despite being recently diverged in evolutionary time. Outlier analyses suggested that the MHC II exon 2 and RCO8640 loci have been affected by different evolutionary processes in the northern and southern cluster. We saw signs of diversifying selection only in the southern populations while all markers seem to be mainly influenced by drift processes in northern populations. However, this pattern could be enforced by the fact that we studied fewer populations in the northern cluster. A study in Scandinavian common frog *R. temporaria* found that the impact of drift was higher in northern than in the southern populations probably because the northern populations were smaller and more isolated [79]. In our study, *R. arvalis* is at the northernmost range of its distribution and we suggest that populations are smaller and the connectivity among populations is poor in the northern cluster. Therefore, drift processes are likely to be more important at high latitudes due to a high degree of population fragmentation and low effective population sizes (even though we cannot find a clear difference concerning effective sizes among southern and northern populations, Table 2).

The microsatellite locus RCO8640 was found to be under diversifying selection and had very low diversity in the northernmost populations (Luleå). This locus is located inside the up-regulated transcription factor gene C/EBP involved in the thyroid hormone pathway [62, 63], which is linked with adaptive divergence in larval developmental rate [56, 57]. We found that RCO8640 was under selection in the southern cluster, possibly suggesting selection on development rate along the southern part of the gradient, as found previously at local and broader geographical scales in northern European anurans (e.g., [57, 80]). While we did not find significant diversifying selection in the northern cluster, F_{ST} was still high along the northern part of the gradient. It would be very interesting to link variation in this locus to phenotypic variation in development rate along the latitudinal gradient.

When analyses were made on all the populations we found evidence of diversifying selection on two (MHC II exon 2 and RCO8640) and signs of stabilizing selection on five loci. We find two reasons which might explain why we find a high number of loci under stabilizing selection: 1) microsatellites were developed using known sequences of coding genes (See; [56, 57], 2) the long gradient with a high global F_{ST} allows for a better the detection of stabilizing selection. When the populations were divided into a northern and southern cluster we found evidence for selection in the southern cluster (three cases of stabilizing and two cases of diversifying selection) while in the northern cluster we could only find

Table 2 Genetic variation at 9 neutral microsatellites in the studied populations

Locality	Sampling area	code	n	NA	As	PA	H_o	H_e	AR	Ne (LDNe)	Ne (CNe)
Altwarmbüchen		A	20	38		8	0.38	0.28	3.08	1.9 (1.0–1.9)	0.7 (0.5–1.0)
Mardorf	Germany	M	20	25	45	1	0.30	0.30	2.10	11.2 (2.3-inf)	68.9 (0.8–346)
Seebeckwiesen		Se	20	36		1	0.33	0.32	2.56	37.1 (8.5 - inf)	8.5 (0.2–31.5)
Sjöhusen		S	20	31		0	0.30	0.29	2.16	28.1 (4 - inf)	inf (inf - inf)
Tvödora	Skåne	T	19	32	31	0	0.34	0.31	2.12	inf (10.1 - inf)	inf (inf - inf)
Räften		R	20	29		0	0.29	0.27	2.17	8 (2.1–51.9)	inf (inf - inf)
Österbybruk		AÖ	20	53		16	0.51	0.43	3.95	1.0 (0.8–1.2)	1.0 (0.7–1.3)
Valsbrunna	Uppsala	V	20	33	46	0	0.32	0.32	2.13	inf (203.4 - inf)	inf (inf - inf)
Holmsjön		H	20	20		1	0.14	0.13	1.62	inf (19.2 - inf)	inf (inf - inf)
Nydalasjön	Umeå	Ny	20	20	17	0	0.15	0.12	1.51	39 (19.6 -inf)	36.7 (0.0–184)
Besbyn		B	20	20		0	0.16	0.09	1.77	37 (2.7-inf)	4 (0.0–20)
Ernäs	Luleå	E	20	17	20	1	0.12	0.08	1.45	inf (0.6-inf)	3.4 (0.0–16.8)
Total			239			25	0.28	0.25	2.22		

The populations are ordered from South to North. n = number of individuals; NA = alleles within a population; As = alleles within a sampling area; PA = private alleles; H_o = observed heterozygosity, H_e = expected heterozygosity; AR = allelic richness; Ne = effective population size by linkagedisequilibrium method (Ne (LDNe)) and by coancestry method (CNe)

signs of stabilizing selection on one locus. These patterns could reflect actual differences among regions in the relative importance of drift and selection but we advise caution when interpreting the results of this study. While drift is predicted to be more important in small and fragmented populations and selection is more important in large connected ones [3], we cannot entirely rule out the possibility that the detected patterns may be partly due to lower sample sizes and fewer population contrasts in the north. So while in line with predictions from population genetic theory, these results should be deemed as tentative.

Combined effects of recolonization history, selection and drift

We suggest that genetic variation among the regions and populations can be explained by complex patterns of selection, drift and the two recolonization routes of Scandinavia since the last glaciation (see [81] for a similar example). Our results provide an example of a situation where the level of adaptive MHC II exon 2 diversity seems to be correlated with neutral diversity among populations. This is not always the case as depletion in overall genetic variation may or may not be correlated with the amount of adaptive genetic variation [4]. Here we found that overall F_{ST} and genetic diversity indexes (AR and H_E) were substantially higher for the MHC locus compared to neutral microsatellite markers (even though correlated). These results suggest that the MHC II exon 2 locus is under diversifying selection and are in agreement with previous studies finding more differentiation in MHC than in neutral loci suggesting adaptation to local parasite faunas (e.g. [27, 78]. However, earlier studies on differentiation at MHC loci show substantial heterogeneity. Other studies have found no difference among population differentiation at MHC and neutral markers indicating the dominant role of drift, and yet others find no differentiation at the MHC indicating balancing selection maintaining genetic variation (see [82] for a recent example, summarized by [83]). There is a need for further studies of the processes shaping within and among population genetic variation in natural populations to further improve our understanding on how genetic variation is geographically portioned and distributed.

