Molecular Ecology Notes (2004) 4, 765–767 doi: 10.1111/j.1471-8286.2004.00744.x

TECHNICAL NOTE

Mitochondrial DNA structure and organization of the control region of South American camelids

M. L. MATÉ, F. DI ROCCO, A. ZAMBELLI and L. VIDAL-RIOJA
Instituto Multidisciplinario de Biología Celular, C. C. 403, 1900 La Plata, Buenos Aires, Argentina

Abstract

This work presents the mitochondrial DNA molecular organization of the control region (CR) of South American camelids. Sequencing of five individuals each of guanaco, llama, alpaca and vicuna species showed that this region spans 1060 bp including three conserved sequence blocks (CSB I–III) adjacent to the tRNAPhe gene, a conserved central domain and one extended termination-associated sequence in the 3′ domain of the CR close to the tRNAPro gene. A repeated array formed by three units of 26 bp was detected between CSB I and II. Alignment of the CR sequences from the four species shows a 337-bp segment that includes most of the nucleotide variability with 10 polymorphic sites. We suggest the use of this segment as a molecular marker to infer data on camelid genetic relationships and population diversity studies.

Keywords: haplotype, mitochondrial DNA control region, molecular marker, South American camelids

Received 20 February 2004; revision received 18 May 2004; accepted 16 June 2004

South American native artiodactyls are represented, among others, by the Tylopoda suborder which comprises two wild forms of camelids, guanaco (*Lama guanicoe*) and vicuna (*Vicugna vicugna*), and their respective domestic derivatives llama (*L. glama*) and alpaca (*L. pacos*). Traditionally, camelids have been the principal source of food and income for high Andean rural communities. Currently, the fineness and high value of their wool have increased the interest of breeders who have extended the exploitation of these species to other regions outside the Andes. In order to achieve the long-term viability and profitable use of the species, a number of issues about breeding and management conditions have been identified by institutions and organizations responsible for their local conservation (Torres 1995). Among them, a thorough knowledge is recommended of the genetic diversity of the populations from which camelids will be harvested for founding new populations or making reintroductions in older populations. In many species, including camelids, nuclear and mitochondrial DNA (mtDNA) markers are currently used to identify both population genetic structure and variability (Nagata *et al*. 1999; Kadwell *et al*. 2001; Bustamante *et al*. 2002, 2003; Nyakaana *et al*. 2002). Within mtDNA, the control region (CR) has been shown to evolve five times faster than coding regions. Therefore, its use is suitable for assessing genetic differentiation, population structure, haplotype distribution and phylogeny. This work is the first report on the sequence structure and organization of the mtDNA CR of the four South American camelids (SAC). Furthermore, in the 5′ domain of the CR we detected the presence of a 337-bp hypervariable segment and propose its use as a molecular marker for camelid studies on both the phylogenetic evolution and genetic diversity of populations.

During this investigation we analysed five unrelated individuals each of guanaco, llama, alpaca and vicuna from Argentina. Genomic DNA was isolated from blood samples as described in Bustamante *et al*. (2002). The complete mitochondrial CR was amplified by polymerase chain reaction using the following flanking primers, forward CTGATAAATCCCATAGAGC and reverse TTTCAGCGCC-TTGCTTTAAG, homologous to sequences of the tRNAPro and tRNAPhe genes, respectively. Amplifications were carried out in 15 μL reactions as reported elsewhere (Bustamante *et al*. 2003). The cycling parameters were 3 min at 94 °C followed by 35 cycles of 94 °C for 1 min, 52 °C for 1 min and 72 °C for 1 min with a final extension at 72 °C for 3 min. The amplification product was run in 1% agarose gel and the single band (about 1200 bp) was cut, purified and ligated into p-GEM T-easy vector (Promega). The transformation of *Escherichia coli* XL1Blue cells was followed by recombinant plasmid purification. Sequencing was performed

