The LINNEAN SOCIETY of London BIOLOGICAL Journal of the Linnean Society

Biological Journal of the Linnean Society, 2013, **110**, 77–90. With 6 figures

Phylogeography of willow grouse (*Lagopus lagopus*) in the Arctic: taxonomic discordance as inferred from molecular data

JACOB HÖGLUND[1]*, BIAO WANG[1], THOMAS AXELSSON[2] and MARÍA QUINTELA[1]†

[1]*Population Biology and Conservation Biology, Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, SE 75236 Uppsala, Sweden*
[2]*Molecular Medicine, Department of Medical Sciences, University Hospital, SE 751 85 Uppsala, Sweden*

Received 21 November 2012; revised 10 March 2013; accepted for publication 10 March 2013

Using independently segregating nuclear single nucleotide polymorphisms (SNPs) and mitochondrial control region sequences, we found an east–west division among sampled willow grouse *Lagopus lagopus* subspecies. This division cut across the range of the subspecies with the largest distribution (*lagopus*) and thus contradicted existing taxonomic classifications. Russian *Lagopus lagopus lagopus* tended to cluster with North American willow grouse partly classified as other subspecies. Scandinavian willow grouse (*L. l. lagopus*) clustered with red grouse from Britain and Ireland (*Lagopus lagopus scoticus* and *Lagopus lagopus hibernicus*) but substructuring confirmed the monophyly of the latter. In North America, we could not detect any major genetic divisions apart from two birds described as *alexandrae* from the Heceta Island (Alaska) when using mitochondrial sequences. Other samples from North America were intermingled regardless of whether they were described as *muriei*, *alexandrae* or *lagopus*. A specimen described as *alexandrae* was to some extent distinct when analysing the SNP data. The genetic analyses indicated some concordance between genetics and taxonomy but not complete congruence. This is particularly evident for mitochondrial DNA network analyses. We suggest that the taxonomy of this species would benefit by a careful re-examination of the available evidence for subspecies. It appears as if subspecies status is a poor proxy for assigning evolutionary significant units and management units in this species.

ADDITIONAL KEYWORDS: *alexandrae* – evolutionary significant unit – *hibernicus* – *kamschatkensis* – *Lagopus lagopus* – *muriei* – phylogeography – *scoticus* – willow grouse subspecies.

INTRODUCTION

To take accurate conservation actions, it is important to correctly identify the taxonomic units relevant for conservation (Moritz, 1994; Crandall *et al*., 2000; Ballard & Whitlock, 2004; Beaumont & Balding, 2004). Many conservation actions still rely on taxonomic subspecies classifications based on morphological characters. Although these may be valid in many instances, they should be confirmed by the genetic data because classifications based purely on morphological data may be misleading as a result of phenotypic plasticity and clinal variation (Storz, 2002; Relethford, 2004). Subsequent to the introduction of genetic techniques in conservation, it has been common practice to use mitochondrial (mt)DNA sequences to infer species and other taxonomic relationships. For taxa to classify as an evolutionary significant unit (ESU), it has been suggested that they need to be reciprocally monophyletic for mtDNA alleles (Moritz, 1994). However, maternally inherited mtDNA has no (or very low) recombination rates and is thus inherited as a single linkage group (i.e. as one gene) (Ballard & Whitlock, 2004). It has long been known that single gene trees may be discordant with

*Corresponding author. E-mail: jacob.hoglund@ebc.uu.se
†Current address: Department of Animal Biology, Plant Biology and Ecology, Faculty of Science, University A Coruña, E-15008 A Coruña, Spain.

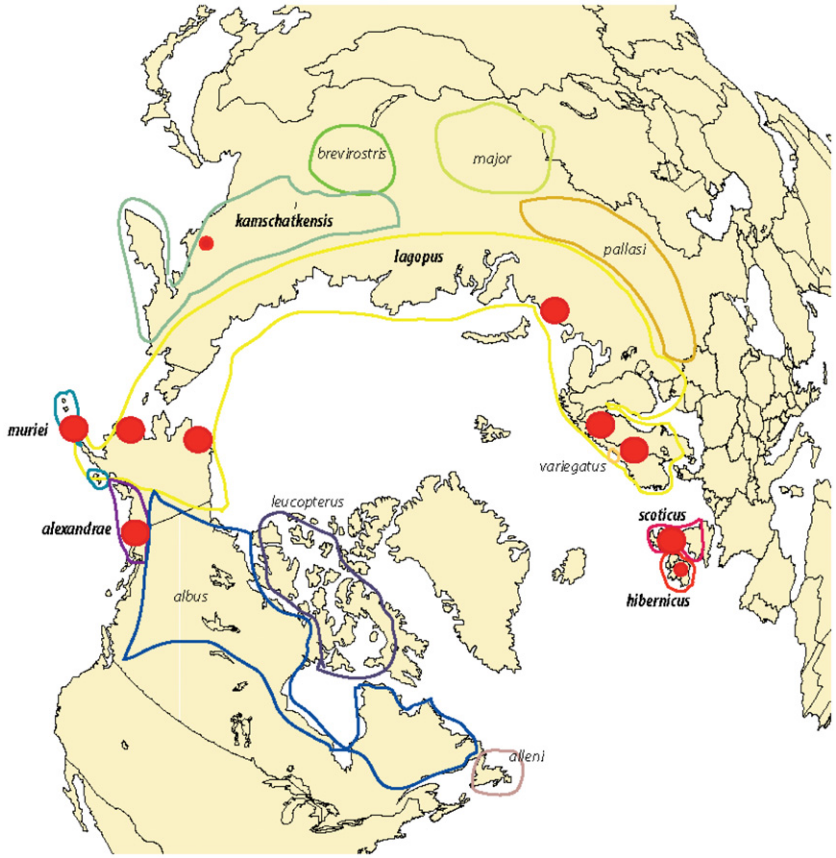

Figure 1. Approximate distribution of *Lagopus lagopus* subspecies (*sensu* Potapov 1985) Subspecies in bold were included in the present study. Large dots indicate several samples from that location; small dots refer to one sample.

the species tree and, consequently, that using a single gene tree may lead to inference of misleading species relationships (Pamilo & Nei, 1988; Maddison, 1997; Edwards & Beerli, 2000). As a result of the stochasticity with respect to the way that two lineages in a population coalesce in the preceding generation, no independently segregating single gene trees will be identical. Therefore, when reconstructing species phylogenies, the use of multiple loci is now recommended (Edwards, Liu & Pearl, 2007). Similarly, when using intraspecific phylogenies in phylogeographical studies, similar problems may arise and gene trees have been reported to be discordant either with morphology (Juan *et al*., 1996; Babik *et al*., 2005) or among markers (Bernatchez *et al*., 1995; Gantenbein & Largiadèr, 2003).

One approach for overcoming the problem of single gene trees would be to use data from many genes to reconstruct the species tree (Edwards *et al*., 2007). A complementary approach is to employ genetic marker data to infer the number of populations among the sampled individuals using criteria such as units in Hardy–Weinberg or linkage equilibrium. These are the criteria used by STRUCTURE (Pritchard, Stephens & Donnelly, 2000) in which the user can choose between different model-based approaches to detect the possible genetic population structure. If found, it is then straightforward to compare these groupings with taxonomic status and/or geographical origin of the analyzed individuals.

