MINISTRY OF EDUCATION
AND TRAINING

VIETNAM ACADEMY
OF SCIENCE AND TECHNOLOGY

GRADUATE UNIVERSITY OF SCIENCE AND TECHNOLOGY

HA HUU HAO

THE STUDY ON Y CHROMOSOME'S MOLECULAR MARKERS FOR APPLICATION IN DNA FORENSIC TESTING

Major : Biotechnology
Code : 9 42 02 01

SUMMARY OF BIOLOGY DOCTORAL THESIS

Hanoi - 2023

The thesis has been completed at: Graduate University of Science and Technology - Vietnam Academy of Science and Technology

Supervisor 1: Prof. Dr. Chu Hoang Ha

Supervisor 2: Assoc. Prof. Dr. Le Van Son

Reviewer 1: ……

Reviewer 2: ……

Reviewer 3: ……

The thesis will be defended at the Board of Examiners of Graduate University of Science and Technology-Vietnam Academy of Science and Technology at……..on…….

The thesis can be referred at:

- Library of Graduate University of Science and Technology
- National Library of Vietnam

INTRODUCTION

1.The necessity of research

Since 1953 with the discovery of the double helix structure of DNA by Watson and Crick, molecular biotechnology has made great strides and has become an effective tool for many scientific disciplines such as medicine, agriculture, archeology, forensics... Especially, in the field of forensic examination, criminal science and DNA technology it also plays a very important role to solve cases such as criminal identification, paternity testing, finding missing people...

DNA analysis not only helps to solve civil cases such as determining paternity for children, inheritance disputes to criminal cases but also DNA is evidence of high legal value and decisive value in legal cases in courts. Moreover, in natural disasters, disasters with numerous deaths or when conventional identification is not possible, DNA testing becomes the most feasible method.

In DNA analysis field, the selection of molecular markers that are stabled and highly polymorphic gene loci, specific to each individual is important for the application of its. There are some different molecular markers. Among them, STR (short tandem repeat) - which are consecutively repeating DNA sequences with each repeat unit of 1 - 6 bp have been studied, selected and become the most popular molecular marker today. The STR loci have been studied, analyzed together to form commercial kits in order to analyze multiple STR loci quickly, accurately, high sensitivity.

Besides the STR locus on the autosomal pair (AS-STR), the study of STR loci on the X and Y sex chromosomes is also attracting the attention of scientists. STRs on the Y chromosome (referred to as Y-STR) have the main characteristic that they are inherited only in the paternal line without recombination through generations. Therefore, there will be similar Y-STR profiles between males of the same paternal line. When the use of chromosomal STRs is often not sufficient to satisfy the examination requirement, further analysis of the Y-STR profile will help to resolve many cases especially for DNA testing of samples from male such as identification determining the relationship according to the paternal line (brother - brother, uncle - grandchild ...), rape cases with male suspects, genetic genealogy according to the father's line...

Currently, Y-STR analysis is mainly performed through commercial kits, in which the most popular being the PowerPlex® Y23 system (Promega) with 23 Y-STRs and the Yfiler™ Plus PCR Amplification Kit (Thermo Fisher Scientific) with 27 Y-STRs. Two kits proved to be effective with different applications: distinguishing male individuals in the population, determining paternity, criminal identification on samples with low DNA concentration, samples mixed from many sources …Because the genetic structure of each population is different, it is necessary to evaluate the polymorphism and suitability of STRs in general and Y-STR in particular before using in the population. Each laboratory and research institute is recommended to build a table of allele distribution frequencies with different human populations in order to serve the calculation of reliability probability of paternity testing.

However, in Vietnam at present, there is no detailed study to investigate the distribution frequency, polymorphism, suitability, applicability of Y-STRs in a representative population of Vietnamese men. The further investigation of Y-STR loci in order to serve to calculate the reliability of Y-STR records,

probability of paternity relationship, application in criminal science is really necessary and meaningful. high practicality. Therefore, we carried out the project "**The study on Y chromosome's molecular markers for application in dna forensic testing**".

2. Research objective

- Select and study the related characteristics (allele distribution frequency, polymorphism, individual discrimination ability...) of 29 Y-STR markers in the Vietnamese male population of Kinh ethnicity.
- Research on some small size markers (mini STR) on Y chromosome.
- Investigate the potential application of Y-STR markers on many different types of samples (mixed samples, poor quality samples, decomposed samples).

3. The main contents of the thesis

To achieve the goal of the study, we have carried out the following main research contents:

- Select research subjects, collect samples and extract DNA.
- Implement the process of analyzing the Y-STR indicator from the research samples. Make a table of allele frequency distribution of Y-STR locus markers and calculate related indices: number of alleles, polymorphism, haplotype diversity, ability to distinguish individuals.
- Selecting and optimizing conditions for amplification of some mini STRs on the Y chromosome and initially applied in DNA assessment.
- Investigate the ability to create Y-STR indicators on many different types of examination samples such as samples from many sources, aged remains, and micro-trace samples.

CHAPTER 1. LITERATURE OVERVIEW

1.1. DNA analysis in forensic examination

1.2. Types of markers

1.3. Overview of the STR marker

1.3.1. Features in the genome

1.3.2. Classification of STR

1.4. STR on the Y - sex chromosome

1.5. Application directions of the STR marker on the Y - sex chromosome

1.5.1. Y-STR in the study of population genetic structure

1.5.2. Application of Y-STR in DNA relationship testing

1.5.3. Applications of the Y-STR in the field of criminal science

1.6. Analysis method of Y-STR markers

1.6.1. Analysis based on capillary electrophoresis

1.6.2. Analysis using commercialized Y-STR kits

1.6.3. Analysis using mini STR strategy

1.7. The importance of calculating allele frequencies for STR markers

1.8. Research status of Y-STR markers

1.8.1. Research situation in the world

1.8.2. Research situation in Vietnam

CHAPTER 2. MATERIALS AND METHODS

2.1. Materials

- Research on Y-STR locus belonging to PPY23 and Yfiler Plus kit: 400 samples of Kinh men collected in northern Vietnam, not related by blood relation. Samples collected include: hair, buccal swab or blood.

- Research on degraded samples: 30 bone samples were collected fromVietnam - Laos Friendship Martyrs Cemetery, Anh Son District, Nghe An Province (15 samples) and Dinh Hoa Martyrs Cemetery, Province Thai Nguyen (15 samples).

- Research on mixture samples: 50 samples of vaginal secretions collected from female victims in rape cases were provided by police agencies that requested DNA testing in the Faculty of Biomedicine - National Institute of Forensic Medicine.

2.2. Methods

2.2.1. DNA extraction method

The study uses 3 methods of DNA extraction depending on the type of sample and research purpose:

- DNA extraction method by Chelex® 100: This is the extraction method used for common samples such as: blood, hair, hair and even tissue samples that have not been decomposed.

- DNA extraction method using Qiamp® DNA micro kit: This method is used for samples with low DNA concentration, samples collected at crime scenes and difficult to extract DNA.

- DNA extraction method by QIAamp® DNA investigator kit of QIAgen - Germany: This method is used for decomposed bone and tooth samples, with low DNA concentration.

2.2.2. DNA quantification method

- For samples extracted by Chelex® 100 method: Measure the concentration of DNA after extraction with a Quantus Fluorometer (Promega) according to the manufacturer's instructions,

- For samples extracted using kit: after DNA extraction, DNA quantification was carried out by Realtime PCR method according to Trio DNA Quantification Kits (Applied Biosystem, USA) on 7500 Real-Time PCR machine (Applied Biosystem, USA).

2.2.3. Gel electrophoresis method

PCR products were detected by electrophoresis on a 6% Polyacrylamide gel.

2.2.4. PCR method

- Analysis of 29 Y-STR indicators: according to the PowerPlex® Y23 System kit (Promega - USA) and Yfiler™ Plus PCR Amplification Kit (Thermo Fisher Scientific - USA), PCR reaction components with two kits comply manufacturer's instructions.

- Analysis of 10 Y-STR mini indicators: primer pair design and PCR reaction cycle optimization

2.2.5. Capillary electrophoresis method

- Analyze 29 Y-STR indicators from 2 kits PPY23 and YPlus according to the manufacturer's instructions

- The product after PCR was mixed with the internal standard scale and Hi-Di Formamide, was placed into the well on a 96-well plate, shocked at 95oC, for 3 minutes. Then immediately transfer to ice for 3 minutes denaturation and into the 3500 Genetic Analyzer capillary electrophoresis system (Applied Biosystems - USA) for analysis with the program installed according to the manufacturer's instructions.

