J. health med. sci.,
4(4):275-280, 2018.

Identification of Llama (*Lama glama*) Class I Mayor Histocompatibility Complex genes and the phylogeny of the Cetartiodactyla Clade

Identificación de Genes de Clase I del Complejo Mayor de Histocompatibilidad en la Llama (*Lama glama*) y la Filogenia de Clado Cetartiodactyla

Natalia Vicentela[1]; Daniel Moraga[2] & Felipe Figueroa[2]

VICENTELA, N.; MORAGA, D. & FIGUEROA, F. Identification of llama (*Lama glama*) class I mayor histocompatibility complex genes and the phylogeny of the Cetartiodactyla clade. *J. health med. sci., 4(4)*:275-280, 2018.

ABSTRACT: The evolutionary history of the American camelids was disrupted by dramatic events of artificial bottlenecks, which resulted in a decline of the genetic variability of the species. Around 80% of the number of individuals in populations of wild and domestic camelids, became extinct during the first 100 years of European settlement and this decrease persisted without interruption until part of the XX century. The consequences of these extinctions in the gene pool of the current camelid populations are unknown. Our goal is to investigate the extent of the genetic variability in camelids through the study of the polymorphism in genes of the Major Histocompatibility Complex (Mhc). The Mhc genes are highly polymorphic, with a large number of alleles and large genetic distances between the alleles. Therefore, the polymorphism Mhc genes have been a tool used to investigate the extent of the genetic variability in several mammalian species as well in teleost fishes. As the first step of our efforts, in this study, we report the identification and gene sequencing of five class I llama alleles at the Major Histocompatibility Complex. The sequences obtained have allowed us to inquire into the phylogenetic relationships between families and species *Cetartiodactyla* and obtain additional information about the mode of evolution of the Mhc complex.

KEY WORDS: camelids, Mhc, class I, evolution.

INTRODUCTION

Of the four South American camelids species, llamas and alpacas (*Lama glama* and *Vicugna pacos*) are domestic, and guanacos (*Lama guanicoe*) and vicuñas (*Vicugna vicugna*) are wild. The species belong to the clade *Cetartiodactyla*, order *Artiodactyla*, suborder *Tylopoda*, family *Camelidae*, and tribe *Lamini*. Archaeological evidence suggests that the domestication process of the llama begun in several places of the Andes between 5.000 and 3.800 years before present (BP). In northeast Argentina and northern Chile domestication would have started at elevations of 3.200 meters above the sea and upper levels and in the Peruvian Andes at elevations over 4.000 meters around 4.000 years ago (Cartajena *et al*., 2007).

The recent evolutionary history of South American camelids has shown dramatic events of artificial genetic bottlenecks that have threatened to the species with evident extinction risk. At the end of the first century of the European conquest, the populations of South American camelids were drastically reduced (Wheeler, 2012). It is estimated that during pre-Columbian times, around 40 million guanacos and 2 million vicuñas inhabited the South American territories (Raedeke, 1979; Brack, 1980). In the second half of the 20th century, the population of guanacos and vicuñas had diminished to critical levels (Koford, 1957; Jungius, 1971). The current guanaco and vicuna populations are not higher than 600.000 and 400.000 animals respectively (Wheeler). A similar situation

[1] Thesis. Biochemistry, Faculty of Health Sciences, Universidad de Antofagasta, Antofagasta, Chile.
[2] Medical School, Universidad de Tarapacá, Arica, Chile.

happened with the domestic camelids and one century after the European settling, about 90 % of llamas and alpacas had disappeared. The genetic impact of these bottlenecks on the genetic diversity of the South American camelids is largely unknown. To get an answer to this and other questions, we have initiated a project to evaluate the extent of genetic diversity throughout the study of polymorphism at the Major Histocompatibility Complex (Mhc) genes.

