JOURNAL OF Evolutionary Biology

doi: 10.1111/jeb.12491

Evolutionary neutrality of mtDNA introgression: evidence from complete mitogenome analysis in roe deer

M. MATOSIUK*, I. N. SHEREMETYEVA†, I. S. SHEREMETYEV†, A. P. SAVELJEV‡ & A. BORKOWSKA*

*Institute of Biology, University of Bialystok, Bialystok, Poland
†Institute of Biology and Soil Science, Far East Branch, Russian Academy of Sciences, Vladivostok, Russia
‡Department of Animal Ecology, Russian Research Institute of Game Management and Fur Farming, Russian Academy of Sciences, Kirov, Russia

Keywords:

Capreolus;
hybridization;
introgression;
mitogenomics;
natural selection;
roe deer.

Abstract

Introgressive hybridization offers a unique platform for studying the molecular basis of natural selection acting on mitogenomes. Most of the mtDNA protein-coding genes are extremely conserved; however, some of the observed variations have potentially adaptive significance. Here, we evaluated whether the evolution of mtDNA in closely related roe deer species affected by widespread mtDNA introgression is neutral or adaptive. We characterized and compared 16 complete mitogenomes of European (*Capreolus capreolus*) and Siberian (*C. pygargus*) roe deer, including four of Siberian origin introgressed into European species. The average sequence divergence of species-specific lineages was estimated at 2.8% and varied across gene classes. Only 21 of 315 fixed differences identified in protein-coding genes represented nonsynonymous changes. Only three of them were determined to have arisen in the *C. pygargus* lineage since the time to the most recent common ancestor (TMRCA) of both *Capreolus* species, reflecting a decelerated evolutionary ratio. The almost four-fold higher d_N/d_S ratio described for the European roe deer lineage is constrained by overall purifying selection, especially pronounced in the *ND4* and *ND5* genes. We suggest that the highly divergent *C. capreolus* lineage could have maintained a capability for genomic incorporation of the well-preserved and almost ancestral type of mtDNA present in *C. pygargus*. Our analyses did not indicate any signs of positive selection for Siberian roe deer mtDNA, suggesting that the present widespread introgression is evolutionarily neutral.

Introduction

A task of evolutionary biology is to understand how the evolution of lineages proceeds within and across species through mechanisms of natural selection and genetic drift which drive the functional diversification of genes over time (Biswas & Akey, 2006; Toll-Riera *et al.*, 2011). Since the emergence of phylogeography as a discipline, mitochondrial DNA has been considered an evolutionarily nearly neutral marker and has been used mainly to infer evolutionary and demographic history of both populations and species (Avise, 2004). The nearly neutral character of mitogenome evolution has been questioned, and there is mounting evidence for both positive and negative selection shaping mitochondrial genome diversity in various taxonomic groups (Ballard & Whitlock, 2004; Meiklejohn *et al.*, 2007; Galtier *et al.*, 2009; Melo-Ferreira *et al.*, 2014).

The relatively small mitochondrial genome, consisting mostly of genes coding proteins that participate in the oxidative phosphorylation (OXPHOS) system, plays a vital role in cell energy production (Burton & Barreto, 2012). Mutations that alter individual fitness and athletic performance or cause diseases have been identified in it (Wallace, 2005). The mitochondrion is characterized by an insufficient DNA repair system and a lack of protective chromosome-associated proteins and is

Correspondence: Maciej Matosiuk, Institute of Biology, University of Bialystok, Swierkowa 20B, 15-950, Bialystok, Poland.
Tel.: +48 85 745 72 89; fax: +48 85 745 73 01;
e-mail: maciej.matosiuk@gmail.com

exposed to rapid molecular evolution. Therefore, it should be subject to strong selection against mutations at evolutionarily constrained sites and against alterations in the tertiary structure of the molecules (Bernt *et al.*, 2013). In most studies, the mtDNA ratio of nonsynonymous substitutions per nonsynonymous site to synonymous substitutions per synonymous site (d_N/d_S ratio) is well below 1 (Bazin *et al.*, 2006). This indicates that the vast majority of nonsynonymous mutations are eliminated by purifying (background) selection, pointing to its predominant role in shaping mtDNA diversity (Bazin *et al.*, 2006; Meiklejohn *et al.*, 2007).

The mitogenome's accelerated mutation ratio creates intense background noise concealing potential positively selected sites of adaptive value. The few studies that analyse the physiology of mitochondrial fitness and the influence of mtDNA components on metabolism, fertility and survivability have demonstrated the mostly disadvantageous effects of specific mitochondrial variants (Smith *et al.*, 2010; Boratyński *et al.*, 2011; Innocenti *et al.*, 2011). Most of the standard methods and approaches to detect positive selection are not applicable to mtDNA or, like the d_N/d_S ratio and McDonald–Kreitman test, are biased towards moderately conserved protein-coding genes (Biswas & Akey, 2006). Insights about positive selection on the mitochondrial level can be gained from phylogeny-wide comparisons of patterns of physicochemical changes of amino acids in the context of their effects on protein structure (da Fonseca *et al.*, 2008; Scott *et al.*, 2011), branch-site tests tracking signs of selection at single codons (Zhang *et al.*, 2005) and potentially by estimating lineage-specific deviations in the gene d_N/d_S ratio, recently applied to a large data set of nuclear genes (Toll-Riera *et al.*, 2011).

Introgressive hybridization offers a platform for studying the molecular basis of natural selection acting on mitogenomes (Boratyński *et al.*, 2014; Melo-Ferreira *et al.*, 2014). Interspecific mtDNA introgression was studied mainly in terms of disruption of co-adaptations between jointly evolving mitochondrial and nuclear genomes (mutually encoding components of the OXPHOS system), leading to mito-nuclear incompatibilities manifested in decreased fitness or hybrid breakdown (Burton & Barreto, 2012; Derr *et al.*, 2012). Adaptive introgression seems to be an essential source of adaptive variation, and there are examples of advantageous nuclear DNA variants from donor species spreading in recipient species (Hedrick, 2013). Despite the lack of unequivocal evidence for adaptive mtDNA introgression, when foreign variants are found deep within the distribution range of other taxa the presumption is that their maintenance or spread may be due to having adaptive value (Toews & Brelsford, 2012 and references therein). It has been suggested that positive selection of mtDNA-encoded proteins played a key role in the successful radiation and diversification of mammalian lineages, enabling them to face changing environmental circumstances and colonize new habitats (da Fonseca *et al.*, 2008).