The willow grouse (or willow ptarmigan), *Lagopus lagopus*, Linnaeus, 1758, is a grouse species found in subalpine/boreal forests and heathlands in both the Nearctic and Palaearctic. Several subspecies are described but their validity should be evaluated (Storch, 2007). Johnsgard (1983) listed 16, Potapov (1985) listed 13, Potapov & Flint (1989) recognized 15, and del Hoyo, Elliot & Sargatal (1994) suggested 19 subspecies worldwide (Fig. 1). Hannon, Martin & Eason (1998) described six subspecies for North America. However, there have been few genetic studies covering large parts of the range of this almost circumpolar species and the subspecies classifications are based on morphological criteria that may be subjected to phenotypic plasticity and clinal variation. The willow grouse is a species that is

Table 1. Origin (and number) of individuals genotyped from each of the subspecies at the respective molecular markers

Species	Subspecies	Single nucleotide polymorphism markers	Mitochondrial DNA
Lagopus lagopus	*lagopus*	Alaska (5)	Alaska (5)
		Russia (6)	Russia (5)
		Scandinavia (51)	Scandinavia (15)
	kamschatkensis	Siberia East (1)	
	scoticus	UK (4)	UK (7)
	hibernicus	Ireland (1)	
	muriei	Alaska (5)	Alaska (4)
	alexandrae	Alaska (4)	Alaska (2)
		Canada (3)	Canada (3)
Lagopus muta			Scandinavia (3)
Hybrid			Scandinavia (1)
	Total	80	45

characterized by large outbreeding populations harbouring large amounts of genetic variation (Berlin, Quintela & Höglund, 2008), although isolated populations that have lost genetic variability have been described (Gyllensten, 1985; Freeland *et al*., 2007; McMahon *et al*., 2012). Willow grouse of both sexes normally attain an all white winter plumage, whereas females are camouflaged grey/brown during the summer and males develop a similar brownish nuptial plumage on the neck, breast, and back. The subspecies found in Britain and Ireland (called red grouse) does not acquire the winter plumage (Skoglund & Höglund, 2010).

The present study aimed to compare mitochondrial and nuclear genetic data with previous taxonomic classifications of a number of willow grouse individuals collected throughout the range of the species. We wanted to compare genetic data from nuclear single nucleotide polymorphisms (SNPs) and mitochondrial control region sequences with taxonomic status and geographical origin. It is assumed that, if the taxonomic status is correct, there will be a correspondence between this and the genetic data. However, if the taxonomic status and the genetic data do not correspond, we expect a better fit with geographical origin because this a widespread circumpolar species in which we expect a gradual change in allele frequencies as the geographical distance between samples increases.

MATERIAL AND METHODS

Samples

Samples from 80 individuals of *Lagopus lagopus* belonging to six different subspecies were collected: *alexandrae* (7), *hibernicus* (1), *kamschatkensis* (1), *lagopus* (62), *muriei* (5), and *scoticus* (4). The geographical distribution of the samples is shown in Table 1 and individuals were collected in North-America (subspecies *lagopus*, *alexandrae*, and *muriei*); Britain and Ireland (subspecies *hibernicus* and *scoticus*); Russia (subspecies *lagopus* and *kamschatkensis*); and Scandinavia (Norway and Sweden, subspecies *lagopus*); for further details about sampling sites, see the Supporting Information (Table S1). Samples from rock ptarmigan (*Lagopus muta*) and a *lagopus* × *muta* hybrid (with *muta* mtDNA) were also included to be used as an outgroup. DNA was extracted either using a salt-extraction procedure (Paxton *et al*., 1996) or the Qiagen DNeasy Blood and Tissue isolation kit (Qiagen Inc.) in accordance with the manufacturer's instructions. To avoid contamination, DNA extractions, pre-polymerase chain reaction (PCR) and post-PCR pipetting were carried out in different rooms and the equipment was sterilized using ultraviolet radiation.

Single nucleotide polymorphism (SNP) genotyping

Twenty-four unlinked SNP markers were selected for multiplex genotyping using the GenomeLab SNPstream system (Beckman Coulter) (Bell *et al*., 2002) available at the SNP & SEQTechnology Platform at Uppsala University (http://www.genotyping.se). SNPs were located in sixteen exons scattered throughout chromosomes of different sizes classes: AKR, APOA, bcl-2, BRIP, CAAX, CXC4, EPN, KELCH, LEPR, MBL, MICRO, PKP4, PPARG, TAR, TRANS, and YTH; for further details about SNP detection in willow grouse, see Berlin *et al*. (2008) and Quintela *et al*. (2010). Primers for multiplex were designed for substitutions flanked by regions of at least 100 bp on

both sides (for SNPs and flanking sequences, see Supporting Information, Table S3).

Analysis of SNP markers

The analysis with the SNP set of markers was performed using two different approaches:

1. Geographical approach: samples were divided into four groups based on the geographical origin of the individuals: Russia ($N = 7$), Scandinavia ($N = 51$), Alaska–Canada ($N = 17$), and the British Isles ($N = 5$).
2. Taxonomic approach: individuals were divided into six groups with respect to the subspecies: *lagopus* ($N = 63$), *alexandrae* ($N = 7$), *muriei* ($N = 5$), *kamschatkensis* ($N = 1$), *scoticus* ($N = 4$), and *hibernicus* ($N = 1$).

Both basic statistics and pairwise F_{ST} were calculated with GENALEX, where the significance of the latter was based on 10 000 permutations. The use of single summary statistics such as F_{ST} (Weir & Cockerham, 1984) or Nei's D (Nei, 1972, 1978) was not considered sufficient to adequately capture interpopulation relationships (Dyer & Nason, 2004). Because the genetic interrelation between populations is measured upon their genotypes at a number n of loci with m independent combination of alleles, each population can be represented in a multidimensional space consisting of m orthogonal axes. To summarize these data in a comprehensible format and better understand the structure of the set of populations investigated in the present study, we performed multivariate analysis using the R package ADEGENET for the individuals genotyped at SNPs (Jombart, 2008).

To further investigate population structure, we identified genetic clusters with the Bayesian model-based clustering algorithms implemented in STRUCTURE, version 2.3.1, under a model assuming admixture and correlated allele frequencies using population information (LOCPRIOR option). Ten runs with a burn-in period consisting of 100 000 replications and a run length of 1 000 000 Markov chain Monte Carlo iterations were performed for a number of K clusters, ranging from 1 to 15 depending on the approach. We then applied the Evanno, Regnaut & Goudet (2005) *ad hoc* summary statistic ΔK, which is based on the rate of change of the 'estimated likelihood' between successive K-values. Simulations indicate that the K-value with a higher ΔK corresponds to the uppermost hierarchical level of population structure (Evanno *et al.*, 2005). Furthermore, we compared the posterior probabilities for the values of K with the highest $P(X|K)$ using a Wilcoxon two-samples test *sensu* Rosenberg *et al.* (2001). Runs were averaged using CLUMPP, version 1.1.1 (Jakobsson & Rosenberg, 2007) using the large K greedy algorithm and the G' pairwise matrix similarity statistics and results were visualized as barplots. STRUCTURE analyses were performed: (1) by adding prior information about the individuals' geographical sampling area; (2) according to subspecies assignment; (3) according to sampling site; and (4) without any priors.

