



# Mitochondrial DNA insertions in the nuclear *Capra hircus* genome

F.Y. Ning<sup>1,2</sup>, J. Fu<sup>1,2</sup> and Z.H. Du<sup>1,2</sup>

<sup>1</sup>China Key Laboratory of Animal Genetics, Breeding and Reproduction, Education Department of Heilongjiang Province, Northeast Agricultural University, Harbin, China

<sup>2</sup>College of Animal Science and Technology, Northeast Agricultural University, Harbin, China

Corresponding author: F.Y. Ning

E-mail: nfy1900@sina.com

Genet. Mol. Res. 16 (1): gmr16018266

Received October 14, 2016

Accepted December 11, 2016

Published January 23, 2017

DOI <http://dx.doi.org/10.4238/gmr16018266>

Copyright © 2017 The Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution ShareAlike (CC BY-SA) 4.0 License.

**ABSTRACT.** Nuclear mitochondrial pseudogenes (numts), originating from mtDNA insertions into the nuclear genome, have been detected in many species. However, the distribution of numts in the newly published nuclear genome of domestic goat (*Capra hircus*) has not yet been explored. We used the entire goat mtDNA sequence and nuclear genome, to identify 118 numts using BLAST. Of these, 79 were able to map sequences to the genome. Further analysis showed that the size of the numts ranged from 318 to 9608 bp, and the homologous identity between numts and their respective corresponding mtDNA fragments varied between 65 and 99%. The identified Yunnan black goat numts covered nearly all the mitochondrial genes including mtDNA control region, and were distributed over all chromosomes with the exception of chromosomes 18, 21, and 25. The Y chromosome was excluded from our analysis, as sequence data are currently not available. Among

the discovered 79 numts that we were able to map to the genome, 26 relatively complete mitochondrial genes were detected. Our results constitute valuable information for subsequent studies related to mitochondrial genes and goat evolution.

**Key words:** Goat genome; Mitochondrial pseudogenes

## INTRODUCTION

Hundreds of DNA sequences homogenous to mtDNA fragments, often referred to as nuclear mitochondrial DNA (numts) or mitochondrial pseudogenes (Lopez et al., 1994), have been verified to exist in nuclear genomes. Since their first discovery (du Buy and Riley, 1967), numts have been found in numerous species (Bensasson et al., 2001; Leister, 2005; Liu and Zhao, 2007; Hassanin et al., 2010; Hazkani-Covo et al., 2010; Nergadze et al., 2010; Song et al., 2013). An increasing research effort is now directed at determining the evolution of numts and their possible function and influence on studies involving mitochondrial genes (Ricchetti et al., 2004; Thalmann et al., 2004; Schmitz et al., 2005; Cleaver et al., 2014).

The mitochondrion plays a central role in cell metabolism and many human metabolic disorders have been reported to be closely related with mtDNA polymorphisms (Greaves and Taylor, 2006; Yu et al., 2014). As in the field of animal genetics, studies of mtDNA polymorphism and genetic variation are also increasing in numbers. Due to their homology with the counterpart mtDNA fragments, numts could severely influence the results of analyses involving polymerase chain reaction (PCR) methods and might even result in completely erroneous conclusions (Wallace et al., 1997; Nergadze et al., 2010). An example of such a case, is the argument about the relationship between Alzheimer's disease and mtDNA polymorphisms (Davis et al., 1997; Hirano et al., 1997; Wallace et al., 1997; Wang et al., 2014). Thus, as ancient mtDNA sequences in the nuclear genome, numts are a potential source of error when identifying mutations in studies of, e.g., animal product traits and human diseases. Therefore, it is very important to clarify the distribution of mtDNA numts.

