*Tạp chí Công nghệ Sinh học* **16**(2): 223-229, 2018

# GENETIC POLYMORPHISM OF 23 Y-CHROMOSOME SHORT TANDEM REPEAT LOCI IN THE KINH POPULATION OF VIETNAM

**Ha Huu Hao[1, ✉], Nguyen Duc Nhu[1], Chu Hoang Ha[2], Le Van Son[2], Le Tuan Anh[1], Dinh Thi Lan[1], Chu Thi Thuy[1], Hoang Thai Thanh[1], Nguyen Thi Phuong Lien[1]**

[1]*National Institute of Forensic Medicine*
[2]*Institute of Biotechnology, Vietnam Academy of Science and Technology*

✉ To whom correspondence should be addressed. E-mail: hahaobio@gmail.com

Received: 11.8.2017
Accepted: 15.6.2018

SUMMARY

Y-chromosome microsatellites or short tandem repeats (STRs) have been proved to be ideal markers to delineate the differences between individuals in human population. Nowadays, Y-STR testing using the PowerPlex® Y23 amplification kit is considered as an extremely sensitive analysis method and has the potential to be used to perform forensic caseworks, and to explore the complexity in population substructures. However, little is known about the forensic Y-chromosome databases in the Vietnam population. In this study, 23 Y-STR loci (DYS576, DYS389I, DYS389 II, DYS448, DYS19, DYS391, DYS481, DYS549, DYS533, DYS438, DYS437, DYS570, DYS635, DYS390, DYS439, DYS392, DYS393, DYS458 DYS456, DYS643, YGATAH4, and DYS385a/b) were investigated in 120 non-related males of the Kinh population in Northern Vietnam using PowerPlex® Y23 system kit (Promega). Our results showed that allele frequencies of 23 loci in the sample population, with the calculated average gene diversity (GD) for each locus, ranged from 0.24 (DYS438) to 0.92 (DYS385a/b). In addition, a total of 120 different haplotypes were found, all of them were unique. Therefore, we found that the haplotype diversity was 1 with a discrimination capacity of 100%, which serves as an essential prerequisite for using Y-chromosomal STR with PowerPlex® Y23 System kit in forensic application in Vietnam. We also compared genetic distances between Kinh population and 10 other neighboring populations from Y-chromosome haplotype reference database (YHRD). The Kinh population is significantly different from other populations. In conclusion, it was indicated that the 23 Y-STR loci were highly genetically polymorphic in the Kinh population in Vietnam and might be of great value in forensic application.

**Keywords**: *Allele, Kinh population, PowerPlex® Y23, STR, Y-chromosome*

## INTRODUCTION

Currently, short tandem repeat (STR) marker analysis is considered as one of the most reliable methods in human identification and forensic investigation. Normally autosomal STRs are used but in some caseworks, such as in rape cases or other cases containing DNA mixtures, autosomal STRs may prove to be limited in resolving a case. In these instances, Y-STRs can be considered as an attractive alternative method. Y-STRs are STRs loci on the Y-chromosome, which are characterized by male inheritance pattern, and remain relatively unchanged from generation to generation (Kareem *et al.,* 2015). Therefore, Y-STRs are not as discriminatory as autosomal STRs, but are useful in establishing potential linkages or for excluding males in some cases. Y-STRs analysis can be used in DNA testing for the individual recognition; identification of groups of paternally related men (paternal lineages); confirmed dead body parts after disasters; study of the male line in anthropology; identifying male suspects in cases like sexual assault, murder, violence etc (Kayser *et al*., 1997; Ballantyne, Kayser, 2012).

The identification of the allele frequency distribution of STR polymorphic loci in human populations (ethnic) is essential in order to determine the reliability and objectivity of the DNA analysis methods as well as the application of the commercial kits in the DNA forensics. In Vietnam, PowerPlex®

Y23 System kit (Promega) with 23 STR loci on Y-chromosome is commonly used in paternal DNA assessments. However, the frequency distribution, relevance and polymorphisms of these loci in Vietnam population have not been studied fully and in detail. Therefore, the aim of the population study was to explore the distribution and polymorphisms of 23 short tandem repeat (STR) loci on the Y-chromosome in the Vietnam male population (Kinh ethnic) and estimate their forensic parameters using PowerPlex® Y23 System Kit.

