

TRÌNH TỰ HOÀN CHỈNH CỦA GENOME TY THỂ MANG HAPLOTYPE HIẾM E4 CỦA CHÓ LƯNG XOÁY PHÚ QUỐC

Trần Hoàng Dũng[✉], Trương Nguyễn Thị Như Mai, Nguyễn Thành Công, Huỳnh Văn Hiếu

Đại học Nguyễn Tất Thành

[✉] Người chịu trách nhiệm liên lạc. E-mail: tranhoangdung1975@yahoo.com

Ngày nhận bài: 21.5.2016

Ngày nhận đăng: 28.12.2016

TÓM TẮT

Chó lưng có xoáy Phú Quốc là nòi chó quý của Việt Nam. Việc phát hiện chó lưng xoáy Phú Quốc mang haplotype hiếm E1 và E4 với tần suất cao khi đánh giá đa dạng di truyền nòi chó này là một phát hiện bất ngờ. Trong nghiên cứu này kỹ thuật primer-walking đã được áp dụng để giải toàn bộ mtDNA (genome ty thể) của chó Phú Quốc mang haplotype E4, B1 và E1 để tìm kiếm các manh mối nguồn gốc chó Phú Quốc. Về cơ bản chiều dài của các mtDNA đều nằm trong khoảng 16,7 kb cụ thể chó E4 là 16.760 bp và B1 là 16.731 bp không khác biệt so với các nòi chó khác trên thế giới. Sự khác biệt chiều dài giữa mtDNA dòng E4 và B1 chủ yếu do sự khác biệt về số lần lặp lại của các motif lặp lại nằm ở vùng điều khiển. Số lượng các gen, mã mở đầu và kết thúc của mtDNA dòng E4 tương đồng với các haplotype phổ biến khác. Khoảng cách di truyền giữa mtDNA chó Phú Quốc nòi E1, E4 và chó Pusang (EU789662) nằm chung trong 1 cụm trên cây tiến hóa lần lượt là 0,072% và 0,24% cho thấy chó Phú Quốc mang haplotype E có quan hệ rất gần với các chó Pungsan ở Triều Tiên. Mặc dù cùng haplogroup nhưng khoảng cách di truyền của chó Phú Quốc dòng E1 và E4 lại ở mức 0,23%. Trong khi đó khoảng cách di truyền giữa chó Phú Quốc dòng E1, E4 so với các chó sói xám nằm ở vị trí phân kỳ chị em trên phả hệ đồ lần lượt là 0,3, và 0,1%. Genome ty thể chó mang dòng E4 là lần đầu tiên được giải mã và phân tích trên thế giới.

Từ khóa: Chó lưng có xoáy Phú Quốc, khoảng cách di truyền, bộ gen ty thể, haplotype E4-E1, kỹ thuật PCR kích thước lớn, kỹ thuật giải trình tự cuốn chiếu

MỞ ĐẦU

Chó Phú Quốc là một loài chó riêng của đảo Phú Quốc (Việt Nam), có đặc điểm phân biệt với các loài chó khác là có các xoáy lông trên lưng. Đến nay thế giới ghi nhận có 3 nòi chó lưng có xoáy là Phú Quốc, chó lưng xoáy Thái Lan và chó lưng xoáy Châu Phi – Rhodesian. Chó Phú Quốc được giới chuyên môn trên thế giới và trong nước đánh giá rất cao, được xem là một giống chó quý hiếm với nhiều ưu điểm nổi trội mà không có giống chó nào sánh bằng. Quan hệ nguồn gốc giữa chó lưng xoáy Phú Quốc và Thái Lan vẫn còn đang nhiều tranh cãi. Ở Việt Nam chưa có những nghiên cứu chuyên sâu về di truyền chó Phú Quốc.

Trong một công bố trước, Thai *et al.*, (2016) đã phát hiện tính đa dạng di truyền của chó lưng xoáy Phú Quốc rất cao. Theo đó, ba mươi mẫu chó Phú Quốc thu tại huyện đảo Phú Quốc mang 11 haplotype

thuộc 3 haplogroup chính là A, B và E, đặc biệt 16,67% mang haplotype E (bao gồm E1 và E4). Tiếp tục tầm soát với các mẫu chó Phú Quốc thu thập tại Tp Hồ Chí Minh, chúng tôi tiếp tục phát hiện 6/16 cá thể chó lưng xoáy Phú Quốc mang haplotype E (Trần Hoàng Dũng, dữ liệu chưa công bố).

Phát hiện chó Phú Quốc mang haplotype E1 và E4 với tần suất rất cao là một phát hiện thú vị mà các nghiên cứu trước đó của Savolainen *et al* (2002), Pang *et al* (2009) và Oskarsson *et al* (2012) không ghi nhận. Vì haplogroup E chỉ phân bố hạn hẹp ở khu vực Đông Bắc Á với các nòi chó ôn đới như Pungsan (Triều Tiên), chó Jindo (Nhật bản), chó Shar Pei (Trung Quốc) với tỷ lệ cực thấp, chỉ chiếm 1- 2% tổng đàn chó thế giới (Oskarsson *et al.*, 2012). Do đó liệu chó Phú Quốc có quan hệ chủng loại như thế nào với các nòi chó ôn đới nói trên như thế nào và tại sao chúng lại hiện diện ở vùng nhiệt đới Đông Nam Á là một câu hỏi đầy lý thú. Do vậy, chúng tôi quyết định giải mã toàn bộ trình tự genome ty thể chó lưng xoáy Phú Quốc thuộc haplotype E và B

phân tích cấu trúc bên trong, so sánh với các mtDNA của các nòi chó nổi tiếng khác trên thế giới nhằm bước đầu nhận định nguồn gốc của chó lưng xoáy Phú Quốc.

Ngoài ra, đến nay cơ sở dữ liệu GenBank chỉ ghi nhận có 1 trình tự genome ty thể chó dòng E1 (EU789662), còn các dòng E2, E3 và E4 thì chưa xuất hiện. Dữ liệu mtDNA của dòng E4 của chúng tôi sẽ góp phần bổ sung nguồn dữ liệu các mtDNA dòng E quý hiếm còn thiếu trên GenBank.