Correspondence: L. Vidal-Rioja. E-mail: lvidalrioja@imbice.org.ar

© 2004 Blackwell Publishing Ltd

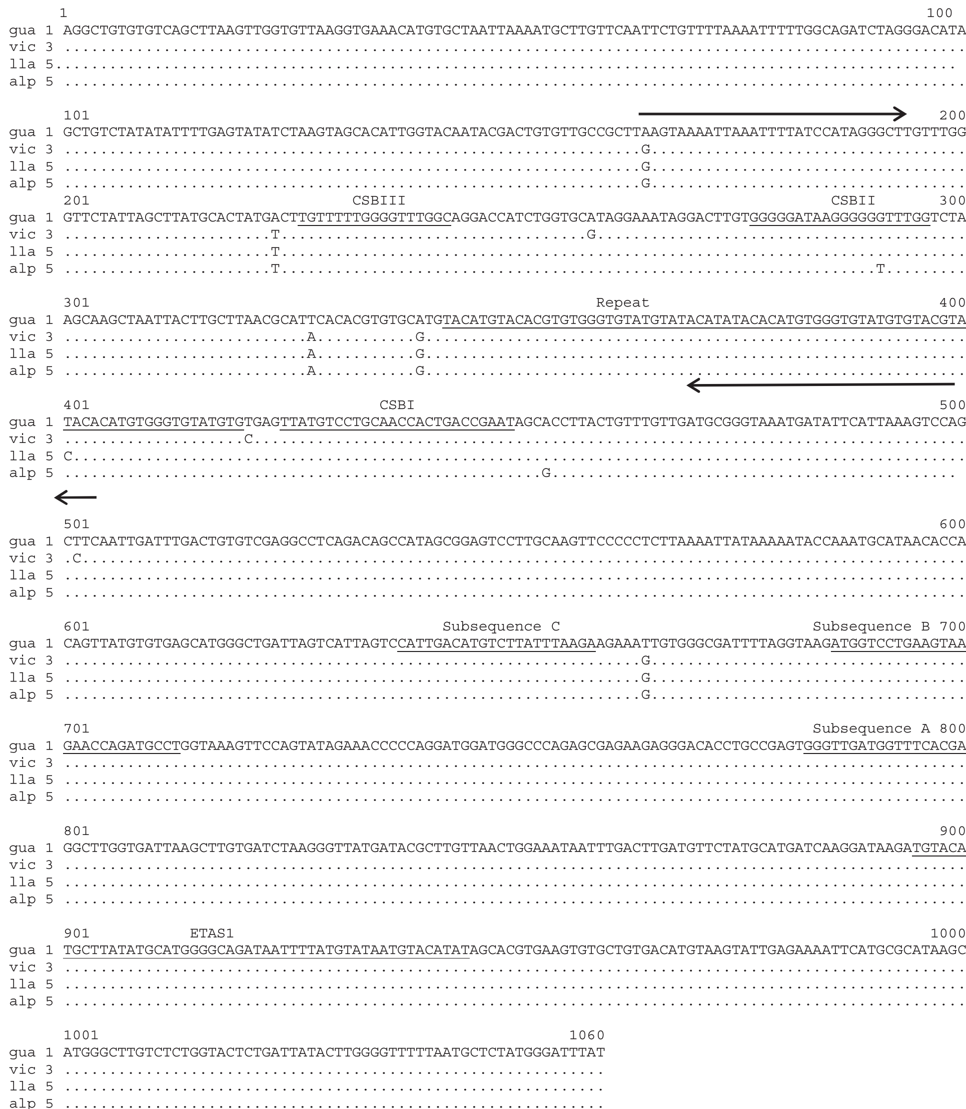

Fig. 1 Representative alignment of the complete control region from one individual of each South American camelid species. CSB I–III, conserved sequence blocks; ETAS I, extended termination-associated sequence I. The repeat includes three 26-bp repetition motifs. Dots represent nucleotide identity with the guanaco 1 reference sequence. Arrows indicate the beginning and end of the hypervariable region. gua, guanaco; vic, vicuna; lla, llama; alp, alpaca.

in an ABI PRISM 377 automatic sequencer (Applied Biosystem). Correlation between the obtained sequences and the mitochondrial CR was confirmed by the nucleotide BLAST program (www.ncbi.nih.gov). Sequences from all 20 SAC analysed were aligned using the multiple alignment option of the CLUSTAL V software (Higgins & Sharp 1988). Repeated sequences were identified and located with the EQUICKTANDEM program from the EMBOSS package (http://www.emboss.org). The nucleotide sequences data reported in this work have been submitted to GenBank under Accession nos AY500893–AY500899 and AY639017. (Users of these data are requested to refer to the present publication.)

In all guanaco, vicuna, llama and alpaca samples analysed, the complete CR was 1060 bp long (Fig. 1). In order to study the molecular organization of camelid CR the consensus sequence was also aligned to CR sequences of species representative of different mammalian orders (Gemmel *et al*. 1996; Sbisà *et al*. 1997). Thus, it was possible to identify a central domain, the conserved sequence blocks (CSB I–III) and the extended termination-associated sequence I as well as a repetitive motif on the left side of the CR (Fig. 1). On the 5′ end of the CR we found three CSB, CSB I–III. Within the central domain we detected three potential secondary structure regions, subsequences A–C (Fig. 1), homologous to those described by Gemmel *et al*. (1996) in

Table 1 Polymorphic sites within the hypervariable segment at the 5′ end of the mitochondrial DNA control region from 20 South American camelids

	1	2	2	2	3	3	4	4	4	5	
	6	2	5	9	2	4	0	2	5	0	
Animal	5	4	9	1	8	0	1	1	4	2	Haplotype
Gua1	A	A	A	G	T	A	T	T	A	T	H1
Gua2	.	.	.	.	.	.	.	.	.	.	H1
Lla1	.	.	.	.	.	.	.	.	.	.	H1
Lla2	.	.	.	.	.	.	.	.	.	.	H1
Lla3	.	.	.	.	.	.	.	.	.	.	H1
Vic1	.	.	.	.	.	.	.	.	.	.	H1
Gua5	G	.	.	.	.	G	.	.	.	.	H2
Lla4	G	T	.	.	A	G	.	.	.	.	H3
Alp2	G	T	.	.	A	G	.	.	.	.	H3
Gua3	G	T	.	.	A	G	.	.	.	.	H3
Gua4	G	T	.	.	A	G	.	.	.	.	H3
Vic2	G	T	.	.	A	G	.	.	.	.	H3
Alp1	G	T	.	T	A	G	.	.	.	.	H4
Lla5	G	T	.	.	A	G	C	.	.	.	H5
Vic4	G	T	.	T	A	G	.	.	G	.	H6
Alp5	G	T	.	T	A	G	.	.	G	.	H6
Vic5	G	T	.	T	A	G	.	.	G	.	H6
Alp3	G	T	G	.	A	G	.	C	.	C	H7
Alp4	G	T	G	.	A	G	.	C	.	C	H7
Vic3	G	T	G	.	A	G	.	C	.	C	H7

