Gene 628 (2017) 63–71

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

Ancient mitochondrial pseudogenes reveal hybridization between distant lineages in the evolution of the *Rupicapra* genus

T. Pérez[a], F. Rodríguez[b], M. Fernández[a], J. Albornoz[a], A. Domínguez[a,*]

[a] *Departamento de Biología Funcional, Área de Genética, Universidad de Oviedo, 33071 Oviedo, Spain*
[b] *Marine Biological Laboratory, Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Woods Hole, MA, United States*

ARTICLE INFO

Keywords:
Reticulate evolution
Male-dispersal
Chamois
Numts
mtDNA
COI
cytb

ABSTRACT

Mitochondrial pseudogenes (numts) inserted in the nuclear genome are frequently found in population studies. Its presence is commonly connected with problems and errors when they are confounded with true mitochondrial sequences. In the opposite side, numts can provide valuable phylogenetic information when they are copies of ancient mitochondrial lineages. We show that *Rupicapra* individuals of different geographic origin from the Cantabrian Mountains to the Apennines and the Caucasus share a nuclear *COI* fragment. The numt copies are monophyletic, and their pattern of differentiation shows two outstanding features: a long evolution as differentiated true mitochondrial lineage, and a recent integration and spread through the chamois populations. The *COI* pseudogene is much older than the present day mitochondrial clades of *Rupicapra* and occupies a basal position within the *Rupicapra-Ammotragus-Arabitragus* node. Joint analysis of this numt and a *cytb* pseudogene with a similar pattern of evolution places the source mitochondrial lineage as a sister branch that separated from the *Ammotragus-Arabitragus* lineage 6 million years ago (Mya). The occurrence of this sequence in the nucleus of chamois suggests hybridization between highly divergent lineages. The integration event seems to be very recent, more recent than the split of the present day mtDNA lineages of *Rupicapra* (1.9 Mya). This observation invites to think of the spread across the genus by horizontal transfer through recent male-biased dispersal.

1. Introduction

Mitochondrial pseudogenes integrated in the nuclear genomes (numts) are frequently detected as unwanted paralogous in PCR-based mitochondrial studies (reviewed in (Bensasson et al., 2001; Ermakov et al., 2015). In addition, numts were detected in scans of most sequenced eukaryotic genomes (Richly and Leister, 2004; Hazkani-Covo et al., 2010; Verscheure et al., 2015). Their abundance varies in different eukaryotic taxa and can derive from all types of mitochondrial sequences, vary in sizes, and bear different degrees of similarity to their mitochondrial counterparts (Hazkani-Covo, 2009; Shi et al., 2016). When transferred to the nucleus, mitochondrial sequences become nonfunctional pseudogenes (Li et al., 1981; López et al., 1997). The mutation rate in the nucleus is lower than that of the mitochondria and hence, when numts are the result of ancient translocations to the nucleus, they give information about ancestral sequences of mitochondrial genes. In that case, the study of numts provides information about their origin and the evolutionary histories and phylogenetic relationships of taxa, given that they represent a fossil lineage that evolved in parallel to their mitochondrial counterparts (Hazkani-Covo, 2009; Baldo et al., 2011; Miraldo et al., 2012; Ko et al., 2015).

Here we analyse a numt of the gene *COI* to investigate the ancestral history of chamois (genus *Rupicapra*). Chamois is a mount caprine found in most of the mountain ranges of southern Eurasia from the Cantabrian Mountains to the Caucasus. The most accepted classification considers two species, *R. pyrenaica* and *R. rupicapra*, (Grubb, 1993; Corlatti et al., 2011): *Rupicapra pyrenaica* (with the subspecies *parva*, *pyrenaica* and *ornata*) from southwestern Europe, and *R. rupicapra* (with the subspecies *cartusiana*, *rupicapra*, *tatrica*, *carpatica*, *balcanica*, *asiatica* and *caucasica*) from northeastern Europe. The mitochondrial phylogeny

Abbreviations: A, adenosine; AIC, Akaike information criterion; bp, base pair; BPP, Bayesian posterior probability; *COI*, cytochrome oxidase I; *cytb*, cytochrome *b*; dCTP, deoxycytidine triphosphate; dNTP, deoxyribonucleoside triphosphate; G, guanosine; HKY, Hasegawa-Kishino-Yano; HMW, high molecular weight; HPD, highest posterior density; I, isoleucine; JC, Jukes-Cantor; Ka, number of non-synonymous substitutions per non-synonymous site; kb, kilobase; Ks, number of synonymous substitutions per synonymous site; kya, thousand years ago; MCMC, Markov chain Monte Carlo; mt, mitochondrial; Mya, million years ago; Myr, million years; NJ, neighbor joining; nt, nucleotide; nuCOI, COI nuclear pseudogene; numt, nuclear mitochondrial pseudogene; P, p-value; SDS, sodium dodecyl sulfate; SSC, saline sodium citrate buffer; TBE, tris-borate-EDTA buffer; u, unit; UPGMA, Unweighted Pair Group Method with Arithmetic Mean; UV, ultraviolet; V, valine

[*] Corresponding author.
E-mail address: sanjurjo@uniovi.es (A. Domínguez).

http://dx.doi.org/10.1016/j.gene.2017.07.035
Received 24 February 2017; Received in revised form 7 July 2017; Accepted 11 July 2017
Available online 12 July 2017

showed three main lineages (Crestanello et al., 2009; Rodríguez et al., 2009; Rodríguez et al., 2010) that were referred to as W, C and E after its restricted geographic distribution in either west, central or east Eurasia (see Rodríguez et al., 2009). The populations in the Iberian Peninsula (*R. pyrenaica parva* and *R. pyrenaica pyrenaica*) and several individuals in the west Alps (*R. rupicapra rupicapra*) grouped into clade W, the population in the Apennines (*R. pyrenaica ornata*) and the small population in the Massif of Chartreuse (*R. rupicapra cartusiana*) were of mitochondrial type C and most of the alpine chamois (*R. rupicapra rupicapra*) and the populations to the East of the Alps (*R. rupicapra tatrica*, *R. rupicapra carpatica*, *R. rupicapra balcanica*, *R. rupicapra asiatica* and *R. rupicapra caucasica*) were of clade E. The divergence between the three main mtDNA clades has been estimated around 1.9 Mya (Lalueza-Fox et al., 2005; Rodríguez et al., 2010; Pérez et al., 2014), at the Early Pleistocene following the recent "Formal Subdivisions of the Pleistocene Series/Epoch" of the Subcommission on Quaternary Stratigraphy (Cohen et al., 2013) that we will adopt along the paper. This is by far older than the age of the most ancient *Rupicapra* fossils in Europe that were discovered in the Balkans and correspond to the beginning of the middle Pleistocene, between 780 and 750 thousand years ago (kya) (Fernández and Cregut, 2007). The analysis of different nuclear markers, contrary to the data obtained from mtDNA, indicated very low divergence among populations even when they belonged to different putative species (Pérez et al., 2002; Rodríguez et al., 2010; Pérez et al., 2011; Pérez et al., 2013; Pérez et al., 2017). A previous analysis of a nuclear copy of *cytb* added more complexity to the history. The nuclear pseudogene originated from a lineage even older than the already old lineages found in the mitochondria (Rodríguez et al., 2007). Here we study a numt of the *COI* gene and check its presence in different populations across Europe. In addition we study the phylogeny of the *COI* numt and the previously identified *cytb* numt within the mitochondrial phylogeny of caprini (Hassanin et al., 2012) in order to investigate the timing of transposition of the numts and the ancient chamois relatives that were involved in the process. Our data show how unexpected findings change the perception of an evolutionary history.

2. Materials and methods

2.1. Samples and DNA extraction

A total of seventy samples were included in the study (see Additional File 1). Twenty eight samples were of the species *R. pyrenaica* (11 from the Cantabrian Mountains, 15 from the Pyrenees and 2 from the Apennines), and 42 of *R. rupicapra* (2 from the Massif of Chartreuse, 24 from the Alps, 1 from the Tatra Mountains, 6 from the Carpathians, 6 from the Balkans, 1 from the Pontic Mountains and 2 from the Caucasus).