## MATERIALS AND METHODS

### Material

This study involved a total of 120 non-related healthy males from Northern Vietnam who were of Kinh ethnicity. Participants were selected based on identification cards from individuals who had DNA testing at National Institute of Forensic Medicine.

### DNA extraction

DNA was extracted from blood, hair, nails, toe nails or buccal swab samples using QIAamp DNA micro kit (QIAgen – Germany).

### PCR amplification

PCR amplification was performed using PowerPlex® Y23 System (Promega Corporation) according to the manufacturer's recommendations. Amplification conditions were as follows: 2.5 µl PowerPlex® Y2310X Primer Pair Mix; 5 µl PowerPlex® Y23 5X Master Mix; DNA template (0.5 ng) and $H_2O$ de-ion up to 17.5 µl. The cycling conditions were as follows: 96°C/2 min, 26 cycles of [94°C/10 sec; 61°C/1 min; 72°C/30 sec]; 60°C/20 min; 4°C soak. Samples were stored at 4°C.

### Electrophoresis

Samples for electrophoresis were prepared according to the manufacturer's recommendations. PCR products were separated and detected by capillary electrophoresis on the ABI 3500 Genetic Analyzer (Applied Biosystems). Collected data were analyzed and haplotypes were obtained using GeneMapper® ID-X 1.3 software (Applied Biosystems).

### Statistical analyses

The generation of the allele frequencies and haplotype frequencies, was facilitated by using the direct gene counting method. Gene diversity (GD) was calculated as $1- \Sigma\ pi^2$, where pi is the allele frequency. Haplotype diversity (HD) was estimated by Nei's formula: $HD = n*(1- \Sigma\ pi_{th}^2)/(n-1)$ (Nei *et al.*, 1987) where n is the sample size and pi is the *i*th's haplotype frequency. The discrimination capacity (DC) was calculated according to the formula DC = h/n, where h is the number of different haplotypes in the observed population (Purps *et al.,* 2014).

AMOVA (Analysis of molecular variance) online tool from Y Chromosome Haplotype Reference Database – YHRD (www.yhrd.org) was used to calculate population pairwise genetic distances (Rst) and associated probability values (P values) between the studied population and the neighboring populations.

## RESULTS AND DISCUSSION

### Allele frequencies of 23 STR loci on the Y chromosome

Using the methods described above we have identified and calculated allele frequency of 23 Y-STR loci from 120 samples, which are representative of the male Vietnam population (Kinh ethnic) (Table 1).

In recent years, Y-STR marker analysis has been increasingly used in forensic science and population studies. However, the number of studies about Y-STR in the Vietnamese population is limited with only two previous studies including 13 and 17 Y-STR (Koji Dewaa *et al.,* 2003; Loi V L *et al.,* 2013). This study gives the first population data for 23 Y-STR loci for the Vietnam population, adding 11 Y-STR loci (DYS576, DYS448, DYS481, DYS549, DYS533, DYS570, DYS635, DYS643, DYS458, DYS456 and YGATAH4) compared with the previous study of Koji Dewaa with 119 male Vietnamese samples, adding 6 Y-STR loci (DYS576, DYS481, DYS549, DYS533, DYS570 and DYS643) compared with the study of Loi V L.

We used Y Chromosome Haplotype Reference Database – YHRD (www.yhrd.org) to compare allele frequency of 23 Y-STR loci in this study with other publications. It is an open access, annotated collection of population samples typed for Y chromosomal sequence variants around the world with the objective to generate reliable frequency estimates for Y-STR Haplotypes and Y-SNP Haplotypes to be used in the quantitative assessment of matches in forensic and kinship cases. The results showed that allele

frequency distribution in this study is similar to the general statistical data in 13 Y-STR loci (including DYS576, DYS389 I, DYS19, DYS391, DYS549, DYS438, DYS437, DYS385a / b, DYS635, DYS390, DYS439, DYS392). However, 10 loci (including DYS448, DYS389II, DYS481, DYS533, DYS570, DYS392, DYS643, DYS393, DYS458, YGATAH4) have allele frequencies different from the general statistics. The differences reflect the characteristics of Y-STR genetic polymorphism in the Vietnamese population when compared with other human populations around the world.