Methods

Sample collection and DNA extraction

R. arvalis has a broad Euroasiatic distribution, from the North Sea coast to Siberia [67]. It lives in diverse habitats, from forests to pastures, and breeds in semi-permanent to permanent ponds and lakes. The centre of the distribution of *Rana arvalis* is located in the area of eastern Germany and western Poland (see distribution map in Sillero et al. [84]). We collected *R. arvalis* eggs in five regions from northern Germany (Hanover) to northern Sweden (Luleå) in 2014 and 2015 (Table 1, Additional file 10: Table S6; Figure 2). The eggs were collected at two sites in each region, with the exception of northern Germany and southern Sweden (Skåne), where samples were collected from three sites. The average distance between collection sites in the same region was 20 km (range 8 to 50 km). The collection sites were ponds situated in flat terrain dominated by mixed forest, pastures and agricultural land. At each site we collected ca. ten eggs from each of 20 freshly laid clutches. The

eggs were transported to Uppsala and kept in a climate room at 16 °C. After hatching the tadpoles (stage 25, [85]) were euthanized with an overdose of MS222, preserved in 96% ethanol and stored at 4 °C until DNA extraction.

Genomic DNA was extracted from 240 tadpoles (one individual/clutch) using a high salt extraction precipitation protocol (modified from [86]). Purity and concentration of DNA were determined with a NanoDrop® 2000, spectrophotometer and Qubit®3.0 fluorometer Quantitation Kit (Invitrogen™). Species verification was carried out by mtDNA cytochrome b amplification followed by the addition of *HaeIII* restriction enzyme [87]. Digestions by *HaeIII* produces different, easily distinguishable banding patterns in *R. arvalis* and *R. temporaria.*

Microsatellite genotyping

We successfully genotyped all individuals at 18 microsatellite loci isolated from different *Rana* species, and tested another set of ten microsatellites without success (Additional file 1: Table S1). Some of the successfully genotyped loci are situated in or near coding regions, increasing the probability of these markers being under selection. PCR amplification was performed individually for each microsatellite. Reactions were performed in a final volume of 20 μl using an ABI 2720 thermal cycler. The PCRs were done using either the Type-it® Microsatellite PCR Kit (Qiagen®, Sollentuna, Sweden) or DreamTaq (Thermo Scientific) following the manufacturer's instructions. Additional file 11: Table S7 specifies the used PCR type, the exact reaction mix and the thermocycling conditions for each PCR reaction. Prior to genotyping, the PCR products were diluted in water (1:10, 1:30 or 1:50). 1 μl was mixed with 9.8 μl Hi-DiTM Formamide (P/N 4311320, Applied Biosystems) and 0.2 μl size standard (GeneScanTM, 600 LIZ®, Thermo Scientific) and run on 3730xl DNA Analyzer (Applied Biosystems). Samples were genotyped using GeneMapper® Software 5 (Life TechnologiesTM).

MHC II exon 2 gene

Single locus amplification and preparation for sequencing

We amplified the complete second exon (272 bp) of the single MHC II gene (corresponding to the β −2 domain) in *R. arvalis*. The primers ELF_1 (3′-GAGGTGATCCCTCCAGTCAGT-5′) and ELR_2 (3′-GCATAGCAGACGGAGGAGTC-5′) were designed based on primers sequences developed for *R. pretiosa* and *R. luteiventris* [88], amplifying a 338 bp fragment with the primers positioned in the flanking intron-exon region (Additional file 12: Figure S5). Both forward and reverse primers were modified for Illumina MiSeq sequencing with an individual 8 bp barcode and a sequence of three N (to facilitate cluster identification). Each amplicon was marked with an individual combination of a forward and a reverse barcode for identification. PCR reactions were conducted in a total volume of 20 μl containing 1 μl of genomic DNA, 2 μl of 10X Dream taq buffer (Thermo scientific lab), 0.4 μl of 2 mM of each dNTP, 0.5 μl of each 10 μM primer (ELF_1 and ELR_2, respectively), 1.5 μl of Bovine serum albumine (BSA; 5 mg/ml) and 0.25 μl of Dream taq DNA polymerase (5 U/μl, Thermo scientific lab) in deionized water. Thermocycling was performed on the ABI 2720 (Applied Biosystems®). An initial denaturation step of 3 min at 95 °C was followed by a touch-down procedure, consisting of 30 s at 95 °C, annealing for 30 s at temperatures decreasing from 63 to 58 °C during the first 5 cycles (one detrimental degree per cycle) and ending with an extension step at 72 °C for 1 min. Thereafter, 30 similar cycles with a consistent annealing temperature of 58 °C followed, and PCR products were stored at 4 °C. All amplifications were carried out using filter tips in separate (pre- and post-PCR) rooms, and negative controls were included in all amplifications to avoid contaminations. PCR products were run and visualized on a 1.5% agarose gel using gelgreen (BIOTIUM). To reduce the number of samples for subsequent purification, 3–9 PCR products with similar concentrations were pooled based on estimations from the gel image. These sample pools were run on 1.5% agarose gel, the target band was excised from the gel and extracted using the MinElute Gel Extraction Kit (Qiagen® Sollentuna, Sweden). The concentration of each sample pool was measured with Quant-iT PicoGreen dsDNA assay kit (Invitrogen Life Technologies, Stockholm, Sweden) on a fluorescence microplate reader (Ultra 384; Tecan Group Ltd., Männedorf, Switzerland). The final amplicon pooling was done according to the measured concentrations and consisted of equimolecular amounts of each sample. A total of eight final amplicon pools were generated, and libraries were prepared using the Illumina Truseq DNA PCR-Free Sample preparation kit (Illumina Inc., San Diego, CA). Four pools each were combined into a Miseq run, and sequencing of two Miseq 250 runs was carried out at the National Genomic Infrastructure (NGI), the SNP&SEQ Technology Platform hosted at SciLifeLab in Uppsala (Sweden).

