

(RESEARCH ARTICLE)



mtDNA haplotype network analysis: Exploring genetic relationships and diversity in dog haplogroups

Quan Ke Thai *, Tuan Thanh Nguyen and Hien Thi Thanh Pham

Faculty of Natural science Education, Saigon University, 273 an Duong Vuong, Ward 3, District 5, Ho Chi Minh City, Vietnam, 70000.

GSC Biological and Pharmaceutical Sciences, 2023, 24(01), 224–232

Publication history: Received on 10 June 2023; revised on 15 July 2023; accepted on 18 July 2023

Article DOI: <https://doi.org/10.30574/gscbps.2023.24.1.0284>

Abstract

The genetic diversity and relationships of dog haplogroups were studied by analyzing the HV1 region of mitochondrial DNA. Previous studies have found six distinct haplogroups (A, B, C, D, E, and F) in dogs. Haplogroups A, B, and C were widely distributed, while haplogroups D, E, F were rare and distributed in restricted regions. In this study, HV1 sequences from global dog populations were collected, categorized into haplotypes, and used to construct haplotype networks. The results showed that haplogroup A was the most prevalent, comprising approximately 72.34% of dogs worldwide. Haplogroups A, B, and C together accounted for around 97.40% of the global dog population. Haplogroups D, E, and F were rare, constituting less than 3% of the dog population. Haplogroups E and F made up only about 1-2%. Number of haplotypes in haplogroups D, E, and F were little, suggesting that these haplotypes were introduced into the canine population more recently with limited time for significant mutations. Analysis of haplotype networks showed that haplotypes A were introduced into the dog populations in the early stage of dog domestication. Haplotype E harboring dogs were genetically close to wolves, suggesting a recent introduction of haplogroup E. Similarly, haplogroup F exhibited a narrow distribution primarily in Japanese dogs, with haplotype F3 identified as the founder haplotype likely introduced from a few wolves carrying the F3 haplotype. Through the analysis of the haplotype network and assessment of the betweenness value, this study has identified important haplotypes contributing significantly to the dog population. These analyses offer valuable insights into the identification of founder haplotypes involved in the formation of dog breeds worldwide, serving as a valuable reference for breed development and genetic studies.

Keywords: Betweenness; dog; Genetic diversity; Haplotype network; mtDNA

1. Introduction

The HV1 region of mitochondrial DNA in dogs displays diverse nucleotide sequences with various mutations including nucleotide insertion or deletion. Each of these unique sequences is known as a haplotype. Savolainen et al. (2002) constructed a phylogenetic tree using HV1 sequences from 654 dogs, representing a wide range of dog breeds[1]. The study revealed six distinct haplogroups, namely A, B, C, D, E, and F, which were also supported by Pang *et al.* (2009) using a larger sample size of 1543 dogs[1, 2]. These findings suggest that the physical characteristics of dog breeds are influenced by migration and interbreeding between breeds rather than the domestication of wolves in different regions.

The survey findings indicate that approximately 72.34% of dogs worldwide belong to haplogroup A, while around 97.40% belong to haplogroup A, B, or C[2]. Haplogroups A, B, and C are widely distributed across the globe, except for haplogroup C, which is absent in the Americas[3]. Conversely, haplogroups D, E, and F are considered rare, comprising less than 3% of the global dog population, with haplogroup E and F accounting for only about 1-2%. Haplotypes associated with haplogroups D, E, and F are primarily restricted to specific regions. For example, haplogroup D is found in Turkey, Spain, and Scandinavia, haplogroup E is present in Siberia, Japan, Korea, Indonesia, Thailand, and Vietnam,