- Data were analyzed using GeneMapper® ID-X 1.3 software (Applied Biosystems - USA).

2.2.6. DNA sequencing method

According to the Sanger gene sequencing method with the main steps: purifying the product after PCR, Performing PCR reaction with BigDye Terminator v3.1 Cycle Sequencing Kit, purifying after running Bigdye, analyzing on the ABI 3500Genetic Analyzer system (Applied Biosystems, Thermo Scientific, USA).

2.3. Analyze and process data

The frequency of each allele in the loci, the frequency of each haplotype, the total number of alleles and the total number of haplotypes in the studies were determined using Excel 2013 and Arlequin 1.3 software.

The gene diversity (GD) of each Y-STR locus was calculated by the formula:

$$GD = n*(1- \Sigma pi2)/(n-1).$$

Haplotype diversity (HD) according to the formula: $HD = n*(1- \Sigma pi_{th}^2)/(n-1)$

Discrimination capacity (DC) is calculated by the formula: DC = h/n

In which n is the number of samples, pi is the frequency of the corresponding allele, p_{th} is the frequency of the corresponding haplotype, h is the number of different haplotypes observed in the population.

In addition, we also used the online tool AMOVA on the webiste http://yhrd.org to calculate the population genetic distance (Rst) and the combined probability value (P value) between the study population and the populations. neighboring populations

CHAPTER 3. RESULTS AND DISCUSSION

3.1. Total DNA extraction

With samples of hair, blood, buccal swab, total DNA after extraction was quantified by measuring absorbance on a Quantus Fluorometer (Promega company - USA). The total DNA yield with high purity ranged from 2.70 to 10.05 ng/µl. These total DNA samples all met the quality and quantity standards to be used as input DNA for the study of 29 Y-STR markers using the PPY23 and YPlus kits..

For samples of poor quality, samples that have been decomposed such as bone samples, samples from crime scenes, the concentration of DNA obtained after extraction is quite low so it is difficult to check the extraction efficiency by gel electrophoresis. Instead, we quantified DNA using Trio DNA Quantification Kits (Thermo Fisher Scientific - USA) on Real time PCR system. This is high sensitivity and is capable of separately quantifying large-sized, small-sized DNA fragments and DNA from Y- chromsome.

Real-time PCR results of bone and tooth samples illustrated R1-R5 (Table 3.1) showed that the content of large-sized DNA (T.Large autosomal) was very little while that of small-sized DNA (T.Small autsomal) was higher. The Degradation Index (DI) is calculated as the ratio between small size DNA content / large size DNA is > 1, indicating that the remains have decomposed DNA. Such samples will be classified as samples used in the study to evaluate the efficiency of Y-STR analysis on highly degraded samples.

Table 3.1. Real-time PCR results of bone and tooth samples are denoted from R1 to R5. T.Large autosomal: large size DNA, T.Small autosomal: small size DNA, T.Y: content of DNA on Y chromosome.

Notation	Target Name	Quantity (ng/ µl)	Degration index
R1	T.Large Autosomal T.Small Autosomal T.Y	0.0137 0.0179 0.0173	1.309
R2	T.Large Autosomal T.Small Autosomal T.Y	0.0065 0.0076 0.0093	1.165
R3	T.Large Autosomal T.Small Autosomal T.Y	0.0124 0.0202 0.0196	1.633
R4	T.Large Autosomal T.Small Autosomal T.Y	0.0063 0.0081 0.0097	1.276
R5	T.Large Autosomal T.Small Autosomal T.Y	0.0094 0.0133 0.0097	1.425

Realtime PCR results of the skeleton sample with symbol R1 form R5 show that the large amount of DNA (T.Large autosomal) is very little while the amount of DNA is small (T.Small autsomal). higher again. The decomposition index (DI-Degradation Index) is calculated as the ratio between the small size DNA content / large size DNA is 2.21 (>1), indicating that the remains have decomposed DNA. Such samples will be classified as samples used in the study to evaluate the efficiency of Y-STR analysis on highly decomposed samples.

3.2. 3. Research results of 29 Y-STR markers from PowerPlex® Y23 System (PPY23) and Yfiler Plus kit

3.2.1. Selection of Y-STR in research

Up to now there are about 400 Y-STR markers have been well studied for their structure and position on chromosomes. However, in order to be selected into a locus with potential applications in DNA identification, it is necessary to meet a number of requirements such as: high polymorphism, relatively large number of alleles, not too complicated repeating structure, which has been studied. clear gene sequence, simple analysis method, accurate reading results. Based on those criteria in this study, we selected 29 Y-STR loci to survey in Kinh male population using two popular commercial Y-STR kits:PowerPlex® Y23 System (Promega, Abbreviation: PPY23) and Yfiler™ Plus PCR Amplification Kit (Thermo Fisher Scientific, Abbreviation: YPlus).

3.2.2. Establishment of Y-STR profiles

Following the supplier's instructions, we obtained Y-STR data from 400 samples that met the requirements for a complete Y-STR profile. Cases that do not meet the requirements such as low peaks, infected samples during manipulation (peak more than 2 allele at common loci and more than 3 allele at loci DYS385ab and DYF387S1ab) will not be included in the statistical data.

3.2.3. Frequency table of 29 Y-STR loci from PowerPlex® Y23 System (PPY23) and YPlus kits

The frequencies of each allele were statistically and calculated using Arlequin 1.3 software to form a table of the distribution frequencies of 29 Y-STRs. The allele frequency table of each locus were also compared with the general data gathered from the data published on the YHRD website (www.yhrd.org) to compare similarities and differences in frequency. allele distribution of 29 Y-STR loci in this study with general data. For each allele locus with the greatest distribution frequency is bolded in the statistical table (Table 3.2).

Comment: Thus, from profiling Y-STR of 400 research samples with 2 sets of PPY23 and YPlus, we have established a table of individual allele frequency distribution of 29 Y-STR loci. The comparison results show that there are 16/29 loci with allele frequency distribution relatively similar to the general data on YHRD. In contrast, 13/29 loci had a different distribution than the general data. In addition, statistics also show that there are 41 rare alleles/29 loci. This shows the difference in the genetic structure of the Kinh and Vietnamese populations when compared with other populations.

Table 3.2: Allele Frequencies for the 29 Y-STR markers in the Kinh population (n = 400)

DYS576			**DYS389 I**			**DYS448**			**DYS389 II**			**DYS19**		
Alen	PPY23	YPlus	Alen	PPY23	YPlus	Alen	PPY23	YPlus	Alen	PPY23	YPlus	Alen	PPY23	YPlus
14	0.000	0.005	11	0.085	0.085	16	0.010		25	0.000	0.005	12		0.005
15	0.030	0.015	12	0.310	0.350	17	0.010	0.045	27	0.050	0.095	13	0.045	0.015
16	0.080	0.060	**13**	**0.385**	**0.385**	**18**	**0.425**	**0.530**	28	0.295	0.260	14	0.120	0.135
17	0.160	0.185	14	0.215	0.180	18.2	0.010		**29**	**0.355**	**0.340**	**15**	**0.415**	**0.450**
18	**0.335**	**0.375**	15	0.005		19	0.235	0.215	30	0.225	0.230	16	0.355	0.330
19	0.265	0.265				19.2	0.005		31	0.060	0.055	17	0.065	0.065
20	0.110	0.080				20	0.180	0.155	32	0.015	0.010			