Miseq data analyses

Sequencing data were extracted from the raw data and combined into single forward reads using FLASH (Magoč and Salzberg 2011), each of the eight amplicon pools were analyzed independently. In total, eight fastq files were generated and transformed into fasta

(multifasta) files using Avalanche NextGen package (DNA Baser Sequence Assembler v4 (2013), Heracle BioSoft, www.DnaBaser.com). The jMHC software [89] was used to remove primer sequences and unique tags, and to generate alignments of all variants per amplicon. Generally, in MHC studies using NGS techniques rigorous quality control and filtering procedures have to be applied to distinguish PCR and sequencing artefacts from true MHC allele. In this study, however, the distinction of true alleles and artefacts was greatly simplified, for three reasons: 1) amplifying a single gene will only yield one or two true MHC alleles per amplicon, and the project was designed in a way that both 2) replication rate and 3) per amplicon coverage was markedly high.

We assigned the two most frequent variants within each amplicon as valid MHC alleles that occurred in at least 3% of the reads [52, 90]. In a single locus system, it is not expected that chimera are generated in higher frequencies than the two parental alleles, and hence, no chimeric sequence that could have been taken for a valid MHC alleles could be detected when thoroughly checking the sequences per amplicon by eye.

If only one sequence exceeded the 3% rule, the amplicon was scored as a homozygote. We discarded amplicons with <300 reads from the analysis for quality reasons. Among amplicons, valid MHC alleles had to be present in at least two amplicons (2-independent-PCR-criterion, [91]), therefore we replicated all amplicons that revealed unique MHC alleles in the second Miseq run, and ran every individual twice in the second Miseq run from independent PCRs. In addition, we used the DOC method, not assuming any specific number of loci to identify and estimate the number of alleles (*Ai*) per individual. This procedure is based on the break point in sequencing coverage between variants within each individual and avoids choosing a subjective threshold to separate true alleles from artefacts. In this procedure, variants are sorted top-down by coverage, followed by the calculation of the coverage break point (DOC statistic) around each variant. The variant with the highest DOC value is assumed to be the last true allele (see [65]).

All valid allele sequences were imported and aligned in MEGA 6.0 [92]. Allele sequences were extensively compared to other sequences for the same locus: giant spiny frog (*Quasipaa spinosa*; GenBank: KM390904.1) natterjack toad (*Epidalea calamita*; GenBank: HQ388291.1), mouse (*Mus musculus*; GenBank: JN948541.1) and turkey (*Meleagris gallopavo*; GenBank: DQ993255.2). According to the GT/AG rule [93] and in concordance with exon 2 in *Odorrana tormota* (JQ918829), the exon boundary was defined 16 bps downstream from intron-exon boundary determined in mouse. We removed the intron sequences for further analyses and ended up with a MHC II exon 2 fragment of 272 bps. Valid MHC exon 2 alleles were named following the nomenclature suggested by Klein [29]: a four digit abbreviation of the species name followed by gene*-numeration, e.g. Raar_DAB*01.

Data analyses

Outlier analyses

We used the programs Lositan [94] and BayeScan v. 2.01 [95] to identify loci under divergent or uniform selection. The null distribution of F_{ST} was simulated with 100,000 iterations implementing a stepwise mutation model (SMM) in Lositan. In Bayescan, outliers were detected by implementing the multinomial Dirichlet model. Outliers identified by at least one of the two methods were deemed as being non-neutral (Additional file 2: Table S2), and the term "neutral loci" is used to refer to the non-outlier loci included in subsequent analyses.

The microsatellite and MHC II exon 2 data set were divided in different subdatasets: the entire gradient with all the populations included in the analyses; a northern cluster (Umeå and Luleå); and a southern cluster (Germany, Skåne and Uppsala). We did the division after analyzing our data (see below) and in accordance with a bidirectional recolonization hypothesis of Scandinavia [67, 68]. For MHC II exon 2 we transformed the sequence data into genotype data using PGBSpider [96] to be able to run the outlier test in the Lositan software while implementing a Stepwise Mutation Model (SMM) (see Fig. 1; Additional file 1: Table S1). We also ran the outlier analyses using an Infinite Alleles Model (IAM). The results identifying outlier loci were independent of the mutation rate model, IAM or SMM (see Additional file 13: Figure S6). Outlier analyses were also repeated excluding the German populations from the data set (See Additional file 14: Table S8 and Additional file 15: Figure S7).

Analyses of microsatellite data

Input files were converted for different analyses program formats using the Excel add-in Microsatellite toolkit [97]. The frequency of null alleles was estimated with two different softwares: FreeNA [98] and Genepop [99]. To assess neutral genetic diversity, expected heterozygosity (H_E), observed heterozygosity (H_O) and allelic richness (AR), allele numbers rarified to the smallest sample size were calculated for each locality in FSTAT 2.9.3.2 [100] Multi-locus means were obtained using Microsatellite toolkit. Effective population size was estimated based on two different methods: the linkage disequilibrium method and the coancestry method using the software Ne estimator [101].

To examine population structure and differentiation, global and pair-wise F_{ST} between all the populations [102] were calculated according to Nei et al. [103], as implemented in the R-package hierfstat [104] We used the ENA correction described in [98] and F'_{ST}/G'_{ST} corrections [105] for all the F_{ST} values when appropriate. We tested for deviation from Hardy-Weinberg equilibrium (HWE) in ARLEQUIN v. 3.5. [106] and tested for Isolation by distance (IBD) by calculating pair-wise F_{ST} and log distance (km) for every population pair. Pair-wise F_{ST} was transformed based on Slatkin's (1995) data adjustment ($F_{ST}/(1-F_{ST})$). The Euclidean distance matrix was estimated using the R package Geosphere (see Hijmans et al. [107]). To test if pair-wise F_{ST} was spatially auto-correlated, Mantel tests were performed in R running the package Adegenet [108].