The Mhc genes code for glycoproteins that bind short peptides derived from self and non-self molecules and display them on the surface of antigen presenting cells (Klein & Sato, 2000). Recognition of Mhc-nonself peptide complexes by T lymphocytes initiates the specific form of the immune response. Mhc genes fall into two classes, I and II, whose products present intra and extracellularly derived peptides, respectively. Mhc class I molecules are heterodimers formed by α and β chains, but only the α chain is part of the Mhc. The α chain has three extracellular domains, known as α1, α2, and α3; the Antigen Recognition Sites (ARS) includes amino acids present in the α1 and α2 domains. The α1 and α2 domains are encoded by the exons 2 and 3 of the class I genes and both show a high degree of polymorphism which is characterized by a large number of alleles and large genetic distances between alleles (Klein & Sato). The Mhc offers new possibilities for the study of camelids evolution because its extensive polymorphism (Klein & Figueroa, 1986; Klein & Sato), evolves trans-specifically (Figueroa *et al*., 1988), making it possible to relate taxa through their sharing of old allelic lineages, and to differentiate them through the segregation of ancestral polymorphisms.

Here we report the initial step of our effort with the cloning and sequencing of exon 2 of class I Mhc genes of the llama.

MATERIAL AND METHOD

Sample and preparation

Total genomic DNA was prepared from a piece of ethanol-preserved lung tissue according to a standard phenol/chloroform extraction, the DNA was precipitated in 100 % ethanol, the precipitate washed twice with 70 % ethanol and finally resuspended in 500 µl of TE (10 mM Tris-HCl pH8.0 1 mM EDTA (Sambrook & Russell, 2001).

First design and amplification

To design new specific Mhc class I primers, we first perform a BLAST search against the GenBank non-redundant artiodactyl nucleotide database for class I Mhc gene sequences and using a bovine sequence as a query. Then, a multiple alignments of the nucleotide sequences was created with the help of the CLUSTALX program (Thompson *et al*., 1997), conserve regions at both ends of the exon 2 of the gene were identified and degenerate oligonucleotides were designed. One of the primer combinations, 5´- GC(C/T)CCCACTC(G/C/T)(A/C)TGA(A/G)(G/C)TATGT-3' (sense) and 5'-CCGGCCTCGCTCTGGTTGTAGTA-3'(antisense) was used in the amplification of the llama DNA by the Polymerase Chain Reaction (PCR).

Standard PCR amplifications (Saiki *et al*., 1988) were performed in the PTC-100 Programmable Thermal Controller (MJR, Bio-Rad, CA, USA). One microliter containing 100 ng of llama genomic DNA was added to a reaction mixture consisting of 1xPCR buffer (50 mM KCl, 1.5 mM MgCl2, 10 mM Tris-HCl, pH 8.3, 0.0001 % gelatin), 0.2 mM of each of the four deoxyribonucleoside triphosphates (Promega WI, USA), 1 mM of each of the sense and antisense primers, and 1 unit of AmpliTaq DNA polymerase (Promega WI, USA). The amplification consisted of DNA denaturation for 5 min at 94°C, followed by 35 cycles, each cycle consisting of 15s denaturation at 94 °C, 30s annealing at 45 °C and 2 min extension at 72 °C. The reactions were completed by a final primer extension for 5 min at 72 °C.

DNA cloning and sequencing

PCR products were separated by electrophoresis in a low-melting point agarose gel (Axygen Biosciences, USA) and the bands were stained with ethidium bromide, excised, and eluted with the aid of the QIAEX II Gel Extraction Kit (Qiagen Hilden, Germany). The eluted DNA was ligated to the pGEM®-T vector with the help of the T Easy Vector System I Kit (Promega WI, USA) and used to transform JM109 (Promega WI, USA) competent cells. Plasmid DNA was prepared with the aid of the E.Z.N.A. plasmid DNA mini II Kit (Promega WI, USA), and resuspended in water at a concentration of 1 µg/µl and sequenced. Sequencing reactions were performed at the Ecology Unit, Pontificia Universidad Católica de Chile.