The study of introgression of Siberian roe deer (*Capreolus pygargus*) mtDNA in the European roe deer (*C. capreolus*) genome could shed light on the molecular basis of adaptive mitogenome evolution of closely related species. The two roe deer species have a very wide distribution in the Palearctic with a narrow contact zone around the banks of the rivers Volga and

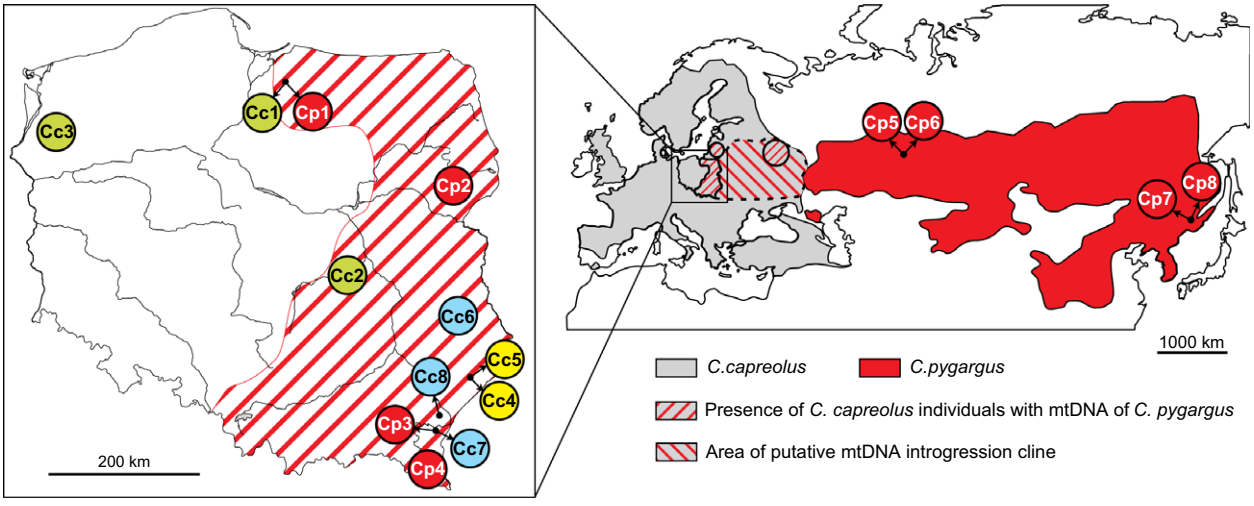

Fig. 1 Distribution of European (*Capreolus capreolus*) and Siberian roe deer (*C. pygargus*) in Eurasia and sampling locations. The proven and predicted areas for *C. pygargus* mtDNA are highlighted within the distribution range of *C. capreolus*. Cc1–Cc8, European roe deer samples (mtDNA Clade Central in green, Clade East in blue and Clade West in yellow). Samples of European roe deer possessing mtDNA of *C. pygargus* (Cp1–Cp4) and Siberian roe deer samples (Cp5–Cp8) in red (details in Table S1).

lower Don in Russia (Danilkin & Hewison, 1996; Fig. 1). They have been distinguished as separate species on the basis of differences in morphology, karyotype, behavioural ecology and no evidence for interbreeding in the wild (Danilkin & Hewison, 1996). Recent studies have found *C. pygargus* mtDNA in European roe deer populations in Moscow area (Zvychaynaya *et al.*, 2011b) and nearly 2000 km from the present distribution range of the Siberian roe deer, in Poland and Lithuania (Lorenzini *et al.*, 2014; Matosiuk *et al.*, 2014). Area of putative mtDNA introgression cline stretches between Poland in the west and the contact zone of these species in the east (Fig. 1). The absence of *C. pygargus* mtDNA haplotypes in contemporary Siberian roe deer, which are widespread in European roe deer populations in Poland and Moscow area, combined with support from ecological niche modelling of the two species, indicated that this instance of hybridization was mostly of natural origin (Zvychaynaya *et al.*, 2011b; Matosiuk *et al.*, 2014). The introgression of the mtDNA genome seems to have been driven by demographic processes in eastward-expanding *C. capreolus* populations, probably after the end of the Younger Dryas (10.8–10.0 ka BP; Matosiuk *et al.*, 2014). Our analyses indicated that some environmental factors (average temperature in January, snow cover depth, number of days with snow cover) could have played a role in maintaining *C. pygargus* mtDNA haplotypes in European roe deer populations (Matosiuk *et al.*, 2014). The last two of these environmental factors considered to be the most important ones in limiting the distribution of both roe deer species (Danilkin & Hewison, 1996). However, there is no direct evidence for whether this instance of widespread introgression of *C. pygargus* mtDNA into *C. capreolus* genome is an effect of adaptive introgression or only a by-product of demographic changes during the range expansion of the European roe deer in the past. To answer the question of how mitochondrial DNA evolved in particular *Capreolus* sp. lineages, we sequenced whole mitogenomes of both roe deer species and assessed (i) divergence between species-specific lineages; (ii) mitochondrial protein evolution, focusing particularly on their evolutionary ratio; and (iii) potential positively selected sites in each lineage, to confirm or reject the hypothesis of adaptive mtDNA introgression.

Materials and methods

Samples

For mitogenome analyses, we selected eight European roe deer specimens representing three haplogroups described in this species (Clade Central, Clade East, Clade West; Randi *et al.*, 2004) and four individuals from Poland possessing mtDNA pertaining to the *C. pygargus* lineage. Phylogenetic relation between selected mitotypes and specific mtDNA lineages was determined in our previous study (Matosiuk *et al.*, 2014). Four Siberian roe deer specimens from western Siberia and the Far East were also included. Species of the specimens were identified by morphological traits (Danilkin & Hewison, 1996). The precise sampling localities for each specimen are given in Fig. 1 and Table S1.

DNA extraction, PCR and sequencing

Genomic DNA was extracted from frozen tissues using the Dneasy Blood & Tissue Kit (Qiagen, Hilden, Germany). To obtain the *Capreolus* sp. mitochondrial genomes, we amplified 18 overlapping fragments in PCRs. The PCRs were performed in a 10-μl reaction volume containing Qiagen Multiplex PCR Master Mix (1×), 0.2 μM of each primer, ~20 ng genomic DNA and RNAse-free water. PCR conditions followed the manufacturer's protocol and varied in annealing temperature depending on the fragment amplified. Table S2 shows the primers used in this study, including 14 primer pairs newly designed in Fast PCR software (Kalendar *et al.*, 2009), their mtDNA genome position, length and annealing temperature. New primer pairs were designed on the basis of available mitogenome of European roe deer (Hassanin *et al.*, 2012) and its closest relatives from *Cervidae* family: water deer (*Hydropotes inermis*, NC018032) and moose (*Alces alces*, JN632595). The amplification products were purified using shrimp alkaline phosphatase (SAP) and Exonuclease I (Thermo Scientific) in the enzymatic reaction, following the manufacturer's protocol. Both strands of amplified products were sequenced with each of the amplification primers and the Big Dye Terminator v3.1 Cycle Sequencing Kit (Life Technologies, Carlsbad, CA, USA) in cycle sequencing reactions including 25 cycles of 95 °C (20 s), 50 °C (15 s) and 60 °C (60 s). An additional primer was applied in sequencing of the fragment spanning between *ND2* and *COX1* genes due to the presence of long homopolymer. Unincorporated dideoxynucleotides were removed using the ExTerminator Kit (A&A Biotechnology, Gdynia, Poland), and the sequencing products were run on a 3130 Applied Biosystems Genetic Analyzer.

Sequence analysis

The obtained DNA sequences were aligned with BioEdit v7.0.4 (Hall, 1999) and revised manually. Mitochondrial sequences of all 13 protein-coding genes and two rRNA subunits were identified using reference gene sequences of *Bos taurus* (KC153977.1). To avoid nuclear DNA sequences of mitochondrial origin (pseudogenes/numts), we checked all coding regions for open reading frames. We also examined overlapping fragments of all obtained sequences to ensure complete sequencing coverage. Additionally, we were able to reconstruct

the cloverleaf secondary structure of 22 tRNAs using tRNA-scan SE 1.21 (Lowe & Eddy, 1997) and pknotsRG (Reeder *et al.*, 2007). Fixed tRNA interspecific differences were counted for acceptor arm stem, D loop, anticodon loop, variable loop and TψC loop. Basic sequence statistics including species-specific lineage polymorphism and nucleotide divergence, number of synonymous and nonsynonymous substitutions, and a sliding window analysis of number of net nucleotide substitutions per site between species-specific lineages were calculated in DnaSP 5.10 (Librado & Rozas, 2009). To determine whether fixed lineage-specific nonsynonymous substitutions alter the properties of the encoded polypeptides, we categorized the amino acid replacements as either transmembrane or nontransmembrane domains of transmembrane topology predicted in Phobius (Käll *et al.*, 2007).