Most of the samples (62 out of the 70) have been included in previous analysis. The DNA for amplification was obtained by different methods as detailed in our previous studies (Pérez et al., 2002; Rodríguez et al., 2010). DNA from soft tissue was extracted either with the phenol/chloroform method (Sambrook et al., 1989), using Chelex following Estoup et al. (1996) or using the 'DNeasy Tissue kit' (Qiagen, Hilden, Germany). Genomic DNA from bone or teeth was extracted from 1 g powdered material following Cattaneo et al. (1995) and further purified with Chelex as described (Pérez et al., 2002).

2.2. Amplification and sequencing of PCR products

We amplified a fragment of 382 bp of the *COI* gene using a primer pair designed from the mitochondrial genome of *Ovis aries* (GenBank Acc. N°. NC_001941) with the aid of the software Amplify (Engels, 2005). The sequences of the primers are as follow: 5′TTCAACCAACCACAAAGATATCGG3′ (CO1F forward primer) and 5′TGCCTGCTAGAGGAGGGTATACG3′ (CO1R reverse primer).

PCR amplifications were performed in 20-μl volume with 0.5 u of Biotools DNA polymerase, 0.5 μM of each primer, 200 μM of each dNTP, 1 × PCR Buffer, and 2.5 mM $MgCl_2$. Cycling parameters were as follow: 1 cycle of 94 °C for 3 min, followed by 35 cycles, each of 94 °C for 15 s; 64 °C for 30 s; and 72 °C for 30 s; followed by 72 °C for 10 min. PCR products were electrophoresed along with size standards in 2% agarose gels in 1 × TBE, stained with ethidium bromide (0.5 μg/ml) and visualized under UV light. The PCR-amplified products were purified with the illustra™ ExoStar™ 1-Step (GE Healthcare). Both strands of PCR products were sequenced with PCR primers and the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). Sequencing products were purified with isopropanol precipitation and sequenced in an ABI 310 Genetic Analyzer (Applied Biosystems).

2.3. Identification of the COI numt

The nucleotide sequences were checked, edited and aligned manually with MEGA vs 6.06 (Kumar et al., 2012). The first inspection of sequences evidenced two unexpected facts: 1) Several sequences had a stop codon due to a substitution at nucleotide 126 of the *COI* gene and the amino acid number 42 of the protein sequence; 2) For several individuals, we obtained chromatograms with the same double peaks repeatedly (the double peaks were not due to contamination). After these observations, we hypothesized that we were dealing with a nuclear pseudogene and performed several analysis in order to confirm this possibility. First, the presence of a nuclear copy of *COI* was shown by Southern blot of genomic DNA, as described below. Further, the individuals that produced double peaked sequences were reanalyzed by cloning the amplified products in order to obtain single sequences, either mitochondrial or nuclear. Amplification products of the expected size were purified using GFX PCR DNA and Gel Band Purification Kit (Amersham Biosciences), and they were directly cloned into the pMOSBlue vector (Amersham Biosciences) and transformed into MOSBlue competent cells according to the supplier's specification. Clones were screened for inserts of the expected size by PCR amplification with the universal primers M13 and T7. Several clones from each sample were sequenced attempting to obtain both the mitochondrial and the nuclear sequences. For sequencing, plasmid DNA was prepared for selected clones (Sambrook et al., 1989) and sequences were determined for both strands using the T7 and M13 universal primers. The BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) was used for sequencing, and the reactions were run on an ABI 310 sequencer.

We obtained both sequences, gene and pseudogene, from a single individual for the subspecies *ornata*, *rupicapra*, *carpatica*, *balcanica* and *caucasica*. In other subspecies where the nuclear copy did not arise fortuitously, we checked for its presence using high molecular weight (HMW) DNA as template for the amplification (only possible in the cases in which we had enough quality and quantity of starting DNA). Genomic DNA was electrophoresed in 0.75% agarose gels and the HMW DNA was extracted from the gel using the GFX PCR DNA and Gel Band Purification Kit. We then amplified the *COI* fragment from this fraction of DNA and proceeded with the sequencing. This way, the non-functional copies corresponding to the nuclear pseudogene were obtained from HMW DNA of specimens of *parva*, *pyrenaica* and *asiatica*. Eventually, we did obtain both sequences, mitochondrial (*COI*) and nuclear (nu*COI*) from 11 individuals covering eight of the ten subspecies.

2.4. Southern blotting

We checked for the presence of nuclear copies of the *COI* gene in five *Rupicapra* specimens from the subspecies *parva*, *pyrenaica*, *ornata*, *carpatica* and *asiatica*. We included the outgroups *Ammotragus lervia* and *Capra hircus* in the analysis. Total genomic DNA was digested with *Pst*I (New England Biolabs®Inc.), a restriction enzyme that does not cut the mtDNA of *Rupicapra*, so that we expected a Southern band of 16.4 kb when hybridized with a *COI* probe. Approximately 5 μg of genomic

DNA were digested with a fivefold excess of enzyme in a total volume of 50 μl, using standard buffers provided by the manufacturer. Digested DNA was separated by electrophoresis in 20-cm, 0.7% agarose gels and TBE buffer (Sambrook et al., 1989). Gels were run at 65 V for 13 h. Gels were depurinated in 0.2 N HCl for 20 min. DNA was transferred to a charged nylon membrane (Sigma) by capillary blotting under alkaline conditions, following Sambrook et al. (1989). DNA was fixed to the membrane by exposing it to UV light. Membranes were prehybridized during 4 h at 65 °C in 6 × SSC, 0.5% SDS, 5 × Denhart containing 100 μg/ml of denatured herring sperm DNA. About 25 ng of probe DNA were 32P-labeled by random priming using 4 μl of High Prime (Roche diagnostics) and 3 μl DCTP in a 20 μl reaction and hybridized overnight at 65 °C in the same solution. Membranes were washed twice at 65 °C for 30 min in 2 × SSC, 0.1% SDS, once in 0.1 × SSC, 0.1% SDS at 65 °C, briefly washed in 0.1 × SSC at room temperature and autoradiographed for 1–4 days at 70 °C with one intensifying screen.

2.5. Phylogenetic analysis

The final dataset consisted of 81 sequences, including gene and pseudogene, with a length of 271 nucleotides corresponding to positions 110 to 380 of the *COI* gene. At first, we chose to use simple models of nucleotide substitution for analyzing phylogenies because differences in genetic estimates of distances is low when one is studying closely related sequences. In addition, statistical prediction based on a model with many parameters is subject to more errors than a simple one (Nei and Kumar, 2000). We constructed a neighbor joining (NJ) tree with MEGA, using the simple Jukes-Cantor (JC) model of nucleotide substitution. The reliability of the nodes was assessed by 1000 bootstrap replicates (Felsenstein, 1985). The sequences grouped into 8 haplotypes of the *COI* gene and 3 haplotypes of the *COI* pseudogene. GenBank Accession Numbers and list of samples bearing each haplotype are given in Additional File 2.

To evaluate the differentiation between the gene and the pseudogene and compare rates of substitution, we limited the dataset to the eleven individuals from which both sequences, *COI* and nu*COI*, were obtained. We constructed a UPGMA tree based in JC distance because it is a simple linear tree that allows to readily check distances between nodes and from nodes to tips. The sequence of *Capra ibex* was used as outgroup. Estimates of nucleotide diversity and divergence within and between groups of sequences corresponding to the gene and the pseudogene were obtained with MEGA under JC. Errors of the estimates were obtained by a bootstrap procedure using 1000 replicates. Synonymous and nonsynonymous distances, both within and between the two groups of sequences, were estimated using the Kumar model (Nei and Kumar, 2000). The rates of substitution in the mitochondrial and nuclear lineages were compared both using the Maximum Likelihood ratio test and Tajima's relative rate test, both implemented in MEGA.