21	0.015	0.015				21	0.115	0.055	33		0.005				
22	0.005					22	0.010								
	DYS391			**DYS481**			**DYS549**			**DYS533**			**DYS438**		
sAlen	PPY23	YPlus	Alen	PPY23	YPlus	Alen	PPY23	YPlus	Alen	PPY23	YPlus	Alen	PPY23	YPlus	
6	0.020		19		0.005	10	0.005		9	0.005		8	0.005	0.005	
8		0.005	21	0.020	0.020	11	0.175		10	0.340	0.390	9	0.005	0.005	
9	0.010	0.025	22	0.085	0.130	**12**	**0.560**		**11**	**0.445**	**0.435**	**10**	**0.840**	**0.825**	
10	**0.630**	**0.615**	**23**	**0.330**	**0.405**	13	0.235		12	0.190	0.165	11	0.120	0.145	
11	0.325	0.330	24	0.295	0.170	14	0.015		13	0.015	0.010	12	0.030	0.015	
12	0.015	0.025	25	0.140	0.115	15	0.010		14	0.005		14		0.005	
			26	0.095	0.105										
			27	0.030	0.035										
			28	0.005	0.015										
	DYS437			**DYS570**			**DYS635**			**DYS390**			**DYS439**		
Alen	PPY23	YPlus	Alen	PPY23	YPlus	Alen	PPY23	YPlus	Alen	PPY23	YPlus	Alen	PPY23	YPlus	
12	0.000	0.005	14	0.000	0.025	18	0.005	0.005	22	0.025	0.035	10	0.035	0.025	
13	0.005	0.000	15	0.040	0.045	19	0.050	0.030	23	0.255	0.180	11	0.255	0.260	
14	**0.730**	**0.730**	**16**	**0.330**	**0.310**	20	0.130	0.110	**24**	**0.360**	**0.385**	**12**	**0.530**	**0.520**	
15	0.255	0.250	17	0.230	0.175	**21**	**0.365**	**0.405**	25	0.305	0.365	13	0.160	0.160	
16	0.010	0.015	18	0.135	0.135	22	0.195	0.205	26	0.055	0.035	14	0.015	0.030	
			19	0.145	0.140	23	0.190	0.195				15	0.005	0.005	
			20	0.075	0.120	24	0.050	0.040							
			21	0.025	0.030	25	0.015	0.010							
			22	0.015	0.010										
			23	0.005	0.010										
	DYS449			DYS518			DYS393			DYS458			DYS627		
Alen	PPY23	YPlus	Alen	PPY23	YPlus	Alen	PPY23	YPlus	Alen	PPY23	YPlus	Alen	PPY23	YPlus	
23		0.005	32	0.005		10	0.000	0.005	12		0.010	13		0.005	
25		0.040	33	0.005		11	0.000	0.025	13	0.005	0.005	14			
26		0.160	34	0.015		**12**	**0.375**	**0.310**	14	0.025	0.015	15			
27		0.080	35	0.060		13	0.180	0.165	15	0.120	0.135	16		0.005	
28		0.090	36	0.120		14	0.400	0.465	16	0.170	0.190	17		0.005	
29		0.105	37	0.120		15	0.045	0.030	17	0.195	0.200	18		0.065	
30		0.095	38	0.095					**18**	**0.310**	**0.295**	19		0.080	
31		**0.145**	39	0.105					19	0.140	0.085	20		0.230	
32		0.125	**40**	**0.150**					20	0.025	0.045	21		0.230	
33		0.050	41	0.145					21		0.015	**22**		**0.235**	
34		0.040	42	0.105					22	0.010	0.000	23		0.100	
35		0.040	43	0.065					23		0.005	24		0.045	
36		0.025	44	0.005											
			45	0.005											
	DYS456			**YGATAH4**			**DYS643**			**DYS460**			**DYS392**		
Alen	PPY23	YPlus	Alen	PPY23	YPlus	Alen	PPY23	YPlus	Alen	PPY23	YPlus	Alen	PPY23	YPlus	
12		0.005	10	0.130	0.125	8	0.015		9		0.110	10		0.010	
13	0.090	0.060	11	0.365	0.425	9	0.075		**10**		**0.645**	11	0.100	0.045	
14	0.180	0.165	**12**	**0.465**	**0.425**	10	0.190		11		0.190	12	0.035	0.030	
15	**0.525**	**0.570**	13	0.040	0.025	11	0.290		12		0.050	**13**	0.655	0.700	
16	0.160	0.155				**12**	**0.350**		13		0.005	14	0.200	0.205	
17	0.025	0.040				13	0.080					15	0.010	0.010	
18	0.020	0.005													

Table 3.3. Allele Frequency distribution of DYS385a/b and DYF387S1

DYS385a/b						**DYF387S1**	
Alen	PPY23	YPlus	Alen	PPY23	YPlus	Alen	YPlus
9, 17	0.005		13,20	0.040	0.035	34,34	0.010
11,11	0.015	0.005	13,21	0.005	0.010	34,36	0.005
11,12		0.015	13,22	0.010	0.005	34,37	0.005
11,13	0.005		14,14		0.005	34,40	0.005
11,14	0.005	0.005	14,15	0.010		35,36	0.005
11,15	0.005		14,16	0.005	0.015	35,37	0.010

11,17	0.005	0.005	14,17	0.005		35,38	0.045
11,18	0.035	0.020	14,18	0.025	0.010	35,39	0.015
11,19	0.010		14,19	0.020	0.025	35,40	0.010
11,20	0.010	0.005	14,20	0.010		36,36	0.040
12,12	0.025	0.010	14,21	0.015		36,37	0.050
12,12.3	0.005		15,15		0.010	36,38	0.100
12,13	0.005	0.010	15,16	0.005	0.010	36,39	0.070
12,14	0.005	0.015	15,17		0.010	36,40	0.025
12,15		0.005	15,18	0.030	0.005	36,41	0.015
12,16	0.015	0.015	15,19	0.020	0.060	37,37	0.090
12,17	0.020	0.020	15,20	0.015	0.025	**37,38**	**0.140**
12,18	0.050	0.030	15,21	0.015	0.015	37,39	0.080
12,19	0.035	0.035	15,22		0.005	37,40	0.015
12,20	0.040	0.055	15,23		0.005	37,41	0.010
12,21	0.010		16,17	0.005		38,38	0.095
13,13	16.000	0.080	16,18	0.005		38,39	0.040
13,14	3.000	0.015	16,19	0.005	0.015	38,40	0.030
13,15		0.015	16,20	0.020	0.015	39,39	0.040
13,16	0.005	0.015	16,21	0.015	0.005	39,40	0.020
13,17	0.050	0.045	17,20	0.010		39,41	0.005
13,18	**0.150**	**0.185**	18,19		0.005	40,40	0.020
13,19	0.120	0.105				40,41	0.005

The Y-STR distribution frequency table in this study is the table with the largest number of research samples and loci surveyed in the Vietnamese population so far with the number of samples and research methods. achieve high reliability. Previous publications for the Vietnamese population were mainly made with 13-17 Y-STRs on a relatively small sample size of less than 200 samples. It is also necessary to build a frequency table of Y-STR loci, to meet the needs of calculating the criminal index related to men in the Vietnamese population. On the general YHRD data, it can be seen that the number of haplotypes published using 2 kits PPY23 and YPlus from studies with worldwide populations is quite large. In contrast to the Vietnamese population before this study, the number of published haplotypes was very limited with only 45 haplotypes from the PPY23 set and 47 halotypes with the YPlus set contributed through a few small studies on the Vietnamese population. Male. With this study, we added 200 haptypes using the PPY23 set to the general YHRD data with the access code YA004576, bringing the total number of hapatterns on the YHRD data in the Vietnamese population to 245 with the PPY23 set.

3.2.4. Characterization of 29 Y-STR markers

For each locus, the number of alleles ranged from 4 to 15 alleles with an average of 7 alleles/1 locus. In which, the PPY23 kit has a total of 144 alleles and the YPlus kit has 183 alleles. With the combination of 2 alleles, DYS385a/b has 55 different allele combinations, DYF387S1a/b has 28 different allele combinations observed. With 2 both kits, DYS458 and DYS570 are the loci with the highest number of alleles with about 10 alleles for each locus. In contrast, DYS437, DYS438, DYS393, DYS390 have the least number of alleles (4-5 alleles/locus). Particularly, the YPlus kit has DYS518, DYS449 are newly

added loci compared to the PPY23 set and are also the loci with the highest number of alleles (DYS518: 14 alleles, DYS449: 13 alleles).

The total number of alleles observed in the study was 209 alleles, significantly more than previously reported with the Vietnamese population. In a survey study on 205 men using the AmpFLSTR™ Yfiler™ PCR Amplification Kit including 17 loci of Toan T.T only 99 alleles were recorded. The number of alleles in the study is also at the same level when compared to other populations in the world. For example, in a survey study on 581 Han ethnic men in Shaanxi province - China using the PPY23 kit there are 237 alleles were observed. The result from a survey of 436 male samples in the United Arab Emirates (UAE) using the YPlus set showed 202 different alleles.