VẬT LIỆU VÀ PHƯƠNG PHÁP

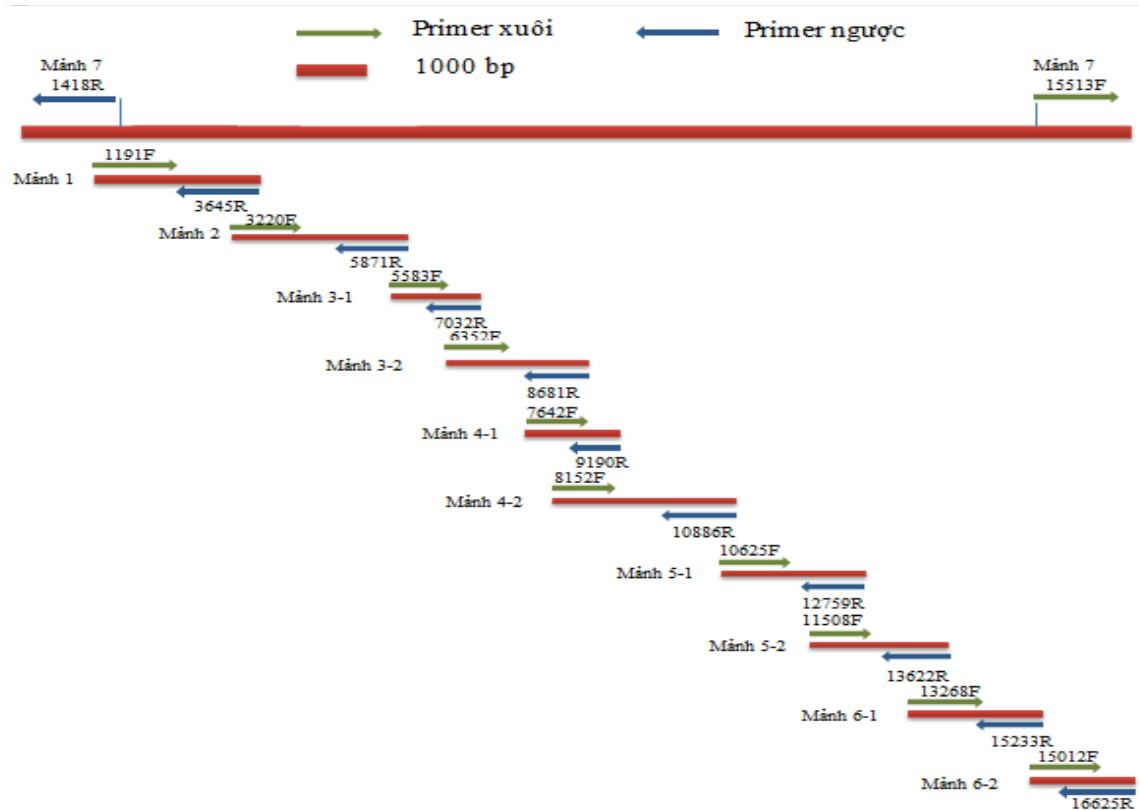
Mẫu vật

Ba cá thể chó Phú Quốc có ký hiệu là PQ1, PQ2 và PQ8 mang haplotype lần lượt E4, B1 và E1 đã

được xác định trước đó (Thai *et al.*, 2016) được chọn để đọc toàn bộ trình tự genome ty thể. Haplotype B1 (chó PQ2) được chọn như một mẫu đối chứng vì đây một haplotype cổ xưa phổ biến ở chó.

Khuếch đại mtDNA

Toàn bộ hệ gen ty thể được khuếch đại dưới dạng các đoạn DNA nhỏ có kích thước từ 1,4 – 3,5 kb chồng lắp lên nhau ít nhất 0,5 kb với các môi được tham khảo (Webb *et al.*, 2009 và Gundry *et al.*, 2007). Các đoạn DNA sau đó sẽ được giải trình tự tại Công ty Macrogen (Hàn Quốc) bằng phương pháp Sanger. Các môi khuếch đại cũng được sử dụng để giải trình tự bên cạnh các môi giải trình tự khác tự thiết kế nằm bên trong đoạn DNA. Trình tự và vị trí của các của các primers được biểu thị ở bảng 1, và hình 1.



Hình 1. Sơ đồ vị trí các cặp môi dùng để khuếch đại các đoạn gen ty thể nhỏ chồng lắp lên nhau của chó PQ1.

Bảng 1. Danh sách mỗi sử dụng khuếch đại và giải trình tự mtDNA chó Phú Quốc.