In *Capra hircus*, the Vietnamese domestic goat, numts were first reported using a low-quality genome (Hassanin et al., 2010). Recently, the whole genome (~2.66 Gb) sequence of a female Yunnan black goat was published (Dong et al., 2013). Whole-genome mapping data facilitated the assembly of super-scaffolds >5X longer based on the N50 metric, compared to scaffolds augmented by fosmid-end sequencing (scaffold N50 = 3.06 Mb, super-scaffold N50 = 16.3 Mb) (Dong et al., 2013). This provides us with an excellent opportunity to analyze the distribution of numts across the whole domestic goat genome. The objective of our study was to describe numt distribution in the genome and to detect any numts corresponding to relatively complete mtDNA genes. In addition to deepening our understanding of numt evolution, this could provide important reference information to avoid PCR amplification errors in studies relating to mtDNA.

## MATERIAL AND METHODS

### Identification and characterization of numts

In our study, BLASTn, the basic local alignment search tool, was used to find regions

of local sequence similarity across the whole Yunnan black goat mitochondrial genome (accession No. KJ940969), in the HGSC. This was done by selecting the *Bos taurus* linear scaffold databases that were released on August 19, 2012 (genome browser v. 1.0; domestic goat (taxid: 9925); CHIR 1.0). In our searches, we set the maximum expectation value to be  $e = 10^{-4}$ , to recover only biologically significant accession hits (Pereira and Baker, 2004; Richly and Leister, 2004), and the Max score was set to  $>250$ . No other filters were used. If the numts' and their mitochondrial counterparts' distances were well-matched, they were joined together and taken as a single event (Woischnik and Moraes, 2002). The identity between the joined numt and the mtDNA counterpart was identified using the DNAMAN v. 5.0 software for Windows, following the suggested sequence alignment method (<http://www.lynnon.com/>).

### Prediction of secondary structures of tRNAs

Prediction of the secondary structure for the mt-tRNA and their corresponding numt-tRNA with a complete tRNA sequence were performed using the DNA M-fold web server (Zuker, 2003). We set 37°C as the folding temperature, while the default values were used for all other parameters.

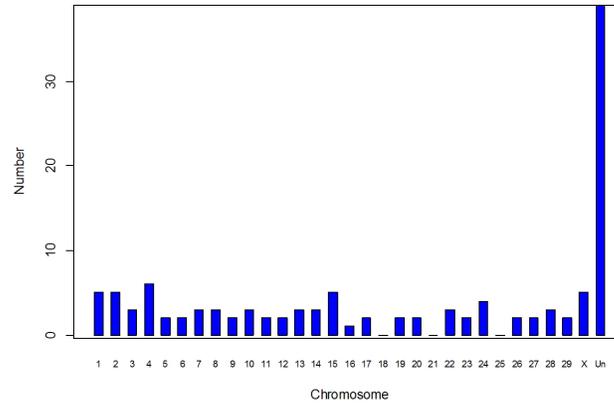
## RESULTS AND DISCUSSION

BLAST identified 200 alignments. After modifying the 200 hits according to the method described above, we identified 118 numts in the female Yunnan black goat nuclear genome (partial results are shown in Table 1 and all results of their counterpart sequences, which cover almost the whole mitochondrial genome, are shown in the [Table S1](#)). Except for the unavailable Y chromosome data, the integrations of mitochondrial fragments occurred across all the domestic Yunnan black goat nuclear genome except for chromosomes 18, 21, and 25 (Figure 1).

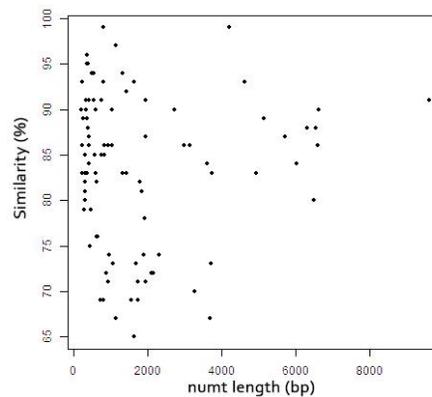
**Table 1.** Results of detected numts in the cattle nuclear genome (BLASTn in the HGSC databases).