**Table 1.** Allele frequencies and gene diversity values of 23 Y-STR loci in the Kinh population of Vietnam.

| **DYS576** | | **DYS389 I** | | **DYS448** | | **DYS389 II** | | **DYS19** | | **DYS391** | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **GD = 0.72** | | **GD = 0.68** | | **GD = 0.63** | | **GD = 0.72** | | **GD = 0.67** | | **GD = 0.50** | |
| Allele (7) | Frequency | Allele (5) | Frequency | Allele (7) | Frequency | Allele (7) | Frequency | Allele (6) | Frequency | Allele (5) | Frequency |
| 15 | 0.008 | 11 | 0.042 | 16 | 0.017 | 26 | 0.008 | 13 | 0.050 | 6 | 0.033 |
| 16 | 0.025 | 12 | 0.333 | 17 | 0.025 | 27 | 0.042 | 14 | 0.075 | 9 | 0.008 |
| 17 | 0.233 | **13** | **0.392** | **18** | **0.550** | 28 | 0.250 | 14.3 | 0.008 | **10** | **0.633** |
| **18** | **0.400** | 14 | 0.225 | 19 | 0.233 | 29 | 0.308 | **15** | **0.400** | 11 | 0.317 |
| 19 | 0.233 | 15 | 0.008 | 20 | 0.117 | **30** | **0.342** | 16 | 0.400 | 12 | 0.008 |
| 20 | 0.092 | | | 21 | 0.050 | 31 | 0.033 | 17 | 0.067 | | |
| 21 | 0.008 | | | 22 | 0.008 | 32 | 0.017 | | | | |
| **DYS481** | | **DYS549** | | **DYS533** | | **DYS438** | | **DYS437** | | **DYS385a/b** | |
| **GD = 0.78** | | **GD = 0.64** | | **GD = 0.62** | | **GD = 0.24** | | **GD = 0.34** | | **GD = 0.92** | |
| Allele (10) | Frequency | Allele (5) | Frequency | Allele (5) | Frequency | Allele (4) | Frequency | Allele (3) | Frequency | Allele (35) | Frequency |
| 17 | 0.008 | 10 | 0.025 | **10** | **0.450** | 8 | 0.008 | **14** | **0.792** | 11,11 | 0.008 |
| 19 | 0.017 | 11 | 0.217 | 11 | 0.408 | **10** | **0.867** | 15 | 0.192 | 11,18 | 0.008 |
| 21 | 0.042 | **12** | **0.517** | 12 | 0.125 | 11 | 0.092 | 16 | 0.017 | 11,20 | 0.008 |
| 22 | 0.117 | 13 | 0.217 | 13 | 0.008 | 12 | 0.033 | | | 12,12 | 0.008 |
| **23** | **0.383** | 14 | 0.025 | 14 | 0.008 | | | | | 12,13 | 0.025 |
| 24 | 0.200 | | | | | | | | | 12,16 | 0.017 |
| 25 | 0.133 | | | | | | | | | 12,18 | 0.050 |
| 26 | 0.042 | | | | | | | | | 12,19 | 0.025 |
| 27 | 0.050 | | | | | | | | | 12,20 | 0.025 |
| 28 | 0.008 | | | | | | | | | 12,24 | 0.008 |
