Mitochondrial DNA insertions in the nuclear Capra hircus genome

F.Y. Ning1,2, J. Fu1,2 and Z.H. Du1,2

1China Key Laboratory of Animal Genetics, Breeding and Reproduction, Education Department of Heilongjiang Province, Northeast Agricultural University, Harbin, China
2College 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
Mitochondrial numts in goat

Genetics and Molecular Research 16 (1): gmr16018266

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).

<table>
<thead>
<tr>
<th>Numt segment corresponding to the mitochondrial genes</th>
<th>mtDNA</th>
<th>Numt</th>
<th>Score</th>
<th>Chromosome</th>
<th>Identity (%)</th>
<th>Accession No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>tRNA&lt;sup&gt;Leu&lt;/sup&gt;-tRNA&lt;sup&gt;His&lt;/sup&gt;</td>
<td>3289</td>
<td>12834</td>
<td>9346</td>
<td>30444</td>
<td>X</td>
<td>KJ940969.1</td>
</tr>
<tr>
<td>tRNA&lt;sup&gt;Cys&lt;/sup&gt;-tRNA&lt;sup&gt;His&lt;/sup&gt;</td>
<td>5179</td>
<td>15836</td>
<td>6620</td>
<td>21279</td>
<td>Y</td>
<td>KJ940969.1</td>
</tr>
<tr>
<td>tRNA&lt;sup&gt;Phe&lt;/sup&gt;-tRNA&lt;sup&gt;Val&lt;/sup&gt;</td>
<td>2351</td>
<td>5652</td>
<td>4202</td>
<td>36709</td>
<td>X</td>
<td>KJ940969.1</td>
</tr>
<tr>
<td>tRNA&lt;sup&gt;Leu&lt;/sup&gt;-tRNA&lt;sup&gt;Tyr&lt;/sup&gt;</td>
<td>1677</td>
<td>6317</td>
<td>4641</td>
<td>15055</td>
<td>X</td>
<td>KJ940969.1</td>
</tr>
<tr>
<td>tRNA&lt;sup&gt;Val&lt;/sup&gt;-tRNA&lt;sup&gt;Ala&lt;/sup&gt;</td>
<td>11459</td>
<td>14190</td>
<td>2732</td>
<td>17247</td>
<td>Y</td>
<td>KJ940969.1</td>
</tr>
<tr>
<td>tRNA&lt;sup&gt;Gln&lt;/sup&gt;-tRNA&lt;sup&gt;Tyr&lt;/sup&gt;</td>
<td>4918</td>
<td>6936</td>
<td>1979</td>
<td>10358</td>
<td>Y</td>
<td>KJ940969.1</td>
</tr>
<tr>
<td>tRNA&lt;sup&gt;His&lt;/sup&gt;-tRNA&lt;sup&gt;Leu&lt;/sup&gt;</td>
<td>11513</td>
<td>13137</td>
<td>1620</td>
<td>50739</td>
<td>Y</td>
<td>KJ940969.1</td>
</tr>
<tr>
<td>tRNA&lt;sup&gt;Phe&lt;/sup&gt;-tRNA&lt;sup&gt;Val&lt;/sup&gt;</td>
<td>11700</td>
<td>13143</td>
<td>1444</td>
<td>67821</td>
<td>Y</td>
<td>KJ940969.1</td>
</tr>
<tr>
<td>tRNA&lt;sup&gt;His&lt;/sup&gt;-tRNA&lt;sup&gt;Leu&lt;/sup&gt;</td>
<td>682</td>
<td>1840</td>
<td>597</td>
<td>21106</td>
<td>Y</td>
<td>KJ940969.1</td>
</tr>
<tr>
<td>tRNA&lt;sup&gt;Phe&lt;/sup&gt;-tRNA&lt;sup&gt;Lys&lt;/sup&gt;</td>
<td>958</td>
<td>1554</td>
<td>597</td>
<td>21106</td>
<td>Y</td>
<td>KJ940969.1</td>
</tr>
<tr>
<td>tRNA&lt;sup&gt;Arg&lt;/sup&gt;-tRNA&lt;sup&gt;Val&lt;/sup&gt;</td>
<td>10750</td>
<td>11153</td>
<td>424</td>
<td>66027</td>
<td>Y</td>
<td>KJ940969.1</td>
</tr>
<tr>
<td>tRNA&lt;sup&gt;Val&lt;/sup&gt;-tRNA&lt;sup&gt;Leu&lt;/sup&gt;</td>
<td>8345</td>
<td>8702</td>
<td>358</td>
<td>15798</td>
<td>Y</td>
<td>KJ940969.1</td>
</tr>
<tr>
<td>tRNA&lt;sup&gt;Leu&lt;/sup&gt;-tRNA&lt;sup&gt;Tyr&lt;/sup&gt;</td>
<td>9524</td>
<td>9551</td>
<td>65</td>
<td>91385</td>
<td>Y</td>
<td>KJ940969.1</td>
</tr>
</tbody>
</table>

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).
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 tRNALeu (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^Phe^ and tRNA^His^; Figure 3B and C).
Figure 3. Predicted secondary tRNA structures and their corresponding nuclear pseudogenes, obtained using M-fold web server. A. tRNA_{Leu}. B. tRNA_{Phe}. C. tRNA_{His}.

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 (Sacccone 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 numt fragment corresponding to the most variable region of the mtDNA in Yunnan black goat (Table S1). Therefore, the existing D-loop numts would not influence results relating to this region. This 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


Genetics and Molecular Research 16 (1): gmr16018266
Mitochondrial numts in goat


**Supplementary material**

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