*Corresponding author: Quan Ke Thai

and haplogroup F is identified in Japan[2, 4]. The widely accepted hypothesis is that domestic dogs have descended from gray wolves, supported by archaeological evidence and genetic studies. However, the exact time and location of dog domestication remain subjects of debate[2, 5-8]. To address this, researchers have utilized different DNA markers, including mitochondrial DNA. Shannon *et al.* analyzed nuclear and mitochondrial DNA from 4676 dogs, indicating Central Asia as the region of dog origin[9]. Vila *et al.* conducted sequence analysis of the HV1 region in wolves and domestic dogs, proposing that dog domestication occurred over 100,000 years ago from wolves[10]. Savolainen *et al.* (2002) challenged the previously mentioned hypothesis by focusing on the HV1 region of mitochondrial DNA. They analyzed a sequence of 582 base pairs (from nucleotide 15458 to nucleotide 16039) in 654 domestic dogs from major breeds worldwide. Their findings indicated that the East Asian domestic dog lineage originated from a single genetic lineage derived from wolves approximately 15,000 years ago[1]. In summary, the hypothesis of dog domestication from wolves has garnered support, but there are differing perspectives on the specific time and location. Shannon *et al.* suggested Central Asia as the origin based on a comprehensive analysis of nuclear and mitochondrial DNA. In contrast, Vila *et al.* proposed an earlier domestication period of over 100,000 years ago. However, Savolainen *et al.* contradicted these ideas, suggesting a more recent domestication event around 15,000 years ago, with the East Asian domestic dog descending from a specific wolf lineage. Pang *et al.* (2009) argued that the studies by Vila *et al.* (1997) and Savolainen *et al.* (2002) had limitations due to small sample sizes, making it challenging to draw accurate global conclusions[2]. However, Pang *et al.* conducted extensive investigations using complete mitochondrial genome analysis of 169 dogs and CR region analysis of 1543 dogs from across the Old World. Their research revealed that domestication occurred approximately 16,300 years ago from a few hundred female wolves. A more recent study by Wang *et al.* (2016) has made significant progress in confirming the origin of domestic dogs. Through analysis of the entire dog genome, the authors not only affirmed previous findings regarding the domestication timeframe but also constructed a migration map elucidating historical dog movements. Their research suggests that domestication took place around 33,000 years ago in Southeast Asia[7]. Approximately 15,000 years ago, a group of dogs began migrating towards the Middle East, Africa, and eventually reached Europe around 10,000 years ago. In the Middle East, some dogs migrated back to the east and interbred with local Asian canines, resulting in a genetically diverse population in northern China before further migration to the new continent. However, Wang *et al.*'s study also highlights unanswered questions regarding dog migration within Africa and within the New World. These areas remain subjects of ongoing research, requiring further investigation to gain a comprehensive understanding of dog migration patterns.

In this study, we employed social network analysis to gain insights into the connections among dog haplotypes across the globe. We collected 512 bp sequences of the HV1 region from canine mitochondrial DNA worldwide, sourced from GenBank. These sequences were categorized into haplotypes and utilized to construct a network representing haplotype relationships. In this network, each haplotype is represented as a node, and a line connects adjacent haplotypes, illustrating the nucleotide differences between them. By examining the closeness and betweenness values, we aimed to pinpoint the significant haplotypes within the network.

2. Material and methods

2.1. Data collection

The reference sequence for domestic dog mitochondrial DNA, consisting of 582 base pairs (GenBank U96639.2), was utilized in this study. Using the BLAST tool, this sequence was compared against the GenBank nucleotide database to identify similar sequences. A stringent E-value parameter of $10e-94$ was employed to ensure high similarity. The majority of sequences obtained from the BLAST results were mitochondrial DNA sequences resembling the standard sequence and originating from individuals of the *Canis lupus*. For each obtained sequence, the corresponding GenBank accession number was utilized to access and download comprehensive information from the GenBank database, including the nucleotide sequence, organism name, and sequence annotation. Only sequences belonging to *Canis lupus* were retained, and the relevant information was separated and used for further analysis.

2.2. Sequence alignment and mutation identification

The collected sequences were compared with the reference sequences (GenBank U96639.2) using ClustalW[11]. Differences of each sequence from the reference sequence would be recorded as mutations. The set of mutations of a sequence would be considered as its mutation profile.

2.3. Haplotype naming

Sequences with haplotypes published before would be assigned appropriate names. Sequences harbouring new mutation profiles would be named using the format XnNum, where X represents the corresponding haplogroup, n

indicates new, and Num denotes the sequence number, for example, An100 is the new haplotype with number 100 in the haplogroup A

2.4. Haplotype network

Due to significant differences in the nucleotide sequences of haplotypes across various haplogroups, separate haplotype networks were constructed for each of the six haplogroups. The haplotype network of each haplogroup was built using the Minimum spanning network method with the support of PopART1.7. The output of PopART was used as input data for Gephi 0.9 software to estimate the centrality of each haplotype in the network.

