

Phân tích đa dạng di truyền hệ gen ty thể và nguồn gốc tiến hóa của sáu giống lợn bản địa Việt Nam

Bùi Anh Tuấn¹, Nguyễn Đức Hiếu², Nghiêm Ngọc Minh²,
Võ Thị Bích Thủy^{2*}

¹Viện Khoa học hình sự, Bộ Công an

²Viện Nghiên cứu hệ gen, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

Ngày nhận bài 26/3/2018; ngày gửi phản biện 30/3/2018; ngày nhận phản biện 2/5/2018; ngày chấp nhận đăng 7/5/2018

Tóm tắt:

Cây phát sinh chủng loại của 33 giống lợn nhà và lợn hoang thuộc nhánh châu Âu và châu Á, trong đó có 6 giống lợn bản địa Việt Nam đã được dựng lên từ dữ liệu trình tự vùng D-loop và vùng mã hóa của hệ gen ty thể. Lần đầu tiên dữ liệu hoàn chỉnh về hệ gen ty thể của 6 giống lợn Ỉ, Móng Cái, Mường Khương, Mường Lay, Hương và Hạ Lang được công bố trên GenBank với các mã số truy cập KX094894, KU556691, KY432578, KX147101, KY964306 và KY800118. Mục tiêu của nghiên cứu này là xác định trình tự hoàn chỉnh hệ gen ty thể của cả 6 giống lợn, chú giải chức năng hệ gen, phân tích đa hình trình tự mtDNA, qua đó làm cơ sở để nghiên cứu phát sinh chủng loại, xác định về nguồn gốc và quan hệ tiến hóa của các giống lợn này với các giống lợn khác trên thế giới, phục vụ trực tiếp cho mục tiêu bảo tồn nguồn gen. Kết quả dựa trên khảo sát về khoảng cách di truyền và mối quan hệ phát sinh phân tử cho biết mối quan hệ di truyền theo dòng mẹ giữa 6 giống lợn bản địa Việt Nam và những giống lợn bản địa này có mối quan hệ gần gũi với các nhóm lợn Nam Trung Quốc và lưu vực sông Hoàng Hà. Những kết quả nghiên cứu của đề tài là nguồn dẫn liệu quan trọng cho các nghiên cứu khác về các giống lợn bản địa ở Việt Nam.

Từ khóa: Hệ gen ty thể, phát sinh chủng loại, *Sus scrofa*, tiến hóa phân tử.

Chỉ số phân loại: 4.6

Đặt vấn đề

Việt Nam có khoảng trên 20 giống lợn bản địa, trong số đó có những giống thuộc Danh mục nguồn gen vật nuôi quý hiếm cần được bảo tồn [1]. Lợn Ỉ nguồn gốc xuất phát từ tỉnh Nam Định, ngày nay ít được nuôi do hiệu quả kinh tế không cao và đang có nguy cơ tuyệt chủng, đặc biệt, lợn Ỉ được đưa vào Danh mục nguồn gen vật nuôi quý hiếm cần được bảo tồn. Lợn Móng Cái là giống lợn nội được hình thành và phát triển lâu đời, xuất xứ từ thành phố Móng Cái, tỉnh Quảng Ninh, với khả năng sinh khá cao. Lợn Mường Khương là một giống lợn gắn liền với đời sống người H'Mông, được nuôi ở nhiều vùng thuộc tỉnh Lào Cai, nhiều nhất là ở huyện Mường Khương [2]. Đây là một trong ba giống lợn quý ở các tỉnh phía Bắc, cũng là một trong ba giống lợn nội chủ yếu làm nền lợn lai kinh tế ở miền Bắc Việt Nam. Giống lợn Hương được nuôi rộng rãi ở địa bàn biên giới phía Bắc thuộc tỉnh Cao Bằng. Đặc điểm của giống lợn Hương là sinh trưởng chậm hơn các giống khác nhưng thuần thực sớm. Giống lợn này có lớp mỡ mang mùi thơm tự nhiên. Đây là giống có sức đề kháng cao, ít bệnh dịch, dễ

nuôi và không kén thức ăn. Lợn Mường Lay được chăn nuôi chủ yếu ở địa bàn thị xã Mường Lay, tỉnh Điện Biên, đây là giống lợn phàm ăn, thích nghi tốt với điều kiện khắc nghiệt, có tính kháng bệnh tốt. Lợn Hạ Lang phân bố ở tỉnh Cao Bằng, cũng như các giống lợn bản địa khác, quần thể lợn Hạ Lang cũng đang ngày càng bị thu hẹp do áp lực của các giống nhập nội. Các giống lợn truyền thống Việt Nam rất nhiều mỡ và sinh trưởng chậm. Do vậy, chúng không phù hợp với nhu cầu phát triển kinh tế. Hiện nay, ở Việt Nam cũng như trên thế giới có xu hướng nuôi lợn lai nhập nội, dẫn đến nguy cơ mất dần đi những giống lợn bản địa. Ngày càng nhiều giống lợn có quy mô quần đàn ở mức bị đe dọa, chủ yếu là do sức ép của quá trình sản xuất lợn với quy mô toàn cầu, buộc người nông dân phải chọn lựa nuôi một số ít giống lợn phổ biến cho hiệu quả kinh tế cao [3].