Mt control region

Sequence assembly and alignment

The 1089-bp mitochondrial control region was amplified using primers and methods *sensu* Baba, Fujimaki & Koike (2001). Resulting products were purified using ExoSAP-IT (GE Healthcare) in accordance with the manufacturer's instructions. Automatic sequencing was performed on a MegaBace 1000 (GE Healthcare).

The sequence assembly of the mitochondrial control region sequences was performed using CODONCODE ALIGNER (CodonCode Corporation). The sequences were then aligned using CLUSTALW (Thompson, Higgins & Gibson, 1994) in CODONCODE ALIGNER and refined by eye. Sequences with many gaps as a result of missing data were removed. All unique sequences were deposited in GenBank (accession codes: JX274545–JX274589).

Phylogenetic analysis

Conventional F-statistics and pairwise F_{ST} were computed from haplotype frequencies and statistical significance was based on 10 100 permutations using ARLEQUIN (Excoffier, 2006).

The haplotype phylogeny of willow grouse was reconstructed using four different methods: Neighbour joining, maximum parsimony (MP), maximum likelihood (ML), and Bayesian inference. We choose four different phylogenetic reconstruction methods to check that relationships are robust and not artefacts of any given phylogenetic reconstruction method. Three rock ptarmigan (*L. muta*) control region sequences were used as outgroup taxa.

Neighbour-joining analyses were performed in MEGA, version 4 (Tamura *et al.*, 2007). Because the variation of intraspecific sequences might be low, the Jukes–Cantor estimate of the number of nucleotide substitutions per site (d) was employed to: (1) avoid incorrect topology resulting from the larger variance introduced by complex model because, if d is less than 0.05, the distance model of p-distance or Jukes–Cantor distance should be used (Nei & Kumar, 2000); (2) evaluate the nodal support of the Neighbour-joining tree because, if d is not bigger than 0.5, the confidence value of bootstrap test (Felsenstein, 1985)

and the interior branch test (Nei, Stephens & Saitou, 1985) is nearly the same and, for closely-related sequences, the interior branch test is more mathematically rigorous (Sitnikova, Rzhetsky & Nei, 1995; Nei & Kumar, 2000). The test was performed for 1000 iterations and a 50% majority-rule was used to generate the consensus tree.

Maximum parsimony analyses were conducted in PAUP* 4.0b10 (Swofford, 2002) using a heuristic search and the stepwise addition option. Sequences were added randomly with ten replicates and the tree bisection–reconnection algorithm was used for branch swapping. Maximum likelihood analyses were performed in PHYML, version 2.4.4 (Guindon & Gascuel, 2003), using the best-fit nucleotide substitution model selected by the Akaike information criterion in MODELTEST, version 3.6 (Posada & Crandall, 1998). The bootstrap method (1000 replicates) was used to evaluate the nodal support of the MP and the ML trees. Bayesian inference was conducted using MrBayes, version 3.1.2 (Huelsenbeck & Ronquist, 2001). The best-fit model of nucleotide substitution was selected in MRMODELTEST, version 2.2 (Nylander, 2004) following the Akaike information criterion. The analysis was run for 500 000 generations, with four parallel chains and a sampling frequency of every 100 generation. The burn-in point at which the sampling of the trees reached a stationary state was estimated using TRACER, version 1.5 (Rambaut & Drummond, 2009). Samples before the stationary point were discarded.

The phylogenetic network analysis was performed in NETWORK, version 4.5.1.0 (http://www.fluxus-engineering.com). The reduced median network option was applied to identify obvious parallel mutations. Then the median-joining method (with default settings) was used to calculate the MP network of the refined dataset (Bandelt, Forster & Röhl, 1999).

RESULTS

SNP MARKERS

The conversion rate of SNP markers was 70.83% (17 out of the 24 multiplexed primers successfully yielded a product) with 100% of reproducibility according to duplicate analysis of 7% of the genotypes. One out of 17 approved assays, CAAX-124 (1.33%), was monomorphic. Summary statistics for all SNPs are provided in the Supporting Information (Table S4).

Geographical approach

Pairwise F_{ST} (Table 2) values ranged from 0 (Russia versus Alaska–Canada) to 0.072 (UK versus Alaska–Canada). The only pairwise comparisons that showed significant structure were North America versus Scandinavia, and the UK. Analyses performed with principal component analysis (PCA) (Fig. 2), pairwise F_{ST}, and STRUCTURE using population information (Fig. 4A) yielded very similar results for most of the pairwise comparisons. The only discrepancies were found in the comparison for Russia–Scandinavia where STRUCTURE detected differences that went unnoticed by PCA (some overlap of the respective 95% inertia ellipses) but mainly for pairwise F_{ST} (0.014, $P > 0.05$). The second disagreement was found in the PCA for Scandinavia–Ireland_UK, where the overlap of the 95% inertia ellipses did not agree either with the nonsignificant F_{ST} (0.007, $P > 0.05$) or with the high similarity in the inferred membership of individuals from both groups according to STRUCTURE. Posterior probabilities for different K values under

Table 2. Single nucleotide polymorphism markers: pairwise F_{ST} between geographical regions and subspecies calculated with GENEALEX (below diagonal) with probability values based on 9999 permutations (above diagonal)

Geographical regions	Scandinavia	Russia	Alaska–Canada	UK–Ireland
Scandinavia	–	0.175	0.000	0.318
Russia	0.014	–	0.429	0.106
Alaska–Canada	**0.067**	0.000	–	0.031
UK–Ireland	0.007	0.048	*0.072*	–
Subspecies	*Lagopus*	*Alexandrae*	*Muriei*	*Scoticus*
Lagopus	–	0.002	0.089	0.334
Alexandrae	**0.141**	–	0.021	0.004
Muriei	0.065	*0.146*	–	0.012
Scoticus	0.016	**0.249**	*0.240*	–

Values shown shown in bold are significantly higher than zero after Bonferroni correction ($\alpha = 0.0083$). Values in italics are significant at $P < 0.05$.

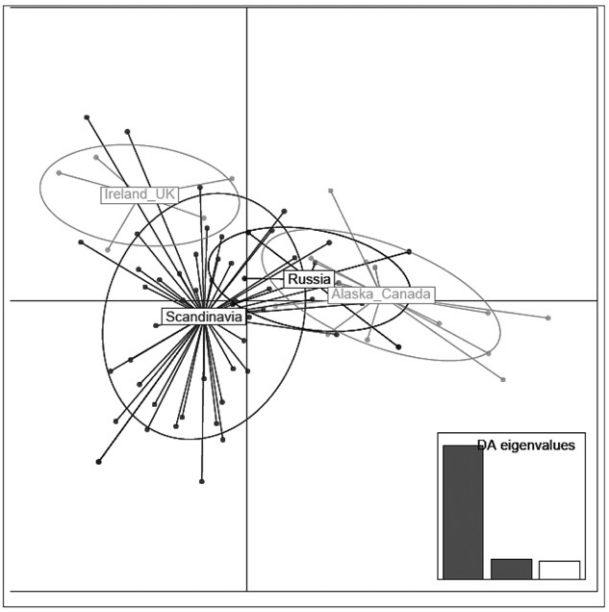

Figure 2. Topology of the four geographical regions obtained by principal component analysis. Eigenvalues corresponding to the represented components are filled in black. Points represent the genotypes; geographical groups are labelled inside their 95% inertia ellipses. The percentage of explained variance was 77.65% and 11.64% for axes 1 and 2, respectively.