3. Results and discussion

3.1. Data collections

Using the BLAST tool found in GenBank's DNA database, 6238 nucleotide sequences derived from individuals of the *Canis lupus* had high similarity of over 90% with the HV1 sequence region of the reference sequence (GenBank U96639.2). Of these, sequences with unclear nucleotide (N) in the HV1 region were removed. The remaining 3946 sequences were grouped into 729 groups based on sequence similarities. Out of these 729 sequence groups, 298 sequence groups were identified as assigned haplotypes, the remaining 431 sequences carried new mutation profile. It is worth noting that all the assigned haplotypes had been recognized in dogs, whereas some unassigned haplotypes had been found only in wolves so far.

Table 1 Number of haplotypes

Haplogroup	assigned	unassigned	haplotype
A	217	353	570
B	46	47	93
C	19	21	40
D	9	3	12
E	4	5	9
F	3	2	5
Total	298	431	729

The majority (77%) of the collected sequences belonged to haplogroup A. These findings are consistent with previous studies. The total number of sequences from the three common haplogroups A, B, and C was 97.3% of the total, whereas the three rare haplogroups D, E, and F account for only 2.7%.

3.2. Haplogroup diversity

Although most of the collected sequences were haplotypes A, all six haplogroups had similar haplotype diversity. The lowest haplotype F had a diversity of only about 0.83, whereas other haplotypes had haplotype diversity from 0.90 to 1.00. However, nucleotide diversities and nucleotide differences were quite different. Compared to the other haplogroups, the nucleotide diversity of haplogroup A was much greater, even twice that of haplogroups B and C. These analysis indicated that haplogroup A sequences were abundant, resulted a diverse populations. This could be inferred that haplogroup A was introduced into domestic dog populations very early and supported Savolainen's hypothesis that a particular haplotype A (A29) is an ancient haplotype.

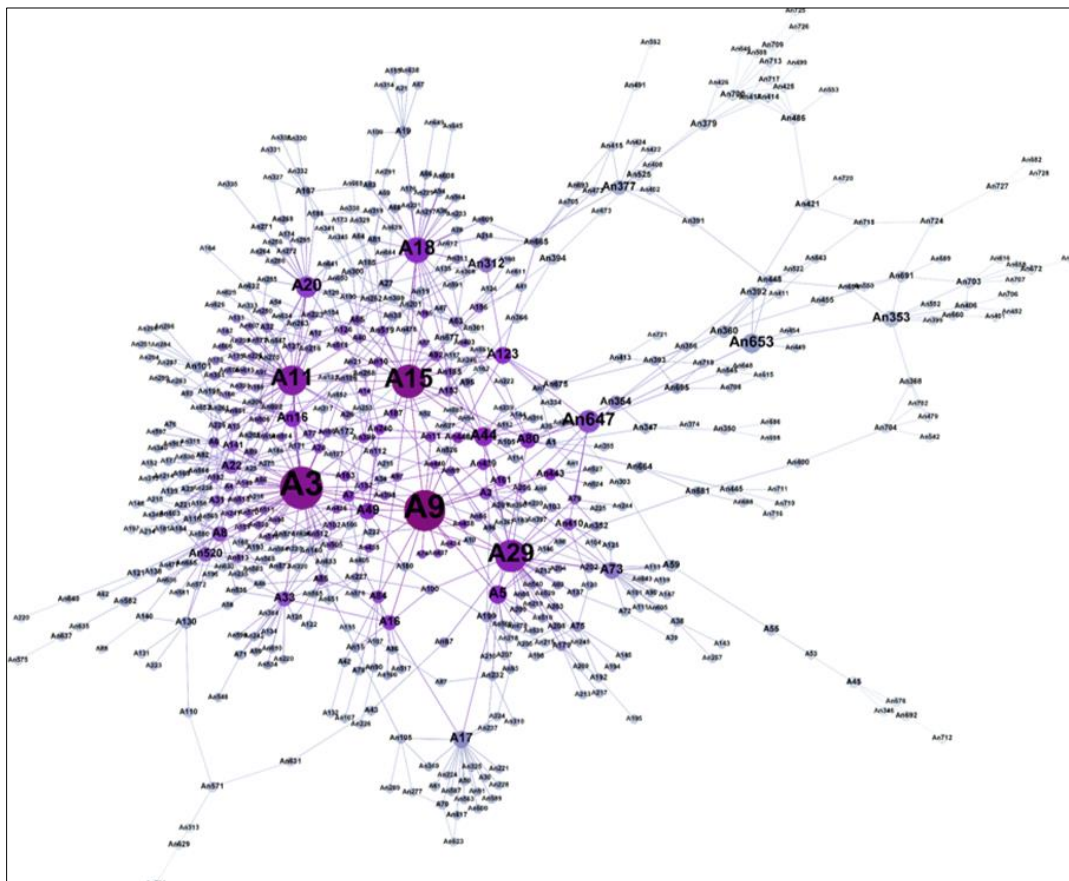
Even though haplogroups B and C exhibit the same level of haplotype diversity as haplogroup A, their nucleotide diversity is quite low. It is possible that these haplotypes were introduced to the dog population long after haplotypes A. Haplogroups E and F presented a unique scenario among all the haplogroups. The low number of sequences collected, coupled with the low haplotype and nucleotide diversity, suggested that these haplotypes had been recently introduced to the domestic dog population. This conclusion is supported by published studies, which indicated that dogs in haplogroups E and F are primarily found in specific regions such as Siberia, Japan, Korea, Indonesia, Thailand, and Vietnam.