We visualized the spatial structure of microsatellite data using discriminant analysis of principal components (DAPC) implemented in Adegenet [108]. The optimal number of clusters for the DAPC was chosen based on the lowest Bayesian Information Criterion (BIC) value for the different clustering solutions, which coincided with a sharp break in the curve of BIC values as a function of k. Genetic clustering was also analyzed using STRUCTURE [109] to find the most likely number of genetic clusters (K) and assign individuals to these clusters assuming an admixture model and correlated allele frequencies with 10,000 burn-in steps and 100,000 MCMC (Markov chain Monte Carlo) steps. We investigated the likelihood of various numbers of K (2–5), following the approach suggested by Evanno et al. [110]. To infer the most likely K we visualized and compared the different likelihood distribution plots and ΔK plots produced by STRUCTURE HARVESTER [111] Runs for all the K were averaged using the LargeGreedy algorithm in CLUMPP software [112], which aligns cluster assignment across replicate analyses and were then visualized with DISTRUCT [113]. We estimated and visualized the gene flow patterns between all pairs of population samples by using divMigrate-online [66], which can detect asymmetric gene flow patterns. We included 12 populations, used G_{ST} (as implemented in divMigrate-online) as the differentiation metric and set the filter threshold to 0.20; 0.30; 0.40, respectively. Confidence limits on gene flow estimates were determined by 1000 bootstrap replicates.

Analyses of MHC II exon 2

To assess genetic diversity in the MHC II exon 2 standard diversity indices (H_E, H_O, and allelic frequencies) were calculated for each locality in Arlequin v 3.5 [106] Allelic richness was calculated in FSTAT 2.9.3.2 [100]. Allele frequency plots were created in R using the "ggplot2" package [114].

Global population differentiation, HWE expected values and a pair-wise distance matrix for all the populations were computed in ARLEQUIN v. 3.5; we ran 1000 permutations to allow calculation of confidence limits. We tested for Isolation by distance (IBD; [115]), as described above for the microsatellite data.

Comparing differentially selected and neutral loci

To compare population differentiation among the differentially selected loci, as revealed by the outlier analyses (above), we plotted the standardized F'_{ST} for MHC exon 2 and RCO8640, respectively, against the differentiation for neutral microsatellites, G'_{ST} [105]. We then computed the restricted major axis regression slopes for the northern and southern population comparisons separately for each comparison with a differentially selected marker. We hypothesized that if population differentiation was stronger among southern populations, the slope of this regression would be steeper than among northern populations.

Conclusion

In summary, populations were shown to be subjected to different selective regimes and combined with different historical demographic patterns affecting the strength of genetic drift, a complex pattern of differentiation have evolved along the gradient. Some loci are more divergent than expected by drift among populations due to diversifying selection while others are more uniform among populations due to stabilizing selection. Our data show a high number of MHC exon 2 alleles in comparison to other European amphibian species. Both overall and MHC genetic variation are lower at northern latitudes which suggest a high risk of extinction when confronted with emerging pathogens and climate change. These results emphasize the importance of latitudinal gradient studies in order to elucidate and understand the evolutionary processes shaping genetic variation among natural populations.

Additional files

Additional file 1: Table S1. Microsatellites used in the study. (PDF 40 kb)

Additional file 2: Table S2. 16 microsatellites outlier analyses results from Lositan and Bayescan. (PDF 31 kb)

Additional file 3: Table S3. 15 microsatellites and MHC II exon 2 outlier analyses results from Lositan and Bayescan. Analyses were performed independently considering: all the gradient, northern populations as a cluster and southern populations as another cluster. (PDF 31 kb)

Additional file 4: Figure S1. Isolation-by-distance a) for MHC class II and b) for microsatellite (MS) shown as Slatkin's linearized pair-wise F_{ST} "($F_{ST}/(1-F_{ST})$)" as a function of the natural logarithm of distance (km) between locality pairs. MHC class II is represented by blue circles and MS are represented by black stars. (PDF 2099 kb)

Additional file 5: Figure S2. Results of STRUCTURE analyses using the Admixture model. Average individual assignment probability (y-axis) of individuals for four values of K. Sampled populations are given below the plot and country, region or province of origin is given above. Delta K and Mean estimation Ln probability of the data are shown for different values of K. (PDF 326 kb)

Additional file 6: Figure S3. Discriminant component analyses (DAPC) of 9 neutral microsatellites for all the individuals. All individuals from the north cluster together to the right of the figure. (PDF 1539 kb)

Additional file 7: Table S4. Pairwise FSTs. a) MHC gene FST pairwise comparisons and b) neutral microsatellite FST pairwise comparisons, respectively. (PDF 49 kb)

Additional file 8: Figure S4. Colour scheme for the allele distribution maps a) MHC gene and b) RCO8640 gene. (PDF 875 kb)

Additional file 9: Table S5. Relative allele frequency for all the MHC alleles found along the latitudinal gradient. N = number of individual per population, AS = Alleles within sampling areas. (PDF 101 kb)

Additional file 10: Table S6. Information about pond locations along the latitudinal gradient. (PDF 48 kb)

Additional file 11: Table S7. PCR conditions for a total volume of 10 μl reaction for the neutral microsatellites finally used in the study. (PDF 69 kb)

Additional file 12: Figure S5. a) Primers were modified for Illumina Miseq consisting of a 8 bps barcode assignment and a sequence of three N Barcodes represented in pink (forward direction; "5- > 3") and in orange (reverse direction; 3- > 5"). The NNN sequence is shown in blue. Possible primer pair combinations are presented in two different boxes (brown and orange, respectively). Every forward primer was combined with 9 different reverse primers for every pool. A total of four pools were constructed in the study: blue (A), purple (B), orange (C) and red (D). b) Outline of the library preparation. (PDF 173 kb)

Additional file 13: Figure S6. Results of outlier analyses by Lositan software in MHC class II exon 2 and RCO8640 according to a) Infinite allele model approach (IAM) b) Stepwise mutation model approach (SMM). (PDF 140 kb)

