



## The elusive nature of adaptive mitochondrial DNA evolution of an Arctic lineage prone to frequent introgression

Melo-Ferreira, Jose; Vilela, Joana; Fonseca, Miguel M.; Rodrigues da Fonseca, Rute Andreia; Boursot, Pierre; Alves, Paulo C.

*Published in:*  
Genome Biology and Evolution

*DOI:*  
[10.1093/gbe/evu059](https://doi.org/10.1093/gbe/evu059)

*Publication date:*  
2014

*Document version*  
Publisher's PDF, also known as Version of record

*Document license:*  
[CC BY-NC](https://creativecommons.org/licenses/by-nc/4.0/)

*Citation for published version (APA):*  
Melo-Ferreira, J., Vilela, J., Fonseca, M. M., Rodrigues da Fonseca, R. A., Boursot, P., & Alves, P. C. (2014). The elusive nature of adaptive mitochondrial DNA evolution of an Arctic lineage prone to frequent introgression. *Genome Biology and Evolution*, 6(4), 886-896. <https://doi.org/10.1093/gbe/evu059>

# The Elusive Nature of Adaptive Mitochondrial DNA Evolution of an Arctic Lineage Prone to Frequent Introgression

José Melo-Ferreira<sup>1,\*†</sup>, Joana Vilela<sup>1,2,†</sup>, Miguel M. Fonseca<sup>1,3</sup>, Rute R. da Fonseca<sup>4,5</sup>, Pierre Boursot<sup>6</sup>, and Paulo C. Alves<sup>1,2,7</sup>

<sup>1</sup>CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, InBIO Laboratório Associado, Portugal

<sup>2</sup>Departamento de Biologia da Faculdade de Ciências da Universidade do Porto, Portugal

<sup>3</sup>Departamento de Bioquímica, Genética e Imunologia, Facultad de Biología, Universidad de Vigo, Spain

<sup>4</sup>Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Denmark

<sup>5</sup>CIMAR/CIIMAR, Centro Interdisciplinar de Investigação Marinha e Ambiental, Universidade do Porto, Portugal

<sup>6</sup>Institut des Sciences de l'Evolution, Université Montpellier 2, CNRS, IRD, France

<sup>7</sup>Wildlife Biology Program, College of Forestry and Conservation, University of Montana, Missoula

\*Corresponding author: E-mail: [jmeloferreira@cibio.up.pt](mailto:jmeloferreira@cibio.up.pt).

†These authors contributed equally to this work.

Accepted: March 24, 2014

**Data deposition:** This project has been deposited at GenBank under the accessions KJ397605, KJ397606, KJ397607, KJ397608, KJ397609, KJ397610, KJ397611, KJ397612, KJ397613, and KJ397614.

## Abstract

Mitochondria play a fundamental role in cellular metabolism, being responsible for most of the energy production of the cell in the oxidative phosphorylation (OXPHOS) pathway. Mitochondrial DNA (mtDNA) encodes for key components of this process, but its direct role in adaptation remains far from understood. Hares (*Lepus* spp.) are privileged models to study the impact of natural selection on mitogenomic evolution because 1) species are adapted to contrasting environments, including arctic, with different metabolic pressures, and 2) mtDNA introgression from arctic into temperate species is widespread. Here, we analyzed the sequences of 11 complete mitogenomes (ten newly obtained) of hares of temperate and arctic origins (including two of arctic origin introgressed into temperate species). The analysis of patterns of codon substitutions along the reconstructed phylogeny showed evidence for positive selection in several codons in genes of the OXPHOS complexes, most notably affecting the arctic lineage. However, using theoretical models, no predictable effect of these differences was found on the structure and physicochemical properties of the encoded proteins, suggesting that the focus of selection may lie on complex interactions with nuclear encoded peptides. Also, a cloverleaf structure was detected in the control region only from the arctic mtDNA lineage, which may influence mtDNA replication and transcription. These results suggest that adaptation impacted the evolution of hare mtDNA and may have influenced the occurrence and consequences of the many reported cases of massive mtDNA introgression. However, the origin of adaptation remains elusive.

**Key words:** mitogenomics, *Lepus*, codon evolution,  $d_N/d_S$ , protein structure, positive selection.

## Introduction

Mitochondria are essential components of the cell as they are responsible for the production of most of cellular energy through the oxidative phosphorylation (OXPHOS) pathway, which takes place in protein complexes embedded in the inner mitochondrial membrane (Saraste 1999). The energy generated along the electron gradient of the OXPHOS chain may be converted in usable energy to the cell, in the form of

ATP, or released as heat by the uncoupling of the OXPHOS, which is, for example, important for the thermogenesis of organisms that inhabit cold environments (Cannon and Nedergaard 2004).

The small circular double-stranded mitochondrial DNA (mtDNA) molecule is present in numerous copies in each mitochondrion and encodes 13 of the dozens of proteins that

constitute the OXPHOS complexes. Even if these proteins are involved in key processes of the cell, the direct impact of mtDNA variation on individual fitness, and thus potentially on adaptive evolution, has traditionally been a matter of debate. The evolution of mtDNA has long been considered neutral, mostly for the convenience of phylogenetic and phylogeographic inferences, but evidence that it is subject to various sorts of selective pressures accumulate in the literature. For example, some deleterious effects of mtDNA variation underlie diseases (e.g., Taylor and Turnbull 2005; Tuppen et al. 2010) or low male fertility (e.g., Smith et al. 2010), as expected given its maternal transmission. Also, mtDNA has been suggested to be involved in adaptation to high altitude (Luo et al. 2008; Hassanin et al. 2009; Scott et al. 2011) or temperature (Ballard and Whitlock 2004; Ballard and Rand 2005; Fontanillas et al. 2005), or to coevolve with nuclear genes due to the functional interdependence of mitochondrial and nuclear genomes (Rand et al. 2004; Willett and Burton 2004; Blier et al. 2006; Dowling et al. 2008). Analyses of codon sequence variation have often revealed a role for positive selection in the evolution of mtDNA, as, for example, in salmon (Garvin et al. 2011), whales (Foote et al. 2011), or along the phylogeny of mammals (da Fonseca et al. 2008), among many others. Also, an exhaustive meta-analysis of intraspecific mtDNA sequence variation has shown that mtDNA diversity cannot be predicted by effective population sizes, a genetic pattern that is compatible with the occurrence of recurrent selective sweeps affecting mtDNA (Bazin et al. 2006). This may question the use of mtDNA for inferences of species historical demography but reveals the great potential of studies that attempt to assess the adaptive impact of mtDNA evolution (Galtier et al. 2009).

The interest of studying the adaptive relevance of mtDNA evolution is reinforced by relatively frequent observations of interspecific mtDNA introgression (see e.g., Toews and Brelsford 2012), which are often not paralleled by equivalent nuclear introgression (e.g., Roca et al. 2005; Melo-Ferreira et al. 2012). Such situations offer natural laboratories to study the effect of mtDNA replacement on the evolutionary trajectory of populations (Ballard and Whitlock 2004). The adaptive nature of mtDNA introgression has been suggested in numerous such situations, as, for example, in fish (Nevado et al. 2009; Mee and Taylor 2012), amphibians (Sequeira et al. 2005; Hofman et al. 2007), birds (Irwin et al. 2009; Brelsford et al. 2011), or mammals (Melo-Ferreira et al. 2011; Reid et al. 2012), among other animal groups. However, disentangling the relative contribution of neutral demography and natural selection to mtDNA introgression remains a major challenge when based on the sole patterns of polymorphism of mtDNA sequence in populations (see e.g., Melo-Ferreira et al. 2011).