DNA ty thể (mtDNA) đã được sử dụng rộng rãi trong các nghiên cứu phát sinh chủng loại vì một số lý do sau: Thứ nhất, sự tiến hóa của mtDNA ở động vật có vú xảy ra trước tiên do sự thay thế từng cặp base đơn, chỉ một tỷ lệ hiếm là các trình tự có sự tái sắp xếp [4]. Thứ hai, tốc độ tiến hóa mtDNA xuất hiện nhanh hơn 10 lần so với DNA

*Tác giả liên hệ: thuybvo@ibt.ac.vn

Genetic diversity of mitochondrial genome and evolutionary origin of six Vietnamese indigenous pig breeds

Anh Tuan Bui¹, Duc Hieu Nguyen²,
Ngoc Minh Nghiem², Thi Bich Thuy Vo^{2*}

¹Institute of Forensic Science, Ministry of Public Security

²Institute of Genome Research, Vietnam Academy of Science and Technology

Received 26 March 2018; accepted 7 May 2018

Abstract:

The phylogenetic trees of 33 domestic and wild boar pig breeds of Asian and European clades, including six Vietnamese indigenous pig breeds were reconstructed from D-loop region and coding region sequence using the Bayesian inference method. It is the first time the complete mitochondrial genome of the I, Mong Cai, Muong Khuong, Muong Lay, Huong, and Ha Lang pig breeds was sequenced and deposited on GenBank (accession numbers: KX094894, KU556691, KY432578, KX147101, KY964306, and KY800118, respectively). The genetic distances and phylogenetic relationships, based on both the mtDNA and the D-loop region, revealed that all of six Vietnamese pig breeds belonged to Asian clade with the close evolutionary relationship to other pig breeds from South China and Chinese Yellow River Valley. The publication of the mitochondrial genome sequences will make an important contribution to elucidating the relationship among Vietnamese indigenous pig breeds and supporting the selection of suitable ones for pig breeding in the area.

Keywords: Mitochondrial genome, molecular evolution, phylogenetic relationship, *Sus scrofa*.

Classification number: 4.6

gen nhân [5]. Thứ ba, mtDNA được di truyền theo dòng mẹ, đơn bội và không có sự tái tổ hợp [6]. Vì những lý do này nên mtDNA được sử dụng như một công cụ cho việc xác định các mối quan hệ giữa các cá thể trong cùng loài và trong số các loài có quan hệ gần với khoảng thời gian phân ly mới gần đây [5].

Việt Nam và khu vực Nam Trung Quốc được cho là một trong những trung tâm thuần hóa sớm nhất của các giống lợn nhà [7]. Giả thiết của Hongo và cộng sự về nguồn gốc của lợn bản địa Việt Nam có thể là hậu duệ của các con lợn rừng và lợn nhà từ một số khu vực ở châu Á, trong đó có một số vùng của Trung Quốc [8]. Nhóm nhà khoa học Nhật Bản và Việt Nam đã sử dụng trình tự phân đoạn (574 bp) mtDNA của một số giống lợn rừng và lợn nhà Việt Nam, so sánh với giống lợn rừng Ryukyu của Nhật Bản. Kết quả của nghiên cứu này chỉ ra rằng, con cháu của tổ tiên giống Ryukyu vẫn cư trú tại Việt Nam, trình tự mtDNA của các giống lợn nhà Việt Nam có độ phân ly rất lớn [9]. Mục đích của nghiên cứu này là giải trình tự hoàn chỉnh hệ gen ty thể, dự đoán cấu trúc, phân tích dữ liệu phân tử, xác định mối quan hệ phát sinh chủng loại sử dụng sự đa hình về trình tự mtDNA, qua đó đánh giá về nguồn gốc tiến hóa của 6 giống lợn bản địa Việt Nam.

Vật liệu và phương pháp nghiên cứu

Vật liệu

Máu ngoại vi được thu thập từ các cá thể lợn được lựa chọn ngẫu nhiên thuộc các giống lợn bản địa dựa trên thông tin điều tra về phả hệ của Viện Chăn nuôi quốc gia.

Trình tự của các giống lợn châu Âu, châu Á được tải về từ GenBank và được xếp vào 5 khu vực địa lý chính [10]: Đông Bắc Á, Khu vực Mekong, Lưu vực sông Hoàng Hà, Nam Trung Quốc, Lưu vực sông Dương Tử và các quốc gia châu Âu.

Phương pháp

Phương pháp tách chiết và nhân bội DNA: DNA tổng số được tách chiết bằng bộ kit GeneJET™ Whole Blood Genomic DNA Purification Mini Kit (Thermo Fisher Scientific, Mỹ). Toàn bộ trình tự mtDNA được khuếch đại bằng PCR với 30 cặp mỗi (bảng 1). Tổng thể tích PCR 25 µl: 12,5 µl GoTaq® Green Master Mix (Promega, Madison, WI, USA), 1,0 µl (20 ng/µl) DNA khuôn, 0,5 µl mỗi mỗi (10 mmol/l), và 10,5 µl dH₂O. Chu trình nhiệt: Biến tính ở 94°C trong 5 phút, 25 chu kỳ: 30 giây ở 94°C, 30 giây ở 53-55°C, 30 giây ở 72°C, và 8 phút ở 72°C. Sản phẩm

Bảng 1. 30 cặp mồi sử dụng cho PCR và vị trí các phân đoạn được khuếch đại.