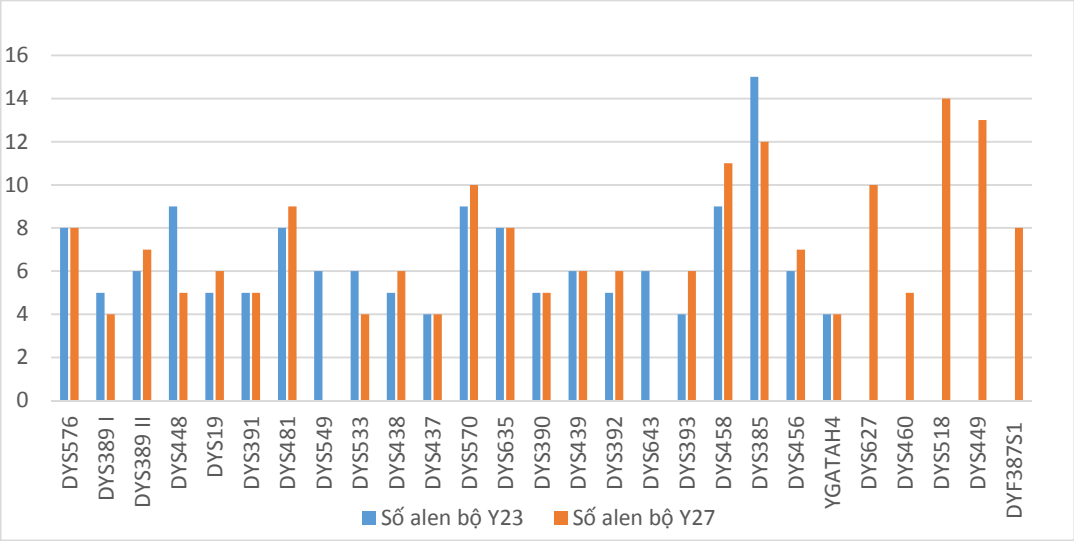

Figure 3.1: Allele numbers of 29 Y-STR loci in PPY23 and YPlus kit

In this study, no case was found outside the allele ladder (off ladder), detected 02 allele variants at locus DYS448 (allele 18.2, 19.2), 1 variant at locus DYS385b (allele 12.3). During the survey, 2 cases of allele loss (also known as null allele) were observed. In the first case, the loss of 18/23 Y-STR loci was detected when analyzed with the PPY23 kit. The other case has null alleles at 4 loci DYS481, DYS576, DYS570, DYS458. Reconfirmation analysis using the YPlus kit also gives similar results. Based on the location of the loci lost on the map of the Y chromosome, it was found that this is a case of large deletion on both the long arm and the short arm of the Y chromosome.

3.2.5. Gene diversity (GD)

From the frequency of allele distribution, we calculated the polymorphism or gene diversity (GD) of each Y-STR locus in this study. The results showed that two loci with a combination of 2 alleles, DYS385a/b and DYS387S1a/b were the highest genetic diversity loci (GD 0.941, 0.934, respectively). With single allele loci shared in two kits, DYS458 and DYS570 is considered as highest polymorphism loci

(mean GD 0.809, 0.808 respectively) and also the loci with the highest number of alleles (9 - 10 alleles, respectively). In contrast, the DYS438 locus was the locus with the lowest polymorphism (GD about 0.28) with only 4-5 alleles detected in the Vietnamese population, followed by DYS437 (GD: 0.490), DYS392 (GD: 0.495), DYS391 (GD:0.507). These are loci with the lowest polymorphism less than 0.5.

Previous studies on the Vietnamese Kinh population also show that DYS437, DYS438 are also loci with the lowest GD value. This result is quite similar to the global analysis of 23 Y-STR loci which showed that the DYS391, DYS437 and DYS438 loci have lower GD values than other Y-STR loci in Asian populations. Frequency survey of 23 Y-STRs from 129 populations in 51 countries by Josephine Purps et al. The comparison results of 23 Y-STR loci polymorphisms with some other countries in the region also show similarities. Specifically, in a study on Filipino populations, DYS391 and DYS438 had the lowest polymorphism, 2 markers DYS385 and DYF387S1 had the highest polymorphism. In a survey of 29 Y-STR loci with the Shanghai population (China) showed that DYS449 was the locus with the highest frequency (GD: 0.8966), whereas DYS438 had the lowest frequency (GD: 0.4186) followed by DYS437. The polymorphisms of the 6 new loci in the YPlus set were compared with the statistics on the Asian population in the US provided by the manufacturer and also showed a correlation in value.

However, the polymorphisms also differed significantly from the general data at 4 loci: DYS392, DYS391, DYS438 and DYS460. In which locus DYS392, DYS438, DYS460 in the Vietnamese population has a much lower polymorphism than other countries in the same region, whereas DYS391 has a higher polymorphism. These differences reflect the characteristics and segregation of genetic capital in the Kinh and Vietnamese populations compared with other populations.

Table 3.4 . The genetic diversity of 29 Y-STR loci in this study and comparison with the average GD value of the Asian population

STT	Locus	PPY23 GD	YPlus GD	Standard deviation		Average GD	GD of Asia population	Standard deviation
1	DYS576	0.776	0.748	0.01		0.762	0.7996	0.02
2	DYS389 I	0.713	0.693	0.01		0.703	0.6564	0.02
3	DYS389 II	0.734	0.751	0.01		0.742	0.6585	0.04
4	DYS448	0.713	0.647	0.03		0.680	0.7650	0.04
5	DYS19	0.685	0.669	0.01		0.677	0.6920	0.01
6	DYS391	0.499	0.514	0.01		**0.507**	**0.4136**	**0.05**
7	DYS481	0.771	0.768	0.00		0.769	0.8416	0.04
8	DYS549	0.603				0.603	0.6425	0.02
9	DYS533	0.653	0.635	0.01		0.644	0.6265	0.01
10	DYS438	0.280	0.300	0.01		**0.290**	**0.5707**	**0.14**

11	DYS437	0.404	0.406	0.00		0.405	0.4891	0.04
12	DYS570	0.795	0.821	0.01		0.808	0.8313	0.01
13	DYS635	0.774	0.745	0.01		0.760	0.7700	0.01
14	DYS390	0.712	0.687	0.01		0.700	0.7325	0.02
15	DYS439	0.630	0.638	0.00		0.634	0.6823	0.02
16	DYS392	0.522	0.467	0.03		**0.495**	**0.7434**	**0.12**
17	DYS643	0.749				0.749	0.7510	0.00
18	DYS393	0.668	0.662	0.00		0.665	0.6597	0.00
19	DYS458	0.806	0.813	0.00		0.809	0.8275	0.01
20	DYS385	0.947	0.936	0.01		0.941	0.9741	0.02
21	DYS456	0.661	0.622	0.02		0.641	0.6093	0.02
22	YGATAH4	0.635	0.626	0.00		0.630	0.6345	0.00
23	DYS627		0.820			0.820	0.8120	0.00
24	DYS460		0.536			**0.536**	**0.6750**	**0.07**
25	DYS518		0.893			0.893	0.8670	0.01
26	DYS449		0.900			0.900	0.8820	0.01
27	DYF387S1		0.934			0.934	0.9450	0.01

The results of sorting by GD value of 29 Y-STR loci in Figure 3.2 show that the number of loci with polymorphism from 0.7 - 1 accounts for the largest proportion with 84.2% (including 16/29 loci with 5 loci with GD > 0.9, 4 loci with GD in the range 0.8-0.9, 7 loci with GD in the range 0.7-0.8), 9 loci with GD between 0.5 - 0.6 accounted for 31% and only 4 loci had GD ≤ 0.5. It shows that most of the loci in the study have high polymorphism, which has great potential in distinguishing male individuals in the surveyed Kinh and Vietnamese populations.

3.2.6. Haplotype diversity (HD) and Discriminating capacity (DC)

With each kit separately, when analyzing 200 research samples we observed 200 different haplotypes, meaning that each haplotype is unique in the survey population. Therefore, the haplotype diversity (HD) is 1.00. Discriminating capacity (DC) was calculated as the ratio of the number of different haplotypes observed in a population to the total number of samples studied. Because the haplotypes are all unique, the ability to distinguish individuals is 100% with the two sets of PPY23 and YPlus. This reflects the high potential of the 29 Y-STR loci in discriminating paternally unrelated males.