To place the *COI* nuclear copy of *Rupicapra* in the mtDNA phylogenetic tree of caprines, it was compared with the corresponding sequences of species of the tribe (Hassanin et al., 2012). Their GenBank Accession Numbers are listed in Additional File 3. We performed a Bayesian analysis using the Monte Carlo Markov chains (MCMC) method implemented in BEAST 2 (Drummond and Rambaut, 2007) and defining the topology previously obtained from complete mitochondrial genomes (Hassanin et al., 2012; Pérez et al., 2014), as prior. We used a relaxed log Normal clock and a calibrated Yule speciation process as priors. The model of nucleotide substitution was the HKY + G with the empirical base frequencies, as determined by the AIC criteria in MEGA 6. The analysis was run for 1 million generations with tree and parameter sampling every 100 generations and a burn-in of 10%. Two independent replicates were run and their sampling distributions were checked for convergence using the software TRACER 1.6 (Rambaut and Drummond, 2009). The sampled trees of the two replicates were combined using the software LogCombiner (within BEAST) and the Maximum clade credibility criterion tree was obtained with TreeAnnotator, using a burn-in of 10% trees and with mean node heights.

We compared the phylogeny obtained for the *COI* numt with the one obtained for a pseudogene of *cytb* that was sequenced in a previous study (Rodríguez et al., 2007). At first, the partial sequences of *cytb* (403 nt) were studied with BEAST as before. Finally, the sequences of the two genes were analysed together in BEAST including the two partitions in the same analysis under linked site and clock models and trees. The parameters of the run were the same as previously described except in that three nodes were calibrated to obtain divergence times for the lineage origin of pseudogenes. Calibrations included soft bounds to account for uncertainty (Ho and Phillips, 2009). Two calibration points were based on the fossil record using the ages and prior probability distributions given in Bibi (2013). These points were crown Bovidae with a normal prior (mean = 18 Ma, standard deviation = 1 Ma), based on *Eotragus noyei*; and crown Caprini with a normal prior (mean = 8.9 Ma, standard deviation = 2 Ma) based on *Aragoral mudejar* (see Additional File 1 in Bibi, 2013). The third calibration point was the node that links the branch of *Rupicapra* with *Ammotragus-Arabitragus* (mean = 7.90, 95% HSP 6.67–9.23). This calibration is based on the previously obtained molecular dating from the complete mitochondrial genome (Pérez et al., 2014). Two independent replicates were run as before, convergence was checked with TRACER, the trees were combined with LogCombiner and the Maximum clade credibility tree was obtained with TreeAnnotator. All trees obtained by the different methods were visualized with FigTree 1.4 (Rambaut, 2006).

3. Results

3.1. NuCOI detection and nucleotide substitution pattern

Our final dataset consisted of 81 sequences, 61 corresponding to the *COI* gene and 20 corresponding to the nuclear pseudogene, nu*COI*. From 11 out of the 70 specimens studied, we obtained the two sequences corresponding to the true gene and the pseudogene. All the 20 copies of the pseudogene presented a stop codon at the nucleotide 126 and hence represent non-functional copies of the gene. The phylogenetic tree of the 81 sequences unambiguously supported two monophyletic clades, one consisting of the nuclear copies and the other consisting of the mitochondrial ones (Fig. 1).

The existence of nuclear copies of *COI* was checked by Southern analysis. Total genomic DNA digested with *Pst*I was electrophoresed and hybridized to a radiolabeled *COI* probe (Fig. 2). Every *Rupicapra* sample shows a band of about 16 kb, corresponding to the mitochondrial gene, and another one of > 50 kb that should correspond to nuclear DNA. The samples of *Ammotragus* and *Capra* do not show the high molecular weight band. *Ammotragus* only presents the mtDNA band at 16 kb and *Capra* has a band of 11.9 kb that corresponds to the fragment containing the *COI* gene obtained after the digestion of its mtDNA with *Pst*I.

Analysis of sequences revealed 11 haplotypes, 8 among the 61 mitochondrial sequences (2 in *R. pyrenaica* and 7 in *R. rupicapra*, 1 shared by individuals of the subspecies *R. pyrenaica parva*, *R. pyrenaica pyrenaica* and *R. rupicapra rupicapra*) and 3 among the 20 sequences corresponding to the nuclear copy, one common to the 16 individuals of *R. rupicapra* (subspecies *rupicapra*, *carpatica*, *balcanica*, *asiatica* and *caucasica*) and one of the specimens of *R. pyrenaica* (subspecies *parva*), and the other two found in individuals of *R. pyrenaica*, one in the subspecies *pyrenaica* and the other in the subspecies *ornata*. Overall, a total of 23 substitutions were observed in 23 sites (Table 1). The average percent divergence between nuclear and mitochondrial copies of the gene was 6.79 ± 1.56 substitutions. A total of 10 mutations were observed among mitochondrial sequences, with an average of 1.19 ± 0.39 substitutions between pairs of individuals. The difference between percent synonymous and non-synonymous mutations was 2.83 ± 1.12. Only 2 mutations were observed between the nu*COI*

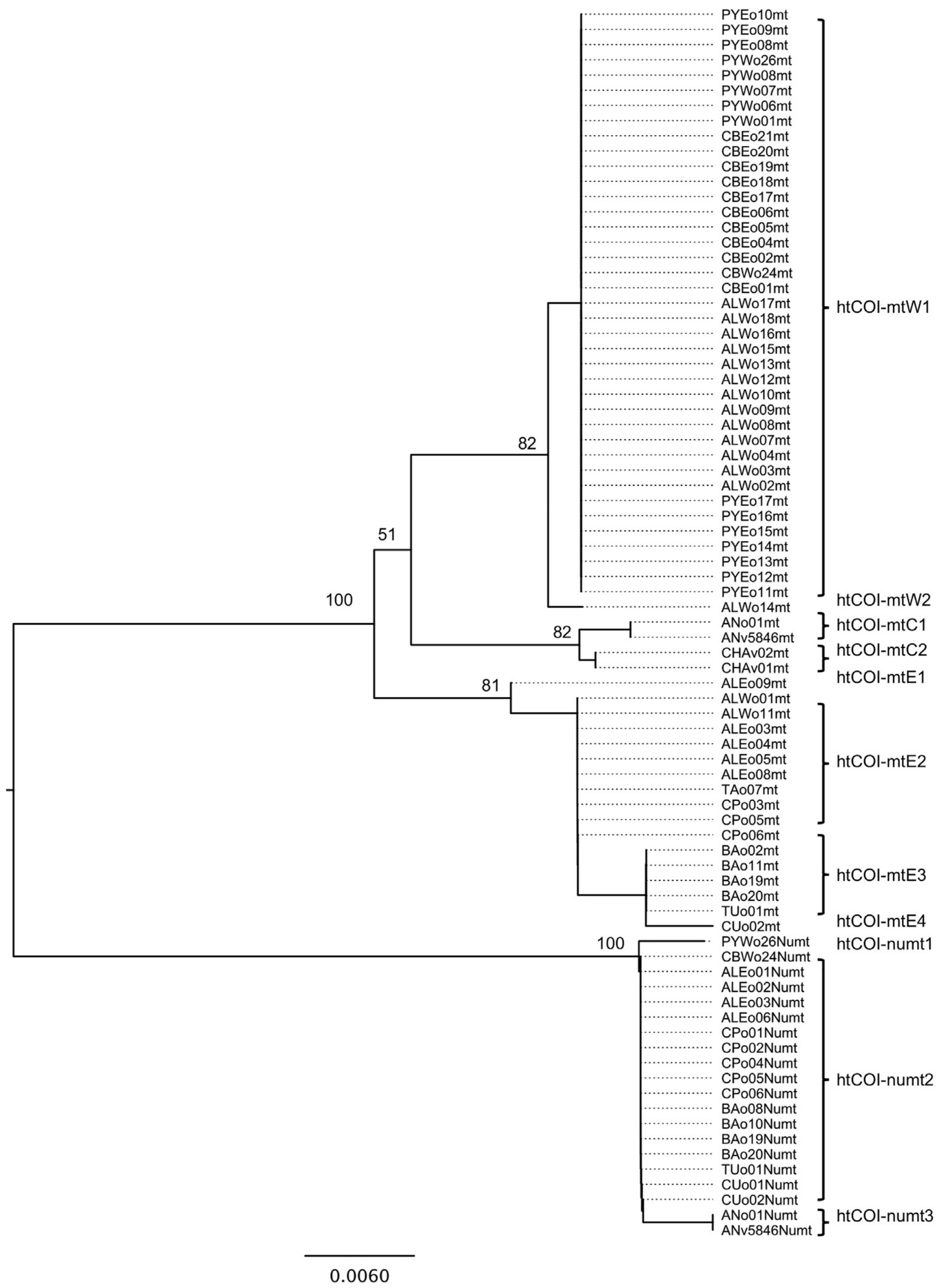

Fig. 1. Neighbor-joining tree based on Jukes-Cantor distance of 81 COI sequences. The two deep clades correspond to the gene and pseudogene. Numbers at nodes indicate percent bootstrap support.

