

KHẢO SÁT SỰ ĐA DẠNG HAPLOTYPE TRÊN 27 LOCUS Y-STR Ở QUẦN THỂ NGƯỜI KINH, SINH SỐNG TẠI TỈNH THANH HÓA

Đỗ Thị Xao Mai^{1*}, Nguyễn Thị Ngọc Ánh¹

TÓM TẮT

Mục tiêu: Khảo sát sự đa dạng haplotype trên 27 locus Y-STR ở quần thể nam giới người Kinh, sinh sống tại tỉnh Thanh Hóa, Việt Nam.

Đối tượng và phương pháp: Nghiên cứu thực nghiệm, phân tích sự đa dạng haplotype trên 27 locus Y-STR (DYS438, DYS627, DYS458, DYS437, DYS391, DYS392, DYS635 (Y GATA C4), DYS19, DYS390, DYS439, DYS456, DYS393, DYS449, DYS387S1 a/b, DYS576, DYS460, DYS533, DYS389 II, DYS570, DYS385 a/b, DYS481, YGATA H4, DYS518, DYS448) ở 200 nam giới không có quan hệ huyết thống, thuộc quần thể người Kinh, sinh sống tại tỉnh Thanh Hóa, từ tháng 01-12/2023.

Kết quả: Thu được 200 haplotype khác nhau, trong đó tất cả 200 haplotype đều là duy nhất ở từng người, không có haplotype nào giống nhau giữa 2 người. Độ đa dạng haplotype (haplotype diversity) chung trong quần thể nghiên cứu bằng 1, với khả năng phân biệt (discriminatory capacity) 100%, mức độ đa dạng haplotype bằng 1. Tần suất phân bố các alen trên 27 locus Y-STR với độ đa dạng gen dao động trong khoảng từ 0,28 (DYS438) đến 0,934 (DYS387S1).

Từ khóa: Alen, haplotype, nhiễm sắc thể Y, STR, quần thể người Kinh ở Thanh Hóa.

ABSTRACT

Objectives: Survey of Haplotype diversity on 27 Y-STR loci in the Kinh males population in Thanh Hoa province, Vietnam.

Subjects and methods: Experimental study and analysis of haplotype diversity on 27 loci. Y-STR (DYS438, DYS627, DYS458, DYS437, DYS391, DYS392, DYS635 (Y GATA C4), DYS19, DYS390, DYS439, DYS456, DYS393, DYS449, DYS387S1 a/b, DYS576, DYS460, DYS533, DYS389 II, DYS570, DYS385 a/b, DYS481, YGATA H4, DYS518, DYS448) in 200 unrelated males of the Kinh ethnic population, living in Thanh Hoa province, from January to December 2023.

Results: We obtained 200 different haplotypes, of which all 200 haplotypes are unique to each person, there is no similar haplotype between 2 people, thus the overall haplotype diversity (HD) in the study population is 1 with discriminatory capacity (DC) of 100%, haplotype diversity is 1. We have calculated the frequency distribution of alleles on 27 Y-STR loci with genetic diversity (GD) ranges from 0.28 (DYS438) to 0.934 (DYS387S1).

Keywords: Allele, Haplotype, Y chromosome, STR, ethnic population in Thanh Hoa province.

Chịu trách nhiệm nội dung: Đỗ Thị Xao Mai, Email: domai1987mifm@gmail.com

Ngày nhận bài: 16/01/2024; mời phản biện khoa học: 01/2024; chấp nhận đăng: 18/4/2024.

¹Viện Pháp y Quân đội.

1. ĐẶT VẤN ĐỀ.

Trên thế giới, công tác giám định ADN được thực hiện với 3 mảng lớn: giám định ADN trên nhiễm sắc thể thường; giám định ADN trên nhiễm sắc thể giới tính và giám định ADN ty thể. Hiện nay, công tác giám định pháp y, đặc biệt là trong khoa học hình sự chủ yếu sử dụng các marker STR (Short tandem repeat) - là các đoạn ADN có cấu trúc lặp lại từ 2-6 bp, có tính bảo thủ cao, được di truyền qua các thế hệ và mang tính đặc trưng cá thể. Tuy

nhien, trong nhiều trường hợp, việc sử dụng các marker STR trên nhiễm sắc thể thường không thể thực hiện và không đem lại kết quả. Lúc này, phân tích ADN trên nhiễm sắc thể Y trở thành công cụ hữu ích. Các đoạn lặp ngắn trên nhiễm sắc thể giới tính Y (Y-STR) có độ dài từ 2-7 bp trên vùng không tái tổ hợp (NRY), chỉ xuất hiện ở nam giới và được truyền nguyên vẹn từ cha sang con trai, trừ trường hợp đột biến [1]. Do đó, phân tích Y-STR là rất quan trọng trong việc nhận dạng người chưa rõ tung tích; trường hợp bị xâm hại tình dục, phân tích

