

STUDY ON GENETIC VARIATIONS OF THE D-LOOP REGION IN THREE VIETNAMESE ETHNIC GROUPS KINH, LOLO AND LAHU

Nguyen Thuy Duong^{1,2,✉}, Nguyen Van Phong^{1,2}, Nguyen Thy Ngoc^{1,3}, Nong Van Hai^{1,2}

¹*Institute of Genome Research, Vietnam Academy of Science and Technology*

Graduate University of Science and Technology, Vietnam Academy of Science and Technology

²*University of Science and Technology of Hanoi, Vietnam Academy of Science and Technology*

✉To whom correspondence should be addressed. E-mail: tdnguyen@igr.ac.vn

Received: 15.4.2020

Accepted: 19.6.2020

SUMMARY

Mitochondrial DNA (mtDNA) sequence analysis has been widely used to investigate population genetic and evolutionary history of different ethnic groups worldwide. In this study, the D-loop region of 119 Vietnamese individuals belonging to three different ethnic groups was sequenced and compared with reference mtDNA on rCRS for genetic variations. Total 218 genetic polymorphisms were found in this population, among which 48 variations appeared with frequencies of more than 0.1. Further statistical analysis showed that there were 23, 13 and 24 single nucleotide polymorphisms (SNP) distributed differently between Kinh vs. Lolo, Kinh vs. Lahu, and Lolo vs. Lahu groups, respectively. The mean pairwise genetic distances between each pair of ethnic groups were 0.0101, 0.0098 and 0.0092 for Kinh - Lolo, Kinh - Lahu, Lolo - Lahu, respectively. This indicated that although the Lolo and Lahu ethnic groups both belong to the Tibeto-Burman language group, the genetic distance between them was no closer than the genetic distances between them and the Kinh group of the Viet-Muong (Austroasiatic) language group. This study was the first research on the D-loop region of the Lolo and Lahu populations in Vietnam. These results provided more data for exploring the genetic background and the history of those ethnic groups as well as other minority ethnic groups in Vietnam.

Keywords: *mtDNA, D-loop, genetic polymorphisms, Kinh, Lolo, Lahu.*

INTRODUCTION

Among the 54 ethnic groups of people in Vietnam, Lolo and Lahu are two ethnic groups of lowest numbers. According to the Report of General Statistics Office of Vietnam 2009, Lolo and Lahu populations in Vietnam were approximately 4550 and 9600, respectively (General Statistics Office of Vietnam, 2010). The Lahu ethnic, mostly living in Lai Chau province, has three sub-groups: black Lahu, yellow Lahu and white Lahu. The two Lolo sub-groups, red Lolo and black Lolo, are more spread over Lai Chau, Ha Giang and Cao Bang provinces (Edmondson, 2003). Culturally, both Lahu and Lolo ethnic groups belong to the Tibeto-Burman language group, a branch of the Sino-Tibetan language family, which has more

than 400 languages widely spoken in East, South and Southeast Asia (Bradley, 1997).

The human mitochondrial genome consists of a closed, circular double-stranded DNA molecule. It has 16569 base pairs encoding two rRNAs, 22 tRNAs and 13 polypeptides (Taanman, 1999). In the mitochondrial DNA (mtDNA), the Displacement loop (D-loop) is a non-coding region, which plays a key role in the human mtDNA replication process (Holt Reyes, 2012). D-loop sequence analysis has become a powerful tool for studying the phylogeography and historical origin of ethnic groups (Bertranpetit *et al.*, 1995), as well as forensic investigation (Chemale *et al.*, 2013).

Previously, several groups have investigated

the D-loop sequence and its variations in ethnic minorities of Vietnam. The first study of this kind in 2005 revealed 32 variations in D-loop region of three Vietnamese individuals, belonging to three ethnic groups (Kinh, Tay and Hmong), among which 8 variations were novel polymorphisms (Huỳnh Thị Thu Huệ *et al.*, 2005). In 2008, a study sequenced the D-loop regions of 78 healthy people of 3 ethnic groups: Kinh, Tay and Muong and detected 1043 polymorphisms, of those 73 polymorphisms were novel (Nguyễn Đăng Tôn *et al.*, 2008). In 2016, sequenced and analyzed mtDNA HV-II fragments in D-loop region of 169 healthy individuals from four ethnic groups belonging to Kinh, Muong, Ede and Jarai. 79 different haplogroups were found among these 169 individuals (Hung *et al.*, 2016). In 2017 a research on 622 individuals from 6 ethnic groups: Kinh, Mong, Tay, Nung, Hoa, Khmer, showed a historical gene flow between these Vietnamese ethnics and other neighboring countries (Pischedda *et al.*, 2017).