| STT | Tên Mồi | Trình tự Mồi | Nguồn | Chức năng |
|-----|---------|-------------------------|-----------------------------|---------------------------------|
| 1 | 42R | GGCATTTCAGTGCCTTGCTT | Gundry <i>et al.</i> (2007) | - Giải trình tự |
| 2 | 318R | GGTTAATCGTATGACCGCGG | Tự thiết kế | - Giải trình tự |
| 3 | 888F | CCATGAAGCACGCACACACC | Tự thiết kế | - Giải trình tự |
| 4 | 997R | AAGCACACCTTCCGGTATG | Webb <i>et al.</i> (2008) | - Giải trình tự |
| 5 | 1191F | GGAGCGATAGAGATAGTACC | Webb <i>et al.</i> (2008) | - Khuếch đại - Giải trình tự |
| 6 | 1418R | CACCAGGCTCGTTAGGCTT | Webb <i>et al.</i> (2008) | - Khuếch đại - Giải trình tự |
| 7 | 1620F | GAAAGCGTTCCAGCTCAACA | Webb <i>et al.</i> (2008) | - Giải trình tự |
| 8 | 1770F | ATCAGGAACGGATAGACCAC | Webb <i>et al.</i> (2008) | - Giải trình tự |
| 9 | 2436F | ACATCCTAATGGTGACAGCAG | Tự thiết kế | - Giải trình tự |
| 10 | 2556R | CATCCCTTGTCCTTTTCGTAC | Webb <i>et al.</i> (2008) | - Giải trình tự |
| 11 | 2976R | AGGGCTAGTGATAGAGCTAG | Tự thiết kế | - Giải trình tự |
| 12 | 3220F | GTCATTTACACTATCCACGC | Webb <i>et al.</i> (2008) | - Khuếch đại - Giải trình tự |
| 13 | 3645R | GGCAACATGTCATATGCATA | Webb <i>et al.</i> (2008) | - Khuếch đại - Giải trình tự |
| 14 | 3945F | CAACTATCATGACAGGAACC | Webb <i>et al.</i> (2008) | - Khuếch đại - Giải trình tự |
| 15 | 4193R | GTGGTTATCATGATGGATGCG | Tự thiết kế | - Giải trình tự |
| 16 | 4581F | TCATCCACCACGACCCTATC | Tự thiết kế | - Giải trình tự |
| 17 | 4793R | GTGCTATATGTGAGTCGCAG | Webb <i>et al.</i> (2008) | - Khuếch đại - Giải trình tự |
| 18 | 5583F | GGAAACTGACTAGTGCCGTT | Webb <i>et al.</i> (2008) | - Khuếch đại - Giải trình tự |
| 19 | 5689R | GCTTCTACCATAGAAGATGC | Tự thiết kế | - Giải trình tự |
| 20 | 5871R | AGTTTGATACTGGGATATTGC | Webb <i>et al.</i> (2008) | - Khuếch đại - Giải trình tự |
| 21 | 6352F | CCAGCTATGCTATGAGCTT | Webb <i>et al.</i> (2008) | - Khuếch đại - Giải trình tự |
| 22 | 6826F | GAGTGACTACATGGATGTCC | Tự thiết kế | - Giải trình tự |
| 23 | 7032R | TTGAAATGGGTACGCCATAG | Webb <i>et al.</i> (2008) | - Khuếch đại - Giải trình tự |
| 24 | 7642F | CCACAGCTTTATACCCATTG | Webb <i>et al.</i> (2008) | - Khuếch đại - Giải trình tự |
| 25 | 7805R | GGATGTATCTAGCTGTGGCA | Tự thiết kế | - Giải trình tự |
| 26 | 8152F | AAAGGGACGAACCTGAGCTC | Webb <i>et al.</i> (2008) | - Khuếch đại - Giải trình tự |
| 27 | 8681R | TGTCAGCGGTCATGGGCTTG | Webb <i>et al.</i> (2008) | - Khuếch đại - Giải trình tự |
| 28 | 9031F | GCCTCTACTCAACACCTCAG | Tự thiết kế | - Giải trình tự |
| 29 | 9190R | CGGAGATTGTAAGATGTCTC | Webb <i>et al.</i> (2008) | - Khuếch đại - Giải trình tự |
| 30 | 10087R | CGATTGGTATCATGCTGGCT | Tự thiết kế | - Giải trình tự |
| 31 | 10625F | GACTAAACGCAGGACTCTAC | Webb <i>et al.</i> (2008) | - Khuếch đại - Giải trình tự |
| 32 | 10886R | GGAGTACAGCGCAAGTACTA | Webb <i>et al.</i> (2008) | - Khuếch đại - Giải trình tự |
| 33 | 11286F | AGCAAGCCTCACAAATCTGG | Tự thiết kế | - Giải trình tự |
| 34 | 11508F | AATGACCTTGACCTACTGC | Webb <i>et al.</i> (2008) | - Khuếch đại - Giải trình tự |
| 35 | 11720R | CCAACGGATTACTTCTATCC | Tự thiết kế | - Giải trình tự |
| 36 | 12529F | GCACAATAGTTGTAGCAGGAG | Tự thiết kế | - Giải trình tự |
| 37 | 12759R | AATGCGTGAGTGACAGATGTG | Webb <i>et al.</i> (2008) | - Khuếch đại - Giải trình tự |
| 38 | 13268F | TTTCATCCTGGCACTAGAAC | Webb <i>et al.</i> (2008) | - Khuếch đại - Giải trình tự |
| 39 | 13622R | GTTACAGGCTGATCATTATTAAT | Webb <i>et al.</i> (2008) | - Khuếch đại |

| | | | | |
|----|--------|------------------------|-----------------------------|---------------------------------|
| | | | | - Giải trình tự |
| 40 | 14253F | CGTCTAACATCTCTGCTTGA | Webb <i>et al.</i> (2008) | - Giải trình tự |
| 41 | 14397R | TCAGCCGTAGTTAACGTCTC | Tự thiết kế | - Giải trình tự |
| 42 | 15012F | CCTATGCTATCCTACGATCC | Webb <i>et al.</i> (2008) | - Khuếch đại - Giải trình tự |
| 43 | 15233R | AAGATTGAAGCGACTTGTC | Webb <i>et al.</i> (2008) | - Khuếch đại - Giải trình tự |
| 44 | 15412F | CCACTATCAGCACCCAAAG | Gundry <i>et al.</i> (2007) | - Giải trình tự |
| 45 | 15513F | GGTAAACCCTTCTCCCCTC | Webb <i>et al.</i> (2008) | - Khuếch đại - Giải trình tự |
| 46 | 15887F | GATCACACATAACTGTGGTG | Tự thiết kế | - Giải trình tự |
| 47 | 16072F | CTCACGCATAARATCAAGGTG | Gundry <i>et al.</i> (2007) | - Giải trình tự |
| 48 | 16114R | CCTGAAACCATTGACTGAATAG | Gundry <i>et al.</i> (2007) | - Giải trình tự |
| 49 | 16625R | AGACTACGAGACCAAATGCG | Webb <i>et al.</i> (2008) | - Khuếch đại - Giải trình tự |

Note: Forward (F): mũi xuôi; Reverse (R): mũi ngược; ký hiệu số phía trước chỉ vị trí mỗi bắt cặp dựa theo vị trí trên bộ gen ty thể chó của Kim *et al.* (1998).

Thành phần hóa chất cho phản ứng PCR

Phản ứng PCR khuếch đại các đoạn genome ty thể chó Phú Quốc có kích thước trên 3 kb được thực hiện bao gồm 2,5 µl đệm PCR (10X); 2,5 µl MgCl₂ (25µM); 0,25 µl Taq Plus DNA polymerase PCR 250UI (hãng ABM - 090120); 2,5 µl dNTP (2µM); 0,4 µl mũi xuôi và mũi ngược (10µM) do hãng SIGMA cung cấp); 1,0 µl DNA tổng số (nồng độ 50 ng/µl); nước cất vừa đủ 25 µl. Với các đoạn genome ty thể dưới 3 kb, phản ứng bao gồm, 12,5 µl Taq DNA pol 2x-preMix (GeneON, Cat.-No: S113.100 res), 0,5 µl (nồng độ 10 µM) mũi mũi ngược và xuôi, 0,7 µl (nồng độ 50 ng/µl) DNA tổng số, bổ sung nước cất tới 25 µl.