Table 2 Diversity indice of haplogroups

Haplogroup	Nucleotide diversity	Haplotype diversity	Nucleotide differences pairwise
A	0.012302 ± 0.006366	0.9900 ± 0.0014	7.418337 ± 3.471493
B	0.006275 ± 0.003540	0.9927 ± 0.0029	3.683208 ± 1.876354
C	0.005362 ± 0.003146	0.9939 ± 0.0065	3.126263 ± 1.651829
D	0.008383 ± 0.004948	1.0000 ± 0.0340	4.878788 ± 2.557764
E	0.008539 ± 0.005030	0.9091 ± 0.0795	4.969697 ± 2.599896
F	0.007338 ± 0.004550	0.8333 ± 0.0980	4.277778 ± 2.337958

3.3. Haplotype network

Published studies has demonstrated that the emergence of various haplogroups could be introduced to the dog population by different wolveshabouring specific haplotypes. Haplotypes in different haplogroups exhibit notable diffences in their nucleotide sequences. For the ease in visualization, in this study, each haplotype network corresponding to each haplogroup were build for further analysis.



The nodes' size were adjusted corresponding to the betweenness

Figure 1 Haplotype network of haplogroup A

The haplotype network of haplogroup A showed its complex nature, characterized by the number of haplotypes and their connections. In this network, each haplotype is represented by a node, while a link between two nodes signifies a close relationship (one nucleotide mutation).The nodes in the network are also resized based on their betweenness value. Upon examining the network, certain haplotypes, namely A3, A9, A11, A15, A18, A20, A29, A44, A73, as well as unassigned haplotypes An647 and An653, exhibited notably high betweenness values and easily recognized within the

network. This observation emphasizes the importance of these haplotypes in the genetic structure of the dog population.

Table 3 20 haplotypes (haplogroup A) with highest betweenness values

No	Haplotype	betweenness	No	Haplotype	betweenness
1	A3	0,29311	11	A5	0,082257
2	A9	0,27878	12	A123	0,071596
3	A15	0,213681	13	A73	0,06781
4	A29	0,203212	14	A17	0,067317
5	A11	0,18036	15	An16	0,067067
6	A18	0,148098	16	A49	0,066323
7	An647	0,119879	17	A80	0,063997
8	A20	0,109476	18	An353	0,063442
9	An653	0,089886	19	An312	0,059721
10	A44	0,088095	20	A8	0,057173

An353 is a haplotype found only in wolf.