One of the tested primer pairs (Table S3) amplified a nuclear pseudogene of mitochondrial origin. The amplified numt spans at least 1300 bp of the *COX2*, *ATP8*, *ATP6* and *COX3* genes. It contained insertion in *COX2* gene changing the reading frame and mutation leading to premature termination in translation of *ATP8* gene. Identical pseudogene variants were obtained for both pure and introgressed European roe deer specimens. They differed from those obtained for Siberian roe deer at 0.75% nucleotide divergence, whereas the average nucleotide divergence of the corresponding mtDNA fragment between species-specific lineages was equal to 2.99%. Our analysis revealed some inconsistency with the first published European roe deer mitochondrial genome (JN632610; Hassanin *et al.*, 2012) which contains a numt in the *COX3* gene. The nuclear sequences of mitochondrial origin obtained in this study were deposited in GenBank (KJ804136–KJ804137).

Phylogenetic analysis

The Bayesian inference of phylogeny in MrBayes v3.2.0 (Ronquist & Huelsenbeck, 2003) was performed using a set of 15374-bp aligned sequences, including all 13 protein-coding genes, 12S rRNA, 16S rRNA and 22 concatenated tRNA genes (reverse complement sequences of L-strand genes were used). The mitochondrial control region, oriL, stop codons, 12 intergenic spacers (1–5 bp long) and intragenic indel polymorphisms in rRNA subunits were excluded from further analyses. Mitochondrial sequence alignment was augmented by adding the mitogenomes of the water deer (NC018032), moose (JN632595), and a more distant even-toed ungulate, the domestic cow (KC153977). To infer phylogeny, we partitioned the data into 6 sets corresponding to the 1st, 2nd and 3rd codon positions in protein-coding genes, 12S rRNA, 16S rRNA and concatenated tRNAs. Analysis for each partition was performed under general time-reversible (GTR) models allowing variation of the gamma distribution of nucleotide sites between partitions and including a proportion of invariable sites (GTR + G + I model). A single MrBayes run was performed using a random starting tree with four default chains (one cold and three heated) consisting of 8 million generations (first 25% discarded as burn-in) sampled every 100 generations.

We also inferred amino acid sequences at the ancestral node to *C. capreolus* and *C. pygargus* species-specific lineages using a maximum likelihood approach and the general reversible mitochondrial (mtREV) model implemented in MEGA 5 (Tamura *et al.*, 2011) with a specified tree representing topology (((*C. capreolus* seqs, *C. pygargus* seqs), *H. inermis*), *A. alces*, *B. taurus*) based on the results of the phylogenetic analysis described above.

Mitochondrial protein evolution

To determine whether mitochondrial DNA in the genus *Capreolus* is evolving under a standard neutral model or rather under natural selection, we performed McDonald & Kreitman's (1991) test for concatenated protein-coding genes as well as for each gene separately, comparing divergence and polymorphism at nonsynonymous sites with those at synonymous sites. We also performed an extension of the McDonald–Kreitman test, comparing the polymorphism of synonymous and nonsynonymous sites in each species-specific lineage with fixed interspecific differences to identify any unusual pattern of evolution as the *C. capreolus* and *C. pygargus* mtDNA lineages split.

To test whether the mtDNA type of *C. capreolus* or *C. pygargus* shows a lineage-specific pattern of evolutionary rates we assessed variation of the ratio of amino acid replacements per nonsynonymous site (d_N) to silent substitutions per synonymous site (d_S) among the branches of the tree with topology inferred by the previous Bayesian approach. Branch-specific d_N/d_S (ω) values were analysed using 12 protein-coding mtDNA genes (*ND6* excluded) and the maximum-likelihood approach implemented in the CODEML program of the PAML package (Yang, 2007). We applied a branch model (free-ratio model) that allowed the ω ratio to vary among branches in the phylogeny (Yang & Nielsen, 1998). To display relationship of haplotypes in species-specific lineages of the genus *Capreolus,* we constructed their median-joining network in NETWORK v4.6.1.0 (http://www.fluxus-engineering.com) based on concatenated amino acid sequences of 12 protein-coding genes.

We employed branch-site model A (Zhang *et al.*, 2005) implemented in CODEML to test for positive selection in both the European and the Siberian roe deer mtDNA lineages. The branch-site model allows variation of d_N/d_S values among different branches as well as sites of a particular lineage (foreground branch), so it can identify a few sites under positive selection

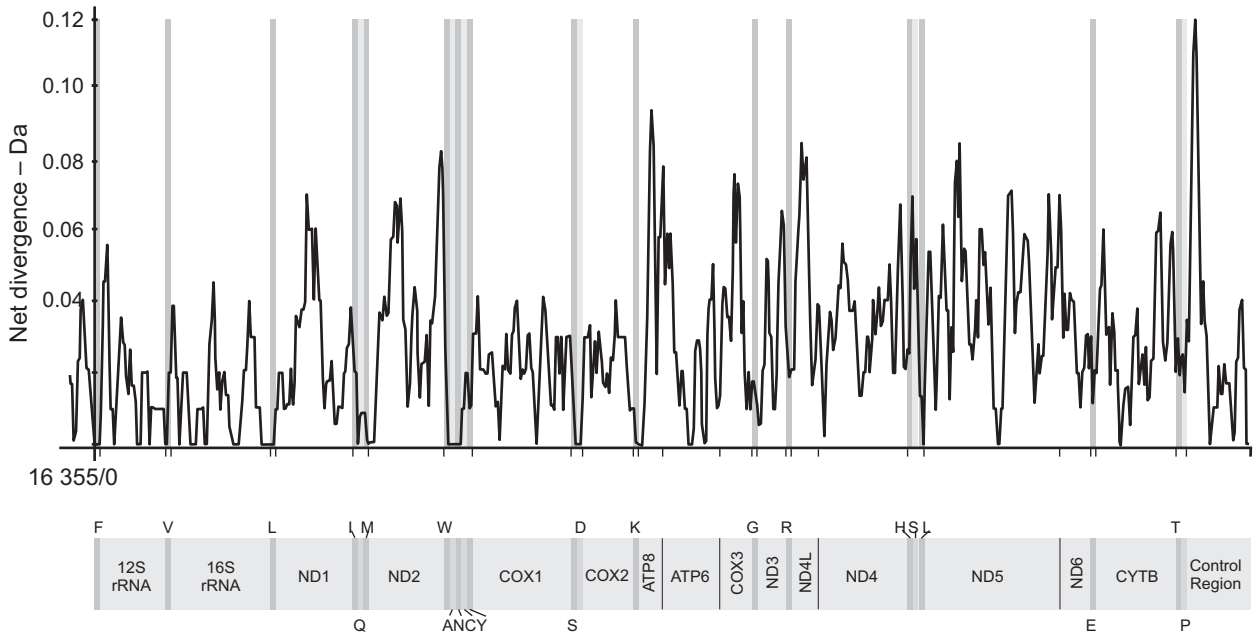

Fig. 2 Gene organization of mitochondrial genomes in genus *Capreolus* and sliding window analysis of net divergence (Da) between species-specific mtDNA lineages of European and Siberian roe deer (window size 100 bp, step size 25 bp). Vertical grey bars (light and dark) highlight the tRNA genes.