Chu trình nhiệt trong phản ứng khuếch đại

Với các đoạn DNA dài trên 3 kb, chu trình nhiệt là tiền biến tính 95 °C trong 5 phút; sau đó 30 chu kỳ bao gồm biến tính 95 °C trong 15 giây, bắt cặp 55 °C trong 30 giây và kéo dài 72 °C trong 3,5 phút; kéo dài sau cùng 72 °C trong 7 phút và trữ mẫu 4 °C. Với các đoạn DNA dài dưới 3 kb, chu trình nhiệt là tiền biến tính 95 °C trong 5 phút; sau đó 30 chu kỳ bao gồm biến tính 95 °C trong 15 giây, bắt cặp 55 °C trong 30 giây và kéo dài 72 °C trong 3 phút; kéo dài sau cùng 72 °C trong 7 phút và trữ mẫu 4 °C.

Giải trình tự

Các đoạn genome ty thể được giải trình tự trực tiếp bằng mũi khuếch đại và mũi trung gian, thực hiện tại Công ty Macrogen (Hàn Quốc).

Hiệu chỉnh trình tự

Trình tự thu được sau khi giải mã đã được kiểm tra độ chính xác của các sắc ký đồ (chromatography) bằng phần mềm ChromasPro 2.0 (Technelysium Pty Ltd), các vùng mơ hồ ở đầu và cuối kết quả giải trình tự bị cắt bỏ. Sau đó kiểm tra các sai lệch giữa hai kết quả giải trình tự từ mũi xuôi và mũi ngược cho cùng 1 đoạn genome bằng phần mềm SeaView; các điểm sai lệch được kiểm tra bằng mắt để ghi nhận kết quả sau cùng. Các đoạn trình tự thô thường được biểu hiện ở dạng sắc ký đồ để nhận diện các base và tính toán chất lượng điểm giải trình tự thể hiện qua "giá trị Q" tương ứng với xác suất lỗi theo quy mô Q10=1/10, Q20=1/100, Q30=1/1000, và tiếp tục như vậy. Quá trình xử lý các khoảng trống đối với các trình tự liên ứng (consensus), nhất là khi có nhiều đoạn đọc chất lượng thấp được thực hiện thủ công.

Lắp ráp genome ty thể

Các đoạn genome ty thể sau đó lắp ráp thành chuỗi trình tự genome ty thể hoàn chỉnh dựa trên trình tự genome ty thể chó tham khảo của Kim *et al.* (1998). Chất lượng của tất cả các đoạn lắp ráp và các đoạn trình tự được kiểm tra lại bằng cách sử dụng phần mềm chuyên dụng như SeaView, Sequencher hoặc CONSED.

Xây dựng bộ dữ liệu mtDNA

Trình tự genome ty thể chó Phú Quốc sau khi giải trình tự, hiệu chỉnh sẽ được gộp chung với trên 100 trình tự mtDNA của các nòi chó được thu thập từ

GenBank để tạo bộ cơ sở dữ liệu bộ gen ty thể của chó. Quá trình sắp cột thẳng hàng (alignment) được thực hiện bằng phần mềm SeaView với thuật toán Muscle, các vùng mờ hồ hoặc không có khả năng sắp xếp được loại bỏ bằng tay trước khi phân tích. Quá trình sắp cột thẳng hàng các mtDNA chó được thực hiện bằng phần mềm MEGA6.0; các mtDNA có kích thước dưới 10 kb không đưa vào nghiên cứu này.

Dò tìm mô hình tiến hoá

Xác định mô hình tiến hoá phù hợp nhất theo chuẩn Akaike Information Criterion (AICc) bằng phần mềm jModelTest phiên bản 0.1.1 (Posada, 2008).

Xây dựng cây phát sinh loài

Phương pháp khả năng tối đa (Maximum Likelihood) được tạo bằng phần mềm PhyML 3.0, Phương pháp khoảng cách (Neighbor-Joining) và Phương pháp tiết giảm tối đa (maximum parsimony) được thực hiện bằng PAUP* 4.0b10 (Swofford, 2003) với các thông số tiến hóa được lấy mô hình tiến hóa tối ưu trước đó. Tất cả được thực hiện với độ tin cậy (bootstrap) là 1000 lần lặp lại. Phương pháp xác suất hậu nghiệm được thực hiện trên phần mềm MrBayes 3.2 (Ronquist & Huelsenbeck, 2003), trong đó các thông số cơ bản được cài đặt theo mẫu

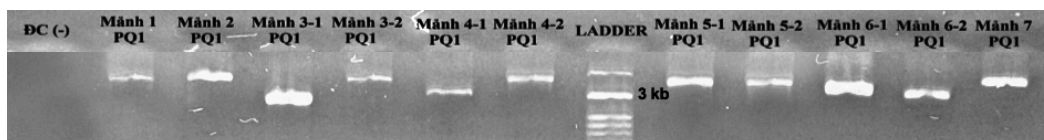
chuẩn. Mẫu được chạy 2.000.000 lần với 1 chuỗi lạnh và 3 chuỗi nóng, cây được lưu sau mỗi 100 lần chạy như mô tả của Hoef-Emden *et al* (2005).

KẾT QUẢ VÀ THẢO LUẬN

Khuếch đại, giải trình tự và lắp ráp toàn bộ hệ gen ty thể chó Phú Quốc

Các đoạn genome ty thể kích thước lớn được khuếch đại thành công như dự kiến (Hình 2), các sản phẩm PCR có chiều dài từ 1,4 đến 3,3 kb được giải trình tự trực tiếp bằng phương pháp Sanger.

Toàn bộ mtDNA của mẫu chó Phú Quốc PQ1 và PQ2 đại diện cho kiểu đơn bội E4, B1 được khuếch đại và giải trình tự như dự kiến lần lượt dài là 16.760 bp 16.731 bp. Trong đó mtDNA của haplotype E4 của chó trong nghiên cứu này là lần đầu tiên được công bố trên thế giới, hiện đang gửi Genbank để cấp mã số truy cập. Riêng mtDNA của chó PQ8 mang haplotype E1 chỉ mới giải mã được 13.149 bp, tuy vậy khi so sánh với các mtDNA mang haplotype E1 có trên GenBank (EU798662) không cho thấy sự khác biệt. Mặc dù mẫu mtDNA của chó PQ8 không được phân tích cấu trúc nhưng vẫn được sử dụng để phân tích phát sinh chủng loài về sau.