| | | | | | | | | | | 13,13 | 0.050 |
| **DYS570** | | **DYS635** | | **DYS390** | | **DYS439** | | **DYS392** | | 13,14 | 0.025 |
| **GD = 0.76** | | **GD = 0.73** | | **GD = 0.66** | | **GD = 0.55** | | **GD = 0.46** | | 13,16 | 0.008 |
| Allele (9) | Frequency | Allele (7) | Frequency | Allele (6) | Frequency | Allele (6) | Frequency | Allele (6) | Frequency | 13,17 | 0.017 |
| 14 | 0.008 | 19 | 0.050 | 21 | 0.017 | 8 | 0.008 | 10 | 0.017 | **13,18** | **0.225** |
| 15 | 0.017 | 20 | 0.100 | 22 | 0.050 | 10 | 0.017 | 11 | 0.067 | 13,19 | 0.058 |
| **16** | **0.383** | **21** | **0.425** | 23 | 0.117 | 11 | 0.267 | 12 | 0.033 | 13,20 | 0.025 |
| 17 | 0.250 | 22 | 0.258 | **24** | **0.475** | **12** | **0.608** | **13** | **0.717** | 13,21 | 0.017 |
| 18 | 0.125 | 23 | 0.108 | 25 | 0.308 | 13 | 0.083 | 14 | 0.150 | 14,15 | 0.008 |
| 19 | 0.100 | 24 | 0.050 | 26 | 0.033 | 14 | 0.017 | 15 | 0.017 | 14,18 | 0.042 |
| 20 | 0.092 | 25 | 0.008 | | | | | | | 14,19 | 0.050 |
| 21 | 0.017 | | | | | | | | | 14,20 | 0.033 |
| 22 | 0.008 | | | | | | | | | 14,21 | 0.025 |
| **DYS643** | | **DYS393** | | **DYS458** | | **DYS456** | | **GATA H4** | | 14,22 | 0.008 |
| **GD = 0.73** | | **GD = 0.64** | | **GD = 0.79** | | **GD = 0.60** | | **GD = 0.61** | | 15,15 | 0.008 |
| Allele (8) | Frequency | Allele (6) | Frequency | Allele (8) | Frequency | Allele (7) | Frequency | Allele (4) | Frequency | 15,17 | 0.025 |
| 6 | 0.008 | 10 | 0.008 | 14 | 0.008 | 13 | 0.050 | 10 | 0.142 | 15,18 | 0.033 |
| 8 | 0.025 | 11 | 0.008 | 15 | 0.167 | 14 | 0.125 | **11** | **0.542** | 15,19 | 0.058 |
| 9 | 0.058 | 12 | 0.283 | 16 | 0.125 | **15** | **0.592** | 12 | 0.275 | 15,20 | 0.025 |
| 10 | 0.125 | 13 | 0.183 | 17 | 0.200 | 16 | 0.167 | 13 | 0.042 | 15,21 | 0.025 |
| 11 | 0.208 | **14** | **0.500** | **18** | **0.342** | 17 | 0.050 | | | 16,17 | 0.017 |
| **12** | **0.442** | 15 | 0.017 | 19 | 0.117 | 18 | 0.008 | | | 16,19 | 0.008 |
| 13 | 0.125 | | | 20 | 0.033 | 19 | 0.008 | | | 16,20 | 0.008 |
| 14 | 0.008 | | | 21 | 0.008 | | | | | 16,21 | 0.008 |
| | | | | | | | | | | 17,21 | 0.008 |