However, until now, there has been no research on the D-loop sequence and its variations in the Lolo and Lahu populations. These two ethnics belong to the same language group, and live close to each other in northeastern Vietnam with many cultural similarities. In this study, we investigated the sequences of the D-loop regions of Lolo and Lahu people and the variations within them. The obtained data was compared with that of Kinh population, which is the majority group of Vietnam.

MATERIALS AND METHODS

Subjects

The study subjects were Vietnamese individuals belonging to 3 ethnic groups: Kinh (51 individuals), Lahu (32 individuals) and Lolo (36 individuals). All individuals were healthy, and unrelated with at least 3 generations (individual/parents/grandparents) belonging to the same ethnic. All participants were clearly informed of the purpose and procedure of this

scientific project and volunteered to provide their blood samples to this study.

Molecular methods

Genomic DNA was extracted from 200 µl of peripheral blood sample by GeneJET Whole Blood Genomic DNA Purification kit (Thermo Fisher Scientific Inc., Vilnius, Lithuania) following the manufacturer's protocol. The genomic DNA concentration and purity were measured using Nanodrop Eppendorf Biospectrometer® basic. The two DNA fragments containing HV-I and HV-II genes of the D-loop region were amplified by polymerase chain reaction (PCR) amplification in 20µl reaction volumes with 50 ng of genomic DNA, 1X PCR buffer, 2.5mM dNTPs and 0,2µM of each primer and 1 U of Taq DNA polymerase (Qiagen). Thermocycling started with a single denaturation step for 5 min at 95°C, followed by 40 cycles x (denaturation for 45s at 95°C, annealing for 30 s at 55°C and extension at 72°C for 1 min); one final extension step was added for 8 mins at 72°C. PCR products were checked using 1.5% agarose gel electrophoresis to verify product size and then purified with the GeneJET PCR Purification Kit (Thermo Scientific). The sequence of corresponding PCR product was obtained through Sanger sequencing method using ABI Big Dye Terminator v3.1 Sequencing Standard Kit (Applied Biosystems, CA) on ABI 3500 Genetic Analyzer sequencer (Applied Biosystems). All the sample sequences were compared with reference mitochondrial human genome rCRS (Revised Cambridge Reference Sequence of the Human Mitochondrial DNA), Genbank ID NC_012920.1 using BioEdit version 7.2.6 software.

Statistical analysis

Fisher exact test was used to test for pairwise statistical differences in frequencies of the genetic variations in the D-loop region between ethnic groups. The obtained data was analyzed using SPSS software version 23.0 (IBM company, Chicago, United States) and Office Excel software version 2013 (Microsoft

Corp, Washington DC, US). All p-values (2 sided) under 0.05 were considered significant.

Comparison of genetic distances between ethnic groups

D-loop sequences from all individuals were loaded into MEGA7 version 7.0.9. The mean pairwise genetic distance was calculated for each pair of ethnic groups.

RESULTS

Genetic polymorphisms found in the three ethnics

In this study, 218 variations in the D-loop

region of 119 individuals from 3 ethnic groups were obtained from Sanger sequencing (Fig 1). The distribution of variations among ethnic groups was shown in Table 1.

Among these 218 genetic variations, 49 polymorphisms were common variations (with frequencies ≥ 0.1 in at least one ethnic group). The 10 most common genetic variations were A73G, A263G, T310C, T489C, T16519C, T16189C, A16183C, T16217C, A16182C, and G16129A (Table 2). Among these, two variations A73G and A263G were found in almost all Vietnamese people (allele frequency: AF ~ 1).