sequences with an average of 0.29 ± 0.10% substitutions between pairs of individuals. Among the 12 mutations that differentiate the copies of the mitochondrial gene and the pseudogene, 10 were synonymous, leading to a difference of 19.82 ± 6.47 between the percent synonymous and non-synonymous substitutions. The non-synonymous include a transition G > A that produced a stop codon at position 76 of the protein. The other non-synonymous substitution was a transition A > G that causes an I > V amino acid substitution at position 57 of the protein. This substitution is present in other caprines including *Ammotragus lervia* and *Arabitragus jayakari*.

Among others, we obtained the sequences of both the gene and the pseudogene (*COI* and nu*COI*) from eleven specimens: four from the species *Rupicapra pyrenaica*, including the subspecies *parva*, *pyrenaica*, and *ornata*, and seven from the species *R. rupicapra*, including the subspecies *rupicapra*, *carpatica*, *balcanica*, *asiatica* and *caucasica*. We constructed a simple UPGMA tree using the sequence of *Capra ibex* as outgroup to check the divergence between the two nominal species, both for the gene and the pseudogene. The hypothesis of equal evolutionary rate was not rejected by the molecular clock test by Maximum Likelihood (P = 1). In the same way, in the relative rate test of Tajima (1993), the mean distance between the outgroup and the pseudogene sequences (9.96 ± 1.93) and between the outgroup and the mitochondrial *COI* sequences (9.88 ± 1.83) did not differ (P ≥ 0.61 in the relative rate test). The UPGMA tree of Fig. 3 shows clearly that the pseudogene originated far before the split of mitochondrial sequences of the genus *Rupicapra*. The group of sequences of the true mitochondrial gene shows a deep subdivision into clades that can't be seen among the sequences of the nuclear pseudogene. The percent diversity within the group of sequences of the gene (1.57 ± 0.48) is higher than for the pseudogene (0.19 ± 0.13).

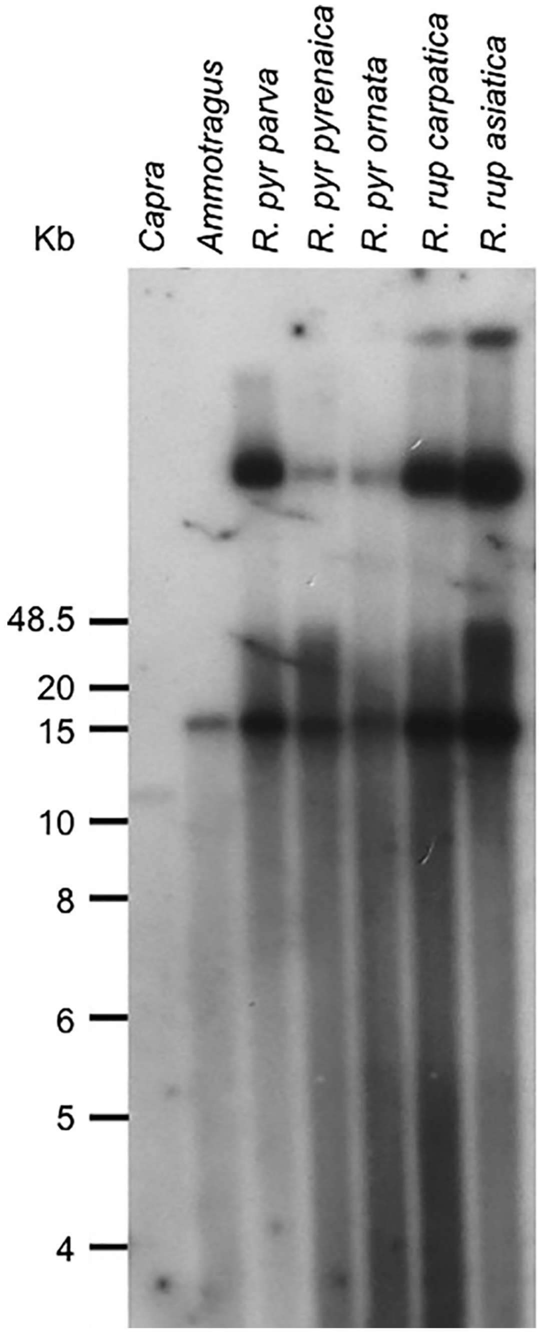

Fig. 2. Southern blot of total genomic DNA digested with *Pst*I and hybridized to a *COI* fragment probe.

3.2. Numt phylogenetic position within the Caprini

The *COI* pseudogene has been shown to be older than the split of *Rupicapra* mtDNA clades and to show a pattern of substitutions proper to a true gene rather than a pseudogene. These features have been previously observed in a nuclear copy of the *cytb* gene (Rodríguez et al., 2007). We studied the position of these two numts within the phylogeny of caprines, at first separately and then in a combined analysis of the two partitions in BEAST.

The *COI* pseudogene is phylogenetically close to the branch leading *Ammotragus* and *Arabitragus* even though the Bayesian posterior probability (BPP) of the grouping is only of the 56% (Fig. 4a). The tree based on *cytb* sequences groups the pseudogene in the same node (*Rupicapra*, *Ammotragus*, *Arabitragus*) but the branch joints to *Rupicapra* with a 57% BPP (Fig. 4b). Finally, the combined analysis joins the branch of the pseudogene to the *Ammotragus-Arabitragus* node with a 69% support. The alternative grouping basal to the branch leading to *Rupicapra* has a BPP of 31%. The age of the split is of 6.36 Mya and the highest posterior density interval of the estimate [95% HPD, 4.90–8.23] overlaps the basal node of the clade *Rupicapra*, *Ammotragus*, *Arabitragus*.

4. Discussion

Mitochondrial pseudogenes are a common component of eukaryotic nuclear genomes. During years their identification was linked to the occurrence of unexpected sequences in phylogenetic studies (López et al., 1994; Arctander, 1995; Pons and Vogler, 2005). Most recently, numts have been detected in genome sequencing projects (Richly and Leister, 2004; Hazkani-Covo et al., 2010) even though next generation sequencing methods, based on the assemblage of small reads, presumably fail to identify many of the numts present in a genome.

Many studies have dealt with numts as artefacts that can confound phylogenetic relationships between taxa (Haran et al., 2015). Others have focused on the possibility of misidentification of species when attempting to use mitochondrial DNA markers for DNA barcoding (Ermakov et al., 2015). These effects would be particularly important when the numts present the evolutionary pattern of true mitochondrial genes, as in the present study, complicating its unmasking. Recommendations to avoid the unwanted inclusion of numts in analyses were given elsewhere (Bensasson et al., 2001; Thalmann et al., 2004). Besides these undesirable effects, numts can be used to identify extinct mitochondrial lineages that provide an insight in the evolution of their carriers (Zischler et al., 1998; Schmitz et al., 2005; Kim et al., 2006; Hazkani-Covo, 2009). In the present study, after obtaining numt sequences as accidental artefacts, we intended to go deeper in their study to obtain information on the evolution of the genus *Rupicapra*.