Cluster map comparisons

We first perform a PSI-BLAST search against the NCBI GenBank non-redundant Cetartiodactyla class I Mhc protein database using a bovine query. We obtained a total of 1043 *Cetartiodactyla* class I exon 2 sequences. From the order *Artiodactyla*, family *Bovidae*: 327 sequences from *Bos Taurus* (Bota), 4 *Bos mutus* (Bomu), 9 *Bos indicus* (Boin), 7 *Bubalus bubalis* (Bubu), 34 *Bison bison* (Bibi), 6 *Bison bonasus* (Bibo), 3 *Capra hircus* (Cahi) and 55 *Ovis aries* (Ovar); family *Cervidae*: 128 sequences from *Cervus nippon* (Ceni) and 8 *Elaphurus davidianus* (Elda); family *Suidae*: 449 sequences from *Sus scrofa* (Susc); From order *Cetacea*, family *Delphinidae*: 4 sequences from *Orcinus orca* (Oror), 7 *Lipotes vexillifer* (Live) and 1 *Tursiops truncates* (Tutr); family *Balaenopteridae*: 1 *Balaenoptera acutorostrata scammoni* (Baac). To cluster the sequences into different groups, we used the CLANS program, a Java application for visualizing protein families based on pairwise similarity (Frickey & Lupas, 2004).

Dendrogram construction

Evolutionary relationships were evaluated by genetic distance methods using the neighbor-joining algorithm for phylogenetic tree construction (Saitou & Nei, 1987). Distances were measured between sequences using the proportion of different amino acids in the a1 domain.

RESULTS

Five different partial exon 2 class I sequences were obtained from the amplified DNA, which together with one *Bos taurus* exon 2 sequence are presented in Fig. 1A. Sequences covered 226 pairs of bases, from the third position of the codon 8 to codon position 83. The sequences were designated as Lagl from *Lama glama* according to the nomenclature of the Mhc genes (Klein *et al*., 1990) and followed by digits 01, 02, 03 and 04. Two 04 sequences were obtained which differ from each other by a single non-synonymous substitution at second position of codon 32. This single substitution was confirmed by independent PCR and we assume that they represent alleles at a single locus and therefore we designated them as 04.01 and 04.02. No stop codons, insertions or deletions are present in any of the partial exon 2 sequences. The llama 01, 02, 03, 04.01 and 04.02 sequences are separated from each other by large genetic distances (Table I). Thus, genetic distances vary from a minimum of 0.004 between sequences *Lagl*_04.01 and *Lagl*_04.02, up to a maximum value of 0.245 between sequences *Lagl*_01 and *Lagl*_04.01. The average genetic distances of the sequence *Lagl*_01 with the remaining is 0.206 and for the sequences, *Lagl*_02, *Lagl*_03, *Lagl*_04.01, and *Lagl*_04.02 compared with the rest are 0.163, 0.199, 0.165 and 0.160, respectively.

The alignment of the deduced amino acid sequences shows that the majority of the differences between the sequences are concentrated in the putative Antigen binding sites (ABS). Thus, 7 out 26 ABS show more than 1 substitution at these sites (Fig. 1B). In the alignment, none of the sequences show traits that preclude the exons to be part of classical Mhc class I genes. In addition, the sequence of exon 2 *Lagl*_01 is identical to the corresponding exon 2 of an alpaca sequence. The latter is a full sequence derived from a genome sequencing *Vicugna pacos* project of the University of Washington and shows all the features of a class I Mhc gene classic (NCBI GenBank acc. number 560986547).