Numt segment corresponding to the mitochondrial genes	mtDNA			Numt			Score	Chromosome	Identity (%)	Accession No.
	Start	End	Length	Start	End	Length				
tRNA <sup>Leu</sup> -tRNA <sup>His</sup>	3289	12834	9546	30444	40051	9608	13564	X	91	NW_005100994.1
Cox3-tRNA <sup>Glu</sup>	9197	15826	6630	2127946	2121330	6617	9032	4	90	NW_005100653.1
tRNA <sup>Leu</sup> -tRNA <sup>Tyr</sup>	2351	6552	4202	36709455	36710270	816	7524	3	99	NW_005100637.1
tRNA <sup>Val</sup> -tRNA <sup>Ala</sup>	1677	6317	4641	15055897	15060529	4633	6897	7	93	NW_005100725.1
None	11459	14190	2732	17247653	17250377	2725	3699	6	90	NW_005100696.1
tRNA <sup>Gln</sup> -tRNA <sup>Tyr</sup>	4918	6896	1979	1035802	1037764	1963	2767	12	91	NW_005100794.1
tRNA <sup>His</sup> -tRNA <sup>Leu</sup>	11518	13137	1620	5073897	5075533	1637	2459	8	93	NW_005100742.1
tRNA <sup>His</sup> -tRNA <sup>Leu</sup>	11700	13143	1444	6782171	6783620	1450	2107	7	92	NW_005100721.1
tRNA <sup>Phe</sup>	682	1840	1159	1064562	1065720	1159	1945	29	97	NW_005100968.1
tRNA <sup>Lys</sup> -ATP8	8720	9752	1033	5448900	5449936	1037	1418	13	90	NW_005100802.1
tRNA <sup>Phe</sup>	958	1554	597	211060	211655	596	816	1	90	NW_005100594.1
ND3-tRNA <sup>Arg</sup>	10730	11153	424	6027147	6027570	424	589	16	91	NW_005100845.1
None	8345	8702	358	157904	157547	358	488	3	90	NW_005100643.1
None	9527	9591	65	7138485	7138819	335	470	9	91	NW_005100754.1

The similarities between the numts and their corresponding mtDNA fragments varied between 65 and 99%, and the numt fragment length ranged between 318 to 9608 bp. Further analysis revealed that the length of numts and similarities between numts and corresponding mtDNA fragments were concentrated from 65 to 99% and 218 to 1963 bp, respectively (Figure 2).

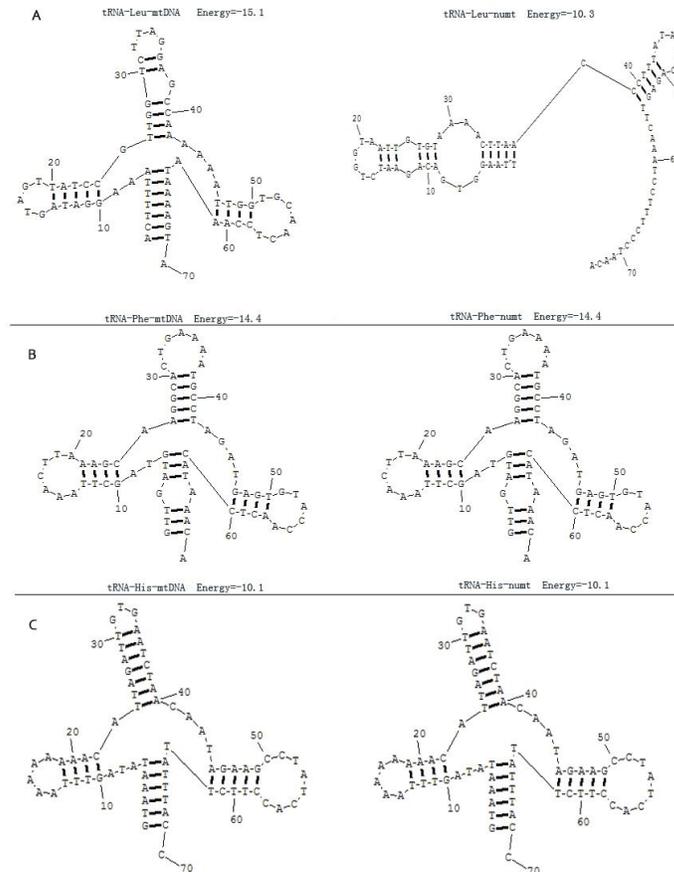


**Figure 1.** Number of numts found in each chromosome. Un indicates the number of numts that did not map on any chromosome (bp).