At first, we must remind the singular evolutionary history of numts. In its origin, a numt arises from a functional mitochondrial gene that transposes to the nucleus, thereafter the evolution of the numt is expected to follow a pattern characterized by a lower evolutionary rate due to the lower mutation rate in the nucleus than in the mitochondrion (Brown et al., 1979), and its evolution is expected to be neutral, given

Table 1
Absolute number of mutations, sequence divergence and estimated number of synonymous (Ks) and non-synonymous (Ka) substitutions per 100 sites in mitochondrial gene *COI* and nuclear pseudogene nu*COI*.

	Total	*COI*	nu*COI*	*COI* vs nu*COI*
N° sequences	81	61	20	
N° haplotypes	11	8	3	
N° mutations	23	10 (1 shared)	2	12
Synonymous	19	9	1	10
Mean divergence	3.22 ± 0.68	1.19 ± 0.39	0.29 ± 0.10	6.79 ± 1.56
Ks	9.48 ± 2.65	3.06 ± 1.19	0.11 ± 0.11	20.64 ± 6.53
Ka	0.45 ± 0.30	0.24 ± 0.24	0.11 ± 0.12	0.82 ± 0.63
Ks-Ka	9.03 ± 2.27	2.83 ± 1.12	−0.005 ± 0.16	19.82 ± 6.47

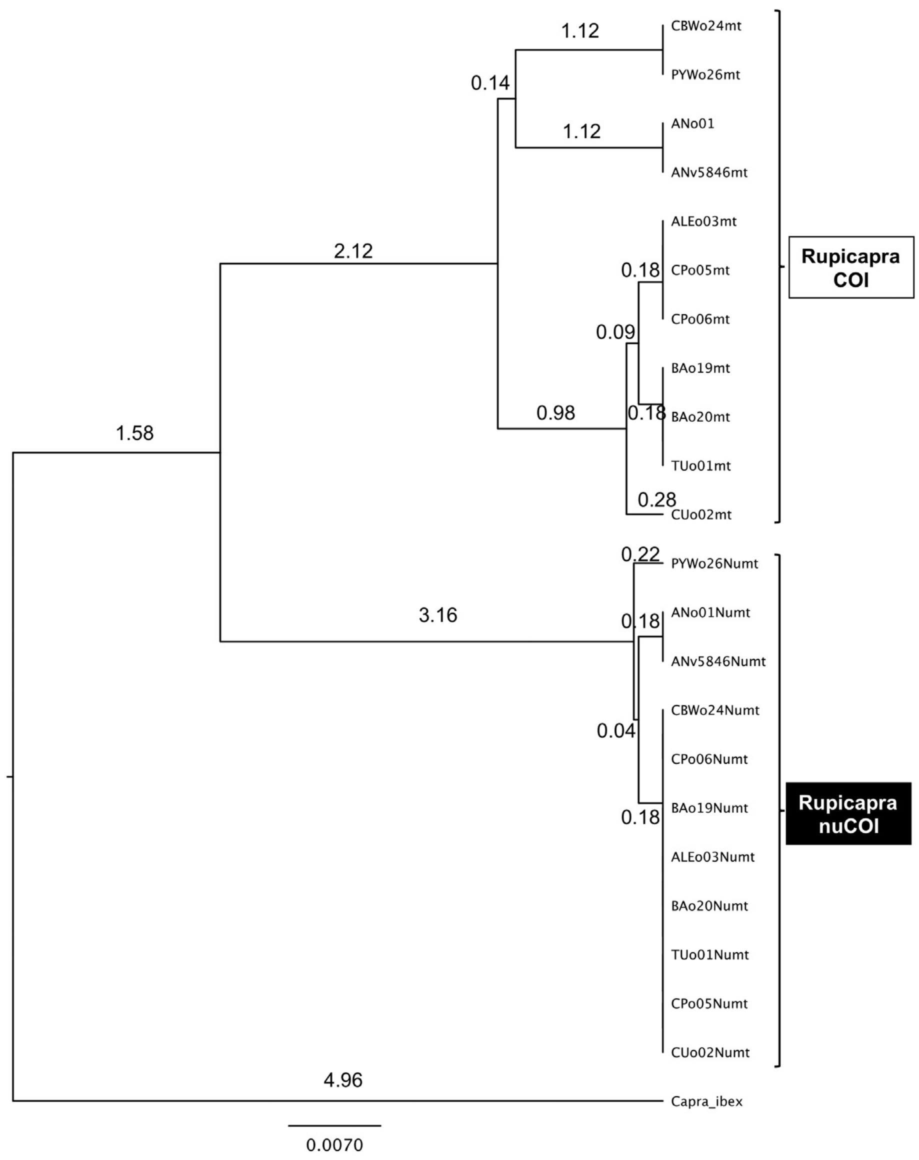

Fig. 3. UPGMA tree of pairs of sequences, gene and pseudogene, obtained from eleven specimens of *Rupicapra*. The tree is based on Jukes-Cantor distances and *Capra ibex* was used as outgroup. Branch lengths are given in percent number of substitutions per nucleotide.

that the copy in the nucleus is non-functional and hence it is not subject to selection (López et al., 1997). The comparison among the true mitochondrial sequences shows a larger number of synonymous than non-synonymous substitutions as expected for a gene subject to purifying selection. There were only two mutations between the sequences of the pseudogene, one synonymous and one non-synonymous, pointing to relaxed selection although the number of mutations is too small to draw any conclusion.

We will consider now the comparison between *COI* mitochondrial sequences versus *nuCOI* sequences. The branches leading to numts in a phylogenetic tree are expected both to be shorter than those of their true mitochondrial correlates and to accumulate non-synonymous mutations. On the contrary, length differences between the branches leading to the *COI* gene and to the nu*COI* pseudogene are non-significant. In addition, ten of the twelve fixed substitutions that differentiate gene and pseudogene are synonymous, as expected if they occurred in a functional mitochondrial sequence. This pattern of evolution is fully coincident with the one that we have described previously for a pseudogene of *cytb* (Rodríguez et al., 2007). Similar facts, long branches and a pattern of substitutions proper of a functional mitochondrial gene, theoretically unexpected for the evolution of numts, were recurrently observed in numt studies (López et al., 1994; Sunnucks and Hales, 1996; DeWoody et al., 1999; Lu et al., 2002; Baldo et al., 2011). Alternative interpretations were given to these observations. Lu et al. (2002) interpreted that the pseudogene originated from a duplication event in the mitochondrial genome of a common ancestor, diverged in the mitochondria as a functional copy, and then inserted into the nuclear genome and pseudogenized. As the authors point out, no duplication in the mitochondria has been described in vertebrates. In addition, it is difficult to explain the loss of the duplicate sequence from the mitochondrion and its simultaneous acquisition by the nucleus. The most likely interpretation is that long numt branches come from source mitochondrial lineages that are now extinct or unsampled (Sunnucks and Hales, 1996) and that are very different from the extant mitochondrial lineage with which they appear now associated. This implies hybridization between lineages and transposition to the nucleus of a mitochondrial sequence highly divergent from the lineage present in the mitochondrion (Rodríguez et al., 2007; Baldo et al., 2011).