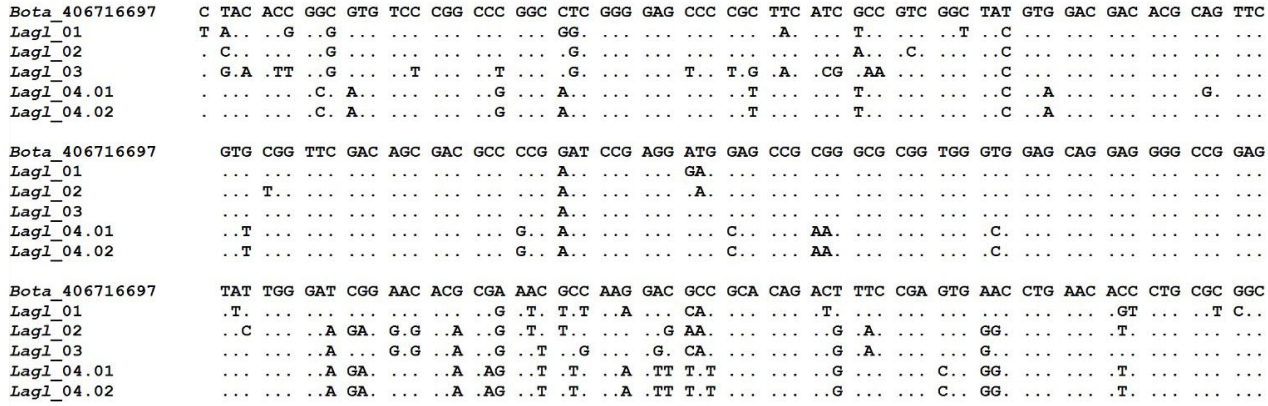

Fig. 1. A. Llama nucleotide exon 2 Mhc class I sequences. The Bota_406716697 sequence corresponds to a cattle sequence deposited in the NCBI database. Identity to the cattle sequence is indicated by dots.

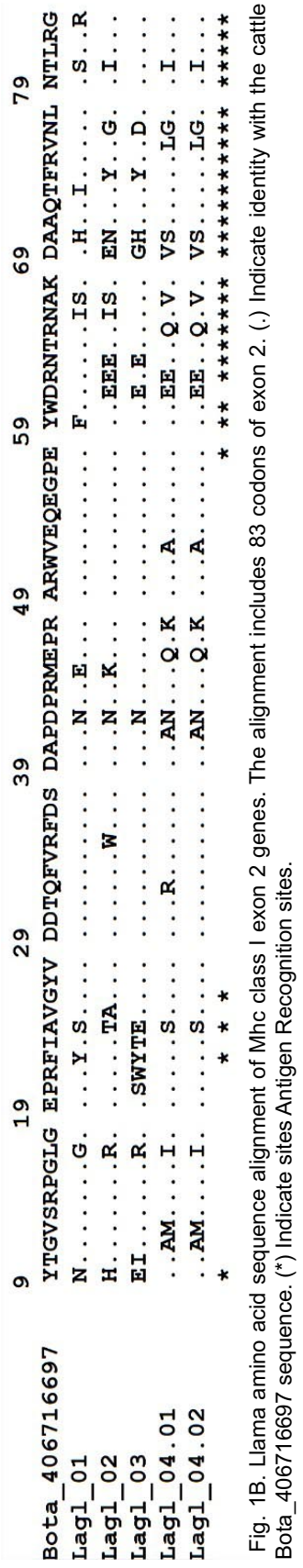

Fig. 1B. Llama amino acid sequence alignment of Mhc class I exon 2 genes. The alignment includes 83 codons of exon 2. (.) Indicate identity with the cattle Bota_406716697 sequence. (*) Indicate sites Antigen Recognition sites.

Table I. The genetic distances of the 5 Lagl nucleotide sequences.