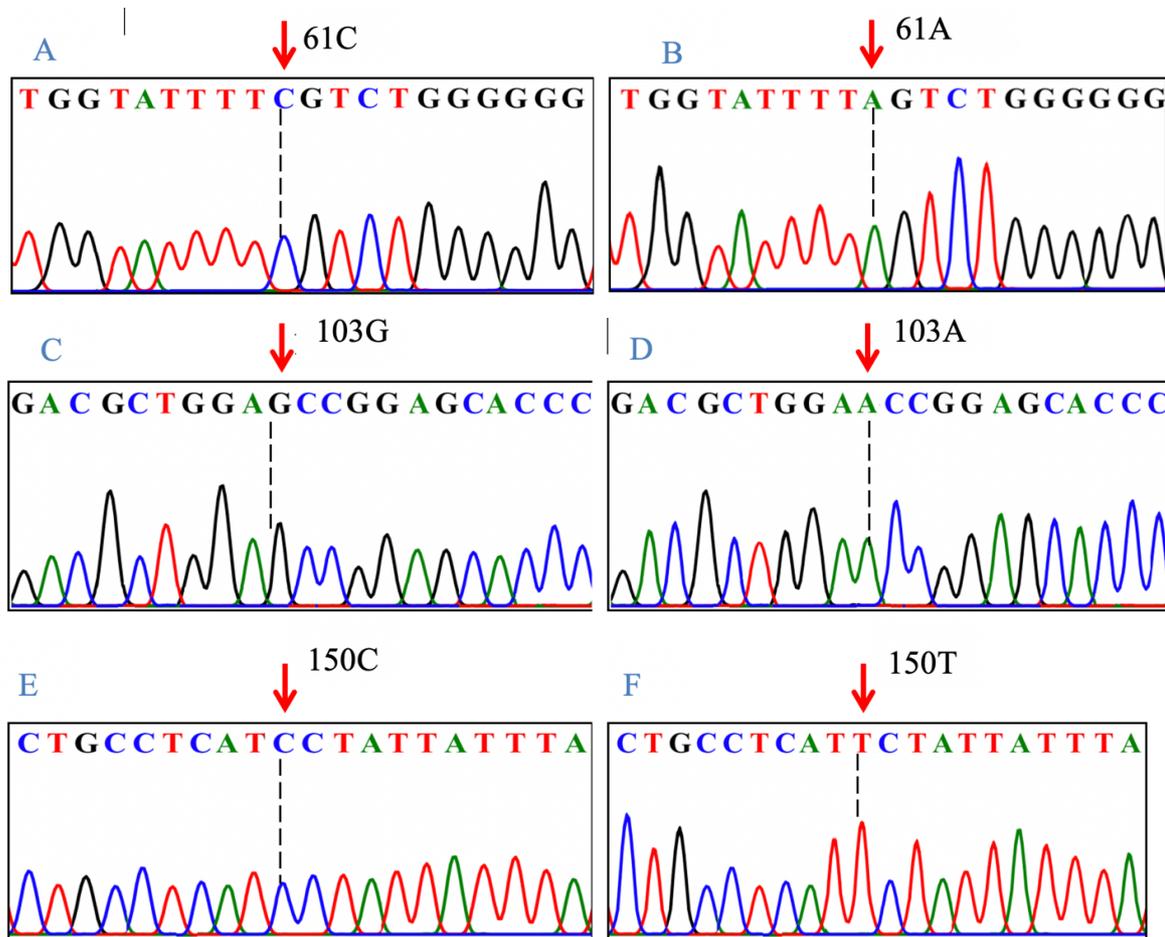


Figure 1. Sanger sequencing results of three representatives. A, B. From Kinh; C, D. From Lahu; E, D. From Lolo.

Table 1. Number of genetic variations in D-loop region found in this study.

Ethnic	Number of individuals	Number of genetic variations in D-loop region		
		Maximum per individual	Minimum per individual	Mean ± S.D
Kinh	51	16	6	11.2±2.3
LaHu	32	15	7	11.4±1.9
LoLo	36	15	7	12.4±3.3
All population	119	16	6	

Table 2. The most 10 popular genetic variations in the studied population.

No	Variation	Frequency in the studied population			
		Kinh ethnic	Lahu ethnic	Lolo ethnic	All population
1	A73G	1	1	1	1
2	A263G	0.98	1	1	0.99
3	T310C	0.78	0.84	0.92	0.84
4	T16519C	0.75	0.63	0.97	0.78
5	T16189C	0.57	0.72	0.72	0.66
6	A16183C	0.47	0.69	0.72	0.61
7	T16217C	0.12	0.66	0.53	0.39
8	A16182C	0.18	0.38	0.53	0.34
9	G16129A	0.29	0.19	0.36	0.29
10	T489C	0.45	0.19	0.03	0.25

Some polymorphisms had different distributions among ethnic groups. Variation T16217C was the major allele in the Lahu group (AF = 0.66) and Lolo group (AF = 0.53), but in the Kinh group it was a minor allele (AF = 0.12). In contrast, the variation T489C was quite common in the Kinh group (AF = 0.45), and relatively rare in the Lolo group (AF = 0.03).