Hares (*Lepus* spp.) are a particularly suitable group of species to study the role of natural selection in the divergence of mtDNA. First, although the genus diversified recently (4–5 Ma; Matthee et al. 2004), over 30 different species are

currently classified, which colonized the entire globe and are well adapted to contrasting environments (Alves and Hackländer 2008). These different habitats encompass environmental pressures that have been suggested to shape adaptive mtDNA evolution (e.g., Ballard and Whitlock 2004), as different species of hares inhabit for instance arid (e.g., *Lepus capensis*), temperate (e.g., *L. granatensis*), or arctic regions (e.g., *L. timidus*). Second, interspecific mtDNA introgression is a common phenomenon among species of hares (reviewed by Alves et al. 2008). It has been described both in current hybrid zones (e.g., Thulin et al. 2006) and in areas of ancient contact between species (Melo-Ferreira et al. 2005). The mountain hare, *L. timidus*, a boreal/arctic species widely distributed all over the northern Palearctic, has often been found to be the donor, and introgression of mtDNA from this species may affect the northern part of the ranges of many temperate species distributed across Eurasia, from the Iberian Peninsula to China (Alves et al. 2008). Considering cases where mtDNA introgression has been demonstrated (Melo-Ferreira et al. 2012) or suspected (Alves et al. 2008), a dozen different species may have captured mtDNA from *L. timidus* through hybridization. Repeated hybridization during the competitive replacement of *L. timidus* by temperate species during deglaciation after the last glacial maximum may explain the ubiquity of mtDNA introgression of *timidus* origin, simply due to drift in the front of the wave of expansion (Melo-Ferreira et al. 2007). However, recurrent introgression of this mtDNA type may also be caused by natural selection (Melo-Ferreira et al. 2011), presumably driven by adaptation to cold. The adaptive introgression hypothesis would receive credit if one could show that adaptation has contributed to the divergence of the donor and receiver species.

In this study, we analyze 11 complete mtDNA sequences from hares (10 newly obtained), including arctic and nonarctic lineages, and assess whether the pattern of codon evolution along the phylogeny uncovers events of positive selection. Second, we evaluate whether the differences detected among lineages imply functional changes of the encoded peptides. Finally, we characterize mtDNA structure and composition of the different lineages. We hypothesize that positive selection may have governed the evolution of mtDNA in hares, particularly in the arctic lineages. We find evidence of positive selection along these lineages and discuss the origins of the inferred adaptations.

## Materials and Methods

### Sampling and Sequencing of the Complete mtDNA

Tissue samples of a total of ten specimens from nine species of hares were used in this study (table 1). This sampling was designed to span the known phylogenetic tree of hares (Melo-Ferreira et al. 2012) and to include native mtDNA of species that are well adapted to arctic (*L. arcticus*, *L. othus*,

**Table 1**

Specimens Used in This Study (All Individuals Except “eur” Were Newly Sequenced)

Species	Conde	Locality	Sequence Length (bp)	GenBank Accession Numbers
<i>Lepus timidus</i>	tim	Finland	17,755	KJ397605
<i>L. corsicanus</i>	cor	Corsica, France	17,056	KJ397606
<i>L. arcticus</i>	arc	Northwest territories, Canada	16,868	KJ397607
<i>L. othus</i>	oth	Alaska, USA	17,288	KJ397608
<i>L. townsendii</i>	tow	Wyoming, USA	17,732	KJ397609
<i>L. granatensis</i>	gra_int	León, Spain	16,916	KJ397610
<i>L. granatensis</i>	gra_nat	Huelva, Spain	17,765	KJ397611
<i>L. capensis</i>	cap	Mauritania	16,887	KJ397612
<i>L. americanus</i>	ame	Montana, USA	17,042	KJ397613
<i>L. californicus</i>	cal	Texas, USA	16,938	KJ397614
<i>L. europaeus</i>	eur	Skane, Sweden	17,734	NC_004028

and *L. timidus*), boreal (*L. americanus*), and temperate (*L. californicus*, *L. capensis*, and *L. granatensis*) environments, and haplotypes introgressed from *L. timidus* into other species (*L. granatensis*, *L. corsicanus*, and possibly *L. townsendii*) (see Melo-Ferreira et al. 2012).

Extraction of total DNA was performed from ear tissues using the EasySpin Genomic DNA Tissue Kit, Citomed, or liver tissues with the QIAGEN DNeasy kit, following the manufacturer's instructions. In *L. granatensis*, we selected specimens harboring either *L. granatensis* or *L. timidus* mtDNA based on the polymerase chain reaction (PCR)-restriction fragment length polymorphism (RFLP) assay described by Melo-Ferreira et al. (2005). The complete mtDNA of the sampled specimens was amplified in two overlapping fragments of approximately 9,000 bp (specimens tim, tow, and gra\_nat in table 1) or three overlapping fragments of approximately 6,000 bp using the primers indicated in [supplementary table S1, Supplementary Material](#) online, and the long range Phusion Flash High-Fidelity PCR Master Mix (Finnzymes) protocol. This long-range PCR approach allowed minimizing the probability of amplifying nuclear copies of mtDNA sequences (NUMTs), because these copies tend to be shorter (Calvignac et al. 2011). Final PCR mix included 6  $\mu$ l of Master Mix and 0.6  $\mu$ l of each primer at 20  $\mu$ M, in a final volume of 8.6  $\mu$ l. The following PCR cycle conditions were used: 10 s at 98 °C; 30 cycles with 1 s at 98 °C for the denaturation, 5 s for the annealing, and 4 min at 72 °C for the extension, and 1 min at 72 °C for the final extension step.

For samples tim, tow, and gra\_nat (table 1), nested PCRs of overlapping portions of approximately 1,200 bp were performed to increase amounts of amplified DNA (see primers in [supplementary table S1, Supplementary Material](#) online). Nested PCRs were performed in a final mix volume of 9.5  $\mu$ l, using 3.5  $\mu$ l of a PCR Master Mix (QIAGEN Multiplex PCR Kit) and 0.2  $\mu$ l of each primer at 20  $\mu$ M, using the following temperature cycles: 15 min at 94 °C; 35 cycles of denaturation/annealing/extension with 30 s at 94 °C for denaturation, 45 s for annealing, and 1.15 min at 72 °C for extension and 20 min

at 60 °C for the final extension step. Nested PCR products were sequenced using the forward and reverse PCR primers following the ABI PRISM BigDye Terminator Cycle Sequencing 3.1 standard protocol and using an ABI PRISM 3130 sequencer (Applied Biosystems).

The long-range PCR products of the remaining specimens (table 1) were sequenced using 454 technology and assembled (using Mira assembler v3.2.1; Chevreux et al. 1999) by Ecogenics (Switzerland). Tablet v.1.12.03.26 (Milne et al. 2010) was used to assess Phred quality scores and coverage along the obtained contigs. Only sites with quality score above 20 and coverage above 10 $\times$  were accepted. Sequenced portions that did not satisfy these criteria were coded as missing data. Those portions with missing data were resequenced using the nested PCR strategy described earlier and Sanger sequencing, which also allowed verifying the concordance of the two sequencing procedures.

### Phylogenetic Analysis

The complete mtDNA sequences of ten specimens sequenced in this work and of *L. europaeus* (GenBank Acc. Nr. NC\_004028) were aligned using ClustalW (Thompson et al. 1994). For the phylogenetic reconstruction, a sequence of the European wild rabbit, *Oryctolagus cuniculus*, complete mtDNA was added to the alignment and used as outgroup (GenBank Acc. Nr. NC\_001913).

The sequences of the 13 mtDNA protein-coding genes were used for phylogeny estimation. The best fit model of evolution for each gene was determined with jModelTest v0.1.1 (Posada 2008), applying the Akaike information criterion. The overlapping regions between ATP8/ATP6, ND4L/ND4, and ND5/ND6 were considered as a separate partition.