ADN trên nhiễm sắc thể Y có thể tách riêng được kiểu gen của nam giới trong mẫu dịch âm đạo và trường hợp mẫu lẫn; xác định quan hệ huyết thống theo dòng cha trong trường hợp người cha đã chết; nghiên cứu sự di cư của con người, nghiên cứu tiến hóa phả hệ [2].

Tại Khoa Xét nghiệm sinh học, Viện Pháp y Quân đội đã và đang sử dụng bộ kit nhân gen Yfiler™ Plus PCR Amplification Kit (hãng Thermo Fisher Scientific) 27 locus. Bộ khuếch đại Yfiler™ Plus là sự kết hợp của bộ AmpFISTR® YFiler® PCR Amplification Kit 17 locus Y-STR với 10 locus Y-STR mới, sử dụng công nghệ 6 màu huỳnh quang, cho độ nhạy, độ chính xác cao nhất hiện nay. Trên thế giới, bộ kit này đang được sử dụng rộng rãi trong giám định pháp y tại các quốc gia phát triển và được FBI đưa vào tiêu chuẩn trong giám định gen hình sự.

Tần số của haplotype trong quần thể rất quan trọng, đặc biệt với tính toán xác suất trong giám định ADN hình sự. Việc tăng cường sử dụng các marker Y-STR của người ứng dụng trong pháp y, trong nghiên cứu nhân chủng học và khảo cổ học đã tạo ra sự nỗ lực hợp tác để thu thập dữ liệu haplotype từ các quần thể khác nhau và để xây dựng các cơ sở dữ liệu Y [3]. Tại Việt Nam hiện nay, các nghiên cứu về tần suất haplotype Y-STR trên các quần thể người tại Việt Nam còn rất hạn chế. Các đề tài trước đó chỉ tập trung nghiên cứu trên quần thể người Kinh ở 3 thành phố lớn (Hà Nội, thành phố Hồ Chí Minh, Đà Nẵng). Chưa có công trình nghiên cứu nào thực hiện trên hệ Yfiler Plus 27 locus gen (chủ yếu thực hiện trên hệ 17 locus). Thanh Hóa là tỉnh có dân số đông thứ 3 trên cả nước, gồm nhiều dân tộc sinh sống (chủ yếu là các dân tộc: Kinh, Mường, Thái, Thổ, Dao, Mông, Khơ Mú). Trong đó, người Kinh chiếm 81,7% dân số toàn tỉnh và có địa bàn phân bố rộng khắp; các dân tộc khác có dân số và địa bàn sống thu hẹp hơn.

Hiện nay, chưa có công trình nghiên cứu nào về tần suất haplotype Y-STR trên người Kinh ở Thanh Hóa, Việt Nam. Chúng tôi tiến hành khảo sát sự đa dạng haplotype trên 27 locus Y-STR (DYS438, DYS627, DYS458, DYS437, DYS391, DYS392, DYS635 (Y GATA C4), DYS19, DYS390, DYS439, DYS456, DYS393, DYS449, DYS387S1 a/b, DYS576, DYS460, DYS533, DYS389II, DYS570, DYS385 a/b, DYS481, YGATA H4, DYS518, DYS448) ở 200 nam giới thuộc quần thể người Kinh sinh sống ở tỉnh Thanh Hóa, nhằm khảo sát, xây dựng bảng phân bố tần suất haplotype, tần suất alen của 27 locus Y-STR; đánh giá các chỉ số đa dạng di truyền quan trọng khi sử dụng 27 locus này ở người Kinh tỉnh tại Thanh Hóa, Việt Nam.