**Figure 2.** Distribution of numt length (bp) and similarity (%) between the numts and their corresponding mtDNA fragment.

The total length of the Yunnan black goat numt fragments was found to be 227,783 bp, about 13.9 times longer than the entire Yunnan black goat mitochondrial genome. We identified only 34 relatively complete mitochondrial gene regions distributed over 79 numts, in which 12 protein-coding genes (ND1, ND2, CoxI, CoxII, CoIII, ATP8, ATP6, ND3, ND4L, ND4, ND6, and Cyt b) and all the 22 mt-tRNA genes were detected, except for srRNA (16s and 12s rRNA). Further analysis of the 12 protein-coding genes revealed multiple-internal termination codons, regardless of whether a universal or vertebrate mitochondrial genetic codon was selected. Thus, the genes could not be translated into functional proteins. In addition, based on the predicted secondary structures, we found that most of the numts could not be folded to perfect structures that would be necessary for active function, for example tRNA<sup>Leu</sup> (Figure 3A). Although our results indicated that most of the numts did not exhibit any functions, some studies have found numts present in predicted gene exons and introns in the human genome (Ricchetti et al., 2004). In some cases, we also obtained an identical secondary tRNA between the mtDNA and chromosomal numts in Yunnan black goat (e.g., tRNA<sup>Phe</sup> and tRNA<sup>His</sup>; Figure 3B and C).



**Figure 3.** Predicted secondary tRNA structures and their corresponding nuclear pseudogenes, obtained using M-fold web server. **A.** tRNA<sup>Leu</sup>. **B.** tRNA<sup>Phe</sup>. **C.** tRNA<sup>His</sup>.

Our results suggest that numts might participate in the function and regulation of genes, and there is therefore a need to study their potential functions. For example, one study has verified nuclear mitochondrial pseudogene effects on DNA barcoding (Kim et al., 2013). When analyzing mtDNA, the mtDNA fragments themselves will greatly predominate the copy number, which should not influence the study results in routine applications. However, there are a few specific situations under which the primers may match numts better than the intended mtDNA, thus resulting in contaminated mtDNA products and a preferential amplification of the numts (Collura and Stewart, 1995; Liu and Zhao, 2007): i) a mutation in the authentic mtDNA sequence at the 3'-most primer position (Zullo et al., 1993), ii) the attempt to amplify mtDNA from cell/tissue types with a reduced mtDNA content (Greenwood and Pääbo, 1999), iii) or during the amplification of ancient mtDNA, especially samples contaminated with endogenous or exogenous DNA (Willerslev and Cooper, 2005). Therefore, the method using sequence-specific primers was developed in genotyping of mtDNA more than 20 years ago (Garritsen et al., 1997; Hori et al., 2003).

Regarding the variable D-loop region, which is often used in analyses of the evolutionary relationship of species (Saccone et al., 1987; Zhang et al., 1995), we did not find a complete numt fragment corresponding to the most variable region of the mtDNA in Yunnan black goat genome (**Table S1**). Therefore, the existing D-loop numts would not influence results relating to this region. The same was observed also for the Cyt b gene, which is also regularly used among and within species studies (Esposti et al., 1993; Kawamura et al., 2014; Kartavtsev et al., 2016).