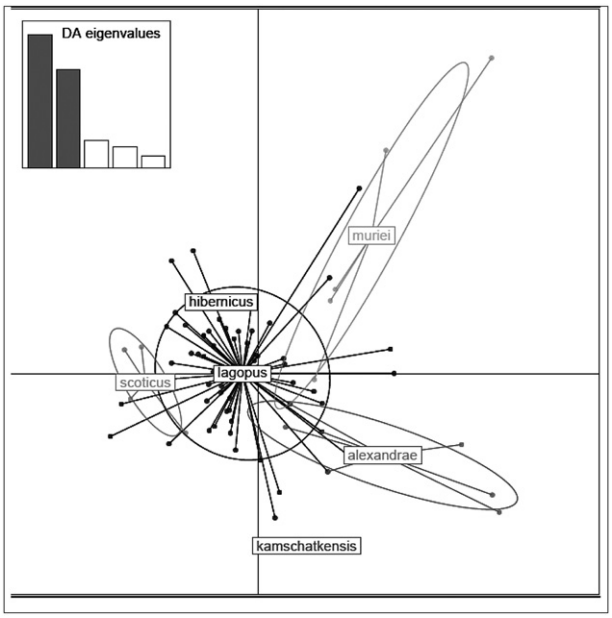

Figure 3. Topology of the six subspecies obtained by principal component analysis. Eigenvalues corresponding to the represented components are filled in black. Points represent the genotypes; subspecies are labelled inside their 95% inertia ellipses. The percentage of explained variance was 45.44% and 33.50% for axes 1 and 2, respectively.

the geographical approach (and subsequently for the taxonomic and sample site approaches) are provided in the Supporting Information (Tables S5 and S6).

Taxonomic approach

Pairwise F_{ST} (Table 2) ranged from 0.016 (*lagopus* versus *scoticus*) to 0.249 (*scoticus* versus *alexandrae*). All values were significantly different from zero, except for *lagopus* versus *scoticus* (0.016, $P > 0.05$) and *lagopus* versus *muriei* (0.065, $P > 0.05$). Although these results were largely supported by PCA (Fig. 3), it is worth noting that the degree of overlap between the aforementioned subspecies was higher than expected, also taking into account that they shared a similar average proportion of membership to the same STRUCTURE cluster (Fig. 4B).

STRUCTURE analyses performed considering the site of sampling site as a prior yielded two main clusters (Fig. 4C). Thus, individuals from the following locations showed more than 0.75 of proportion of membership to cluster I: site 1 (Ireland), site 9 (Naryan Mar and Ostrov Kolguyev, North-West Russia), site 10 (Magadan, Eastern Siberia), site 11 (Atqasuk, North Alaska), site 12 (Heceta Island, South-East Alaska), site 13 (Zarembo Island, South-East Alaska), and site 14 (Chilkat Pass, North-West BC, Canada), whereas individuals from site 2 (Orkney, South-York-Dales, Islay_Scotland), site 4 (Vålådalen, Tjallingen, Sweden), and site 5 (Tjuoltadalen, Sweden) showed a proportion of membership higher than 0.75 to cluster II.

STRUCTURE analyses performed without prior information yielded three main clusters. However, when plotting individual membership proportions to each cluster, we could not detect any geographical or taxonomic signal and individuals were all assigned to each cluster with equal probability.

Mitochondrial sequences

Sequence variation

The final data matrix included 41 individual 1089-bp long sequences (Table S2) corresponding to nucleotides 61 to 1140 of the complete mitochondrial control region of willow grouse (based on a comparison with GenBank accession number AJ297169 (Lucchini *et al*., 2001). We found 42 polymorphic sites and 30 haplotypes. Within the nominal subspecies, one haplotype was shared by two individuals from Scandinavia and one from Russia. Similarly, in Alaska, another haplotype was shared by two specimens of *lagopus* and one specimen of *muriei*. Descriptive statistics of genetic diversity for geographical regions and subspecies, respectively, are given in Table 3. The geographical approach for genetic differentiation based on pairwise F_{ST} for mtDNA

(Table 4A) was consistent with the results obtained with SNP markers, with the exception of the comparison for Scandinavia–Britain, which suggested monophyly for the British birds for mtDNA (Fig. 5). However, when considering the taxonomic approach, discrepancies between both markers appeared to arise. Thus, despite the high values found for pairwise F_{ST} using mtDNA, the low sample sizes hampered revealing any statistical significance (Table 4B). Therefore, it is necessary to interpret these results with caution because the apparent lack of structure between taxa can be spurious.

Phylogeny

The Jukes–Cantor estimate of the mean ± SE number of nucleotide substitutions per site (d) was 0.015 ± 0.002. Therefore, the p-distance model was used to reconstruct a Neighbour-joining tree (Nei & Kumar, 2000), and the interior branch method was selected to test the reliability of the tree (Sitnikova *et al*., 1995). For the ML analysis, the best-fit substitution model was HKY+I+G (base frequencies: A, 0.3319, C, 0.1471, G, 0.2575, T, 0.2635; transition/transversion ratio = 3.2187; proportion of invariable site = 0.8386; gamma shape parameter = 0.9443). For the Bayesian inference, the best-fit substitution model was HKY+I+G (proportion of invariable site = 0.8386; gamma shape parameter = 0.9443). The sampling of the trees became stationary after 460 generations. After discarding the initial burn-in, the remaining 4500 generations were used to determine the posterior probability distribution.

The Neighbour-joining tree clearly suggested that willow grouse haplotypes are separated into two major clades. One clade consists of sequences from the British Isles and Scandinavia, whereas the second clade consists of sequences from Russian and North-American willow grouse (Fig. 5). The other phylogenetic trees using MP, ML, and Bayesian methods resulted into a tree topology similar to that of Neighbour-joining method, although with lower support values (see Supporting Information, Figs S1–S3). The median network analyses suggested a similar grouping with clear geographical separation. Scandinavian and British haplotypes were separated but, together, were distinct from those from Russia and North America. North American and Russian sequences also showed internal structure (Fig. 6).

DISCUSSION

A major finding of the present study is that both the nuclear and mitochondrial data suggest that willow grouse are separated into two main genetic groups: one in the western part of the range, including the British Isles and Scandinavia, and another from Russia eastwards in North America. However, we were unable to obtain samples from eastern North America and southern Russia and so many described subspecies were not included in the present study. It is thus possible that further geographical structuring is present within this species. We did find some further geographical structuring in the mitochondrial data in red grouse from the British Isles as they clustered together in a clade with some but not all sequences from Scandinavia. Divergence between Scandinavian and British red/willow grouse was confirmed previously using nuclear markers (Quintela *et al*., 2010). This east–west subdivision is in disagreement with taxonomic classifications in that the widespread continental subspecies is paraphyletic.