The nu*COI* sequences present a mean divergence of 6.8% with the current mitochondrial sequences of *Rupicapra*. This value is remarkably close to the divergence of 7.2% previously reported between a *cytb* pseudogene and the corresponding mitochondrial gene. The patterns of evolution of the two pseudogenes were also similar and both of them were shown to hybridize to the same specific nuclear DNA band of high molecular weight obtained from the digestion of genomic DNA with *Pst*I. It can be presumed that the two sequences come from the same

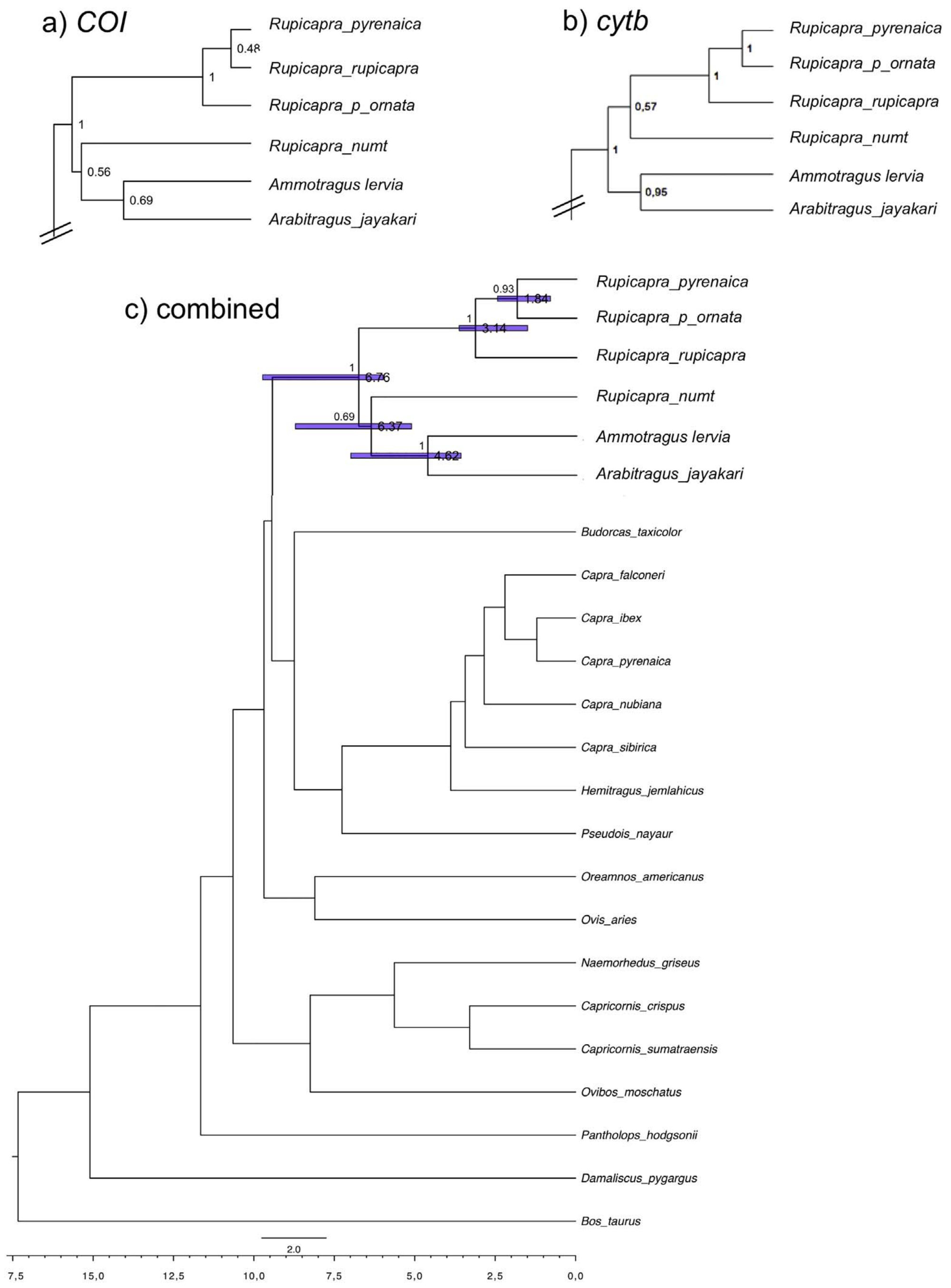

Fig. 4. Bayesian trees relating the sequences of the pseudogenes with the true mitochondrial genes of caprines. a) *COI* tree, b) *cytb* tree, c) Joint analysis of the two partitions. Numbers at the nodes are Bayesian posterior probabilities (BPP). In (c) the mean age estimate for each node is given in million years, with 95% credibility intervals indicated by the blue bars. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

source ancestor. The simultaneous Bayesian phylogenetic analysis of the two pseudogenes supports a species tree in which the numt is grouped to the *Rupicapra-Ammotragus-Arabitragus* clade at a basal position. The split of the putative ancestor of the pseudogenes from the *Ammotragus-Arabitragus* branch was 6.36 Mya [95% HPD: 4.90–8.23 Mya]. Although we were not able to perform an analysis on the presence/absence of numts at the individual level over the different populations/subspecies of *Rupicapra*, we did uncover the presence of nu*COI* in the subspecies *parva*, *pyrenaica*, *ornata*, *rupicapra*, *carpatica*, *balcanica*, *asiatica* and *caucasica*. That means that the same numt is widely spread over the chamois populations and is present in populations belonging to the extant three different mitochondrial clades that did diverge 1.9 Mya (Pérez et al., 2014).

The question of when the mitochondrial sequence leading to the present day pseudogenes has been translocated into the nucleus and how it spread over chamois populations is not easy to solve. The most recent common ancestor of the caprine contributing the numts and the current chamois occurred at the split of *Rupicapra*, *Ammotragus*, *Arabitragus*, 7.5 Mya [95% HPD: 6.24–8.81 Mya]. How the new pseudogene persisted and spread along the population is difficult to explain given that the fate of a single new mutation that occurs in a population must in most cases be its loss (Hedrick, 2010). The persistence of the new insertion in a population and its eventual fixation is a highly improbable event even if it is beneficial (Hedrick, 2010).

We can consider three possibilities for the origin and spread of the pseudogene of the genus *Rupicapra* as depicted in the Fig. 5: a) the translocation of the mtDNA to the nucleus occurred in the ancestral lineage of *Rupicapra* and from this initial event it became fixed in the population and went through the same historical events as the mitochondrial gene given that both copies resided in the same specimens; b) the pseudogene originated after the hybridization with a highly divergent lineage, now extinct or unsampled, that occurred before the

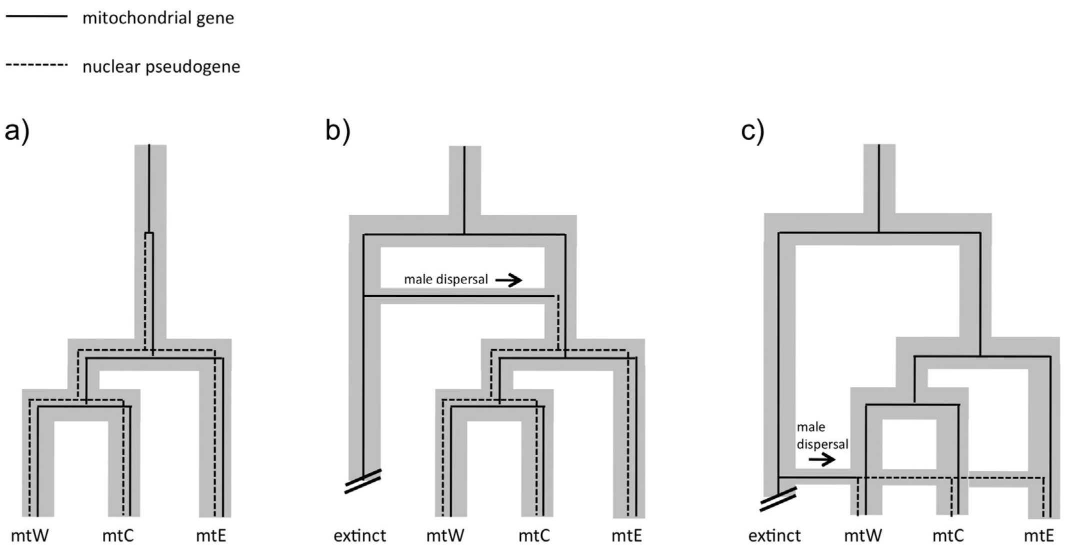

Fig. 5. Alternative hypotheses for the origin or the numt lineage of *Rupicapra*. a) The numt originated in the ancestral lineage of *Rupicapra* and from this initial event it became fixed in the population; b) The numt originated subsequent to hybridization with a distant lineage before the split of the three extant mtDNA lineages; c) The numt originated subsequent to hybridization with a distant lineage after the split of the extant mtDNA lineages and then spread across populations through male-dispersal. Solid lines indicate true mtDNA lineages. Dotted lines are used to represent numt sequences.