This result is also consistent with the study of Toan T.T when surveying on the Kinh population, Vietnam with the combination of 17 Y-STR in the Yfiler set also gave 100% discrimination ability. Besides, in Koji Dewa's study, a survey of 119 samples on the Vietnamese population with a combination of 13 Y-STRs showed only 95% discrimination. This again suggests that increasing the number of analyzed Y-STRs will contribute to increased haplotype diversity and discriminant ability.

Survey studies with the PPY23 kit on 129 different populations around the world also show that Asian populations have the highest discrimination ability (> 97%), followed by Europe and Latin America (DC approx. 96%) and finally Africa (DC about 85%).

3.3. Research on mini Y-STR markers

3.3.1. Mini Y-STR selection

In commercial kits, different Y-STR loci are often amplified simultaneously in one reaction PCR so in order to distinguish during capillary electrophoresis requires loci has different size. As a result, some Y-STR loci which are quite large (>200 bp) are difficult to analyze successfully in samples with degraded DNA, much broken DNA. In the world, recognizing the importance of STR in general in analyzing degraded samples, many laboratories have researched and developed kits including mini STRs such as the AGCU Mini-STR Amplification Kit (AGCU). ScienTech Incorporation, China), AmpFLSTR MinifilerTM kit (Applied Biosystem) includes 8 STRs in the nucleus and a sex-determining Amelogenin indicator. However, these are mostly autosomal STRs. Until now there is no commercialized kit specifically for mini STR in Y chromosome. Mini Y-STRs with short size below 200 bp are suitable for degraded biological samples with much broken DNA.

Currently in Vietnam, the number of research works on the Y-STR locus is very limited and there is no research on the design of primer pairs to amplify mini Y-STR locus in order to analyzed degraded samples. Therefore, in order to achieve good results in DNA remain analysis at the National Institute of Forensic Medicine, we tried to study new mini Y-STR loci with potential applications in forensic medicine.

In this study, we selected loci with high polymorphism in the Vietnamese population but had large sizes in commercial kits and tried to minimize the size of Y-STR loci to achieve high success rate of analysis with decomposed samples. Primer pairs are designed to amplify 10 Y-STR loci include 5 loci already present in PPY23 kit and 5 new loci not yet available in commercial kit. 5 new loci including: DYS505, DYS522, DYS508, DYS460, DYS388 are high polymorphism with potential application in Y-STR analysis according to a number of published studies.

3.3.2. Optimization of PCR reaction conditions

PCR reaction was investigated with 10 primer pairs in 10 separate PCR reactions. The annealing temperature of each locus ranged from 56 to 58 oC. The PCR reaction cycle is optimized with conditions including: primer concentration, primer attachment temperature, chain elongation time. The product after PCR was tested on 6% Polyacrylamide gel. The electrophoresis images show clear PCR product bands, no extra bands and the correct size as previously reported. The locus sizes ranged from 91-213bp. This suggests that the designed primer pairs are available for the amplification of the desired Y-STR loci.

PCR reactions were examined in 10 separate PCR reactions. The PCR reaction cycle is optimized with conditions including: primer concentration, primer attachment temperature, chain elongation time. The product after PCR was tested on 6% Polyacrylamide gel.

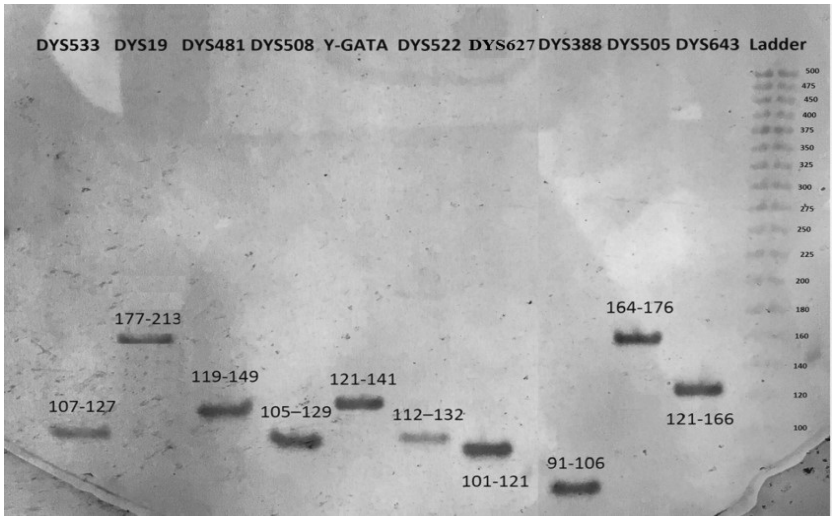

Figure 3.2: Electrophoresis image of 10 Y-STR loci PCR products on 6% Polyacrylamide gel using ILS 500 standard ladder (Promega - USA).

Performing single PCR reactions with the same sample requires a lot of chemicals, time and effort. The multi-primer PCR reaction is designed based on mixing multiple primer pairs into the same reaction to amplify multiple Y-STR loci simultaneously. The principle to select Y-STR loci to amplify together is that the loci must have similar annealing temperatures, otherwise the PCR products must have different sizes to ensure separation during electrophoresis. check. In this study, because the Y-STR loci are in the range of 100-200 bp in size, the size difference is not much, so the number of loci in the multiplex PCR reaction does not exceed 4 loci.

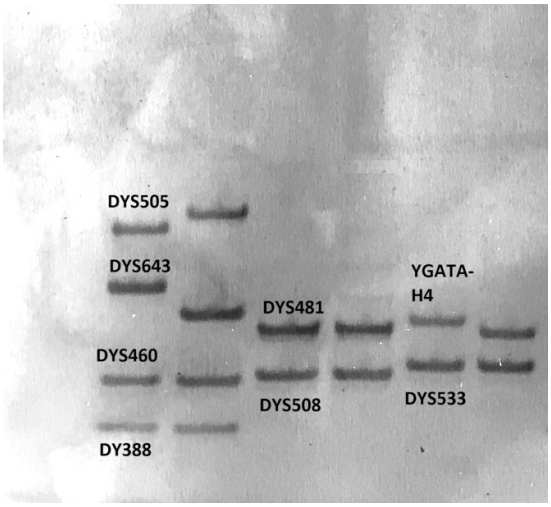

Figure 3. 3. Polyacrylamide gel electrophoresis images of multiplex PCR reactions

We optimized primer concentration, annealing temperature and successfully amplified 8 out of 10 Y-STR loci through 3 different multiplex PCR reactions. Multiplex PCR with multiple primer pairs is also the basis for creating commercial STR and Y-STR kits that can analyze up to 27 loci through a single PCR reaction.

3.3.3. Sequencing to determine the repeat structure of new Y-STR loci

Sequencing results on ABI 3500, Pop7, Cappilarry 50 cm with the company's running program for sequences with clear wave peaks. The sequencing also allows to clearly identify the repeating motif, the number of characteristic repeats of each Y-STR locus. For example, the locus DYS505 has a repeating structure of TCCT with a repeat count of 15 or the allele 15 (Figure 3.4).

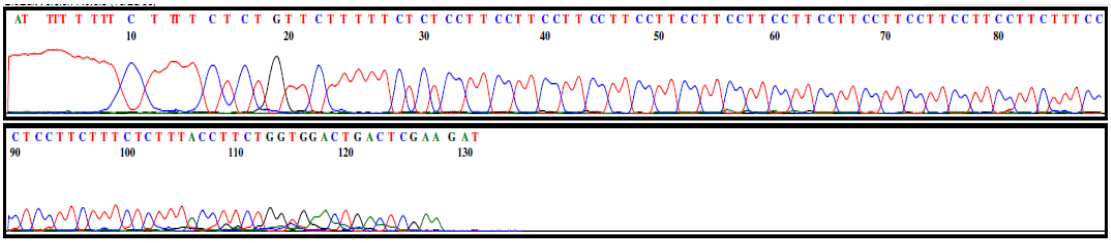

Figure 3.4. Results of sequencing the DYS505 . locus

The results of sequencing the DYS505 locus showed a fragment size of more than 170 bp and a TCCT repeat structure. The sequencing also allows to clearly identify the repeating motif, the number of characteristic repeats of each Y-STR locus. For example, the DYS505 locus has a repeat structure of TCCT with a repeat count of 15 or the 15 allele (Figure 3.4). This is also the basis for building allele ladders for new loci that are not yet available in commercial kits. Counting the number of repeats of the core structure by sequencing method also coincided with the analysis by commercial kit as illustrated locus DYS643 (Figure 3.5).