TT	Trình tự mồi (5'-3')		T _a (°C)	Vị trí các amplicon
	Mồi xuôi	Mồi ngược		
D-loop	AGGAGACTAATCCGCCAT	GCGGATACTTGCATGTGT	54	16556-1306
1	ACTAAGTCAATGCCTATTCTG	CAATGTATGAAACCTCAG	54	16298-548
2	CTACACAATAACCTCCCATA	TGGCAGGAGATTACCAACT	54	1107-1490
3	GCTCATAAGCCCTTGCTC	ATTCTTTCATCTTCCCTT	54	1377-2415
4	CACAACCATGCAAGAAGAGACA	ACAACCGATCATCACCAGGC	54	2141-2634
5	CCGTAAGGGAAGATGAAAG	TATGGTTATTTGACTGGT	54	2393-3493
6	CCGTGCAAGGTAGCATA	CCAACATCGAGGTCGTAA	55	3189-3606
7	TGGGGTGACCTCGGAGTAC	AATATGGCGAAAGGTCCGG	54	3423-4589
8	CGAGCAGTAGCCCAACA	GGTCGTATCGGAATCGTG	55	4321-4771
9	GTATCAGGCTTTAACGTAGA	TGGTAATACTGCTGTCATTC	55	4543-5671
10	CACAGAAGCAGCCACAAA	ATGGGATAGGATAAAGT	55	5242-5782
11	ACATAGGATGAATGACAGC	TGGTGGAGTAGTCAGAAAC	55	5643-6831
12	GCACTGCCTTGAGCCTAC	GTGTTGAGTTGCGGTCT	55	6599-7160
13	CCCATTATGATGGGGTTT	TGCTGTGTATGCGTCAGGAT	55	6724-7857
14	CACCTTGTAAATCATATCGTAG	TAGTTGGAAAGGTAAGC	53	7747-8223
15	TTATCTCACTAACAGCAG	TTGAGTTCGGTTGATTCTG	55	7885-9086
16	GCTTCATGCCAATGTAC	TTATAGCGGAATCCTGTG	55	8816-9478
17	GCAAGCCAGAAATCAACCG	CGAGGAGGATTGAGGTGTT	55	9061-10214
18	ATACCACATAGTAAACCAA	CCTGTAGCCACAAGAAA	55	9820-10404
19	CTAAACACCTCAATCCTCC	TTGGACGTAATCGGTACCG	55	10193-11341
20	CCTTCAGGGTTACTTAT	TTCGGGTTGTGGTTTCTT	53	11113-11632
21	CGGTACCGATTACGTCAA	CCGATTAGATTGATGATG	55	11323-12488
22	ACCAGCTCTACTGCTTA	GAGGCTTGTATGTTTGA	55	12172-12644
23	ATGATGACTAATAGCAAGCC	GGGATGTAGTCCGAATTG	55	12429-13627
24	CATCGGAGACATTGGATT	AGTTGGCTTGAAGTTGAG	55	13462-13863
25	CCTACTCTAGCTGCAGCAG	ATTATGGAGATTACTCGTGG	55	13579-14765
26	TCCGATCATCATTACTA	TTTATGGTGGACTTGGGT	55	14576-15187
27	TAATTACCACGAGTAATCTC	TTCTACGAGTCTGTTCCG	55	14740-15827
28	GGAGCATCCATATCTTT	GGTGTAGTTGTCTGGGTCT	53	15597-16112
29	TCGTAGAATGAATCTGAGG	GGTGATACGCATGTTGACTG	55	15820-301

PCR được tinh sạch, sau đó giải trình tự bằng phương pháp Sanger [11] trên máy phân tích gen ABI 3500 Genetic Analyzer (Applied Biosystems, Mỹ).

Phương pháp lắp ghép, giống hàng trình tự, phân tích và chú giải hệ gen: Trình tự vùng D-loop và vùng mã hóa được lắp ghép bằng sử dụng phần mềm EditSeq (DNASTAR Inc., Madison, WI, Mỹ) [12] và phần mềm DNADragon v.1.6.0

(SequentiX, Đức). Trình tự D-loop và trình tự mã hóa của toàn bộ hệ gen ty thể của 6 giống lợn bản địa Việt Nam, các giống lợn châu Âu và giống lợn thuộc nhóm châu Á được giống hàng đa trình tự, sử dụng thuật toán MUSCLE [13]. Phân tích và chú giải hệ gen của các giống lợn bản địa Việt Nam bằng công cụ trực tuyến Dogma và MITOS Web Server [14, 15]. Tất cả các chú giải được kiểm tra bằng công cụ BLAST trên GenBank [16, 17]. Toàn bộ 22 gen tRNA được dự đoán cấu trúc bậc hai sử dụng công cụ trực tuyến MITOS Web Server.

Phương pháp xác định đa hình trình tự và phân tích phát sinh chủng loại: Xác định các vị trí đa hình, vị trí SNP trên trình tự D-loop của các giống lợn được tiến hành sử dụng phần mềm DnaSP v6. [18] và SeqMan v7.1.0 (DNASTAR Inc., Madison, WI, Mỹ). Khoảng cách di truyền được tính toán sử dụng thuật toán hai thông số của Kimura trong phần mềm MEGA [13].

Trình tự của vùng D-loop và toàn bộ vùng mã hóa của hệ gen ty thể được sử dụng riêng biệt làm dữ liệu đầu vào để dựng cây phát sinh chủng loại. Phân tích phát sinh phân tử được tiến hành dựa trên dữ liệu rời rạc sử dụng phương pháp suy luận Bayes theo mô hình Hasegawa-Kishino-Yano [19]. Xác suất hậu nghiệm của cây được tính với chuỗi Markov Chain Monte Carlo (MCMC) 10000000 [20] sử dụng phần mềm BEAST v1.8.3 [21]. Sau đó, cây tốt nhất sẽ được tìm ra bằng phần mềm Tree Annotater v.1.8.4. Cuối cùng, phần mềm Figure Tree v1.4.2 được sử dụng để đọc tệp tin kết xuất cho việc vẽ cây phát sinh. Cây được xác định gốc sử dụng trình tự tương đồng đối chứng của lợn hoang Malaysia (*Sus barbatus*).