The east–west separation reported in the present study approximately corresponds to the eastern boundary of the European ice sheet during the last glacial maximum 20 000 years BP (Hewitt, 2000). We can therefore tentatively suggest that Western Europe and Scandinavia were recolonized from a refugium that was distinct from the unglaciated Beringial region in Eastern Siberia and Alaska that might have served as the refugium for Russian and North American lineages.

Dating the split between the western and eastern clade of *lagopus* is difficult because molecular clock calibrations for the avian Control Region revealed varying results (from 0–38% divergence per 1 million years in different avian genera (Ruokonen & Kvist, 2002). In the present study, we report 0.015 substitutions/site. A 2% divergence corresponds to mutations accumulating at a rate of approximately one every 3500 years. This would give an estimate of 50 000 years for the timing of the split between the western and eastern clade of willow grouse. If the control region evolves five- to ten-fold faster (Ruokonen & Kvist, 2002), we obtain an estimate of 5000–10 000 years ago. This approximately corresponds in time to the retreat of the inland ice in Europe since the Last Glacial Maximum.

Recent analyses using Bayesian estimation of species trees suggested that Scandinavian willow and British red grouse coalesce before the Pleistocene glaciations (more than 20 000 years ago) (Quintela *et al*., 2010). This clear separation between these taxa was found by using 76 unlinked SNPs harboured in 13 protein-coding loci. In the present study, using some of the same SNPs but including a wider sampling and STRUCTURE analyses, we could not confirm this separation. We suggest that this is most likely because of differences among the studies with respect to the SNP typing method. In the present study, we could not use all 76 SNPs of the previous study in which we employed a direct sequencing

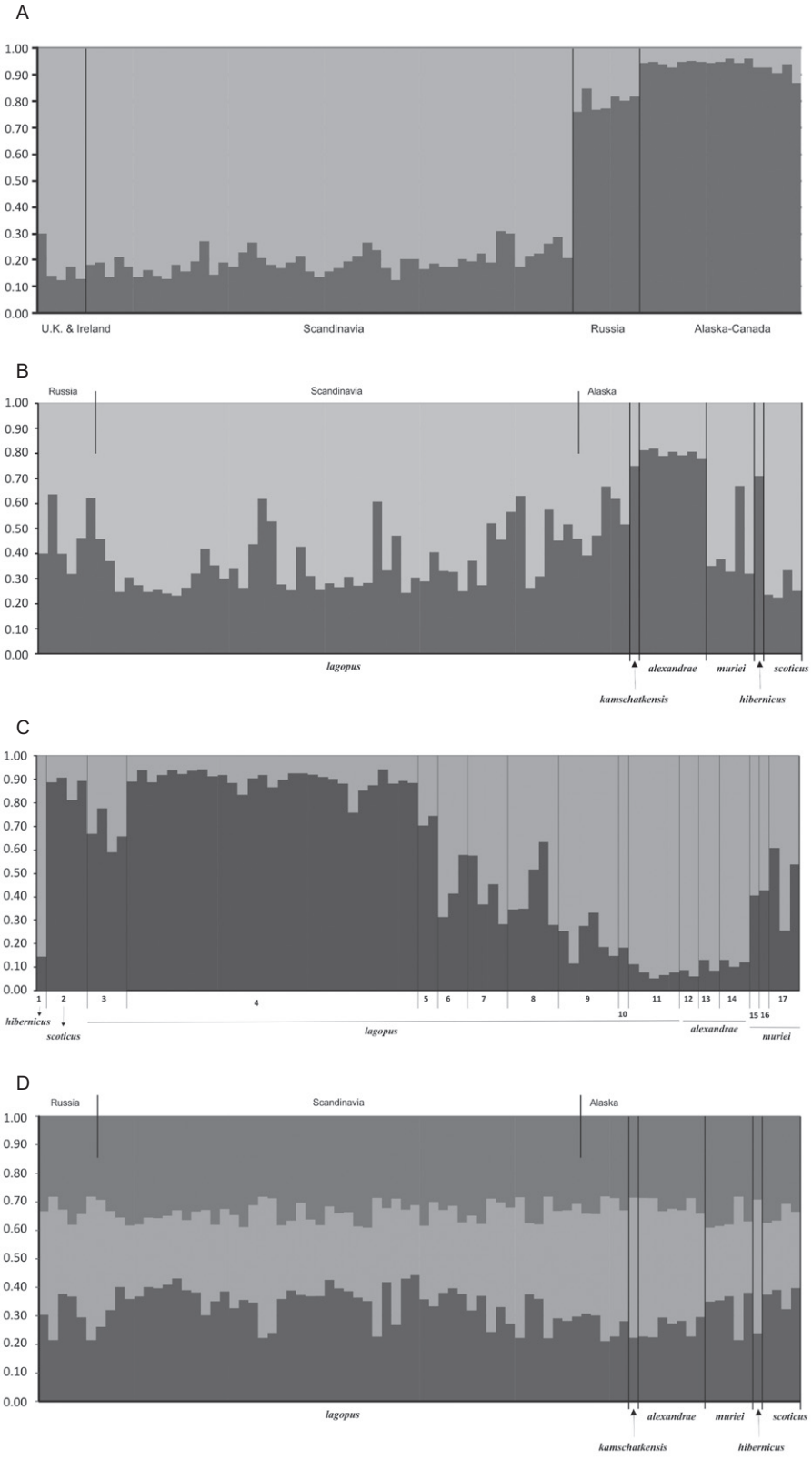

Figure 4. Inferred ancestry assessed with STRUCTURE using single nucleotide polymorphism markers under different approaches (for details, see Material and methods). A, geographical approach: individuals were grouped into their four geographical regions of origin. B, taxonomic approach: individuals were divided into six subspecies (individuals of the nominal subspecies were ordered in the barplot in accordance with the geographical origin of the samples. C, sampling sites approach (with the taxonomic adscription included). Sites correspond to: 1, Ireland; 2, Orkney, South-Yorkshire Dales, Islay, Scotland; 3, Kaiseloukta, Kaisejåkkå, Lulep, Sarta, Sweden; 4, Vålådalen, Tjallingen, Sweden; 5, Tjuoltadalen, Sweden; 6, Jämtland, Sweden; 7, Ylivinsa, Sweden; 8, Norway; 9, Naryan Mar and Ostrov Kolguyev, North-West Russia; 10, Magadan, East Siberia; 11, Atqasuk, North Alaska; 12, Heceta Island, South-East Alaska; 13, Zarembo Island, South-East Alaska; 14, Chilkat Pass, North-West BC, Canada; 15, Schumagin Island, Alaska; 16, Maclaren River Valley Alaska; 17, South-West Birdsall Island, Izembek National Wildlife Refuge. D, no priors.