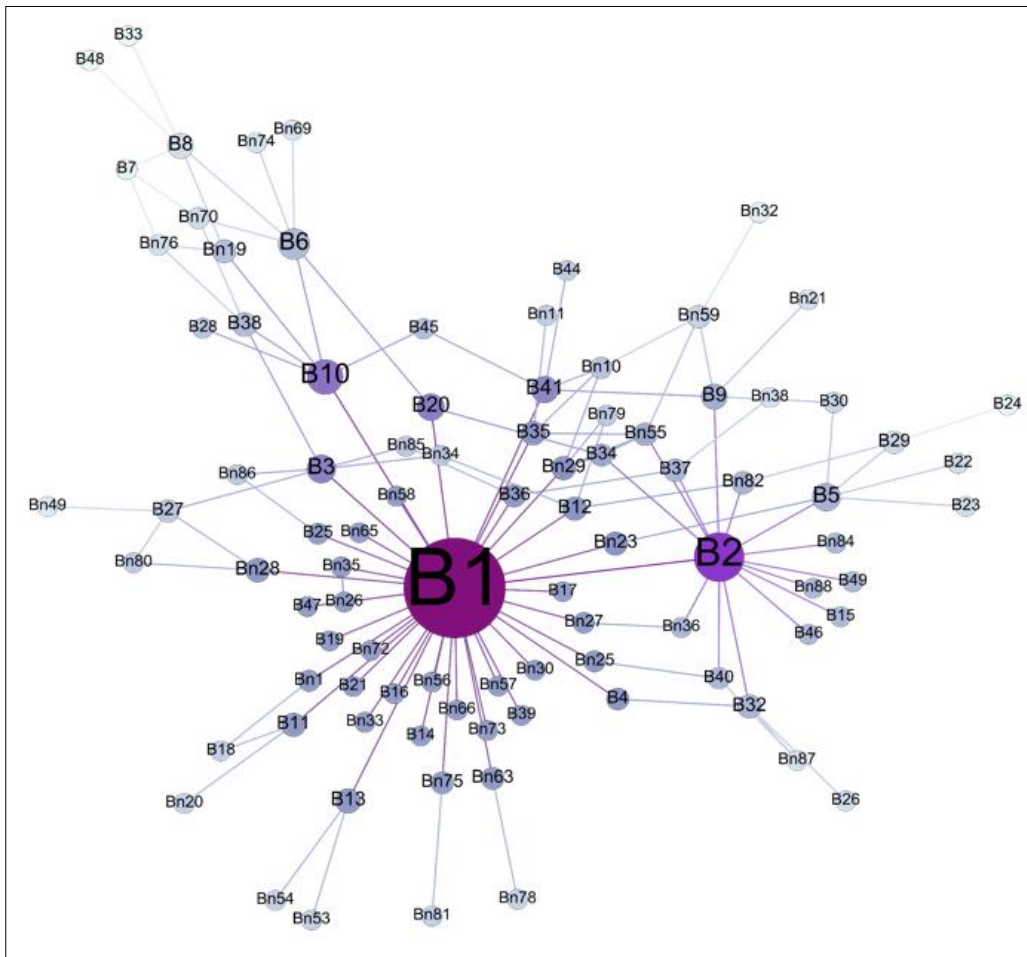


Figure 2 Haplotype network of haplogroup B

The haplotype network of haplogroup B exhibited a clearly structure with a fewer number of haplotypes. Notably, haplotypes B1, B2, B6, B10, and B20 stood out prominently within this network. Among them, haplotypes B1 and B2 assumed significant roles. Haplotype B1 located at the central position in the network, being only a maximum of three intermediate haplotypes away from other haplotypes within the network. This network, along with the mentioned diversity values, indicated that a limited number of haplotypes, potentially B1 and B2, were introduced into the dog population long after the introduction of haplotypes from haplogroup A. In a similar scenario to haplogroup B, the haplotypes within haplogroup C displayed a close relationship to each other. The four main haplotypes C1, C2, C3, and C17, exhibited one nucleotide different to adjacent haplotype, possessed the highest betweenness values among all the haplotypes, indicating their significance within the network. This observation suggests that at least one of these four haplotypes (as founder haplotype) was likely introduced into the dog population during a same timeframe as the haplotypes from group B.