(Yang *et al.*, 2005). In this model, the null hypothesis assumes that all branches (background and foreground) evolve under negative selection or neutrality, opposed against the alternative hypothesis in which the foreground branch is evolving under positive selection. Subsequent likelihood ratio tests (LRT) were performed to test the significance of detection of sites under positive selection on the selected branch. Branch-site tests for positive selection with the most widespread European roe deer lineage (Clade Central) as foreground branch were also performed.

TreeSAAP software (Woolley *et al.*, 2003) was used for detection of directional positive selection in each protein-coding gene separately. TreeSAAP categorizes the physicochemical properties of amino acid replacements into eight magnitude categories and then determines whether the observed magnitude of amino acid changes deviates significantly from neutral expectations. Only when significantly more replacements in higher magnitude categories are inferred than are expected by chance, they could indicate directional positive selection (McClellan & Ellison, 2010). Thus, only changes with magnitude category ≥ 6 and $P \leq 0.001$ were considered as an indication of positive selection. To strengthen accuracy and to avoid false–positive results, we excluded 11 of 31 properties analysed by TreeSAAP which had an accuracy of detection of selection lower than 85%, according to a methodology improvement described by McClellan & Ellison (2010).

Results

Mitochondrial genome organization and divergence

We sequenced 16 mitochondrial genomes in the genus *Capreolus* (GenBank accession numbers: KJ681480–KJ681495; Table S1), including the first complete Siberian roe deer mtDNA genome. The roe deer mitogenomes ranged from 16352 to 16359 bp in length (Table S1) and the gene organization was typical of metazoans (Bernt *et al.*, 2013), shown in Fig. 2. After exclusion of indels, they comprised 16346 nucleotides and showed 855 polymorphic sites (861 mutations) consisting of 663 parsimony-informative sites and 192 singletons.

Species-specific lineage sequences differed on average at ~546.8 nucleotide sites, reflected in 0.028 net nucleotide substitutions per site across the mitochondrial genome. Net divergence (Da) varied substantially between genes; however, it was lowest in concatenated tRNAs and two rRNA subunits (under 0.017) and exceeded 0.050 at *ATP8* and *ND4L* (Table 1). Variation of net divergence between *C. capreolus* and *C. pygargus* shown in sliding window analysis (window size 100 bp, step size 25 bp; Fig. 2) revealed conserved and highly divergent regions. Most of the former were within ribosomal subunit genes and in adjacent regions of tRNAs. Regions of the highest window frame divergence in protein-coding genes (over 0.080) were at *ND2*, *ATP8*, *ND4L* and *ND5* (Fig. 2).

Table 1 Summary statistics describing the divergence in mitochondrial genes between European and Siberian roe deer lineages.

Gene	Length [bp]	*n*	Da	D_s	D_n
ND1	956	31.9	0.029	23	1
ND2	1042	49.8	0.039	30	4
COX1	1545	39.3	0.022	29	1
COX2	684	18.3	0.021	12	0
ATP8	201	11.5	0.052	7	3
ATP6	681	23.0	0.028	14	1
COX3	784	28.0	0.031	20	0
ND3	346	15.6	0.039	7	4
ND4L	297	17.4	0.054	14	0
ND4	1378	56.1	0.034	39	1
ND5	1821	84.3	0.040	58	4
ND6	528	21.0	0.035	17	0
cytb	1140	44.1	0.030	24	2
12S rRNA	955	18.4	0.017	–	–
16S rRNA	1564–1571	23.6	0.013	–	–
tRNAs	1513–1516	23.9	0.013	–	–
noncoding DNA	985–987	40.9	0.030	–	–
Total	16352–16359	546.8	0.028	294	21

n: mean number of nucleotide differences; Da: number of net nucleotide substitutions per site between species-specific lineages, D_s: number of fixed synonymous substitutions; D_n: number of fixed nonsynonymous substitutions.

Table 2 Divergence of mitochondrial components of electron transport system (ETS) complexes between European and Siberian roe deer lineages.

ETS complex	mtDNA genes	Concatenated sequence [bp]	Da	D_n	D_n per 1000 bp
ETS I	*ND1*, *ND2*, *ND3*, *ND4L*, *ND4*, *ND5*, *ND6*	6368	0.037	14	2.20
ETS III	*cytb*	1140	0.030	2	1.75
ETS IV	*COX1*, *COX2*, *COX3*	3013	0.024	1	0.33
ETS V	*ATP8*, *ATP6*	882	0.034	4	4.54

Da: number of net nucleotide substitutions per site between species-specific lineages; D_n: number of fixed nonsynonymous substitutions.

Analyses of protein-coding genes divided into the mitochondrial components of electron transport system (ETS) complexes showed divergence highest in complex I (Da = 0.037) and lowest in complex IV (Da = 0.024; Table 2). Substantial divergence in protein-coding genes resulted in fixed substitutions at 315 nucleotide sites (41 at 1st, 7 at 2nd and 267 at 3rd codon position), causing permanent replacement of 21 amino acids between species-specific lineages (Table 1). Eleven of 22 studied tRNAs contained 18 fixed differences between lineages. Seven of them were identified in the TψC loop and four in each acceptor stem and D loop of tRNAs. Only two fixed substitutions were found in an anticodon loop and one in a variable loop. Divergence was highest in tRNA-His, which contained four fixed nucleotide changes (Table S4).

MtDNA protein evolution in genus Capreolus

A global McDonald–Kreitman test of concatenated sequences of all mitochondrial protein-coding genes revealed an overall excess of amino acid polymorphism and a deficiency of interspecific fixed amino acid replacements, suggesting that purifying selection shaped the genetic variation of mitochondrial genome components in *Capreolus* (Neutrality Index: NI = 2.35; Fisher's exact test, two-tailed *P*-value: $P = 0.0013$). When protein-coding genes were examined separately, signs of purifying selection were found for three adjacent genes: *ND4* (NI = 9.07; $P = 0.0210$), *ND5* (NI = 5.40; $P = 0.0028$) and *ND6* showing no interspecific differences (NI value not obtained; $P = 0.0118$). After applying the Bonferroni correction for multiple tests ($\alpha = 0.0038$), only the values for *ND5* were significant.

More insights into the potential lineage-specific evolutionary pattern came from comparing polymorphism at synonymous and nonsynonymous sites in a selected lineage with their divergence from the other lineage in the extension of the McDonald–Kreitman test. These tests indicated an evident excess of amino acid polymorphism in the European roe deer lineage and the significant effect of purifying selection on genetic variation at both *ND4* (NI = 16.36, $P = 0.0033$) and *ND5* (NI = 7.33, $P = 0.0020$; Table S5). *ND6* also displayed purifying selection in this lineage ($P = 0.0033$) but NI values were not estimated due to the absence of fixed interspecific nonsynonymous replacements. Analysis of protein-coding gene polymorphism in the Siberian roe deer lineage showed no significant signs of purifying selection (Table S6).

We compared the polymorphism of concatenated gene sequences in the lineages against their interspecific divergence. The comparison indicated that the Siberian roe deer mitochondrial lineage evolved under neutrality. In the European roe deer lineage, the ratio of diversity to divergence between the two classes of sites (synonymous and nonsynonymous) was biased towards higher polymorphism at amino acid level, indicating the pressure of purifying selection (NI = 3.56; $P < 0.0001$).

Maximum likelihood analysis of ancestral amino acid states at the node preceding the split of the *C. capreolus* and *C. pygargus* mitochondrial lineages revealed a great disparity. Most of the ancestral states (17 of 21) were maintained in the *C. pygargus* lineage, with only three ancestral states assigned to the *C. capreolus* lineage (one ancestral state was not assigned to either lineage due to an ambiguous phylogenetic pattern and/or possible ancestral polymorphism; Table 3). Such a disparity in

Table 3 Fixed nonsynonymous substitutions between species-specific lineages in genus *Capreolus*, their peptide chain locations, inferred ancestral states (shaded), and impact on local physicochemical properties.