Table 3. Descriptive statistics of mitochondrial DNA genetic diversity for geographical regions (top) and subspecies (bottom)

	Sample size	Haplotype diversity (h)	Nucleotide diversity (π)	Number of pairwise differences (K)
Geographical regions				
Britain	7	0.857	0.00167	1.810
Scandinavia	14	0.934	0.00355	3.857
Russia	5	1.000	0.00267	2.900
Alaska–Canada	14	0.956	0.00533	5.780
Subspecies				
Lagopus lagopus scoticus	7	0.857	0.00167	1.810
Lagopus lagopus lagopus	24	0.975	0.00590	6.399
Lagopus lagopus alexandrae	5	0.900	0.00756	8.200
Lagopus lagopus muriei	4	1.000	0.00230	2.500

Table 4. Mitochondrial DNAcontrol region: pairwise F_{ST} calculated with ARLEQUIN and computed from haplotype frequencies between geographical regions and subspecies (below diagonal) and probability values based on 10 100 permutations (above diagonal)

Geographical regions	Scandinavia	Russia	Alaska–Canada	UK–Ireland
Scandinavia	–	0.3661	0.0020	0.0072
Russia	0.0129	–	0.1843	0.1867
Alaska–Canada	**0.0550**	0.0244	–	0.0143
UK–Ireland	**0.1004**	0.0736	**0.0884**	–
Subspecies	*Lagopus*	*Alexandrae*	*Muriei*	*Scoticus*
Lagopus	–	0.0255	0.5708	0.0041
Alexandrae	**0.0581**	–	0.4449	0.0714
Muriei	0.0000	0.0531	–	0.2292
Scoticus	**0.0783**	0.1231	0.0812	–

Values shown shown in bold are significantly higher than zero.

method. Instead, we applied the SNPStream typing method and were able to design primers for only a part of the SNPs; furthermore, we lost SNPs as a result of unreliable typing. Thus, given that the analyses of the present study are based on a smaller number of SNPs, the results should be interpreted with some caution.

For the mitochondrial data, we found that Britain and Scandinavia formed one clade and Russia and North America formed another both in the phylogenetic reconstructions and in the haplotype network analyses. These results were confirmed by STRUCTURE analyses of the nuclear SNP data when using geographical origin as a prior. With the taxonomic

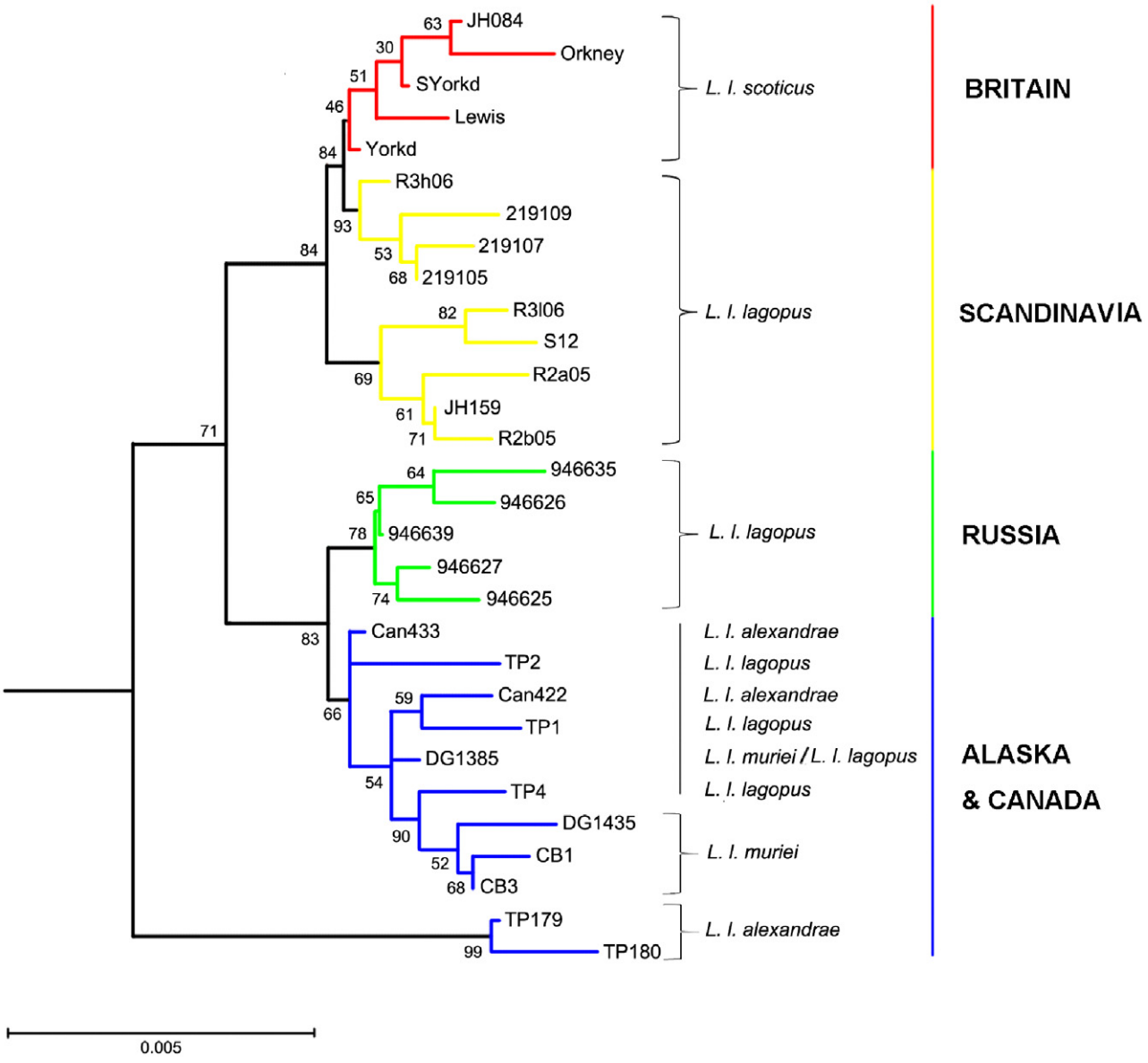

Figure 5. Neighbour-joining tree of mitochondrial control region haplotypes of willow grouse. Subspecies names and phylogenetic clades are indicated to the right-hand side. Rock ptarmigan served as an outgroup and connected at the base of the tree (not shown). The support value of the nodes is from the interior branch test (1000 replicates). The haplotype DG1385 includes the sequences of individuals from two subspecies, *Lagopus lagopus lagopus* and *Lagopus lagopus muriei*. Branch lengths are proportional to the number of substitutions per site (scale bar = 0.005 substitutions/site).

approach, we could only detect a weak signal for *alexandrae* when $K = 2$. The mitochondrial data further suggested that the British birds formed a clade within the partly Scandinavian–British clade, confirming the monophyly of red grouse.