	Lagl_01	Lagl_02	Lagl_03	Lagl_04.01	Lagl_04.02
Lagl_01	****				
Lagl_02	0.152	****			
Lagl_03	0.191	0.147	****		
Lagl_04.01	0.245	0.179	0.232	****	
Lagl_04.02	0.239	0.174	0.226	0.004	****

DISCUSSION

The exon 2 of Mhc class I sequences *Lagl*_01 through *Lagl*_04 described in this study, were obtained after the PCR amplification of DNA isolated from a single llama. The sequences are separated from each other by large genetic distances and most likely all four represent alleles at different class I loci (Table I). To get an insight on the phylogenetic relationships of the llama sequences among themselves and with those from other Cetartiodactyla species, we first perform a Psi-BLAST search of the NCBI GenBank protein database, using a cattle class I exon 2 sequence as a query. The search identified 1043 *Cetartiodactyla* sequences, which were then retrieved from the GenBank and along with the llama sequences analyzed by the phylogenetic CLANS program.

A graph layout of all *Cetartiodactyla* sequences generated by the CLANS application is shown in Fig. 2. The Mhc exon 2 class I sequences, were clustered into 2 main groups, designated in the graph as cluster A and B. The A Cluster enclosed a total of 550 sequences which belong to the Bovidae, Cervidae and Cetacea families with 404, 133 and 13 sequences respectively. All 402 sequences included in Cluster B belong to the family Suidae (domestic pigs). The cluster A is further divided into three subgroups. In the Bovidae A-1 subgroup, 228 sequences were originated from the cattle (Bota), 5 from zebu (Boin), 4 from water buffalo (Bubu), 27 from American bison (Bibi), 5 from European bison (Bibo), 2 from Yak (Bomu), 2 from goat (Cahi) and 49 from sheep (Ovar). In the Cervidae A-1 subgroup 6 sequences were derived from Pere David´s deer (Elda) and 78 from syka deer (Ceni). In the Cetacea A-1 subgroup 4 sequences were originated from 4 killer whales (Oror) 1 from northern minke whale (Baac), 7 from Chinese river dolphins (Live) and 1 from bottlenosed dolphin (Tutr). The *Bovidae* A-2 subgroup enclosed 59 Bota, 1 Bubu, 4 Bibi, 1 Bibo, and 3 Boin sequences, and 12 Ceni sequences are enclosed the *Cervidae* A-2 subgroup. The A-3 subgroup included 6 Bota, 2 Bomu, 2 Elda, 1 Ovar and 1 Cahi sequences (*Bovidae*) and 34 Ceni sequences (*Cervidae*). The subgroups B-1 and B-2 enclosed 390 and 12 pig sequences respectively. Besides the main clusters, the graph layout in Fig. 2 shows 8 minor clusters and several singletons. Four of the minor clusters correspond to pig and 4 to cattle sequences.

In the graph layout of Fig. 2, all but one of the llama sequences are unconnected with the main or minor *Cetartiodactyla* clusters. The exception is the sequence *Lagl*_02 which appears loosely related to the A-1 cluster. All Llama sequences except *Lagl*_04.01 and 04.02 are unrelated to each other which is in good agreement with the long genetic distances that separate them. The exon 2 llama *Lagl*_01 and alpaca *Vipa*_01 sequences appear as a singleton (their nucleotide sequences are identical) and the *Lagl*_02 appear closely related to