Phylogenetic inference was performed using the Bayesian methodology implemented in BEAST v1.7.4 (Drummond et al. 2012), applying the appropriate models determined with jModelTest, or the next most parameterized model available in BEAST. The uncorrelated lognormal relaxed clock (Drummond et al. 2006) and the Yule tree prior were used.

Three independent runs of 100,000,000 Markov chain Monte Carlo (MCMC) generations, sampling at every 10,000 generations, were performed and the stability of the runs checked using Tracer v1.5 (Rambaut and Drummond 2009). The first 10% of each run was discarded as burn-in, the results of the three runs were combined using LogCombiner v1.7.4, and the information of the trees summarized using TreeAnnotator v1.7.4, both part of the BEAST package (Drummond et al. 2012).

### Tests of Selection

The effect of natural selection on the evolution of the mtDNA protein-coding genes was assessed by comparing the number of nonsynonymous changes per nonsynonymous sites ( $d_N$ ) with that of synonymous changes per synonymous site ( $d_S$ ) using maximum likelihood and Bayesian approaches. Purifying selection is characterized by a  $d_N/d_S$  ratio ( $\omega$ )  $< 1$ , whereas values  $> 1$  are indicative of positive selection. The phylogeny estimated in this work, unrooted and excluding the outgroup, was used in all analyses of selection. Also, all methodologies described below were applied to three different alignments: 1) the 13 protein-coding genes concatenated; 2) protein-coding genes concatenated by respiratory complex; and 3) single genes. These separate analyses allowed verifying whether natural selection affected preferentially different OXPHOS complexes or genes. Also, the concatenation of genes results in larger data sets and higher number of variable codons, which may help improve the power of the tests (Yang and dos Reis 2011).

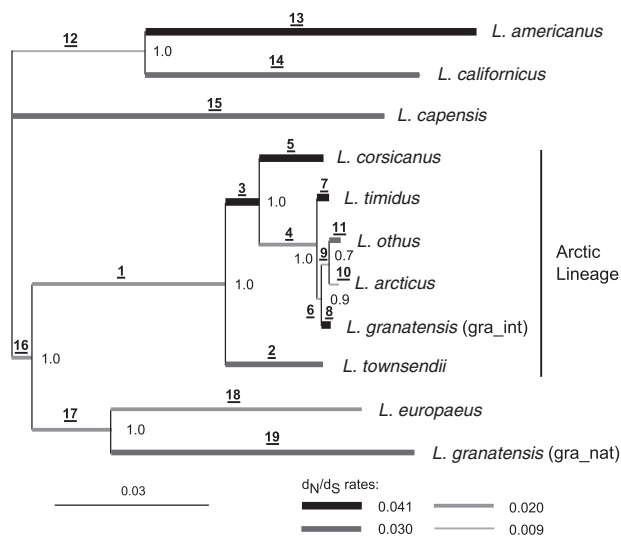
First, we assessed whether selection could be detected at individual codons, independently of the branch of the phylogeny, using the sites models (Nielsen and Yang 1998; Yang et al. 2000) implemented in CODEML, in the PAML 4 package (Yang 2007). Different  $\omega$  starting values were used to avoid local optima. Nested null (restricting  $\omega$  to values  $\leq 1$ ) and positive selection (allowing  $\omega$  to take values  $> 1$ ) models were compared using a likelihood ratio test (LRT): M1a versus M2a, M7 versus M8, and M8a versus M8.

We further assessed the impact of selection along the mtDNA phylogeny of hares using additional codon models to estimate the rates of synonymous and nonsynonymous substitutions (Pond and Frost 2005c). Four methods implemented on the DATAMONKEY web server (<http://www.datamonkey.org/> [last accessed April 15, 2014]; Pond and Frost 2005b), Single Likelihood Ancestral Counting (SLAC), Fixed Effects Likelihood (FEL), Random Effects Likelihood (REL), and PRoperty Informed Models of Evolution (PRIME), were used. After reconstructing ancestral sequences, SLAC compares normalized expected and observed numbers of synonymous and nonsynonymous substitutions per variable site. FEL compares the instantaneous synonymous site rate ( $\alpha$ ) and the instantaneous nonsynonymous site rate ( $\beta$ ) on a per site basis, without assuming a prior  $d_N/d_S$  distribution. In

PRIME,  $\beta$  also depends on the properties of the amino acid variants, analyzing two sets of five amino acid properties (Atchley et al. 2005; Conant et al. 2007). REL allows for heterogeneity both in synonymous and nonsynonymous rates, by fitting discrete distributions of rates across sites and then inferring the rate per site. Sites with cut-off values of  $P < 0.25$  in SLAC and FEL (in small data sets, high nominal  $\alpha$ -levels in these tests are advised; Pond and Frost 2005c) and  $P < 0.05$  in PRIME and Bayes factors  $> 50$  in REL were considered as candidates to have evolved under positive selection. In all analyses performed in DATAMONKEY, the most suited model of evolution for each data set, directly estimated on this web server, was used.

Branch models (Yang 1998; Yang and Nielsen 1998), which allow  $\omega$  to vary along branches of the phylogeny, were also fit to the data. The likelihood of a single  $\omega$  (null model) was tested against the likelihood of a model allowing different  $\omega$  in each branch (full model) using an LRT. In case of rejection of the null model, models considering two  $\omega$  rates were tested. These simplified models were based on the hypothesis that a distinct selection regime could affect the mtDNA lineage of species adapted to arctic or boreal conditions: arctic lineages versus the rest, and arctic lineage and *L. americanus* (hereafter named arctic/boreal group) versus the rest. Finally, the GA-branch model selection analysis (Pond and Frost 2005a) implemented in DATAMONKEY was used to assess which combinations of rates per branches allowed the best fit to the data without prior assumptions.

The methods described earlier are best suited to detect events of pervasive and recurrent positive selection because 1) site-based methods average  $\omega$  over the entire phylogeny (Yang et al. 2000), whereas 2) branch-based methods average  $\omega$  over all codons in a protein (Yang 1998). These methods may thus be complemented with branch-site tests that focus on particular residues along specific lineages, which provide increased power to detect episodic selection (Yang et al. 1995; Murrell et al. 2012). Thus, the branch-sites model (Yang et al. 2005; Zhang et al. 2005), implemented in CODEML, was also used here. The likelihood of a model (A) in which site classes in a background lineage are restricted to values  $\leq 1$ , while site classes of the foreground lineage are allowed to take values  $> 1$ , is compared with the likelihood of a null model that fixes  $\omega$  at 1 in the foreground lineage using an LRT. This method has been shown to be rather robust to low numbers of codons in the alignment (Yang and dos Reis 2011). All individual branches were tested as possible foreground lineage. In addition, numerous combinations of branches were tested as the foreground lineage, among which were branches 1–11 (arctic lineage), branches 1–11 and 13 (arctic/boreal group), branches 3 and 5 (evolution of lineage introgressed into *L. corsicanus*), 3 and 4 (evolution of the native arctic species, *L. timidus*, *L. othus* and *L. arcticus*), and 6 and 8 (evolution of lineage introgressed into *L. gransis*) (see branch numbers in fig. 1). The Bayes empirical



**Fig. 1.**—Unrooted phylogeny of 11 hare specimens (codes in parenthesis correspond to those indicated in table 1) based on 13 protein-coding mtDNA sequences, estimated using Bayesian inference. Branch number is indicated on the respective branch and posterior probabilities next to the nodes. Two North American hares, *Lepus americanus* and *L. californicus* are the first to split, followed by the African *L. capensis* and then the European *L. granatensis* (native haplotype) and *L. europaeus*. Finally, one clade named “arctic” includes the sequences from the arctic/boreal *L. timidus*, *L. arcticus*, and *L. othus*, and also from *L. corsicanus* and *L. granatensis* (temperate species affected by ancient mtDNA introgression of *L. timidus* origin) (Melo-Ferreira et al. 2012), and *L. townsendii* (a North American species for which mtDNA introgression from *L. timidus* has been suspected but was never demonstrated) (Alves et al. 2008; Melo-Ferreira et al. 2012). Branch thickness and gray shade indicate the inferred  $d_N/d_S$  rates that provide the best fit to the data as estimated by the GA-branch model selection scheme in DATAMONKEY.