There might be a number of reasons for the relatively poor taxonomic signal in the SNP data. First, as noted above, our analyses are based on a limited number of SNPs and taxon sampling. Although we know that the SNPs included here segregate independently as a result of a large effective population size and high recombination rates (Berlin *et al.*, 2008), it might be necessary to include a larger number of SNPs before any firm conclusions can be drawn. Second, subspecies classifications might not correspond with genetic differences among populations. Rather, similar habitat and microclimate regimes might produce similar phenotypes, although the sampling locations were far apart. Studies of the closely-related blue grouse (Barrowclough *et al.*, 2004)

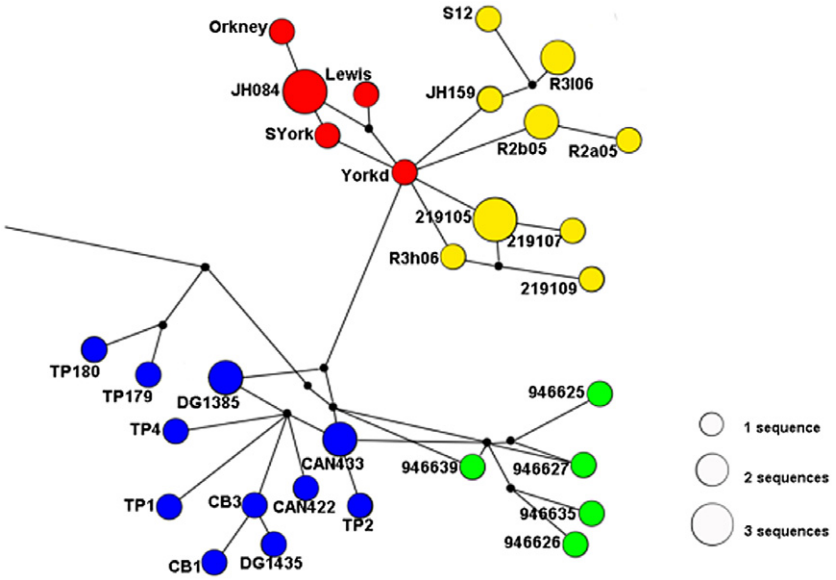

Figure 6. Median-joining network (MJN) of mitochondrial control region haplotypes of willow grouse. Red, Britain; yellow, Scandinavia; green, Russia; blue, Alaska and Canada. Black, median vectors (mv) that represent hypothetical missed or unsampled ancestral haplotypes; purple, rooting taxa. The size of the circle is proportional to the number of individuals that share the haplotype.

and sage grouse (Young *et al*., 2000) revealed further substructuring among recognized subspecies, suggesting that further genetic structuring beyond the subspecies level may be common among grouse species.

In a large outbreeding population such as the willow grouse, we would expect isolation by distance and, indeed, this was previously found in Scandinavia (Quintela *et al*., 2010). A possible consequence of isolation-by-distance is clinal variation in morphological traits. This would call for caution when assigning subspecies status to local populations within a continuous range. Thus, a few of the widespread subspecies with no clear geographical distinction from neighbouring subspecies (such as *lagopus* and *alexandrae*) would be predicted to have high levels of gene flow and low genetic differentiation. By contrast, island subspecies and otherwise isolated subspecies (such as *scoticus* and *hibernicus*) would be predicted to show limited gene flow with other subspecies and be more genetically distinct. Our mitochondrial network analyses confirmed these patterns to some extent, whereas more extensive nuclear data are required to address this issue in earnest. This also corresponds with the conservation status of willow grouse subspecies. Although willow grouse represents a common species throughout most of its range, some populations and subspecies inhabiting islands have been are classified as locally threatened, such as red grouse on Ireland (*Lagopus hibernicus*; McMahon *et al*., 2012). However, our data suggest that none of these threatened populations, nor the large continental subspecies would classify as reciprocally monophyletic for mtDNA, which has been suggested to be a criterion for assigning ESU-status to subspecies (Moritz, 1994). However, these threatened island populations probably classify as management units (Palsbøll, Bérubé & Allendorf, 2007).

In conclusion, we found an east–west division among the sampled willow grouse subspecies. This division cut across the range of one of the subspecies (*lagopus*) with the largest distribution and thus contradicted existing taxonomic classifications. In North America, we could not detect any major genetic divisions apart from two birds described as *alexandrae* from the Heceta Island (Alaska) when using mitochondrial sequences. The other samples from North America were intermingled regardless of whether they were described as *muriei*, *alexandrae* or *lagopus*. The specimens described as *alexandrae* were to some extent distinct when analysing the SNP data. Even though the fit between taxonomy and genetics might be considered poor, the genetic analyses indicated some concordance between genetics and taxonomy, although not complete congruence. This was particularly evident for mtDNA network analyses. It is evident that the taxonomy of this species would benefit by a careful re-examination of the available evidence for subspecies designation. More extensive sampling among and within putative subspecies and more nuclear markers might help to resolve some of the uncertainties discerned by the present study, although it appears as if subspecies status is a poor

proxy for assigning ESUs (Moritz, 1994) and management units (Palsbøll *et al*., 2007) in this species.

ACKNOWLEDGEMENTS

The study was supported from a grant from the Swedish Research Council (to J.H.). We thank Eleanor Jones for kindly commenting on a previous version of the manuscript. We thank John A. Allen and two anonymous reviewers for their helpful comments.

REFERENCES

Baba Y, Fujimaki Y, Koike H. 2001. Genetic variability in the mitochondrial control region of the Japanese rock ptarmigan *Lagopus mutus japonicus*. *Japanese Journal of Ornithology* **50:** 53–64.

Babik W, Branicki W, Crnobrnja-Isailovic J, Cogalniceanu D, Sas I, Olgun K, Poyarkov NA, Garcia-Paris M, Arntzen JW. 2005. Phylogeography of two European newt species – discordance between mtDNA and morphology. *Molecular Ecology* **14:** 2475–2491.

Ballard JWO, Whitlock MC. 2004. The incomplete natural history of mitochondria. *Molecular Ecology* **13:** 729–744.

Bandelt HJ, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16:** 37–48.

Barrowclough GF *et al*. 2004. Phylogeographic structure, gene flow and species status in blue grouse (*Dendragapus obscurus*). *Molecular Ecology* **13:** 1911–1922.

Beaumont MA, Balding DJ. 2004. Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology* **13:** 969–980.

Bell PA, Chaturvedi S, Gelfand CA, Huang CY, Kochersperger M, Kopla R, Modica F, Pohl M, Varde S, Shao R, Zhao X, Boyce-Jacino MT. 2002. SNPstream® UHT: ultra-high throughput SNP genotyping for pharmacogenomics and drug discovery. *BioTechniques* **32:** S70–S77.

Berlin S, Quintela M, Höglund J. 2008. A multilocus assay reveals high nucleotide diversity and limited differentiation among Scandinavian willow grouse (*Lagopus lagopus*). *BMC Genetics* **9:** 89.

Bernatchez L, Glemet H, Wilson CC, Danzmann RG. 1995. Introgression and fixation of Arctic char (*Salvelinus alpinus*) mitochondrial genome in an allopatric population of brook trout (*Salvelinus fontinalis*). *Canadian Journal of Fisheries and Aquatic Sciences* **52:** 179–185.

Crandall KA, Bininda-Emonds ORP, Mace GM, Wayne RK. 2000. Considering evolutionary processes in conservation biology. *Trends in Ecology & Evolution* **15:** 290–295.

Dyer RJ, Nason JD. 2004. Population graphs: the graph theoretic shape of genetic structure. *Molecular Ecology* **13:** 1713–1727.

Edwards SV, Beerli P. 2000. Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution* **54:** 1839–1854.

Edwards SV, Liu L, Pearl DK. 2007. High-resolution species trees without concatenation. *Proceedings of the National Academy of Sciences of the United States of America* **104:** 5936–5941.

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

Excoffier L. 2006. *Arlequin Version 3.1: An Integrated software package for Population genetics*. Bern: University of Bern.

Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39:** 783–791.

Freeland J, Andersson S, Allen D, Looney D. 2007. Museum samples provide novel insights into the taxonomy and genetic diversity of Irish red grouse. *Conservation Genetics* **8:** 695–703.