Gene	Site	Location	Ancestral state	Cp-type	Cc-type	Magnitude of change†
ND1	311	Transmembrane	Leu	Leu	Met	
ND2	149	Transmembrane	Ile	Ile	Met	
	152	Transmembrane	Asn	Asp	Asn	
	324	Transmembrane	Thr	Thr	Met	
	334	Transmembrane	Thr	Ile	Thr	6 *pK'**; 8 R_a*
COX1	415	Transmembrane	Val	Val	Met	
ATP8	39	Nontransmembrane	Pro	Pro	Ser	
	42	Nontransmembrane	Thr	Thr	Ala	6 *Pα**
	48	Nontransmembrane	Asn	Asn	Ser	
ATP6	49	Nontransmembrane	Val/Met	Met	Val	
ND3	4	Nontransmembrane	Met	Met	Val	
	82	Nontransmembrane	Asn	Asn	Asp	
	85	Transmembrane	Gly	Gly	Asn/Asp	
	89	Transmembrane	Thr	Thr	Ile	
ND4	421	Nontransmembrane	His	Tyr	His	
ND5	62	Nontransmembrane	Ile	Ile	Val	8 *pK'**
	180	Transmembrane	Ile	Ile	Val	8 *pK'**
	267	Nontransmembrane	Thr	Thr	Met	6 *Pα**
	489	Transmembrane	Met	Met	Thr/Ala	6 *Pα**
cytb	295	Transmembrane	Val	Val	Ile	8 *pK'**
	349	Transmembrane	Thr	Thr	Ala	6 *Pα**

Cp-type – *C. pygargus* type ancestral state; Cc-type – *C. capreolus* type ancestral state.
**P* < 0.05.
†Magnitude of amino acid changes inferred with TreeSAAP: magnitude category and affected physicochemical amino acid property. *Pα*: alpha-helical tendencies; *pK'*: equilibrium constant (ionization of COOH); R_a: solvent accessible reduction ratio.

the amount of amino acid modifications was reflected in differential d_N/d_S ratios of species-specific lineages. The European roe deer lineage displayed an almost four-fold higher ω ratio (0.044) than the Siberian roe deer lineage (0.012; Fig. 3a) during the time to the most recent common ancestor (TMRCA) of *Capreolus* sp. to its separation into present-day clades. The background d_N/d_S ratio across the reconstructed phylogeny (Fig. 4) was estimated at 0.025. The differences in evolutionary rates become even more evident within species-specific lineages, where Clade Central, Clade East and Clade West of *C. capreolus* give d_N/d_S ratios of

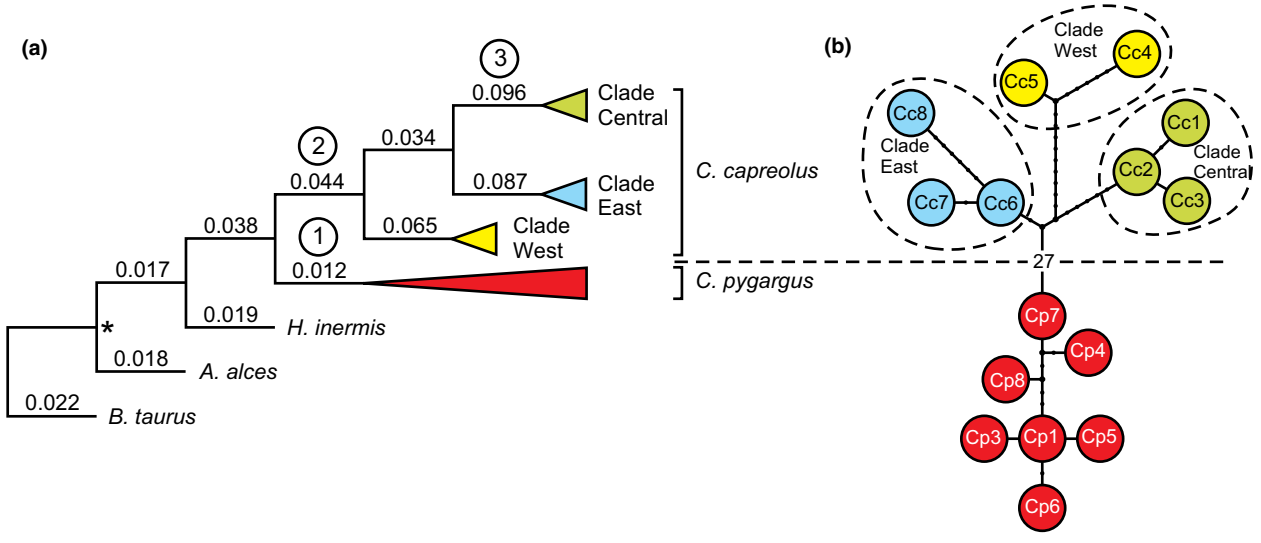

Fig. 3 Lineage-specific d_N/d_S based on analysis of a 10848-bp concatenated alignment of 12 protein genes (a) and respective median-joining network of amino acid sequences in genus *Capreolus* (b). Numbering indicates branches of *C. pygargus* lineage (1), *C. capreolus* lineage (2) and Clade Central (3) selected as foreground branches in branch-site tests for positive selection.

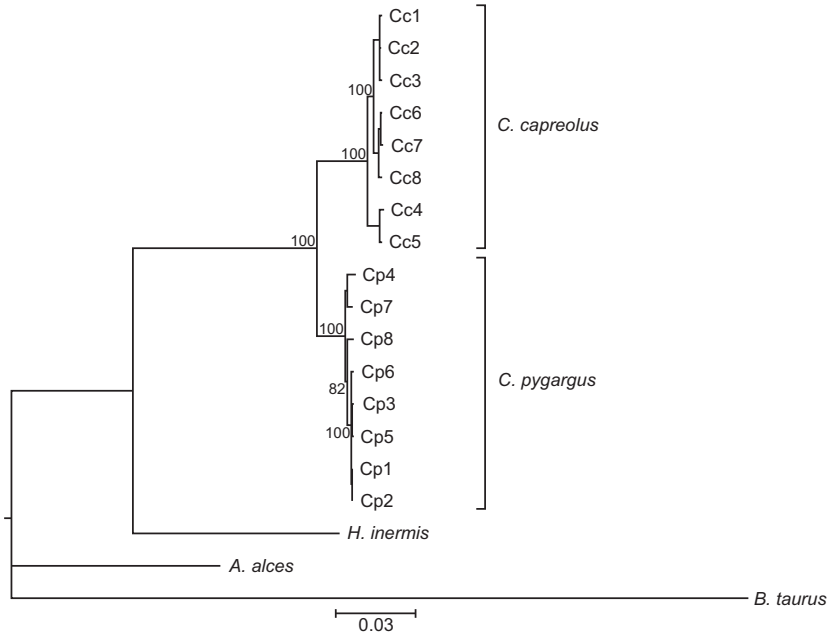

Fig. 4 Phylogenetic relationships among 16 mitogenomes of *C. capreolus* and *C. pygargus* obtained in this study. Tree topology and branch lengths are based on maximum likelihood analysis. Credibility values are given at nodes. Haplotypes Cp1, Cp2, Cp3 and Cp4 represent Siberian roe deer lineage found in European roe deer.