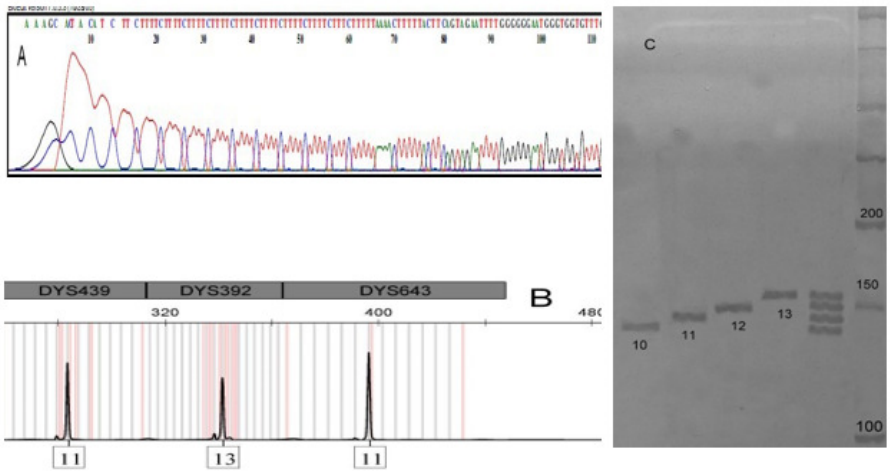

Figure 3.5. Result of making mini standard scale Y-STR with DYS643

A. The results of sequencing with the number of repeats CTTTT is 11, B. The results of analysis using the PPY23 set with the number of alleles is 11, C: The indicator standard scale DYS643 with 4 alleles 10, 11,12, 13

In this study, with samples that have been identified alleles, we built the allele scale of the DYS643 marker (available in the PPY23 set) based on the mini Y-STR method. This allele scale will serve as a basis for determining the exact number of alleles without the need for re-sequencing in future analyses, as well as the basis for distinguishing allele differences between samples

3.3.4. Analytical results of 10 mini Y-STRs on degraded samples

Based on the multi-primer PCR method with 10 pairs of primers mentioned above, we have applied it in the analysis of 50 decomposed research samples, which are bone and tooth samples of martyrs collected from cemeteries. With the tooth and bone samples decomposed, the amplification success rate of 10 mini Y-STRs ranged from 44% to 82%. Specifically: 2 loci DYS533 and DYS481 achieved the highest success rate (80%, 76.7%, respectively) with locus size only about 100 - 150 bp. Followed by the loci DYS388, DYS508, DYS460, YDATA-H4 achieved a success rate of 66.7 - 73.3% with amplification size range from 91 - 141 bp. Two loci DYS522, DYS505 with sizes about 112 - 176 bp only achieved 53.3% and 56.7% success rates. Particularly, two large-sized loci, DYS19 (177-213 bp) and DYS643 (121 - 166 bp), achieved relatively low amplification rates, 43.3% and 50%, respectively.

Table 3.5. Statistical table of successful amplification rates of 10 mini Y-STR loci on 30 samples

Locus	Size in PPY23 (bp)	Size in YPlus (bp)	Mini Y-STR size (bp)	Success rate
DYS533	245-285	338-379	107-127	80.0%
DYS19	312-352	184-224	177-213	43.3%
DYS481	139-184	207-252	119-149	76.7%
DYS643	368-423		121-166	50.0%
Y-GATA-H4	374-414	236-264	121-141	70.0%
DYS460		79-109	101-121	60.0%
DYS505			164-176	56.7%
DYS508			105–129	73.3%
DYS388			91-106	66.7%
DYS522			112–132	53.3%

This can be explained because the larger the locus (ex DYS19 ~200bp locus), the lower the amplification success rate. Especially with the remains with DNA that has been decomposed and broken many times.

In addition, another factor affecting the PCR performance is the quality of the research sample. In the study with long-term remains that were heavily decomposed (high decomposition index), with much broken DNA, it was difficult to amplify all Y-STR loci. Especially in Vietnam's hot and humid climate,

many types of microorganisms in the ground greatly affect the quality of bone samples. Based on the research results, it can be seen that there are at least 9 unsuccessful samples when using mini STR. It can be seen that the application of mini Y-STR solves many situations, especially in the assessment of degraded samples.

3.4. Efficacy of Y-STR markers in DNA testing

3.4.1. Effective in analyzing degraded samples

After successfully studying the technique of amplifying Y-STR minis with the corresponding primer pairs, we tried to apply it to a number of specific assessment cases at the Faculty of Medicine - Biology, National Institute of Forensic Medicine.

3.4.1.1. In the assessment work

Currently, every year at the Faculty of Biomedical Sciences, the Department of Biomedical Sciences is carrying out the task of examining the case from the decision to solicit the expertise of the police agency. Many of the samples collected at the crime scene were collected some time after the crime occurred, causing the quality of the DNA on the specimen to be intact or the amount of DNA collected on the sample to be very small (the sample was in the form of microscopic traces). This makes it very difficult to assess and draw conclusions. With the effectiveness of the mini-study Y-STR, we have applied it in practice and obtained certain results that can be illustrated through a specific case. In these cases, the comparison test between the field records or the victim's samples with Y-STR records by 2 sets of PPY23 and YPlus did not increase enough at all loci, especially loci with large size, so it is not enough basis to compare with the suspect sample. Then, further analysis of the Y-STR minis will help add data to the Y-STR file, allowing the conclusion that the sample collected at the scene (the victim) is the same/different from the sample collected from the suspect. This is an effective measure in a number of cases that have been examined.

3.4.1.2. In the assessment of martyrs' identities

In the assessment of the identity of the remains of martyrs which are samples with much damaged, broken DNA. Normally, to perform the autopsy, it will be based on analysis and comparison of sequences of HV1, HV2 supervariable regions on mitochondrial DNA to determine the maternal lineage relationship. However, in some cases, when analyzing HV1, HV2 regions on mitochondrial DNA, there is a very small difference at 1-2 positions. This difference is most likely due to differences in bloodlines between the two individuals or due to mitochondrial mutations, amination due to prolonged exposure of samples to unfavorable environmental conditions. With these cases, there will not be enough grounds to conclude the maternal lineage relationship. Or there are cases where there is only a comparison sample according to the paternal line (sons, grandsons...) then the mitochondrial DNA analysis method cannot be applied. Further analysis of the mini Y-STR will then provide additional data that is of great value in reaching the final conclusion.

3.4.1.3. The effectiveness of 2 kits in analyzing degraded samples

In addition to the Y-STR minis, the PPY23 and YPlus kits also include a number of alleles with both small size and genetic diversity, which will be the ideal gene loci for additional analysis in the identification cases. DNA with samples has been decomposed a lot, DNA has been broken into small fragments such as in bone samples, samples after disasters, natural disasters.. To evaluate the analytical potential of degraded samples, we filter them separately. Data collection of Y-STRs with size < 220 bp belonging to 2 kits when analyzing with 400 research samples and calculating related genetic indices.

Table 3.6. Comparison of genetic parameters obtained from the set of Y-STR markers with size < 220 bp belonging to PPY23 and Yplus kit

	Haplotype from 8 loci with size <220 bp belonging to PPY23	Haplotype from 10 loci with size <220 bp belonging to YPlus
Appears only in 1 individual	188	178
Appears in 2 individual	3	4
Appears in 3 individual	2	1
Appears in 4 individual	0	1
Appears in 5 individual	0	0
Appears in 6 individual	0	0
Appears in 7 individual	0	1
Appears in 8 individual	0	0
Appears in 9 individual	0	0
Appears in 10 individual	0	0
Total number of loci	8	10
Total number of observed	193	185
HD	0.99954	0.99829
DC	97%	93%

PPY23 kit includes 8 loci with size < 220 bp (DYS576, DYS389I, DYS391, DYS481, DYS570, DYS635, DYS393 and DYS458). YPlus set includes 10 loci with size < 220 bp (DYS576, DYS389I, DYS458, DYS456, DYS390, DYS570, DYS437, DYS393, DYS439 and DYS460). The results in Table 3.5 show that with the combination of Y-STRs belonging to the PPY23 set, the same haplotype can appear in 2, 3 individuals while the YPlus set had the same haplotype in 2, 3, 4 and 7 individuals. The number of haplotypes observed in the PPY23 and YPlus sets was 193, 185 haplotypes, respectively (in total of 200 analyzed samples). The discriminant ability of the PPY23 (97%) is also significantly more than that of the YPlus (93%). Thus, Y-STR collection < 220 bp in size in the PPY23 set has a better discriminant ability than that in the YPlus set although PPY23 has number of Y-STRs smaller than in YPlus. This is explained

by the fact there are 2 small size loci with high polymorphism: DYS481 (GD: 0.77) and DYS635 in PPY23 significantly increase the ability to differentiate.