Gantenbein B, Largiadèr CR. 2003. The phylogeographic importance of the Strait of Gibraltar as a gene flow barrier in terrestrial arthropods: a case study with the scorpion *Buthus occitanus* as model organism. *Molecular Phylogenetics and Evolution* **28:** 119–130.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52:** 696–704.

Gyllensten U. 1985. Temporal allozyme frequency changes in density fluctuating populations of willow grouse (*Lagopus lagopus L*). *Evolution* **39:** 115–121.

Hannon SJ, Martin K, Eason PK. 1998. Willow ptarmigan. In: Poole A, Gill F, eds. *The birds of North America*. 369. Philadelphia, PA: The birds of North America.

Hewitt G. 2000. The genetic legacy of the Quaternary ice ages. *Nature* **405:** 907–913.

del Hoyo J, Elliot A, Sargatal J. 1994. *Handbook of the birds of the world,* Vol. 2. *New World vultures to guineafowl*. Barcelona: Lynx Ediciones.

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17:** 754–755.

Jakobsson M, Rosenberg NA. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23:** 1801–1806.

Johnsgard PA. 1983. *The grouse of the world*. Lincoln, NE: University of Nebraska Press.

Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24:** 1403–1405.

Juan C, Oromi P, Hewitt GM. 1996. Mitochondrial DNA sequence variation and phylogeography of *Pimelia* darkling beetles on the Island of Tenerife (Canary Islands). *Heredity* **77:** 589–598.

Lucchini V, Höglund J, Klaus S, Swensson J, Randi E. 2001. Historical biogeography and a mitochondrial DNA phylogeny of grouse and ptarmigan. *Molecular Phylogenetics and Evolution* **20:** 149–162.

Maddison WP. 1997. Gene trees in species trees. *Systematic Biology* **46:** 523–536.

McMahon BJ, Johansson MP, Piertney SP, Buckley K, Höglund J. 2012. Genetic variation among endangered Irish red grouse (*Lagopus lagopus hibernicus*) populations: implications for conservation and management. *Conservation Genetics* **13:** 639–647.

Moritz C. 1994. Applications of mitochondrial-DNA analysis in conservation: a critical review. *Molecular Ecology* **3:** 401–411.

Nei M. 1972. Genetic distance between populations. *American Naturalist* **106:** 283–292.

Nei M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89:** 583–590.

Nei M, Kumar S. 2000. *Molecular evolution and phylogenetics*. Oxford: Oxford University Press.

Nei M, Stephens JC, Saitou N. 1985. Methods for computing the standard errors of branching points in an evolutionary tree and their application to molecular data from humans and apes. *Molecular Biology and Evolution* **2:** 66–85.

Nylander JAA. 2004. *MrModeltest*, Version 2.2. Program distributed by the author. Uppsala: Evolutionary Biology Centre, Uppsala University.

Palsbøll PJ, Bérubé M, Allendorf FW. 2007. Identification of management units using population genetic data. *Trends in Ecology & Evolution* **22:** 11–16.

Pamilo P, Nei M. 1988. Relationships between gene trees and species trees. *Molecular Biology and Evolution* **5:** 568–583.

Paxton RJ, Thorén PA, Tengö J, Estoup A, Pamilo P. 1996. Mating structure and nestmate relatedness in a communal bee, *Andrena jacobi* (Hymenoptera, Andrenidae), using microsatellites. *Molecular Ecology* **5:** 511–519.

Posada D, Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14:** 817–818.

Potapov RL. 1985. *Fauna of the USSR*. Leningrad: Soviet Acedemy of Sciences (in Russian).

Potapov RL, Flint VE. 1989. *Handbuch der Vögel der Sowjetunion. Band 4 Galliformes, Gruiformes*. Lutherstadt Wittenberg: Ziemsen Verlag.

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

Quintela M, Berlin S, Wang B, Höglund J. 2010. Genetic diversity and differentiation among *Lagopus lagopus* populations in Scandinavia and Scotland: evolutionary significant units confirmed by SNP markers. *Molecular Ecology* **19:** 2380–2393.

Rambaut A, Drummond AJ. 2009. *Tracer: MCMC Trace Analysis Tool*, Version 1.5.0. Available at: http://tree.bio.ed.ac.uk/software/tracer/

Relethford JH. 2004. Boas and beyond: migration and craniometric variation. *American Journal of Human Biology* **16:** 379–386.

Rosenberg NA, Burke T, Elo K, Feldman MW, Freidlin PJ, Groenen MAM, Hillel J, Mäki-Tanila A, Tixier-Boichard M, Vignal A, Wimmers K, Weigend S. 2001. Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds. *Genetics* **159:** 699–713.

Ruokonen M, Kvist L. 2002. Structure and evolution of the avian mitochondrial control region. *Molecular Phylogenetics and Evolution* **23:** 422–432.

Sitnikova T, Rzhetsky A, Nei M. 1995. Interior-branch and bootstrap tests of phylogenetic trees. *Molecular Biology and Evolution* **12:** 319–333.

Skoglund P, Höglund J. 2010. Sequence polymorphism in candidate genes for differences in winter plumage between Scottish and Scandinavian willow grouse (*Lagopus lagopus*). *PLoS ONE* **5:** e10334.

Storch I. 2007. Conservation status of grouse worldwide: an update. *Wildlife Biology* **13:** 5–12.

Storz JF. 2002. Contrasting patterns of divergence in quantitative traits and neutral DNA markers: analysis of clinal variation. *Molecular Ecology* **11:** 2537–2551.

Swofford DL. 2002. *PAUP*. Phylogenetic Analysis Using Parsimony (and Other Methods)*. Sunderland, MA: Sinauer Associates.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. *Molecular Biology and Evolution* **24:** 1596–1599.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22:** 4673–4680.

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

Young JR, Braun CE, Oyler-McCance SJ, Hupp JW, Quinn TW. 2000. A new species of sage-grouse (Phasianidae: *Centrocercus*) from Southwestern Colorado. *The Wilson Bulletin* **112:** 445–453.

SUPPORTING INFORMATION

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

Figure S1. Maximum parsimony tree of mitochondrial control region haplotypes of willow grouse (maximum parsimony tree): length (L) = 109, consistency index (CI) = 0.789 and retention index (RI) = 0.917.
Figure S2. Maximum likelihood tree of mitochondrial control region haplotypes of willow grouse.
Figure S3. Bayesian tree of mitochondrial control region haplotypes of willow grouse.
Table S1. Information about sampling location, subspecies classification of the studied individuals, and source for DNA extraction.

Table S2. List of names of the mitochondrial control region haplotypes of willow grouse.
Table S3. Single nucleotide (SNP) positions (substitutions noted in brackets in bold) and flanking regions used to design the twenty-four multiplexed primers (bold and italics indicate the codes of those that were successfully amplified). The SNP code consists of the exon name and the position of the substitution (bp).
Table S4. Summary statistics for SNPs (mean ± SE) calculated with GENALEX using geographical and taxonomic approaches, respectively: observed heterozygosity (H_O), unbiased expected heterozygosity (UH_E), number of alleles (A), and proportion of polymorphic loci.
Table S5. Posterior probabilities for different values of K under four approaches of STRUCTURE data analysis.
Table S6. Mean ± SD value of r after 10 runs per K with different STRUCTURE analysis approaches.