Hình 2. Kết quả khuếch đại toàn bộ hệ gen ty thể chó PQ1.

Phân tích thành phần và trật tự gen của mtDNA chó PQ1 mang haplotype E1 và chó PQ2 mang haplotype B1

Thành phần base

Thành phần base của DNA ty thể chó mang haplotype E4 có T chiếm 28,7%, C chiếm 25,5%, A chiếm 31,6% và G chiếm 14,2%. Ở chó B1, tỷ lệ này lần lượt là 28,7%; 25,5%; 31,6% và 14,1%. Riêng chó E1, mặc dù trình tự chỉ mới đọc được 13.149 nhưng tỷ lệ nucleotide cũng được tính toán và giá trị

nhận được là 28,9%; 25,9%; 31,1% và 14,0%. Tỷ lệ này tương đương với tỷ lệ nucleotide trung bình cho các giống chó trên thế giới là 28,8%; 25,5%; 31,7% và 14,1% (Kim *et al.*, 1998).

Hệ gen ty thể chó E4 và B1 chứa 13 gen mã hóa protein, 22 gen RNA vận chuyển (tRNA), 2 gen ribosome (12S và 16S rRNA) và một vùng không mã hóa gọi là vùng kiểm soát. Sự tổ chức và điều khiển của chúng giống như những động vật có vú khác. Đặc điểm hệ gen ty thể chó E4 được ghi nhận và thể hiện trong bảng 2.

Bảng 2. Đặc tính của genome ty thể chó Phú Quốc mang haplotype E4.

| STT | Tên gen | Vị trí | Chiều dài | Mã khởi đầu | Mã kết thúc |
|-----|--|-------------|-----------|-------------|-------------|
| 1 | tRNA-Phe | 1-69 | 69 | | |
| 2 | 12S rRNA | 70-1023 | 954 | | |
| 3 | tRNA-Val | 1024-1090 | 67 | | |
| 4 | 16S rRNA | 1091-2671 | 1581 | | |
| 5 | tRNA-Leu (UUR) | 2672-2746 | 75 | | |
| 6 | NADH dehydrogenase subunit 1 (ND1) | 2750-3704 | 957 | ATG | ATA |
| 7 | tRNA-Ile | 3705-3773 | 69 | | |
| 8 | tRNA-Gln | 3770-3844 | 75 | | |
| 9 | tRNA-Met | 3846-3915 | 70 | | |
| 10 | NADH dehydrogenase subunit 2 (ND2) | 3916-4958 | 1044 | ATA | ACT |
| 11 | tRNA-Trp | 4957-5025 | 68 | | |
| 12 | tRNA-Ala | 5039 -5107 | 69 | | |
| 13 | tRNA-Asn | 5109-5180 | 72 | | |
| 14 | Origin of L- strand replication (OLR) | 5181-5217 | 37 | | |
| 15 | tRNA-Cys | 5214-5281 | 68 | | |
| 16 | tRNA-Tyr | 5282-5349 | 68 | | |
| 17 | Cytochrome c oxidase subunit 1 (COI) | 5351-6895 | 1545 | ATG | TAA |
| 18 | tRNA-Ser (UCN) | 6893-6963 | 71 | | |
| 19 | tRNA-Asp | 6968-7035 | 68 | | |
| 20 | Cytochrome c oxidase subunit II (COII) | 7036-7719 | 684 | ATG | TAA |
| 21 | tRNA-Lys | 7737-7803 | 67 | | |
| 22 | ATPase 8 | 7805-8008 | 204 | ATG | TAA |
| 23 | ATPase 6 | 7966-8646 | 681 | ATG | TAA |
| 24 | Cytochrome c oxidase subunit III (COIII) | 8646-9429 | 784 | ATG | CC T |
| 25 | tRNA-Gly | 9430-9497 | 68 | | |
| 26 | NADH dehydrogenase subunit 3 (ND3) | 9499-9843 | 347 | ATA | AAT |
| 27 | tRNA-Arg | 9844-9912 | 69 | | |
| 28 | NADH dehydrogenase subunit 4L (ND4L) | 9913-10209 | 297 | ATG | TAA |
| 29 | NADH dehydrogenase subunit 4 (ND4) | 10205-11582 | 1378 | ATG | ACT |
| 30 | tRNA-His | 11583-11650 | 71 | | |
| 31 | tRNA-Ser (AGY) | 11651-11710 | 60 | | |
| 32 | tRNA-Leu (CUN) | 11711-11780 | 70 | | |
| 33 | NADH dehydrogenase subunit 5 (ND5) | 11782-13602 | 1821 | ATA | TAA |
| 34 | NADH dehydrogenase subunit 6 (ND6) | 13586-14113 | 528 | TTA | CAT |
| 35 | tRNA-Glu | 14114-14182 | 69 | | |
| 36 | Cytochrome b | 14187-15326 | 1140 | ATG | AGA |
| 37 | tRNA-Thr | 15327-15396 | 70 | | |
| 38 | tRNA-Pro | 15396-15461 | 66 | | |
| 39 | Control region | 15462-16760 | 1299 | | |

Các gen mã hóa protein

Tổng chiều dài của mtDNA mã hóa cho 13 gen mã hóa protein ở chó E4 và chó B1 là 11.406 bp. Gen dài nhất là ND5 (1821 bp), gen ngắn nhất là ATPase8 (204 bp). Hệ gen của cả chó E4 và B1 sử dụng 3 bộ mã khởi đầu: 9 gen sử dụng ATG; ND2, ND3, ND5 sử dụng ATA và ND6 sử dụng TTA. Các bộ mã kết thúc của 7 gen được kết thúc đầy đủ (TAA, TAG hoặc AGA) và 6 gen (ND1, ND2, COIII, ND3, ND4, ND6) sử dụng bộ mã kết thúc không đầy đủ. Như trong một số động vật có vú khác, có bốn khung đọc chồng lấp lên nhau trong chó PQ1 và chó PQ2, ATPase8/ATPase6, ATPase6/CoIII, ND4/ND4L và ND5/ND6, chồng lên nhau 43, 1, 7 và 17 bp, theo thứ tự tương ứng.