In the upper part of the table the vertical numbers indicate the positions of nucleotides relative to the reference guanaco 1 sequence. Below are shown the nucleotide changes and the resulting haplotypes. Dots represent identity with the reference sequence.

Ornithorhynchus anatinus. At the 3′ end of the CR we identified a 52-bp sequence over 95% similar to the extended termination-associated sequence conserved domain described by Sbisà *et al*. (1997) in 10 different mammalian orders (data not shown). All SAC examined here share a 78 nucleotide repetitive motif located between the CSB I and II (Fig. 1). This element comprises three imperfect repeats of 26 bases, each differing in sequence but not length when compared with each other and among individuals. These differences, however, are smaller than those present in one alpaca reported by Ursing *et al*. (2000) who found a 30-bp motif repeated six times. The consensus unit of our camelids is TAC**R**T**R**TACACGTGTGGGTGTATGT**R**. The analysis of the repeated unit showed that this sequence might have self-complementary ends (TACAT/ATGTA) able to form hairpin loops. Starting at nucleotide 165, all camelid individuals reported here show a segment of 337 bp which concentrates 10 polymorphic sites (seven transitions and three transversions). Based on the sequence of this region and regarding the haplotype of a wild guanaco as reference (H1), we defined seven haplotypes which differ from the reference by 0–2.07% (Table 1). Therefore, this hypervariable segment constitutes a suitable genetic marker to be applied in genetic variability searching among camelids. In addition, the typing of individuals and/or populations with this marker may be useful for designing breeding strategies and sustainable management conditions in these species.

Acknowledgements

The authors are grateful to Dr D. Delamo (CENPAT, Puerto Madryn), Dr J. von Thüngen (INTA-Bariloche), Dr B. Carpinetti (Estación ECAS, Buenos Aires), Dr V. Roig (Mendoza Zoo) and Eng. H. Lamas (INBIAL, Jujuy) for providing the Argentinean camelid blood samples used in this investigation. They also thank the National Node EMBnet of Argentina for allowing the use of the EMBOSS package software and B. Tosti for the English revision. The help and advice of Prof. Ulfur Arnason (University of Lund, Sweden) are much appreciated. This study was supported by grants from ANPCyT, CONICET and CICPBA of Argentina.

References

Bustamante A, Zambelli A, De Lamo D, von Thüngen J, Vidal-Rioja L (2002) Genetic variability of guanaco and llama populations in Argentina. *Small Ruminant Research*, **44**, 97–101.

Bustamante A, Maté ML, Zambelli A, Vidal-Rioja L (2003) Isolation and characterization of 10 polymorphic dinucleotide microsatellite markers for llama and guanaco. *Molecular Ecology Notes*, **3**, 68–69.

Gemmel NJ, Western PS, Watson JM, Marshal Graves JA (1996) Evolution of the mammaliam mitochondrial control region — comparisons of control region sequences between monotreme and therian mammals. *Molecular Biology and Evolution*, **13** (6), 798–808.

Higgins D, Sharp PM (1988) CLUSTAL V: a package for performing multiple sequence alignment on a microcomputer. *Gene*, **73**, 475–493.

Kadwell M, Fernández M, Stanley H *et al.* (2001) Genetic analysis reveals the wild ancestors of the llama and the alpaca. *Proceedings of the Royal Society of London, B*, **268**, 2575–2584.

Nagata J, Masuda R, Tamate H *et al.* (1999) Two genetically distinct lineages of the Sika deer, *Cervus nippon*, in Japanese islands: comparison of mitochondrial D-loop region sequences. *Molecular Phylogenetics and Evolution*, **13** (3), 511–519.

Nyakaana S, Arctander P, Siegismund HR (2002) Population structure of the African savannah elephant inferred from mitochondrial control region sequences and nuclear microsatellite loci. *Heredity*, **89**, 90–98.

Sbisà E, Tanzariello F, Reyes A, Pesole G, Saccone C (1997) Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. *Gene*, **205**, 125–140.

Torres H (1995) Criterios de UICN 1992a, b, para el manejo sustentable de especies silvestres. In: *Técnicas para el Manejo del Guanaco* (ed. Puig S), Chap. 3, pp. 51–53. UICN Quito, Ecuador.

Ursing BM, Slack KE, Arnason U (2000) Subordinal artiodactyls relationships in the light of phylogenetic analysis of 12 mitochondrial protein-coding genes. *Zoologica Scripta*, **29**, 83–88.