沙漠研究 **26**-4, 223-226 (2017)
Journal of Arid Land Studies
http://dx.doi.org/10.14976/jals.26.4_223

DNA Analyses of Camels

Naruya SAITOU*[1] and Shayire SHOKAT[1]

Abstract: Family Camelidae includes Guanaco and Vicuna in South America and two-humped camels (*Camelus bactrians* and *C. ferus*) and one-humped camels (*C. dromedarius*) are distributed from Eurasia to Northern Africa. We reviewed studies on mitochondrial and nuclear DNA of camels. We collected 37 complete mitochondrial DNA sequences of Camelid including those of now extinct *Camelops* which distributed in North America from DDBJ/EMBL/GenBank International Nucleotide Sequence Database. Neighbor-joining trees were constructed for these sequences, and evolution of family Camelidae and genus *Camelus* are discussed with special reference to demographic changes of *C. bactrians* and *C. dromedarius*.

Key Words: Camel, DNA, evolution, Kazakhstan.

1. Introduction

Family Camelidae belongs to Artiodactyla (Tree of Life web project; http://tolweb.org/tree/). Present-day Camelidae is divided into humped type (genus *Camelus*) and non-humped types (genus *Lama* and genus *Vicugna*), and the former is distributed in Old World, while the latter is distributed in New World. Humped Camelidae are generally called "camels" in common English. *Camelus dromedarius* is one-humped camel, and is distributed from North Africa to Central Asia. This species consists of ~90% of all camels in the world. *Camelus bactrianus* is two-humped camel, and is distributed from Central Asia to East Asia. There is another two-humped camel species in Mongolia and some part of China; *Camelus ferus*. While *Camelus dromedarius* and *Camelus bactrianus* are all domesticated, *Camelus ferus* remains wild. There are four kinds of non-humped Camelidae in South America; alpaca, guanaco, llama, and, vicuna. Alpaca and llama are domesticated animals, while guanaco and vicuna are wild. Stanley *et al.* (1994) determined cytochrome b gene sequences of mitochondrial DNA (mtDNA) for those four kinds of camelids in South America, and showed that alpaca mtDNA is clustered with that of vicuna, and llama mtDNA with that of guanaco. DNA study of Kadwell *et al.* (2001) comparing mtDNA cytochrome b gene sequences and four nuclear microsatellite loci confirmed this, though bidirectional hybridization between these two clusters were also observed.

The first nucleotide sequence of Camelidae was reported by Irwin *et al.* (1991). They compared ~1100 bp mtDNA cytochrome b gene of 20 mammalian species including *Camelus dromedarius*. Camel and dolphin were clustered in the maximum parsimony tree and the neighbor-joining tree when mostly nonsynonymous changes were considered, though its statistical support was low. Later molecular studies supported phylogenetic position of Cetacea (dolphins and whales) within Artiodactyla (*e.g.*, Shimamura *et al.* 1997). Price *et al.* (2005) presented a Cetartiodactyla phylogeny in which Camelidae was basal.

The complete mtDNA genome sequence of alpaca was reported by Ursing *et al.* (2000). After that, numerous complete mtDNA sequences were determined for Camelidae. We will produce phylogenetic trees from these sequences, and will discuss the molecular evolution of family Camelidae. Nuclear genome sequences for three species of Camelidae are now available, so we will also discuss their features and possible scenario of the camel domestication process.

2. Materials and Methods

We used complete mtDNA sequences of *Camelus dromedarius* (DDBJ/EMBL/GenBank International Nucleotide Sequence Database accession number JN632608) reported by Hassanin *et al.* (2012) as query, and conducted BLAST (Altschul *et al.* 1990) homology search using DDBJ web system (http://ddbj.nig.ac.jp/blast/blastn?lang=ja). We retrieved 37 mtDNA complete genome sequences from the camelid, and they are listed in **Table 1**. They include sequences from 4 *Lama*, 2 *Vicugna*, 3 *Camelops*, 3 *Camelus ferus*, 10 *Camelus bactrianus*, and 15 *Camelus dromedarius* individuals. These 37 sequences were published in Ursing *et al.* (2000), Arnason *et al.* (2004), Cui *et al.* (2007), Ji *et al.* (2009), Di Rocco *et al.* (2010), Hassanin *et al.* (2012), Heintzman *et al.* (2015), as well as unpublished studies (or published only in DDBJ/EMBL/GenBank International Nucleotide Sequence Database) by Huang *et al.*, Mohandesan *et al.*, Tahmoorespur *et al.*, Yasue *et al.*, and Zhang *et al* (see Table 1). These sequences were multiply aligned by using

* Corresponding Author: saitounr@nig.ac.jp
1111 Yata, Mishima, 411-8540, Japan Tel: ++81-55-981-6790/-6789 Fax: +81-55-981-6790/-6789
1) Division of Population Genetics, National Institute of Genetics

(Received, January 1st, 2017; Accepted, January 5th, 2017)

Table 1. Complete mitochondrial DNA genome sequences used in this study.

