

**BỘ GIÁO DỤC
VÀ ĐÀO TẠO**

**VIỆN HÀN LÂM KHOA HỌC
VÀ CÔNG NGHỆ VIỆT NAM**

HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ



Lưu Thủy Tiên

**NGHIÊN CỨU CẤU TRÚC DI TRUYỀN QUẦN THỂ NGƯỜI Ê ĐÊ,
VIỆT NAM VÀ MỐI QUAN HỆ VỚI CÁC QUẦN THỂ KHÁC
CÙNG KHU VỰC ĐÔNG NAM Á VÀ CÙNG NGŨ HỆ NAM ĐẢO**

LUẬN VĂN THẠC SĨ NGÀNH SINH HỌC

LƯU THỦY TIÊN

SINH HỌC THỰC NGHIỆM

2023

Hà Nội – 2023

**BỘ GIÁO DỤC
VÀ ĐÀO TẠO**

**VIỆN HÀN LÂM KHOA HỌC
VÀ CÔNG NGHỆ VIỆT NAM**

HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ



Lưu Thủy Tiên

**NGHIÊN CỨU CẤU TRÚC DI TRUYỀN QUẦN THỂ NGƯỜI Ê ĐÊ, VIỆT
NAM VÀ MỐI QUAN HỆ VỚI CÁC QUẦN THỂ KHÁC CÙNG KHU VỰC
ĐÔNG NAM Á VÀ CÙNG NGŨ HỆ NAM ĐẢO**

Chuyên ngành : Sinh học thực nghiệm

Mã số: 8420114

LUẬN VĂN THẠC SĨ NGÀNH SINH HỌC

NGƯỜI HƯỚNG DẪN KHOA HỌC :

GS.TS Chu Hoàng Hà

Hà Nội - 2023

LỜI CAM ĐOAN

Tôi xin cam đoan đề tài nghiên cứu trong luận văn này là công trình nghiên cứu của tôi dựa trên những tài liệu, số liệu do chính tôi tự tìm hiểu và nghiên cứu. Chính vì vậy, các kết quả nghiên cứu đảm bảo trung thực và khách quan nhất. Đồng thời, kết quả này chưa từng xuất hiện trong bất cứ một nghiên cứu nào. Các số liệu, kết quả nêu trong luận văn là trung thực nếu sai tôi hoàn chịu trách nhiệm trước pháp luật.

Tác giả

Lưu Thủy Tiên

LỜI CẢM ƠN

Đầu tiên tôi xin được gửi lời cảm ơn đến thầy hướng dẫn GS.TS Chu Hoàng Hà, nếu không có sự hỗ trợ tận tình trong suốt quá trình tôi sẽ không thể hoàn thành được luận văn này. Tôi xin cảm ơn các nhân viên và lãnh đạo Trung tâm Giám định DNA và Phòng thí nghiệm trọng điểm Công nghệ Gen, Viện Công nghệ sinh học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam, đã giúp đỡ tôi có thêm nhiều kiến thức và kinh nghiệm trong mọi bước tiến hành luận văn. Bên cạnh đó, tôi cũng xin được cảm ơn Viện Hàn lâm Khoa học và Công nghệ Việt Nam và các thành viên trong đề tài “Nghiên cứu giải trình tự gen các mẫu xương khảo cổ tại Việt Nam nhằm cung cấp thông tin di truyền cho nghiên cứu đa dạng sinh học người và khảo cổ học”, mã số đề tài: DL0000.08/20-22 giúp tôi đạt được những kết quả trong luận văn này.

Tiếp đến tôi xin gửi lời cảm ơn đến ban Lãnh đạo, phòng Đào tạo, các phòng chức năng của Học viện Khoa học và Công nghệ để luận văn được hoàn thành.

Cuối cùng, tôi xin chân thành cảm ơn khi nhận được sự ủng hộ, giúp đỡ từ gia đình và bạn bè trong suốt quá trình làm luận văn.

MỤC LỤC

MỞ ĐẦU	1
Chương 1. TỔNG QUAN NGHIÊN CỨU	2
1.1. Người Ê Đê ở Việt Nam	2
1.1.1. Các dân tộc ở Việt Nam.....	2
1.1.2. Người Ê Đê – lịch sử và văn hóa	4
1.2. Nghiên cứu cấu trúc di truyền quần thể sử dụng chỉ thị STR.....	7
1.2.1. Chỉ thị STR.....	7
1.2.2. Cấu trúc di truyền quần thể	12
1.2.3. Phương pháp phân tích cấu trúc di truyền giữa các quần thể	13
Chương 2. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU	17
2.1. ĐỐI TƯỢNG NGHIÊN CỨU	17
2.2. PHƯƠNG PHÁP NGHIÊN CỨU	18
2.2.1. Locus STR sử dụng	18
2.2.2. Phân tích mối quan hệ di truyền giữa các quần thể	18
2.2.2.1. Tính khoảng cách di truyền Fst đôi một	18
2.2.2.2. Phân tích cấu trúc quần thể	19
2.2.3. Xây dựng cây phân loại bằng phương pháp Neighbor-joining.....	20
Chương 3. KẾT QUẢ VÀ THẢO LUẬN	22
3.1. Phân tích cấu trúc di truyền dựa trên hai hệ dòng của quần thể người Ê Đê	22
3.1.1. Phân tích chỉ số khoảng cách di truyền Fst.....	22
3.1.2. Kết quả phân tích thành phần chính.....	25
3.1.3. Phân nhóm quần thể bằng STRUCTURE.....	28

3.2. Phân tích cấu trúc di truyền theo khu vực phân bố của quần thể người Ê Đê	30
3.2.1. Phân tích chỉ số khoảng cách di truyền Fst.....	31
3.2.3. Phân nhóm quần thể bằng STRUCTURE.....	36
3.2.4. Phân tích kết quả xây dựng cây phân loại.....	38
KẾT LUẬN VÀ KIẾN NGHỊ.....	42
PHỤ LỤC	47

DANH MỤC CÁC KÝ HIỆU, CHỮ CÁI VIẾT TẮT

STT	Tên viết tắt	Tên đầy đủ	Nghĩa tiếng việt
1	CODIS	Combined DNA Index System	Hệ thống Chỉ số DNA Kết hợp
2	DNA	Deoxyribonucleic acid	Deoxyribonucleic acid
3	MP	Match probability	Chỉ số khả năng trùng hợp ngẫu nhiên
4	NST	Nhiễm sắc thể	
5	NJ	Neighbor-joining	
6	PCA	Principal components analysis	Phân tích thành phần chính
7	PCR	Polymerase chain reaction	Phản ứng khuếch đại luân nhiệt
8	PD	Power of discrimination	Chỉ số phân biệt
9	PE	Power of exclusion	Chỉ số loại trừ
10	PIC	Polymorphic information content	Chỉ số đa hình
11	STR	Short tandem repeats	Các đoạn lặp lại song song ngắn

DANH MỤC BẢNG

Bảng 1. 1. Danh sách các locus có trong Bộ khuếch đại VeriFiler Express PCR	10
Bảng 2. 1. Các quần thể tham chiếu.....	18
Bảng 3. 1. Bốn nhóm người Ê Đê được chia theo địa bàn cư trú.....	31

DANH MỤC HÌNH

Hình 1.1. Bản đồ các dân tộc Việt Nam [1].....	3
Hình 1.2. 13 locus STR cốt lõi CODIS với vị trí tương ứng trên NST [16].....	9
Hình 3.1. Biểu đồ heatmap giá trị Fst đôi một của 2 nhóm quần thể Ê Đê với các quần thể tham chiếu	24
Hình 3.2. Biểu đồ scree trực quan tỉ lệ các thành phần chính.....	25
Hình 3.3. Chất lượng biểu diễn của các biến trên sơ đồ nhân tố (\cos^2)	26
Hình 3.4. Kết quả phân tích PCA hai nhóm Nie và Mlo với hai thành phần chính đầu tiên	27
Hình 3.5. Kết quả trung bình giá trị $L(K)$ với 10 lần lặp lại K	28
Hình 3.6. Kết quả biểu hiện cho A. K từ 1 đến 10, B. $K=5$	29
Hình 3.7. Biểu đồ heatmap giá trị Fst đôi một của 4 nhóm quần thể Ê Đê với các quần thể tham chiếu	Error! Bookmark not defined.
Hình 3.8. Biểu đồ scree trực quan tỉ lệ các thành phần chính.....	33
Hình 3.9. Chất lượng biểu diễn của các biến trên sơ đồ nhân tố (\cos^2)	34
Hình 3.10. Phân tích PCA của 4 nhóm EDE1, EDE2, EDE3 và EDE4 với hai thành phần chính đầu	35
Hình 3.11. Giá trị trung bình $L(K)$ với 10 lần lặp lại của K	36
Hình 3.12. Kết quả biểu hiện cho A. K từ 1 đến 10 và B. $K=5$	37
Hình 3.13. Cây phân loại của quần thể nghiên cứu với các quần thể thuộc 5 ngữ hệ	39
Hình 3.14. Cây Neighbor-Joining của quần thể nghiên cứu với các quần thể thuộc Nam Đảo.....	40

MỞ ĐẦU

Dân tộc Ê Đê là một trong 54 dân tộc được công nhận tại Việt Nam. Người Ê Đê nói tiếng Ê Đê, một ngôn ngữ thuộc nhóm ngôn ngữ Malay – Polynesis thuộc ngữ hệ Nam Đảo. Người Ê Đê là cư dân đã có mặt lâu đời ở miền Trung Tây Nguyên. Trong văn hóa và nghệ thuật của người Ê Đê có thể thấy rõ nguồn gốc hải đảo của họ. Người Ê Đê có lịch sử phát triển phức tạp, bị ảnh hưởng nhiều bởi các nhóm dân tộc khác nhau. Những giả thuyết về nguồn hải đảo của người Ê Đê chủ yếu dựa trên phân tích văn hóa, nghệ thuật thông qua sử thi và nghệ thuật kiến trúc, nghệ thuật tạo hình dân gian. Do đó nghiên cứu về cấu trúc di truyền quần thể người Ê Đê tại Việt Nam là hết sức cần thiết, bởi chúng giúp bổ sung cho những giả thuyết về cội nguồn dân tộc Ê Đê. Ngoài ra, hiện nay ở Việt Nam cũng như trên thế giới chưa có bất kỳ một nghiên cứu chính thức nào về cấu trúc quần thể người Ê Đê. Nghiên cứu này sẽ phân tích cấu trúc di truyền của quần thể người Ê Đê đồng thời đánh giá mối quan hệ di truyền với các quần thể cùng khu vực Đông Nam Á và cùng ngữ hệ Nam Đảo cũng như đánh giá tác động của yếu tố dòng họ và địa bàn cư trú đến cấu trúc di truyền quần thể người Ê Đê.

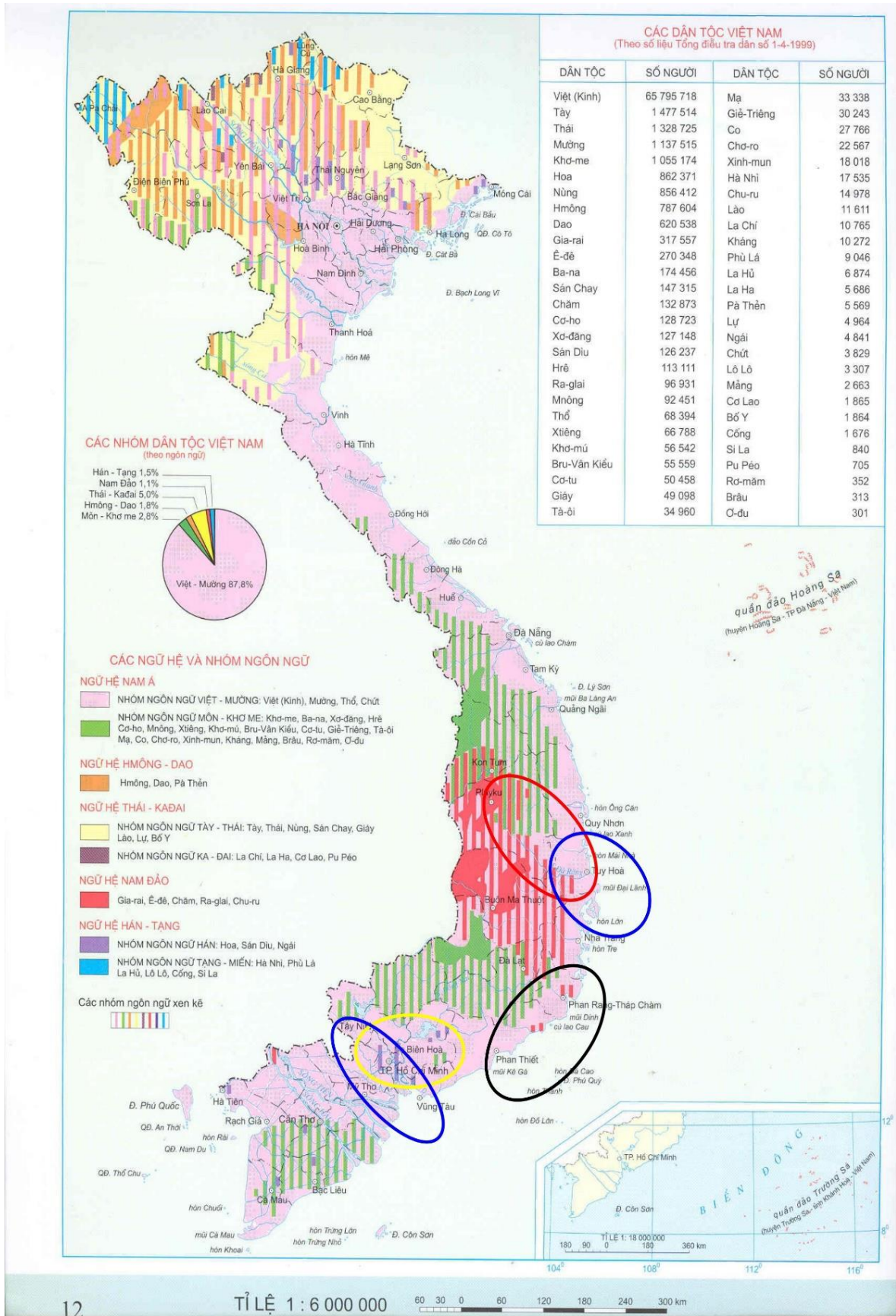
Chương 1. TỔNG QUAN NGHIÊN CỨU

1.1. Người Ê Đê ở Việt Nam

1.1.1. Các dân tộc ở Việt Nam

Dân tộc – tộc người là một cộng đồng người có mối liên hệ chặt chẽ, bền vững, có sinh hoạt kinh tế chung, có ngôn ngữ riêng và những nét văn hóa đặc thù, xuất hiện sau bộ lạc, bộ tộc và tạo thành một quần thể sinh học. Việt Nam là ngôi nhà chung của 54 dân tộc anh em. Sống trên mảnh đất Đông Dương - nơi cửa ngõ nối Đông Nam Á lục địa với Đông Nam Á hải đảo, Việt Nam là nơi giao lưu của các nền văn hoá trong khu vực với 5 nhóm ngữ hệ và 8 nhóm ngôn ngữ [1].

- Ngữ hệ Nam Á (Austroasiatic – AA)
 - Nhóm Việt - Mường : Chứt, Kinh, Mường, Thổ
 - Nhóm Môn – Khome : Ba na, Brâu, Bru-Vân kiều, Chơ-ro, Co, Cơ-ho, Cơ-tu, Gié-triêng, Hrê, Kháng, Khmerer, Khơ mú, Mạ, Mảng, M'Nông, Ô-đu, Rơ-măm, Tà-ôi, Xinh-mun, Xơ-đăng, Xtiêng
- Ngữ hệ Thái – Kadai (Tai-Kadai – TK)
 - Nhóm Tày – Thái : Bô Y, Giáy, Lào, Lự, Nùng, Sán Chay, Tày, Thái.
 - Nhóm Kadai : Cờ LaoIsan, La Chí, La ha, Pu péo.
- Ngữ hệ HMông – Dao (Hmong-Mien – HM) : Dao, Mông, Pà thên
- Ngữ hệ Nam Đảo (Austronesian-- AN) : Chăm, Chu-ru, Ê Đê, Gia-rai, Ra-glai.
- Ngữ hệ Hán – Tạng (Sino-Tibetan – ST)
 - Ngôn ngữ Hán : Hoa, Ngái, Sán diu.
 - Ngôn ngữ Tạng: Cống, Hà nhì, La hủ, Lô lô, Phù lá, Si la.
- Nhóm Nam Đảo có 5 dân tộc là: Chăm, Chu-ru, Ê Đê, Gia-rai, Ra-glai.



Hình 1.1. Bản đồ các dân tộc Việt Nam [1]

1.1.2. Người Ê Đê – lịch sử và văn hóa

Văn hóa

Người Ê Đê còn gọi là Anăk Ea Đê, Ra Đê (hay Rhađê). Người Ê Đê nói tiếng Ê Đê, một ngôn ngữ thuộc nhóm ngôn ngữ Malay – Polynesia thuộc ngữ hệ Nam Đảo. Theo Tổng điều tra Dân số và Nhà ở Việt Nam năm 2019 tại Việt Nam có dân số 398,671 người; trong đó 90.1% tổng số người Ê Đê tại Việt Nam sinh sống tại Đắk Lắk, số còn lại sinh sống tại các tỉnh Phú Yên, Đắk Nông [2]... Người Ê Đê thuộc nhóm người cư trú lâu đời nhất ở khu vực Đông Nam Á (nhóm Nam Đảo) với độ dày lịch sử, phong tục và văn hóa rất đặc trưng. Dấu vết về nguồn gốc hải đảo của dân tộc Ê Đê đã phản ánh lên từ các sử thi và trong nghệ thuật kiến trúc, nghệ thuật tạo hình dân gian. Cho đến nay, cộng đồng Ê Đê vẫn còn là một xã hội đang tồn tại những truyền thống đậm nét mẫu hệ ở nước ta. Con mang họ mẹ và người phụ nữ làm chủ trong hôn nhân. Người Ê Đê có kho tàng văn học truyền miệng rất phong phú, đặc biệt là các sử thi nổi tiếng mà cho đến nay vẫn còn đặt ra nhiều câu hỏi cho các nhà nghiên cứu văn hóa. Nhạc cụ, trang phục dân tộc, tục thờ cúng ... có nhiều điểm tương đồng với các nhóm dân tộc ở các đảo của Philippines hay Indonesia.