Kết quả

Thành phần, cấu trúc của hệ gen

Thành phần hệ gen ty thể hoàn chỉnh của 6 giống lợn bản địa gồm Ỉ, Móng Cái, Mường Khương, Mường Lay, Hương và Hạ Lang có 13 gen mã protein, 22 gen RNA vận chuyển (tRNA), 2 gen RNA ribosome (rRNA) và các vùng không mã hóa. Toàn bộ 22 gen tRNA của cả 6 giống bản địa đã được dự đoán có chung cấu trúc bậc hai (cỏ ba lá) với các gen tRNA có chiều dài từ 59 đến 75 bp. Cấu trúc hệ gen mtDNA của các giống lợn bản địa đã được chú thích rõ (bảng 2). Thành phần và cấu trúc hệ gen ty thể của 6 giống lợn bản địa Ỉ, Móng Cái, Mường Khương, Mường Lay, Hương và Hạ Lang tương tự như của các giống lợn nhà khác [22].

Bảng 2. Cấu trúc hệ gen ty thể của 6 giống lợn bản địa Việt Nam.

Gene	Codon			Chức năng	Vị trí											
	Start	Stop	Anti-codon		Hạ Lang		Hương		Mường Lay		Í		Móng Cái		Mường Khương	
					Start	Stop	Start	Stop	Start	Stop	Start	Stop	Start	Stop	Start	Stop
D-loop				H	1	1285	1	1315	1	1304	1	1295	1	1275	1	1244
tRNA <i>Phe</i>			GAA	H	1286	1355	1316	1385	1305	1374	1296	1365	1276	1345	1245	1314
12S rRNA				H	1356	2318	1386	2349	1375	2336	1366	2325	1346	2305	1315	2274
tRNA <i>Val</i>			TAC	H	2318	2385	2349	2416	2336	2403	2327	2394	2307	2374	2276	2343
16S rRNA				H	2384	3955	2415	3986	2402	3972	2395	3964	2375	3944	2344	3912
tRNA <i>Leu2</i>			TAA	H	3956	4030	3987	4061	3973	4047	3965	4039	3945	4019	3913	3987
ND1	ATG	TAG		H	4033	4989	4064	5020	4056	5000	4042	4998	4022	4978	3990	4946
tRNA <i>Ile</i>			GAT	H	4988	5056	5019	5087	5005	5073	4997	5065	4977	5045	4945	5013
tRNA <i>Gln</i>			TTG	L	5054	5126	5085	5157	5071	5143	5063	5135	5043	5115	5011	5083
tRNA <i>Met</i>			CAT	H	5128	5197	5159	5228	5145	5214	5137	5206	5117	5186	5085	5154
ND2	ATA	TAG		H	5198	6241	5229	6272	5215	6253	5207	6250	5187	6230	5155	6198
tRNA <i>Trp</i>			TCA	H	6240	6307	6271	6338	6258	6325	6249	6316	6229	6296	6197	6264
tRNA <i>Ala</i>			TGC	L	6314	6381	6345	6412	6332	6399	6323	6390	6303	6370	6271	6338
tRNA <i>Asn</i>			GTT	L	6383	6457	6414	6488	6401	6475	6392	6466	6372	6446	6340	6414
tRNA <i>Cys</i>			GCA	L	6490	6555	6521	6586	6508	6573	6499	6564	6479	6544	6447	6512
tRNA <i>Tyr</i>			GTA	L	6556	6620	6587	6651	6574	6638	6565	6629	6545	6609	6513	6577
COX1	ATG	TAA		H	6622	8166	6653	8197	6640	8178	6631	8175	6611	8155	6579	8123
tRNA <i>Ser2</i>			TGA	L	8170	8238	8201	8269	8188	8256	8179	8247	8159	8227	8127	8195
tRNA <i>Asp</i>			GTC	H	8246	8313	8277	8344	8264	8331	8255	8322	8235	8302	8203	8270
COX2	ATG	T--		H	8314	9001	8345	9032	8332	9012	8323	9010	8303	8990	8271	8958
tRNA <i>Lys</i>			TTT	H	9002	9068	9033	9099	9020	9086	9011	9077	8991	9057	8959	9025
ATPase8	ATG	TAA		H	9070	9273	9101	9304	9088	9282	9079	9282	9059	9262	9027	9230
ATPase6	ATG	TAA		H	9231	9911	9262	9942	9249	9923	9240	9920	9220	9900	9188	9868
COX3	ATG	T--		H	9911	10694	9942	10725	9929	10711	9920	10703	9900	10684	9868	10651
tRNA <i>Gly</i>			TCC	H	10695	10763	10726	10794	10713	10782	10704	10772	10684	10752	10652	10720
ND3	ATA	T--		H	10764	11109	10795	11140	10783	11127	10773	11118	10753	11099	10721	11066
tRNA <i>Arg</i>			TCG	H	11111	11179	11142	11210	11130	11198	11120	11188	11100	11168	11068	11136
ND4l	GTG	TAA		H	11180	11476	11211	11507	11214	11492	11189	11485	11169	11465	11137	11433
ND4	ATG	T--		H	11470	12847	11501	12878	11489	12856	11479	12856	11459	12836	11427	12804
tRNA <i>His</i>			GTG	H	12848	12916	12879	12947	12867	12935	12857	12925	12837	12905	12805	12873
tRNA <i>Ser1</i>			GCT	H	12917	12975	12948	13006	12936	12994	12926	12984	12906	12964	12874	12932
tRNA <i>Leu1</i>			TAG	H	12976	13045	13007	13076	12995	13064	12985	13054	12965	13034	12933	13002
ND5	ATA	TAA		H	13046	14866	13077	14897	13065	14880	13055	14875	13035	14855	13003	14823
ND6	ATG	TAA		L	14853	15380	14881	15408	14876	15391	14859	15386	14842	15369	14807	15334
tRNA <i>Glu</i>			TTC	L	15378	15446	15409	15477	15398	15466	15387	15455	15367	15435	15335	15403
Cytb	ATG	AGA		H	15451	16590	15482	16621	15471	16604	15460	16599	15440	16579	15408	16547
tRNA <i>Thr</i>			TGT	H	16591	16658	16622	166689	16611	16678	16600	16667	16580	16647	16548	16615
tRNA <i>Pro</i>			TGG	L	16658	16722	16689	16753	16678	16742	16667	16731	16647	16711	16615	16679