Các gen tRNA và rRNA

Các giới hạn và chiều của các gen tRNA ở chó được xác định bằng sự liên hệ so sánh với những gene tRNA của chó Sapsaree (NC_002008.4). Chiều dài tổng hợp của các gen tRNA của chó E4 và B1 là 1.517 nucleotide. Như ở động vật có vú khác, 22 gen tRNA đã được tìm thấy trong mtDNA chó E4 và B1, và có sự tổ chức tương tự nhau. Chiều dài của chúng dao động trong khoảng từ 60 đến 75 bp. Các gen 12S và 16S rRNA được phân cách bởi các gen tRNA-Val và bị giới hạn phía các bên kia bằng các gen tRNA-Phe và tRNA-Leu, tương ứng. Độ dài của gen 12S rRNA là 954 bp, gen 16S rRNA là 1.580 bp, giống nhau cho cả chó PQ1 và chó PQ2 và giống với chó Sapsaree trong nghiên cứu của Kim *et al* (1998).

Vùng kiểm soát (control region – CR hay D-loop)

Vùng CR của mtDNA của chó E4 dài 1.299 bp và của chó B1 dài 1.270 bp. Sự khác nhau về chiều dài là do sự thay đổi số lần lặp lại của một motif gồm 10 nucleotide 5'-GTACAC-GT(G/A)C-3' nằm giữa vùng bảo tồn I và II (Conserved sequence block – CSB). Ở đây, chó E4 lặp lại motif này trên 33 lần, và chó B1 là 30 lần. Với chó E1, mặc dù chưa đọc hoàn chỉnh toàn bộ mtDNA, nhưng vùng kiểm soát cũng được chúng tôi so sánh cho thấy kích thước vùng CR của chó E4 và E1 tương tự nhau.

Ngạc nhiên hơn nữa và đoạn dư đôi của vùng CR ở chó nòi E1 và E4 (khoảng 29 nucleotides) cũng được tìm thấy ở mtDNA của nòi chó Golden Retriever (FJ817363), nòi Old English Sheepdog (AY656742.1), sói xám (Grey wolf) mẫu

(AM711902 và NC_009686) là những nòi chó săn hoặc chăn cừu cổ xưa. Tuy các motif lặp lại này không đại diện thực sự cho phép nhận dạng haplotype nhưng các thông tin trên cũng cho thấy phần nào điều thú vị về mtDNA của chó Phú Quốc.

Phân tích phát sinh chủng loài dựa trên mtDNA

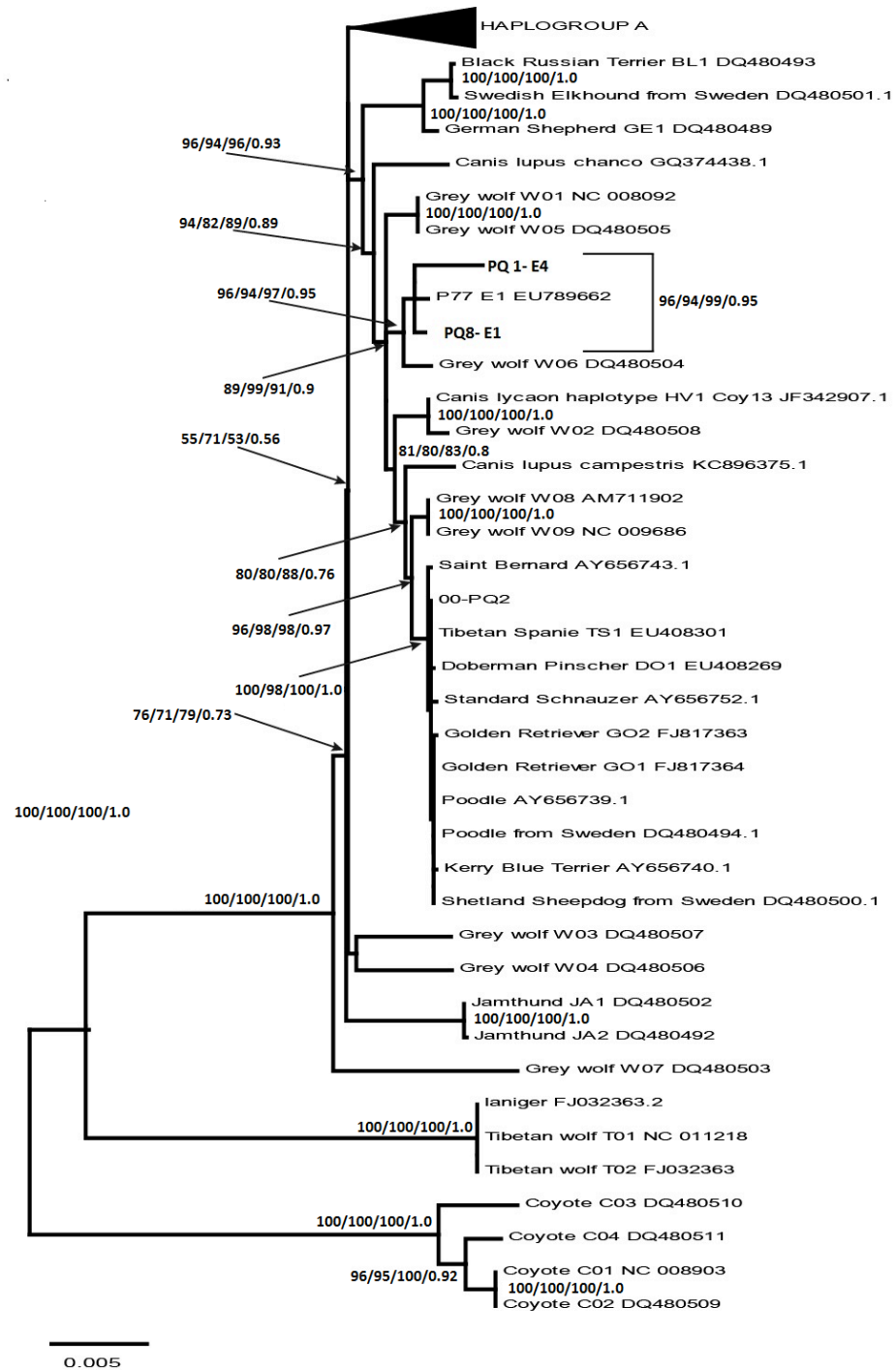
Xây dựng cơ sở dữ liệu cho phân tích phát sinh chủng loài

Trong các nghiên cứu trước đó (Bjornerfeldt *et al.*, 2006; Pang *et al.*, 2009; Oskarsson *et al.*, 2012) nhóm chó mang kiểu đơn bội E rất ít được chú ý do số lượng mẫu rất hiếm. Trong nghiên cứu này, một trình tự mtDNA mang kiểu đơn bội E1 từ giống chó Pungsan ở Hàn Quốc (ký hiệu mẫu P77, EU789662) được tìm thấy như một đối chứng quý giá. Do vậy trong phân tích này, chúng tôi quyết định đưa trình tự mtDNA chó lung xoáy Phú Quốc mang haplotype E1 vào phân tích, mặc dù chiều dài chỉ mới đọc khoảng 13 kb.