Additional file 14: Table S8. 15 microsatellites and MHC II exon 2 outlier analyses results from Lositan and Bayescan. Analyses were performed independently considering: all the gradient and southern population, excluding all the German locations. Overall *F*ST for the gradient excluding all population from Germany was 0.35. (PDF 17 kb)

Additional file 15: Figure S7. Results of outlier analyses by Lositan software for 15 microsatellites neutral markers, MHC class II exon 2 and RCO8640 according to the Stepwise mutation model approach (SMM) and excluding all the German populations. (PDF 856 kb)

Acknowledgments
We thank Alex Richter-Boix and Soraya Rouifed for invaluable help in the field. Many thanks to Patrik Rödin Mörch and Peter Halvarsson for help on the statistical analyses, to Emilio Virgós Cantalapiedra for the helpful comments on the manuscript and to Kevin Mulder for the help with the primer design. We are thankful to Anders Hallengren (Länstyrelsen Skåne), Anders Forsgren and Stefan Andersson (Piteå Kommun), Thomas Brandt (Ökologische Schutzstation Steinhuder Meer e.V), and Bernd Rittberg and Andreas Jacob (Authority for Nature Protection of Lower Saxony and the Hannover region) for helping with the sampling locations and permissions. Funding was provided by The Swedish Research Council Formas (grant 146400178 to JH), Stiftelsen Oscar och Lili Lamms Minne (grant DO2013-0013 to JH), Stiftelsen för Zoologisk Forskning and Swedish Research Council (grant 621-2013-4503 to AL).

Funding
Funding was provided by The Swedish Research Council Formas (grant 146,400,178 to JH), Stiftelsen Oscar och Lili Lamms Minne (grant DO2013–0013 to JH), Stiftelsen för Zoologisk Forskning and Swedish Research Council (grant 621–2013-4503 to AL) for the study design, data collection and analyses.

Availability of data and materials
The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Authors' contributions
MC collected the data, did the practical work in the laboratory, analyzed the data and wrote the manuscript. EZL did some of the practical work in the laboratory and analyzed some data as a master project. EL participated in data collection. YML, AL and JH made substantial contributions to conception, design, analysis and interpretation of data. All the authors read and approved the final version of the manuscript.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Animal Ecology/Department of Ecology and Genetics, Uppsala University, Norbyvägen 18D, 75236 Uppsala, Sweden. [2]Present address: Department of Neurogenetics of Vocal Communication, Max Planck Institute of Psycholinguistics, Box 310, 6500 Nijmegen, Netherlands. [3]Present address: Université Claude Bernard - Lyon I, CNRS, UMR 5023 – LEHNA, 3–6, rue Raphaël Dubois - Bâtiments Darwin C and Forel, 69622 Villeurbanne Cedex 43, Boulevard du 11 novembre, 1918 Lyon, France.

Received: 17 April 2017 Accepted: 26 July 2017
Published online: 14 August 2017

References
1. Nei M. Molecular evolutionary genetics. New York City: Columbia university press; 1987.
2. Lynch M, Walsh B. The origins of genome architecture. Sunderland: Sinauer Associates; 2007.
3. Kimura M. The neutral theory of molecular evolution. Cambridge: Cambridge University Press; 1983.
4. Hedrick PW. Evolutionary genetics of the major histocompatibility complex. Am Nat. 1994;143(6):945–64.
5. Sagarin RD, Gaines SD. The abundant centre distribution: to what extent is it a biogeographical rule? Ecol Lett. 2002;5(1):137–47.
6. Vucetich JA, Waite TA. Spatial patterns of demography and genetic processes across the species' range: null hypotheses for landscape conservation genetics. Conserv Genet. 2003;4(5):639–45.
7. Parsons PA. Extreme environmental change and evolution. Cambridge: Cambridge University Press; 1997.
8. Eckert C, Samis K, Lougheed S. Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. Mol Ecol. 2008;17(5):1170–88.
9. Wright S. Evolution in Mendelian populations. Genetics. 1931;16(2):97–159.
10. Dobzhansky T, Dobzhansky TG. Genetics and the origin of species. New York City: Columbia University Press; 1937.
11. Mayr E. Systematics and the origin of species, from the viewpoint of a zoologist. Cambridge: Harvard University Press; 1942.
12. Hewitt GM. Speciation, hybrid zones and phylogeography—or seeing genes in space and time. Mol Ecol. 2001;10(3):537–49.
13. Hewitt GM. The structure of biodiversity–insights from molecular phylogeography. Front Zool. 2004;1(1):4.
14. Hampe A, Petit RJ. Conserving biodiversity under climate change: the rear edge matters. Ecol Lett. 2005;8(5):461–7.
15. Höglund J. Evolutionary conservation genetics. Oxford: Oxford University Press; 2009.
16. Hewitt GM. Some genetic consequences of ice ages, and their role in divergence and speciation. Biol J Linn Soc. 1996;58(3):247–76.