rRNA: ribosomal RNA; 16S rRNA: rRNA tiểu phần lớn; 12S rRNA: rRNA tiểu phần nhỏ; tRNA: RNA vận chuyển và các từ in nghiêng là mã của các amino acid; ND1-6 và ND4l: genes mã hóa nicotinamide dinucleotide dehydrogenase các tiểu phần 1 đến 6 và 4l; ATPase6 và 8: các gene mã hóa adenosine triphosphatase tiểu phần 6 và 8; COX1 đến 3: các gene mã hóa các tiểu phần cytochrome c oxidase I đến III; Cytb: gene mã hóa cytochrome b. T-- thể hiện ở bộ ba tận cùng không hoàn thiện; H, L: tương ứng là sợi nặng và sợi nhẹ.

Thành phần các loại nucleotide A, C, G, và T trong trình tự hệ gen của 6 giống lợn được liệt kê ở bảng 3.

Bảng 3. Tỷ lệ thành phần các loại base trong trình tự hệ gen ty thể của 6 giống lợn bản địa Việt Nam.

	A (%)	C (%)	G (%)	T (%)	G+C (%)
Ỉ	34,64	26,24	13,33	25,75	39,57
Móng Cái	34,70	26,20	13,30	25,79	39,50
Mường Khương	34,68	26,19	13,31	25,81	39,50
Hạ Lang	34,67	26,20	13,32	25,78	39,55
Hương	34,65	26,22	13,352	25,78	39,57
Mường Lay	34,70	26,19	13,32	25,79	39,51

Phân tích đa dạng di truyền

Độ đa dạng di truyền đối với trình tự vùng D-loop của các giống lợn bản địa Việt Nam được sử dụng trong nghiên cứu này có 25 vị trí đa hình trên tổng số 1184 vị trí (chiều dài contig), trong đó có 5 vị trí có mặt ít nhất 2 biến thể (Ỉ với 3 vị trí, Hạ Lang và Mường Khương mỗi giống có 1 vị trí) (xem bảng 4).

Bảng 4. Các vị trí SNP trình tự vùng D-loop của 6 giống lợn bản địa Việt Nam.

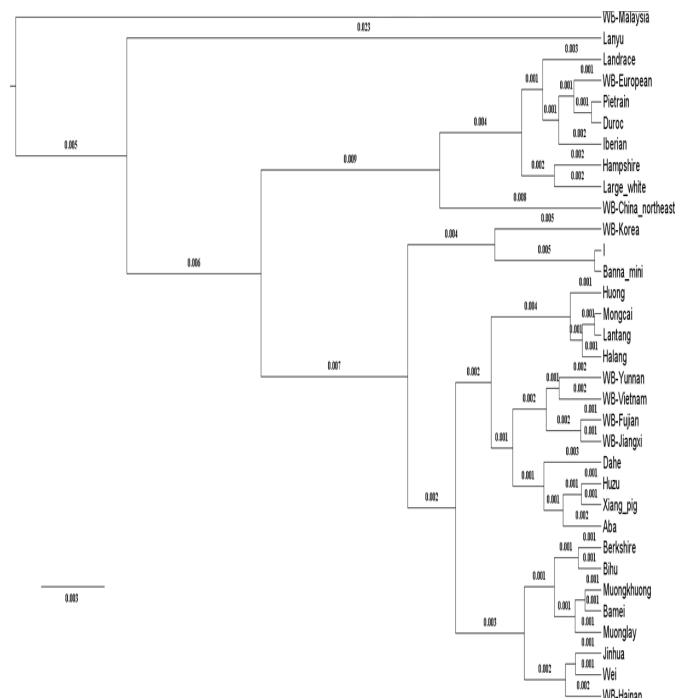
STT	Giống	Ỉ	Mường Khương	Mường Lay	Móng Cái	Hạ Lang	Hương
Vị trí							
1	24	G/A	-	-	-	-	-
2	183	T/C	-	-	-	-	-
3	215	T/C	-	-	-	-	-
4	242	T/C	T/C	T/C	T/C	-	-
5	280	-	C/T	C/T	-	-	-
6	407	-	-	-	T/C	T/C	T/C
7	454	-	C/T	C/T	C/T	C/T	C/T
8	503	A/G	-	-	-	-	-
9	562	T/C	-	-	-	-	-
10	706	G	G	-	G	-	-
11	714	A	-	A	-	-	-
12	736	-	-	A	-	-	-
13	734	G/G/A	-	-	-	-	-
14	744	A	-	-	-	-	-
15	746	-	-	A	-	-	-
16	754	G/G/A	-	-	-	-	-
17	756	-	-	A	-	-	-
18	764	G/G/A	-	-	-	-	-
19	766	-	-	A	-	-	-
20	774	A	-	-	-	-	-
21	776	-	-	A	-	-	-
22	791	C	-	-	-	-	-
23	813	-	T/T/C	-	-	-	-
24	1032	A/G	-	-	-	-	-
25	1165	-	-	-	-	A/C/A	-

"T/C": vị trí có 1 biến thể thay thế nucleotide; "G/G/A": vị trí có mặt ít nhất 2 biến thể; "-": vị trí không có đa hình.