Bayes method was used to identify sites under positive selection in case of rejection of the null model (Yang et al. 2005).

In addition, the mixed effects model of evolution (MEME), a branch-site method incorporated in the DATAMONKEY server, was used to test for both pervasive and episodic diversifying selection (Murrell et al. 2012). MEME is a generalization of FEL but models variable  $\omega$  across lineages at individual sites, restricting  $\omega$  to be  $\leq 1$  in a proportion  $p$  of branches and unrestricted at a proportion  $(1-p)$  of branches per site. Positive selection was inferred with this method for a  $P$  value  $< 0.05$ .

TreeSAAP (Woolley et al. 2003), which takes into account the magnitude of the impact of the amino acid replacements on local physicochemical properties, was also applied here. Radical magnitudes of changes  $\geq 6$ , with  $P$  value  $\leq 0.001$ , were here considered as indicating directional positive selection for a given physicochemical property. Also, only amino acid properties that had an accuracy of detection of selection higher than 85% were considered (McClellan and Ellison

2010), and therefore, 11 of the 31 properties analyzed by TreeSAAP were excluded.

### Analyses of Protein Function and Three-Dimensional Structure

The potential impact of the observed amino acid substitutions was assessed considering their location relative to known functional domains of the proteins and the physicochemical properties of the amino acid, such as size and charge. A three-dimensional protein homology model was built for subunits, whenever appropriate templates were available, and the amino acid substitutions tested in silico. Protein models were built with PHYRE 2 (<http://www.sbg.bio.ic.ac.uk/phyre2>, last accessed April 7, 2014; Kelley and Sternberg 2009) for the following subunits (PDBid of the respective template in parenthesis): ND2 (3RKO), ND4 (3RKO), ND5 (3RKO), COXI (1V54), COXII (1V55), COXIII (1V54), and CYTB (3CWB). We have mutated each residue in the structure into the alternative amino acid, and the result was inspected site by site, verifying predictable structural alterations due to steric clashes and inspecting possible changes in polarity close to relevant protein regions such as proton channels and redox centers. This way it was possible to evaluate whether the detected amino acid changes are or not potentially neutral from the protein function point of view.

### Analysis of mtDNA Composition and Structure

Sequence annotation was performed by using MITOS (Bernt et al. 2012) to identify protein-coding genes and rRNAs and ARWEN (Laslett and Canback 2008) to identify tRNA genes. Gene limits were checked by comparison with orthologous mtDNA sequences of published Leporidae species (*L. europaeus* and *O. cuniculus*) and by performing sequence similarity in National Center for Biotechnology Information using BLASTX (Altschul et al. 1997) against the nonredundant protein sequences database. Analyses of variation in start and stop codon usage of protein-coding genes and in limits of overlapping genes were performed with in-house Perl scripts. The origin of light-strand replication was manually identified and the minimum free energy of its secondary structure was calculated by using the Vienna RNA Websuite, DNA energy parameters (Gruber et al. 2008). Pseudo-tRNAs or cloverleaf secondary structures—within the major mtDNA noncoding region (the control region or D-loop)—were identified using ARWEN.

## Results

### Complete mtDNA Sequence Data and Phylogenetic Inference

Sequences of the nearly complete mitogenomes of ten specimens of hares were newly obtained in this study. Total sizes vary from 16,868 bp in *L. arcticus* to 17,765 bp in native

*L. granatensis* (table 1, where GenBank accession numbers are shown). Some portions of the control region were incomplete, and the serine tRNA was incomplete in *L. arcticus* and *L. californicus*. No differences were detected whenever portions of the mitogenomes were sequenced both with Sanger and 454 strategies (total of 19,176 bp resequenced). The long-range PCR approach employed prior to sequencing is expected to avoid the amplification of NUMTs. In keeping, no protein-coding genes displayed premature stop codon or disruption of the reading frame. Also, sequences of the overlapping PCR fragments were identical, which would be unexpected if some fragment was of nuclear origin.

The unrooted phylogenetic tree displaying the evolutionary relationships inferred for the mtDNA between the 11 hare specimens analyzed in this study (10 newly sequenced) estimated using BEAST v1.7.4 (Drummond et al. 2012) is depicted in figure 1 (see rooted version in supplementary fig. S1, Supplementary Material online; the evolutionary models selected for each sequence partition are shown in supplementary table S2, Supplementary Material online). The obtained

topology agrees with previous inferences based on shorter mtDNA fragments (Melo-Ferreira et al. 2012).

Evidence of Natural Selection

We applied several methods designed to detect evidence of positive selection affecting mtDNA protein-coding genes throughout the phylogenetic tree of hares. Only the sites model implemented in CODEML (Nielsen and Yang 1998; Yang et al. 2000) did not provide evidence of positive selection. The suggestions of positive selection of the several site-specific analyses are summarized in table 2, where only sites with more than one method suggesting evolution under positive selection are shown (complete results are in supplementary table S3, Supplementary Material online). The branch of the phylogenetic tree where positive selection was inferred is indicated whenever a particular method allowed such inference—for example, branch-sites model of CODEML (Yang et al. 2005; Zhang et al. 2005), MEME (Murrell et al. 2012), or TreeSAAP (Woolley et al. 2003). For the remaining methods (REL, FEL, SLAC, and PRIME; Pond and Frost 2005c), only the

Table 2

Sites Affected by Positive Selection According to Seven Different Methods (Only Sites with Suggestion of Selection by Two or More Methods Are Shown; See Complete Results in supplementary table S3, Supplementary Material online, Where the Amino Acid Variation Is Indicated)

Gene	Site	Tests of Selection <sup>a</sup>							
		SLAC <sup>b</sup>	REL <sup>b</sup>	FEL <sup>b</sup>	PRIME <sup>c</sup>	MEME <sup>d</sup>	Branch Sites <sup>d</sup>	TreeSAAP	
						Branch <sup>d</sup>		Magnitude of Change <sup>e</sup>	
ATP8	38	x <sup>fgh</sup>	—	—	—	—	—	(1; 5; 15) <sup>gh</sup>	6 (1; 15) ↓ P <sub>α</sub> ; 6 (5) ↑ P <sub>α</sub>
	49	—	x <sup>fgh</sup>	x <sup>fg</sup>	x <sup>h</sup> (P: -7.891)	(3; 4; 5) <sup>fgh</sup>	(3+5) <sup>gh*</sup>	—	—
CYTb	43	—	—	—	x <sup>f</sup> (V: -6.997)	(7) <sup>fh</sup>	—	(7) <sup>fh</sup>	8 ↑ pK'
ND1	49	—	x <sup>h</sup>	x <sup>f</sup>	—	—	—	—	—
ND2	92	—	x <sup>h</sup>	x <sup>f</sup>	—	—	—	—	—
ND4	20	—	x <sup>fh</sup>	x <sup>f</sup>	—	—	—	—	—
	187	x <sup>fgh</sup>	x <sup>fh</sup>	x <sup>f</sup>	—	—	—	(1; 13) <sup>h</sup>	6 ↓ P <sub>α</sub>
	337	—	x <sup>fh</sup>	x <sup>f</sup>	—	—	—	(3; 18) <sup>fgh</sup>	8 ↑ pK'
ND4L	351	—	x <sup>fh</sup>	x <sup>f</sup>	—	—	—	(1; 13) <sup>fgh</sup>	8 ↓ pK'
	6	—	x <sup>f</sup>	x <sup>f</sup>	—	—	—	(2; 13) <sup>fgh</sup>	8 ↓ pK'
ND5	21	—	x <sup>fh</sup>	—	—	—	—	(1; 14) <sup>fgh</sup>	8 ↑ pK'
	410	—	x <sup>fh</sup>	—	—	—	—	(19) <sup>fgh</sup>	8 ↑ pK'
ND6	531	x <sup>fgh</sup>	x <sup>fh</sup>	x <sup>fgh</sup>	x <sup>gh</sup> (P: 20.000; V: -10.177; C/pHi: 18.683)	—	—	(5; 11; 13; 15) <sup>fh</sup>	6 ↓ P <sub>α</sub>
	14	—	x <sup>h</sup>	—	—	—	—	(2; 15) <sup>gh</sup>	8 ↑ pK'
	108	—	x <sup>gh</sup>	x <sup>f</sup>	—	—	—	—	—