17. Hewitt GM. Post-glacial re-colonization of European biota. Biol J Linn Soc. 1999;68(1–2):87–112.
18. Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF. Comparative phylogeography and postglacial colonization routes in Europe. Mol Ecol. 1998;7(4):453–64.
19. Kohn MH, Murphy WJ, Ostrander EA, Wayne RK. Genomics and conservation genetics. Trends Ecol Evol. 2006;21(11):629–37.
20. van Tienderen PH, de Haan AA, Van der Linden CG, Vosman B. Biodiversity assessment using markers for ecologically important traits. Trends Ecol Evol. 2002;17(12):577–82.
21. Bekessy SA, Ennos RA, Burgman MA, Newton AC, Ades PK. Neutral DNA markers fail to detect genetic divergence in an ecologically important trait. Biol Conserv. 2003;110(2):267–75.
22. Holderegger R, Kamm U, Gugerli F. Adaptive vs. neutral genetic diversity: implications for landscape genetics. Landsc Ecol. 2006;21(6):797–807.
23. Väli Ü, Einarsson A, Waits L, Ellegren H. To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? Mol Ecol. 2008;17(17):3808–17.
24. Sommer S. Major histocompatibility complex and mate choice in a monogamous rodent. Behav Ecol Sociobiol. 2005;58(2):181–9.
25. Piertney S, Oliver M. The evolutionary ecology of the major histocompatibility complex. Heredity. 2006;96(1):7–21.
26. Faulks L, Östman Ö. Adaptive major histocompatibility complex (MHC) and neutral genetic variation in two native Baltic Sea fishes (perch *Perca fluviatilis* and zander *Sander lucioperca*) with comparisons to an introduced and disease susceptible population in Australia (*P. fluviatilis*): assessing the risk of disease epidemics. J Fish Biol. 2016;88(4):1564–83.
27. Meyer-Lucht Y, Mulder KP, James MC, McMahon BJ, Buckley K, Piertney SB, Höglund J. Adaptive and neutral genetic differentiation among Scottish and endangered Irish red grouse (*Lagopus lagopus scotica*). Conserv Genet. 2016; 17(3):615–30.
28. Rózsa J, Strand TM, Montadert M, Kozma R, Höglund J. Effects of a range expansion on adaptive and neutral genetic diversity in dispersal limited Hazel grouse (*Bonasa bonasia*) in the French Alps. Conserv Genet. 2016;17(2):401–12.
29. Klein J. Biology of the mouse histocompatibility-2 complex. Principles of immunogenetics applied to a single system. Berlin: Springer-Verlag; 1975.
30. Hood L, Steinmetz M, Malissen B. Genes of the major histocompatibility complex of the mouse. Ann Rev Inm. 1983;1(1):529–68.
31. Kaufman J, Flajnik M, Du Pasquier L, Riegert P. *Xenopus* MHC class II molecules. I. Identification and structural characterization. J Inmunol. 1985; 134(5):3248–57.
32. Hughes AL. Natural selection and the diversification of vertebrate immune effectors. Immunol Rev. 2002;190(1):161–8.
33. Rock KL, Reits E, Neefjes J. Present yourself! By MHC class I and MHC class II molecules. Trends Inmunol. 2016;37(11):724–37.
34. Neefjes J, Ploegh H. Assembly and intracellular transport of MHC molecules. In: The HLA system in clinical transplantation. Berlin: Springer; 1993.
35. Boyce WM, Hedrick PW, Muggli-Cockett NE, Kalinowski S, Penedo MCT, Ramey RR. Genetic variation of major histocompatibility complex and microsatellite loci: a comparison in bighorn sheep. Genetics. 1997;145(2):421–33.
36. Lei W, Zhou X, Fang W, Lin Q, Chen X. Major histocompatibility complex class II DAB alleles associated with intestinal parasite load in the vulnerable Chinese egret (*Egretta eulophotes*). Ecol Evol. 2016;6(13):4421–34.
37. Meyer-Lucht Y, Sommer S. MHC diversity and the association to nematode parasitism in the yellow-necked mouse (*Apodemus flavicollis*). Mol Ecol. 2005;14(7):2233–43.
38. Savage AE, Zamudio KR. MHC genotypes associate with resistance to a frog-killing fungus. Proc Natl Acad Sci. 2011;108(40):16705–10.
39. Savage AE, Zamudio KR. Adaptive tolerance to a pathogenic fungus drives major histocompatibility complex evolution in natural amphibian populations. Proc R Soc B. 2016;283(1827):20153115.
40. Froeschke G, Sommer S. Insights into the complex associations between MHC class II DRB polymorphism and multiple gastrointestinal parasite infestations in the striped mouse. PLoS One. 2012;7(2):e31820.
41. Hughes AL. MHC polymorphism and the design of captive breeding programs. Conserv Biol. 1991;5(2):249–51.
42. Hedrick PW. Conservation genetics: where are we now? Trends Ecol Evol. 2001;16(11):629–36.
43. Stuart SN, Chanson JS, Cox NA, Young BE, Rodrigues AS, Fischman DL, Waller RW. Status and trends of amphibian declines and extinctions worldwide. Science. 2004;306(5702):1783–6.
44. Berger L, Speare R, Daszak P, Green DE, Cunningham AA, Goggin CL, Slocombe R, Ragan MA, Hyatt AD, McDonald KR. *Chytridiomycosis* causes amphibian mortality associated with population declines in the rain forests of Australia and central America. Proc Natl Acad Sci. 1998;95(15):9031–6.