Ê Đê là một trong những tộc người bản địa ở Đắk Lắk, có nền văn hóa dân gian đậm đà bản sắc dân tộc. Người Ê Đê yêu ca hát, thích tấu nhạc và thường rất có năng khiếu về lĩnh vực này. Cồng chiêng, sáo, gôc, kni, đình năm là các loại nhạc cụ phổ biến của người Ê Đê và được nhiều người yêu thích. Kiến trúc nhà dài nhiều gian độc đáo, vũng chãi mang đậm tính mẫu hệ. Bộ luật tục hàng trăm điều, phản ánh những luật lệ nghiêm ngặt của cộng đồng thị tộc người Ê Đê cổ đại; với hệ thống nghi lễ dân gian như mừng lúa mới, cúng bến nước, cúng cầu mưa, cúng thần gió, rước hồn lúa, lễ đặt tên, thổi tai, lễ cưới, lễ bỏ mả, lễ rước kpan...Ấn tượng hơn cả là kể khan (hát kể), một hình thức sinh hoạt văn hóa cộng đồng, thu hút đông đảo dân làng tham gia, có ý nghĩa giáo dục truyền thống, bồi dưỡng nhân cách, ý chí kiên cường, tình yêu núi rừng, buôn làng của mọi thành viên trong cộng đồng. Tuy đã qua hàng thế kỷ với nhiều biến động lịch sử, nhưng ngày nay đồng bào Ê Đê vẫn còn giữ được những phong tục và hình thức sinh hoạt văn hóa cộng đồng mang đậm bản sắc [3].

Xã hội

Êđê là xã hội mẫu hệ và mẫu quyền điển hình nhất ở Tây Nguyên. Mọi quy tắc ứng xử trong cộng đồng xã hội, trong gia đình đều tuân theo một hệ thống luật tục (klei bhiăn) được truyền từ đời này sang đời khác. Người Êđê có nhiều dòng họ nhưng phần lớn các họ đều xuất phát từ hai dòng họ gốc là Nie hoặc Mlo, từ đó phân ra nhiều dòng họ, chi họ nhánh. Xã hội truyền thống của người Êđê mang tính chất tự quản cao. Trong tổ chức tự quản của tộc người này có sáu nhân vật đại diện : người đầu buôn, người giúp việc trực tiếp cho người đầu buôn, người xử kiện, người chăm lo công việc sản xuất, thầy thuốc và người trông coi về mặt quân sự[4, 5].

Ngôn ngữ

Tiếng nói của người Ê Đê thuộc dòng ngôn ngữ Malayo –Polynesia (ngữ hệ Nam Đảo), có quan hệ gần gũi với tiếng J'rai, Chăm, Raglai, Churu, Malaysia, Indonêsia, Philippin. Ngôn ngữ Ê Đê là ngôn ngữ đơn lập, phát triển theo xu hướng đơn âm tiết. Đây là kết quả của sự ảnh hưởng ngôn ngữ dòng Môn-Khmer. Do sự tiếp xúc với tiếng Việt -một ngôn ngữ Đơn tiết Điển hình cho nên quá trình Đơn tiết hóa ở tiếng Ê Đê diễn ra khá mạnh. Người Ê Đê tuy có chia thành các nhóm phương ngữ nhưng đến nay sự pha trộn giữa tiếng của nhóm người này với nhóm người kia diễn ra khá phổ biến do quá trình tiếp xúc văn hóa cộng cư với các nhóm khác và các tộc người khác [5].

Lịch sử

Vào đầu Công nguyên, Chiêm Thành là một trong hai vương quốc của người Malayo - Polynesia lớn trên bán đảo Đông Dương. Quốc gia này mang nhiều ảnh hưởng của nền văn minh Ấn Độ với lãnh thổ nằm dọc theo các đồng bằng eo hẹp miền Trung đến chân dãy Trường Sơn về phía Tây. Sinh hoạt chính của người Malayo - Polynesia là trồng lúa nước và buôn bán. Để mở rộng nguồn hàng, người Malayo - Polynesia mở rộng tầm kiểm soát lên các vùng rừng núi đồng thời khuất phục luôn các nhóm dân cư bản địa đã có mặt từ trước, điển hình là nhóm Ê Đê Bih. Tên gọi Ê Đê được cho là xuất phát từ Rang Đê trong bia ký của người Champa. Trong Bia Po Nagar tại tháp Po Nagar đã ghi chép về việc nhóm Rang Đê ở Ea Trang (Nha Trang) bị thu phục bởi vua Champa. Mặt khác, cuối thế kỷ VII, quân Java của Indonesia từ Biển Đông lại tràn vào

đánh phá Ea ryu (Phú Yên) và Kauthara- Ea Trang (Khánh Hòa), khiến phần lớn dân Chiêm Thành phải chạy lên cao nguyên M'Đrak. Người Rang Đê được cho là tổ tiên của người Ê Đê và Jarai, đã được ghi chép khá nhiều trong các bia ký Champa. Cuộc Nam tiến của người Việt xuống đất Champa đã tạo ra làn sóng người Rang Đê di chuyển lên vùng bình nguyên Cheo Reo, dần dần tách ra thành dân tộc Jarai như hiện nay.

Năm 1832, vương quốc Champa suy tàn, người Tây Nguyên sinh sống hoàn toàn độc lập, không trực thuộc bất kỳ chính quyền nào. Giai đoạn này họ gần như không có va chạm với bất kỳ nền văn hóa nào, kể cả người Kinh và Xiêm La. Tiếp đó, dưới sự cai trị của thực dân Pháp, Tây Nguyên trở thành lãnh thổ tự trị gọi là “Pays Montagnard du Sud Indochinois” (Xứ Thượng Miền Nam Đông Dương), viết tắt là PMSI. Người Pháp hạn chế tối đa, nếu không muốn nói là cấm hẳn, mọi phong trào di dân của người Kinh vào lãnh thổ này. Người Kinh chủ yếu là nhân viên hành chính và quân sự, chỉ tập trung quanh các thị trấn lớn như Đà Lạt, Buôn Ma Thuột. Tóm lại, giai đoạn 1832 đến 1954, cộng đồng người Tây Nguyên sống hoàn toàn biệt lập với người Kinh đồng bằng.

Sau hiệp định Genève năm 1954, Xứ Thượng Miền Nam Đông Dương bị xóa bỏ. chính quyền Ngô Đình Diệm đã đưa khoảng 700.000 lên lập nghiệp ở Tây Nguyên, Sài Gòn và vùng ven đô. Sự hiện diện quá đông đảo của những di dân cũng như chính sách Việt hóa triệt để làm đảo lộn lối sống cổ truyền của những nhóm dân cư bản địa đã có mặt từ trước, sinh hoạt văn hóa truyền thống của người Tây Nguyên mai một dần. Giai đoạn này người Tây Nguyên nói chung bị ảnh hưởng sâu sắc bởi người Kinh.

Có thể nói, quá trình lịch sử phức tạp với các cuộc xâm chiếm, việc pha trộn giữa các nhóm, tự nhiên hay bị ép buộc hay việc thay đổi địa bàn cư trú khiến cho việc duy trì tính thuần chủng của người Ê Đê trở nên khó khăn. Ngoài ra, về mặt nhân chủng học, người Ê Đê không thuần chủng, bề ngoài có nhiều nét khác nhau, đó là do kết quả của nhiều thế kỉ pha trộn với các nhóm người khác nhau.

1.2. Nghiên cứu cấu trúc di truyền quần thể sử dụng chỉ thị STR

1.2.1. Chỉ thị STR

STR là đoạn đa hình nằm trong vùng không mã hóa, có trình tự là các đoạn lặp lại của một đoạn trình tự nucleotide có độ dài khoảng 2 – 6 bp. Các STR rất phổ biến trong hệ gen người, từ những năm 1990 đã có hàng chục nghìn STRs trên các nhiễm sắc thể (NST) được phát hiện. Trong quá trình phân bào, các đoạn STR này không bị phân cắt, chúng có tính bảo thủ cao, di truyền từ bố mẹ sang con cái và khác biệt rõ ràng giữa các cá thể không có quan hệ huyết thống trực hệ. Do đó các cá thể này sẽ mang bộ số lượng đoạn lặp lại khác nhau của các STR [6, 7]. Trước đây, các STR được cho là không được phiên mã và do đó không liên quan đến biểu hiện gen [6, 7]. Tuy nhiên, ngày càng có nhiều bằng chứng cho thấy các trình tự DNA không mã hóa như STR có thể tham gia vào quá trình điều hòa gen thông qua các cơ chế khác nhau [8, 9].

Các chỉ thị STR có tần suất đột biến cao với 10^{-6} đến 10^{-2} qua mỗi thế hệ. Chakraborty và cộng sự đã chỉ ra rằng trong các locus STR không gây bệnh của con người, các lần lặp lại dinucleotide cho thấy tỷ lệ đột biến cao nhất, trong khi tỷ lệ đột biến của các chuỗi tetra-nucleotide thấp hơn 50%. Đầu những năm 1970 [10]. Hiện nay có ba cơ chế gây đột biến STR được đề xuất: (1) trao đổi chéo không đồng đều trong giảm phân; (2) cơ chế chuyển vị lại; (3) sao chép trượt sợi. Trong đó, sao chép trượt sợi được coi là mô hình chính của đột biến STR [11–13].

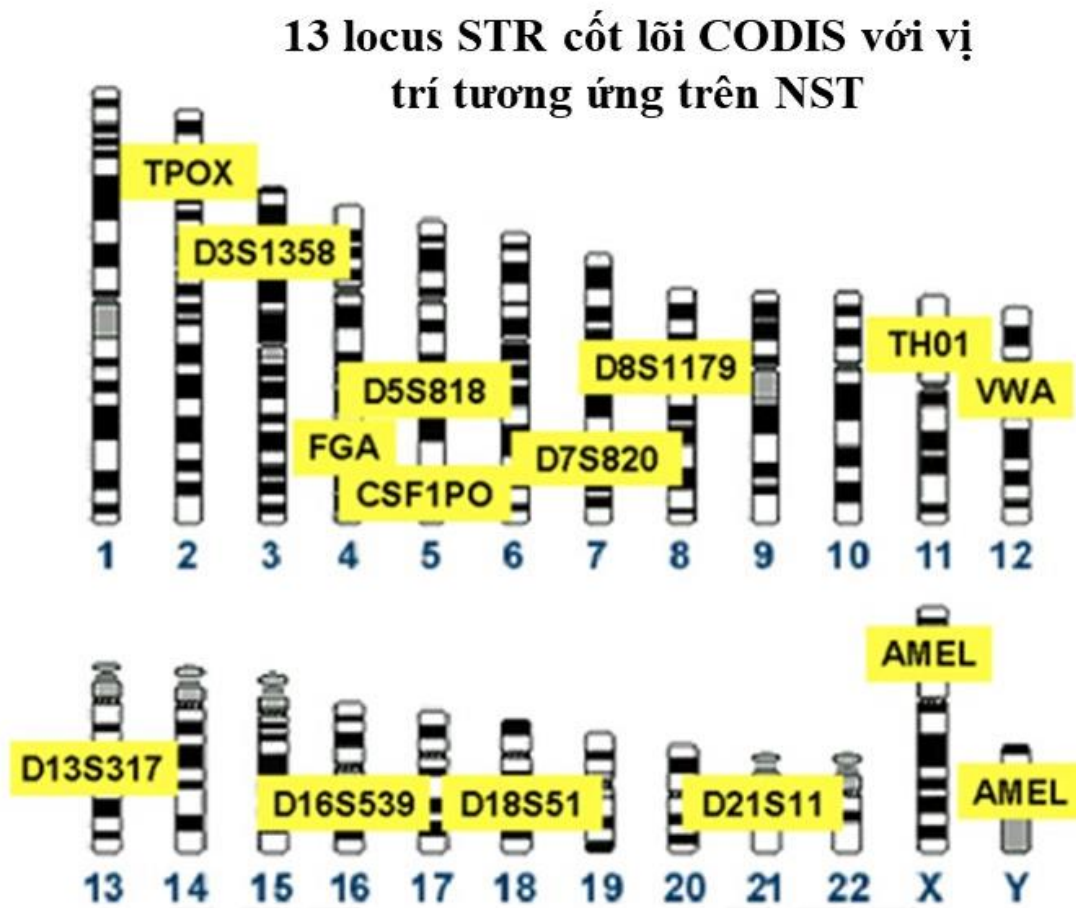
Danh pháp hay tên của từng đoạn STR được đặt theo tên của gen nếu locus này nằm một phần hoặc nằm toàn bộ trong gen. Ví dụ chỉ thị STR TH01 có nguồn gốc từ tên gen tổng hợp enzym tyrosine hydroxylase của người, nằm trên NST số 11. Chữ "TH" xuất phát từ chữ cái đầu tyrosine hydroxylase. Phần "01" của ký hiệu "TH01" xuất phát từ vùng intron 1 của gen tổng hợp enzym tyrosine hydroxylase. Các trình tự DNA nằm ngoài vùng gen thì được xác định tên bằng vị trí của chúng trên NST.

Thông thường cấu trúc của các locus STR là các đoạn lặp đơn giản, ví dụ như locus D3S1768, D19S253 và D10S2325 với 4-5 bp lặp lại. Tuy nhiên, một vài chỉ thị có cấu trúc tương đối phức tạp. Một số locus có cùng độ dài đoạn lặp nhưng khác nhau về trình tự, ví dụ như D2S1360 và D12S1064. Có nhóm locus khác nhau về thành phần motif (lặp tri, tetra, hay penta) hoặc đơn vị lặp chứa điểm đa hình như D8S1132, D9S1118 và D12S391. Cũng có locus là phức hợp của cấu trúc lặp cố định và biến đổi như đoạn Mycl-1 hay D11S554. Những locus có cấu trúc phức tạp tuy có khả năng phân loại cao nhưng lại khó cho quá trình khuếch đại[6].

Dự án STR năm 1997 đã lựa chọn ra 13 chỉ thị di truyền STR cốt lõi CODIS (Combined DNA Index System) [14, 15]. Các loci CODIS được lựa chọn là CSF1PO, FGA, TH01, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51 và D21S11 (Hình 1.2). Sau đó, năm 2017, FBI, Mỹ đã bổ sung thêm 7 loci STR là D1S1656, D2S441, D2S1338, D10S1248, D12S391, D19S433, DNA D22S1045 để tăng hiệu quả phân loại cá thể.

Mỗi STR là đa hình, nhưng số lượng alen rất ít. Thông thường, mỗi alen STR sẽ được chia sẻ bởi khoảng 5–20% cá nhân. Sức mạnh của phân tích STR bắt nguồn từ việc kiểm tra đồng thời nhiều locus STR. Càng nhiều vùng STR được sử dụng, khả năng phân biệt các cá thể càng cao. Một trong những điểm mạnh của phân tích STR là ở khả năng phân biệt thống kê của nó. Vì 20 locus hiện đang được sử dụng để phân biệt trong CODIS được di truyền độc lập, xác suất thống kê để các cá thể không liên quan về huyết thống có chung kiểu gen ở cả 20 locus là gần như bằng 0. Ví dụ, với một người có ba locus ABC đều di truyền độc lập, thì xác suất một người khác không có quan hệ huyết thống có cùng ba locus ABC sẽ bằng xác suất có locus A nhân với xác suất có locus B nhân với xác suất có locus C. Tuy nhiên, trong thực tế, có nhiều yếu điểm trong việc áp dụng STR để phân biệt cá thể. Một trong những yếu điểm của việc áp dụng STR đến từ việc không thể biết tần suất trùng khớp hồ sơ DNA có thể xảy ra trong dân số nói chung, hoặc nói cách khác, khả năng trùng khớp locus giữa hai cá thể không có chung quan hệ huyết thống trong trường hợp thực tế.

Với các đặc tính như độ đa dạng, đa hình cao, trung tính về mặt chọn lọc và kích cỡ đủ nhỏ cho quá trình khuếch đại, dữ liệu STR được sử dụng rộng rãi trong nhiều nghiên cứu khác nhau như xác định vị trí gen, sinh học pháp y, chuẩn đoán bệnh hay trong những nghiên cứu di truyền quần thể. Tuy nhiên, việc áp dụng STR vào di truyền quần thể đôi khi cần thêm những thông tin về đột biến STR. Sự khác biệt về kích thước giữa hai alen STR có thể chứa thông tin về phát sinh loài. Tuy nhiên do mức độ phổ biến của các sự kiện đột biến khác nhau có thể khác nhau đáng kể giữa các nhóm cũng như cùng một loại alen được tìm thấy ở các cá thể hoặc quần thể khác nhau có thể bắt nguồn từ các quá trình tiến hóa khác nhau nên chúng có thể dẫn đến các ước tính sai lệch về cấu trúc di truyền [15].



Hình 1.2. 13 locus STR cốt lõi CODIS với vị trí tương ứng trên NST [16]

Trong nghiên cứu này, 23 locus STR được khuếch đại sử dụng bộ khuếch đại VeriFiler Express PCR. Đây là một bộ STR khuếch đại trực tiếp với khả

năng phân biệt cao, hiệu suất vượt trội và thời gian cho kết quả nhanh. Bộ VeriFiler Express cho phép khuếch đại đa thành phần của 22 locus STR có trong bộ GlobalFiler Express, cũng như các locus Penta D và Penta E có khả năng phân biệt cao và hai dấu hiệu giới tính (Amelogenin, Yindel). Tuy nhiên, trong quá trình phân tích và so sánh, do các quần thể tham chiếu không có đủ dữ liệu của 23 locus STR, nên chỉ có 15 locus STR (gồm 13 locus STR) được sử dụng trong phân tích tiếp theo. Thông tin về vị trí các locus cũng như các allen trong ladder và thuốc nhuộm đánh dấu được liệt kê ở bảng 1.1 dưới đây.

Bảng 1.1. Danh sách các locus có trong Bộ khuếch đại VeriFiler Express PCR [17]

Locus	Vị trí	Các alen có trong thang chuẩn
D3S1358	3p21.31	9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20
vWA	12p13.31	11,12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24
D16S539	16q24.1	5, 8, 9, 10, 11, 12,13, 14, 15
CSF1PO	5q33.3-34	6, 7, 8, 9, 10, 11, 12, 13, 14, 15
TPOX	2p23-2per	5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15
Y indel	Yq11.221	1, 2
Amelogenin	X: p22.1-22.3 Y: p11.2	X, Y
D8S1179	8q24.13	5, 6, 7, 8, 9 10, 11, 12, 13, 14, 15, 16, 17, 18, 19
D21S11	21q11.2-q21	24, 24.2, 25, 26, 27, 28, 28.2, 29, 29.2, 30, 30.2, 31, 31.2, 32, 32.2, 33, 33.2, 34, 34.2, 35, 35.2, 36, 37, 38

D18S51	18q21.33	7, 9, 10, 10.2, 11, 12, 13, 13.2, 14, 14.2, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27
Penta E	15q26.2	5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26
D2S441	2p14	8, 9, 10, 11, 11.3, 12, 13, 14, 15, 16, 17
D19S433	19q12	6, 7, 8, 9, 10, 11, 12, 12.2, 13, 13.2, 14, 14.2, 15, 15.2, 16, 16.2, 17, 17.2, 18.2, 19.2
TH01	11p15.5	4, 5, 6, 7, 8, 9, 9.3, 10, 11, 13.3
FGA	4q28	13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 26.2, 27, 28, 29, 30, 30.2, 31.2, 32.2, 33.2, 42.2, 43.2, 44.2, 45.2, 46.2, 47.2, 48.2, 50.2, 51.2
D22S1045	22q12.3	8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19
D5S818	5q21-31	7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18
D13S317	13q22-31	5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16
D7S820	7q11.21-22	6, 7, 8, 9, 10, 11, 12, 13, 14, 15
D6S1043	6q16.1	9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25
D10S1248	10q26.3	8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19
D1S1656	1q42.2	9, 10, 11, 12, 13, 14, 14.3, 15, 15.3, 16, 16.3, 17, 17.3, 18.3, 19.3, 20.3
D12S391	12p13.2	14, 15, 16, 17, 18, 19, 19.3, 20, 21, 22, 23, 24, 25, 26, 27

D2S1338	2q35	11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28
Penta D	21q22.3	2.2, 3.2, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17

1.2.2. Cấu trúc di truyền quần thể

Cấu trúc quần thể là sự khác biệt về tần số alen giữa các quần thể con (subpopulation) trong một quần thể lớn hoặc giữa các quần thể với nhau. Trong một quần thể giao phối ngẫu nhiên, tần số alen dự kiến sẽ gần giống nhau giữa các nhóm. Ngược lại, với các quần thể giao phối không ngẫu nhiên, có thể phát sinh sự khác biệt về tần số alen giữa các nhóm. Ví dụ, do ảnh hưởng về mặt địa lý như một con sông khiến các cá thể ở hai bên bờ ít giao phối với nhau, do đó khi có đột biến mới, qua nhiều thế hệ nó có thể trở nên phổ biến ở một nhóm, trong khi nhóm còn lại hoàn toàn không phát sinh đột biến.