It is commonly believed that numts were formed by non-homologous recombination (Henze and Martin, 2001; Woischnik and Moraes, 2002), which are independent and continuous on-going events (Mourier et al., 2001). Once numts are integrated in the nuclear genome, they and their mitochondrial counterparts would evolve independently and showing different patterns (Lopez et al., 1994; Henze and Martin, 2001; Kazazian, 2004). These different patterns could be used for investigating the relationship between numts and their corresponding mtDNA sequences; for example, the time point of the mtDNA insertion in the nuclear genome and the evolutionary relationship between the sequences located in the mitochondrial and nuclear genomes (Schmitz et al., 2005; Kim et al., 2006). However, to date, the mechanisms responsible for the transfer of DNA fragments from mitochondria to nuclei are still elusive (Hazkani-Covo et al., 2010).

In our study, 118 highly homogenous numt fragments were identified between mitochondria and nuclear genome in Yunnan black goat. These numts included most tRNAs and protein coding genes, except for rRNAs in the mitochondrial DNA. The relatively complete numts could serve as powerful tools to analyze the relationship of different goat species and reconstruct the phylogenetic network by comparing their mtDNA counterparts.

## Conflicts of interest

The authors declare no conflict of interest.

## REFERENCES

- Bensasson D, Zhang D, Hartl DL and Hewitt GM (2001). Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends Ecol. Evol. (Amst.)* 16: 314-321. [http://dx.doi.org/10.1016/S0169-5347\(01\)02151-6](http://dx.doi.org/10.1016/S0169-5347(01)02151-6)
- Cleaver JE, Brennan-Minnella AM, Swanson RA, Fong KW, et al. (2014). Mitochondrial reactive oxygen species are scavenged by Cockayne syndrome B protein in human fibroblasts without nuclear DNA damage. *Proc. Natl. Acad. Sci. USA* 111: 13487-13492. <http://dx.doi.org/10.1073/pnas.1414135111>
- Collura RV and Stewart CB (1995). Insertions and duplications of mtDNA in the nuclear genomes of Old World monkeys and hominoids. *Nature* 378: 485-489. <http://dx.doi.org/10.1038/378485a0>
- Davis RE, Miller S, Herrstadt C, Ghosh SS, et al. (1997). Mutations in mitochondrial cytochrome c oxidase genes segregate with late-onset Alzheimer disease. *Proc. Natl. Acad. Sci. USA* 94: 4526-4531. <http://dx.doi.org/10.1073/pnas.94.9.4526>
- Dong Y, Xie M, Jiang Y, Xiao N, et al. (2013). Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). *Nat. Biotechnol.* 31: 135-141. <http://dx.doi.org/10.1038/nbt.2478>
- du Buy HG and Riley FL (1967). Hybridization between the nuclear and kinetoplast DNA's of *Leishmania enriettii* and between nuclear and mitochondrial DNA's of mouse liver. *Proc. Natl. Acad. Sci. USA* 57: 790-797. <http://dx.doi.org/10.1073/pnas.57.3.790>
- Esposti MD, De Vries S, Crimi M, Ghelli A, et al. (1993). Mitochondrial cytochrome b: evolution and structure of the protein. *Biochim. Biophys. Acta* 1143: 243-271. [http://dx.doi.org/10.1016/0005-2728\(93\)90197-N](http://dx.doi.org/10.1016/0005-2728(93)90197-N)
- Garritsen HS, Szufflad P, Sibrowski W and Dzik WH (1997). A sequence-specific polymerase chain reaction assay for mitochondrial DNA polymorphisms in human platelets and white cells. *Transfusion* 37: 1012-1019. <http://dx.doi.org/10.1046/j.1537-2995.1997.371098016438.x>