Phân tích về quan hệ phát sinh chủng loại

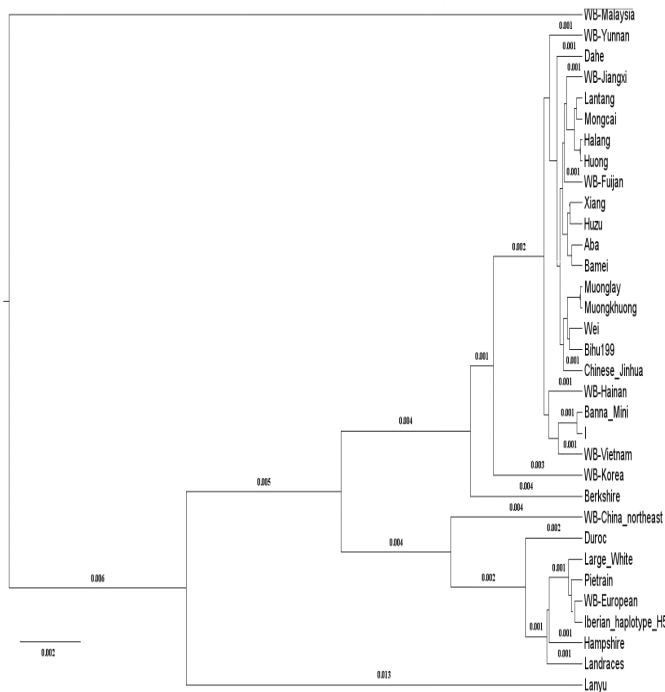
Trình tự của giống lợn hoang Malaysia (WB-Malasia) được chọn lựa là nhóm ngoại (outgroup) bởi nó được biết đến là có sự khác biệt với nhóm lợn hoang châu Âu và châu Á, thường được sử dụng trong các nghiên cứu trước đây về phát sinh chủng loại ở lợn [10, 23]. Phân tích về phát sinh chủng loại của trình tự vùng D-loop (hình 1) và trình tự mtDNA hoàn chỉnh (hình 2) thể hiện có 3 nhánh chính riêng biệt (một nhánh châu Âu, hai nhánh châu Á). Nhánh châu Âu bao gồm các giống lợn kiểu Âu như: Berkshire, Duroc, Hampshire, Landrace, Pietrain Large, White Iberian, WB-European. Nhánh châu Á 1 gồm chủ yếu các con lợn nhà Trung Quốc, nhánh châu Á 2 tập hợp chủ yếu các giống lợn hoang châu Á. Hai giống lợn bản địa của Việt Nam là Hương và Hạ Lang cùng nhóm với lợn Lantang ở miền nam Trung Quốc. Giống Mường Lay cùng nhánh phụ với giống Bamei ở lưu vực sông Hoàng Hà, Trung Quốc.

Khoảng cách di truyền trung bình tương đối gần trong nhóm các con lợn Việt Nam là $0,0039 \pm 0,00112$ (trị số trung bình $\pm$ độ lệch chuẩn) so với khoảng cách di truyền trung bình chung của nhóm lợn Việt Nam với các giống lợn châu Âu, châu Á là $0,01279 \pm 0,00196$. Phân tích khoảng cách cặp giữa các giống lợn cho thấy, lợn Mường Lay có khoảng cách gần nhất với lợn Bamei (0,00104) của Trung



Hình 1. Quan hệ phát sinh chủng loại được phân tích sử dụng phương pháp suy luận Bayes bằng phần mềm BEAST v1.8.3 [21]. Cây phát sinh chủng loại được dựng lên sử dụng phần mềm Tree Annotator, thông qua việc so sánh các trình tự ở vùng kiểm soát của hệ gen ty thể của 6 giống lợn bản địa Việt Nam và các giống lợn châu Âu, châu Á.

Quốc, lợn Hương và Hạ Lang có khoảng cách di truyền bằng 0,0000 (có sự đồng nhất 100% về trình tự). Lợn Hạ Lang và Hương có khoảng cách di truyền ngắn nhất với lợn Lantang (0,00104). Khoảng cách di truyền của giống lợn Ỉ là xa nhất so với 5 giống lợn bản địa Việt Nam còn lại, cụ thể là giống lợn Ỉ có khoảng cách di truyền với Hương và Hạ Lang là 0,0081; khoảng cách với 3 giống Móng Cái, Mùong Khương và Mùong Lay là 0,00782. Quan hệ giữa các giống lợn này được thể hiện rõ trên cây phát sinh chủng loại ở phần kết quả dưới đây, khi giống lợn Ỉ nằm phân nhánh xa hơn so với các giống còn lại của Việt Nam.



Hình 2. Quan hệ phát sinh chủng loại được phân tích sử dụng phương pháp suy luận Bayes bằng phần mềm BEAST v1.8.3 [21]. Cây phát sinh chủng loại được dựng lên sử dụng phần mềm Tree Annotator, thông qua việc so sánh các trình tự ở vùng mã hóa hoàn chỉnh của hệ gen ty thể của 6 giống lợn bản địa Việt Nam và các giống lợn châu Âu, châu Á.