Cấu trúc quần thể phổ biến trong các quần thể tự nhiên và là kết quả của các quá trình chọn lọc tự nhiên, trôi dạt di truyền, đột biến và dòng gen. Sự trôi dạt di truyền và chọn lọc tự nhiên thúc đẩy việc tạo ra cấu trúc di truyền, ngược lại dòng gen đồng nhất các quần thể con [18]. Các rào cản đối với dòng gen có thể là các yếu tố tự nhiên như sông, các dãy núi hoặc là các yếu tố ngoại lai như tỷ lệ săn bắn hoặc nhóm động vật ăn thịt. Ngoài ra còn có các giới hạn tự nhiên do khả năng phân tán của sinh vật bị hạn chế như ở thực vật và động vật nhỏ. Trong một số trường hợp, dòng gen giữa các vùng có thể giảm xuống bằng 0, nhưng thông thường dòng gen chỉ bị giảm ở một mức độ nào đó. Khi dòng gen bị giảm, cơ hội cho sự khác biệt về tần số alen thông qua sự trôi dạt di truyền được tăng cường, kéo theo sự phát triển của cấu trúc di truyền. Thông qua việc phân tích các dữ liệu di truyền, có thể suy ra quá trình thay đổi của tần số alen. Những suy luận này cung cấp thông tin quan trọng trong nghiên cứu đa dạng, đặc biệt là để quản lý hiệu quả các loài có nguy cơ tuyệt chủng.

Việc phân tích cấu trúc quần thể dựa trên chỉ thị di truyền có thể được ứng dụng trong nhiều lĩnh vực như trong nghiên cứu tương quan toàn bộ nhiễm

sắc thể (GWAS) và nghiên cứu pháp y [19]. Trong nghiên cứu tương quan toàn bộ nhiễm sắc thể, các nghiên cứu sàng lọc các biến dị nhiễm sắc thể quen thuộc của nhiều cá thể khác nhau để xem khác biệt nào tương quan tới một đặc tính cụ thể nào đó, ví dụ như sự liên kết giữa các chỉ thị di truyền và bệnh. Sự hiện diện của cấu trúc quần thể có thể dẫn đến sự liên kết giả giữa bệnh và chỉ thị di truyền. Do đó, phân tích cấu trúc quần thể là điều kiện tiên quyết cho các nghiên cứu lập bản đồ liên kết để tránh tạo ra các mối tương quan giả hoặc thiếu các mối tương quan thực sự, giúp giảm tỷ lệ dương tính giả. Trong pháp y, việc xác định cấu trúc dân số là điều kiện tiên quyết để phát triển các bảng tham chiếu. Các bảng tham chiếu bao gồm một tập hợp các dấu hiệu di truyền có thể cung cấp thông tin về một cá thể. [20]

1.2.3. Phương pháp phân tích cấu trúc di truyền giữa các quần thể

Hiện nay, có một số phương pháp được sử dụng rộng rãi để phát hiện và đánh giá cấu trúc di truyền.

Phương pháp đầu tiên là thống kê F của Wright, đặc biệt là Fst [21]. Đây là một thống kê được sử dụng rộng rãi trong nghiên cứu quần thể và di truyền tiến hóa. Năm 1940 và 1950, hai nhà khoa học đã giới thiệu thống kê F như một công cụ để mô tả sự phân chia đa dạng di truyền trong và giữa các quần thể. Trong một bài báo xuất bản năm 1931, Wright đã tính toán về các quá trình gây ra sự khác biệt di truyền giữa các quần thể. Ông đã chỉ ra rằng mức độ khác biệt di truyền giữa các quần thể có mối quan hệ có thể dự đoán được với tốc độ của các quá trình tiến hóa quan trọng (di cư, đột biến và trôi dạt). Wright đã giới thiệu Fst là một trong ba tham số có liên quan đến nhau để mô tả cấu trúc di truyền của quần thể lưỡng bội. Các tham số này là: Fit, mối tương quan giữa các giao tử trong một cá thể so với toàn bộ quần thể; Fis, mối tương quan giữa các giao tử trong một cá thể so với quần thể con mà cá thể đó thuộc về; và Fst, mối tương quan giữa các giao tử được chọn ngẫu nhiên từ trong cùng một quần thể con so với toàn bộ quần thể [22]. Ở đây, các tham số này được xác định theo sự khác biệt của tần số kiểu gen so với định luật Hardy–Weinberg. Trong đó, giá trị Fst là một trong những số liệu thống kê được sử dụng rộng rãi nhất. Fst có giá trị từ 0 đến 1. Trong đó, giá trị 0 cho thấy hai quần thể giao phối tự

do và ngẫu nhiên với nhau. Ngược lại giá trị 1 cho thấy hai quần thể không chia sẻ alen với nhau hay hai quần thể hoàn toàn biệt lập và không giao phối với nhau

Phương pháp thứ hai sử dụng cách tiếp cận phi tham số. Năm 2006, Liu và Zhao đã đề xuất hai bước phân tích phi tham số để phân tích cấu trúc dân số quần thể từ dữ liệu di truyền. Giai đoạn đầu tiên là giảm kích thước của bộ dữ liệu kiểu gen bằng các phương pháp phân tích đa biến như phân tích thành phần chính (PCA). Phương pháp này chiếu các mẫu lên một loạt các trục trục giao, mỗi trục được tạo thành từ sự kết hợp tuyến tính của các giá trị alen hoặc các điểm đa hình đơn nucleotide (SNP). Các trục này được chọn sao cho hình chiếu của các mẫu dọc theo trục đầu tiên (hoặc thành phần chính đầu tiên) giải thích phương sai lớn nhất có thể có trong dữ liệu giữa tất cả các trục. Tương tự như vậy, phép chiếu các mẫu lên trục thứ hai sẽ tối đa hóa phương sai cho tất cả các trục có thể vuông góc với trục thứ nhất cho các thành phần tiếp theo. Giai đoạn thứ hai liên quan đến việc áp dụng các thuật toán phân cụm để xác định cấu trúc quần thể con từ dữ liệu đã giảm chiều. Một phương pháp không giả định cấu trúc khác cũng được sử dụng là phương pháp dựng cây phát sinh sử dụng dữ liệu khoảng cách di truyền bằng các thuật toán khác nhau như neighbour joining [12]. Các phương pháp phi tham số có nhiều ưu điểm, bao gồm chi phí tính toán hiệu quả và không yêu cầu giả định mô hình hóa. Các phương pháp phi tham số có chi phí tính toán hiệu quả hơn so với các phương pháp có tham số, làm cho phương pháp thứ nhất trở nên khả thi hơn để phân tích các tập dữ liệu lớn [23]. Ngoài ra, các phương pháp phi tham số không đưa ra bất kỳ giả định nào về dữ liệu di truyền, đây là một lợi thế lớn so với các phương pháp tham số. Do đó, khi có sẵn một lượng lớn dữ liệu kiểu gen, các phương pháp phi tham số được ưu tiên hơn, vì không cần xác minh các giả định của Hardy-Weinberg và trạng thái cân bằng liên kết. Hơn nữa, vì các phương pháp này không phụ thuộc vào việc ước tính tần số alen, nên chúng không bị ảnh hưởng khi quần thể có ít cá thể đại diện. Phương pháp này cũng được sử dụng để suy luận về các sự kiện tiến hóa. Ví dụ: sự phân tán của các phép chiếu mẫu dọc theo một đường được cho là dấu hiệu của các mẫu được trộn lẫn giữa hai quần

thể, trong khi mối tương quan giữa các thành phần chính và trục địa lý đã được hiểu là bằng chứng cho các làn sóng di cư [24]

Một phương pháp dựa trên mô hình thay thế được phát triển gần đây bởi Pritchard và cộng sự. (2000) và được triển khai bởi phần mềm STRUCTURE giúp phân định các cụm cá thể trên cơ sở kiểu gen của chúng tại nhiều locus bằng phương pháp Bayesian. Phần mềm STRUCTURE sử dụng ước tính Markov Chain Monte Carlo (MCMC), bắt đầu với chỉ định ngẫu nhiên các cá thể vào một số nhóm được xác định trước, sau đó ước tính tần số biến thể của nhóm và các cá thể được chỉ định lại dựa trên các ước tính tần suất đó. Điều này được lặp lại nhiều lần trong bước burning để các kết quả tiến gần đến giá trị ước tính đáng tin cậy của tần số alen đáng tin cậy trong quần thể và xác suất cá thể thuộc quần thể đấy. Phần mềm tính toán khả năng xảy ra của dữ liệu đối với một dải giá trị K (số quần thể) bằng cách tính xác suất hậu nghiệm của K , được gọi là X , ký hiệu $X|K$. Vì K không phải là một giá trị tuyệt đối nên mọi giá trị cần được xem xét cẩn thận và tính đến các đặc điểm của quần thể được lấy mẫu. Phần mềm STRUCTURE tính toán khả năng xảy ra của dữ liệu (X) cho phạm vi giá trị K và số thực của K được xác định bởi giá trị tối đa của $\ln P(X|K)$. Tuy nhiên, Evanno và cộng sự đã phát hiện ra rằng giá trị cực đại không phải lúc nào cũng cung cấp đúng giá trị K trong thực tế. Thay vào đó, giá trị tối đa của tốc độ thay đổi (ΔK) trong $\ln P(X|K)$ giữa các giá trị K kế tiếp sẽ suy ra chính xác số cụm di truyền thực sự trong dữ liệu [25].

1.2.4. Tình hình nghiên cứu trong và ngoài nước

Người Ê Đê có lịch sử phát triển phức tạp, bị ảnh hưởng nhiều bởi các nhóm dân tộc khác nhau. Những giả thuyết về nguồn hải đảo của người Ê Đê chủ yếu dựa trên phân tích văn hóa, nghệ thuật thông qua sử thi và nghệ thuật kiến trúc, nghệ thuật tạo hình dân gian. Những nghiên cứu về dân tộc Ê Đê thường tập trung vào ngôn ngữ hoặc văn hóa cũng như chế độ mẫu hệ hết sức đặc trưng của dân tộc này. Hiện nay ở Việt Nam, có rất ít những nghiên cứu về dữ liệu di truyền của quần thể người Ê Đê tại Việt Nam. Nghiên cứu của Macholdt dựa trên dữ liệu mt DNA và nhiễm sắc thể Y cho thấy quần thể người Ê Đê có sự trao đổi mạnh mẽ với quần thể ở Gia Lai [26]. Hay bài báo của

Pham và cộng sự đã cung cấp cơ sở dữ liệu STR nhiễm sắc thể thường đầu tiên của người Ê Đê cũng như đánh giá mối quan hệ của quần thể này với một vài các quần thể thuộc cùng ngữ hệ [27]. Tuy nhiên, các nghiên cứu này chưa phân tích sâu về cấu trúc nội tại của quần thể người Ê Đê. Với lịch sử phát triển phức tạp, gắn liền với các quá trình thực dân hóa hoặc cưỡng bức di dời, và sự kết hợp với các dân tộc khác, người Ê Đê có sự đa dạng trong văn hóa cũng như về mặt nhân chủng học. Do đó, nghiên cứu về sự đa dạng của quần thể người Ê Đê là rất cần thiết vì chúng có thể bổ sung cho các giả thuyết sẵn có về sự phát triển của dân tộc này.

Chương 2. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1. ĐỐI TƯỢNG NGHIÊN CỨU

Bộ dữ liệu kiểu gen STR của 397 cá thể người Ê Đê thu thập năm 2020 tại tỉnh Đắk Lắk, Việt Nam và không có quan hệ huyết thống trực hệ liên tiếp ba đời gần nhất. Việc thu thập mẫu được căn cứ theo hồ sơ nhân thân được cung cấp bởi Công an các quận, huyện trực thuộc các tỉnh, thành phố. Các mẫu thu thập được lưu trữ tại Trung tâm giám định DNA, Viện Công nghệ sinh học.

Bộ dữ liệu bao gồm kết quả khuếch đại 23 locus STR nhiễm sắc thể thường gồm D3S1358, vWA, D16S539, CSF1PO, TPOX, D8S1179, D21S11, D18S51, D2S441, D19S433, TH01, FGA, D22S1045, D5S818, D13S317, D7S820, D10S1248, D1S1656, D12S391, D2S1338, D6S1043, Penta D, Penta E. Trong đó, tổng cộng 255 alen được tìm thấy trong quần thể người Ê Đê, với trung bình là 11,087 alen trên mỗi locus. Sự hiện diện của alen 12.1 của D7S820, alen 5 của D16S539, alen 27 của D2S1338 và alen 17.2 của D19S433 ở người Ê Đê và sự vắng mặt của chúng trong các quần thể Nam Đảo khác là rất đáng chú ý. Trong số các alen hiếm này, alen 27 của D2S1338 và alen 17.2 của D19S433 đã được quan sát thấy ở người Kinh Việt Nam, đã được báo cáo trong các nghiên cứu trước đây. Hệ số dị hợp quan sát (H_{obs}) thấp nhất ở locus TPOX và cao nhất ở locus FGA. Chỉ số phân biệt kết hợp (CPD) và chỉ số loại trừ kết hợp lần lượt là 0,999 999 999 999 999 999 999 999 950 và 0,999 999 999 999 999 998, cho thấy mức độ phân biệt cao của 23 locus STR. Trong đó, locus D12S391 được coi là chỉ thị mạnh của người Ê Đê với các giá trị PD (0,9631), PE (0,8776) và PIC (0,8401) cao nhất trong tất cả các locus. 22 trong số 23 locus ở trạng thái cân bằng Hardy-Weinberg, ngoại trừ D10S1248, với giá trị p đáng kể ngay cả sau khi áp dụng hiệu chỉnh Bonferroni ($\alpha = 0,05/23$). Do đó, locus này sẽ không được sử dụng trong các nghiên cứu như phân nhóm bằng STRUCTURE. Tất cả các locus được kế thừa độc lập và không tìm thấy giá trị mang ý nghĩa thống kê nào trong phép thử mất cân bằng liên kết [27].

Dựa theo thông tin về dòng họ và địa bàn cư trú, quần thể nghiên cứu được chia thành các nhóm nhỏ. Các nhóm này được phân tích cấu trúc di

truyền và so sánh với các quần thể tham chiếu khác cùng ngữ hệ Nam Đảo và cùng khu vực địa lý.

2.2. PHƯƠNG PHÁP NGHIÊN CỨU

2.2.1. Locus STR sử dụng

Tuy bộ dữ liệu bao gồm 23 locus STR nhiễm sắc thể thường nhưng trong các phân tích tiếp theo chỉ có 15 locus STR được lựa chọn để phân tích gồm D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX, D18S51, D5S818 và FGA vì chúng bao gồm đủ 13 locus được CODIS khuyến cáo (thêm hai locus là D2S1338, D19S433) và chúng có ở tất cả các quần thể tham chiếu khác. Ngoài ra, bộ locus này cũng được dùng để phân tích cấu trúc của quần thể ở nhiều nghiên cứu khác [28–31].

2.2.2. Phân tích mối quan hệ di truyền giữa các quần thể

Để xác định mối quan hệ quần thể, nghiên cứu sử dụng hai phương pháp phân nhóm dựa trên khoảng cách di truyền và phân nhóm dựa trên mô hình theo giả thuyết về cấu trúc quần thể.

2.2.2.1. Tính khoảng cách di truyền Fst đôi một

Để đánh giá mối tương quan giữa dòng họ và địa bàn cư trú với cấu trúc di truyền của người Ê Đê, quần thể nghiên cứu được phân chia thành các nhóm nhỏ dựa trên hệ dòng (Nie và Mlo) và địa bàn cư trú (EDe1, EDe2, EDe3, EDe4). Các nhóm quần thể tham chiếu được sử dụng được thống kê ở bảng dưới đây.

Bảng 2. 1. Các quần thể tham chiếu

STT	Quần thể	Kích cỡ mẫu	Ký hiệu	Vị trí	Ngữ hệ	Tài liệu
1	Bố Y – Hà Giang	41	BY_HG	Việt Nam	Thái - Kadai	[28]
2	Bố Y – Lào Cai	55	BY_LC	Việt Nam	Thái - Kadai	[28]

3	Li	50	LI	Trung Quốc	Thái - Kadai	[32]
4	BlackTai	50	BT	Thái Lan	Thái - Kadai	[31]
5	Lao Isan	50	LAO	Thái Lan	Thái - Kadai	[31]
6	Lue	50	LUE	Thái Lan	Thái - Kadai	[29]
7	Laotian	50	LAT	Lào	Thái - Kadai	[31]
8	Kinh	50	KINH	Việt Nam	Nam Á	[33]
9	Mon	47	MON	Thái Lan	Nam Á	[31]
10	Khmer	50	KHM	Thái Lan	Nam Á	[31]
11	Bru	50	BRU	Thái Lan	Nam Á	[31]
12	Tibetan	21	TIB	Trung Quốc	Hán - Tạng	[34]
13	Nu	34	NU	Trung Quốc	Hán - Tạng	[34]
14	Han	50	HAN	Trung Quốc	Hán - Tạng	[34]
15	Yi	50	YI	Trung Quốc	Hán - Tạng	[34]
16	Ami	50	AMI	Đài Loan	Nam Đảo	[35]
17	Paiwan	38	PAI	Đài Loan	Nam Đảo	[35]
18	Atayal	50	ATA	Đài Loan	Nam Đảo	[35]
19	Bunun	24	BUN	Đài Loan	Nam Đảo	[35]
20	Brunei	50	Brunei	Brunei	Nam Đảo	[36]
21	Marquesan	78	MAR	Marquesans	Nam Đảo	[37]

Với phương pháp dựa trên khoảng cách di truyền, giá trị khoảng cách di truyền đôi một được tính toán bằng phần mềm Arlequin và được trực quan hóa bằng gói ggplot2 của phần mềm R [38]. Sau đó, ma trận được vẽ theo hai chiều sử dụng phương pháp phân tích thành phần chính PCA và trực quan hóa bằng gói FactoMineR của phần mềm R [39].