Note: The table shows allele frequencies for each investigated locus except for DYS385a/b, for which genotype frequencies were calculated for the combination of the two alleles GD – gene diversity. Major allele frequencies per locus are in bold.

**Gene and Haplotype diversity**

In this study, we detected 165 alleles at the 23 Y-STR loci in the Kinh population of Vietnam. Apart from DYS385a/b with 35 combinations of the two alleles, the most polymorphic locus was DYS481 with 10 alleles. The least polymorphic loci were DYS437 with 3 alleles, GATA-H4 and DYS438 with 4 alleles (Table 1). This data was compared with Purps's study, in which the data was collected from 129 populations of 54 countries and showed that the most polymorphic loci were DYS385a/b and DYS481. However, according to the Purps's statistics, the least polymorphic loci were DYS391 and DYS393 as opposed to this study (Purps *et al*., 2014).

Table 1 also lists the Y chromosome gene diversity values. For isolated microsatellite loci gene diversity ranged between 0.3137 (DYS437) and 0.8012 (DYS385 a/b when approached as genotype). The initial analysis of GD values indicated that the highest GD was detected at DYS385a/b loci with a value of 0.92 and the lowest GD at DYS438 locus with a value of 0.24, which is consistent with the polymorphism findings presented above in this study. This result was in concordance with the previously published data for the Vietnamese population provided by Dewaa *et al.* (2003) with GD ranging from 0.33 (DYS438) to 0.95 (DYS385a/b). However, when compared with other populations, there is a difference in the frequency of each alleles as well as the polymorphism of each locus. For example, according to Dogan's study in the Turkish population, the most polymorphic locus was DYS458 with 12 alleles (GD = 0.81) and the least polymorphic locus was DYS391 with 3 alleles (GD = 0.47) (Dogan *et al.*, 2014).

We found a total of 120 haplotypes which demonstrates that all 120 samples had unique haplotypes. Accordingly, the haplotype diversity in the studied population was 1,000 with a discrimination capacity of 1. It indicates the ability of the PowerPlex® kit Y23 System to discriminate among male individuals.

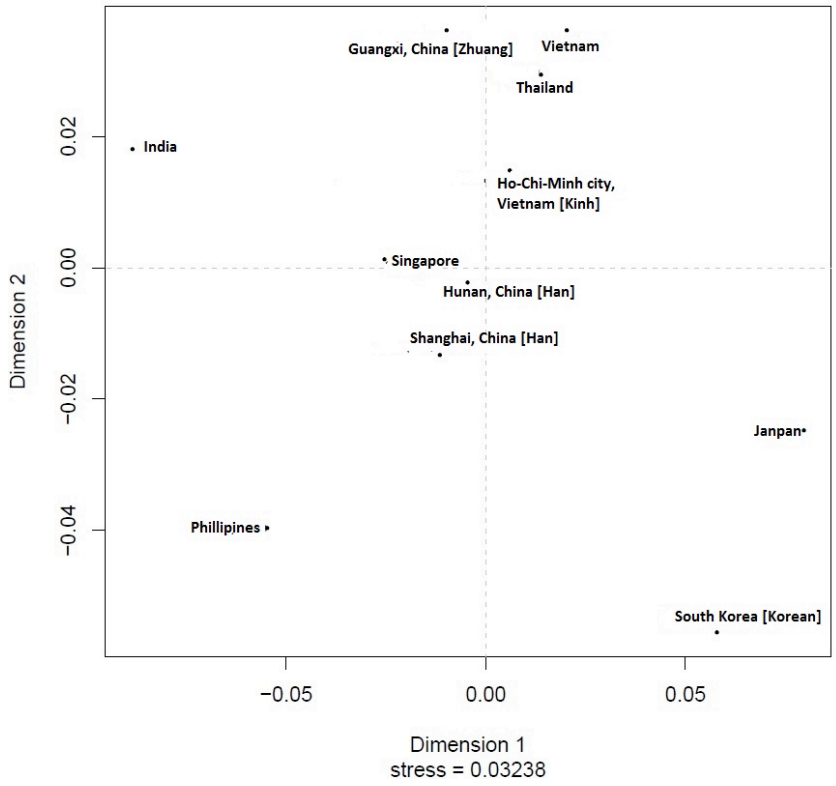


**Figure 1**. Multi-dimensional scaling (MDS) plot based on population pairwise Rst values between compared populations.

**Comparative analysis of genetic distance among Vietnamese and neighboring populations**

Genetic distance is the term used to describe the number of differences or mutations between two sets of Y-chromosome DNA or mitochondrial DNA test results. Populations with many similar alleles have small genetic distances. This indicates that they are closely related and have a recent common ancestor. Genetic distance is useful for reconstructing the history of populations (Nei *et al.,* 1987).

