Indian Journal of Forensic Medicine and Toxicology
Volume 16 No. 3, July-September 2022
DOI: http://dx.doi.org/xxxxx/ijfmt.xxxx.16322.62

Allele Frequencies and Forensic Statistical Parameters for 21 Autosomal Short Tandem Repeats (STRs) loci in Northern Thai Population

Supakit Khacha-ananda[1,2*], Yutti Amornlertwatana[1], Phatcharin Mahawong[3], Nitaya Khamkomkai[4]

[1]Instructor, Department of Forensic Medicine, Faculty of Medicine, Chiang Mai University, Chiang Mai, 50200, Thailand, [2]Instructor, Research Center in Bioresources for Agriculture, Industry and Medicine, Chiang Mai University, Chiang Mai, Thailand, [3]Medical Technologist, Department of Forensic Medicine, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand, [4]Scientist, Department of Forensic Medicine, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand.

How to cite this article: Supakit Khacha-ananda, Yutti Amornlertwatana, Phatcharin Mahawong et al. Allele Frequencies and Forensic Statistical Parameters for 21 Autosomal Short Tandem Repeats (STRs) loci in Northern Thai Population. Indian Journal of Forensic Medicine and Toxicology 2022;16(3):354-359.

Abstract

Background: Although the genetic information of STRs has been established over the entire world, the different characteristics between populations influenced the STRs information. FBI's guideline recommended to expand additional number of STRs loci to increase power of discrimination and exclusion. Unfortunately, the current data of STRs in northern Thai population have not fully covered additional loci. The aim of this study is to obtain accurate allele frequencies of STRs among northern Thai population.

Methods: The genetic profiles ofunrelated individuals were characterized by Investigator 24 plex Go kit. Allele frequencies and forensic statistical parameters were calculated within GenoProof®3and Arlequin ver 3.5.2.2.

Results: The SE33 and TH01 loci represented largest and lowest number of different alleles, respectively. There was no significant deviation from the Hardy–Weinberg equilibrium (HWE) after Bonferroni correction in all loci (p=0.002). The SE33 locus showed the greatest PD and PE, whereas the TPOX represented the lowest PD and PE. The studied population (northern Thai) appear to be most closely related to previously reported populations containing of Chinese, Japanese, and Vietnamese based on allele frequency.

Conclusion: Our results recommended that current autosomal STRs data extended the application of STRs typing in parentage analyses and human identification among the localized population in northern region of Thailand due to highly informative polymorphic data.

Keywords: short tandem repeats; northern Thai; allele frequencies; forensic genetic.

Introduction

Short tandem repeats (STRs), also known as microsatellites, is a unique region on DNA strand.[1] This region shows the high mutation rates, so it contributes to frequent mutation and great polymorphism more than other regions in DNA.[2]

Corresponding Author: Supakit Khacha-Ananda, Instructor, Department of Forensic Medicine, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand.

Email: supakit.kh@cmu.ac.th

Due to high variable polymorphism and small DNA fragment in STRs region, STRs typing has been applied to provide useful information for human identity testing including DNA data basing, forensic casework, missing persons/mass disaster victim identification, and parentage testing.[3] In 1997, the Federal Bureau of Investigation (FBI) announced the selection of 13 loci on STR regions containing of CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, FGA, TH01, TPOX, and VWA. The latest recommendations regarding the NIST highlighted the need for increased number of STR loci. Therefore, STR typing were extended to seven new additional loci containing D1S1656, D2S441, D2S1338, D10S1248, D12S391, D19S433, and D22S1045 in order to decrease the number of adventitious matches, increase international compatibility, and the power of discrimination for criminal and missing person cases.[4]

Thailand is geographically located on mainland Southeast Asia. The regions of Thailand can be divided into four main regions including north, northeast, center, and southaccording to topography [5].Because of the difference in environment and geographical features, Thai population is diverse in ethnicity, race, and languages including the genetic divergence. Moreover, the people in each region of Thailand tend to have specific characteristics and appearancesbecause of different demography, histories, and genetic structures among major Thai populations.[6,7] Therefore, it was probable that this phenomenon was probable to influence the variation in genetic pattern of autosomal STRincluding the allele frequencies of STRs in individual population. It is important to understand the genetic structure in individual population for providing the database and scientifically reliable results for forensic genetic purposes.