Sequence ID	INSD Accession Number	Organism description#	Ref.*
Lama_1	Y19184	Lama pacos	1
Lama_2	AJ566364	Lama pacos	2
Lama_3	AP003426	Lama glama	3
Lama_4	EU681954	Lama guanicoe	4
Vicugna_1	KU168760	Vicugna pacos huacaya	5
Vicugna_2	FJ456892	Vicugna vicugna isolate 214v	4
Camelops_1	KR822420	Camelops cf. hesternus	6
Camelops_2	KR822421	Camelops cf. hesternus	6
Camelops_3	KR822422	Camelops cf. hesternus	6
C_ferus_1	EF212038	Camelus bactrianus ferus isolate W1	7
C_ferus_2	EF507800	Camelus ferus isolate W2	8
C_ferus_3	EF507801	Camelus ferus isolate W3	8
C_bact_1	AP003423	Camelus bactrianus	3
C_bact_2	EF212037	Camelus bactrianus isolate D1	7
C_bact_3	EF507798	Camelus bactrianus isolate D2	8
C_bact_4	EF507799	Camelus bactrianus isolate D3	8
C_bact_5	KX554925	Camelus bactrianus haplotype 1	9
C_bact_6	KX554926	Camelus bactrianus haplotype 2	9
C_bact_7	KX554927	Camelus bactrianus haplotype 3	9
C_bact_8	KX554928	Camelus bactrianus haplotype 4	9
C_bact_9	KX554929	Camelus bactrianus haplotype 5	9
C_bact_10	KX554930	Camelus bactrianus haplotype 6	9
C_drom_1	EU159113	Camelus dromedarius	10
C_drom_2	JN632608	Camelus dromedarius isolate Morocco	11
C_drom_3	KU605072	Camelus dromedarius isolate Drom439	12
C_drom_4	KU605073	Camelus dromedarius isolate Drom795	12
C_drom_5	KU605074	Camelus dromedarius isolate Drom796	12
C_drom_6	KU605075	Camelus dromedarius isolate Drom797	12
C_drom_7	KU605076	Camelus dromedarius isolate Drom801A	12
C_drom_8	KU605077	Camelus dromedarius isolate Drom802	12
C_drom_9	KU605078	Camelus dromedarius isolate Drom806	12
C_drom_10	KU605079	Camelus dromedarius isolate Drom816	12
C_drom_11	KU605080	Camelus dromedarius isolate Drom820	12
C_drom_12	KX554931	Camelus dromedarius haplotype 1	13
C_drom_13	KX554932	Camelus dromedarius haplotype 2	13
C_drom_14	KX554933	Camelus dromedarius haplotype 3	13
C_drom_15	KX554934	Camelus dromedarius haplotype 4	13

Following definition line of database entry.

* 1: Ursing *et al.* (2000), 2: Arnason *et al.* (2004), 3: Yasue *et al.* (unpublished), 4: Di Rocco *et al.* (2010), 5: Zhang *et al.* (unpublished), 6: Heintzman *et al.* (2015), 7: Cui *et al.* (2007), 8: Ji *et al.* (2009), 9: Tahmoorespur *et al.* (unpublished), 10: Huang *et al.* (unpublished), 11: Hassanin *et al.* (2012), 12: Mohandesan *et al.* (unpublished), 13: Tahmoorespur *et al.* (unpublished).

MUSCLE (Edgar, 2004), pairwise number of nucleotide substitutions were estimated by using Tamura and Nei (1994), and phylogenetic trees were constructed using the neighbor-joining method (Saitou and Nei, 1987). These are implemented in MEGA6 (Tamura *et al.* 2013). Please refer to Saitou (2013) for basic concepts of multiple alignment and phylogenetic tree construction.