Table 4 20 haplotypes (haplogroup B) with highest betweenness value

No	Haplotype	betweenness	No	Haplotype	Betweenness
1	B1	0,793008	11	B13	0,043239
2	B2	0,286933	12	B38	0,037836
3	B10	0,14641	13	Bn28	0,037294
4	B6	0,103023	14	B35	0,036948
5	B3	0,079805	15	B12	0,034461
6	B5	0,075029	16	B32	0,034376
7	B20	0,065832	17	B11	0,03237
8	B41	0,063001	18	Bn29	0,031932
9	B9	0,050913	19	Bn23	0,030908
10	B8	0,050306	20	Bn19	0,028111

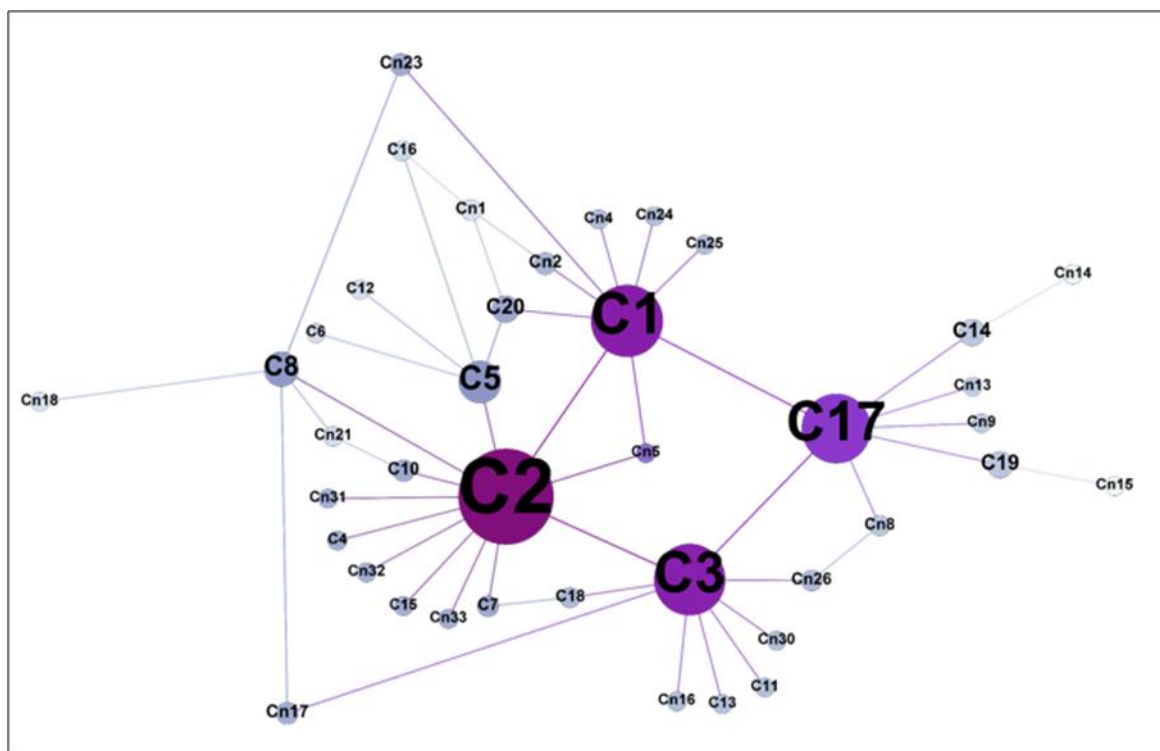


Figure 3 Haplotype network of haplogroup C

Table 5 20 haplotypes (haplogroup C) with highest betweenness value

No	Haplotype	betweenness	No	Haplotype	betweenness
1	C2	0,534683	11	Cn2	0,020794
2	C1	0,370355	12	Cn23	0,02037
3	C3	0,365227	13	C10	0,018264
4	C17	0,35018	14	C7	0,013315
5	C5	0,159784	15	Cn1	0,012191
6	C8	0,108277	16	Cn26	0,010999
7	C20	0,053689	17	C16	0,008243
8	C14	0,051282	18	C18	0,008187
9	C19	0,051282	19	Cn8	0,007917
10	Cn17	0,021645	20	Cn21	0,002249

Haplogroup D consisted of a smaller number of haplotypes. Like haplogroups B and C, the assigned haplotypes with high betweenness, namely D3, D5, D6, and D1, exhibited significance within the network. In contrast, the unassigned haplotypes had relatively minor roles, indicating their recent emergence. Information obtained from published studies provided a good explanation for this network. These haplotypes which distributed in restricted locations - Turkey, Spain, and Scandinavia - seemly introduced into the dog population long after the introduction of haplotypes from group A, hence experiencing fewer mutations.