The database of allele frequencies worldwide has been provided by many research studies. They demonstrated that the allele frequency pattern was different between populations, which some alleles were predominate or absent in each populations.[8] Currently available official allele frequencies for northern Thai populationwere latest published in 2006 with three STR lociobtained from only 203 unrelated northern Thai.[9] There was also an official report in 2006 regarding 15 STR lociobtained from only 210 unrelated Thai individuals.[10] Nevertheless, the current data of autosomal STR in northern Thai population have not fully covered the seven new additional autosomal STR according to the recommendation of FBI CODIS Core Loci Working Group's progress. Therefore, the goal of this study isto obtain more accurate allele frequencies of 21 autosomal STRs loci among localized people who lived in northern Thailandin order toincrease the amount of genetic data available and update the specific northern Thai frequency databases for all autosomal STR loci.

Material and Methods

Ethics approval and consent to participate:

Written informed consent was obtained from all subjects in this study. This study was carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) and approved by the Ethics Committee of the Faculty of Medicine, Chiang Mai University, Thailand (Study code: No.012/2019).

Sample collection

Biological buccal swab samples were collected from 280 unrelated individuals (126 male and 154 female) living in the provinces of northern region of Thailandand have been declared ethnicity as Thai. The informed consent forms were signed by all participants.

STR typing

The genomic DNA was extracted from buccal swab usingChelex 100 method.[11] Genotype at 21 autosomal STRs loci (CSF1PO, D1S1656, D2S441, D2S1338, D3S1358, D5S818, D7S820, D8S1179, D10S1248, D12S391, D13S317, D16S539, D18S51, D19S433, D21S11, D22S1045, FGA, SE33, TH01, TPOX, and vWA) and amelogenin were obtained using Investigator 24plex Go kit (QIAGEN, Germany) according to the manufacturer instructions. The DNA amplification was conducted on 9600 Thermal Cyclers (Applied Biosystems, CA, USA). The amplification products were separated on the ABI Prism 3130 Genetic Analyzer (Applied Biosystems,CA, USA). GeneMapper™ ID-X Software v1.6 (Thermo Fisher Scientific, MS, USA) was used for raw data analysis. The positive and negative control DNA included in the kit was also analyzed. The quality control was performed for day-to-day quality control, laboratory internal controls, kit controls, and genotypes for overlapping loci between kits. Moreover, the quality control for

STR typing was regularly annual participated in proficiency testing program provided by Forensic DNA Network of Thailand organization.

Statistical analysis

For all autosomal 21STRs loci of Investigator 24plex Go kit, the allele frequencies, polymorphic information content (PIC), power of discrimination (PD), power of exclusion (PE), paternity index (PI), observed heterozygosity (Ho), expected heterozygosity (He), and Hardy-Weinberg equilibrium (HWE)were calculated withinGenoProof®3 (Qualitype GmbH, Germany) and Arlequin ver 3.5.2.2.[12] The neighbor-joining dendrogram was constructed based on $D_A$genetic distance results using POPTREE software.[13] The tree was constructed with allele frequency data based on pairwise $D_A$distance of 15 STR loci (CSF1P0, D2S1338, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D19S433, D21S11, FGA, TH01, TPOX, and vWA) for all populations which were previously reportedincluding northeast Thai[14], central Thai[14] southern Thai[14] Laotian[15], Vietnamese[16], Malaysian[17], Han Chinese [18], Yunnan Zhuang Chinese[19], Yunnan Dai Chinese[19], Yunnan Hani Chinese[19], Japanese[20] and Hongkong Chinese[21]

Results

The largest number of different alleles was found on locus SE33, followed by FGA and D21S11, whereas the lowest number of different alleles was also observed onlocus TH01.The most frequent allele among all allelic variantswas the allele 8 at TPOX with allele frequency 0.5661.There weredistinct microvariant alleles found in eight loci of 21 STRs markers, with one being located at TH01, three at D2S441 and D7S820, four at D1S1656, six at D19S433, seven at D21S11, seven at FGA, and nineteen at SE33.