Based on pairwise Rst comparisons and AMOVA tool, present haplotype data of Vietnamese were compared with 10 other previously published populations which include: Hunan - China (Jiang *et al.,* 2017), Guangxi Zhuang - China (Luo *et al.,* 2015), Shanghai - China (Li *et al*., 2016), South Korea (Kim *et al*., 2008), Singapore (Yong *et al*., 2006), Ho Chi Minh city in Vietnam (Kinh ethnic), India (Yadav *et al.,* 2011), Japan (Mizuno *et al*., 2008), Phillipines (Miranda *et al*., 2001) and Thaland (Siriboonpiputtana *et al.*, 2010). The smaller the pairwise Rst values and the bigger the associated p-values, the closer it is between pairs of populations. The result was listed in Table 2. It was found that, the Vietnamese population was significantly different from those of India (p = 0.0000), Japan (p = 0.000), Philippines (p = 0.0000), Singapore (p = 0.0000) and South Korea (p = 0.0000). However, comparison of the Y-STR data suggests that there were no significant differences between Vietnamese population in this study and Guangxi - Zhuang (China), Thailand (p = 0.0301, Rst = 0.0060), Ho Chi Minh City (Kinh ethnic) (p = 0.0364, Rst = 0.0161). This comparison is also similar to previous studies and suggests that there is a close genetic distance between the Vietnamese population and Guangxi Zhuang - China population (Luo *et al.*, 2015) and Thailand population (Miranda-Barros *et al.,* 2016). The MDS plots stated visualize the genetic variation between the studied populations (Fig. 1).

**Table 2.** Analysis of molecular variance pairwise distances based on Rst values between Vietnam population from the present study and selected populations.

| Population | Vietnam | Hunan, China [Han] | Shanghai, China [Han] | Guangxi, China [Zhuang] | South Korea [Korean] | Ho-Chi-Minh City, Vietnam [Kinh] | India | Japan | Philippines | Singapore | Thailand |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Vietnam** | - | 0.0000 | 0.0000 | 0.0030 | 0.0000 | 0.0364 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0301 |
| **Hunan, China [Han]** | 0.0605 | - | 0.0059 | 0.0000 | 0.0000 | 0.0786 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| **Shanghai, China [Han]** | 0.0850 | 0.0058 | - | 0.0000 | 0.0000 | 0.0073 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| **Guangxi, China [Zhuang]** | 0.0138 | 0.0441 | 0.0594 | - | 0.0000 | 0.0159 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0016 |
| **South Korea [Korean]** | 0.1068 | 0.0726 | 0.0753 | 0.1160 | - | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| **Ho-Chi-Minh City, Vietnam [Kinh]** | 0.0161 | 0.0098 | 0.0249 | 0.0203 | 0.0709 | - | 0.0000 | 0.0000 | 0.0000 | 0.0047 | 0.1016 |
| **India** | 0.1320 | 0.0923 | 0.0730 | 0.1029 | 0.1677 | 0.0922 | - | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| **Japan** | 0.0970 | 0.0805 | 0.0928 | 0.1035 | 0.0405 | 0.0727 | 0.1568 | - | 0.0000 | 0.0000 | 0.0000 |
| **Philippines** | 0.1349 | 0.0532 | 0.0458 | 0.0932 | 0.1496 | 0.0767 | 0.1012 | 0.1457 | - | 0.0000 | 0.0000 |
| **Singapore** | 0.0687 | 0.0171 | 0.0154 | 0.0419 | 0.0968 | 0.0231 | 0.0410 | 0.0975 | 0.0369 | - | 0.0000 |
| **Thailand** | 0.0060 | 0.0403 | 0.0644 | 0.0080 | 0.1039 | 0.0078 | 0.1181 | 0.0935 | 0.1077 | 0.0471 | - |

Note: P values are shown above the diagonal and Rst values below it. Compared population including China, South Korea, Ho Chi Minh city (Vietnam), India, Japan, Phillipines, Singapore and Thailand.

## CONCLUSION

In conclusion, this study revealed a high genetic diversity among the 120 male participants from the Kinh population, with a haplotype diversity of 1 accompanied with a discrimination capacity of 1. Consequently, it can be suggested that the 23 Y-STR loci present in the Powerplex ® Y23 System kit are ideal for casework analysis in Vietnam; since these loci have demonstrated the capability of differentiating between male individuals of different paternal lineages within this population. In addition, the statistical results and forensic parameters generated in this study will no doubt drastically improve the reliability of statistical assessments of allele and haplotype frequencies during routine casework analysis.