Bàn luận

Phân tích cấu trúc, tổ chức hệ gen ty thể của 6 giống lợn bản địa Việt Nam là Ỉ, Móng Cái, Mùong Khương, Mùong Lay, Hương và Hạ Lang cho thấy có sự tương đồng với cấu trúc hệ gen ty thể của các loài động vật có vú khác. Thành phần các loại base có trong hệ gen ty thể của cả 6 giống lợn đều theo hướng giàu A+T (trên 60%), có sự tương đồng với các nhóm lợn châu Á khác [22]. Sự phân bố của hầu hết các gen đều nằm trên chuỗi H, ngoại trừ gen ND6 và tám gen tRNA nằm trên chuỗi L.

Cả hai cây phát sinh chủng loại đều cho thấy sự tách biệt rõ ràng giữa hai nhánh lợn châu Âu và châu Á, hai nhánh đã có sự phân ly từ thời gian khá lâu (khoảng 746000 năm

trước) [24] so với thời điểm các giống lợn nhà được thuần hóa là khoảng 9000 năm trước đây. Các giống lợn bản địa Việt Nam đều có mối quan hệ về dòng mẹ gần nhau so với các giống lợn châu Á. Kết quả xác định khoảng cách di truyền và phân tích về quan hệ phát sinh chủng loại cho thấy có sự tương đồng với các nghiên cứu khác về nguồn gốc của lợn bản địa Việt Nam có liên quan đến lợn châu Á [24]. Cây phát sinh đã chỉ ra các giống lợn bản địa làm đối tượng nghiên cứu ở cùng các nhánh phụ với các giống lợn Nam Trung Quốc và lưu vực sông Hoàng Hà với khoảng cách di truyền tương đối gần. Điều này có thể đưa ra một nhận định tương đối vững chắc về nguồn gốc của 6 giống lợn bản địa Việt Nam là bắt nguồn từ lợn châu Á, có thể từ lợn Trung Quốc. Lợn Móng Cái, Hương và Hạ Lang nằm cùng một nhánh phụ với khoảng cách di truyền gần, điều này phù hợp với sự tương đồng về đặc điểm phân bố địa lý và đặc điểm về hình thái học. Lợn Mùong Lay và Mùong Khương ở các nhánh chị em có khoảng cách di truyền rất gần (0,0005). Giống lợn Ỉ có khoảng cách xa nhất với các giống lợn bản địa Việt Nam còn lại và có khoảng cách di truyền gần nhất với lợn Banamini. Hai giống lợn bản địa Hương và Hạ Lang cùng nhóm với lợn Lantang ở miền nam Trung Quốc. Giống Mùong Lay cùng nhánh phụ với giống Bamei ở lưu vực sông Hoàng Hà, Trung Quốc. Như dự đoán, có khoảng cách lớn giữa lợn Việt Nam - Trung Quốc với nhóm lợn châu Âu tương đồng với sự phân bố các giống theo vùng địa lý. Nhóm lợn bản địa Việt Nam có mối quan hệ khá gần với các nhóm lợn Trung Quốc, điều này đưa tới một giả thiết về sự di cư gần liền với các hoạt động giao thương của hai nước có chung đường biên giới với lịch sử lâu đời.

Hai giống bản địa của tỉnh Cao Bằng là Hạ Lang và Hương có độ tương đồng về trình tự D-loop là 100%, phù hợp với một số đặc điểm tương đồng giữa hai giống lợn về hình thái, bên cạnh đó đã có một số tài liệu đề cập đến nguồn gốc gần gũi hoặc có chung nguồn gốc của hai giống lợn này [25, 26]. Thông qua phân tích phát sinh chủng loại có thể nhận định rằng hai giống Hạ Lang và Hương có chung nguồn gốc tổ tiên, phân ly ở cùng một thời điểm tiến hóa. Tuy nhiên, cần có những nghiên cứu sâu hơn về tiến hóa, có thể sử dụng các chỉ thị di truyền để làm sáng tỏ luận điểm trên.

Kết luận

Kết quả nghiên cứu xác định dữ liệu hệ gen ty thể hoàn chỉnh và phân tích về quan hệ phát sinh chủng loại của 6 giống lợn là Ỉ, Móng Cái, Mùong Khương, Mùong Lay, Hương và Hạ Lang là nguồn cơ sở dữ liệu quan trọng trong các nghiên cứu về phát sinh chủng loại, tiến hóa phân tử, các nghiên cứu khác nhằm nhận diện, đánh giá và sử dụng giống lợn bản địa Việt Nam, từ đó đóng góp một cách hiệu quả cho việc bảo tồn và sử dụng nguồn gen này.