- Greaves LC and Taylor RW (2006). Mitochondrial DNA mutations in human disease. *IUBMB Life* 58: 143-151. <http://dx.doi.org/10.1080/15216540600686888>
- Greenwood AD and Pääbo S (1999). Nuclear insertion sequences of mitochondrial DNA predominate in hair but not in blood of elephants. *Mol. Ecol.* 8: 133-137. <http://dx.doi.org/10.1046/j.1365-294X.1999.00507.x>
- Hassanin A, Bonillo C, Nguyen BX and Cruaud C (2010). Comparisons between mitochondrial genomes of domestic goat (*Capra hircus*) reveal the presence of numts and multiple sequencing errors. *Mitochondrial DNA* 21: 68-76. <http://dx.doi.org/10.3109/19401736.2010.490583>
- Hazkani-Covo E, Zeller RM and Martin W (2010). Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. *PLoS Genet.* 6: e1000834. <http://dx.doi.org/10.1371/journal.pgen.1000834>
- Henze K and Martin W (2001). How do mitochondrial genes get into the nucleus? *Trends Genet.* 17: 383-387. [http://dx.doi.org/10.1016/S0168-9525\(01\)02312-5](http://dx.doi.org/10.1016/S0168-9525(01)02312-5)
- Hirano M, Shilbans A, Mayeux R, Davidson MM, et al. (1997). Apparent mtDNA heteroplasmy in Alzheimer's disease patients and in normals due to PCR amplification of nucleus-embedded mtDNA pseudogenes. *Proc. Natl. Acad. Sci. USA* 94: 14894-14899. <http://dx.doi.org/10.1073/pnas.94.26.14894>
- Hori K, Shin WS, Hemmi C, Toyo-oka T, et al. (2003). High fidelity SNP genotyping using sequence-specific primer elongation and fluorescence correlation spectroscopy. *Curr. Pharm. Biotechnol.* 4: 477-484. <http://dx.doi.org/10.2174/1389201033377391>
- Kartavtsev YP, Sharina SN, Saitoh K, Imoto JM, et al. (2016). Phylogenetic relationships of Russian far eastern flatfish (Pleuronectiformes, Pleuronectidae) based on two mitochondrial gene sequences, Co-1 and Cyt-b, with inferences in order phylogeny using complete mitogenome data. *Mitochondrial DNA* 27: 667-678. <http://dx.doi.org/10.3109/19401736.2014.913139>
- Kawamura K, Ueda T, Arai R and Smith C (2014). Phylogenetic relationships of bitterling fishes (Teleostei: Cypriniformes: Acheilognathinae), inferred from mitochondrial cytochrome B sequences. *Zool. Sci.* 31: 321-329. <http://dx.doi.org/10.2108/zs130233>
- Kazian Jr HH (2004). Mobile elements: drivers of genome evolution. *Science* 303: 1626-1632. <http://dx.doi.org/10.1126/science.1089670>
- Kim JH, Antunes A, Luo SJ, Menninger J, et al. (2006). Evolutionary analysis of a large mtDNA translocation (numt) into the nuclear genome of the *Panthera* genus species. *Gene* 366: 292-302. <http://dx.doi.org/10.1016/j.gene.2005.08.023>
- Kim SJ, Lee KY and Ju SJ (2013). Nuclear mitochondrial pseudogenes in *Austinograea alayseae* hydrothermal vent crabs (Crustacea: Bythograeidae): effects on DNA barcoding. *Mol. Ecol. Resour.* 13: 781-787. <http://dx.doi.org/10.1111/1755-0998.12119>
- Leister D (2005). Origin, evolution and genetic effects of nuclear insertions of organelle DNA. *Trends Genet.* 21: 655-663. <http://dx.doi.org/10.1016/j.tig.2005.09.004>
- Liu Y and Zhao X (2007). Distribution of nuclear mitochondrial DNA in cattle nuclear genome. *J. Anim. Breed. Genet.* 124: 264-268. <http://dx.doi.org/10.1111/j.1439-0388.2007.00674.x>
- Lopez JV, Yuhki N, Masuda R, Modi W, et al. (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.
- Mourier T, Hansen AJ, Willerslev E and Arctander P (2001). The Human Genome Project reveals a continuous transfer of large mitochondrial fragments to the nucleus. *Mol. Biol. Evol.* 18: 1833-1837. <http://dx.doi.org/10.1093/oxfordjournals.molbev.a003971>
- Nergadze SG, Lupotto M, Pellanda P, Santagostino M, et al. (2010). Mitochondrial DNA insertions in the nuclear horse genome. *Anim. Genet.* 41 (Suppl 2): 176-185. <http://dx.doi.org/10.1111/j.1365-2052.2010.02130.x>
- Pereira SL and Baker AJ (2004). Low number of mitochondrial pseudogenes in the chicken (*Gallus gallus*) nuclear genome: implications for molecular inference of population history and phylogenetics. *BMC Evol. Biol.* 4: 17. <http://dx.doi.org/10.1186/1471-2148-4-17>
- Ricchetti M, Tekaia F and Dujon B (2004). Continued colonization of the human genome by mitochondrial DNA. *PLoS Biol.* 2: E273. <http://dx.doi.org/10.1371/journal.pbio.0020273>
- Richly E and Leister D (2004). NUMTs in sequenced eukaryotic genomes. *Mol. Biol. Evol.* 21: 1081-1084. <http://dx.doi.org/10.1093/molbev/msh110>
- Saccone C, Attimonelli M and Sbisà E (1987). Structural elements highly preserved during the evolution of the D-loop-containing region in vertebrate mitochondrial DNA. *J. Mol. Evol.* 26: 205-211. <http://dx.doi.org/10.1007/BF02099853>
- Schmitz J, Piskurek O and 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. <http://dx.doi.org/10.1007/s00239-004-0293-3>
- Song S, Jiang F, Yuan J, Guo W, et al. (2013). Exceptionally high cumulative percentage of NUMTs originating from linear mitochondrial DNA molecules in the *Hydra magnipapillata* genome. *BMC Genomics* 14: 447. <http://dx.doi.org/10.1186/1471-2164-14-447>