Statistical analysis

The result showed that 23 variations were distributed differently between Kinh and Lolo ethnic groups, among these three were strongly significant (p < 0.0001), being T489C, T16217C and C16223T. The number of different variations between Kinh group and Lahu group was 13, in which two were strongly

significant: T16217C and A16235G. Between Lolo group and Lahu group, there were 24 variations with significant differences in distribution. However, none of these had p – value less than 0.0001 (Table 3).

Genetic distances

The pairwise genetic distances were calculated pairwise using MEGA7 between four groups: rCRS (reference), Kinh ethnic, Lahu ethnic, and Lolo ethnic. Result showed that genetic distances were similar between comparisons (Table 4). Among the ethnic groups of Vietnam, the distances between Kinh - Lahu and Kinh - Lolo were approximately 0.01 (0.0098 and 0.0101, respectively). Meanwhile, the genetic distance between Lahu - Lolo was slightly smaller: 0.0092.

Table 3. Statistical analysis of the variation distribution between ethnic groups.

Variations	Kinh - Lolo ^(*)	Kinh - Lahu ^(*)	Lolo - Lahu ^(*)
C150T	0.001	0.017	
T152C	0.003		0.003
A183G	0.0004		0.005
T199C	0.02	0.031	
A214G	-	0.001	0.007
A244G	0.0004		0.005
G247A	-	0.01	0.029
A263G	-		
T310C	-		
A368G	-	0.001	0.007
A374G	0.0004		0.005
T489C	0.000014	0.014	0.031
C16108T	0.043	0.026	
C16147T	-	0.000365	0.001
T16178C	-	0.001	0.007
A16180C	0.006		0.029
A16181C	0.006		0.029
A16182C	0.001	0.043	
A16183C	0.019		
C16184A	-	0.000172	0.001
G16213A	0.006		0.029
T16217C	0.00003	0.0000003	
C16223T	0.000019		0.003
C16232A	0.015		
C16234T			0.029
A16235G		0.00006	0.001
T16249C			0.029
C16266A			0.037
G16274A	0.002		0.02
A16289G	0.0004		0.005
C16292T	0.006		0.029
T16297C	0.033		
C16301T	0.002		0.005
T16311C	0.02		
T16519C	0.005		0.000279

^(*) p-value was calculated by Fisher exact test (2-sided).

Table 4. Mean genetic distance between each group of ethnic.

	rCRS	Kinh	Lahu	Lolo
rCRS				
Kinh	0.0082			
Lahu	0.008	0.0098		
Lolo	0.0088	0.0101	0.0092	

DICUSSION

In the three Vietnamese ethnic groups, almost all individuals carried the two D-loop mtDNA variations A73G and A263G (only 1 individual in Kinh group carried A263). This result was similar to previous studies in other Vietnamese ethnic groups such as Muong, Tay, Hmong (Huỳnh Thị Thu Huệ *et al.*, 2005; Trần Thị Thúy Hằng *et al.*, 2017). These two variations were also found in the Guangxi Bama population of China at a low ratio: 10 carriers / 113 individuals (Yang *et al.*, 2012). Thus, these variations can be specific to the Vietnamese population.

Among the most 10 popular variations in this study (Table 2), some polymorphisms were differently distributed between ethnics, especially T16217C, A16182C and T489C (described above). Besides, statistical analysis showed that there were 23, 13 and 24 genetic variations significantly different between Kinh - Lolo, Kinh - Lahu and Lolo - Lahu, respectively. These findings revealed many noticeable differences in the mtDNA genetic background of these three Vietnamese ethnic groups. Genetic distance analysis between three groups also indicated that although Lolo and Lahu ethnic groups belong to the same language group, the distance between them was not significantly different from genetic distances between Kinh - Lolo and Kinh - Lahu (0.0092 compared to 0.0101 and 0.0098, respectively).