NOTE.—SLAC, REL, FEL (Pond and Frost 2005c); PRIME (Pond and Frost 2005b); branch sites (Yang et al. 2005; Zhang et al. 2005); MEME (Murrell et al. 2012); and TreeSAAP (Woolley et al. 2003).

<sup>a</sup>Null model rejection thresholds: TreeSAAP, 0.001; SLAC and FEL, 0.25; PRIME, 0.05; MEME 0.05; branch sites, 0.05; Bayes Factor for REL = 50.

<sup>b</sup>In SLAC, REL, and FEL, rejection of null model is indicated by "x."

<sup>c</sup>In PRIME, rejection of null model is indicated by "x," and the property under selection and its weight are indicated in parenthesis. C, charge; pHi, isoelectric point; P, polarity; V, volume.

<sup>d</sup>In the branch sites, MEME, and TreeSAAP tests, the branch(es) where positive selection was inferred is indicated (see fig. 1 for correspondence); the Bayes empirical Bayes (Yang et al. 2005) posterior probabilities of codon detection of the branch sites results are \*0.967 (<sup>g</sup>) and 0.893 (<sup>h</sup>).

<sup>e</sup>Magnitude of amino acid changes inferred with TreeSaap: increase (↑) and decrease (↓) in amino acid properties in a given branch number is indicated. P<sub>α</sub>: alpha helical tendencies; pK': equilibrium constant (ionization of COOH).

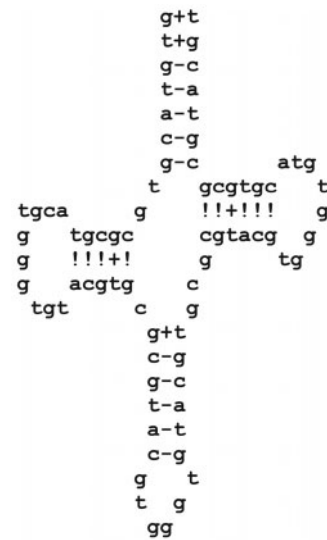
Data sets: <sup>f</sup>individual mtDNA protein-coding genes, <sup>g</sup>mtDNA protein-coding genes concatenated by OXPHOS complex, and <sup>h</sup>all mtDNA protein-coding genes concatenated.

amino acid substitutions are shown in table 2. A total of 18 different codons in eight genes were here suggested to have evolved under recurrent positive selection by REL, FEL, or SLAC, of which ten sites were concordant in two of the methods (table 2). PRIME suggested that adaptive evolution may have affected five sites (indicated by the negative weight of amino acid properties “isoelectric point,” “volume,” and “polarity”; see table 2 and [supplementary table S3, Supplementary Material](#) online), three of which were also pinpointed by other analyses. MEME, which is particularly sensitive to events of episodic selection, suggested that six sites of five genes evolved under positive selection, which may have affected several lineages along the evolution of hares. Only two codons (in the ATP8 and ND4 genes) were suggested as a focus of positive selection by the branch-sites models analysis of CODEML (table 2; [supplementary table S3, Supplementary Material](#) online). TreeSAAP, which incorporates the magnitude of the physicochemical impact of the amino acid substitutions in the analysis, identified numerous codons (103) as possible focus of positive selection. Among the properties in which radical changes were suggested, “alpha helical tendencies” and “equilibrium constant (ionization of COOH)” were the most represented ([supplementary table S3, Supplementary Material](#) online). Several codons (15), in the ATP8, CYTB, and six of the seven NADH dehydrogenase complex genes, were repeatedly reported as positively selected by more than one methodology (table 2). Most notably, codon 49 of ATP8 and codon 531 of ND5 were indicated by five of the eight site-specific methods used here and codon 187 of ND4 by four of these methods.

In the branch model analysis of CODEML (Yang 1998; Yang and Nielsen 1998), the null model of a single  $\omega$  along the phylogeny was rejected only in the analysis of the complete mtDNA protein-coding genes ( $P < 0.05$ ). The model considering two  $\omega$  rates along the phylogeny—in the arctic/boreal group (arctic lineage and *L. americanus*) and elsewhere—was significantly favored over the null model ( $P < 0.05$ ). The model selection procedure employed using GA-branch suggested that the model that optimizes different  $\omega$  rates along the tree is composed of four different rates (fig. 1). In any case, the inferred  $\omega$  rate was consistently well below 1 in all estimates obtained.

#### Little Impact of Mutations on Protein Structure and Function

The predicted impact of amino acid substitutions on the protein function was evaluated for subunits with suitable homologs with available protein structure (see [supplementary table S4, Supplementary Material](#) online). This analysis suggested all substitutions to be neutral from a functional point of view, both regarding potential local changes in structure arising from steric clashes resulting from changes in size, and the changes in polarity and charge nearby the



**Fig. 2.**—Graphical representation of *Lepus granatensis* (*gra\_int*) introgressed mtDNA control region cloverleaf structure (at positions 16224–16303). Nucleotide complementarities are represented by “!” or “-,” whereas nucleotide mismatches are depicted by “+.”

redox centers and putative proton channels, where even small changes may have a large impact.

#### mtDNA Composition

All mitogenomes sequenced have the same gene arrangement and content typically found in mammalian mtDNAs ([supplementary fig. S2, Supplementary Material](#) online). No variation was found in start codons of protein-coding genes, but different stop codons (TAA and TAG) were identified within four genes: COX1, COX2, ND5, and ND6. This variation should not have functional consequences, as different mammalian mtDNA stop codons are not expected to significantly change gene expression.

A region inside the major noncoding portion of the mtDNA was predicted to form stable stem-and-loop secondary structures with minimum free energy prediction varying between  $-123.60$  and  $-49.74$  kcal/mol. One or more copies of a conserved cloverleaf structure (pseudo-tRNA; fig. 2) in this region were inferred only in the haplotypes from the arctic species (*L. arcticus*: three copies, *L. othus*: three copies, and *L. timidus*: one copy), *L. granatensis* carrying the introgressed *L. timidus* haplotype (five copies), *L. corsicanus* (four copies), and *L. townsendii* (two copies). When two or more cloverleaf structures were present, they partially overlapped.

#### Discussion

mtDNA may play an important role in the adaptation to particular habitats, and groups of species that live in contrasting ecological conditions are valuable models to assess this role.



We used here a combination of phylogenetic analyses of codon evolution and theoretical inferences of the impact of amino acid substitutions on protein structure and function to understand whether natural selection governed the evolution of mtDNA in hares, and to provide insights on the origins of the adaptation.