3. Results

We first constructed a neighbor-joining tree (not shown) for all the 37 mtDNA genome sequences listed in Table 1. Because most of *C. bactrianus* and *C. dromedarius* sequences were very similar, we chose C_bact_1 and C_drom_1 as representatives of *C. bactrianus* and *C. dromedarius* sequences, and constructed a neighbor-joining tree of these 14 sequences (**Fig. 1**). All internal branches showed 100% bootstrap probabilities. We can recognize three clusters in this tree; one-humped and two-humped camels (*C. ferus*, *C. bactrianus*, and *C. dromedarius*), extinct genus *Camelops*, and South American non-humped Camelidae (genus *Lama* and *Vicugna*). The tree topology is consistent with that of Figure 2 of Heinzman *et al.* (2015). In both trees, two *Vicugna* lineages were polyphyletic, which suggests existence of interspecific hybridization between Lama and *Vicugna*. Existence of hybridization is consistent with a previous study (Kadwel *et al.*, 2001). The number of nucleotide substitutions per site between South American non-humped Camelidae and Old World camels was ~0.16, while that number between one-humped and two-humped camels was ~0.07.

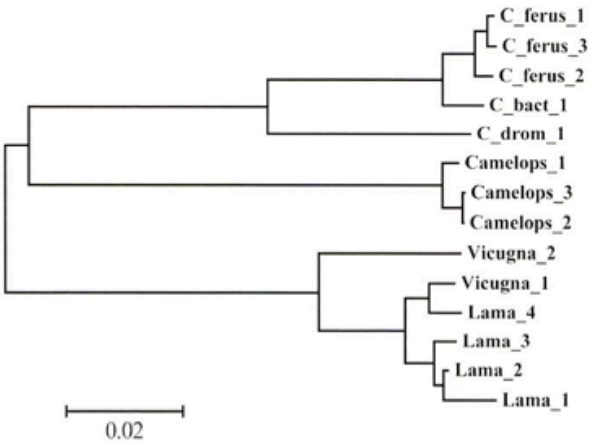

Fig. 1. A phylogenetic tree of 14 complete mtDNA sequences of camelid species. Sequence IDs are the same with those of Table 1. All branches were supported by 100% bootstrap probabilities.

We then focused on camels distributed in Old World. Complete mtDNA sequences of 3 *C. ferus*, 10 *C. bactrianus*, and 15 *C. dromedarius* were compared, and the resulting phylogenetic tree is shown in **Figure 2**. Bootstrap probabilities (in %) higher than 90% are shown in this tree. There are three lineages in *C. dromedarius*; C_drom 9, C_drom_12 - C_drom_15, and the remaining 10 *C. dromedarius* sequences. C_drom 9 is basal among the 15 *C. dromedarius* mtDNA sequences, and the number of nucleotide substitutions per site between C_drom 9 and the remaining 14 *C. dromedarius* was ~0.007. C_ferus_2 was basal among the 3 *C. ferus* mtDNA sequences, and the number of nucleotide substitutions per site between C_ferus_2 and the remaining two *C. ferus* sequences was ~0.006. Ten *C. bactrianus* mtDNA sequences were very close with each other, and the number of nucleotide substitutions per site between the basal C_bact_1 and the remaining 9 sequences was ~0.0009.

4. Discussion

Genome sequences of a two-humped camel (*Camelus bactrianus*) were determined by The Bactrian Camels Genome Sequencing and Analysis Consortium (2012) and by Wu *et al.* (2014), and those of the one-humped camel (*Camelus dromedarius*) were also determined (Wu *et al.*, 2014; Fitak *et al.*, 2016) as well as those of alpaca (*Vicugna pacos*) by Wu *et al.* (2014). The divergence time of *Camelus dromedarius* and *Camelus bactrianus* was estimated to be 4.4 (confidence interval: 1.9-7.2) million years ago, while Old World and New World Camelidae divergence time was estimated to be 16.3 (confidence interval: 9.4-25.3) million years ago (Wu *et al.*, 2014).

Demographic history of three Camelid species genomes was estimated by Wu *et al.* (2014) by using PSMC software (Li and Durbin, 2011). However, there was no description