- Thalmann O, Hebler J, Poinar HN, Pääbo S, et al. (2004). Unreliable mtDNA data due to nuclear insertions: a cautionary tale from analysis of humans and other great apes. *Mol. Ecol.* 13: 321-335. <http://dx.doi.org/10.1046/j.1365-294X.2003.02070.x>
- Wallace DC, Stuard C, Murdock D, Schurr T, et al. (1997). Ancient mtDNA sequences in the human nuclear genome: a potential source of errors in identifying pathogenic mutations. *Proc. Natl. Acad. Sci. USA* 94: 14900-14905. <http://dx.doi.org/10.1073/pnas.94.26.14900>
- Wang D, Liu L, Zhu X, Wu W, et al. (2014). Hesperidin alleviates cognitive impairment, mitochondrial dysfunction and oxidative stress in a mouse model of Alzheimer's disease. *Cell. Mol. Neurobiol.* 34: 1209-1221. <http://dx.doi.org/10.1007/s10571-014-0098-x>
- Willerslev E and Cooper A (2005). Ancient DNA. *Proc. Biol. Sci.* 272: 3-16. <http://dx.doi.org/10.1098/rspb.2004.2813>
- Woischnik M and Moraes CT (2002). Pattern of organization of human mitochondrial pseudogenes in the nuclear genome. *Genome Res.* 12: 885-893. <http://dx.doi.org/10.1101/gr.227202>. Article published online before print in May 2002
- Yu Q, Dong Y, Xu N, Qian K, et al. (2014). A novel role of the ferric reductase Cfl1 in cell wall integrity, mitochondrial function, and invasion to host cells in *Candida albicans*. *FEMS Yeast Res.* 14: 1037-1047.
- Zhang DX, Szymura JM and Hewitt GM (1995). Evolution and structural conservation of the control region of insect mitochondrial DNA. *J. Mol. Evol.* 40: 382-391. <http://dx.doi.org/10.1007/BF00164024>
- Zuker M (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31: 3406-3415. <http://dx.doi.org/10.1093/nar/gkg595>
- Zullo S, Kennedy JL, Gelernter J, Polymeropoulos MH, et al. (1993). Eliminating mitochondrial DNA competition for nuclear DNA primers. *PCR Methods Appl.* 3: 39-45. <http://dx.doi.org/10.1101/gr.3.1.39>

## Supplementary material

**Table S1.** Results of detected numts in the cattle nuclear genome (BLASTn in the HGSC database).