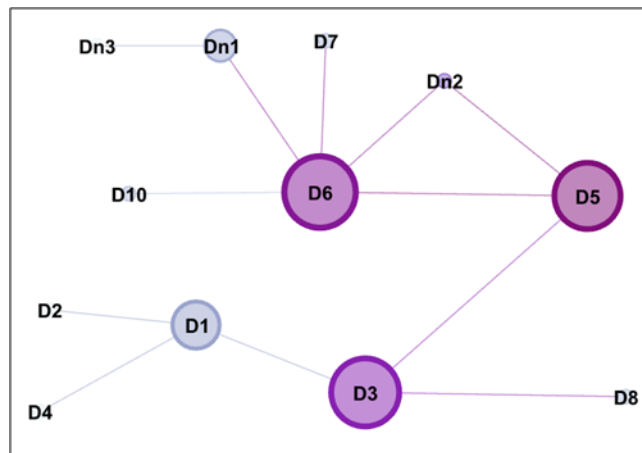


Figure 4 Haplotype network of haplogroup D

Table 6 Haplotypes with highest betweenness value in network of haplogroup D (A), haplogroup E (B), haplogroup F (C)

A			B			C		
No	Haplotype	Betweenness	No	Haplotype	Betweenness	No	Haplotype	Betweenness
1	D6	0,6	1	E2	0,714286	1	F3	1
2	D3	0,563636	2	En6	0,535714	2	Fn1	0
3	D5	0,545455	3	E1	0,464286	3	F2	0
4	D1	0,345455	4	En5	0,25	4	Fn2	0
5	Dn1	0,181818	5	En7	0,25	5	F1	0

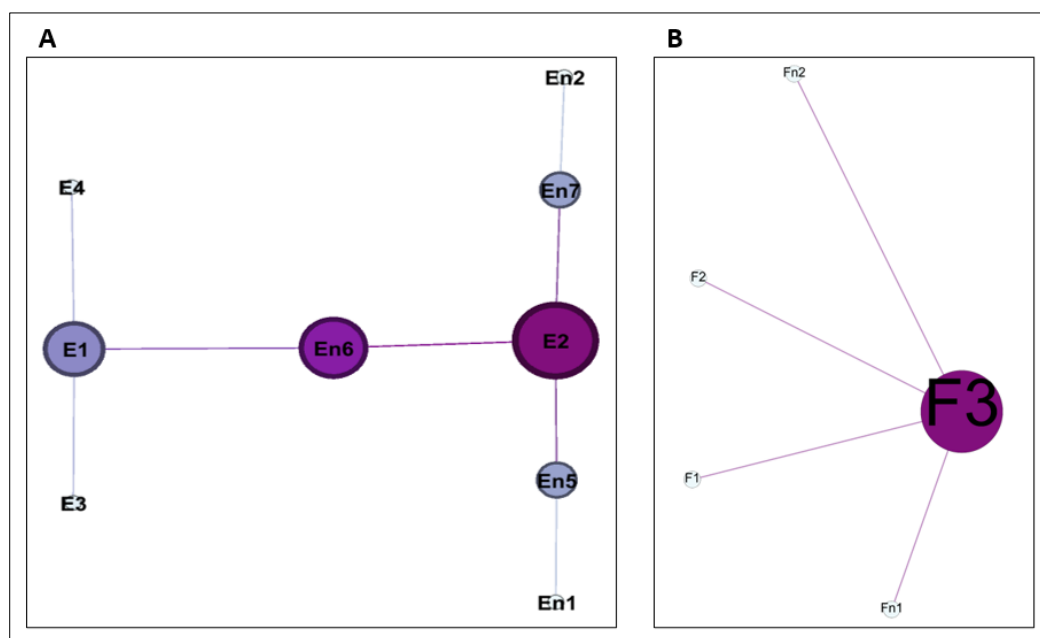


Figure 5 Haplotype network of haplogroup E (A) and haplogroup F (B)