2.2.2.2. Phân tích cấu trúc quần thể

Với phương pháp dựa trên mô hình, nghiên cứu sử dụng phương pháp phân nhóm Bayes bằng phần mềm STRUCTURE v2.3.4 với mô hình quần thể ADMIXTURE. Các quần thể tham chiếu được liệt kê tại bảng 2.1. Để ước tính giá trị K, 10 lần lặp lại của K từ 1 đến 10 đã được thực hiện với 10.000 bước Markov Chain Monte Carlo (MCMC) cho mỗi lần chạy. Thông thường, STRUCTURE chỉ sử dụng duy nhất dữ liệu di truyền để đánh giá quần thể. Tuy nhiên, trong nhiều trường hợp, có tín hiệu thường quá yếu để mô hình tiêu

chuẩn của STRUCTURE có thể phát hiện cấu trúc. Để cải thiện hiệu suất, Hubisz đã phát triển các mô hình mới sử dụng thông tin vị trí để hỗ trợ phân nhóm. Các mô hình mới có thể cung cấp suy luận chính xác về cấu trúc quần thể và tổ tiên của từng cá nhân trong các tập dữ liệu với tín hiệu yếu. Cơ sở lý luận của mô hình này như sau: Thông thường, STRUCTURE giả định rằng tất cả các phân vùng của các cá nhân đều có khả năng tiên nghiệm gần như bằng nhau. Vì có vô số các phân vùng có thể có nên cần dữ liệu có hàm lượng thông tin cao để phân tích có ý nghĩa. Ngược lại, các mô hình LOCPRIOR cho rằng trong thực tế các cá nhân từ cùng một địa điểm lấy mẫu thường đến từ cùng một quần thể. Do đó, các mô hình LOCPRIOR được thiết lập để hy vọng rằng các vị trí lấy mẫu có thể cung cấp thông tin về tổ tiên. Một số ưu điểm khác của mô hình LOCPRIOR là (i) chúng không có xu hướng tìm cấu trúc khi không có cấu trúc nào; (ii) phần mềm có thể bỏ qua thông tin lấy mẫu khi tổ tiên của các cá nhân không tương quan với các địa điểm lấy mẫu; và (iii) các mô hình cũ và mới về cơ bản đưa ra các câu trả lời giống nhau khi tín hiệu về cấu trúc dân số rất mạnh [40]. STRUCTURE HARVESTER được sử dụng để xác định giá trị K tối ưu. Kết quả phân tích Structure sau đó được biểu diễn bằng phần mềm CLUMPAK [41]

2.2.3. Xây dựng cây phân loại bằng phương pháp Neighbor-joining

Trong phân tích mối quan hệ di truyền của quần thể người Ê Đê với các quần thể khác, cây phân loại Neighbor-joining (NJ) dựa trên chỉ số khoảng cách di truyền Nei [42] được xây dựng sử dụng phần mềm POPTREEv2 và được trực quan bằng phần mềm FIGTREEv1.4.4. Cây phân loại đầu tiên được xây dựng với các quần thể thuộc 5 ngữ hệ có ở Việt Nam, gồm:

- Ngữ hệ Nam Đảo: Ami (AMI), Atayal (ATA), Bunun (BUN), Paiwan (PAI) [35], Samoa (SAM), Bali (BAL), Java (JAV) [43], Philippines (PHI) [44], Malaysian Malay (MAL) [45], Madagascar (MAD), Tonga (TON) [46], Marquesan (MAR) [37], Brunei (BRU)[36].
- Ngữ hệ Hán – Tạng: Tibetan (TIB), Nu (NU), Hán (HAN), Yi (YI)
- Ngữ hệ H' Mông – Dao: H' Mông [47], Dao [47]
- Ngữ hệ Thái – Kadai: Bô Y- Hà Giang (BY-HG), Bô Y – Lào Cai (BY-LC),

- BlackTai (BT), Thái (THA) [47], Tày (TAY) [47], Nùng (NUNG) [47].
- Ngữ hệ Nam Á: Laotian (LAT), LaoIsan(LAO), Khmer – Việt Nam (KHM-VN) [47], Kinh (KINH).

Sau đó, để đánh giá mối quan hệ giữa các nhóm Nam Đảo, cây phân loại thứ hai được xây dựng cho phân tích mối quan hệ giữa quần thể nghiên cứu với 13 quần thể Nam Đảo như trên.

Chương 3. KẾT QUẢ VÀ THẢO LUẬN

3.1. Phân tích cấu trúc di truyền dựa trên hai hệ dòng của quần thể người Ê Đê

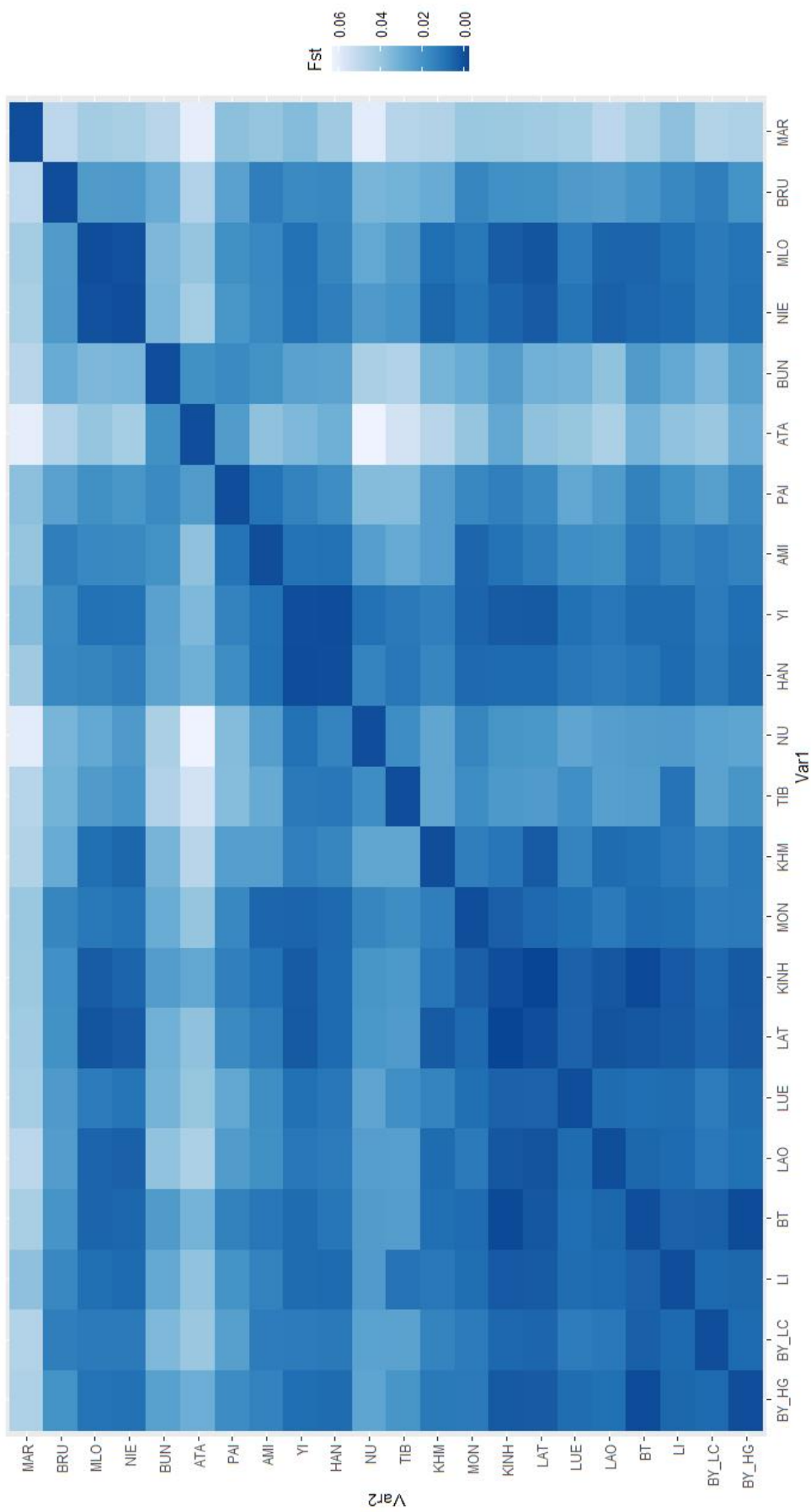
Theo các nhà dân tộc học, tuy đồng bào Ê Đê sinh sống tại Việt nam có khoảng 102 dòng họ khác nhau nhưng các dòng họ này đều được hình thành từ hai hệ dòng gốc là Nie và Mlo. Theo nguyên tắc phát triển dòng họ, các họ mới được tiếp nối trên cơ sở hai hệ dòng này. Tuy nhiên, trong thực tế, quá trình phát triển các dòng họ mới cũng như mối quan hệ giữa hệ dòng với các dòng họ, các chi họ là rất phức tạp. Có thể nói, sự phát triển dòng họ với người Ê Đê không đơn thuần là “sự tiếp nối” mà còn là “sự tạo mới” và chúng thường gắn liền với sự phát triển buôn làng. Việc xác định thuộc hệ dòng sẽ ảnh hưởng đến hôn nhân của các cá thể thuộc chi họ đấy, vì theo luật tục của người Ê Đê, hôn nhân cùng hệ dòng bị cấm tuyệt đối. Trước đây, luật tục này được tuân thủ một cách ngặt nghèo khi hôn nhân giữa những người cùng dòng họ nhưng không có quan hệ huyết thống cũng bị cấm. Trong từng buôn làng - đơn vị xã hội cơ bản và duy nhất của người Ê Đê cũng sẽ có các chi, họ thuộc cả hai dòng họ với một chi họ là hạt nhân. Tuy nhiên, sau 1975, với các chính sách về phát triển kinh tế và xu hướng toàn cầu hóa, buôn làng cổ truyền Tây Nguyên có xu hướng mở hơn với nhiều thay đổi, đặc biệt là trong quan điểm về hôn nhân giữa các dòng họ. Sự phân biệt giữa các nhóm Ê Đê dần trở nên mờ nhạt và ý thức coi mình là dân tộc Ê Đê nói chung trở nên phổ biến. Ngoài ra, việc kết hôn với các nhóm dân tộc khác cũng trở nên phổ biến hơn. Do đó, ảnh hưởng của quan điểm duy trì kết nối giữa hai hệ dòng có thể bị suy giảm. Xuất phát từ cơ sở trên, nghiên cứu này sử dụng quy tắc phân chia dòng họ của người Ê Đê để đánh giá mối tương quan giữa dòng họ và cấu trúc di truyền của quần thể người, quần thể nghiên cứu được phân chia theo hai hệ dòng lớn là Mlo và Nie dựa trên tên họ của người cho mẫu ký hiệu lần lượt là Nie và Mlo. Các họ được phân chia dựa trên nghiên cứu về mặt xã hội học trước đó [3–5].

3.1.1. Phân tích chỉ số khoảng cách di truyền Fst

Dữ liệu của 20 quần thể tham chiếu được sử dụng trong phân tích (Xem bảng 2.1). Trong tổng 231 giá trị Fst thu được, có 219 giá trị (94.8%) có ý nghĩ

thống kê và 12 giá trị (5.2%) không có ý nghĩa thống kê ($p > 0.05$). Nhìn chung, phần lớn các quần thể đều có khác biệt di truyền không lớn ($F_{st} < 0.001$) (Phụ lục 1). Hình 3.1 là biểu đồ heatmap giá trị khoảng cách di truyền F_{st} được tính toán theo cặp đôi một giữa các quần thể bằng phần mềm Arlequin. Trong đó, giá trị F_{st} càng thấp thì màu xanh dương biểu hiện càng đậm, tức hai quần thể càng gần gũi về mặt di truyền. Ngược lại, giá trị F_{st} càng cao thì màu xanh dương biểu hiện càng nhạt, tức hai quần thể xa nhau về mặt di truyền. Trong đó, hai quần thể nghiên cứu Nie và Mlo cho thấy khoảng cách gần gũi về mặt di truyền với nhau và với các quần thể cùng khu vực như Bó Y_Hà Giang, Bó Y_Lào Cai, LI, BT, LaoIsan, LUE, Laotian, Kinh, Mon, Khmer (được biểu hiện bằng màu xanh đậm). Kết quả này cho thấy khoảng cách di truyền giữa các quần thể bị ảnh hưởng lớn bởi vị trí địa lý. Nhận định này có thể giải thích dễ dàng bởi sự cách biệt về khoảng cách địa lý có thể cô lập một quần thể sinh học, như các quần thể sinh vật trên đảo, dẫn đến hạn chế khả năng ngẫu phối. Ngược lại, các quần thể sinh sống gần nhau có xu hướng ngẫu phối tự do trong tự nhiên. Đối với quần thể người, các sự kiện di nhập dân diễn ra phổ biến, nhất là ở các vùng địa lý thuận lợi, dẫn đến vốn gen giữa các nhóm người dễ dàng pha trộn vào nhau.

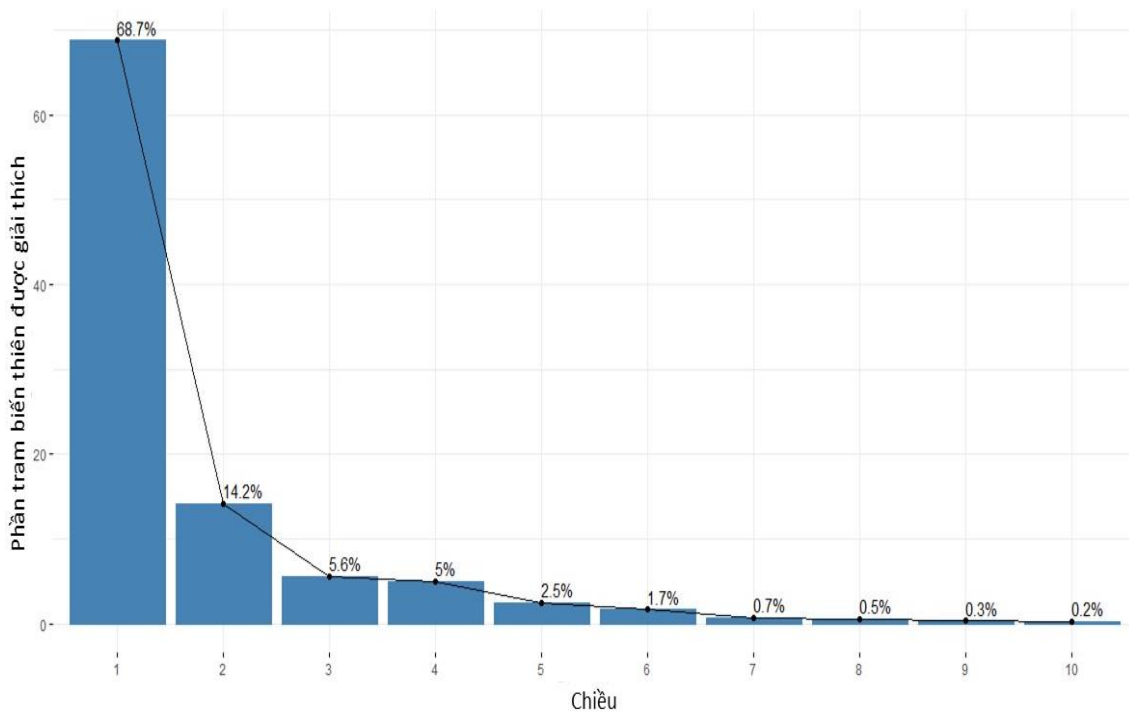
Mặt khác, các quần thể thuộc cùng ngữ hệ Nam Đảo như nhóm quần thể thu mẫu ở Đài Loan (Ami, Atayal, Paiwan, Bunun), quần thể Brunei, quần thể Marquesan hay hai nhóm quần thể nghiên cứu Nie và Mlo cho thấy sự khác biệt di truyền lớn hơn với nhóm các quần thể có vị trí địa lý gần nhau. Cụ thể, có những quần thể cho thấy sự khác biệt rõ ràng về mặt di truyền rõ ràng với các quần thể khác như quần thể Marquesan hay quần thể Atayal (biểu hiện bằng màu xanh nhạt).



Hình 3. 1. Biểu đồ heatmap giá trị F_{st} đôi một của 2 nhóm quần thể Ê Đê với các quần thể tham chiếu

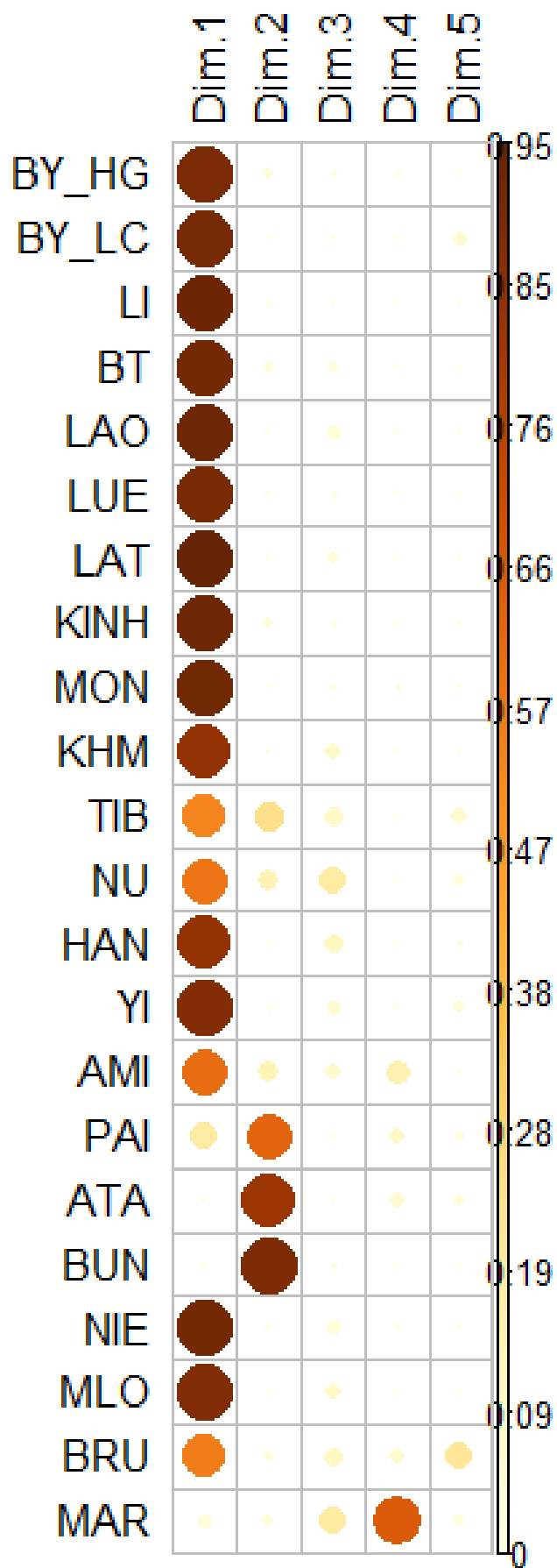
3.1.2. Kết quả phân tích thành phần chính

Bảng giá trị khoảng cách di truyền của 20 quần thể được sử dụng cho phân tích thành phần chính (PCA). Đầu tiên, lượng biến thiên được biểu hiện bởi mỗi chiều (thành phần chính) được kiểm tra (Hình 3.2). Trong đó, các thành phần chính đầu tiên cho thấy lượng biến thiên biểu hiện lớn nhất. Cụ thể, hai thành phần chính đầu tiên biểu hiện lên đến 82,9%, một giá trị biểu hiện tương đối cao. Không có một quy định nào về giá trị chấp nhận được của tỷ lệ biểu hiện của các thành phần chính, tuy nhiên thông thường tỉ lệ biểu hiện của ba thành phần chính đầu lên tới 70% được coi là chấp nhận được.