2. ĐỐI TƯỢNG, PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Đối tượng nghiên cứu

200 mẫu sinh phẩm của 200 nam giới (người Kinh ở Thanh Hóa, Việt Nam), khỏe mạnh, không có quan hệ huyết thống (trong đó có 96 mẫu niêm mạc miệng, 55 mẫu tóc có chân, 36 mẫu máu lấy bằng thẻ FTA, 13 mẫu móng tay), từ tháng 01 đến tháng 12/2023. Mẫu được thu và lưu trữ tại Khoa Xét nghiệm sinh học, Viện Pháp y Quân đội.

2.2. Phương pháp nghiên cứu

- Tách chiết ADN: ADN của các mẫu niêm mạc miệng, tóc, móng tay, máu được tách chiết bằng Chelex 100, bổ sung protein K, ủ ở 56°C, sau đó đun cách thủy 8 phút, li tâm thu ADN [5].

- Định lượng ADN: các mẫu sau tách chiết được định lượng bằng Quantifiler® Trio DNA Quantification kit trên máy Realtime PCR 7500 System; sau đó, phân tích bằng phần mềm Realtime HID V1.2 (hãng Thermo Fisher Scientific).

- Khuếch đại ADN: ADN được khuếch đại sử dụng Yfiler™ Plus PCR Amplification Kit (Thermo Fisher Scientific), với tỉ lệ: 2,5 µl mỗi, 7,5 µl Mastermix, 05 ng ADN, H₂O đủ 25 µl. Chu trình nhiệt: 95°C/1 phút, 30 chu kì (94°C/4 giây; 61,5°C/1 phút); 60°C/22 phút; 4°C/∞. Mẫu được bảo quản ở 4°C.

- Điện di mao quản: mẫu được trộn với liz600 và Hidi; sau đó, được sốc nhiệt ở 95°C/3 phút, nhấc nhanh khay vào nước đá 0°C. Mẫu được điện di mao quản bằng máy 3500 Genetic (Thermo Fisher Scientific). Phân tích kết quả bằng phần mềm GeneMapper®ID-X trên máy giải trình tự ABI 3500 (sử dụng kích thước chuẩn GeneScan-600 LIZ, Applied Biosystems, Hoa Kỳ) và thang alen chuẩn được nhà sản xuất cung cấp theo kit Yfiler plus.

- Phân tích thống kê: tần suất alen và tần suất haplotype được tính toán bằng phương pháp đếm trực tiếp.

+ Chỉ số đa dạng gen - Gene diversity (GD) được tính bằng công thức: $GD = 1 - \sum P_i^2$

Trong đó: P_i là tần suất alen thứ i .

+ Chỉ số đa dạng haplotype - haplotype diversity (HD) được tính theo công thức:

$HD = n^*(1 - \sum P_i^2)/(n-1)$. Trong đó: P_i là tần suất haplotype thứ i , n là số lượng mẫu nghiên cứu.

+ Chỉ số khả năng phân biệt - Discrimination capacity (DC) được tính theo công thức:

$$DC = h/n.$$

Trong đó: h là số lượng haplotype khác nhau trong quần thể nghiên cứu, n là số lượng mẫu [4].

Sử dụng công cụ trực tuyến YHRD (www.yhrd.org) để thực hiện đánh giá khoảng cách di truyền của quần thể nghiên cứu với các quần thể khác đã được công bố trên cơ sở dữ liệu.

- Đạo đức: nghiên cứu được Hội đồng khoa học Viện Pháp y Quân đội thông qua. Thông tin về mẫu và các loại mẫu sử dụng trong nghiên cứu được sự cho phép của hội đồng khoa học Viện. Mẫu chỉ dùng cho mục đích nghiên cứu, không dùng cho

bất kì mục đích nào khác. Kết quả nghiên cứu và thông tin mẫu hoàn toàn bảo mật.

3. KẾT QUẢ NGHIÊN CỨU VÀ BÀN LUẬN

Tách chiết, Realtime PCR, khuếch đại, phân tích kết quả hoàn chỉnh trên 200 mẫu nghiên cứu, thu được 200 kiểu gen Y-STR của mẫu nghiên cứu. Trong đó, không có trường hợp nào bị mất alen (alen drop out) và không có trường hợp nào 2 alen trên 1 locus.