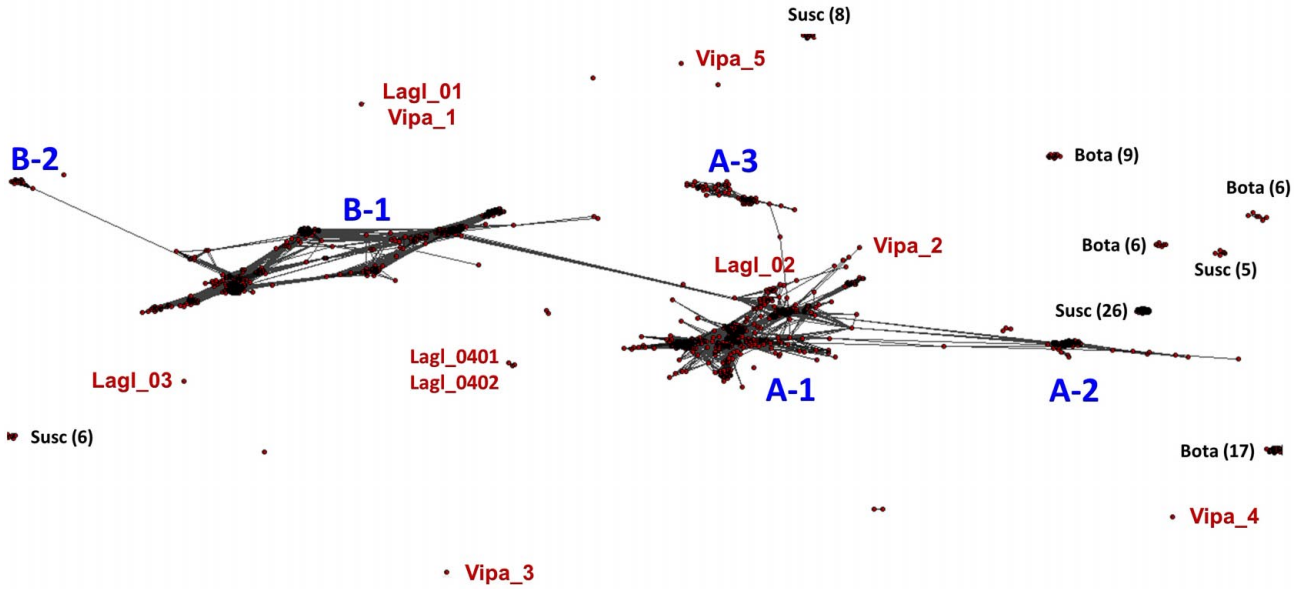

Fig. 2. Graph layout of the five exon 2 sequences from the llama together with 1049 *Cetartiodactyla* homologous sequences extracted from the NCBI database, obtained after the analysis by the program CLANS.

the Alpaca *Vipa*_2. The other 3 alpaca sequences are unrelated to each other and to the other llama sequences. Thus, 6 out of 10 camelids exon 2 available sequences, appear to be unrelated to each other, they are separated by large genetic distances and therefore we could conclude that they belong to different Mhc class I loci. The graph in Fig. 2 suggests the presence in the llamas and alpacas of four and five loci respectively.

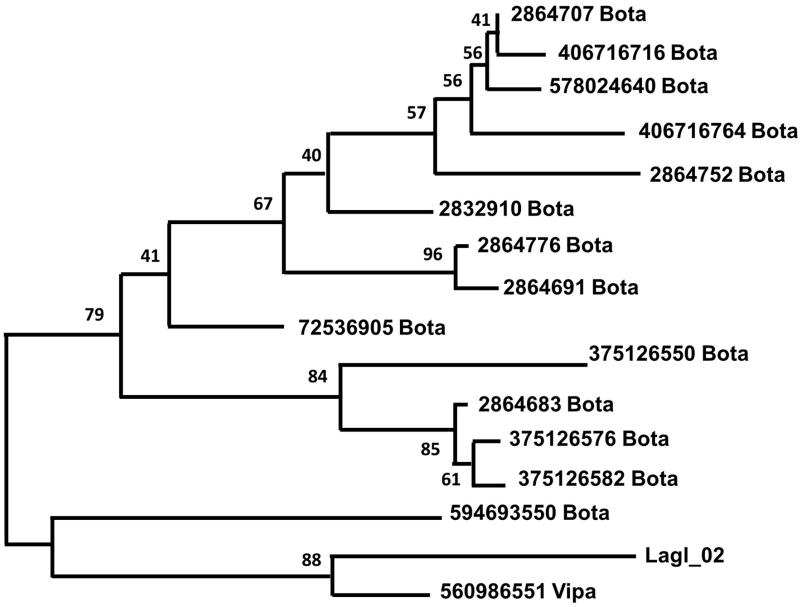

Fig. 3. Amino acid phylogenetic tree of llama Lagl_02 sequence and its closest cattle relatives on the Cluster A-1. Numbers stand for Bootstraps values after 500 replications.