Furthermore, the outcome of this study represented great progression in the field of Y chromosome-related testing of males in Kinh population, Vietnam.

**Acknowledgments**: *This work was supported by National Institute of Forensic Medicine (NIFM) funding. We would like to thank Bio-medicine Department staff, Dr. Nguyen Duc Nhu, NIFM and Prof. Chu Hoang Ha, Institute of Biotechnology, Vietnam Academy of Science and Technology for providing technical assistance.*

## REFERENCES

Ballantyne KN, Kayser M (2012) Additional Y-STRs in Forensics: Why, Which, and When. *Forensic Sci Rev* 24(1): 63–78.

Dewaa K, Rand S, Hohoff C, Brinkmann B (2003) 13 Y-chromosomal STRs in a Vietnamese population. *Forensic Sci Int Genet 9* (1239): 315–317.

Dogan S, Primorac D, Marjanović D (2014) Genetic analysis of haplotype data for 23 Y-chromosome short tandem repeat loci in the Turkish population recently settled in Sarajevo, Bosnia and Herzegovina. *Croat Med J* 55(5): 530–536.

Jiang W, Gong Z, Rong H, Guan H, Zhang T, Zhao Y, Fu X, Zha L, Jin C, Ding Y (2017) Population genetics of 26 Y-STR loci for the Han ethnic in Hunan province, China. *Int J Legal Med* 131(1): 115–117.

Kareem MA, Hussein AO, Hameed IH (2015) Y-Chromosome short typing technology, locus information and allele frequency in different population. A review. *Afr J Biotechnol* 14(27): 2175–2178.

Kayser M, Caglia A, Corach D et al. (1997) Evaluation of Y-chromosomal STRs: a multicenter study. *Int J Legal Med* 110: 125-133, 141–149.

Kim SH, Kim NY, Hong SB, Cho NS, Kim JJ, Han MS, Kim W (2008) Genetic polymorphisms of 16 Y chromosomal STR loci in Korean population. *Forensic SciInt Genet* 2(2): 9–10.

Li L, Yu G, Li S, Jin L (2016) Genetic analysis of 17 Y-STR loci from 1019 individuals of six Han populations in East China. *Forensic SciInt Genet* 20: 101–102.

Loi V L (2013) Calculate the frequency of the 17 Y-STR haplotype for assessment at the Institute of Criminal Science. *Master thesis in Biological Sciences, Insitute of ecology and biological resources, Thai Nguyen University, Viet Nam*.

Luo H, Song F, Zhang L, Hou Y (2015) Genetic polymorphism of 23 Y-STR loci in the Zhuang minority population in Guangxi of China. *Int J Legal Med* 129(4): 737–738.

Miranda-Barros F, Romanini C, Peréz LA, Nhu ND, Phan TD, Carvalho EF, Vullo C, Gusmão L (2016) Y Chromosome STR haplotypes in different ethnic groups of Vietnam *Forensic Sci Int Genet* 22: e18–20.

Miranda JJ, Benecke M, Hidding M, Schmitt C (2001) Y-chromosomal short tandem repeat haplotypes at the loci DYS393, DYS19, DYS392, and DYS385-I/II, DYS390, DYS389-I/II, and DYS391 in a Filipino population sample. *J Forensic Sci* 46(5): 1250–1253.

Mizuno N, Nakahara H, Sekiguchi K, Yoshida K, Nakano M, Kasai K (2008) 16 Y chromosomal STR haplotypes in Japanese. *Forensic SciInt Genet* 174(1): 71–76.

Nei M (1987) Molecular Evolutionary Genetics. *New York: Columbia University Press*.

Purps J (2014) A global analysis of Y-chromosomal haplotype diversity for 23 STR loci. *Forensic Sci Int Genet* 12: 12–23.