0.096, 0.087 and 0.065, respectively (Fig. 3a). The ω ratios within the *C. pygargus* lineage were extremely low. Higher evolutionary rates in the European roe deer lineage and decelerated evolution of the Siberian roe deer lineage were also reflected in the reconstruction of the median-joining network of concatenated amino acid sequences (Fig. 3b).

Despite the low d_N/d_S ratio values across the examined phylogeny, Bayes empirical Bayes (BEB) analysis (Yang *et al.*, 2005) of branch-site model A implemented in CODEML indicated two sites in the *C. pygargus* lineage, six in the *C. capreolus* lineage, and four in Clade Central of *C. capreolus* that may have been positively selected (Table S7). The significance of any detected site could not be confirmed in subsequent likelihood ratio tests. Tests for functional divergence performed in TreeSAAP indicated that 8 sites (in *ND2*, *ND5*, *ATP8* and *cytb* genes) representing fixed substitutions between species-specific lineages of roe deer may lead to functional changes (Table 3). Seven of them appeared in *C. capreolus* and only one in *C. pygargus* mtDNA lineage. However, none of these sites were identified as fixed due to directional positive selection.

Discussion

Mitogenome divergence

The level of described mitogenome divergence between species-specific lineages in *Capreolus* is low. Comparably, low differentiation was noted between major matrilineal phylogroups within mammal species and sister-species pairs that emerged in the Pleistocene (Avise *et al.*, 1998; Green *et al.*, 2008). Climatic oscillations in the quaternary had a great impact on the shaping of genetic variability in temperate zone species (Hewitt, 2004), including the European roe deer (Sommer *et al.*, 2009). Long periods of isolation in spatially restricted Iberian and Balkan refugia led to significant diversification and gave rise to three major phylogroups described in the European roe deer: Clade Central, Clade East and Clade West (Randi *et al.*, 2004). Our results strongly support that division and distinguish the haplogroups by both molecular (0.42–0.82% mean divergence) and protein evolution measures (Fig. 3b). Unlike the European roe deer, the first specieswide reconstruction of the phylogeny of the Siberian roe deer indicated more complex relationships between haplotypes (Vorobieva *et al.*, 2011). Populations of Siberian roe deer are phylogenetically heterogeneous, and there is little evidence for existence of large phylogeographic structure. This could be a result of high migration rates observed in this species (Vorobieva *et al.*, 2011) and high connectivity of populations during glacial periods, revealed by species distribution modelling (Matosiuk *et al.*, 2014). Applying another mitochondrial marker (cytochrome *b*) in addition to the commonly used control region Zvychaynaya *et al.* (2011a) revealed the presence of three distinguishable haplogroups in *C. pygargus*. Employing whole mitogenome sequencing in our study considerably improves inference of deep level phylogenetic relationships within Siberian roe deer lineage. However, a full grasp of the complex phylogeny and phylogeography of *C. pygargus* will require more comprehensive sampling.

The estimated average divergence between *C. capreolus* and *C. pygargus* mtDNA lineages was 2.8%, but heterogeneity of divergence was clearly observed across the mitogenomes. Nonuniform rates of molecular evolution noted between different gene classes and ETS complexes in this study accurately reflected constraints resulting from their functions. We also found ribosomal RNA subunits (12S and 16S) and 22 tRNAs required for mitochondrial protein synthesis to be highly conserved. Functional constraints were evident in tRNA genes, as intraspecific polymorphism just slightly exceeded interspecific divergence (24 polymorphic sites *vs* 18 fixed differences). The level of tRNA divergence between species-specific *Capreolus* lineages was half of that reported in American bison and its introgressed lineage with cattle ancestry (Douglas *et al.*, 2011).

The pattern of divergence between studied mitochondrial components of the electron transport system, indicating relative mutation rates following the order *ND* > *ATP* > *cytb* > *COX* (Table 2), corresponded to the constraints needed to maintain the stability and functional efficiency of OXPHOS complexes (da Fonseca *et al.*, 2008; Montooth *et al.*, 2009). Faster evolution of mitochondrial components of NADH dehydrogenase (ETS I) as compared with cytochrome *c* oxidase subunits (ETS IV) has been reported in various taxonomic groups (Ballard, 2000; Zhang & Broughton, 2013). Nonsynonymous mutations in *ND* genes tend to have a minor impact on the functional properties of proteins (e.g. Blier *et al.*, 2006), whereas even a single mutation in *COX* genes can have a disproportionate effect on fitness (Scott *et al.*, 2011). A recent study gives evidence for compensatory evolution of nuclear components of the cytochrome *c* oxidase complex, minimizing the slightly deleterious effects of mtDNA mutations (Osada & Akashi, 2012). Molecular divergence followed that pattern in our study, and most of the fixed differences were in *ND* genes and the fewest in *COX* genes. When we considered relative sequence length, however, the nonsynonymous substitution ratio was highest in analysed *ATP* genes, twice that of *ND* and *cytb* genes and over thirteen times that of *COX* genes. An increased ratio in the ATP synthase complex can be explained by the stochastic effect of three fixed amino acid substitutions in the shortest protein-encoding mitochondrial gene, *ATP8*. Nevertheless, there is clear disproportion in accumulation of nonsynonymous changes between cytochrome *c* oxidase and other mitochondrial electron transport system complexes, suggesting its high evolutionary conservatism in roe deer.

Mitochondrial protein evolution

Mitochondrial protein evolution analyses shed further light on the possible scenario underlying the widespread introgression of Siberian roe deer mtDNA in European roe deer populations (Zvychaynaya *et al.*, 2011b; Matosiuk *et al.*, 2014). Overall, the very low d_N/d_S ratio in both studied mitochondrial lineages indicates a strong effect of purifying selection on the shaping of mtDNA diversity in *Capreolus*, supporting suggestions that it is the main driver of mitogenome evolution (Meiklejohn *et al.*, 2007). Selective pressure across the reconstructed phylogeny was even higher (very low d_N/d_S) than reported for primates, considered a group with very strong purifying selection on mitochondrial proteins (Hasegawa *et al.*, 1998).

Surprisingly, a closer look at the patterns of intraspecific polymorphism and interspecific divergence in the extension of the McDonald–Kreitman test indicated that overall purifying selection did not have a significant effect on shaping variation in the mitochondrial lineages of both roe deer species. Significant signs of background selection were found only in the *C. capreolus* lineage, and they were pronounced in *ND4* and *ND5*. In this case, selection prevents accumulation of slightly deleterious nonsynonymous mutations in subunits suggested to actively participate in proton pumping (Brandt, 2006). No evident signs of selection were found in other genes, most probably due to the test's low sensitivity to restricted levels of polymorphism and divergence which is common in species with relatively young radiation history (Melo-Ferreira *et al.*, 2014).

Although lacking significant purifying selection, the mitogenome of the *C. pygargus* lineage has remained almost unaltered since the TMRCA of *Capreolus* species. We found that only three amino acid replacements in genes of ETS I (*ND* genes) were fixed in the Siberian roe deer lineage since then. Nonsynonymous substitutions in *ND* genes have milder effects and are more likely to accumulate than in *COX* genes, as suggested by Montooth *et al.* (2009) based on data for insect taxa (Bazin *et al.*, 2006). More evidence for faster accumulation of replacement mutations in ETS I comes from fish and mammals (Zhang & Broughton, 2013).