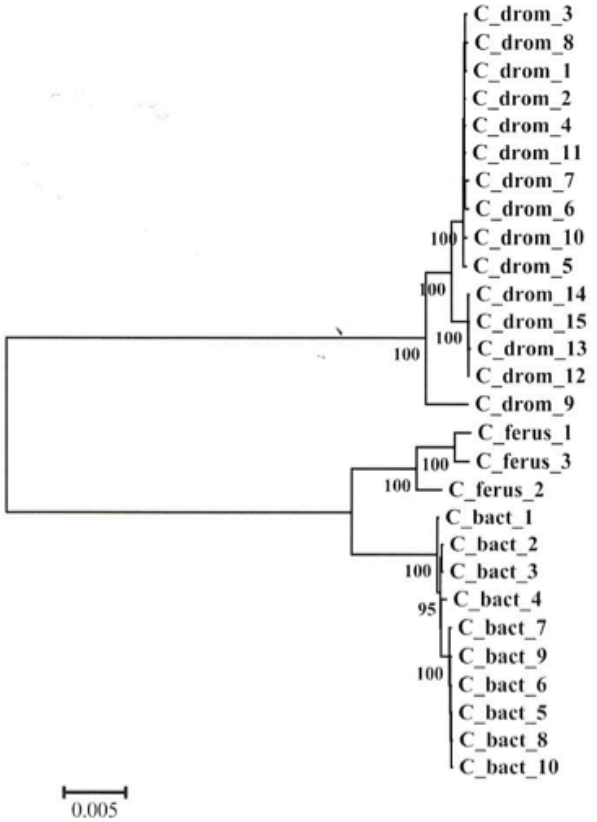

Fig. 2. A phylogenetic tree of 27 complete mtDNA sequences of one-hump and two-hump camels. Sequence IDs are the same with those of Table 1. Bootstrap probabilities higher than 90% are shown.

on generation time nor mutation rate in Wu *et al.* (2014). In contrast, Fitak *et al.* (2016) clearly mentioned generation time as 5 years and mutation rate as 2.5 x 10-8 /site/generation in their Figure 4 legend. Because of large discrepancies on time estimates of *Camelus dromedarius* demographic changes between these two studies (Wu *et al.*, 2014; Fitak *et al.*, 2016), we have to be careful about interpretation of demographic changes, although both studies estimated long-term decline of dromedary camel effective population size. This long-term decline was also the case for bactrian two-humped camel (Wu *et al.*, 2014). This suggests that, contrary to naive belief, camels did not "adapt" to their environments. Rather, both one-humped and two-humped camels were in danger of extinction just before they were domesticated. It should also be noted that Almathen *et al.* (2016) examined 1,083 *Camelus dromedarius* DNA (both mtDNA and nuclear microsatellite DNA polymorphisms), and the "restocking from the wild" hypothesis was supported.

We estimated the number of nucleotide substitution between Old World and New World Camelidae mtDNA sequences to be ~0.16. Therefore, the approximate substitution rate for Camelid mtDNA complete sequence becomes 5×10^{-9}/site/year (= 0.16/[2×16.3 million]). If we use

this rate, the mtDNA sequence divergence time between *Camelus dromedarius* and *Camelus bactrianus* is estimated to be 7 million years (= $0.07/[2\times5\times10^{-9}]$) under the assumption of constant evolutionary rate within Camelidae. This estimate is similar to that (8 million years) by Cui *et al.* (2007) who used only cyt *b* sequences, but is much larger than that (4.4 million years) obtained from genome data (Wu *et al.* 2014), though is barely within the confidence interval (1.9-7.2 million years). In any case, if we use the evolutionary rate of 5×10^{-9}/site/year, coalescent times for *C. dromedarius*, *Camelus bactrianus*, and *C. ferus* are estimated to be 700,000 years, 600,000 years, and 90,000 years, respectively. Under the simplified model of the coalescent theory, an expected coalescent time for mtDNA is N generations ago, where N is the effective population size. Assuming 5 years as the generation time of camels (Fitak *et al.* 2016), N for *C. dromedarius*, *C. bactrianus*, and *C. ferus* become 140,000, 18,000, and 120,000, respectively. Wu *et al.* (2014) estimated the temporal changes of effective population sizes of *C. dromedarius* and *C. bactrianus*. The long-term (up to one 100,000 years ago) harmonic mean (Wright, 1938) for both species are much smaller than 50,000. The same estimate for *C. dromedarius* based on estimates by Fitak *et al.* (2016) was much smaller. MtDNA can be considered to be one locus, and any population genetic estimate based solely on mtDNA is expected to contain a large variance. Therefore, discrepancy between mtDNA and nuclear DNA on effective population size estimation is not surprising.

Fertile hybrids between *C. bactrians* and *C. dromedarius* are produced in camel ranches in Kazakhstan. Blood samples from many camels were collected from various camel farms in Kazakhstan with help of researchers in Kazakhstan. We are now determining nucleotide sequences of camel mtDNA. While mtDNA data are used only as phylogenetic markers, nuclear genome data comparison is expected to pinpoint genomic region(s) which controls hump formation.

Acknowledgments

Our camel DNA study is a part of the integrated study of camels in Central Asia organized by Professor Kaoru Imamura of Nagoya Gakuin University. Professors Buho Hoshino and Tomomi Ishii of Hokkaido Raiuno Gakuin University are also members of this project. Professor Nurtazin Sabir and Mr. Azim Baibagyssov of Al-Farabi Kazakh National University are main collaborators of this project in Kazakstan. Mrs. Masako Mizuguchi of Saitou Laboratory at National Institute of Genetics helped us in sequence analyses. We thank all of these people. This study was partially supported by grant-in-aid for scientific research 26300013 given to Professor Kaoru Imamura.