This result is also consistent with previous publications showing the high potential of the two kits in discriminating degraded samples. The Y-STR loci are also suitable for analyzing samples of victims in natural disasters, disasters, samples from war victims.

3.4.2 Effective use of Y-STR in mixed sample analysis

In this study, we analyzed 50 samples with both PPY23 and YPlus kits to evaluate the ability to generate complete Y-STR profiles of both kits. These samples were all collected in the vagina of female victims from the assessment of rape/sexual assault cases at the Faculty of Biomedicine, National Institute of Forensic Medicine. Analysis of nuclear STRs using the Powerplex® fusion system kit (Promega) in these samples all revealed that the DNA profiles obtained were mainly in the form of mixture samples between males and females and could not be used for individual traceability.

Table 3.7. Analysis results of 50 mixture samples using PPY23 and YPlus

	Number of samples with no locus detected	Number of samples with some loci detected	Number of samples with full loci detected
PPY23	9 (18%)	5 (10%)	36 (72%)
YPlus	12 (24%)	11 (22%)	27 (54%)
Average	21.0%	16.0%	63.0%

The results show that using PPY23 kit gives more complete Y-STR profiles than YPlus kit. In contrast, the number of samples that could detect only a few Y-STRs and not detect Y-STR profiles of the YPlus kit was higher than that of the PPY23 kit. There are several reasons for the results. First, because the number of Y-STR loci in the YPlus kit is higher than that of the PPY23 kit, theoretically the success rate of amplification of all Y-STRs would be lower. Second, most of the newly added loci in the YPlus set are large in size, making it difficult to amplify successfully, especially those with highly damaged DNA damage or samples with low DNA concentration. In addition, DNA samples derived from crime scenes often contain PCR inhibitors such as tannic acid and humic acid. In published studies on PPY23 and YPlus sets, the two kits have similar sensitivities, but the PPY23 kit shows a higher tolerance to humic acids than the YPlus kit.

Thus, further analysis of Y-STR data helps to supplement information, increase the rate of successful resolution in cases with mixed samples, significantly supporting the conclusion of the case. At Faculty of Biomedicine since 2018 we have added Y-STR data in the assessment work and obtained certain results. Based on the complete Y-STR profiles from the scene sample or victime's samples, we compared it with the sample from the suspect to consolidate the results and make a final conclusion. With Y-STR

records showing only a few loci, it is also effective in excluding those who are not the real suspects of the case.

3.4.3. Effective in increasing the ability to distinguish between individuals

Compared with the previous kits, which is the most popular Yfiler kit with 17 Y-STR markers, the PPY23 and YPlus kits added 12 more loci. The newly added loci have brought significant advantages to the two kits above. These are also the loci that were investigated for the first time in the Kinh population, Vietnam when compared with previous studies.

The addition of 2 more multiple copies of loci in the YPlus and PPY23 kits (which are loci with more than 1 version of STR in different positions on the Y chromosome) compared to the previous kits. DYF387S1a/b and DYSS385a/b proved to be loci with high genetic diversity that significantly contributed to the increased discriminating ability of this kit. In addition, among 29 Y-STR markers, 7 Y-STRs are loci with rapid mutation rate (rapidly mutating Y-STR, referred to as RM Y-STR). PPY23 kit adds 02 RM Y-STR is DYS570, DYS576; YPlus kit adds all 7 RM Y-STR.. RM Y-STR plays an important role in distinguishing between closely related males (e.g. brother - younger brother, father). - boy). To evaluate the effectiveness of the newly added loci in the PPY23 and YPlus kits, we also compared the HD, DC values with different combinations of Y-STR commonly used in criminal analysis. Combinations include:

- The minimum haplotype set (MHT - European Minimal Haplotype) includes 9 loci: DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392 and DYS393.

- The haplotype set as recommended by the Association of Scientists Working in the Field of DNA Analysis (SWGDAM): includes 9 loci of the MHT set and 2 loci: DYS438, DYS439.

- The AmpFLSTR™ Yfiler™ PCR Amplification Kit (referred to as Yfiler, ThermoFisher for short) was developed in 2004 including 11 loci as recommended by SWGDAM above and 6 new loci with high polymorphism: DYS437, DYS448, DYS456 , DYS458, Y GATA H4, DYS635.

- Set of PPY23 and set of YPlus: include 17 loci mentioned above and add new loci.

The comparison results show that increasing the number of Y-STR markers in the analysis (from 9 markers in the basic Y Haplotype set to 23-27 markers in the YPlus set) will significantly increase the number of unique haplotypes (only occurring in a single male individual in the study population), reducing the number of haplotypes occurring in 2 or more individuals. Thereby increasing the total number of observed haplotypes from 171 -173 haplotypes (Set of 9 loci) -> 193 - 196 haplotypes (Set of 17 loci) to 200 haplotypes (Set of 23 and 27 loci). Since then, the haplotype diversity and the ability to distinguish are also increased. Specifically, the DC value increased from 86% in the MHT set -> 98% in the Yfiler set to 100% in the PPY23 and YPlus sets.

Table 3.8. Comparison of the number of observed haplotypes and statistical indicators obtained from the set of Y-STR markers in haplotype sets from 200 samples data with the set of PPY23

Number of haplotypes observable	Set of haplotypes minimum (9 locus)	Set of haplotypes SWGDAM (11 locus)	Set of haplotypes Yfiler (17 locus)	Set of haplotypes PPY23 (23 locus)
Appears only in 1 individual	153	169	192	200
Appears in 2 individual	11	11	4	0
Appears in 3 individual	5	0	0	0
Appears in 4 individual	0	1	0	0
Appears in 5 individual	2	1	0	0
Appears in 6 individual	0	0	0	0
Total number of locus	9	11	17	23
Total number of observed haplotypes	171	182	196	200
HD	0.9977	0.9986	0.9958	1
DC	86%	91%	98%	100%

From the results of the study, we also boldly applied the analysis of Y-STR indicators from two kits PPY23 and Yfiler Plus in the assessment work at the Faculty of Biomedical Sciences and achieved many results. In pedigree assessment, the Y-STR indicators are used to assess in cases where there is no longer a comparable father model, to determine the blood relationship between people of the same father line as brother - brother, grandfather. /uncle/uncle – grandson. Especially for some cases of paternity testing with differences at 1 or 2 loci - suspected due to mutation or analysis error, we also applied extensive analysis of Y-STR set to ensure ensure the most accurate conclusion.

3.4.4. Application of the Y-STR markers to compare genetic distances between populations

Genetic distance is a measure of the genetic divergence between species or between populations within a species, whether the distance measures time from common ancestor or degree of differentiation.[2] Populations with many similar alleles have small genetic distances. This indicates that they are closely related and have a recent common ancestor. Because of the male-specific inheritance pattern and the haploidentical of the Y chromosome, markers on the Y chromosome are more sensitive than markers on autosomal chromosome in assessing genetic drift and the founder effect. The Y-STR is widely used to assess structure or differences in populations by testing pairwise genetic distance. In this study, the online tool AMOVA on www.yhrd.org was used to calculate genetic distance based on pairwise comparison (Rst value) and probability of association (P value) among Vietnamese populations. In the study with previously published data of 18 other populations around the world. The comparative population list includes 8 Asian

populations (2: Chinese, 3: Singaporean, 2: Thai, 1: Indian, 1: Filipino, 1: Iraqi, 1: Japanese, 1: Korean), 1 European (Belgium), 1 African (Namabian) population.