The nearly ancestral mitogenome of the *C. pygargus* lineage contrasts with the highly variable one in the *C. capreolus* lineage. To emphasize the disproportion, the most distant mtDNA-encoded protein variants differed by 9 amino acids in both the whole *C. pygargus* lineage and in intraspecific *C. capreolus* lineage, Clade East (Fig. 3b). The disproportion reflects the high discrepancy in evolutionary ratios (d_N/d_S) between species-specific mitochondrial lineages since the TMRCA (Fig. 3a). Significant differences in evolutionary ratios between lineages most likely indicate changes in the impact of natural selection over time (Czelusniak *et al.*, 1982). Higher d_N/d_S ratio, which could result from adaptive evolution at some sites, was determined for *C. capreolus* lineage. Lower ω value found in the *C. pygargus* lineage gives evidence against the hypothesis of adaptive mtDNA introgression into the European roe deer lineage.

We suggest two main evolutionary consequences for the high diversification and accelerated evolutionary ratio found in the European roe deer mitochondrial lineage. First, overall purifying selection acts to maintain the stability and efficiency of OXPHOS components to avoid functional disruption of co-adapted mito-nuclear complexes and to keep their interactions in equilibrium, something difficult to attain by nuclear compensatory evolution alone. Second, divergent *C. capreolus* lineages could have maintained a capability for genomic incorporation of the well-preserved and almost ancestral type of mtDNA present in *C. pygargus*. The nearly ancestral type of proteins encoded by mtDNA, together with the absence of evidence for positive selection on the molecular level in the *C. pygargus* lineage, suggests that the maintenance of foreign genetic variants in the European roe deer genome is evolutionarily neutral. It has been proposed that slightly deleterious variants arising through mutational load accumulation may be easier to replace with those from a closely related species (Rieseberg, 2009; Llopart *et al.*, 2014). Thus, the lower mutational load of the *C. pygargus* mitogenome may have driven the widespread introgression in the European roe deer. The conserved type of mitogenome found in the Siberian roe deer may indicate strictly coevolved interactions with nuclear electron transport system components that could withstand diverged variants. That is, the most plausible explanation for the unidirectional character of introgression indicated by molecular analyses of Siberian roe deer populations in the contact zone of roe deer species (Danilkin *et al.*, 2012).

Acknowledgments

We thank Dr Sergey Anchugov (Kurgan, Russia) for providing samples of *C. pygargus*, and Magdalena Świsłocka, Magdalena Czajkowska, Norbert Duda and Piotr Rode for their help in collecting tissues of *C. capreolus* from Poland. We also thank two anonymous reviewers for their useful comments about this study. Michael Jacobs and Agata Borkowska line-edited the paper for submission. This work was financially supported by the National Science Centre of Poland (Grant no. N N304 390838 to AB). MM is a beneficiary of the project 'Scholarships for PhD students of Podlaskie Voivodeship' cofinanced by European Social Fund, Polish Government and Podlaskie Voivodeship.

References

Avise, J.C. 2004. *Molecular Markers, Natural History, and Evolution*, 2nd edn. Sinauer Associates, Sunderland, MA.

Avise, J.C., Walker, D. & Johns, G.C. 1998. Speciation durations and Pleistocene effects on vertebrate phylogeography. *Proc. Biol. Sci.* **265**: 1707–1712.

Ballard, J.W.O. 2000. Comparative genomics of mitochondrial DNA in members of the *Drosophila melanogaster* subgroup. *J. Mol. Evol.* **51**: 48–63.

Ballard, J.W.O. & Whitlock, M.C. 2004. The incomplete natural history of mitochondria. *Mol. Ecol.* **13**: 729–744.

Bazin, E., Glémin, S. & Galtier, N. 2006. Population size does not influence mitochondrial genetic diversity in animals. *Science* **312**: 570–572.

Bernt, M., Braband, A., Schierwater, B. & Stadler, P.F. 2013. Genetic aspects of mitochondrial genome evolution. *Mol. Phylogenet. Evol.* **69**: 328–338.

Biswas, S. & Akey, J.M. 2006. Genomic insights into positive selection. *Trends Genet.* **22**: 437–446.

Blier, P.U., Breton, S., Desrosiers, V. & Lemieux, H. 2006. Functional conservatism in mitochondrial evolution: insight from hybridization of Arctic and brook charrs. *J. Exp. Zool.* **306**: 425–432.

Boratyński, Z., Alves, P.C., Berto, S., Koskela, E., Mappes, T. & Melo-Ferreira, J. 2011. Introgression of mitochondrial DNA among *Myodes* voles: consequences for energetics? *BMC Evol. Biol.* **11**: 355.

Boratyński, Z., Melo-Ferreira, J., Alves, P.C., Berto, S., Koskela, E., Pentikäinen, O.T. *et al.* 2014. Molecular and ecological signs of mitochondrial adaptation: consequences for introgression? *Heredity* **113**: 277–286.

Brandt, U. 2006. Energy converting NADH:quinone oxidoreductase (complex I). *Annu. Rev. Biochem.* **75**: 69–92.

Burton, R.S. & Barreto, F.S. 2012. A disproportionate role for mtDNA in Dobzhansky-Muller incompatibilities? *Mol. Ecol.* **21**: 4942–4957.

Czelusniak, J., Goodman, M., Hewett-Emmett, D., Wiess, M.L., Venta, P.J. & Tashian, R.E. 1982. Phylogenetic origins and adaptive evolution of avian and mammalian haemoglobin genes. *Nature* **298**: 297–300.

Danilkin, A.A. & Hewison, A.J.M. 1996. *Behavioural Ecology of Siberian and European Roe Deer*. Chapman and Hall, London.

Danilkin, A.A., Zvychaynaya, E.Y. & Kholodova, M.V. 2012. What roe deer inhabits the Volga Region. *Vestn. okhotovedeniya* **9**: 200–208. [in Russian with English summary].

Derr, J.N., Hedrick, P.W., Halbert, N.D., Plough, L., Dobson, L.K., King, J. *et al.* 2012. Phenotypic effects of cattle mitochondrial DNA in American Bison. *Conserv. Biol.* **26**: 1130–1136.

Douglas, K.C., Halbert, N.D., Kolenda, C., Childers, C., Hunter, D.L. & Derr, J.N. 2011. Complete mitochondrial DNA sequence analysis of *Bison bison* and bison-cattle hybrids: function and phylogeny. *Mitochondrion* **11**: 166–175.

da Fonseca, R.R., Johnson, W.E., O'Brien, S.J., Ramos, M.J. & Antunes, A. 2008. The adaptive evolution of the mammalian mitochondrial genome. *BMC Genomics* **9**: 119.

Galtier, N., Nabholz, B., Glémin, S. & Hurst, G.D.D. 2009. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Mol. Ecol.* **18**: 4541–4550.

Green, R.E., Malaspinas, A.-S., Krause, J., Briggs, A.W., Johnson, P.L.F., Uhler, C. *et al.* 2008. A complete neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* **134**: 416–426.

Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**: 95–98.

Hasegawa, M., Cao, Y. & Yang, Z. 1998. Preponderance of slightly deleterious polymorphism in mitochondrial DNA:

nonsynonymous/synonymous rate ratio is much higher within species than between species. *Mol. Biol. Evol.* **15**: 1499–1505.

Hassanin, A., Delsuc, F., Ropiquet, A., Hammer, C., van Vuuren, B.J., Matthee, C. *et al.* 2012. Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by comprehensive analysis of mitochondrial genomes. *C. R. Biol.* **335**: 32–50.

Hedrick, P.W. 2013. Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Mol. Ecol.* **22**: 4606–4618.

Hewitt, G.M. 2004. Genetic consequences of climatic oscillations in the Quarternary. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **359**: 183–195.

Innocenti, P., Morrow, E.H. & Dowling, D.K. 2011. Experimental evidence supports a sex-specific selective sieve in mitochondrial genome evolution. *Science* **332**: 845–848.