The number of Mhc class I and class II varies from species to species. A physical map of the class I region, has revealed the presence of 7 loci of the domestic pig (Lunney, *et al*. 2009) and mapping studies have suggested that there are at least 6 loci in the cattle (Babiuk *et al*. 2007). The graph shown in Fig. 2, no taking in account singletons suggests the presence of 6 loci in both the pig and the cattle. However, cattle sequences in the A-1 cluster of Fig. 2 may correspond to alleles at different class I loci. A dendrogram of the cattle sequences from the A-1 cluster together with the *Lagl*_02 and *Vipa*_2 sequences revealed the presence in the tree of three main branches. If each of them corresponds to a different lineage or loci await for to be established (data not shown). One of these branches included *Lagl*_01 and *Vipa*_2 sequences. A phylogenetic tree of the sequences of this branch revealed that both camelid sequences are separated from each other and with the cattle sequences by large genetic distance (Fig. 3). The variable number of loci on different species of vertebrates would have been created through a process of duplications, deletions and mutations referred to as the Accordion or the Birth-and-Death models of evolution (Nei & Hughes, 1992; Klein *et al*., 1993).

Finally, the graph layout appears to confirm the phylogenetic relationships of camelids with the other members of the Cetartiodactyla clade (Zhou *et al*., 2011). In the current phylogeny of Cetartiodactyla, *Bovidea, Cervidea,* and *Cetacea* form a monophyletic clade which is a sister group of the Suidae family members and followed by camelids at the very base of the phylogenetic tree.

VICENTELA, N.; MORAGA, D. & FIGUEROA, F. Identificación de genes de clase I del complejo mayor de histocompatibilidad en la Llama (*Lama glama*) y la filogenia de Clado Cetartiodactyla*. J. health med. sci., 4(4)*: 275-280, 2018.

RESUMEN: La historia evolutiva de los camélidos americanos fue alterada por dramáticos eventos de cuellos de botella artificiales, que dieron lugar a una disminución en la variabilidad genética de las especies. Alrededor de un 80 % de las poblaciones de camélidos, tanto silvestres como domésticos, se extinguieron durante los primeros 100 años de la colonización europea y esta disminución persistió sin interrupción hasta el transcurso de la segunda mitad del siglo XX. Las consecuencias de esas extinciones en el patrimonio genético de las poblaciones actuales de camélidos son desconocidas. Nuestro objetivo es investigar la amplitud de la variabilidad genética en los camélidos a través del estudio del polimorfismo en los genes del complejo mayor de histocompatibilidad (Mhc). Los genes Mhc se caracterizan por su elevado polimorfismo, que se manifiesta en un gran número de alelos y en grandes distancias genéticas entre los alelos. Es por ello, que el polimorfismo de los genes del complejo Mhc ha sido una herramienta utilizada para investigar la extensión de la variabilidad genética en diversas especies de mamíferos y teleósteos. Como un primer paso, en el presente estudio reportamos la identificación y secuenciación génica de cinco alelos de clase I del Complejo Mayor de Histocompatibilidad de la llama. Las secuencias obtenidas nos han permitido indagar en las relaciones filogenéticas entre familias y especies *Cetartiodactyla* y obtener información adicional sobre el modo de evolución del complejo Mhc.

PALABRAS CLAVE: camélidos, Mhc, clase I, evolución.

REFERENCES

Babiuk, S.; Horseman, B.; Zhang, C.; Bickis, M.; Kusalik, A.; Schook, L. B.; Abrahamsen, M. S. & Pontarollo, R. BoLA class allele diversity and polymorphism in a herd of cattle. *Immunogenetics, 59(2)*:167-76, 2007.