3.1. Kết quả phân tích thống kê

Bảng 1. Tần suất phân bố chung của 23 locus Y-STR trong 200 mẫu nam giới người Kinh ở Thanh Hóa, Việt Nam (ô màu vàng là các alen hiếm phát hiện trong quần thể nghiên cứu)

| ALEN | LOCUS | | | | | | | |
|------|--------|---------|---------|--------|--------|--------|--------|--------|
| | DYS576 | DYS389I | YGATAH4 | DYS448 | DYS391 | DYS460 | DYS458 | DYS19 |
| 6 | | | | | 0,015 | | | |
| 9 | | | | | 0,02 | 0,135 | | |
| 10 | | | 0,2 | | 0,535 | 0,63 | | |
| 11 | | 0,115 | 0,455 | | 0,41 | 0,205 | | |
| 12 | | 0,245 | 0,275 | | 0,02 | 0,03 | | |
| 13 | | 0,435 | 0,065 | | | | 0,005 | 0,03 |
| 14 | | 0,175 | 0,005 | | | | | 0,115 |
| 15 | | 0,03 | | | | | 0,075 | 0,305 |
| 16 | 0,065 | | | | | | 0,165 | 0,065 |
| 17 | 0,225 | | | 0,02 | | | 0,21 | 0,065 |
| 18 | 0,34 | | | 0,57 | | | 0,375 | |
| 19 | 0,255 | | | 0,17 | | | 0,1 | |
| 20 | 0,105 | | | 0,165 | | | 0,05 | |
| 21 | 0,01 | | | 0,075 | | | 0,01 | |
| 22 | | | | | | | 0,01 | |
| ALEN | LOCUS | | | | | | | |
| | DYS456 | DYS439 | DYS438 | DYS392 | DYS393 | DYS570 | DYS437 | DYS533 |
| 9 | | 0,005 | 0,01 | | | | | 0,005 |
| 10 | | 0,025 | 0,84 | 0,005 | | | | 0,525 |
| 11 | | 0,25 | 0,13 | 0,055 | 0,02 | | | 0,525 |
| 12 | | 0,56 | 0,02 | 0,035 | 0,265 | | | 0,13 |
| 13 | 0,02 | 0,145 | | 0,73 | 0,135 | | | 0,02 |
| 14 | 0,22 | 0,015 | | 0,165 | 0,54 | 0,005 | 0,745 | |
| 15 | 0,615 | | | 0,005 | 0,04 | 0,035 | 0,25 | |
| 16 | 0,11 | | | 0,005 | | | 0,005 | |
| 17 | 0,03 | | | | | 0,145 | | |
| 18 | 0,005 | | | | | 0,105 | | |
| 19 | | | | | | 0,085 | | |
| 20 | | | | | | 0,045 | | |
| 21 | | | | | | 0,02 | | |

| ALEN | LOCUS | | | | | | | |
|------|--------|----------|--------|--------|--------|--------|--------|--|
| | DYS635 | DYS389II | DYS627 | DYS390 | DYS518 | DYS449 | DYS481 | |
| 17 | | | 0,015 | | | | | |
| 18 | | | 0,045 | | | | | |
| 19 | 0,035 | | 0,08 | | | | | |
| 20 | 0,11 | | | | | | | |
| 21 | 0,425 | | | | | | 0,03 | |
| 22 | 0,225 | | 0,2 | | | | 0,165 | |
| 23 | 0,13 | | 0,045 | 0,16 | | | 0,475 | |
| 24 | 0,065 | | 0,01 | 0,455 | | | 0,15 | |
| 25 | 0,01 | | 0,005 | 0,375 | | 0,035 | 0,085 | |
| 26 | | | 0,005 | 0,01 | | 0,28 | 0,065 | |
| 27 | | 0,065 | | | | 0,12 | 0,015 | |
| 28 | | 0,24 | | | | 0,045 | 0,005 | |
| 29 | | 0,34 | | | | 0,08 | 0,005 | |
| 31 | | 0,305 | | | | 0,095 | | |
| 32 | | 0,04 | | | | 0,085 | | |
| 33 | | 0,01 | | | | 0,045 | | |
| 34 | | | | | 0,02 | 0,05 | | |
| 35 | | | | | 0,08 | 0,02 | | |
| 36 | | | | | 0,06 | 0,02 | | |
| 37 | | | | | 0,08 | | | |
| 38 | | | | | 0,08 | | | |
| 39 | | | | | 0,11 | | | |
| 40 | | | | | 0,175 | | | |
| 41 | | | | | 0,205 | | | |
| 42 | | | | | 0,13 | | | |
| 43 | | | | | 0,05 | | | |
| 44 | | | | | 0,01 | | | |