divergence of the extant mitochondrial clades; c) as in (b) the pseudogene originated after hybridization but this occurred after the split of the three mitochondrial clades of *Rupicapra* and then spread all over the divergent populations by horizontal transfer through sexual reproduction. In the case (a) the translocation event can be seen as a founder event and the inserted sequence needs to be very different from the sequence that lasted in the mitochondrion, the observed difference seems too high even for a large ancestral *Rupicapra* population. From this ancestral insertion event the population should be supposed to suffer a bottleneck with few founder individuals that could lead to a rapid genetic differentiation (Baker and Moeed, 1987), but this would affect both the nuclear and the mitochondrial sequences. After that, the sequences of the gene and pseudogene would evolve within the same chamois populations and hence the outcome of differentiation will be totally different to the observed as the nuclear sequence should follow a reduced evolutionary rate under neutrality contrasting with the evolution in the mitochondrion. The scenarios depicted in (b) and (c) propose the hypothesis of a more recent insertion of mitochondrial sequences into the nucleus to explain their long evolution as true mitochondrial sequences, their reduced diversity and the lack of a structure that correlates with the three present day mitochondrial lineages. The huge differentiation with the extant mitochondrial sequences implies that hybridization between highly divergent lineages should have occurred.

The generation of numts is thought to occur after the release of mtDNA sequences in the cytoplasm and their posterior translocation and insertion into the nuclear DNA (Blanchard and Schmidt, 1996; Bensasson et al., 2001). In mammals, the transfer of mtDNA to nucleus can occur preferentially when sperm mtDNA is released shortly after penetration of the egg by the sperm (Leister, 2005). The sperm mitochondria are subjected to proteolysis inside the egg cytoplasm. It is interesting to note that the mechanism of elimination of sperm mitochondrial DNA is species specific and that leakage of parental mtDNA occurs in interspecific crosses (Kaneda et al., 1995; Sutovsky et al., 2000). It is then conceivable that the integration of numts is linked to the hybridization event between a female with the extant mitochondrial lineage and a distantly related male from which the pseudogene originated.

So we propose that the origin of the pseudogene is linked to a hybridization event between distant populations. Under the alternative depicted in (b), we would expect the pseudogene branch to be shorter than their mitochondrial companions since numts must have evolved into the nucleus during at least 1.9 Myr. The long branches of the pseudogenes and reduced differentiation between numt copies are compatible with a recent transposition into the nucleus and its spread through sexual reproduction as represented in Fig. 5c. This is in accordance with the male-biased dispersal trough Europe during the late Pleistocene that we had hypothesized after the observation of very little diversity in the genus for nuclear markers (Pérez et al., 2017) and in particular for markers of the Y chromosome (Pérez et al., 2011).

This study shows how the fortuitous finding of mitochondrial pseudogenes reveals central facts in the evolution of a genus. The identification and study of numts in genome sequencing projects and its posterior screening in populations will provide valuable information about past diversity of mitochondrial sequences and phylogeny of extant organisms, especially regarding the role of hybridization among divergent lineages in the evolution.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gene.2017.07.035.

Acknowledgements

This work was funded by Ministerio de Economía y Competitividad, Spain (Grant number CGL2011-25117).

For providing chamois samples the authors are indebted to the following institutions: the Regional Governments of Principado de Asturias (Consejería de Agricultura) and Aragón (Diputación General de Aragón), the game wardens of Asturias and Aragón, Camino Real Hunting, and the following people: Sabine Hammer, Lucas Rossi, Juan Bejar, Paloma Barrachina, Jacques Michallet, Haritakis. Papaioannou, Tomas Skalski, Jason Badridze and Alvaro Mazón. We thank Sara de Albornoz for the correction and improvement of our English.

References

Arctander, P., 1995. Comparison of a mitochondrial gene and a corresponding nuclear pseudogene. Proc. Biol. Sci. 262, 13–19.
Baker, A.J., Moeed, A., 1987. Rapid genetic differentiation and founder effect in colonizing populations of common mynas (*Acridotheres tristis*). Evolution 41, 525–538.
Baldo, L., de Queiroz, A., Hedin, M., Hayashi, C.Y., Gatesy, J., 2011. Nuclear-mitochondrial sequences as witnesses of past interbreeding and population diversity in the jumping bristletail *Mesomachilis*. Mol. Biol. Evol. 28, 195–210.
Bensasson, D., Zhang, D.X., Hartl, D.L., Hewitt, G.M., 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends Ecol. Evol. 16, 314–321.

Bibi, F., 2013. A multi-calibrated mitochondrial phylogeny of extant Bovidae (Artiodactyla, Ruminantia) and the importance of the fossil record to systematics. BMC Evol. Biol. 13, 166.

Blanchard, J.L., Schmidt, G.W., 1996. Mitochondrial DNA migration events in yeast and humans: integration by a common end-joining mechanism and alternative perspectives on nucleotide substitution patterns. Mol. Biol. Evol. 13, 893.

Brown, W.M., George Jr., M., Wilson, A.C., 1979. Rapid evolution of animal mitochondrial DNA. Proc. Natl. Acad. Sci. U. S. A. 76, 1967–1971.

Cattaneo, C., Smillie, D.M., Gelsthorpe, K., Piccinini, A., Gelsthorpe, A.R., Sokol, R.J., 1995. A simple method for extracting DNA from old skeletal material. Forensic Sci. Int. 74, 167–174.

Cohen, K.M., Finney, S.C., Gibbard, P.L., Fan, J.-X., 2013. The ICS international chronostratigraphic chart. Episodes 36, 199–204.

Corlatti, L., Lorenzini, R., Lovari, S., 2011. The conservation of the chamois *Rupicapra* spp. Mammal Rev. 41, 163–174.

Crestanello, B., Pecchioli, E., Vernesi, C., Mona, S., Martínková, N., Janiga, M., Hauffe, H.C., Bertorelle, G., 2009. The genetic impact of translocations and habitat fragmentation in chamois (*Rupicapra*) spp. J. Hered. 100, 691–708.

DeWoody, J.A., Chesser, R.K., Baker, R.J., 1999. A translocated mitochondrial cytochrome b pseudogene in voles (Rodentia: Microtus). J. Mol. Evol. 48, 380–382.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214.

Engels, B., 2005. Amplify 3.

Ermakov, O.A., Simonov, E., Surin, V.L., Titov, S.V., Brandler, O.V., Ivanova, N.V., Borisenko, A.V., 2015. Implications of hybridization, NUMTs, and overlooked diversity for DNA barcoding of Eurasian ground squirrels. PLoS One 10, e0117201.

Estoup, A., Largiader, C.R., Perrot, E., Chourrout, D., 1996. Rapid one-tube DNA extraction for reliable PCR detection of fish polymorphic markers and transgenes. Mol. Mar. Biol. Biotechnol. 5, 295–298.

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

Fernández, P., Cregut, E., 2007. Les Caprinae (Rupicaprini, Ovibovini, Ovini et Caprini) de la séquence pléstocène de Kozarnika (Bulgarie du Nord). Rev. Paléobiol. 26, 425–503.

Grubb, P., 1993. Orden Artiodactila, Mammal Species of the World. Smithsonian Institution Press, Washington, pp. 374–414.

Haran, J., Koutroumpa, F., Magnoux, E., Roques, A., Roux, G., 2015. Ghost mtDNA haplotypes generated by fortuitous NUMTs can deeply disturb infra-specific genetic diversity and phylogeographic pattern. J. Zool. Syst. Evol. Res. 53, 109–115.

Hassanin, A., Delsuc, F., Ropiquet, A., Hammer, C., Jansen van Vuuren, B., Matthee, C., Ruiz-Garcia, M., Catzeflis, F., Areskoug, V., Nguyen, T.T., Couloux, A., 2012. Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. C. R. Biol. 335, 32–50.