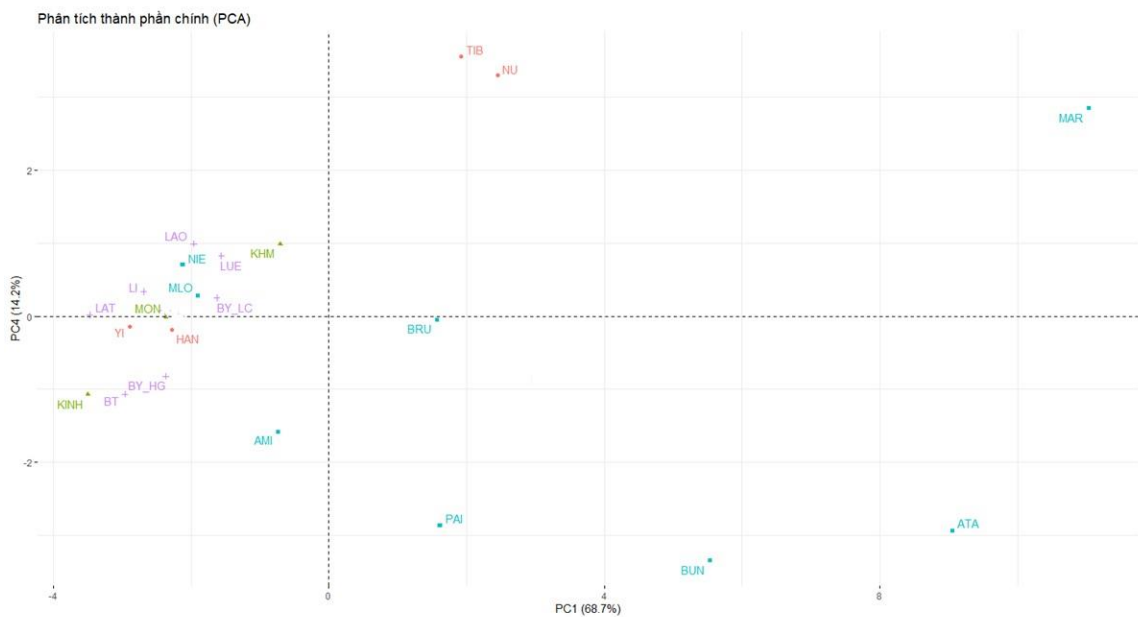
Hình 3. 2. Biểu đồ scree trực quan tỉ lệ các thành phần chính

Hình 3.3 là kết quả biểu hiện chất lượng biểu diễn của các biến trên sơ đồ nhân tố, còn được gọi là \cos^2 (cosin bình phương, tọa độ bình phương). Giá trị \cos^2 phản ánh khả năng đại diện cho các biến của các thành phần chính. Trong đó, khả năng đại diện càng tốt thì màu nâu sẽ càng đậm và ngược lại. Như có thể thấy trên hình, hầu hết các mẫu đều có vòng tròn màu nâu đậm với hai thành phần chính đầu tiên. Điều này phản ánh khả năng đại diện tốt của hai thành phần chính đầu cho các mẫu.



Hình 3. 3. Chất lượng biểu diễn của các biến trên sơ đồ nhân tố (\cos^2)

Hình 3.4 là kết quả phân tích PCA của hai quần thể nghiên cứu và 20 quần thể tham chiếu. Nhìn chung, phân tích PCA không cho thấy một phân cụm cụ thể nào. Tuy nhiên, có thể thấy, hai nhóm quần thể nghiên cứu Nie và Mlo nằm gần nhau và gần các nhóm quần thể ở cùng khu vực địa lý như Việt Nam, Lào, Thái Lan và Trung Quốc (trừ TIB và NU) ở khu vực bên trái của biểu đồ. Điều này cho thấy sự ảnh hưởng của khoảng cách di truyền đến mối quan hệ di truyền giữa các quần thể. Ngược lại, phân tích chưa quan sát được mối liên hệ giữa khoảng cách di truyền và ngữ hệ giữa các quần thể. Ví dụ như ở quần thể Nam Đảo, các quần thể (biểu hiện với màu xanh dương) nằm rải rác khắp bản đồ ở các góc phần tư khác nhau. Đặc biệt quần thể MAR nằm cách biệt hẳn với các nhóm ở rìa bản đồ cho thấy sự khác biệt về di truyền với các nhóm quần thể khác.

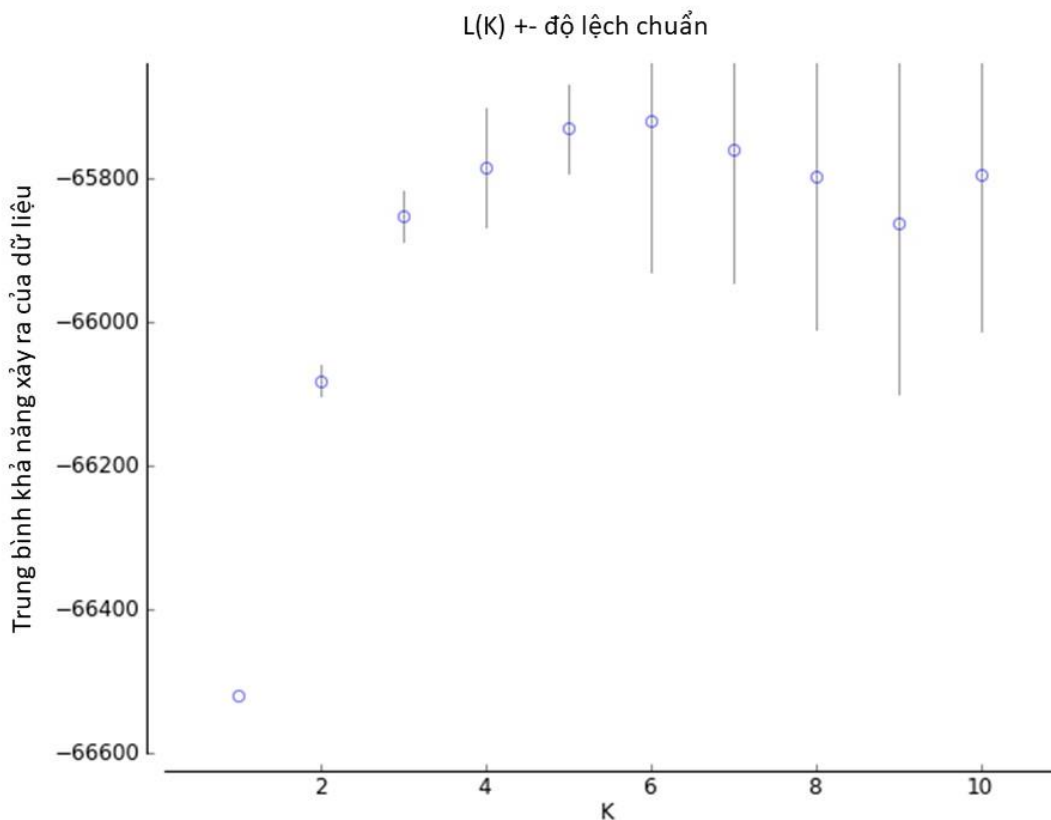


Hình 3. 4. Kết quả phân tích PCA hai nhóm Nie và Mlo với hai thành phần chính đầu tiên

Màu đỏ: nhóm ngữ hệ Hán – Tạng. Màu xanh nước biển: nhóm ngữ hệ Nam Đảo. Màu xanh lá: nhóm ngữ hệ Nam Á. Màu tím: nhóm ngữ hệ Thái - Kadai

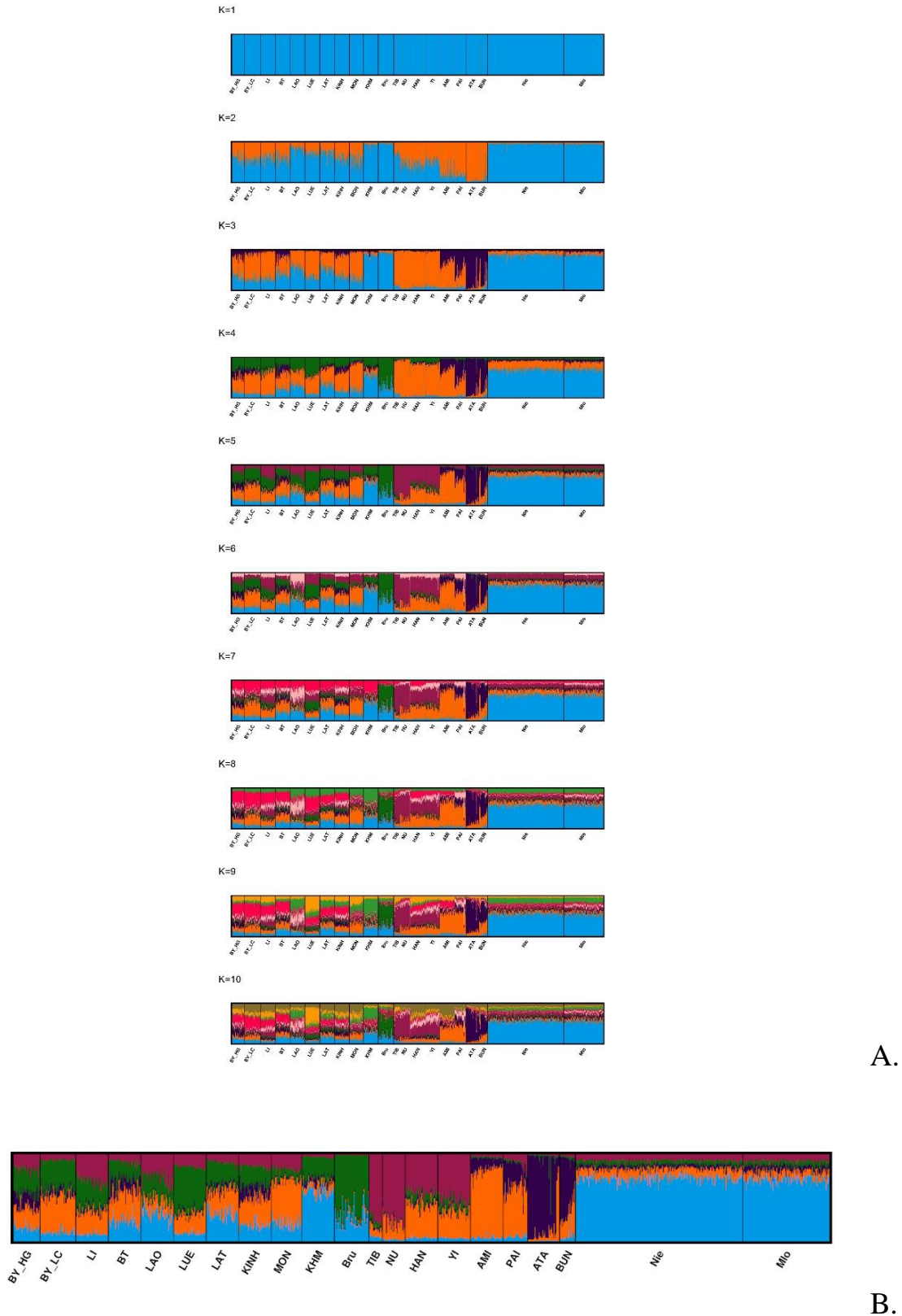
3.1.3. Phân nhóm quần thể bằng STRUCTURE

Để ước tính cấu trúc quần thể, phương pháp phân cụm Bayesian được triển khai với phần mềm STRUCTURE. Do giá trị F_{st} trung bình thấp (0,0188), mô hình LOCPRIOR được sử dụng để bổ sung thông tin về vị trí lấy mẫu [48]. Phương pháp này trước đây được báo cáo là hiệu quả cho các nhóm quần thể ít khác biệt. Dựa trên giá trị trung bình của xác suất của giá trị $L(K)$, giá trị $K = 5$ được xác định là tối ưu (Hình 3.5).



Hình 3. 5. Kết quả trung bình giá trị $L(K)$ với 10 lần lặp lại K

Kết quả của STRUCTURE cũng ủng hộ về tính khác biệt của các quần thể thuộc ngữ hệ Nam Đảo cũng như tính đồng nhất về di truyền của hai nhóm Ê Đê (Hình 3.6). Ở $K=5$, có một thành phần di truyền chung được biểu thị bằng cụm xanh nước biển ở các quần thể thu mẫu ở Việt Nam, Lào và Thái Lan. Cụm này xuất hiện với tỷ lệ cao hơn ở hai nhóm Ê Đê và thấp hơn ở các nhóm Đông Nam Á khác. Ngoài ra, hai quần thể BRU và KHMER chỉ đặc trưng với hai cụm xanh lá và xanh nước biển. Các nhóm quần thể Trung Quốc như TIB, NU, HAN và YI đặc trưng bởi cụm màu tím hồng.



Hình 3. 6. Kết quả biểu hiện cho A. K từ 1 đến 10, B. K=5

Mỗi cá thể được đại diện bởi một cột duy nhất và chia thành các phân đoạn có kích thước và màu sắc tương ứng với tỷ lệ của một cụm cụ thể. Các quần thể được phân cách bằng các vạch đen

Các quần thể Đài Loan nói tiếng Nam Đảo cũng cho thấy sự khác biệt di truyền tương đối rõ ràng giữa họ với các quần thể khác. Với tỷ lệ cụm vàng cam hoặc tím cao vượt trội đã phản ánh sự khác biệt di truyền giữa nhóm này với các quần thể khác. Không chỉ vậy, dù cùng được coi là nhóm người Nam Đảo gốc, hai nhóm Ami, Paiwan và Atayal, Bunun cho thấy sự khác biệt về mặt di truyền với cụm màu vàng cam chiếm ưu thế ở nhóm Ami, Paiwan và cụm màu tím chiếm ưu thế ở nhóm Atayal, Bunun.

Với hai nhóm Nie và Mlo thuộc quần thể người Ê Đê, họ có cấu trúc di truyền tương đồng nhau với cụm màu xanh dương đặc trưng chiếm tới 80% và sự hòa trộn theo tỷ lệ từ cao đến thấp của các cụm màu vàng cam, màu tím hồng, màu xanh lá và màu tím đậm. Như vậy, có thể nói hai quần thể này bị ảnh hưởng về mặt di truyền bởi nhóm Ami, Paiwan nhiều hơn là nhóm Atayal và Bunun. Hai nhóm này Nie và Mlo cũng cho thấy sự pha trộn với nguồn gen mà chiếm đa số trong nhóm Sino-Tibetan (màu tím hồng và màu cam).

Sự tương đồng về mặt di truyền của hai nhóm Nie và Mlo có thể được giải thích bởi văn hóa và tục lệ của người Ê Đê. Theo quan niệm truyền thống, người Ê Đê cho rằng việc kết hôn đồng tộc là loạn luân và sẽ khiến cho các vị thần nổi giận, ảnh hưởng đến cả buôn làng. Tính chặt chẽ của luật cấm hôn nhân dòng ngoài còn được thể hiện ở việc cấm những người cùng họ dù không có quan hệ huyết thống. Người Ê Đê cũng có ý thức cao về vấn đề dòng họ và hệ dòng của mình. Thông tin về dòng tộc sẽ được người thế hệ trước truyền lại cho thế hệ sau [3]. Do đó, hai nhóm Nie và Mlo thường xuyên có sự trao đổi về mặt di truyền. Điều này dẫn đến sự tương đồng về mặt di truyền của hai nhóm này.

3.2. Phân tích cấu trúc di truyền theo khu vực phân bố của quần thể người Ê Đê

Dựa trên địa bàn cư trú, người Ê Đê được phân biệt thành các nhóm với tập tục, ngôn ngữ, tín ngưỡng và có thể là cả sự phân biệt về mặt di truyền. Quần thể người Ê Đê được chia thành 4 quần thể dựa theo 4 khu vực là khu vực

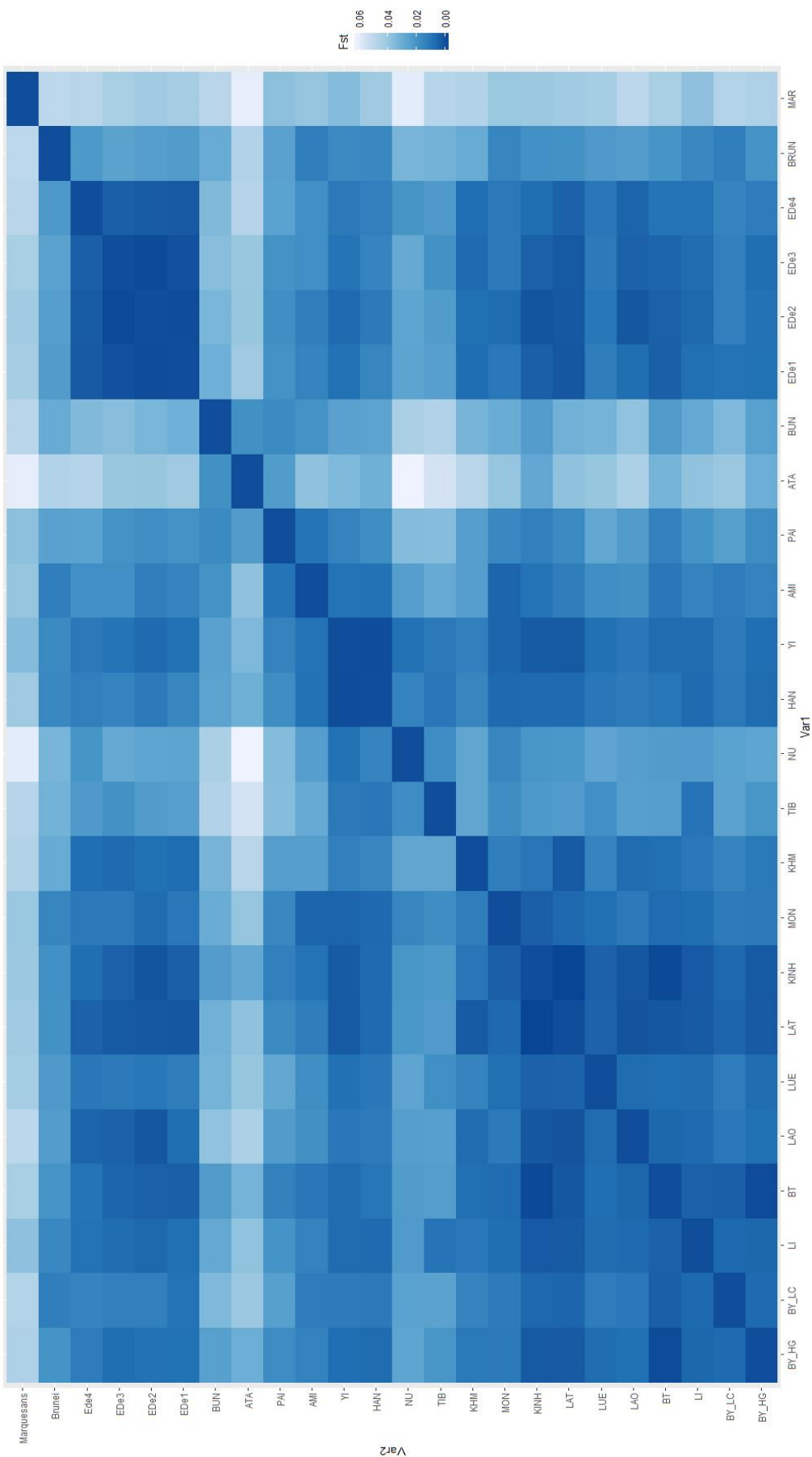
trung tâm (EDE1), các huyện giáp Gia Lai (EDE2), các huyện giáp Đăk Nông (EDE3) và các huyện giáp Khánh Hòa (Bảng 3.1).