The forensic statistical parameters for allautosomal STRs loci are provided in Table 1.The Horanged between 0.6179 (TPOX) and 0.9500 (SE33), theHe ranged from 0.5945 (TPOX) to 0.9431 (SE33). All autosomal STRs markers were classified as a highly information with PIC value more than 0.6 except for locus TPOX. The potential power of a genetic marker to differentiate between any two people chosen at random was expressed as PD. The high discriminating power can be used to discriminate between members of the population. The PD ranged from 0.5613 (TPOX) to 0.9929 (SE33) with the value of 1 for the combined PD. The PE was investigated to evaluate the loci or system efficiency for exclusion of a non-related individual in paternity testing. The SE33 locus showed the greatest PE in studied population, whereas TPOX showed the lowest PE value.Our study revealed that there was no significant deviation from HWE for autosomal STR loci (p-value greater than 0.05) except locus D13S317, D16S539, D18S51, and D19S433. However, no significant deviation from HWE was found after Bonferroni correction (p=0.002). As a result, it was established that the 21 autosomal STRs loci were in equilibrium.

Regarding genetic relatedness, our informative study was compared with previous data. The Neighbor-joining tree based on D_A distance represented four close cluster including northeast Thai and Laotian, Han Chinese and Japan, Yunnan Zhuang Chinese and Yunnan Hani Chinese, southern Thai and Malaysian. The D_A distance of the studied population showed that the studied population (northern Thai) and other populations consisting of Chinese, Japanese, and Vietnamese had the genetic relatedness.

Discussion

The discovery of genetic profiles in specific population is expected to expand in a future since the globalization affected the increase of population number and rapid migration between countries.[8] To increase the power of discrimination for criminal and missing person cases as well as decrease the number of adventitious matches, the FBI launched the seven additional core loci.[4] The allele frequency of our studied population in northern region of Thailand was slightly different from otherreported population in Thailand. A possible reason was that Thailand has a diversity of ethnicity, culture, and linguistics due to the migration and settlement of different ancestor populations. The northern Thai population was recently originated from the Khon Mueang population speaking a Tai-Kadai language which had migrated from the population in southern China, however Mon group who drifted from Myanmar was a major group of Central Thai.[23, 24] Moreover, the other reason to influence the variation in autosomal STR is to mutation process especially strand-slippage replication.[25]

According to seven additional analyzed STR loci here, we observed microvariant alleles in D1S1656, D2S441, and D19S433 loci. A microvariant was occurred during one of the repeat units contains only one, two or three bases of the repeat motif.[26] The microvariants in D1S1656 was found to be 15.3, 16.3,

17.3, and 18.3 as well as D2S441 locus with 11.3 which were not observed and reported in STR database of National Institute of Standards and Technology (NIST), U.S. department of Commerce.

Table 1: Forensic descriptive statistic of 21 autosomal STR loci among unrelated northern Thai population (n = 280)