TÀI LIỆU THAM KHẢO

- [1] Bộ Nông nghiệp và Phát triển nông thôn (2005), *Danh mục nguồn gen vật nuôi quý hiếm cần bảo tồn*, Quyết định số 88/2005/QĐ-BNN ngày 27/12/2005.
- [2] T.Q. Dang Nguyen, N.K. Tich, B.X. Nguyen, M. Ozawa, K. Kikuchi, N. Manabe, J. Ratky, Y. Kanai, T. Nagai (2010), "Introduction of various vietnamese indigenous pig breeds and their conservation by using assisted reproductive techniques", *Journal of Reproduction and Development*, **56**, pp.5-31.
- [3] B.M. Epstein (1986), "Pig", *Evolution of Domesticated Animals*, John Wiley & Sons, Incorporated, pp.145-162.
- [4] D.R. Wolstenholme (1992), "Animal mitochondrial DNA: structure and evolution", *International Review of Cytology*, **141**, pp.173-216.
- [5] W.M. Brown, M.Jr. George, A.C. Wilson (1979), "Rapid evolution of animal mitochondrial DNA", *Proceedings of the National Academy of Sciences of the United States of America*, **76**, pp.1967-1971.
- [6] J.C. Avise (1993), "Molecular Markers, Natural History and Evolution", *Chapman and Hall, New York*,
- [7] P.J. Piper, H. Matsumura, D. Bulbeck (2017), "New perspectives in Southeast Asian and Pacific prehistory", *Acton ACT: ANU Press*, <http://press-files.anu.edu.au/downloads/press/n2320/pdf/book.pdf?referer=2320>.
- [8] H. Hongo, N. Ishiguro, T. Watanobe, N. Shigehara, T. Anezaki, V.T. Long, D.V. Binh, N.T. Tien, N.H. Nam (2002), "Variation in mitochondrial DNA of Vietnamese pigs: relationships with Asian domestic pigs and Ryukyu wild boars", *Zoological Science*, **19**, pp.1329-1335.
- [9] N. Ishiguro, M. Sasaki, M. Iwasa, N. Shigehara, H. Hongo, T. Anezaki, V.T. Long, D.T.B. Lan, P.T. Long (2008), "mtDNA variation in Vietnamese pigs, with particular emphasis on the genetic relationship between wild boars from Vietnam and the Ryukyu Islands", *Mammal Study*, **33**, pp.51-58.
- [10] G. Yu, H. Xiang, J. Wang, X. Zhao (2013), "The phylogenetic status of typical Chinese native pigs: analyzed by Asian and European pig mitochondrial genome sequences", *Animal Science and Biotechnology*, **4**, pp.4-9.
- [11] F. Sanger, S. Nicklen, A.R. Coulson (1977), "DNA sequencing with chain-terminating inhibitors", *Proceedings of The National Academy of Sciences*, **74**, pp.5463-5467.
- [12] J. Hein, J. Stovlbaek (1996), "Combined DNA and protein alignment", *Methods in Enzymology*, **266**, pp.402-418.
- [13] S. Kumar, G. Stecher, K. Tamura (2016), "MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets", *Molecular Biology and Evolution*, **33**, pp.1870-1874.
- [14] M. Bernt, A. Donath, F. Jühling, F. Externbrink, C. Florentz, G. Fritzsch, J. Pütz, M. Middendorf, P.F. Stadler (2013), "MITOS: Improved de novo metazoan mitochondrial genome annotation", *Molecular Phylogenetics and Evolution*, **69**, pp.313-319.
- [15] T. Seemann (2014), "Prokka: rapid prokaryotic genome annotation", *Bioinformatics*, **30**, pp.2068-2069.
- [16] S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Research*, **25**, pp.3389-3402.
- [17] D.A. Benson, M. Cavanaugh, K. Clark, I. Karsch-Mizrachi, D.J. Lipman, J. Ostell, E.W. Sayers (2013), "GenBank", *Nucleic Acids Research*, **41**, pp.D36-D42.
- [18] J. Rozas, A. Ferrer-Mata, J.C. Sanchez-DelBarrio, S. Guirao-Rico, P. Librado, S.E. Ramos-Onsins, A. Sanchez-Gracia (2017), "DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets", *Molecular Biology and Evolution*, **34**, pp.3299-3302.
- [19] M. Hasegawa, H. Kishino, T.A. Yano (1985), "Dating of the human-ape splitting by a molecular clock of mitochondrial DNA", *Journal of Molecular Evolution*, **22**, pp.160-174.
- [20] J.P. Huelsenbeck, F. Ronquist (2001), "MRBAYES: Bayesian inference of phylogenetic trees", *Bioinformatics*, **17**, pp.754-755.
- [21] A.J. Drummond, M.A. Suchard, D. Xie, A. Rambaut (2012), "Bayesian phylogenetics with BEAUti and the BEAST 1.7", *Molecular Biology and Evolution*, **29**, pp.1969-1973.
- [22] L.Y. Wang, Y.L. Chai, H.M. Ma (2016), "The complete sequence of the mitochondrial genome of Duroc pig (Sus Scrofa)", *Mitochondrial DNA. Part A, DNA Mapping, Sequencing, and Analysis*, **27**, pp.3-4.
- [23] G.S. Wu, Y.G. Yao, K.X. Qu, Z.L. Ding, H. Li, M.G. Palanichamy, Z.Y. Duan, N. Li, Y.S. Chen, Y.P. Zhang (2007), "Population phylogenomic analysis of mitochondrial DNA in wild boars and domestic pigs revealed multiple domestication events in East Asia", *Genome Biology*, **8**, pp.1-12.
- [24] E. Giuffra, J.M. Kijas, V. Amarger, O. Carlborg, J.T. Jeon, L. Andersson (2000), "The origin of the domestic pig: independent domestication and subsequent introgression", *Genetics*, **154**, pp.1785-1791.
- [25] Đ.T. Năm (2005), *Nuôi thử nghiệm giống lợn Hương quý hiếm của Trung Quốc tại Cao Bằng*, <http://khcncaobang.gov.vn>.
- [26] Bộ Khoa học và Công nghệ (2016), *Quyết định số 1011/QĐ-BKHCN ngày 4/5/2016 về việc phê duyệt danh mục đặt hàng nhiệm vụ quý gen cấp quốc gia xét giao trực tiếp bắt đầu thực hiện từ năm 2016*, <http://thuvienphapluat.vn>.