Brack, A. *Conservación de la vicuña en el Perú*. Ministerio de Agricultura y Alimentación, Lima, 1980. 33 pp.

Cartajena, I.; Núñez, L. & Grosjean, M. Camelid domestication on the western slope of the Puna de Atacama, northern Chile. *Anthropozoologica, 42*:155-73, 2007.

Figueroa, F.; Gunther, E. & Klein, J. MHC polymorphism pre-dating speciation. *Nature, 335*:265-7, 1988.

Frickey, T. & Lupas, A. CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics, 20(18)*: 3702-04, 2004.

Jungius, H. The Vicuna in Bolivia: The status of an endangered species and recommendations for its conservation. *Mamm. Biol., 36*:129-46, 1971.

Klein, J. & Figueroa, F. Evolution of the major histocompatibility complex. *Crit. Rev. Immunol., 6(4)*:295-386, 1986.

Klein, J.; Bontrop, R. E.; Dawkins, R. L.; Erlich, H. A.; Gyllensten, U. B.; Heise, E. R.; Jones, P. P.; Parham, P.; Wakeland, E. K. & Watkins. D. I. Nomenclature for the major histocompatibility complexes of different species: a proposal. *Immunogenetics, 31(4)*:217-9, 1990.

Klein, J.; Ono, H.; Klein, D. & O'hUigin, C. *The accordion model of MHC evolution.* Progress in Immunology. VIII, London, Academic Press, 1993. 137-143 pp.

Klein, J. & Sato, A. The HLA System. *N. Engl. J. Med., 343(10)*:702-9, 2000.

Koford, C. B. The vicuña and the puna. *Ecol. Monogr., 27(2)*:153-219, 1957.

Lunney, J.K; Ho, C-S.; Wysocki, M. & Smith, D. M. Molecular genetics of the swine major histocompatibility complex, the SLA complex. *Dev. Comp. Immunol., 33(3)*:362-374, 2009.

Nei, M. & Hughes, A. L. *Balanced polymorphism and evolution by the birth-and-death process in the MHC loci*. Pp. 27-38. In Tsuji, K.; Aizawa, M. & Sasazuki, T. (eds). 11th Histocompatibility Workshop & Conference, Oxford, Oxford University Press, 1992.

Raedeke, K. J. *Population dynamics and socio ecology of the guanaco (Lama guanicoe) of Magallanes, Chile*. University Microfilms International, Michigan, Ann Arbor, 1979. 404 pp.

Saiki, R. K.; Gelfand D. H.; Stoffel, S.; Scharf, S. J.; Higuchi, R.; Horn, G. T.; Mullis, K. B. & Erlich, H. A. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science, 239(4839)*:487-91, 1988.

Saitou, N. & Nei, M. The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees. *Mol. Biol. Evol., 4(4)*:406-25, 1987.

Sambrook, J. & Russell, D. In: *Molecular Cloning: A Laboratory Manual.* 3rd ed. New York, Cold Spring Harbor Laboratory Press, 2001.

Thompson, J. D.; Gibson, T. J.; Plewniak, F.; Jeanmougin, F. & Higgins, D. G. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res., 25(24)*:4876-82, 1997.

Wheeler, J. C. South American camelids – past, present and future. *J. Camel. Sci., 5*:1-24, 2012.

Zhou, X.; Xu, S.; Yang, Y.; Zhou, K. & Yang, G. Phylogenomic analyses and improved resolution of Cetartiodactyla. *Mol. Phylogenet. Evol., 61(2)*:255-64, 2011.

Dirección para correspondencia:
Felipe Figueroa
Medical School
Universidad de Tarapacá
Arica
CHILE

Email: ffigueroauta@gmail.com

Recibido : 17-08-2018
Aceptado: 26-09-2018