45. Fisher MC, Garner TW, Walker SF. Global emergence of *Batrachochytrium dendrobatidis* and amphibian *Chytridiomycosis* in space, time, and host. Annu Rev Microbiol. 2009;63:291–310.
46. Carey C, Cohen N, Rollins-Smith L. Amphibian declines: an immunological perspective. Dev Comp Immunol. 1999;23(6):459–72.
47. Ellison AR, Savage AE, DiRenzo GV, Langhammer P, Lips KR, Zamudio KR. Fighting a losing battle: vigorous immune response countered by pathogen suppression of host defenses in the chytridiomycosis-susceptible frog *Atelopus zeteki*. Genes Genomes Gen. 2014;4(7):1275–89.
48. Ellison AR, Tunstall T, DiRenzo GV, Hughey MC, Rebollar EA, Belden LK, Harris RN, Ibáñez R, Lips KR, Zamudio KR. More than skin deep: functional genomic basis for resistance to amphibian *chytridiomycosis*. Gen Biol Evol. 2015;7(1):286–98.
49. Bataille A, Cashins SD, Grogan L, Skerratt LF, Hunter D, McFadden M, Scheele B, Brannelly LA, Macris A, Harlow PS. Susceptibility of amphibians to *chytridiomycosis* is associated with MHC class II conformation. Proc R Soc B. 2015;282(1805):20143127.
50. Pearman PB, Garner TW. Susceptibility of Italian agile frog populations to an emerging strain of *Ranavirus* parallels population genetic diversity. Ecol Lett. 2005;8(4):401–8.
51. Teacher AG, Garner TW, Nichols RA. Evidence for directional selection at a novel major histocompatibility class I marker in wild common frogs (*Rana temporaria*) exposed to a viral pathogen (*Ranavirus)*. PLoS One. 2009;4(2):e4616.
52. Babik W, Pabijan M, Arntzen JW, Cogalniceanu D, Durka W, Radwan J. Long-term survival of a urodele amphibian despite depleted major histocompatibility complex variation. Mol Ecol. 2009;18:769–81.
53. Zeisset I, Beebee TJ. Drift rather than selection dominates MHC class II allelic diversity patterns at the biogeographical range scale in natterjack toads *Bufo calamita*. PLoS One. 2014;9(6):e100176.
54. Höglund J, Wengström Å, Rogell B, Meyer-Lucht Y. Low MHC variation in isolated island populations of the Natterjack toad (*Bufo calamita*). Conserv Genet. 2015;16(4):1007–10.
55. Lind M, Johansson F. The degree of adaptive phenotypic plasticity is correlated with the spatial environmental heterogeneity experienced by island populations of *Rana temporaria*. J Evol Biol. 2007;20(4):1288–97.
56. Richter-Boix A, Quintela M, Segelbacher G, Laurila A. Genetic analysis of differentiation among breeding ponds reveals a candidate gene for local adaptation in *Rana arvalis*. Mol Ecol. 2011;20(8):1582–600.
57. Richter-Boix A, Quintela M, Kierczak M, Franch M, Laurila A. Fine-grained adaptive divergence in an amphibian: genetic basis of phenotypic divergence and the role of nonrandom gene flow in restricting effective migration among wetlands. Mol Ecol. 2013;22(5):1322–40.
58. Laugen AT, Laurila A, Räsänen K, Merilä J. Latitudinal countergradient variation in the common frog (*Rana temporaria*) development rates–evidence for local adaptation. J Evol Biol. 2003;16(5):996–1005.
59. Orizaola G, Quintela M, Laurila A. Climatic adaptation in an isolated and genetically impoverished amphibian population. Ecography. 2010;33(4):730–7.
60. Buchholz DR, Paul BD, Fu L, Shi Y-B. Molecular and developmental analyses of thyroid hormone receptor function in *Xenopus laevis*, the African clawed frog. Gen Comp Endocrinol. 2006;145(1):1–19.
61. Buchholz DR, Paul BD, Shi Y-B. Gene-specific changes in promoter occupancy by thyroid hormone receptor during frog metamorphosis implications for developmental gene regulation. J Biol Chem. 2005;280(50):41222–8.
62. Atkinson BG, Warkman AS, Chen Y. Thyroid hormone induces a reprogramming of gene expression in the liver of premetamorphic *Rana catesbeiana* tadpoles. Wound Repair Regen. 1998;6(4):323–37.
63. Chen Y, Atkinson BG. Role for the *Rana catesbeiana* homologue of C/EBP α in the reprogramming of gene expression in the liver of metamorphosing tadpoles. Devel Gen. 1997;20(2):152–62.
64. Wilson DJ, McVean G. Estimating diversifying selection and functional constraint in the presence of recombination. Genetics. 2006;172(3):1411–25.
65. Lighten J, Oosterhout C, Paterson IG, McMullan M, Bentzen P. Ultra-deep Illumina sequencing accurately identifies MHC class IIb alleles and provides evidence for copy number variation in the guppy (*Poecilia reticulata*). Mol Ecol Resour. 2014;14(4):753–67.
66. Sundqvist L, Keenan K, Zackrisson M, Prodöhl P, Kleinhans D. Directional genetic differentiation and relative migration. Ecol Evol. 2016;6(11):3461–75.