Kalendar, R., Lee, D. & Schulman, A.H. 2009. FastPCR Software for PCR primer and probe design and repeat search. *Genes Genomes Genomics* **3**: 1–14.

Käll, L., Krogh, A. & Sonnhammer, E.L.L. 2007. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res.* **35**: W429–W432.

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

Llopart, A., Herrig, D., Brud, E. & Stecklein, Z. 2014. Sequential adaptive introgression of the mitochondrial genome in Drosophila yakuba and Drosophila santomea. *Mol. Ecol.* **23**: 1124–1136.

Lorenzini, R., Garofalo, L., Qin, X., Voloshina, I. & Lovari, S. 2014. Global phylogeography of the genus *Capreolus* (Artiodactyla: Cervidae), a Palaearctic meso-mammal. *Zool. J. Linn. Soc. Lond.* **170**: 209–221.

Lowe, T.M. & Eddy, S.R. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**: W955–W964.

Matosiuk, M., Borkowska, A., Świsłocka, M., Mirski, P., Borowski, Z., Krysiuk, K. *et al.* 2014. Unexpected population genetic structure of European roe deer in Poland: an invasion of the mtDNA genome from Siberian roe deer. *Mol. Ecol.* **23**: 2559–2572.

McClellan, D.A. & Ellison, D.D. 2010. Assessing and improving the accuracy of detecting protein adaptation with the TreeSAAP analytical software. *Int. J. Bioinform. Res. Appl.* **6**: 120–133.

McDonald, J.H. & Kreitman, M. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**: 652–654.

Meiklejohn, C.D., Montooth, K.L. & Rand, D.M. 2007. Positive and negative selection on the mitochondrial genome. *Trends Genet.* **23**: 259–263.

Melo-Ferreira, J., Vilela, J., Fonseca, M.M., da Fonseca, R.R., Boursot, P. & Alves, P.C. 2014. The elusive nature of adaptive mitochondrial DNA evolution of an arctic lineage prone to frequent introgression. *Genome Biol. Evol.* **6**: 886–896.

Montooth, K.L., Abt, D.N., Hofmann, J.W. & Rand, D.M. 2009. Comparative genomics of *Drosophila* mtDNA: novel features of conservation and change across functional domains and lineages. *J. Mol. Evol.* **69**: 94–114.

Osada, N. & Akashi, H. 2012. Mitochondrial-nuclear interactions and accelerated compensatory evolution: evidence from the primate cytochrome c oxidase complex. *Mol. Biol. Evol.* **29**: 337–346.

Randi, E., Alves, P.C., Carranza, J., Milosevic-Zlatanovic, S., Sfougaris, A. & Mucci, N. 2004. Phylogeography of roe deer (*Capreolus capreolus*) populations: the effects of historical genetic subdivisions and recent nonequilibrium dynamics. *Mol. Ecol.* **13**: 3071–3083.

Reeder, J., Steffen, P. & Giegerich, R. 2007. pknotsRG: RNA pseudoknot folding including near-optimal structures and sliding windows. *Nucleic Acids Res.* **35**: W320–W324.

Rieseberg, L.H. 2009. Evolution: replacing genes and traits through hybridization. *Curr. Biol.* **19**: R119–R122.

Ronquist, F. & Huelsenbeck, J.P. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.

Scott, G.R., Schulte, P.M., Egginton, S., Scott, A.L.M., Richards, J.G. & Milsom, W.K. 2011. Molecular evolution of cytochrome C oxidase underlies high-altitude adaptation in the bar-headed goose. *Mol. Biol. Evol.* **28**: 351–363.

Smith, S., Turbill, C. & Suchentrunk, F. 2010. Introducing mother's curse: low male fertility associated with an imported mtDNA haplotype in a captive colony of brown hares. *Mol. Ecol.* **19**: 36–43.

Sommer, R.S., Fahlke, J.M., Schmolcke, U., Benecke, N. & Zachos, F.E. 2009. Quaternary history of the European roe deer *Capreolus capreolus*. *Mammal Rev.* **39**: 1–16.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**: 2731–2739.

Toews, D.P.L. & Brelsford, A. 2012. The biogeography of mitochondrial and nuclear discordance in animals. *Mol. Ecol.* **21**: 3907–3930.

Toll-Riera, M., Laurie, S. & Alba, M.M. 2011. Lineage-specific variation in intensity of natural selection in mammals. *Mol. Biol. Evol.* **28**: 383–398.

Vorobieva, N.V., Sherbakov, D.Y., Druzhkova, A.S., Stanyon, R., Tsybankov, A.A., Vasil'ev, S.K. *et al.* 2011. Genotyping of *Capreolus pygargus* fossil DNA from Denisova Cave reveals phylogenetic relationships between ancient and modern populations. *PLoS ONE* **6**: e24045.

Wallace, D.C. 2005. A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine. *Annu. Rev. Genet.* **39**: 359–407.

Woolley, S., Johnson, J., Smith, M.U., Crandall, K.A. & McClellan, D.A. 2003. TreeSAAP: selection on amino acid properties using phylogenetic trees. *Bioinformatics* **19**: 671–672.

Yang, Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**: 1586–1591.

Yang, Z. & Nielsen, R. 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *J. Mol. Evol.* **46**: 409–418.

Yang, Z., Wong, W.S.W. & Nielsen, R. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* **22**: 1107–1118.

Zhang, F. & Broughton, R.E. 2013. Mitochondrial-nuclear interactions: compensatory evolution or variable functional constraint among vertebrate oxidative phosphorylation genes? *Genome Biol. Evol.* **5**: 1781–1791.

Zhang, J., Nielsen, R. & Yang, Z. 2005. Evaluation of an improved branch-site likelihood method for detecting

positive selection at the molecular level. *Mol. Biol. Evol.* **22**: 2472–2479.

Zvychaynaya, E.Y., Danilkin, A.A., Kholodova, M.V., Sipko, T.P. & Berber, A.P. 2011a. Analysis of the variability of the control region and cytochrome *b* gene of mtDNA of *Capreolus pygargus* Pall. *Biol. Bull.* **38**: 434–439.

Zvychaynaya, E.Y., Kiryakulov, V.M., Kholodova, M.V. & Danilkin, A.A. 2011b. Roe deer (*Capreolus*) from Moscow area: analysis of mitochondrial control region polymorphism. *Vestn. okhotovedeniya* **8**: 168–172. [in Russian with English summary].

Supporting information

Additional Supporting Information may be found in the online version of this article:

Table S1 List of the 16 mitochondrial genomes of genus *Capreolus* obtained in this study, with their affinity to species-specific mtDNA lineage, isolate source, sampling locality and assignment to major haplogroups of European roe deer (after Randi *et al.*, 2004).

Table S2 List of primer pairs used for PCR and sequencing of mtDNA in genus *Capreolus*.

Table S3 Primer pair amplifying pseudogene in genus *Capreolus*.

Table S4 Predicted changes in tRNA structure between *C. capreolus* and *C. pygargus* mtDNA lineages.

Table S5 Polymorphism within *C. capreolus* and divergence to *C. pygargus* at synonymous and nonsynonymous sites in protein-coding mtDNA genes.

Table S6 Polymorphism within *C. pygargus* and divergence to *C. capreolus* at synonymous and nonsynonymous sites in protein-coding mtDNA genes.

Table S7 Potential positively selected sites detected by branch-site model A implemented in CODEML (PAML) with respective Bayes empirical Bayes (BEB) values.

Data deposited at Dryad: doi:10.5061/dryad.tm8k3

Received 3 April 2014; revised 17 August 2014; accepted 1 September 2014