Hazkani-Covo, E., 2009. Mitochondrial insertions into primate nuclear genomes suggest the use of numts as a tool for phylogeny. Mol. Biol. Evol. 26, 2175–2179.

Hazkani-Covo, E., Zeller Rm Fau-Martin, W., Martin, W., 2010. Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. PLoS Genet. 6.

Hedrick, P.W., 2010. Genetics of Populations, 4th ed. Jones and Bartlett Publishers, Sudbury, Mass.

Ho, S.Y., Phillips, M.J., 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. Syst. Biol. 58, 367–380.

Kaneda, H., Hayashi, J., Takahama, S., Taya, C., Lindahl, K.F., Yonekawa, H., 1995. Elimination of paternal mitochondrial DNA in intraspecific crosses during early mouse embryogenesis. Proc. Natl. Acad. Sci. U. S. A. 92, 4542–4546.

Kim, J.H., Antunes, A., Luo, S.J., Menninger, J., Nash, W.G., O'Brien, S.J., Johnson, W.E., 2006. Evolutionary analysis of a large mtDNA translocation (numt) into the nuclear genome of the *Panthera* genus species. Gene 366, 292–302.

Ko, Y.-J., Yang, E.C., Lee, J.-H., Lee, K.W., Jeong, J.-Y., Park, K., Chung, O., Bhak, J., Lee, J.-H., Yim, H.-S., 2015. Characterization of cetacean Numt and its application into cetacean phylogeny. Genes & Genomics 37, 1061–1071.

Kumar, S., Stecher, G., Peterson, D., Tamura, K., 2012. MEGA-CC: computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis. Bioinformatics 28, 2685–2686.

Lalueza-Fox, C., Castresana, J., Sampietro, L., Marques-Bonet, T., Alcover, J.A., Bertranpetit, J., 2005. Molecular dating of caprines using ancient DNA sequences of *Myotragus balearicus*, an extinct endemic Balearic mammal. BMC Evol. Biol. 5.

Leister, D., 2005. Origin, evolution and genetic effects of nuclear insertions of organelle DNA. Trends Genet. 21, 655–663.

Li, W.H., Gojobori, T., Nei, M., 1981. Pseudogenes as a paradigm of neutral evolution. Nature 292, 237–239.

López, J.V., Yuhki, N., Masuda, R., Modi, W., O'Brien, S.J., 1994. Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. J. Mol. Evol. 39, 174–190.

López, J.V., Culver, M., Stephens, J.C., Johnson, W.E., O'Brien, S.J., 1997. Rates of nuclear and cytoplasmic mitochondrial DNA sequence divergence in mammals. Mol. Biol. Evol. 14, 277–286.

Lu, X.M., Fu, Y.X., Zhang, Y.P., 2002. Evolution of mitochondrial cytochrome B pseudogene in genus Nycticebus. Mol. Biol. Evol. 19, 2337–2341.

Miraldo, A., Hewitt, G.M., Dear, P.H., Paulo, O.S., Emerson, B.C., 2012. Numts help to reconstruct the demographic history of the ocellated lizard (*Lacerta lepida*) in a secondary contact zone. Mol. Ecol. 21, 1005–1018.

Nei, M., Kumar, S., 2000. Molecular Evolution and Phylogenetics. Oxford University Press, New York.

Pérez, T., Albornoz, J., Domínguez, A., 2002. Phylogeography of chamois (*Rupicapra* spp.) inferred from microsatellites. Mol. Phylogenet. Evol. 25, 524–534.

Pérez, T., Hammer, S.E., Albornoz, J., Domínguez, A., 2011. Y-chromosome phylogeny in the evolutionary net of chamois (genus *Rupicapra*). BMC Evol. Biol. 11, 272.

Pérez, T., Essler, S., Palacios, B., Albornoz, J., Domínguez, A., 2013. Evolution of the melanocortin-1 receptor gene (MC1R) in chamois (*Rupicapra* spp.). Mol. Phylogenet. Evol. 67, 621–625.

Pérez, T., González, I., Essler, S.E., Fernández, M., Domínguez, A., 2014. The shared mitochondrial genome of *Rupicapra pyrenaica ornata* and *Rupicapra rupicapra cartusiana*: old remains of a common past. Mol. Phylogenet. Evol. 79, 375–379.

Pérez, T., Fernández, M., Hammer, S.E., Domínguez, A., 2017. Multilocus intron trees reveal extensive male-biased homogenization of ancient populations of chamois (*Rupicapra* spp.) across Europe during Late Pleistocene. PLoS One 12, e0170392.

Pons, J., Vogler, A.P., 2005. Complex pattern of coalescence and fast evolution of a mitochondrial rRNA pseudogene in a recent radiation of tiger beetles. Mol. Biol. Evol. 22, 991–1000.

Rambaut, A., 2006. FigTree: Tree Figure Drawing Tool, Version 1.4.2. Institute of Evolutionary Biology, University of Edinburgh.

Rambaut, A. and Drummond, A.J.: Tracer Institute of Evolutionary Biology, University of Edinburgh (2009).

Richly, E., Leister, D., 2004. NUMTs in sequenced eukaryotic genomes. Mol. Biol. Evol. 21, 1081–1084.

Rodríguez, F., Albornoz, J., Domínguez, A., 2007. Cytochrome b pseudogene originated from a highly divergent mitochondrial lineage in genus *Rupicapra*. J. Hered. 98, 243–249.

Rodríguez, F., Hammer, S., Pérez, T., Suchentrunk, F., Lorenzini, R., Michallet, J., Martinkova, N., Albornoz, J., Domínguez, A., 2009. Cytochrome b phylogeography of chamois (*Rupicapra* spp.). Population contractions, expansions and hybridizations governed the diversification of the genus. J. Hered. 100, 47–55.

Rodríguez, F., Pérez, T., Hammer, S.E., Albornoz, J., Domínguez, A., 2010. Integrating phylogeographic patterns of microsatellite and mtDNA divergence to infer the evolutionary history of chamois (genus *Rupicapra*). BMC Evol. Biol. 10, 222.

Sambrook, J., Fritsch, E., Maniatis, T., 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, New York.

Schmitz, J., Piskurek, O., Zischler, H., 2005. Forty million years of independent evolution: a mitochondrial gene and its corresponding nuclear pseudogene in primates. J. Mol. Evol. 61, 1–11.

Shi, H., Dong, J., Irwin, D.M., Zhang, S., Mao, X., 2016. Repetitive transpositions of mitochondrial DNA sequences to the nucleus during the radiation of horseshoe bats (Rhinolophus, Chiroptera). Gene 581, 161–169.

Sunnucks, P., Hales, D.F., 1996. Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus Sitobion (Hemiptera: Aphididae). Mol. Biol. Evol. 13, 510–524.

Sutovsky, P., Moreno, R.D., Ramalho-Santos, J., Dominko, T., Simerly, C., Schatten, G., 2000. Ubiquitinated sperm mitochondria, selective proteolysis, and the regulation of mitochondrial inheritance in mammalian embryos. Biol. Reprod. 63, 582–590.

Tajima, F., 1993. Simple methods for testing the molecular evolutionary clock hypothesis. Genetics 135, 599–607.

Thalmann, O., Hebler, J., Poinar, H.N., Paabo, S., Vigilant, L., 2004. Unreliable mtDNA data due to nuclear insertions: a cautionary tale from analysis of humans and other great apes. Mol. Ecol. 13, 321–335.

Verscheure, S., Backeljau, T., Desmyter, S., 2015. In silico discovery of a nearly complete mitochondrial genome Numt in the dog (*Canis lupus familiaris*) nuclear genome. Genetica 143, 453–458.

Zischler, H., Geisert, H., Castresana, J., 1998. A hominoid-specific nuclear insertion of the mitochondrial D-loop: implications for reconstructing ancestral mitochondrial sequences. Mol. Biol. Evol. 15, 463–469.