### Nonneutral Evolution of mtDNA in Hares

Estimates of  $\omega$  rates along the branches of hares' mtDNA phylogeny were consistently below 1, which is indicative of pervasive purifying selection affecting all mtDNA protein-coding genes (fig. 1). This result is not surprising because purifying selection has been shown to be a predominant force shaping mitogenomic evolution (Ruiz-Pesini et al. 2004; Bazin et al. 2006; Meiklejohn et al. 2007; Rutledge et al. 2010; Sun et al. 2011). However, we found the  $\omega$  rates to vary along the tree. Considering either a simplified two-rate model (arctic/boreal group and rest of the tree; [supplementary table S5, Supplementary Material](#) online) or the four-rate model (based on the GA-branch model selection procedure; fig. 1), higher rates were generally estimated in branches of the arctic clade and of *L. americanus*, a boreal North American species also well adapted to cold climates (fig. 1). Note, however, that branches represented by a single genome may not be representative of the respective evolutionary lineage, and inferences of natural selection must, in these cases, be looked with caution. Branch 3, which represents a subset of the arctic clade, was the sole interior branch where the higher class  $\omega$  was estimated. Increased  $\omega$  could result from adaptive evolution at some sites, in the face of the dominant purifying selection regime, or from relaxation of purifying selection in some lineages (Bazin et al. 2006; Meiklejohn et al. 2007). The latter phenomenon should, however, affect all genes alike (Tomasco and Lessa 2011), and the rate variation we find among genes in that respect may thus result from branch- and gene-specific positive selection ([supplementary tables S5 and S6, Supplementary Material](#) online).

Purifying and positive selection may occur at a given site at different times during the evolution of a group (Crandall et al. 1999; Chen and Sun 2011), and such situations may complicate inferences of positive selection due to a decreased sensitivity of the tests (Chen and Sun 2011). Also, the history of hare radiation is relatively recent (4–5 Ma; Matthee et al. 2004), and the relatively low levels of divergence among lineages (average 0.69 nucleotide substitutions per codon over the whole hare phylogeny) leads to a reduced power of the tests of selection (Anisimova and Yang 2007). Still, site- and branch-site-based tests of selection were able to detect various instances of sites with high  $d_N/d_S$ , many reported by different methodologies (15 sites; table 2), and thus that possibly evolved under positive selection.

We addressed this question by applying methods that account for site- and branch-specific evolution that are better

suited at detecting episodic selection on a small fraction of amino acids (Zhang et al. 2005; Murrell et al. 2012). We paid particular attention to the hypothesis that selection may have had preferential impact in the arctic clade. Indeed, many of the reports of positive selection involve the arctic clade or sites that distinguish this clade from the remaining hares (e.g., internal branches 1, 3, and 4; table 2 and [supplementary table S3, Supplementary Material](#) online).

Our results suggest that positive selection may have had a particular effect on ATP8, CYTB, and several of the NADH dehydrogenase genes (table 2). Previous studies have suggested that selection in these genes may be related with the efficiency of energy conversion (ATP8; Shen et al. 2010; Zsurka et al. 2010), which may influence adaptation to cold climates (Wallace 2007), the establishment of the proton gradient necessary for ATP production (CYTB; McClellan et al. 2005), or the efficiency of the proton-pump (Brandt 2006). Interestingly, we found that 30 of the sites reported under positive selection in this work coincide with other studies (compiled in [supplementary table S7, Supplementary Material](#) online), some of which proposing a relationship with climate adaptation (e.g., Zsurka et al. 2010). It is thus striking that the site-specific selection reported in ATP8 lied on internal branches of the arctic clade (table 2). In some cases, the reported mutations lie on introgressed haplotypes (branches 5 and 8; table 2 and [supplementary table S3, Supplementary Material](#) online), but these were represented by single individuals in our phylogeny. It is thus difficult to assess whether this sampling is representative of the whole introgressed haplotypes and if these inferences indicate changes in the selection regime induced by introgression. The results obtained here do not demonstrate an adaptive drive of mtDNA reticulation (suggested e.g., by Alves et al. 2008; Melo-Ferreira et al. 2011) but show that positive selection may have punctuated the evolution of mtDNA, most notably in the lineages of species thriving in arctic regions, eventually by adaptation to cold as suggested in other species (Fontanillas et al. 2005). This therefore puts forward a possible impact of the high frequencies of mtDNA of arctic origin on the evolution and adaptation of the temperate species where they are found introgressed.

### On the Origin of mtDNA Adaptation in Hares

The inference of positive selection on the evolution of mtDNA in hares, affecting specific sites, some of which associated with lineages of species adapted to boreal climates, questions the impact of such mutations on the structure and function of the encoded proteins. We modeled the protein structure of the encoded proteins of the different lineages and attempted to predict the impact of amino acid mutations, given their location and consequences on the protein structure and its physicochemical properties. Even though the TreeSAAP analysis suggested that several of the changes could have had a

radical impact on the biochemical properties of the amino acid in question (table 2), we assessed that their impact on the biochemical structure of the encoded proteins should be residual. This analysis was based on the current knowledge on functional aspects of mtDNA-encoded proteins (which are quite conserved from mammals to bacteria) and on available structural models (in several cases, including ATP8, the protein three-dimensional structures could not be modeled). However, mtDNA-encoded proteins are involved in complex interactions with nuclear DNA encoded proteins, and there is evidence of cytonuclear coevolution (Rand et al. 2004). This interaction is not restricted to the OXPHOS pathway but includes many other key processes from the transcription and translation of mtDNA proteins to the regulation of the communication between the organelles (Timmis et al. 2004; Cannino et al. 2007; Christian and Spremulli 2012; Xu et al. 2012). Only an integrated modeling of the cytonuclear protein–protein interactions could provide an accurate view of the impact of mtDNA nonsynonymous changes, and the functional impact of the amino acid changes may have been underestimated by our analysis.

Analyses of nonprotein-coding portions of the mtDNA may provide additional insights on the nature of mitochondrial adaptation in hares, because mutations in the regulatory regions of mtDNA may have an effect on mtDNA activity (Suissa et al. 2009). Although most of the noncoding part of hare mtDNA was highly conserved across lineages, a putative cloverleaf structure, with variable copy number, was inferred in the control region of the arctic lineage, including the introgressed *L. granatensis* and *L. corsicanus* (fig. 2). The copies were estimated to reach high negative free energy in its secondary structure, which suggests high stability and high probability of spontaneous formation. Such unusual structures may influence the efficiency of mtDNA replication and transcription (Brown et al. 1986; Pereira et al. 2008) and thus regulate mtDNA content in cells, which may impact the efficiency of mitochondrial function (Suissa et al. 2009), and contribute to adaptation (Cheng et al. 2013). The presence of this structure only in the arctic lineage and the apparent variable copy number is intriguing and raises the interesting possibility that it may have contributed to regulatory-driven adaptation. This should be the focus of future research.

## Conclusion

This study provides new insights into the adaptive nature of mtDNA evolution, which may have facilitated the colonization of hares of diverse and contrasting environments. Still, the origin of the inferred positive selection remains elusive, given the small predicted impact of the observed amino acid substitutions in protein function. We have, however, only evaluated this impact in the mtDNA encoded proteins, which are part of complex networks with a far greater number of proteins encoded by nuclear DNA. The frequent occurrence of

mtDNA introgression among hares provides natural mtDNA replacement experiments that are valuable to investigate how coevolution and gene regulation in these networks contributed to mtDNA evolution. Two major questions are raised by the occurrence of these massive interspecific mtDNA transfers: 1) was introgression favored by natural selection? 2) and if not, could it have a secondary impact on adaptation and nucleocytoplasmic coevolution? This work reinforces the need to address such questions.