Bảng 2. Tần suất phân bố các alen thuộc locus *DYS385a/b* và locus *DYF387S1*

| Locus <i>DYS385a/b</i> | | Locus <i>DYF387S1</i> | |
|------------------------|------------|-----------------------|---------------|
| Alen | Tần suất | Alen | Tần suất |
| 11 | 0,0325 | 34 | 0,0075 |
| 12 | 0,06 | 35 | 0,0675 |
| 13 | 0,3 | 36 | 0,1825 |
| 14 | 0,0525 | 37 | 0,2775 |
| 15 | 0,0675 | 37,3 | 0,005 |
| 16 | 0,05 | 38 | 0,195 |
| 17 | 0,05 | 39 | 0,155 |
| 18 | 0,1775 | 40 | 0,0875 |
| 19 | 0,12 | 41 | 0,0225 |
| 20 | 0,065 | | |
| 21 | 0,02 | | |
| 22 | 0,005 | | |

Bảng 3. Tần suất phân bố các tổ hợp alen thuộc locus DYS385a/b.

| Haplotype | Tần suất | Haplotype | Tần suất |
|-----------|----------|-----------|----------|
| 11,12 | 0,01 | 13,21 | 0,01 |
| 11,14 | 0,01 | 13,22 | 0,005 |
| 11,18 | 0,025 | 14,15 | 0,005 |
| 11,19 | 0,01 | 14,17 | 0,005 |
| 11,20 | 0,01 | 14,18 | 0,025 |
| 12,12 | 0,005 | 14,19 | 0,03 |
| 12,13 | 0,005 | 14,20 | 0,005 |
| 12,14 | 0,005 | 14,21 | 0,01 |
| 12,15 | 0,005 | 15,15 | 0,005 |
| 12,16 | 0,02 | 15,16 | 0,015 |
| 12,17 | 0,015 | 15,18 | 0,025 |
| 12,18 | 0,01 | 15,19 | 0,03 |
| 12,19 | 0,015 | 15,20 | 0,03 |
| 12,20 | 0,02 | 15,21 | 0,015 |
| 12,21 | 0,005 | 16,16 | 0,005 |
| 13,13 | 0,02 | 16,18 | 0,005 |
| 13,14 | 0,01 | 16,19 | 0,005 |
| 13,16 | 0,025 | 16,20 | 0,015 |
| 13,17 | 0,065 | 16,22 | 0,005 |
| 13,18 | 0,265 | 17,18 | 0,005 |
| 13,19 | 0,145 | 17,20 | 0,01 |
| 13,20 | 0,03 | 20,20 | 0,005 |

Bảng 4. Tần suất phân bố các tổ hợp alen thuộc locus DYS387S1

| Haplotype | Tần suất | Haplotype | Tần suất |
|-----------|----------|-----------|----------|
| 34,37 | 0,01 | 36,41 | 0,01 |
| 34,38 | 0,005 | 37,37 | 0,1 |
| 35,35 | 0,005 | 37,38 | 0,115 |
| 35,36 | 0,02 | 37,39 | 0,095 |
| 35,37 | 0,015 | 37,40 | 0,04 |
| 35,38 | 0,045 | 37,41 | 0,015 |
| 35,39 | 0,025 | 38,38 | 0,045 |
| 35,40 | 0,02 | 38,39 | 0,025 |
| 36,36 | 0,035 | 38,40 | 0,015 |
| 36,37 | 0,06 | 39,39 | 0,035 |
| 36,38 | 0,09 | 40,40 | 0,04 |
| 36,39 | 0,095 | 40,41 | 0 |
| 36,40 | 0,02 | 41,41 | 0,01 |

Bảng 5. Các alen có tần số thấp trong quần thể nghiên cứu

| Locus | Alen | Locus | Alen |
|----------|----------------|-----------|------------|
| DYS576 | 21 | DYS635 | 25 |
| YGATA-H4 | 14 | DYS389II | 32 |
| DYS458 | 13, 19, 21, 22 | DYS627 | 24, 25, 26 |
| DYS456 | 18 | DYS390 | 26 |
| DYS439 | 9 | DYS518 | 44 |
| DYS392 | 10, 15, 16 | DYS481 | 28, 29, 30 |
| DYS570 | 14 | DYS385a/b | 22 |
| DYS437 | 16 | DYS387S1 | 34, 37, 3 |
| DYS533 | 9 | | |