References

Almathen F., Charruau P., Mohandesan E. *et al.* (2016) Ancient and modern DNA reveal dynamics of domestication and cross-continental dispersal of the dromedary. Proc. *Natl. Acad. Sci. USA*, **113**(24): 6707–6712.

Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. (1990): Basic local alignment search tool. *J. Mol. Biol.*, **215**: 403-410.

Arnason U., Gullberg A., Janke A. (2004): Mitogenomic analyses provide new insights into cetacean origin and evolution. *Gene*, **333**: 27-34.

The Bactrian Camels Genome Sequencing and Analysis Consortium (2012): Genome sequences of wild and domestic bactrian camels. *Nat. Commun.*, **3**:1202.

Cui P., Ji R., Ding F., Qi D., Gao H., Meng H., Yu J., Hu S., Zhang H. (2007): A complete mitochondrial genome sequence of the wild two-humped camel (*Camelus bactrianus* ferus): an evolutionary history of camelidae. *BMC Genomics*, **8**: 241.

Di Rocco F., Zambelli A., Mate L., Vidal-Rioja L. (2010): The complete mitochondrial DNA sequence of the guanaco (*Lama guanicoe*): comparative analysis with the vicuna (*Vicugna Vicugna*) genome. *Genetica*, **138**(8): 813-818.

Edgar R.C. (2004): MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl. Acids Res.*, **32**: 1792-1797.

Fitak R.R., Mohandesan E., Corander J., Burger P.A. (2016): The de novo genome assembly and annotation of a female domestic dromedary of North African origin. *Mol. Ecol. Resource*, **16**: 314-324.

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.

Heintzman P.D., Zazula G.D., Cahill J.A., Reyes A.V., MacPhee R.D.E., Shapiro B. (2015): Genomic data from extinct North American *Camelops* revise camel evolutionary history. *Mol. Biol. Evol.*, **32**: 2433-2440.

Irwin D., Kocher D., Wilson A.C. (1991): Evolution of the cytochrome b gene of mammals. *J. Mol. Evol.*, **32**: 128-144.

Ji R., Cui P., Ding F., Geng J., Gao H., Zhang H., Yu J., Hu S., Meng H. (2009): Monophyletic origin of domestic bactrian camel (*Camelus bactrianus*) and its evolutionary relationship with the extant wild camel (*Camelus bactrianus* ferus). *Animal Genet.*, **40**: 377-382.

Kadwell M., Fernandez M., Stanley H.F., Baldi R., Wheeler J.C., Rosadio R., Bruford W. (2001): Genetic analysis reveals the wild ancestors of the llama and the alpaca. *Proc. R. Soc. London B*, **268**: 2575-2584.

Li H., Durbin R. (2011): Inference of human population history from individual whole-genome sequences. *Nature*, **475**: 493-496.

Price S.A., Bininda-Emonds O.R.P., Gittleman J.L. (2005): A complete phylogeny of the whales, dolphins and even-toed hoofed mammals (*Cetartiodactyla*). *Biol. Rev.*, **80**: 445-473.

Saitou N. (2013): *Introduction to Evolutionary Genomics*. Springer, London.

Saitou N., Nei M. (1987): The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, **4**: 406-425.

Shimamura M., Yasue H., Ohshima K., Abe H., Kato H., Kishiro T., Goto M., Munechika I., Okada N. (1997): Molecular evidence from retroposons that whales form a clade within even-toed ungulates. *Nature*, **388**: 666-670.

Stanley H.F., Kadwell M., Wheeler J.C. (1994): Molecular evolution of the family Camelidae: a mitochondrial DNA study. *Proc. R. Soc. London B*, **256**: 1-6.

Tamura K., Nei M. (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Bio.l Evol.*, **10**(3): 512-526.

Tamura K., Stecher G., Peterson D., Filipski A., Kumar S. (2013): MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.*, **30**: 2725-2729.

Ursing B.M., Slack K.E., Arnason U. (2000): Subordinal artiodactyl relationships in the light of phylogenetic analysis of 12 mitochondrial protein-coding genes. *Zool. Scr.*, **29**: 83-88.

Wright S. (1938): Size of population and breeding structure in relation to evolution. *Science*, **87**: 430-431.

Wu H., Guang X., Al-Fageeh M.B. *et al.* (2014): Camelid genomes reveal evolution and adaptation to desert environments. *Nat. Commun.*, **5**: 5188.