Bảng 3. 1. Bốn nhóm người Ê Đê được chia theo địa bàn cư trú

KH	Vị trí	Huyện thành phần
EDE1	Trung tâm	Cư M'gar, Krông Buk, thị xã Buôn Hồ, Krông Pắc, Cư Kuin
EDE2	Giáp Gia Lai	Ea H'leo, Ea Kar, Ea Súp, Krông Năng
EDE3	Giáp Đăk Nông	Thành phố Buôn Ma Thuột, Krông Ana, Buôn Đôn
EDE4	Giáp Khánh Hòa	Krông Bông, M'Drăk

3.2.1. Phân tích chỉ số khoảng cách di truyền Fst

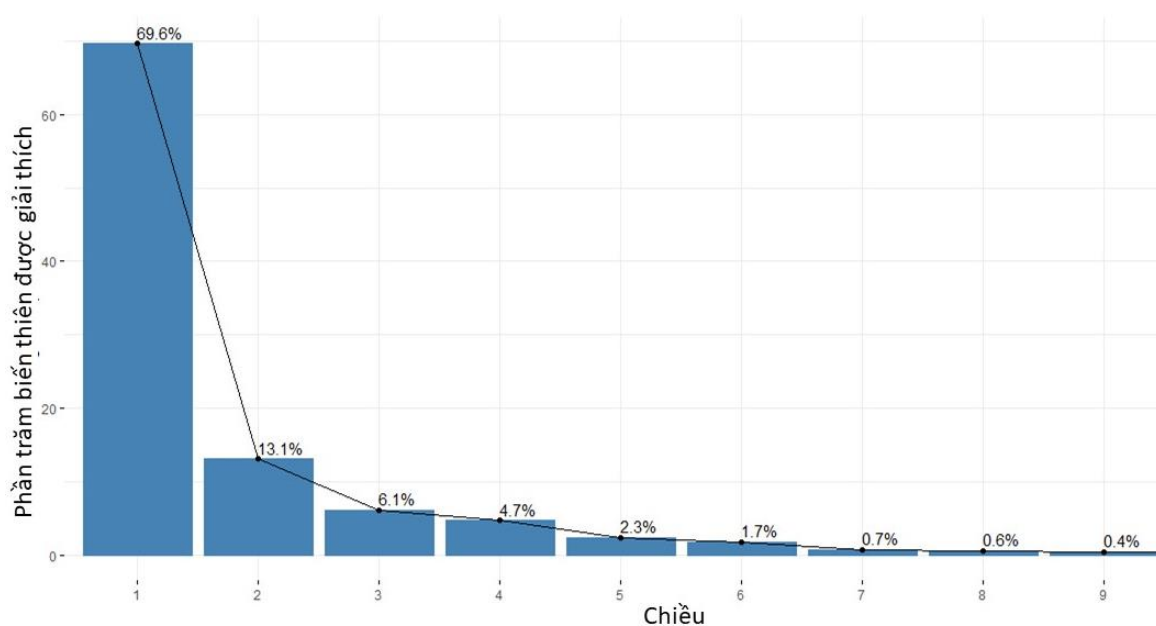
Tương tự như phân tích dựa trên hệ dòng, 20 quần thể tham chiếu được sử dụng trong phân tích. Trong tổng 276 giá trị Fst thu được, có 258 giá trị (93.47%) có ý nghĩa thống kê và 12 giá trị (6.43%) không có ý nghĩa thống kê ($p > 0.05$) (Phụ lục 2). Trong đó, phần lớn các quần thể đều có khác biệt di truyền không lớn ($F_{st} < 0.001$). Hình 3.7 là bản đồ heatmap giá trị Fst đôi một của các quần thể. Trong đó, giá trị Fst càng thấp thì màu xanh dương biểu hiện càng đậm, tức hai quần thể càng gần gũi về mặt di truyền. Ngược lại, giá trị Fst càng cao thì màu xanh dương biểu hiện càng nhạt, tức hai quần thể xa nhau về mặt di truyền. Nhìn chung, các quần thể có vị trí địa lý gần nhau sẽ gần gũi hơn về mặt di truyền, ví dụ như nhóm các quần thể thu mẫu ở Đông Nam Á cho màu xanh dương đậm hơn. Ngược lại, các nhóm quần thể Marquesan hay quần thể Atayal với màu xanh nhạt phản ánh sự khác biệt về mặt di truyền với các quần thể khác. Bốn nhóm quần thể Ê Đê cho thấy sự gần gũi di truyền với nhau và với các nhóm quần thể cùng khu vực Đông Nam Á khác. Một điều khá thú vị là quần thể EDE4 phân tách với 3 quần thể còn lại với khoảng cách di truyền lần lượt là 0.00275, 0.00348, 0.0044. Ba quần thể còn lại đều cho giá trị Fst theo từng cặp không có ý nghĩa thống kê (giá trị $p > 0.05$) (Phụ lục 2)



Hình 3. 7. Biểu đồ heatmap giá trị Fst đôi một của 4 nhóm quần thể Ê Đê với các quần thể tham chiếu

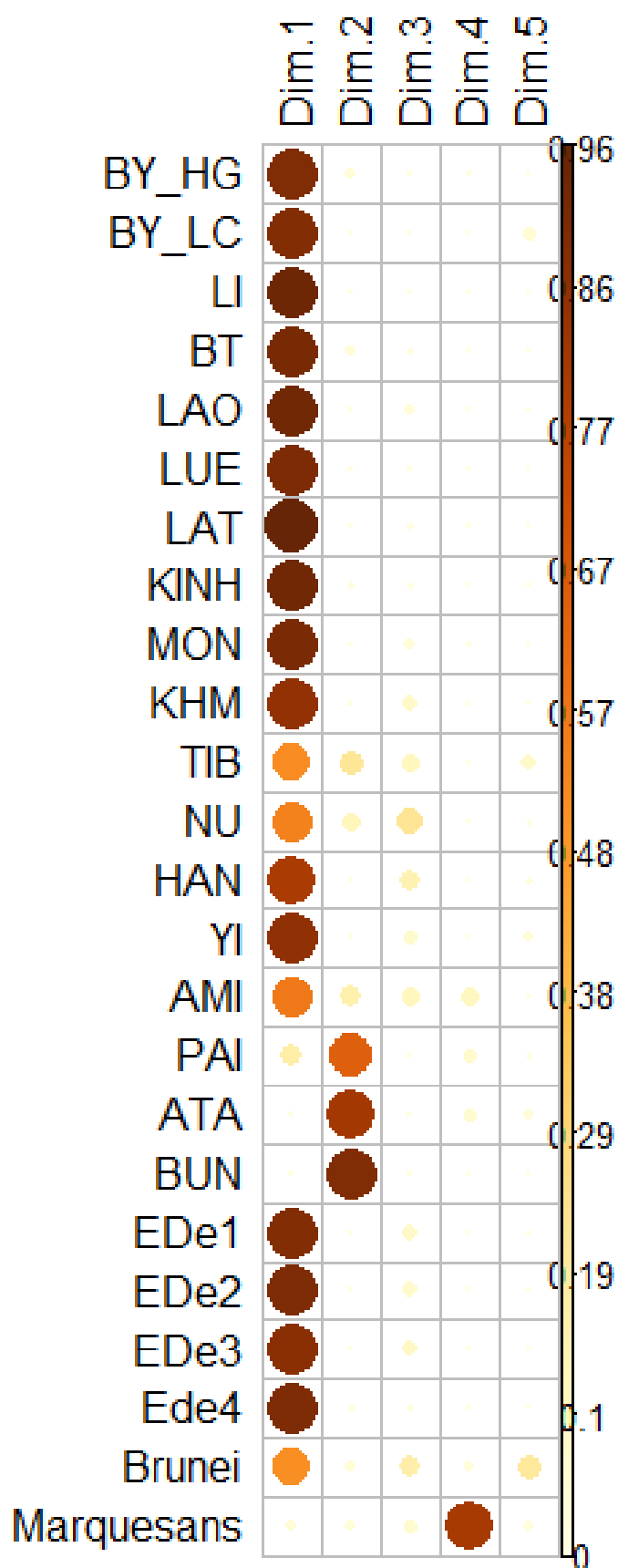
3.2.2. Kết quả phân tích thành phần chính

Bảng giá trị khoảng cách di truyền của 24 quần thể được sử dụng cho phân tích thành phần chính. Yếu tố đầu tiên cần được đánh giá là phần trăm biến thiên của các thành phần chính. Giá trị này càng cao tức khả năng biểu hiện của các thành phần chính càng tốt. Thành phần chính đầu tiên sẽ giải thích được nhiều phần trăm biến thiên nhất, sau đó giảm đều ở các thành phần chính tiếp theo. Thông thường, hai thành phần chính được lựa chọn để biểu hiện các mẫu nên có tổng phần trăm biến thiên biểu hiện cao hơn 70%.



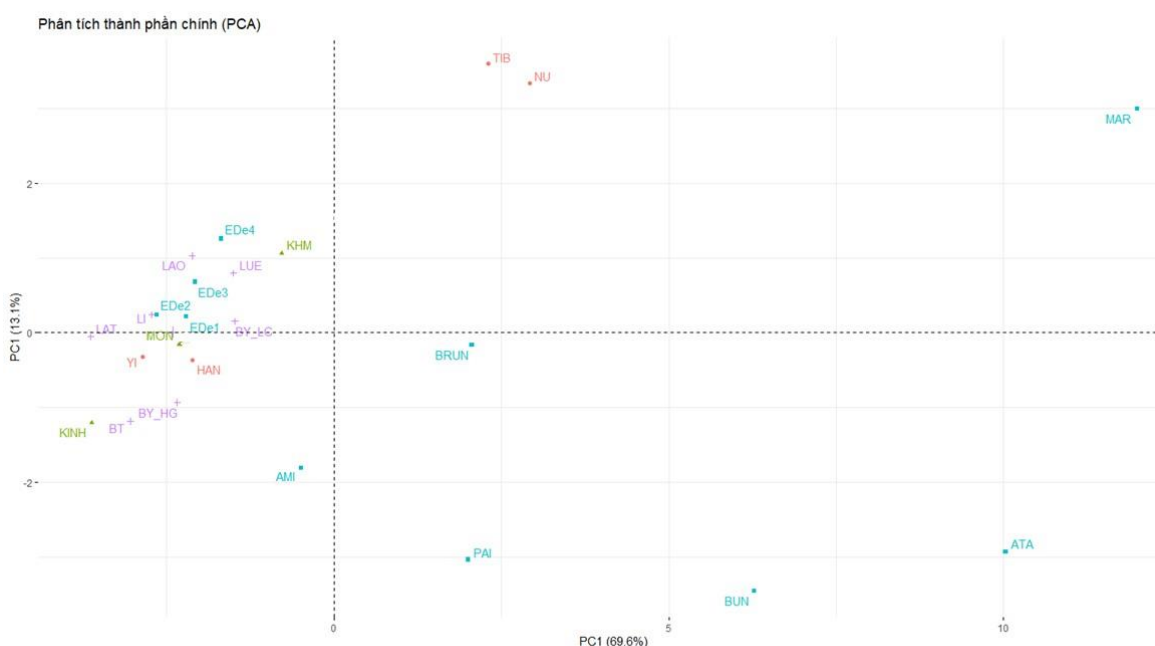
Hình 3. 8. Biểu đồ scree trực quan tỉ lệ các thành phần chính

Hình 3.8 là biểu đồ trực quan tỉ lệ biến thiên được giải thích của các thành phần chính. Trong đó, các thành phần chính đầu tiên có khả năng biểu hiện tương đối tốt, với các giá trị biểu hiện của các chiều lần lượt là chiều 1 (69.6%), chiều 2 (13.1%), chiều 3 (6.1%), chiều 4 (4.7%) và chiều 5 (2.3%). Hai chiều đầu tiên được lựa chọn để biểu hiện mẫu do chúng có tổng phần trăm biến thiên được biểu hiện cao nhất (82.7%).



Hình 3. 9. Chất lượng biểu diễn của các biến trên sơ đồ nhân tố (\cos^2)

Một yếu tố khác cần được đánh giá là chất lượng biểu diễn của các biến trên sơ đồ nhân tố. Yếu tố này được phản ánh qua giá trị \cos^2 với khả năng đại diện càng tốt thì tương ứng với vòng tròn màu nâu càng đậm và ngược lại. Hình 3.8 cho thấy hầu hết các mẫu đều có vòng tròn màu nâu đậm với hai thành phần chính đầu tiên. Điều này phản ánh khả năng đại diện tốt của hai thành phần chính đầu cho các mẫu.



Hình 3. 10. Phân tích PCA của 4 nhóm EDE1, EDE2, EDE3 và EDE4 với hai thành phần chính đầu

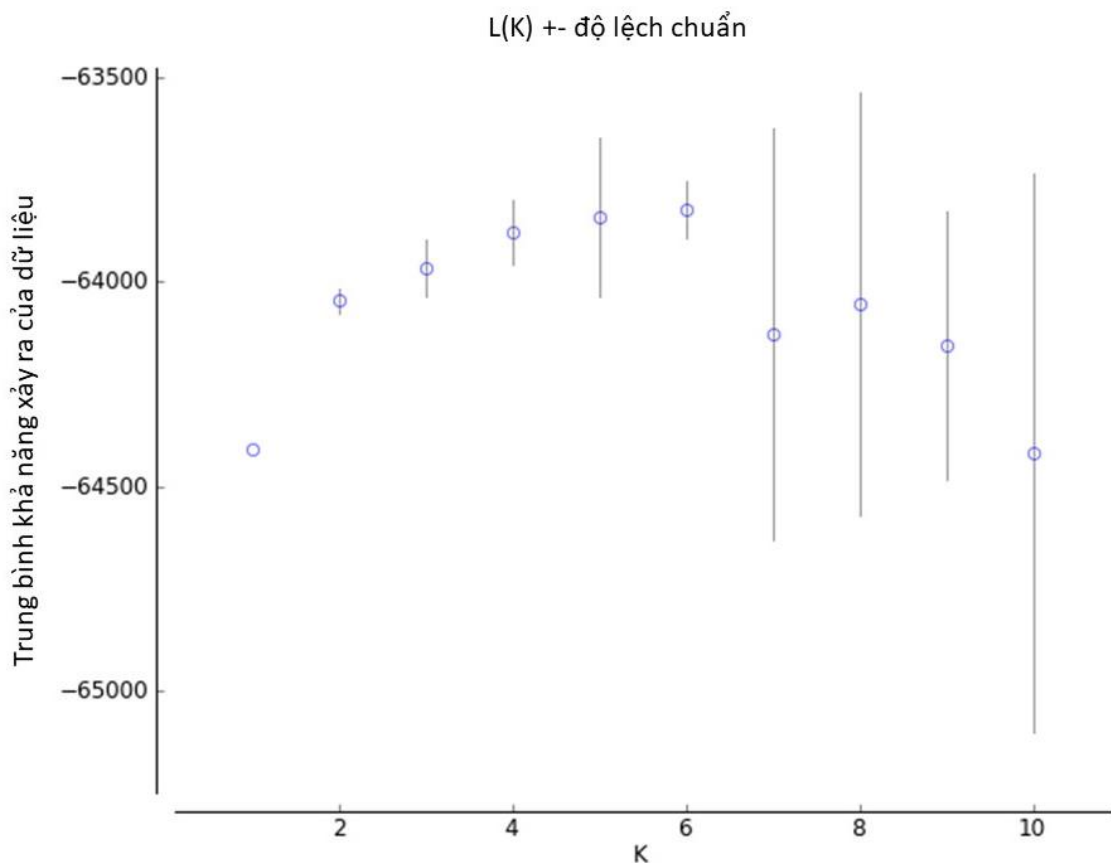
Màu đỏ: nhóm ngữ hệ Hán – Tạng. Màu xanh nước biển: nhóm ngữ hệ Nam Đảo. Màu xanh lá: nhóm ngữ hệ Nam Á. Màu tím: nhóm ngữ hệ Thái - Kadai

Nhìn chung, phân tích PCA không cho thấy một phân cụm cụ thể nào. Tương tự như kết quả phân tích giá trị F_{st} đôi một, các nhóm quần thể thu mẫu ở khu vực Đông Nam Á và Trung Quốc nằm tụ lại về một khu (trừ TIB và NU), cho thấy sự gần gũi về mặt di truyền giữa các nhóm này. Ngược lại, các quần thể thuộc ngữ hệ Nam Đảo phân bố nằm rải rác khắp bản đồ ở các góc phần tư khác nhau. Đặc biệt quần thể MAR nằm cách biệt hẳn với các nhóm ở rìa bản đồ cho thấy sự khác biệt về di truyền với các nhóm quần thể khác. Tập trung vào bốn nhóm quần thể EDE1, EDE2, EDE3 và EDE4, ta có thể thấy cả bốn