Parameters	Polymorphic information content (PIC)	Power of exclusion (PE)	Paternity index (PI)	Power of discrimination (PD)	Combined PD	Combined PE	Observed heterozygosity (Ho)	Expected heterozygosity (He)	p-value
CSF1PO	0.6786	0.4667	1.8129	0.8174	1	1	0.71786	0.7255	0.22323
D10S1248	0.7318	0.542	2.1603	0.876	1	1	0.775	0.76992	0.43333
D12S391	0.8294	0.6898	3.2766	0.9481	1	1	0.80714	0.84892	0.29462
D13S317	0.7627	0.5872	2.4223	0.903	1	1	0.78571	0.795	0.04332
D16S539	0.7431	0.5581	2.2478	0.886	1	1	0.78929	0.77895	0.0075
D18S51	0.8428	0.712	3.5371	0.9559	1	1	0.82143	0.86018	0.00313
D19S433	0.8098	0.6564	2.9459	0.934	1	1	0.77143	0.83176	0.02393
D1S1656	0.8158	0.6631	3.0074	0.935	1	1	0.79643	0.83524	0.28387
D21S11	0.8151	0.6647	3.0226	0.9373	1	1	0.825	0.83607	0.96152
D22S1045	0.7205	0.5251	2.0736	0.864	1	1	0.73571	0.76024	0.76074
D2S1338	0.852	0.7269	3.7353	0.9604	1	1	0.85	0.86769	0.7459
D2S441	0.7685	0.5959	2.4793	0.9078	1	1	0.80357	0.79976	0.09985
D3S1358	0.6826	0.476	1.8512	0.8284	1	1	0.75	0.73122	0.4548
D5S818	0.7581	0.5804	2.3795	0.8994	1	1	0.775	0.79129	0.21171
D7S820	0.7476	0.5604	2.261	0.8845	1	1	0.77143	0.78025	0.96815
D8S1179	0.8342	0.6994	3.385	0.9527	1	1	0.85357	0.85381	0.34405
FGA	0.8783	0.7719	4.4818	0.9728	1	1	0.86786	0.89003	0.08972
SE33	0.9382	0.8805	8.5319	0.9929	1	1	0.95	0.94308	0.81684
THO1	0.6324	0.4047	1.5848	0.7535	1	1	0.65357	0.70516	0.18592
TPOX	0.5362	0.283	1.2297	0.5613	1	1	0.61786	0.59447	0.85243
vWA	0.7553	0.5753	2.3486	0.8958	1	1	0.83214	0.78851	0.42017

The forensic statistical parameters showed that all autosomal STRs markers were classified as a highly information except for locus TPOX. Consistently with previous report, TPOX loci was also reported to have a lower PIC in Thai population.[10] Iraqi Kurds population in Kurdistan Region-Iraq [27], Chinese Han population[28], and population from Kingdom of Bahrain[29]. Moreover, we found that the greatest and lowest of PE were observed in locus SE33 and TPOX, respectively which was consistent with previous report in population of Thailiving in southern border provinces of Thailand (Pattani, Yala, Narathiwat provinces).[30] Obtained result is according to the previously observed from other publication in in Ecuadorian and Bahraini population. It recommended that SE33 was the best STRs loci for human identification with the highest value of PIC, PD, and PE.[29,31]

The D_A distances showed that the studied population (northern Thai) and other populations consisting of Chinese, Japanese, and Vietnamese had the genetic relatedness. Our study clearly showed that the large genetic distance of northern Thai and southern Thai population, possible due to the difference in culture, linguist, and religion since the major population in southern Thai is Thai-Malay Muslims (MUS) which was descended from population in India and distributed over Malay Peninsula according to historical evidence.[32] This assumption was confirmed by the finding of close genetic relationship between southern Thai and Malaysian population in our study. Moreover, the southern Thai people showed the close genetic structure with Indian population since the migration of the Austroasiatic linguistic population into Southeast Asian region.[33] However, some studies found the close relationship between Chinese and northern Thai because the gene flow from southern China to the northern Thai people over the past one thousand years.[18,33]

Conclusion

Our result recommended that 21 autosomal STRs loci were suitable for paternity testing and human identification among the localized population in northern region of Thailand due to high genetic polymorphism and information.

Conflict of interest: The authors declare that they have no conflicts of interest.

Source of Funding: This work was supported by Faculty of Medicine, Chiang Mai University, grant no. 102/2562.