Trong quần thể nghiên cứu, tổng số alen quan sát được là 168 alen; số lượng alen trên mỗi locus nằm trong khoảng từ 3 (locus DYS437) đến 12 (locus DYS385ab, locus DYS449); số lượng alen trung bình trên mỗi locus là 6; locus đa hình nhất là DYS387S1. Một số alen thấy xuất hiện với tần suất rất cao trong quần thể nghiên cứu (alen 10 ở locus DYS438 có tần suất cao nhất là 0,84; alen 14 ở DYS437 với tần suất 0,745 và tần suất alen thấp nhất là 0,005 cho 27 alen khác nhau).

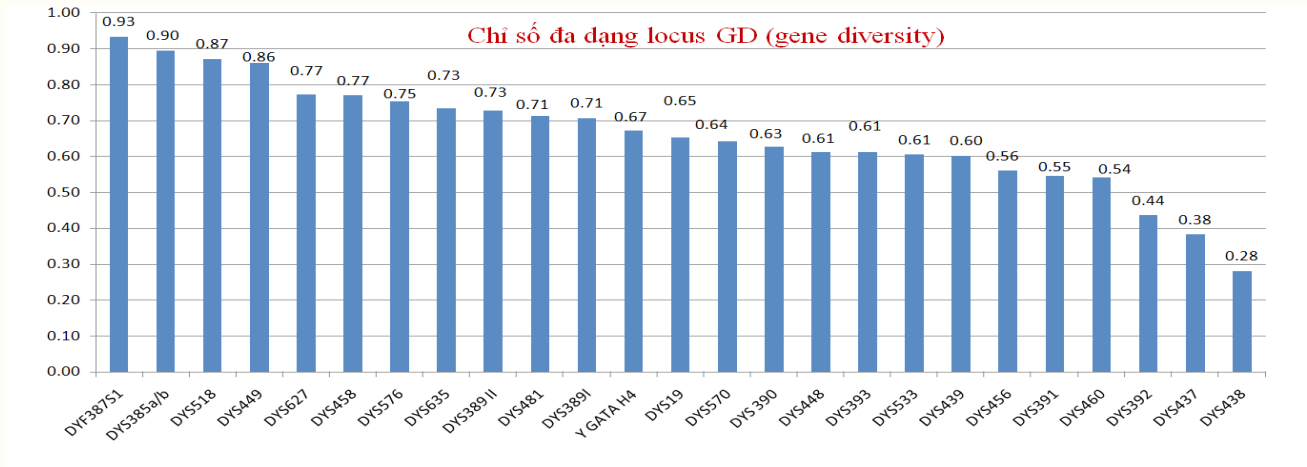
Trong nghiên cứu này, chúng tôi không phát hiện ra alen lạ nào không có trong thang alen tiêu chuẩn ở các mẫu khác nhau và không có locus nào 2 alen. Mức độ đa hình của một locus cụ thể có thể được đo bằng chỉ số đa dạng gen (chỉ số đa dạng locus). Chúng tôi phát hiện ra chỉ số đa dạng gen (GD) nằm trong khoảng từ 28% đối với locus DYS438 đến 93,4% đối với locus DYS387S1. Dựa vào các chỉ số này, có thể sắp xếp độ đa dạng gen từ cao đến thấp của 27 locus nghiên cứu lần lượt là: DYS387S1 *ab*, DYS385 *ab*, DYS518, DYS449, DYS627, DYS458, DYS576, DYS635, DYS389II, DYS481, DYS389I, YGATA H4, DYS19, DYS570, DYS390, DYS448, DYS393, DYS533, DYS439, DYS456, DYS391, DYS460, DYS392, DYS437, DYS438. Chỉ số đa dạng Haplotype và chỉ số khả năng phân biệt với tổ hợp 27 locus Y-STR là 1.

Quần thể nghiên cứu có tần số alen tối thiểu - MAF (Minimum allele frequency) là $5/2N = 0,0125$. Các alen xuất hiện với tần suất $\leq 0,0125$ được coi là các alen có tần suất rất thấp, ít xuất hiện trong

quần thể (có 27/168 alen, chiếm 16% trong quần thể nghiên cứu). Ví dụ alen 24, 25, 26 của locus DYS627; alen 10, 15, 16 của locus DYS392; alen 9 của locus DYS533; alen 18 của locus DYS456.