Smaller Rst and larger P-value represent close genetic distance and no significant difference between populations. The results showed that the Kinh (Vietnamese) population in the study was significantly different from the African and European populations (Namibia and Belgium) with Rst > 0.2, P < 0.001. In contrast, the Kinh and Vietnamese populations are genetically closer to the Thai and Han (Chinese) populations than to other Asian populations. Among them, the genetic distance between Kinh (Vietnamese) and Thai people (Rst = 0.0085, P = 0.0157) is the closest. The close genetic distance between Kinh, Vietnamese and Thai populations has also been mentioned in several studies. Specifically, the Vietnamese genome project published in 2019 shows that the Kinh and Thai people have a high similarity in their genomes and a close evolutionary relationship. Clustering or splitting of populations can be visualized using multidimensional scaling (MDS) analysis based on pairwise genetic distance similarity between populations in multidimensional space in AMOVA tool (Figure 3.6).

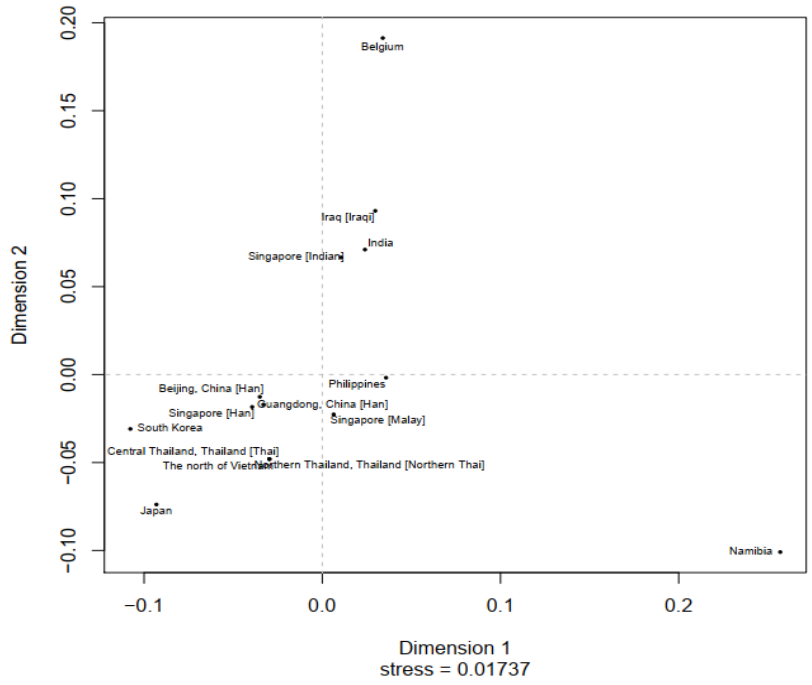

Figure 3.6: Diagram showing the genetic distance between populations in 2D space

More importantly, the genetic distance comparison between the baseline Y haplotype (MHT) set of 400 samples from this study and other Vietnamese populations published on the YHRD showed that the difference in Rst values is not significant. significantly. This suggests a minimum distance between the sample groups and shows that the ethnic populations in Vietnam are generally relatively similar, with little difference in genetic structure.

Through the above analysis, it can be said that the application potential of Y-STR markers is very large, contributing to additional data in many different types of DNA tests. The results of the thesis with a total of 29 Y-STRs belonging to 2 popular commercial kits and 10 Y-STR mini-indicators have been studied, the survey has shown that the Y-STR markers have polymorphism and high efficiency. High amplification results applied on many species are the basis for the selection and development of specific analytical indicators for Vietnamese men.

CHAPTER IV. CONCLUSIONS AND RECOMMENDATIONS

Conclusions

From the research results in the thesis, we would like to draw some conclusions as follows:

1. In the study, we investigated the characteristics of 29 Y-STR markers in the male population distributed in the North, Vietnam. Specifically:

- The frequency table reflects the characteristics in the genetic structure of the Kinh, Vietnamese male population when compared with other populations and is significant in the calculation of the paternity index.

- The survey results also show that some markers with high polymorphism > 0.7 (accounting for 55.2% of the total number of loci surveyed) are promising loci with high polymorphism, having great potential in distinguishing fish. male in the Kinh population, Vietnam.

- Survey results using two popular Y-STR kits: PowerPlex®Y23 system (Promega) and Yfiler™ Plus PCR Amplification Kit (Thermo Fisher Scientific) also show the ability to distinguish 100% of individuals in collection of 400 research samples.

3. Survey results on actual samples show that 29 Y-STR and 10 mini Y-STR indicators have high potential in many application directions:

- In paternity testing, analysis of samples of poor quality, degraded samples, samples mixed from multiple sources of DNA. Depending on the analysis needs, the appropriate type of Y-STR indicator can be selected.

- Research on population genetic structure: the results of comparing the frequency and distribution of 29 Y-STR markers in the study showed that the genetic structure is quite homogenous among ethnic groups in the Vietnamese and Vietnamese population. have the closest genetic distance to the Thai population.

Recommendations

The research team would also like to make some recommendations to improve the research in the future:

1. Conduct more survey on the allele distribution frequency of Y-STR loci with other ethnic groups common in the Vietnamese population, especially the frequency of loci with high polymorphism.

2. Conduct further identification of some new Y-STRs that are not yet available in commercial kits with potential for application in the Vietnamese population.

NEW CONTRIBUTIONS OF THE THESIS

1 - This is the first study in Vietnam to comprehensively survey 29 Y-STR markers in the Kinh male population, Vietnam with a total of 400 research samples and a larger number of Y-STR markers than previous study in Vietnam. The study has successfully established the allele frequency distribution table of 29 Y-STR markers which is the basis for calculating the rarity of a DNA profile in the population as well as the kinship index between DNA profiles. This study shows Y-STR markers have a large number of alleles, high polymorphism, and have great application potential in discriminating between male individuals such as DYS385, DYF387S1, DYS518, DYS 627, DYS458 and others. It also founds Y-STRs which have rare allele in the population.

2 - This is the first study in Vietnam to survey mini Y-STRs - which are short-sized indicators below 200 bp, highly effective in analyzing degraded samples compared to common Y-STR markers in commercial kits.

3 - The study also investigated the potential applications of Y-STR markers in the field of DNA forensic testing in Vietnam, including: paternity testing, assessment in criminal science especially in the case of mixed samples between men and women, assessment of degraded samples. Besides, the results from 400 Y-STR profiles in this study are also applied to build the genetic distance between the Vietnamese population and other population groups in the world.

LIST OF PUBLICATIONS

1. Ha Huu Hao, Le Tuan Anh, Chu Thi Thuy, Dinh Thi Lan, Hoang Thai Thanh, Nguyen Thi Phuong Lien, Pham Thi Tra (2018), "Study on polymorphism of 6 new Y - STR loci in Kinh population in Vietnam", *Journal of Practical Medicine*, volume 7 (1073), 6-9.

2. Ha Huu Hao, Nguyen Duc Nhu, Chu Thi Thuy, Dinh Thi Lan, Hoang Thai Thanh, Nguyen Thi Phuong Lien, Chu Hoang Ha, Le Van Son (2018), "Genetic characteristics of 27 locus Y-STR in Kinh population, Vietnam using Yfiler Plus PCR Amplication Kit", *Scientific report, National Conference on Biotechnology Science 2018, 564-569.*

3. Hao Huu Ha, Trang Hong Nguyen, Linh Huyen Tran, Hanh Thi Hong Nguyen, Ha Hoang & Hoang Ha Chu (2019), "Genetic characteristics of 23 Y-chromosomal STRs in the Kinh population in Northern Vietnam", *International Journal of Legal Medicine*, volume 133: 1403–1404.

4. Ha Huu Hao, Le Tuan Anh, Chu Thi Thuy, Dinh Thi Lan, Hoang Thai Thanh, Nguyen Thi Phuong Lien (2019), "Research and application of mini Y-STR in the assessment of degraded samples at the National Institute of Forensic Medicine", *Journal of Practical Medicine*, volume 11 (1128), 6-9.

5. Hao Huu Ha, Nhu Nguyen Duc, Anh Le Tuan, Thuy Thi Chu, Lan Dinh Thi, Lien Nguyen Thi Phuong, Thanh Hoang Thai (2020), A case study on sexual assault involving an individual with a large 24 Y chromosome deletion and two X chromosomes genotype. *The Asian sciences network newsletter,*Issue 10: 24-25.