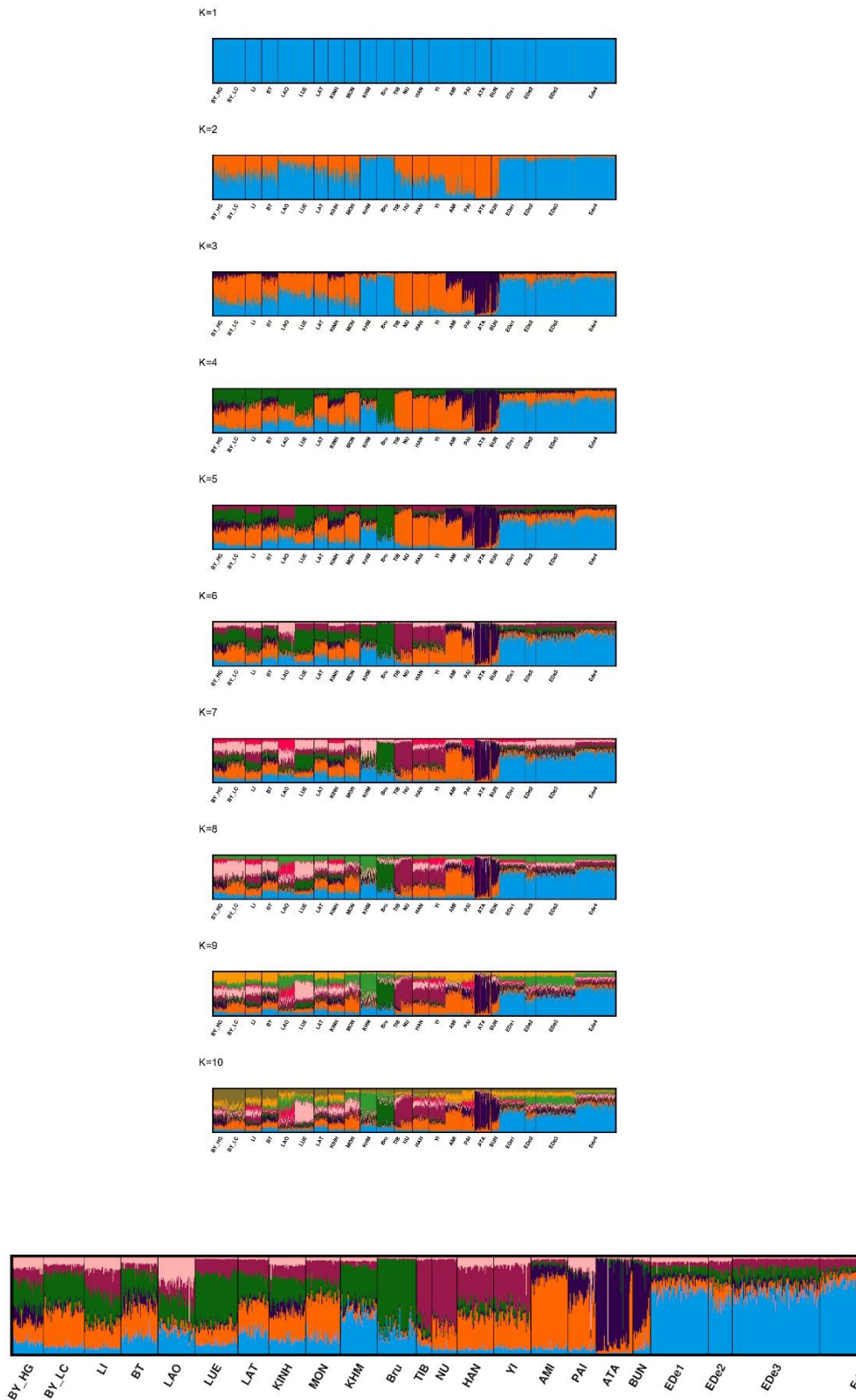
nhóm nằm gần nhau và nằm trên một góc phần tư trên cùng bên trái. Đây cũng là vị trí tập trung của các quần thể thuộc khu vực Đông Nam Á và Trung Quốc.

3.2.3. Phân nhóm quần thể bằng STRUCTURE

Để ước tính cấu trúc quần thể, phương pháp phân cụm Bayesian được triển khai với phần mềm STRUCTURE. Do giá trị F_{st} trung bình thấp (0,0188), mô hình LOCPRIOR được sử dụng để bổ sung thông tin về vị trí lấy mẫu. Giá trị $L(K)$ đạt giá trị lớn nhất tại $K = 6$, phản ánh 6 quần thể con là giá trị quần thể con tối ưu (Hình 3.10).



Hình 3. 11. Giá trị trung bình $L(K)$ với 10 lần lặp lại của K



Hình 3.12. Kết quả biểu hiện cho A.K từ 1 đến 10 và B. K=5

Mỗi cá thể được đại diện bởi một cột duy nhất và chia thành các phân đoạn có kích thước và màu sắc tương ứng với tỷ lệ của một cụm cụ thể. Các quần thể được phân cách bằng các vạch đen

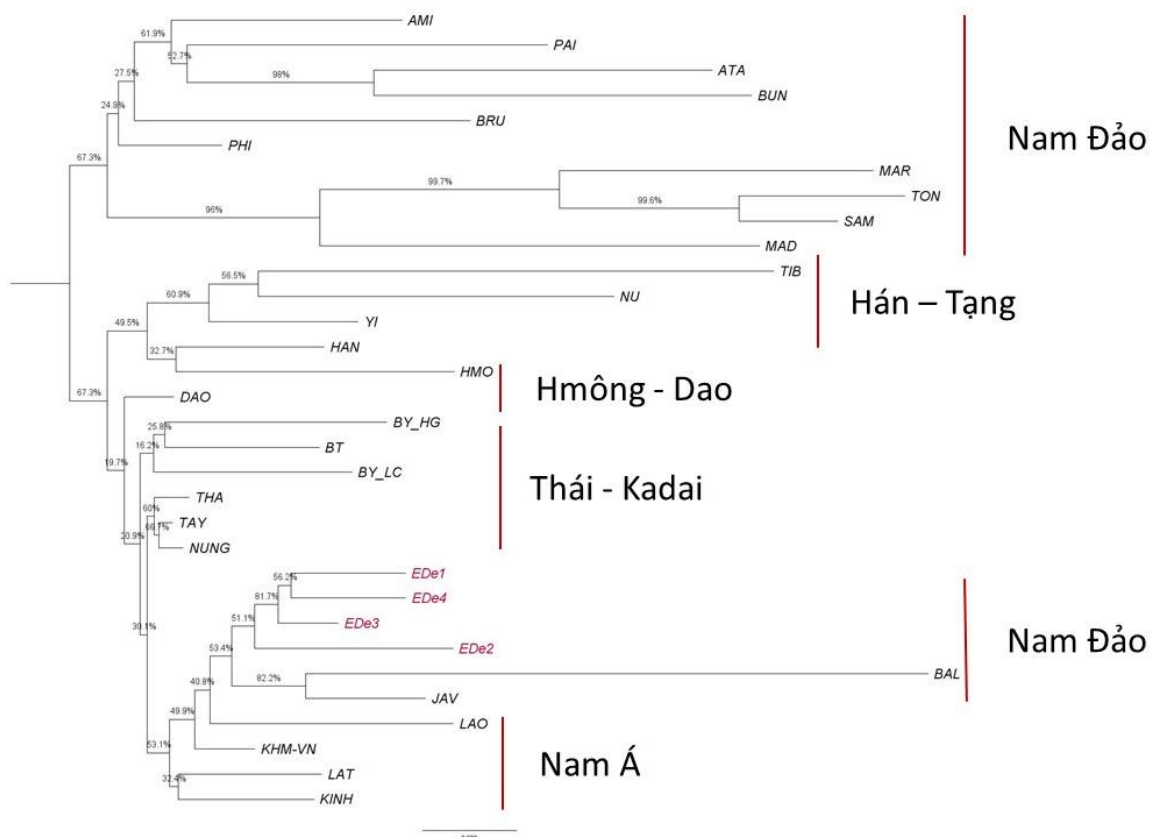
Kết quả của STRUCTURE cũng ủng hộ về tính khác biệt của các quần thể thuộc ngữ hệ Nam Đảo cũng như cho thấy sự khác biệt nhỏ về di truyền của bốn nhóm Ê Đê (Hình 3.11). Ở $K=6$, cụm màu xanh nước biển xuất hiện ở hầu hết các quần thể, trong đó chiếm tỷ trọng cao ở nhóm Ê Đê (~70%), tiếp đến là nhóm quần thể Đông Nam Á (10-30%) và thấp nhất ở nhóm quần thể Trung Quốc và Nam Đảo. Nhóm quần thể Nam Đảo cho thấy sự phân tách rõ rệt về mặt di truyền với cụm màu vàng cam (biểu hiện cao ở quần thể Ami và Paiwan) và cụm màu tím đậm (biểu hiện cao ở quần thể Atayal và Bunun). Cụm màu hồng đậm chiếm tỉ lệ cao ở các quần thể Trung Quốc và phân bố rải rác ở các nhóm quần thể khác.

Với bốn nhóm người Ê Đê, tuy rằng các quần thể này có cấu trúc di truyền tương tự nhau với cụm màu xanh dương chiếm ưu thế nhưng ta vẫn có thể quan sát được sự khác biệt về tỉ lệ giữa các nhóm. Ở quần thể EDE1 và EDE2 xuất hiện tỉ lệ nhỏ của cụm màu hồng nhạt trong khi hai quần thể còn lại gần như không có. Một điều khá thú vị ở quần thể EDE2 là nhóm có tỉ lệ cụm màu xanh nước biển thấp nhất và tỉ lệ cụm màu cam cao nhất so với 3 nhóm còn lại. Cụm này được quan sát nhiều ở nhóm quần thể Ami và Paiwan. Đây là kết quả đáng chú ý, phân tích mtDNA và SNPs cho thấy nhóm Ami còn lưu giữ nguồn gen từ nhóm săn bắt hái lượm bản địa tại khu vực Đông Nam Á-Thái Bình Dương. Do đó, có thể nhóm EDE2 vẫn còn giữ nguồn gen này trong hệ gen của mình. Tuy nhiên, cần có thêm nghiên cứu mở rộng trên hệ gen để có thể khẳng định thêm về quan sát này. Ngoài ra, ở quần thể này gần như không xuất hiện cụm màu tím, đặc trưng bởi nhóm Atayal và Bunun. Tương tự, ở quần thể EDE4 cũng cho thấy tỷ lệ rất nhỏ của cụm màu tím. Mặt khác hai nhóm EDE1 và EDE3 cho thấy tỷ lệ hai cụm màu vàng cam và tím đậm tương đương nhau.

3.2.4. Phân tích kết quả xây dựng cây phân loại

Trong tiến trình lịch sử cũng như hướng di cư của quần thể gốc ở Đài Loan, sự xuất hiện của các quần thể như Philippine và Indonesia đóng vai trò quan trọng trong sự hình thành và phát triển quần thể người Ê Đê. Ngoài ra, các nhóm quần thể trong nước cũng có thể ảnh hưởng đến sự phát triển của quần thể này. Tuy nhiên, do không có dữ liệu kiểu gen locus STR của nhiều

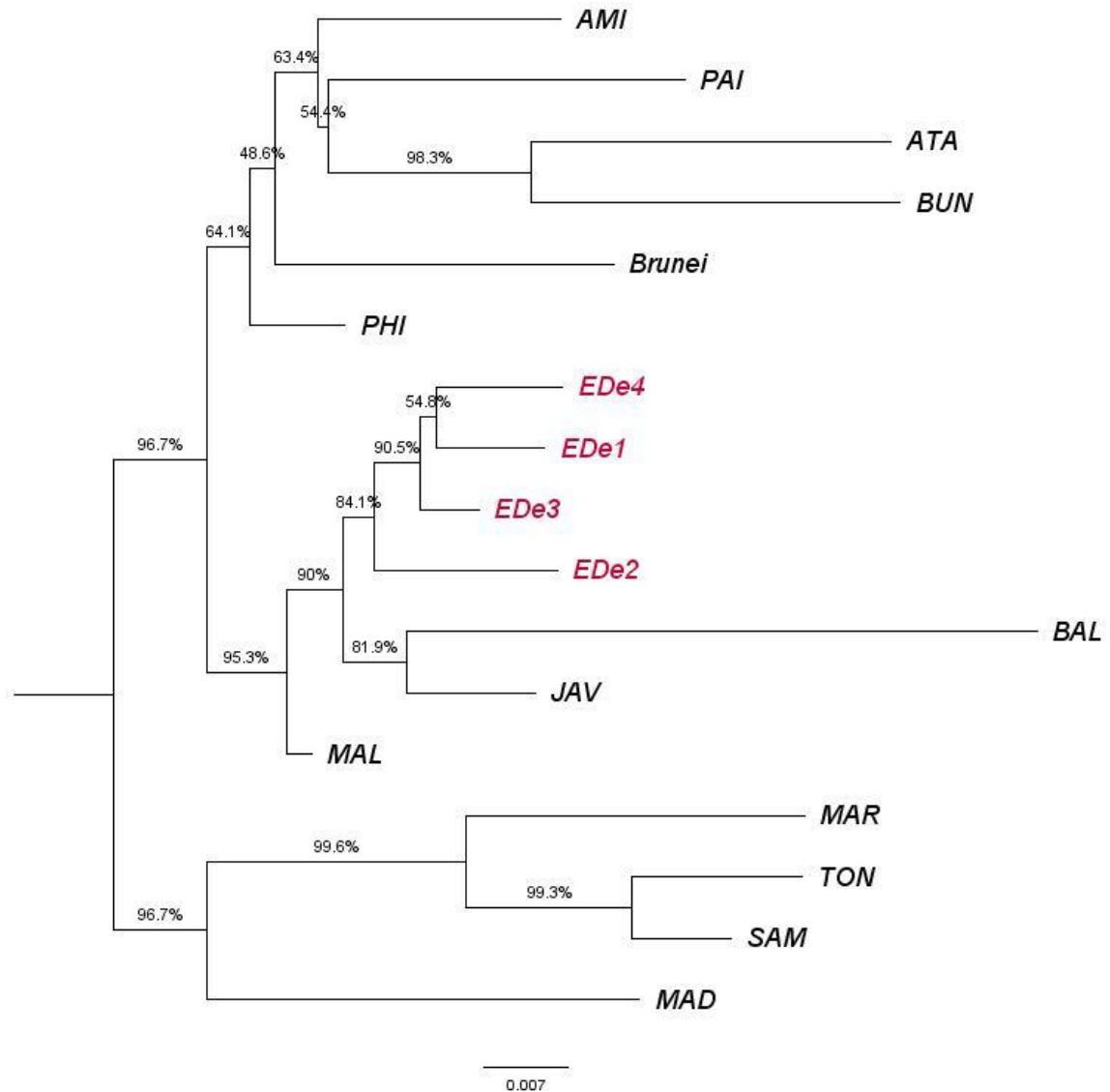
quần thể Nam Đảo như Philippine, Indonesia hay các quần thể ở Việt Nam như Khmer, H'mong, Thái, Tày nên không thể phân tích hay so sánh sâu hơn về mối quan hệ giữa họ và quần thể người Ê Đê. Phương pháp dựng cây phân loại Neighbor-joining dựa trên chỉ số khoảng cách di truyền Nei [42] được lựa chọn để đánh giá mối quan hệ của quần thể nghiên cứu với nhiều quần thể khác do dữ liệu đầu vào chỉ yêu cầu bảng tần số alen.



Hình 3. 13. Cây phân loại của quần thể nghiên cứu với các quần thể thuộc 5 ngữ hệ

Với cây phân loại đầu tiên, quần thể nghiên cứu được so sánh với các nhóm quần thể thuộc 5 ngữ hệ chính ở Việt Nam là Nam Đảo, Nam Á, Hmong – Dao, Thái – Kadai và Hán Tạng. Các nhóm quần thể ở Đông Á và Đông Nam Á cùng thuộc một nhánh lớn với đủ cả 5 ngữ hệ. Trong những nghiên cứu trước, các quần thể ở hai khu vực này đã cho thấy sự gần gũi về mặt di truyền [49]. Trong đó, ba nhóm Thái Kadai, Nam Á và Nam Đảo ở Đông Nam Á (trừ Philippine) phân thành một nhánh do có sự trao đổi nguồn gen trong quá khứ, như đã quan sát được trong nghiên cứu của McColl và cộng sự, 2018 [50]. Mặt

khác, nhóm quần thể Nam Đảo hiện đang phân tách thành hai nhóm khác nhau với nhóm quần thể ở Đài Loan và Polynesia thuộc cùng một nhánh. Điều này cho thấy mặc dù bị ảnh hưởng của vị trí địa lý, giữa các nhóm Nam Đảo vẫn cho thấy sự liên kết về di truyền.



Hình 3. 14. Cây Neighbor-Joining của quần thể nghiên cứu với các quần thể thuộc Nam Đảo

Để quan sát rõ hơn mối quan hệ di truyền giữa các quần thể thuộc ngữ hệ Nam Đảo, cây phân loại gồm 4 nhóm Ê Đê và 13 quần thể thuộc ngữ hệ này được xây dựng. Các quần thể Nam Đảo ở khu vực Polynesia gồm Madagascar, Tonga, Samoa, Marquesan nằm phân tách hẳn với các quần thể khác với chỉ số bootstrap là 96.7%. Quan sát này đã được thấy ở nhiều nghiên cứu trước đó

[37, 46]. Các nhóm quần thể Nam Đảo ở châu Á chia thành hai nhóm lớn là nhóm quần thể Đài Loan, Philippine và các nhóm quần thể khác ở Đông Nam Á, bao gồm cả bốn nhóm Ê Đê. Nhiều nghiên cứu dựa trên các dữ liệu phân tử khác nhau đã chỉ ra rằng quần thể ở Philippine có sự kế thừa về mặt di truyền của nhóm quần thể gốc Nam Đảo ở Đài Loan, ngược lại với các nhóm quần thể Nam Đảo ở Đông Nam Á thì chủ yếu sự kế thừa về ngôn ngữ [51]. Mặt khác, cả bốn nhóm Ê Đê thể hiện mối quan hệ di truyền gần gũi hơn với người Malaysia và Indonesia, phản ánh mối liên hệ địa lý và lịch sử [52] tương tự với phân tích của Pham và cộng sự (2022) [27].

KẾT LUẬN VÀ KIẾN NGHỊ

1. KẾT LUẬN

Hai nhóm Nie và Mlo đồng nhất về mặt di truyền với sự pha trộn về nguồn gen của các nhóm khác nhau. Hai nhóm quần thể này gần gũi với nhóm quần thể ở Đông Nam Á hơn là nhóm quần thể cùng ngữ hệ Nam Đảo. Điều này cho thấy hệ dòng không có ảnh hưởng đến cấu trúc di truyền của quần thể người Ê Đê.

Mặt khác, bốn nhóm EDE1, EDE2, EDE3 và EDE4 tuy tương đồng về cấu trúc di truyền nhưng vẫn quan sát được sự khác biệt nhỏ giữa các nhóm. Kết quả này cho thấy ảnh hưởng của vị trí địa lý đến cấu trúc di truyền của quần thể người. Ngoài ra, bốn nhóm quần thể người Ê Đê phân nhánh với nhóm Nam Đảo ở Malaysia và Indonesia, cho thấy mối quan hệ di truyền gần gũi với hai nhóm này.