The haplotype network of haplogroup E exhibits a simple structure comprising only nine nodes. Among these nodes, haplotypes E2, En6, E1, En5, and En7 stood out with high betweenness values. The network revealed that E1 served as the ancestor for both E4 and E3. The difference between E1 and E4 was a single nucleotide at position 15955, while E1 and E3 differed by two nucleotides at positions 15464 (insertion) and 16003. Haplotypes E1 and E2 were distinguished by five nucleotides, with an intermediate haplotype, En6, differentiating by two nucleotides from E1 and three nucleotides from E2. Interestingly, while E1 and E2 had been observed in both wolves and domestic dogs, En6 had only been reported in wolves. This suggested that the E haplotypes were likely introduced into the dog population relatively recently, without sufficient time for mutations to form a new haplotype. It is reasonable to assume that dogs harbouring haplotype E are genetically close to wolves. A similar scenario was observed in the haplotype network of haplogroup F, which distributed narrowly among Japanese dogs. Haplotype F3 appeared to be the founding haplotype of this haplogroup and had been recently incorporated into the canine population, likely originating from a wolf (or a small group of wolves) carrying the F3 haplotype.

4. Conclusion

Through the analysis of the haplotype network and assessment of the betweenness value, this study has identified important haplotypes contributing significantly to the dog population. These analyses offer valuable insights into the identification of founder haplotypes involved in the formation of dog breeds worldwide, serving as a valuable reference for breed development and genetic studies.

Compliance with ethical standards

Acknowledgments

This study is financially supported by the Saigon University - Vietnam (Project no. TĐ2020-46).

Disclosure of conflict of interest

The authors declare that there is no conflict of interest.

Statement of ethical approval

The present research work does not contain any studies performed on animals/humans subjects by any of the authors.

References

- [1] Savolainen P, Zhang YP, Luo J, Lundeberg J, Leitner T. Genetic evidence for an East Asian origin of domestic dogs. *Science*. 2002;298(5598):1610-3.
- [2] Pang J-F, Kluetsch C, Zou X-J, Zhang A-b, Luo L-Y, Angleby H, et al. mtDNA data indicate a single origin for dogs south of Yangtze River, less than 16,300 years ago, from numerous wolves. *Molecular biology and evolution*. 2009;26(12):2849-64.
- [3] Smalling BB, Satkoski JA, Tom BK, Szeto WY, Erickson B, Spear T, et al. Geographic differences in mitochondrial DNA (mtDNA) distribution among United States (US) domestic dog populations. *The Open Forensic Science Journal*. 2010;3:22-32.
- [4] Klutsch CF, Seppala EH, Fall T, Uhlen M, Hedhammar A, Lohi H, et al. Regional occurrence, high frequency but low diversity of mitochondrial DNA haplogroup d1 suggests a recent dog-wolf hybridization in Scandinavia. *Animal genetics*. 2011;42(1):100-3.
- [5] Klütch CFC, Savolainen P. *Geographical Origin of the Domestic Dog*. eLS: John Wiley & Sons, Ltd; 2011.
- [6] Ding ZL, Oskarsson M, Ardalan A, Angleby H, Dahlgren LG, Tepeli C, et al. Origins of domestic dog in southern East Asia is supported by analysis of Y-chromosome DNA. *Heredity*. 2012;108(5):507-14.
- [7] Wang G-D, Zhai W, Yang H-C, Wang L, Zhong L, Liu Y-H, et al. Out of southern East Asia: the natural history of domestic dogs across the world. *Cell Research*. 2016;26(1):21-33.
- [8] Frantz LA, Mullin VE, Pionnier-Capitan M, Lebrasseur O, Ollivier M, Perri A, et al. Genomic and archaeological evidence suggest a dual origin of domestic dogs. *Science*. 2016;352(6290):1228-31.
- [9] Shannon LM, Boyko RH, Castelhana M, Corey E, Hayward JJ, McLean C, et al. Genetic structure in village dogs reveals a Central Asian domestication origin. *Proceedings of the National Academy of Sciences*. 2015;112(44):13639-44.
- [10] Vila C, Amorim IR, Leonard JA, Posada D, Castroviejo J, Petrucci-Fonseca F, et al. Mitochondrial DNA phylogeography and population history of the grey wolf *canis lupus*. *Molecular ecology*. 1999;8(12):2089-103.
- [11] Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007;23(21):2947-8.