A recent research on mtDNA diversity in Chinese ethnic populations revealed that the Sali ethnic group (a branch of Lolo in China) was quite close, in terms of genetic background, to the Mongolian and Uygur ethnic group

(genetic distance = 0.00042 and 0.00018, respectively) (Yao *et al.*, 2002). In 2010, a phylogenetic study on the HLA gene also indicated that the Lolo (Yi) people in Yunnan, China, despite being Southeast Asian, had some common characteristics with northeast Asian people (Zhu *et al.*, 2010). Meanwhile, the Lahu ethnic group, of the same language family with Lolo, was demonstrated to possess more genetical characteristics of southern Asians than northern Asians in several studies. Phylogenetic analysis of the HLA-DRB1 genetic polymorphisms in Lahu population in Yunnan, China showed that there were not many similarities between this ethnic group and the northern Asian groups (Jia *et al.*, 2002), while another phylogeographic study on mtDNA showed that most Lahu subjects carried original southern Asian haplogroups (Yao & Zhang, 2002).

Although the Lolo and Lahu people have inhabited Northern Vietnam for a long time, their origin and date of migration into Vietnam are still unclear. Our study, being the first research on their mtDNA, provided the data for further genetical as well as anthropological research on these questions.

CONCLUSION

In this research, we studied the genetic variations in the D-loop regions of 119 individuals of 3 different Vietnamese ethnic groups (Kinh, Lolo and Lahu) through Sanger sequencing method. This is the first study that investigated mtDNA genetic variations in Lolo and Lahu populations. The D-loop DNA sequences showed that although they belong to

the same language group, there were still many distinctive differences in mitochondrial DNA between them. These results provided a better understanding on the genetic background and genetic variations of the three Vietnamese ethnic groups.

Acknowledgement: *This research was supported by a grant from the Ministry of Science and Technology project DTDL/CN.60/19. The authors wish to thank all the volunteers and the blood donors that participated in our study.*

REFERENCES

- Bertranpetit J, Sala J, Calafell F, Underhill PA, Moral P, Comas D (1995) Human mitochondrial DNA variation and the origin of Basques. *Ann Human Genet* 59(1): 63-81.
- Bradley D (1997) Tibeto-Burman languages and classification. In Bradley D (Ed.) *Papers in Southeast Asian Linguistics* (14): 1-71.
- Chemale G, Paneto GG, Menezes MA, de Freitas JM, Jacques GS, Cicarelli RM, Fagundes PR (2013) Development and validation of a D-loop mtDNA SNP assay for the screening of specimens in forensic casework. *Forensic Sci Int Genet* 7(3): 353-358.
- Edmondson, Jerold A. 2003. Three Tibeto-Burman languages of Vietnam. In: David Bradley and Boyd Michaelovsky and Randy LaPolla and Graham Thurgood (eds.), *Language variation: papers on variation and change in the Sinosphere and in the Indosphere in honour of James A. Matisoff*, 305-320. Canberra: Research School of Pacific and Asian Studies, Australian National University
- General Statistics Office of Vietnam (2010) The 2009 Vietnam Population and Housing census: Completed results. *Central Population and Housing Census Steering Committee*: 3-5.
- Holt IJ, Reyes A (2012) Human mitochondrial DNA replication. *Cold Spring Harb Perspect Biol*: 4(12).
- Hung DM, Ha NH, Khoi PN, Nhung VP, Phong NV, Duong NT, Ton ND (2016) Genetic variation of mitochondrial sequence-hv2 in Vietnamese populations. *Acad J Biol* 38(2): 243-249.
- Huynh Thi Thu Hue, Hoang Thi Thu Yen, Nguyen Dang Ton, Le Thi Thu Hien, Nguyen Dinh Cuong, Phan Van Chi, Nong Van Hai (2005) Analysis of the mitochondrial DNA D-loop region in five unrelated Vietnamese individuals. *J Biotechnol* 3(1): 15-22.
- Jia ZJ, Fu YG, Pan DJ, Liu ZH, Chen WM, Lin JH, Xu AL (2002) Determination of HLA-DRB1 gene polymorphism by PCR-SBT in Lahu ethnic group of Yunnan, China. *Hereditas* 24(2): 131-136.
- Nguyen Dang Ton, Nguyen Thi Tu Linh, Vu Hai Chi, Tran Thi Ngoc Diep, Dich Thi Kim Huong, Bui Thi Tuyet, Nong Van Hai (2008) Single nucleotide polymorphism of the mitochondrial DNA of Vietnamese individuals. *J Biotechnol* 6(4): 559-590.
- Pischedda S, Barral-Arca R, Gomez-Carballa A, Pardo-Seco J, Catelli ML, Alvarez-Iglesias V, Salas A (2017) Phylogeographic and genome-wide investigations of Vietnam ethnic groups reveal signatures of complex historical demographic movements. *Sci Rep* 7(1): 12630.
- Taanman JW (1999) The mitochondrial genome: structure, transcription, translation and replication. *Biochim Biophys Acta* 1410(2): 103-123.
- Tran Thi Thuy Hang, Tran Huy Thinh, Tran Van Khanh (2017) Studying single nucleotide polymorphism of the mitochondrial DNA HVI and HVII regions in Kinh and Muong groups. *J Med Res* 106(1)
- Yang X, Wang X, Yao H, Deng J, Jiang Q, Guo Y, Jiang H (2012) Mitochondrial DNA polymorphisms are associated with the longevity in the Guangxi Bama population of China. *Mol Biol Rep* 39(9): 9123-9131.
- Yao YG, Nie L, Harpending H, Fu YX, Yuan ZG, Zhang YP (2002) Genetic relationship of Chinese ethnic populations revealed by mtDNA sequence diversity. *Am J Phys Anthropol* 118(1): 63-76.
- Yao YG, Zhang YP (2002) Phylogeographic analysis of mtDNA variation in four ethnic populations from Yunnan Province: new data and a reappraisal. *J Hum Genet* 47(6): 311-318.
- Zhu BF, Yang G, Shen CM, Qin HX, Liu SZ, Deng YJ, Lucas R (2010) Distributions of HLA-A and -B alleles and haplotypes in the Yi ethnic minority of Yunnan, China: relationship to other populations. *J Zhejiang Univ Sci B* 11(2): 127-135.