Đây là các alen hiếm gặp trong quần thể người Kinh ở tỉnh Thanh Hóa, Việt Nam. Số lượng alen có tần số thấp thấy nhiều nhất ở locus DYS458; không phát hiện thấy ở locus DYS391, DYS448.

- Chỉ số GD của từng locus:



Hình 1. Biểu đồ cột theo giá trị chỉ số GD của từng locus.

Để thấy sự tương đồng và khác biệt về phân bố tần suất alen, chúng tôi đã sử dụng cơ sở dữ liệu tham chiếu haplotype nhiễm sắc thể Y - YHRD (www.yhrd.org) để so sánh tần suất alen của 27 locus Y-STR với các nghiên cứu khác đã được công bố từ trước. Đây là một bộ cơ sở dữ liệu có chú thích, truy cập mở số mẫu quần thể trên khắp thế giới với mục tiêu để tạo ước tính tần suất đáng tin cậy cho Y-STR haplotypes, được sử dụng trong đánh giá khả năng trùng khớp trong pháp y và các trường hợp xác định quan hệ họ hàng. Kết quả cho thấy, phân bố tần suất alen trong nghiên cứu này tương tự như dữ liệu thống kê chung ở 14 locus Y-STR (bao gồm DYS387S1 a/b, DYS449, DYS627, DYS635, DYS389II, DYS481, DYS389I, DYS390, DYS439, DYS456, DYS391, DYS437, DYS438). Tuy nhiên, 13 locus (bao gồm DYS576, YGATA-H4, DYS448, DYS460, DYS458, DYS19, DYS392, DYS393, DYS570, DYS533, DYS518, DYS385 a/b) có tần suất alen khác với tần suất chung trong số liệu thống kê. Sự khác biệt phản ánh đặc điểm đa hình di truyền Y-STR ở quần thể người Kinh ở tỉnh Thanh Hóa, Việt Nam khi so sánh với các quần thể khác trên thế giới.

3.2. Phân tích so sánh khoảng cách di truyền giữa người Kinh ở Thanh Hóa, Việt Nam và các quần thể người trong khu vực

Bảng 6. Khoảng cách di truyền giữa một số quần thể người Việt Nam

| Quần thể | Thanh Hóa, Việt Nam | Miền Bắc, Việt Nam | Hà Nội, Việt Nam | Việt Nam (người Kinh) | TP. Hồ Chí Minh, Việt Nam |
|---------------------------------|---------------------|--------------------|------------------|-----------------------|---------------------------|
| Thanh Hóa, Việt Nam | - | 0,1634 | 0,6759 | 0,0331 | 0,3075 |
| Miền Bắc, Việt Nam | 0,0015 | - | 0,0276 | 0,1955 | 0,4549 |
| Hà Nội, Việt Nam | 0,00192 | 0,0193 | - | 0,0574 | 0,0399 |
| Việt Nam (Kinh) | 0,0028 | 0,0029 | 0,0163 | - | 0,0336 |
| Thành phố Hồ Chí Minh, Việt Nam | 0,00147 | 0,0015 | 0,0101 | 0,0253 | - |

So sánh tại bảng 6 cho thấy quần thể người Kinh ở Thanh Hóa nghiên cứu và các quần thể người Việt Nam khác đã được công bố trên YHRD có sự chênh lệch giá trị Rst không đáng kể. Điều này cho thấy, các quần thể dân tộc trong Việt Nam nói chung tương đối tương đồng, ít có sai khác trong cấu trúc di truyền.

Có sự sai lệch khoảng cách di truyền, điều đó chứng tỏ, quần thể người Việt Nam khá đa dạng, nên việc nghiên cứu với số lượng mẫu lớn và nhiều quần thể người Việt Nam khác nhau sẽ quan sát được đầy đủ cấu trúc di truyền hơn.