References

1. Durbin RM, Altshuler D, Durbin RM, et al. A map of human genome variation from population-scale sequencing. *Nature* 2010; 467: 1061-1073. DOI: 10.1038/nature09534.
2. Gemayel R, Vinces MD, Legendre M, et al. Variable tandem repeats accelerate evolution of coding and regulatory sequences. *Annu Rev Genet* 2010; 44: 445-477. DOI: 10.1146/annurev-genet-072610-155046.
3. Fan H and Chu J. A brief review of short tandem repeat mutation. *GPB* 2007; 5: 7-14. DOI: 10.1016/S1672-0229(07)60009-6.
4. Hares DR. Selection and implementation of expanded CODIS core loci in the United States. *Forensic Science International: Genetics* 2015; 17: 33-34. DOI: https://doi.org/10.1016/j.fsigen.2015.03.006.
5. Cosslett TL and Cosslett PD. The setting of mainland southeast Asian countries: Cambodia, Laos, Thailand, and Vietnam. In: Cosslett TL and Cosslett PD (eds) *Sustainable Development of Rice and Water Resources in Mainland Southeast Asia and Mekong River Basin*. Singapore: Springer Singapore, 2018.
6. Chaikhambung J and Tuamsuk K. Knowledge classification on ethnic groups in Thailand. *CCQ* 2017; 55: 89-104. DOI: 10.1080/01639374.2016.1271074.
7. Silva NM, Pereira L, Poloni ES, et al. Human Neutral Genetic Variation and Forensic STR Data. *PLoS One* 2012; 7: 1-11. DOI: 10.1371/journal.pone.0049666.
8. Delgado E and Neyra CD. Allele frequencies of 21 autosomal STR markers in a mixed race Peruvian population applied to forensic practice. *Span J Leg Med* 2019; 45: 92-97. DOI: https://doi.org/10.1016/j.remle.2018.09.001.
9. Bhoopat T, Sriduangkaew S and Steger HF. STR loci D10S2325, D16S539 and D19S253: Northern Thai population data. *Leg Med (Tokyo)* 2006; 8: 306-307. 2006/10/13. DOI: 10.1016/j.legalmed.2006.07.002.
10. Rerkamnuaychoke B, Rinthachai T, Shotivaranon J, et al. Thai population data on 15 tetrameric STR loci-D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX, D18S51, D5S818 and FGA. *Forensic Sci Int* 2006; 158: 234-237. 2005/07/09. DOI: 10.1016/j.forsciint.2005.05.020.
11. Walsh PS, Metzger DA and Higuchi R. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* 1991; 10: 506-513. 1991/04/01.
12. Excoffier L and Lischer HE. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 2010; 10: 564-567. 2011/05/14. DOI: 10.1111/j.1755-0998.2010.02847.x.

13. Takezaki N, Nei M and Tamura K. POPTREE2: Software for constructing population trees from allele frequency data and computing other population statistics with Windows interface. *Mol Biol Evol* 2010; 27: 747-752. 2009/12/22. DOI: 10.1093/molbev/msp312.

14. Shotivaranon J, Chirachariyavej T, Leetrakool N, et al. DNA database of populations from different parts in the Kingdom of Thailand. *Forensic Sci Int Genet* 2009; 4: e37-38. 2009/12/02. DOI: 10.1016/j.fsigen.2009.02.009.

15. Srithawong S, Muisuk K, Srikummool M, et al. Genetic structure of the ethnic Lao groups from mainland Southeast Asia revealed by forensic microsatellites. *Ann Hum Genet* 2020; 84: 357-369. DOI: https://doi.org/10.1111/ahg.12379.

16. Tran HL, Ta TAM, Nguyen NN, et al. Population genetic data of 22 autosomal STR loci for the Mong people in Vietnam. *Leg Med* 2021; 48: 1-5. DOI: https://doi.org/10.1016/j.legalmed.2020.101825.

17. Rashid MNA, Mahat NA, Khan HO, et al. Population data of 21 autosomal STR loci in Malaysian populations for human identification. *Int J Legal Med* 2020; 134: 1675-1678. DOI: 10.1007/s00414-020-02279-z.

18. Yang L, Zhang X, Zhao L, et al. Population data of 23 autosomal STR loci in the Chinese Han population from Guangdong Province in southern China. *Int J Legal Med* 2018; 132: 133-135. 2017/04/24. DOI: 10.1007/s00414-017-1588-4.