ĐA HÌNH VÙNG D-LOOP HỆ GEN TY THỂ CỦA CÁC CÁ THỂ DÂN TỘC KINH, LÔ LÔ VÀ LA HỦ

Nguyễn Thùy Dương^{1,2}, Nguyễn Văn Phòng^{1,2}, Nguyễn Thy Ngọc^{1,2}, Nông Văn Hải^{1,2}

¹*Viện Nghiên cứu hệ gen, Viện Hàn lâm Khoa học và Công nghệ Việt Nam*

²*Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam*

³*Trường Đại học Khoa học và Công nghệ Hà Nội, Viện Hàn lâm Khoa học và Công nghệ Việt Nam*

TÓM TẮT

Vùng D-loop hệ gen ty thể được sử dụng rộng rãi trong các nghiên cứu về di truyền học quần thể cũng như tiến hóa. Trong nghiên cứu này, chúng tôi đã xác định trình tự vùng D-loop của 199 cá thể Việt Nam thuộc ba dân tộc Kinh, Lô Lô và La Hủ và so sánh với trình tự hệ gen ty thể tham chiếu (rCRS) để tìm các đa hình di truyền. Kết quả đã phát hiện 218 điểm đa hình xuất hiện với tần suất hơn 0,1. Phân tích thống kê cho thấy số điểm đa hình có tần suất thu được khác nhau có ý nghĩa thống kê giữa hai dân tộc: Kinh - Lô Lô, Kinh - La Hủ, và Lô Lô - La Hủ lần lượt là 23, 13 và 24 điểm. Khoảng cách di truyền trung bình giữa hai dân tộc Kinh - Lô Lô, Kinh - La Hủ, và Lô Lô - La Hủ lần lượt là 0,0101, 0,0098 và 0,0092. Điều này cho thấy mặc dù hai dân tộc Lô Lô và La Hủ cùng thuộc nhóm ngôn ngữ Tibeto-Burman nhưng khoảng cách di truyền giữa hai dân tộc này không gần hơn khoảng cách di truyền giữa họ và dân tộc Kinh thuộc nhóm ngôn ngữ Việt-Mường. Nghiên cứu này là nghiên cứu đầu tiên về vùng D-loop trên hai dân tộc Lô Lô và La Hủ tại Việt Nam. Những kết quả này cung cấp thêm dữ liệu tạo nền tảng cho các nghiên cứu về di truyền và lịch sử của hai dân tộc Lô Lô và La Hủ cũng như các nhóm dân tộc thiểu số khác ở Việt Nam.

Từ khoá: *Hệ gen ty thể, D-loop, đa hình, dân tộc Kinh, dân tộc Lô Lô, dân tộc La Hủ.*