Để xây dựng bộ cơ sở dữ liệu cho phân tích phát sinh chủng loài, trên 150 trình tự mtDNA của các loài thuộc họ nòi chó nổi tiếng trong loài *Canis familiaris* trên thế giới được thu thập. Sử dụng chó hoang (*Canis latrans*) làm nhóm ngoại. Các trình tự mtDNA có chiều dài gần tương tự như của mtDNA E1 được sử dụng phân tích để hạn chế việc loại bỏ các điểm mang thông tin. Các mtDNA có trình tự dưới 12 kb bị loại bỏ; các mtDNA có khoảng trống gây ảnh hưởng đến tổng số điểm phân tích cũng được loại bỏ. Các khoảng trống và vùng lặp lại ở khu vực bảo tồn bị loại bỏ hoàn toàn. Do vậy bộ dữ liệu mtDNA của chó trong nghiên cứu này chứa 72 trình tự với độ dài vùng phân tích 12.564 ký tự.

Phân tích phát sinh chủng loài

Phả hệ đồ (phylogram) dựa trên trình tự mtDNA cho thấy các haplogroup A, B, C và E tách ra rõ rệt như quan sát trên phả hệ đồ sử dụng trình tự vùng CR như phân tích trước đó (Thai *et al.*, 2016). Chó PQ2 nằm chung với các giống chó mang kiểu đơn bội B; chó PQ1 và PQ8 mang kiểu đơn bội E4 và E1 nằm chung cụm với chó P77 mang haplotype E1 với độ tin cậy (bootstrap) rất cao (96/94/97/0.95). Ngoài ra, trên phả hệ đồ còn cho thấy chó sói xám W06 (mã số DQ480504) nằm ở vị trí chị em so với nhóm E với độ tin cậy gần như tuyệt đối (99/99/91/0.99). Được biết sói xám W06 được (Bjornerfeldt *et al.*, 2006) thu mẫu tại Thụy Điển (Hình 3).



Hình 3. Cây phả hệ (phylogenetic tree) mô tả chi tiết vị trí phân bố của chó lưng có xoáy Phú Quốc trong các nhóm đơn bội B, C và E (không mô tả nhóm A) trên cây tiến hóa có gốc Maximum Likelihood được suy luận từ 72 trình tự vùng phân tích là toàn bộ mtDNA có độ dài 12.564 ký tự (khoảng trống bị loại bỏ hoàn toàn). Mô hình tiến hóa là HKY+G+I (-lnL = 25852.5445, G= 0.1025 và I = 65.0545) được tính bằng phần mềm jModelTest dựa theo chuẩn AICc. Giá trị bootstrap từ trái qua phải là Maximum-likelihood, Neighbor-Joining, Maximum Parsimony không ràng buộc và xác suất hậu nghiệm Bayes (BP - Bayesian probability) Tỷ lệ xích biểu thị số biến đổi tại mỗi vị trí.

Tính toán khoảng cách di truyền giữa chó Phú Quốc dòng E4, E1 và B1 trong nghiên cứu này cũng đã cho ra nhiều kết quả bất ngờ. Chó Phú Quốc dòng B1 tương đồng 100% (khác biệt 0,0%) so với chó Tibetan Spanie TS1 (EU408301), đây là nòi chó cảnh Tây Tạng được xem là dòng chó cổ xưa của Trung Quốc (Li *et al.*, 2011). Haplotype B1 cũng được xem là haplotype cổ nhất trong các haplotype dạng B.

Mặc dù cùng haplogroup nhưng khoảng cách di truyền của chó Phú Quốc dòng E1 và E4 lại ở mức 0,23%.

Điều không ngạc nhiên là tỷ lệ khác biệt giữa mtDNA chó Phú Quốc nòi E1 và chó Pusang (EU789662) là thấp nhất, chỉ 0,072%. Trong khi đó khoảng cách giữa chó Phú Quốc nòi E4 và chó Pusang là 0,24%.

Giá trị khác biệt di truyền cũng được quan sát giữa Chó Phú Quốc dòng E1, E4 so với các chó sói xám W06 (mã số DQ480504) nằm ở vị trí phân kỳ chi em trên phả hệ đồ lần lượt là 0,3, và 0,1%

Ghi nhận của Klütsch *et al* (2011) cho rằng do số lượng cá thể nhóm D, E và F khá nhỏ (không quá 5% tổng đàn chó thể giới) nên việc truy tìm nguồn gốc của 3 nhóm này rất khó khăn. Tuy thế, do các nhóm này chỉ tìm thấy ở khu vực địa lý giới hạn như khu vực Đông Bắc Á nên điều này gợi ý rằng chúng xuất phát từ phép lai giữa chó nhà và chó sói địa phương trong khoảng thời gian gần đây. Đây được gọi là hiện tượng thuần hóa chó sói lần 2, và vì mới xuất hiện từ lần thuần hóa thứ 2 nên các kiểu đơn bội mới sẽ rất khó phát tán ra vùng địa lý khác vốn đã có các kiểu đơn bội khác ngự trị từ trước là A, B và C. Hơn nữa, nếu một haplotype xuất phát từ quá trình thuần hóa chó sói ở một khu vực không có chó, khi đó tần suất của kiểu haplotype mới sẽ chiếm 100% ở quần thể chó ban đầu và tần suất này sẽ chỉ giảm đi khi chó mang haplotype khác tràn đến xâm chiếm lãnh thổ. Giải thích này hoàn toàn phù hợp trong trường hợp chó Pungsan mang kiểu đơn bội E1 được ghi nhận là chó lai giữa chó sói và chó nhà ở Triều Tiên và chỉ xuất hiện ở khu vực Triều Tiên.