## Supplementary Material

Supplementary tables S1–S7 and figures S1 and S2 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

## Acknowledgments

This work was supported by FEDER (Fundo Europeu de Desenvolvimento Regional) funds through the Programa Operacional Factores de Competitividade (COMPETE program; FCOMP-01-0124-FEDER-014266) and Portuguese national funds through the FCT (Fundação para a Ciência e a Tecnologia; PTDC/BIA-EVF/111931/2009 research project), by Programa Operacional Potencial Humano (POPH)–Quadro de Referência Estratégico Nacional (QREN) funds from the European Social Fund and Portuguese Ministério da Educação e Ciência (FCT, SFRH/BPD/43264/2008 and SFRH/BPD/70654/2010 post-doctoral grants to J.M.-F. and M.M.F., respectively, and SFRH/BSAB/1278/2012 sabbatical grant to P.C.A.), and by project “Genomics and Evolutionary Biology” cofinanced by North Portugal Regional Operational Programme 2007/2013 (ON.2 – O Novo Norte), under the National Strategic Reference Framework (NSRF), through the European Regional Development Fund (ERDF). The *Lepus arcticus* and *Lepus othus* tissue samples were yielded by the University of Alaska Museum (original codes: arc – UAM47253, oth – UAM42143). The authors also thank José Carlos Brito, Ellen Cheng, Christian Gortazar, Klaus Hackländer, L. Scott Mills, Christian Pietri, and Jorge Pires for providing samples. They thank Josefa González and three anonymous reviewers for helpful suggestions that helped improving the quality of this work. This research was conducted in the scope of the International Associated Laboratory (LIA) “Biodiversity and Evolution.”

## Literature Cited

- Altschul SF, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389–3402.
- Alves PC, Hackländer K. 2008. Lagomorph species: geographical distribution and conservation status. In: Alves PC, Ferrand N, Hackländer K, editors. *Lagomorph biology: evolution, ecology, and conservation*. Berlin (Germany): Springer-Verlag. p. 395–405.
- Alves PC, Melo-Ferreira J, Freitas H, Boursot P. 2008. The ubiquitous mountain hare mitochondria: multiple introgressive hybridization in

- hares, genus *Lepus*. *Philos Trans R Soc Lond B Biol Sci*. 363: 2831–2839.
- Anisimova M, Yang Z. 2007. Multiple hypothesis testing to detect lineages under positive selection that affects only a few sites. *Mol Biol Evol*. 24: 1219–1228.
- Atchley WR, Zhao J, Fernandes AD, Druke T. 2005. Solving the protein sequence metric problem. *Proc Natl Acad Sci U S A*. 102:6395–6400.
- Ballard JWO, Rand DM. 2005. The population biology of mitochondrial DNA and its phylogenetic implications. *Annu Rev Ecol Syst*. 36: 621–642.
- Ballard JWO, Whitlock MC. 2004. The incomplete natural history of mitochondria. *Mol Ecol*. 13:729–744.
- Bazin E, Glemin S, Galtier N. 2006. Population size does not influence mitochondrial genetic diversity in animals. *Science* 312:570–572.
- Bernt M, et al. 2012. MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol Phylogenet Evol*. 69:313–319.
- Blier PU, Breton S, Desrosiers V, Lemieux H. 2006. Functional conservatism in mitochondrial evolution: insight from hybridization of arctic and brook charrs. *J Exp Zool B Mol Dev Evol*. 306B:425–432.
- Brandt U. 2006. Energy Converting NADH: quinone oxidoreductase (complex I). *Annu Rev Biochem*. 75:69–92.
- Brelsford A, Mila B, Irwin DE. 2011. Hybrid origin of Audubon's warbler. *Mol Ecol*. 20:2380–2389.
- Brown GG, Gadaleta G, Pepe G, Saccone C, Sbisà E. 1986. Structural conservation and variation in the D-loop-containing region of vertebrate mitochondrial DNA. *J Mol Biol*. 192:503–511.
- Calvignac S, Konecny L, Malard F, Douady CJ. 2011. Preventing the pollution of mitochondrial datasets with nuclear mitochondrial paralogs (numts). *Mitochondrion* 11:246–254.
- Cannino G, Di Liegro CM, Rinaldi AM. 2007. Nuclear–mitochondrial interaction. *Mitochondrion* 7:359–366.
- Cannon B, Nedergaard J. 2004. Brown adipose tissue: function and physiological significance. *Physiol Rev*. 84:277–359.
- Chen J, Sun Y. 2011. Variation in the analysis of positively selected sites using nonsynonymous/synonymous rate ratios: an example using influenza virus. *PLoS One* 6:e19996.
- Cheng YT, Liu J, Yang LQ, Sun C, Kong QP. 2013. Mitochondrial DNA content contributes to climate adaptation using Chinese populations as a model. *PLoS One* 8:e79536.
- Chevreaux B, Wetter T, Suhai S. 1999. Genome sequence assembly using trace signals and additional sequence information. *Proc German Conf Bioinform (GCB)*. 99:45–56.
- Christian BE, Spremulli LL. 2012. Mechanism of protein biosynthesis in mammalian mitochondria. *Biochim Biophys Acta*. 1819:1035–1054.
- Conant GC, Wagner GP, Stadler PF. 2007. Modeling amino acid substitution patterns in orthologous and paralogous genes. *Mol Phylogenet Evol*. 42:298–307.
- Crandall KA, Kelsey CR, Imamichi H, Lane HC, Salzman NP. 1999. Parallel evolution of drug resistance in HIV: failure of nonsynonymous/synonymous substitution rate ratio to detect selection. *Mol Biol Evol*. 16: 372–382.
- da Fonseca RR, Johnson WE, O'Brien SJ, Ramos MJ, Antunes A. 2008. The adaptive evolution of the mammalian mitochondrial genome. *BMC Genomics* 9:119.
- Dowling DK, Friberg U, Lindell J. 2008. Evolutionary implications of non-neutral mitochondrial genetic variation. *Trends Ecol Evol*. 23:546–554.
- Drummond AJ, Ho SY, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol*. 4:e88.
- Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol*. 29:1969–1973.
- Fontanillas P, Depraz A, Giorgi MS, Perrin N. 2005. Nonshivering thermogenesis capacity associated to mitochondrial DNA haplotypes and gender in the greater white-toothed shrew, *Crocidura russula*. *Mol Ecol*. 14:661–670.
- Footo AD, et al. 2011. Positive selection on the killer whale mitogenome. *Biol Lett*. 7:116–118.
- Galtier N, Nabholz B, Glemin S, Hurst GD. 2009. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Mol Ecol*. 18: 4541–4550.
- Garvin MR, Bielawski JP, Gharrett AJ. 2011. Positive Darwinian selection in the piston that powers proton pumps in complex I of the mitochondria of Pacific salmon. *PLoS One* 6:e24127.
- Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL. 2008. The Vienna RNA websuite. *Nucleic Acids Res*. 36:W70–W74.
- Hassanin A, Ropiquet A, Couloux A, Cruaud C. 2009. Evolution of the mitochondrial genome in mammals living at high altitude: new insights from a study of the tribe Caprini (Bovidae, Antilopinae). *J Mol Evol*. 68:293–310.
- Hofman S, et al. 2007. Phylogeography of the fire-bellied toads *Bombina*: independent Pleistocene histories inferred from mitochondrial genomes. *Mol Ecol*. 16:2301–2316.
- Irwin DE, Rubtsov AS, Panov EN. 2009. Mitochondrial introgression and replacement between yellowhammers (*Emberiza citrinella*) and pine buntings (*Emberiza leucocephalos*) (Aves: Passeriformes). *Biol J Linn Soc*. 98:422–438.
- Kelley LA, Sternberg MJE. 2009. Protein structure prediction on the web: a case study using the Phyre server. *Nat Protoc*. 4:363–371.
- Laslett D, Canback B. 2008. ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics* 24: 172–175.
- Luo YJ, et al. 2008. Mitochondrial genome analysis of *Ochotona curzoniae* and implication of cytochrome c oxidase in hypoxic adaptation. *Mitochondrion* 8:352–357.
- Matthee CA, van Vuuren BJ, Bell D, Robinson TJ. 2004. A molecular supermatrix of the rabbits and hares (Leporidae) allows for the identification of five intercontinental exchanges during the Miocene. *Syst Biol*. 53: 433–447.
- McClellan DA, Ellison DD. 2010. Assessing and improving the accuracy of detecting protein adaptation with the TreeSAAP analytical software. *Int J Bioinform Res Appl*. 6:120–133.
- McClellan DA, et al. 2005. Physicochemical evolution and molecular adaptation of the cetacean and artiodactyl cytochrome b proteins. *Mol Biol Evol*. 22:437–455.
- Mee JA, Taylor EB. 2012. The cybrid invasion: widespread postglacial dispersal by *Phoxinus* (Pisces: Cyprinidae) cytoplasmic hybrids. *Can J Zool*. 90:577–584.
- Meiklejohn CD, Montooth KL, Rand DM. 2007. Positive and negative selection on the mitochondrial genome. *Trends Genet*. 23: 259–263.
- Melo-Ferreira J, Alves PC, Rocha J, Ferrand N, Boursot P. 2011. Interspecific X-chromosome and Mitochondrial DNA introgression in the Iberian hare: selection or allele surfing? *Evolution* 65:1956–1968.
- Melo-Ferreira J, Boursot P, Suchentrunk F, Ferrand N, Alves PC. 2005. Invasion from the cold past: extensive introgression of mountain hare (*Lepus timidus*) mitochondrial DNA into three other hare species in northern Iberia. *Mol Ecol*. 14:2459–2464.
- Melo-Ferreira J, et al. 2007. The rise and fall of the mountain hare (*Lepus timidus*) during Pleistocene glaciations: expansion and retreat with hybridization in the Iberian Peninsula. *Mol Ecol*. 16:605–618.
- Melo-Ferreira J, et al. 2012. Recurrent introgression of mitochondrial DNA among hares (*Lepus* spp.) revealed by species-tree inference and coalescent simulations. *Syst Biol*. 61:367–381.
- Milne I, et al. 2010. Tablet—next generation sequence assembly visualization. *Bioinformatics* 26:401–402.
- Murrell B, et al. 2012. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet*. 8:e1002764.
- Nevado B, et al. 2009. Complete mitochondrial DNA replacement in a Lake Tanganyika cichlid fish. *Mol Ecol*. 18:4240–4255.