19. Zhang X, Du L, Huang Z, et al. Genetic variation of 20 autosomal STR loci in three ethnic groups (Zhuang, Dai and Hani) in the Yunnan province of southwestern China. *Forensic Sci Int Genet* 2017; 31: 41-42. 2017/07/09. DOI: 10.1016/j.fsigen.2017.06.005.

20. Tie J, Wang X and Oxida S. Genetic polymorphisms of 15 STR loci in a Japanese population. *J Forensic Sci* 2006; 51: 188-189. 2006/01/21. DOI: 10.1111/j.1556-4029.2005.00037.x.

21. Lin SW, Lam TT and Ip SCY. Population data of 23 autosomal STR loci in Hong Kong Chinese. *Forensic Sci Int Genet* 2019; 39: e24-e25. 2018/12/07. DOI: 10.1016/j.fsigen.2018.11.018.

22. Pilav A, Pojskić N, Ahatović A, et al. Allele frequencies of 15 STR loci in Bosnian and Herzegovinian population. *Croat Med J* 2017; 58: 250-256. DOI: 10.3325/cmj.2017.58.250.

23. Kutanan W, Kampuansai J, Colonna V, et al. Genetic affinity and admixture of northern Thai people along their migration route in northern Thailand: evidence from autosomal STR loci. *J Hum Genet* 2011;56:130-137. DOI: 10.1038/jhg.2010.135.

24. Srithawong S, Muisuk K, Srikummool M, et al. Close genetic relationship between central Thai and Mon people in Thailand revealed by autosomal microsatellites. *Int J Legal Med* 2021; 135: 445-448. DOI: 10.1007/s00414-020-02290-4.

25. Huang Q, Xu F, Shen H, et al. Mutation patterns at dinucleotide microsatellite loci in humans. *Am J Hum Genet* 2002; 70: 625-634. 2002/01/15. DOI: 10.1086/338997.

26. 26. Butler JM. *Fundamentals of forensic DNA typing*. USA: Elsevier, 2009.

27. 27. Murad MJ and Amin BK. A microsatellite study for determination of allelic variation of Kurdish population-Kurdistan region-Iraq. *AIP Conference Proceedings* 2017; 1888: 020036. DOI: 10.1063/1.5004313.

28. Huang Q, López JCE, Baeza C, et al. Genetic polymorphism of 15 STR loci in Chinese Han population from Shanghai municipality in East China. *Forensic Sci Int Genet Suppl Ser* 2013; 7: 31-34. DOI: 10.1016/j.fsigen.2012.10.006.

29. Al-Snan NR, Messaoudi S, R. Babu S, et al. Population genetic data of the 21 autosomal STRs included in GlobalFiler kit of a population sample from the Kingdom of Bahrain. *PLoS One* 2019; 14: 1-14. DOI: 10.1371/journal.pone.0220620.

30. Boonderm N, Suriyanratakorn D, Sangpueng S, et al. Population genetic data of 21 STR markers in Thais of southern border provinces of Thailand. *Forensic Sci Int Genet Suppl Ser* 2017; 6: 523-525. DOI: https://doi.org/10.1016/j.fsigss.2017.09.205.

31. Gaviria A, Vela M, Morejon G, et al. Analysis of the most efficient autosomal strs and genetic data for the locus SE33 in ecuadorian population. *Forensic Sci Int Genet Suppl Ser* 2015; 5: 93-95. DOI: https://doi.org/10.1016/j.fsigss.2015.09.038.

32. Kutanan W, Kitpipit T, Phetpeng S, et al. Forensic STR loci reveal common genetic ancestry of the Thai-Malay Muslims and Thai Buddhists in the deep Southern region of Thailand. *J Hum Genet* 2014; 59: 675-681. DOI: 10.1038/jhg.2014.93.

33. Vongpaisarnsin K, Listman JB, Malison RT, et al. Ancestry informative markers for distinguishing between Thai populations based on genome-wide association datasets. *Leg Med (Tokyo)* 2015; 17: 245-250. 2015/03/12. DOI: 10.1016/j.legalmed.2015.02.004.