67. Babik W, Branicki W, Sandera M, Litvinchuk S, Borkin L, Irwin J, Rafiński J. Mitochondrial phylogeography of the moor frog, *Rana arvalis*. Mol Ecol. 2004;13(6):1469–80.
68. Knopp T, Merilä J. The postglacial recolonization of northern Europe by *Rana arvalis* as revealed by microsatellite and mitochondrial DNA analyses. Heredity. 2009;102(2):174–81.
69. Somero G. The physiology of climate change: how potentials for acclimatization and genetic adaptation will determine 'winners' and 'losers'. J Exp Biol. 2010;213(6):912–20.
70. Radwan J, Biedrzycka A, Babik W. Does reduced MHC diversity decrease viability of vertebrate populations? Biol Conserv. 2010;143(3):537–44.
71. Siddle HV, Kreiss A, Eldridge MD, Noonan E, Clarke CJ, Pyecroft S, Woods GM, Belov K. Transmission of a fatal clonal tumor by biting occurs due to depleted MHC diversity in a threatened carnivorous marsupial. Proc Natl Acad Sci. 2007;104(41):16221–6.
72. Doherty PC, Zinkernagel RM. Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. Nature. 1975;256(5512):50–2.
73. Hughes AL, Nei M. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature. 1988;335(6186):167–70.
74. Dionne M, Miller KM, Dodson JJ, Bernatchez L. MHC standing genetic variation and pathogen resistance in wild Atlantic salmon. Phi Trans Roy Soc London B: Biol Sci. 2009;364(1523):1555–65.
75. Dionne M, Miller KM, Dodson JJ, Caron F, Bernatchez L. Clinal variation in MHC diversity with temperature: evidence for the role of host–pathogen interaction on local adaptation in Atlantic salmon. Evolution. 2007;61(9): 2154–64.
76. Takahata N, Nei M. Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. Genetics. 1990;124(4):967–78.
77. Gutierrez-Espeleta GA, Hedrick PW, Kalinowski ST, Garrigan D, Boyce WM. Is the decline of desert bighorn sheep from infectious disease the result of low MHC variation? Heredity. 2001;86(4):439–50.
78. Ekblom R, Saether SA, Jacobsson P, Fiske P, Sahlman T, Grahn M, Kålås JA, Höglund J. Spatial pattern of MHC class II variation in the great snipe (*Gallinago media*). Mol Ecol. 2007;16(7):1439–51.
79. Johansson M, Primmer CR, Merila J. History vs. current demography: explaining the genetic population structure of the common frog (*Rana temporaria*). Mol Ecol. 2006;15(4):975–83.
80. Palo JU, Schmeller DS, Laurila A, Primmer CR, Kuzmin SL, Merilä J. High degree of population subdivision in a widespread amphibian. Mol Ecol. 2004;13(9):2631–44.
81. Bonin A, Taberlet P, Miaud C, Pompanon F. Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). Mol Biol Evol. 2006;23(4):773–83.
82. Rico Y, Morris-Pocock J, Zigouris J, Nocera JJ, Kyle CJ. Lack of spatial immunogenetic structure among wolverine (*Gulo gulo*) populations suggestive of broad scale balancing selection. PLoS One. 2015;10(10):e0140170.
83. Strand TM, Segelbacher G, Quintela M, Xiao L, Axelsson T, Höglund J. Can balancing selection on MHC loci counteract genetic drift in small fragmented populations of black grouse? Ecol Evol. 2012;2(2):341–53.
84. Sillero N, Campos J, Bonardi A, Corti C, Creemers R, Crochet P-A, Isailović JC, Denoël M, Ficetola GF, Gonçalves J. Updated distribution and biogeography of amphibians and reptiles of Europe. Amphibia-Reptilia. 2014;35(1):1–31.
85. Gosner KL. A simplified table for staging anuran embryos and larvae with notes on identification. Herpetologica. 1960;16(3):183–90.
86. Paxton R, Thorén P, Tengö J, Estoup A, Pamilo P. Mating structure and nestmate relatedness in a communal bee, *Andrena jacobi* (Hymenoptera, Andrenidae), using microsatellites. Mol Ecol. 1996;5(4):511–9.
87. Palo JU, Merilä J. A simple RFLP method for identification of two ranid frogs. Conserv Genet. 2003;4(6):801–3.
88. Mulder KP, Cortazar-Chinarro M, Harris DJ, Crottini A, Grant EHC, Fleischer RC, Savage AE. Evolutionary dynamics of an expressed MHC class IIβ locus in the Ranidae (*Anura*) uncovered by genome walking and high-throughput amplicon sequencing. Dev Comp Immunol. 2017;76:177–88.
89. Stuglik MT, Radwan J, Babik W. jMHC software assistant for multilocus genotyping of gene families using next-generation amplicon sequencing. Mol Ecol Resour. 2011;11(4):739–42.
90. Galan M, Guivier E, Caraux G, Charbonnel N, Cosson J-F. A 454 multiplex sequencing method for rapid and reliable genotyping of highly polymorphic genes in large-scale studies. BMC Genomics. 2010;11(1):296.
91. Zagalska-Neubauer M, Babik W, Stuglik M, Gustafsson L, Cichoń M, Radwan J. 454 sequencing reveals extreme complexity of the class II major Histocompatibility complex in the collared flycatcher. BMC Evol Biol. 2010; 10(1):395.
92. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 2013;30(12):2725–9.
93. Mount SM. A catalogue of splice junction sequences. Nucleic Acids Res. 1982;10(2):459–72.
94. Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G. LOSITAN: a workbench to detect molecular adaptation based on a F(st)-outlier method. BMC Bioinf. 2008;9:323.
95. Foll M, Gaggiotti O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics. 2008;180(2):977–93.
96. Lischer H, Excoffier L. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics. 2012;28(2):298–9.
97. Park SDE. Trypanotolerance in West African cattle and the population genetic effects of selection. Dublin: University of Dublin; 2001.
98. Chapuis M-P, Estoup A. Microsatellite null alleles and estimation of population differentiation. Mol Biol Evol. 2007;24(3):621–31.
99. Rousset F. genepop'007: a complete re-implementation of the genepop software for windows and Linux. Mol Ecol Resour. 2008;8(1):103–6.
100. Goudet J. FSTAT (version 1.2): a computer program to calculate F-statistics. J Heredity. 1995;86(6):485–6.
101. Waples RS, Do C. LDNE: a program for estimating effective population size from data on linkage disequilibrium. Mol Ecol Resour. 2008;8(4):753–6.
102. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution. 1984;38(6):1358–70.
103. Nei M, Feldman MW. Identity of genes by descent within and between populations under mutation and migration pressures. Theor Pop Biol. 1972; 3(4):460–5.
104. Goudet J. Hierfstat, a package for R to compute and test hierarchical F-statistics. Mol Ecol Res Resources. 2005;5(1):184–6.
105. Meirmans PG, Hedrick PW. Assessing population structure: FST and related measures. Mol Ecol Resour. 2011;11(1):5–18.
106. Excoffier L, Lischer HE. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and windows. Mol Ecol Resour. 2010;10(3):564–7.
107. Hijmans RJ, Williams E, Vennes C. Geosphere: spherical trigonometry. R package version 1.3–11. 2014.
108. Jombart T. Adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics. 2008;24(11):1403–5.
109. Pritchard JK, Wen W, Falush D. Documentation for STRUCTURE software: version 2.2. Chicago: University of Chicago; 2007.
110. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol. 2005;14(8):2611–20.
111. Earl DA. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour. 2012;4(2):359–61.
112. Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics. 2007;23(14):1801–6.
113. Rosenberg N. Documentation for distruct software: version 1.1. Michigan: University of Michigan; 2007.
114. Wickham H. ggplot2: elegant graphics for data analysis. Berlin: Springer; 2016.
115. Wright S. Isolation by distance. Genetics. 1943;28(2):114.
116. Gasc J-P, Cabela A, Crnobrnja-Isailovic J, Dolmen D, Grossenbacher K, Haffner P, Lescure J, Martens H, Martínez Rica JP, Maurin H. Atlas of amphibians and reptiles in Europe. Bonn: Societas Europaea Herpetologica; 1997.