Bảng 7. Khoảng cách di truyền giữa quần thể người Kinh tại Thanh Hóa, Việt Nam với một số quần thể người trong khu vực.

| | Thanh Hóa, Việt Nam | India | Philippines | Thái Lan | Singapore [Han] | South Korea | Japan | China [Han] |
|---------------------|---------------------|---------|-------------|----------|-----------------|-------------|--------|-------------|
| Thanh Hóa, Việt Nam | - | 0 | 0 | 0,0167 | 0 | 0 | 0 | 0 |
| India | 0,136 | - | 0 | 0 | 0 | 0 | 0 | 0 |
| Philippines | 0,1069 | 0,0939 | - | 0 | 0 | 0 | 0 | 0 |
| Thái Lan | 0,0062 | 0,0119 | 0,818 | - | 0 | 0 | 0 | 0 |
| Singapore [Han] | 0,0385 | 0,12335 | 0,0768 | 0,0225 | - | 0 | 0 | 0,0565 |
| South Korea | 0,08 | 0,1705 | 0,1506 | 0,086 | 0,1609 | - | 0 | 0 |
| Japan | 0,0836 | 0,17 | 0,1457 | 0,0807 | 0,0919 | 0,0444 | - | 0 |
| China [Han] | 0,0292 | 0,0899 | 0,0665 | 0,0195 | 0,0052 | 0,0673 | 0,0808 | - |

Giá trị *Rst* giữa quần thể người Kinh tại Thanh Hóa (Việt Nam) trong nghiên cứu này với các quần thể được chọn (giá trị *p* hiển thị phía trên đường chéo, giá trị *Rst* phía dưới).

Kết quả trong bảng 7 cho thấy, quần thể người Kinh (Thanh Hóa, Việt Nam) khác biệt có ý nghĩa nhất so với quần thể người Ấn Độ và Philippines với $Rst > 0,1$; $p < 0,001$. Ngược lại, khoảng cách di truyền giữa quần thể người Kinh (Thanh Hóa, Việt Nam) so với quần thể người Thái Lan và người Hán (Trung Quốc) gần hơn so với các quần thể người châu Á khác. Trong đó, khoảng cách di truyền giữa quần thể người Kinh (Thanh Hóa, Việt Nam) và người Thái Lan là gần gũi nhất, với $Rst = 0,0062$ và $p = 0,0167$.

4. KẾT LUẬN.

Kết quả nghiên cứu cho thấy quần thể người Kinh ở Thanh Hóa, Việt Nam có chỉ số đa dạng haplotype tương ứng $HD = 1$ (không có haplotype nào trùng nhau trong số 200 mẫu nghiên cứu); chỉ số phân biệt $DC = 100\%$, chỉ số đa dạng locus GD giao động từ 0,28 (locus DYS43) đến 0,934 (locus DYF387S1).

Quần thể người Kinh ở Thanh Hóa, Việt Nam có tính đa hình cao ở các locus Y-STR; trong đó, locus đa hình nhất là DYF387S1a/b, locus ít đa hình nhất là locus DYS438.

Kết quả so sánh tần suất, độ phân bố 27 locus Y-STR trong nghiên cứu cho thấy, cấu trúc di truyền khá đồng nhất giữa quần thể người Kinh ở Thanh Hóa với các quần thể người Kinh ở Việt Nam. So

với các nước trong khu vực, quần thể người Kinh ở Thanh Hóa, Việt Nam có khoảng cách di truyền gần gũi nhất với quần thể người Thái Lan.

TÀI LIỆU THAM KHẢO

- Butler J.M (2003), "Recent developments in Y-short tandem repeat and Y-single nucleotide polymorphism analysis", *Forensic Science Rev*, 2003, 15: 91-111.
- Gusmão L, Butler J.M, Carracedo A, Gill P, Kayser M, Morling N, Prinz M, Roewer L, Tyler-Smith C, Schneider P.M (2006), "DNA commission of the international Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-Strs in forensic analysis", *Forensic Sci Int*, 2006, 157:187-97.
- Ballantyne K.N, Kayser M (2012), "Additional Y-STRs in Forensics: Why, Which, and When", *Forensic Sci Rev*, 24 (1): 63-78
- Purps J (2014), "A global analysis of Y-chromosomal haplotype diversity for 23 STR loci", *Forensic Sci Int Genet*, 12: 12-23.
- Walsh P.S, Metzger D.A, Higuchi R (1991), "Chelex 100 as a medium for simple extraction of DNA for PCR based typing from forensic material", *Biotechniques*, 1991, 10: 506-13. □