- Nielsen R, Yang Z. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148:929–936.
- Pereira F, et al. 2008. Evidence for variable selective pressures at a large secondary structure of the human mitochondrial DNA control region. *Mol Biol Evol.* 25:2759–2770.
- Pond SLK, Frost SDW. 2005a. A genetic algorithm approach to detecting lineage-specific variation in selection pressure. *Mol Biol Evol.* 22:478–485.
- Pond SLK, Frost SDW. 2005b. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21:2531–2533.
- Pond SLK, Frost SDW. 2005c. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol.* 22:1208–1222.
- Posada D. 2008. jModelTest: phylogenetic model averaging. *Mol Biol Evol.* 25:1253–1256.
- Rambaut A, Drummond AJ. 2009. Tracer v1.5. Available from: <http://beast.bio.ed.ac.uk/software/tracer>.
- Rand DM, Haney RA, Fry AJ. 2004. Cytonuclear coevolution: the genomics of cooperation. *Trends Ecol Evol.* 19:645–653.
- Reid N, Demboski JR, Sullivan J. 2012. Phylogeny estimation of the radiation of western North American chipmunks (*Tamias*) in the face of introgression using reproductive protein genes. *Syst Biol.* 61:44–62.
- Roca AL, Georgiadis N, O'Brien SJ. 2005. Cytonuclear genomic dissociation in African elephant species. *Nat Genet.* 37:96–100.
- Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V, Wallace DC. 2004. Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science* 303:223–226.
- Rutledge LY, Patterson BR, White BN. 2010. Analysis of *Canis* mitochondrial DNA demonstrates high concordance between the control region and ATPase genes. *BMC Evol Biol.* 10:215.
- Saraste M. 1999. Oxidative phosphorylation at the fin de siècle. *Science* 283:1488–1493.
- Scott GR, et al. 2011. Molecular evolution of cytochrome c oxidase underlies high-altitude adaptation in the bar-headed goose. *Mol Biol Evol.* 28:351–363.
- Sequeira F, Alexandrino J, Rocha S, Arntzen JW, Ferrand N. 2005. Genetic exchange across a hybrid zone within the Iberian endemic golden-striped salamander, *Chioglossa lusitanica*. *Mol Ecol.* 14:245–254.
- Shen YY, et al. 2010. Adaptive evolution of energy metabolism genes and the origin of flight in bats. *Proc Natl Acad Sci U S A.* 107:8666–8671.
- 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.
- Suissa S, et al. 2009. Ancient mtDNA genetic variants modulate mtDNA transcription and replication. *PLoS Genet.* 5:e1000474.
- Sun Y-B, Shen Y-Y, Irwin DM, Zhang Y-P. 2011. Evaluating the roles of energetic functional constraints on teleost mitochondrial-encoded protein evolution. *Mol Biol Evol.* 28:39–44.
- Taylor RW, Turnbull DM. 2005. Mitochondrial DNA mutations in human disease. *Nat Rev Genet.* 6:389–402.
- 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 Res.* 22:4673–4680.
- Thulin CG, Fang M, Averianov AO. 2006. Introgression from *Lepus europaeus* to *L. timidus* in Russia revealed by mitochondrial single nucleotide polymorphisms and nuclear microsatellites. *Hereditas* 143:68–76.
- Timmis JN, Ayliffe MA, Huang CY, Martin W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat Rev Genet.* 5:123–135.
- Toews DPL, Brelsford A. 2012. The biogeography of mitochondrial and nuclear discordance in animals. *Mol Ecol.* 21:3907–3930.
- Tomasco IH, Lessa EP. 2011. The evolution of mitochondrial genomes in subterranean caviomorph rodents: adaptation against a background of purifying selection. *Mol Phylogenet Evol.* 61:64–70.
- Tuppen HA, Blakely EL, Turnbull DM, Taylor RW. 2010. Mitochondrial DNA mutations and human disease. *Biochim Biophys Acta.* 1797:113–128.
- Wallace DC. 2007. Why do we still have a maternally inherited mitochondrial DNA? Insights from evolutionary medicine. *Annu Rev Biochem.* 76:781–821.
- Willett CS, Burton RS. 2004. Evolution of interacting proteins in the mitochondrial electron transport system in a marine copepod. *Mol Biol Evol.* 21:443–453.
- Woolley S, Johnson J, Smith MJ, Crandall KA, McClellan DA. 2003. TreeSAAP: selection on amino acid properties using phylogenetic trees. *Bioinformatics* 19:671–672.
- Xu F, Addis JB, Cameron JM, Robinson BH. 2012. LRPPRC mutation suppresses cytochrome oxidase activity by altering mitochondrial RNA transcript stability in a mouse model. *Biochem J.* 441:275–283.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586–1591.
- Yang Z, dos Reis M. 2011. Statistical properties of the branch-site test of positive selection. *Mol Biol Evol.* 28:1217–1228.
- Yang Z, Kumar S, Nei M. 1995. A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics* 141:1641–1650.
- Yang Z, Nielsen R. 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *J Mol Evol.* 46:409–418.
- Yang Z, Nielsen R, Goldman N, Pedersen AM. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155:431–449.
- Yang Z, Wong WS, Nielsen R. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol Biol Evol.* 22:1107–1118.
- Yang ZH. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol.* 15:568–573.
- 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.
- Zsurka G, et al. 2010. Distinct patterns of mitochondrial genome diversity in bonobos (*Pan paniscus*) and humans. *BMC Evol Biol.* 10:270.

Associate editor: Josefa Gonzalez