Việc chó Phú Quốc mang haplotype E4 và E1 trong các nghiên cứu trước đó (Thai *et al.*, 2016) là một điều đáng quan tâm để truy tìm nguồn gốc chó Phú Quốc. Trình tự mtDNA của chó Phú Quốc dòng E1 có tỷ lệ tương đồng 99,928 % với chó Punsang là một ngạc nhiên lớn. Điều này tạo tiền đề cho chúng tôi tiến hành các phân tích sâu hơn về sau.

KẾT LUẬN

Chiến thuật primer-walking được áp dụng để giải trình tự genome ty thể chó Phú Quốc thành công cho thấy có thể áp dụng một kỹ thuật đơn giản, hiệu quả nhằm thu nhận các thông tin từ một trong những bộ gen có kích thước nhỏ nhất phục vụ cho nghiên cứu chuyên sâu. Genome ty thể chó Phú Quốc mang haplotype E4, có kích thước 16.760 nucleotide và các thành phần không khác biệt so với các nòi chó khác trên thế giới. Việc phát hiện chó lung xoáy Phú Quốc có mang các haplotype hiếm E1 và E4 đồng thời có quan hệ gần với nòi chó quý Pungsan ở Triều Tiên là một phát hiện thú vị, hé mở dần nguồn gốc của nòi chó đặc hữu Việt Nam.

Lời cảm ơn: Chúng tôi chân thành cảm tạ Bộ Khoa học & Công nghệ đã tài trợ cho Đề tài cấp Nhà nước trong Chương trình Quỹ Gen để thực hiện nghiên cứu này (Hợp đồng số 01/2015 – HĐ-NVQG).

TÀI LIỆU THAM KHẢO

- Gundry RL, Allard MW, Moretti TR, Honeycutt RL, Wilson MR, Monson KL, Foran DR (2007). Mitochondrial DNA analysis of the domestic dog: Control region variation within and among breeds. *J Forensic Sci.* 52(3): 562-572.
- Hoef-Emden K, Tran H-D, Melkonian M (2005). Lineage-specific variations of congruent evolution among DNA sequences from three genomes, and relaxed selective constraints on *rbcl* in *Cryptomonas* (Cryptophyceae). *BMC Evol Biol* 5: 56.
- Kim KS, Lee SE, Jeong HW, Ha JH (1998) The complete nucleotide sequence of the domestic dog (*Canis familiaris*) mitochondrial genome. *Mol Phylogenet Evol.* 10(2): 210-220.
- Li YL, Li Q, Zhao X, Xie Z, Xu Y (2011) Complete sequence of the Tibetan Mastiff mitochondrial genome and its phylogenetic relationship with other Canids (Canis, Canidae). *Animal* 5(1): 18-25.
- Oskarsson MC, Kluttsch CF, Boonyaparakob U, Wilton A, Tanabe Y, Savolainen P (2012) Mitochondrial DNA data indicate an introduction through Mainland Southeast Asia for Australian dingoes and Polynesian domestic dogs. *Proc Biol Sci.* 279(1730): 967-974.
- Pang JF, Kluttsch C, Zou XJ, Zhang AB, Luo LY, Angleby H, Ardalán A, Ekstrom C, Skollermo A, Lundeberg J, Matsumura S, Leitner T, Zhang YP, Savolainen P (2009) mtDNA data indicate a single origin for dogs south of Yangtze River, less than 16,300 years

ago, from numerous wolves. *Mol Biol Evol.* 26(12): 2849-2864.

Ronquist FL, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19(12): 1572-1574.

Savolainen P, Zhang YP, Luo J, Lundeberg J, Leitner T (2002) Genetic evidence for an East Asian origin of domestic dogs. *Science*, 298(5598): 1610-1613.

Thai KQ, Nguyen VT, Tran NT, Huynh VH, Chung AD, Tran HD (2016) Evaluation of genetic diversity of Phu Quoc ridgeback dogs based on mitochondrial DNA Hypervariable-1 region. *J Biotech* 14(1A): 245-253.

Webb KM, Allard MW (2009) Identification of forensically informative SNPs in the domestic dog mitochondrial control region. *J Forensic Sci.* 54(2): 289-304.

COMPLETE MITOCHONDRIAL GENOME SEQUENCES OF RARE HAPLOTYPE E4 FROM PHU QUOC RIDGEBACK DOG

Tran Hoang Dung, Truong Nguyen Thi Nhu Mai, Nguyen Thanh Cong, Huynh Van Hieu

Nguyen Tat Thanh University

SUMMARY

Phu Quoc ridgeback dog is an invaluable species in Vietnam. The discovery of high-frequency rare haplotypes E4 and E1 in Phu Quoc ridgeback dog while evaluate analyzing its genetic distances is an unexpected event. In this research, the primer-walking strategy was used to sequence the whole mtDNA of Phu Quoc ridgeback dogs carrying E4, B1 and E1 haplotypes to explore trace of Phu Quoc ridgeback dog's origin. These mtDNA's length (16,760 bp for PQ1 and 16,731 for PQ2) fall into the range of 16.7 kb, particularly, 16.760 bp in E4 dogs and 16.731 bp in B1 dogs and is basically similar to the ones of other dog breeds. The difference in length between E4 and B1 haplotype is mainly due to the difference of motif repeat frequency in the control region. The number of genes, starting and ending codes of E4's mtDNA are similar to other common haplotypes. Genetic distances between Phu Quoc ridgeback dogs' E1, E4 haplotypes and Pungsan dogs (EU789662) in the same clade in the phylogenetic tree are 0,072% and 0,24%, respectively, showing that Phu Quoc dogs with E haplotypes are very close relatively to Pungsan dogs in North-Korea. Although belonging to the same haplogroup, the genetic distance between Phu Quoc dogs E1 and E4 haplotype is 0,23%. In the other hand, genetic distances between E1, E4 haplotypes Phu Quoc ridgeback dogs and the closest grey wolves on phylogenetic tree are 0,3% and 0,1%, respectively. This is the first mitochondrial genome of E4 haplotype dogs to be sequenced and analyzed.

Keywords: *Phu Quoc ridgeback dog, genetic distances, mitochondrial genome, haplotype E4-E1, Primer-walking approach, longrange-PCR*