Siriboonpiputtana T, Jomsawat U, Rinthachai T, Thanakitgosate J, Shotivaranon J, Limsuwanachot N, Polyorat P, Rerkamnuaychoke B (2010) Y-chromosomal STR haplotypes in Central Thai population. *Forensic Sci Int Genet* 4(3): e71–72.

Yadav B, Raina A, Dogra TD (2011) Haplotype diversity of 17 Y-chromosomal STRs in Saraswat Brahmin Community of North India. *Forensic Sci Int Genet* 5(3): e63–70.

Yong RY, Lee LK, Yap EP (2006) Y-chromosome STR haplotype diversity in three ethnic populations in Singapore. *Forensic Sci Int* 159(2-3): 244–257.

*Tạp chí Công nghệ Sinh học* **16**(2): 223-229, 2018

# ĐA HÌNH DI TRUYỀN CỦA 23 LOCUS STR TRÊN NHIỄM SẮC THỂ Y TRONG QUẦN THỂ NGƯỜI DÂN TỘC KINH TẠI VIỆT NAM

**Hà Hữu Hảo[1], Nguyễn Đức Nhự[1], Chu Hoàng Hà[2], Lê Văn Sơn[2], Lê Tuấn Anh[1], Đinh Thị Lan[1], Chu Thị Thủy[1], Hoàng Thái Thanh[1], Nguyễn Thị Phương Liên[1]**

[1] *Viện Pháp y Quốc gia*
[2]*Viện Công nghệ sinh học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam*

TÓM TẮT

Các microsatellite hay còn gọi là các đoạn lặp lại ngắn trên nhiễm sắc thể Y (Y-STR) được xem là chỉ thị (marker) lý tưởng cho việc phân biệt các cá thể trong quần thể người. Hiện nay, việc phân tích Y-STR sử dụng bộ kit PowerPlex® Y23 được coi là phương pháp có độ nhạy cao được ứng dụng trong các vụ án hình sự và khám phá sự phức tạp trong cấu trúc di truyền của quần thể. Tuy nhiên, tại Việt Nam cơ sở dữ liệu về các locus trên nhiễm sắc thể Y còn rất ít. Trong nghiên cứu này, 23 locus Y-STR (bao gồm DYS576, DYS389I, DYS389 II, DYS448, DYS19, DYS391, DYS481, DYS549, DYS533, DYS438, DYS437, DYS570, DYS635, DYS390, DYS439, DYS392, DYS393, DYS458, DYS456, DYS643, YGATAH4 và DYS385a/b) từ 120 mẫu nam giới không có quan hệ huyết thống thuộc quần thể người Kinh tại miền Bắc Việt Nam được tiến hành khảo sát sử dụng bộ kit PowerPlex® Y23 System (Promega). Kết quả chúng tôi đã tính toán được tần suất phân bố các alen trên 23 locus Y-STR với độ đa dạng gen (GD) dao động trong khoảng từ 0.24 (DYS438) đến 0.92 (DYS385a/b). Thống kê cho thấy có 120 haplotype khác nhau trong đó tất cả 120 haplotype đều là duy nhất ở từng người, không có haplotype giống nhau giữa 2 người bởi vậy độ đa dạng haplotype (haplotype diversity - HD) chung trong quần thể nghiên cứu là 1 với khả năng phân biệt (discriminatory capacity - DC) là 100% cho thấy tiềm năng cao của bộ kit PowerPlex® Y23 System trong các ứng dụng hình sự tại Việt Nam. Từ dữ liệu phân tích chúng tôi cũng đã so sánh được khoảng cách di truyền giữa quần thể người Việt Nam trong nghiên cứu với 10 quần thể người ở các nước lân cận từ dữ liệu YHRD. Kết quả cho thấy quần thể người Kinh có khác biệt ý nghĩa so với các quần thể người khác. Tóm lại nghiên cứu này đã chỉ ra độ đa hình di truyền của 23 locus trên NST Y với giá trị cao để ứng dụng trong phân tích hình sự.

**Từ khóa**: *Allele, Nhiễm sắc thể Y, PowerPlex® Y23, Quần thể người Kinh, STR.*