2. KIẾN NGHỊ

Mở rộng nghiên cứu di truyền của các nhóm dân tộc có sự tương đồng về văn hóa, lịch sử hay ngôn ngữ với người Ê Đê như quần thể người Giarai, Bana, Cho-ru, Chăm, Mnông,... Việc phân tích mối tương quan di truyền giữa họ có thể bổ sung thêm các bằng chứng về lịch sử phát triển của người Ê Đê.

Mở rộng nghiên cứu di truyền với các chỉ thị có tính bảo thủ cao như mtDNA hay NST Y để phân tích về sự xuất hiện và nguồn gốc của người Ê Đê ở Việt Nam.

DANH MỤC TÀI LIỆU THAM KHẢO

1. Bộ giáo dục và Đào Tạo, Atlas Địa lý Việt Nam, p 16
2. Tổng cục thống kê, Tổng điều tra dân số và nhà ở năm 2019
3. Vũ Đình Long (1997) Gia đình và hôn nhân truyền thống ở các dân tộc Malayô - Polynexia Trường Sơn - Tây Nguyên. NXB Khoa học Xã Hội
4. Đặng Minh Tâm (2018) Tên họ người Ê Đê – Nguồn gốc, cấu tạo và cách sử dụng. *HNUE J Sci* 63:86–96. <https://doi.org/10.18173/2354-1069.2018-0031>
5. Đặng Minh Tâm (2019) Tên riêng người Ê Đê ở Tây Nguyên. Trường Đại học Sư Phạm Hà Nội
6. Tilanus MGJ (2006) Short tandem repeat markers in diagnostics: What's in a repeat? *Leukemia* 20:1353–1355. <https://doi.org/10.1038/sj.leu.2404273>
7. Butler JM (2004) Short tandem repeat analysis for human identity testing. *Curr Protoc Hum Genet* Chapter 14:1–22. <https://doi.org/10.1002/0471142905.hg1408s41>
8. Sawaya S, Bagshaw A, Buschiazzi E, et al (2013) Microsatellite Tandem Repeats Are Abundant in Human Promoters and Are Associated with Regulatory Elements. *PLoS One* 8:. <https://doi.org/10.1371/journal.pone.0054710>
9. Chen HY, Ma SL, Huang W, et al (2016) The mechanism of transactivation regulation due to polymorphic short tandem repeats (STRs) using IGF1 promoter as a model. *Sci Rep* 6:1–10. <https://doi.org/10.1038/srep38225>
10. Chakraborty R, Kimmel M, Stivers DN, et al (1997) Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. *Proc Natl Acad Sci U S A* 94:. <https://doi.org/10.1073/pnas.94.3.1041>
11. Kornberg A, Bertsch Ll, Jackson Jf, Khorana Hg (1964) enzymatic synthesis of deoxyribonucleic acid, xvi. oligonucleotides as templates and the mechanism of their replication. *Proc Natl Acad Sci U S A* 51:315–323. <https://doi.org/10.1073/pnas.51.2.315>
12. Huang QY, Xu FH, Shen H, et al (2002) Mutation patterns at dinucleotide microsatellite loci in humans. *Am J Hum Genet* 70:625–634. <https://doi.org/10.1086/338997>
13. Nadir E, Margalit H, Gallily T, Ben-Sasson SA (1996) Microsatellite spreading in the human genome: Evolutionary mechanisms and structural implications. *Proc Natl Acad Sci U S A* 93:6470–6475. <https://doi.org/10.1073/pnas.93.13.6470>
14. Butler JM (2006) Genetics and genomics of core short tandem repeat loci used in human identity testing. *J Forensic Sci* 51:253–265.

- <https://doi.org/10.1111/j.1556-4029.2006.00046.x>
15. Panneerchelvam S, Norazmi MN (2003) Forensic DNA profiling and database. *Malaysian J Med Sci* 10:20–26
 16. NIST 13 CODIS Core STR Loci. <https://www.nist.gov/>. Accessed 12 Jun 2023
 17. Thermo Fisher Scientific, VeriFiler Express PCR Amplification Kit User Guide. 138
 18. Slatkin M (1987) Quantitative genetics of heterochrony. *Evolution* (N Y) 41:799–811. <https://doi.org/10.1111/j.1558-5646.1987.tb05854.x>
 19. Kidd KK, Pakstis AJ, Speed WC, et al (2006) Developing a SNP panel for forensic identification of individuals. *Forensic Sci Int* 164:20–32. <https://doi.org/10.1016/j.forsciint.2005.11.017>
 20. Kidd KK, Speed WC, Pakstis AJ, et al (2014) Progress toward an efficient panel of SNPs for ancestry inference. *Forensic Sci Int Genet* 10:23–32. <https://doi.org/https://doi.org/10.1016/j.fsigen.2014.01.002>
 21. Wright S (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* (N Y) 19:395–420. <https://doi.org/10.1111/j.1558-5646.1965.tb01731.x>
 22. Wright S (1951) The genetical structure of populations. *Ann Eugen* 15:323–354. <https://doi.org/10.2307/2407273>
 23. Paschou P, Ziv E, Burchard EG, et al (2007) PCA-correlated SNPs for structure identification in worldwide human populations. *PLoS Genet* 3:1672–1686. <https://doi.org/10.1371/journal.pgen.0030160>
 24. Elhaik E (2022) Principal Component Analyses (PCA)-based findings in population genetic studies are highly biased and must be reevaluated. Nature Publishing Group UK
 25. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol Ecol* 14:2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
 26. Macholdt E, Arias L, Duong NT, et al (2020) The paternal and maternal genetic history of Vietnamese populations. *Eur J Hum Genet* 28:636–645. <https://doi.org/10.1038/s41431-019-0557-4>
 27. Dung Pham P, Luc Hoang T, Tra Le K, et al (2022) The first data of allele frequencies for 23 autosomal STRs in the Ede ethnic group in Vietnam. *Leg Med* 57:102072. <https://doi.org/10.1016/j.legalmed.2022.102072>
 28. Tran LH, Chu PTM, Nguyen TH, et al (2022) Genetic structure and population connection of two Bouyei populations in northern Vietnam based on short tandem repeat analysis. *Am J Hum Biol* 34:1–10. <https://doi.org/10.1002/ajhb.23702>
 29. Kutan W, Kampuansai J, Colonna V, et al (2011) Genetic affinity and

- admixture of northern Thai people along their migration route in northern Thailand: Evidence from autosomal STR loci. *J Hum Genet* 56:130–137. <https://doi.org/10.1038/jhg.2010.135>
30. Srikumool M, Srithawong S, Muisuk K, et al (2022) Forensic and genetic characterizations of diverse southern Thai populations based on 15 autosomal STRs. *Sci Rep* 12:1–11. <https://doi.org/10.1038/s41598-021-04646-1>
 31. Srithawong S, Muisuk K, Srikumool M, et al (2020) Genetic structure of the ethnic Lao groups from mainland Southeast Asia revealed by forensic microsatellites. *Ann Hum Genet* 84:357–369. <https://doi.org/10.1111/ahg.12379>
 32. Fan H, Du Z, Wang F, et al (2021) The forensic landscape and the population genetic analyses of Hainan Li based on massively parallel sequencing DNA profiling. *Int J Legal Med* 135:. <https://doi.org/10.1007/s00414-021-02590-3>
 33. Tran HL, Nguyen HT, Pham TT, et al (2019) Allele frequencies for 22 autosomal STRs in the Kinh population in Vietnam. *Int J Legal Med* 133:1761–1762. <https://doi.org/10.1007/s00414-018-01996-w>
 34. Kraaijenbrink T, Zuniga S, Su B, et al (2008) Allele frequency distribution of 21 forensic autosomal STRs in 7 populations from Yunnan, China. *Forensic Sci Int Genet* 3:. <https://doi.org/10.1016/j.fsigen.2008.02.007>
 35. Mirabal S, Cadenas AM, Garcia-Bertrand R, Herrera RJ (2013) Ascertaining the role of Taiwan as a source for the Austronesian expansion. *Am J Phys Anthropol* 150:551–564. <https://doi.org/10.1002/ajpa.22226>
 36. Hakim HM, Khan HO, Ismail SA, et al (2020) Forensic parameters and ancestral fractions in the Kedayan population inferred using 21 autosomal STR loci. *Meta Gene* 25:. <https://doi.org/10.1016/j.mgene.2020.100741>
 37. Palencia-Madrid L, Baeta M, Villaescusa P, et al (2019) The Marquesans at the fringes of the Austronesian expansion. *Eur J Hum Genet* 27:801–810. <https://doi.org/10.1038/s41431-019-0336-2>
 38. Hadley Wickham (2009) *ggplot2 - Elegant Graphics for Data Analysis*. Springer New York, NY
 39. Lê S, Josse J, Husson F (2008) FactoMineR: An R package for multivariate analysis. *J Stat Softw* 25:1–18. <https://doi.org/10.18637/jss.v025.i01>
 40. Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information - HUBISZ - 2009 - Molecular Ecology Resources - Wiley Online Library. *Mol Ecol Resour* 9:1322–1332

41. Kopelman NM, Mayzel J, Jakobsson M, et al (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol Ecol Resour* 15:1179–1191. <https://doi.org/10.1111/1755-0998.12387>
42. Nei M, Chesser Rk (1983) Estimation of fixation indices and gene diversities. *Ann Hum Genet* 47:253–259. <https://doi.org/10.1111/j.1469-1809.1983.tb00993.x>
43. Shepard EM, Chow RA, Suafo'a E, et al (2005) Autosomal STR variation in five Austronesian populations. *Hum Biol* 77:825–851. <https://doi.org/10.1353/hub.2006.0019>
44. De Ungria MCA, Roby RK, Tabbada KA, et al (2005) Allele frequencies of 19 STR loci in a Philippine population generated using AmpFISTR multiplex and ALF singleplex systems. *Forensic Sci Int* 152:281–284. <https://doi.org/10.1016/j.forsciint.2004.09.125>
45. Seah LH, Jeevan NH, Othman MI, et al (2003) STR data for the AmpFISTR identifier loci in three ethnic groups (Malay, Chinese, Indian) of the Malaysian population. *Forensic Sci Int* 138:134–137. <https://doi.org/10.1016/j.forsciint.2003.09.005>
46. Regueiro M, Mirabal S, Lacau H, et al (2008) Austronesian genetic signature in East African Madagascar and Polynesia. *J Hum Genet* 53:106–120. <https://doi.org/10.1007/s10038-007-0224-4>
47. Trịnh TT (2019) Nghiên cứu đặc trưng tần suất alen hệ nhận dạng STR (AmpFISTR® identifier® plus kit) của một số tộc người Việt Nam để ứng dụng trong giám định ADN. Đại học Khoa học Tự Nhiên - Đại học Quốc Gia Hà Nội
48. Porras-Hurtado L, Ruiz Y, Santos C, et al (2013) An overview of STRUCTURE: Applications, parameter settings, and supporting software. *Front Genet* 4:1–13. <https://doi.org/10.3389/fgene.2013.00098>
49. The HUGO Pan-Asian SNP Consortium (2013) Mapping Human Genetic Diversity in Asia. *Science* (80-) 2556
50. Mccoll H, Racimo F, Vinner L, et al (2018) The prehistoric peopling of Southeast Asia. *Hum Genomics* 92:88–92
51. Soares PA, Trejaut JA, Rito T, et al (2016) Resolving the ancestry of Austronesian-speaking populations. *Hum Genet* 135:309–326. <https://doi.org/10.1007/s00439-015-1620-z>
52. Bellwood P (1997) Prehistory of the Indo-Malaysian Archipelago. ANU Press

TIB	0.020 91	0.025 86	0.009 79	0.023 99	0.024 7	0.018 9	0.023 04	0.022 25	0.018 1	0.027 18	0	+	+	+	+	+	+	+	+	+	+	+
NU	0.026 96	0.025 64	0.022 98	0.023 47	0.024 23	0.027 13	0.021 38	0.020 93	0.015 71	0.027 3	0.017 97	0	+	+	+	+	+	+	+	+	+	+
HAN	0.007 35	0.011 25	0.007 15	0.010 31	0.011 97	0.010 77	0.007 24	0.007 19	0.006 69	0.015 51	0.010 65	0.014 87	0	-	+	+	+	+	+	+	+	+
YI	0.008 17	0.011 81	0.007 42	0.007 38	0.010 74	0.008 75	0.002 93	0.003 24	0.005 54	0.013 39	0.011 44	0.009 03	-0.00 017	0	+	+	+	+	+	+	+	+
AMI	0.014 27	0.012 31	0.014 15	0.010 71	0.019 05	0.018 32	0.012 52	0.009 3	0.005 81	0.024 31	0.029 25	0.024 36	0.008 88	0.009 31	0	+	+	+	+	+	+	+
PAI	0.017 44	0.024 87	0.020 26	0.014 02	0.023 41	0.027 94	0.017 02	0.013 17	0.016 31	0.024 15	0.036 01	0.035 69	0.017 77	0.014 22	0.009 66	0	+	+	+	+	+	+
ATA	0.030 56	0.040 6	0.038 46	0.032 57	0.045 11	0.039 73	0.038 18	0.028 04	0.039 27	0.048 49	0.055 76	0.063 16	0.031 2	0.034 22	0.038 09	0.023 57	0	+	+	+	+	+
BUN	0.024 92	0.034 2	0.028 34	0.023 24	0.038 58	0.032 36	0.031 68	0.023 64	0.029 88	0.032 87	0.046 53	0.044 72	0.026 19	0.025 6	0.019 82	0.016 87	0.019 19	0	+	+	+	+
NIE	0.009 09	0.011 67	0.007 14	0.006 23	0.004 62	0.010 01	0.003 14	0.005 53	0.009 51	0.006 15	0.020 61	0.022 62	0.013 08	0.009 27	0.016 54	0.021 21	0.043 33	0.033 43	0	-	+	+
MLO	0.009 6	0.011 61	0.008 38	0.005 43	0.005 36	0.011 88	0.001 75	0.003 56	0.010 82	0.008 42	0.023 26	0.028 37	0.015 23	0.008 79	0.016 27	0.019 06	0.039 24	0.034 05	0.000 69	0	+	+
BRU	0.020 26	0.013 04	0.016 38	0.020 81	0.023 65	0.022 87	0.019 44	0.019 22	0.015 73	0.029 48	0.032 22	0.033 21	0.016 21	0.016 9	0.012 85	0.025 56	0.046 54	0.029 51	0.022 71	0.023 21	0	+
MAR	0.045 68	0.046 94	0.037 91	0.044 34	0.049 43	0.043 25	0.041 86	0.040 79	0.040 4	0.046 59	0.048 2	0.059 92	0.041 55	0.035 75	0.039 45	0.037 6	0.060 83	0.048 72	0.044 29	0.042 81	0.049 75	0

KH M	0.01 119	0.01 502	0.01 113	0.00 827	0.00 745	0.01 511	0.00 312	0.01 044	0.01 311	0	+	+	+	+	+	+	+	+	+	+	+	+	+	+
TIB	0.02 091	0.02 586	0.00 979	0.02 399	0.02 47	0.01 89	0.02 304	0.02 225	0.01 81	0.02 718	0	+	+	+	+	+	+	+	+	+	+	+	+	+
NU	0.02 696	0.02 564	0.02 298	0.02 347	0.02 423	0.02 713	0.02 138	0.02 093	0.01 571	0.02 73	0.01 797	0	+	+	+	+	+	+	+	+	+	+	+	+
HA N	0.00 735	0.01 125	0.00 715	0.01 031	0.01 197	0.01 077	0.00 724	0.00 719	0.00 669	0.01 551	0.01 065	0.01 487	0	-	+	+	+	+	+	+	+	+	+	+
YI	0.00 817	0.01 181	0.00 742	0.00 738	0.01 074	0.00 875	0.00 293	0.00 324	0.00 554	0.01 339	0.01 144	0.00 903	- 0.00 017	0	+	+	+	+	+	+	+	+	+	+
AM I	0.01 427	0.01 231	0.01 415	0.01 071	0.01 905	0.01 832	0.01 252	0.00 93	0.00 581	0.02 431	0.02 925	0.02 436	0.00 888	0.00 931	0	+	+	+	+	+	+	+	+	+
PAI	0.01 744	0.02 487	0.02 026	0.01 402	0.02 341	0.02 794	0.01 702	0.01 317	0.01 631	0.02 415	0.03 601	0.03 569	0.01 777	0.01 422	0.00 966	0	+	+	+	+	+	+	+	+
AT A	0.03 056	0.04 06	0.03 846	0.03 257	0.04 511	0.03 973	0.03 818	0.02 804	0.03 927	0.04 849	0.05 576	0.06 316	0.03 12	0.03 422	0.03 809	0.02 357	0	+	+	+	+	+	+	+
BU N	0.02 492	0.03 42	0.02 834	0.02 324	0.03 858	0.03 236	0.03 168	0.02 364	0.02 988	0.03 287	0.04 653	0.04 472	0.02 619	0.02 56	0.01 982	0.01 687	0.01 919	0	+	+	+	+	+	+
EDe 1	0.00 896	0.00 931	0.00 828	0.00 455	0.00 803	0.01 248	0.00 232	0.00 436	0.01 099	0.00 759	0.02 464	0.02 66	0.01 575	0.00 903	0.01 483	0.01 984	0.04 149	0.03 105	0	-	-	+	+	+
EDe 2	0.00 903	0.01 324	0.00 642	0.00 463	0.00 214	0.01 049	0.00 244	0.00 173	0.00 732	0.00 849	0.02 351	0.02 703	0.01 161	0.00 7	0.01 313	0.01 847	0.03 981	0.03 341	- 0.00 043	0	-	+	+	+
EDe 3	0.00 81	0.01 345	0.00 762	0.00 579	0.00 5	0.01 152	0.00 291	0.00 484	0.01 172	0.00 721	0.01 946	0.02 898	0.01 474	0.00 977	0.01 893	0.01 988	0.04 035	0.03 648	0.00 083	- 0.00 106	0	+	+	+

Ede 4	0.01 233	0.01 481	0.00 931	0.00 968	0.00 551	0.01 077	0.00 505	0.00 804	0.01 136	0.00 824	0.02 256	0.02 054	0.01 346	0.01 121	0.01 871	0.02 564	0.04 773	0.03 46	0.00 275	0.00 348	0.00 44	0	+	+
Bru	0.02 026	0.01 304	0.01 638	0.02 081	0.02 365	0.02 287	0.01 944	0.01 922	0.01 573	0.02 948	0.03 222	0.03 321	0.01 621	0.01 69	0.01 285	0.02 556	0.04 654	0.02 951	0.02 346	0.02 427	0.02 577	0.02 222	0	+
Ma r	0.04 568	0.04 694	0.03 791	0.04 434	0.04 943	0.04 325	0.04 186	0.04 079	0.04 04	0.04 659	0.04 82	0.05 992	0.04 155	0.03 575	0.03 945	0.03 76	0.06 083	0.04 872	0.04 344	0.04 192	0.04 434	0.04 858	0.04 975	0