



Research article

Population data for 12 X-STRs loci in Malaysian Malay and Chinese populations

Aedrianeer Reeza Alwi^{a,b}, Naji Arafat Mahat^{c,d,e,f,*}, Faezah Mohd Salleh^{a,d},
Seri Mirianti Ishar^g, Mohammad Rahim Kamaluddin^{h,**},
Mohd Radzniwan A. Rashidⁱ, Sharifah Nany Rahayu Karmilla Syed Hassan^j,
Sasitaran Iyavoo^k

^a Department of Biosciences, Faculty of Science, Universiti Teknologi Malaysia (UTM), 81310 Johor Bahru, Johor, Malaysia

^b Department of Chemistry Malaysia Johor State, Jalan Abdul Samad, 80100 Johor Bahru, Johor, Malaysia

^c Department of Chemistry, Faculty of Science, Universiti Teknologi Malaysia (UTM), 81310 Johor Bahru, Johor, Malaysia

^d Investigative and Forensic Sciences Research Group, Universiti Teknologi Malaysia (UTM), 81310 Johor Bahru, Johor, Malaysia

^e Centre for Sustainable Nanomaterials, Ibnu Sina Institute for Scientific and Industrial Research, Universiti Teknologi Malaysia (UTM), 81310 Johor Bahru, Johor, Malaysia

^f Centre of Research for Fiqh Forensics and Judiciary, Faculty of Syariah and Law, Universiti Sains Islam Malaysia (USIM), Bandar Baru Nilai, 71800, Nilai, Negeri Sembilan, Malaysia

^g Forensic Science Program, Faculty of Health Sciences, Universiti Kebangsaan Malaysia (UKM), 43600 Bangi, Selangor, Malaysia

^h Centre for Research in Psychology and Human Well-Being, Faculty of Social Sciences and Humanities, Universiti Kebangsaan Malaysia (UKM), 43600 Bangi, Selangor, Malaysia

ⁱ Family Medicine Unit, Faculty of Medicine and Health Science, Universiti Sains Islam Malaysia (USIM), Bandar Baru Nilai, 71800, Nilai, Negeri Sembilan, Malaysia

^j Faculty of Medicine, Universiti Sultan Zainal Abidin (UniSZA), Medical Campus, 20400 Kuala Terengganu, Terengganu, Malaysia

^k School of Natural Sciences, College of Health and Science, University of Lincoln, Lincoln, United Kingdom

ARTICLE INFO

Keywords:

X-STRs
Forensic statistical parameters
Malay
Chinese
Malaysia

ABSTRACT

The utilization of X-chromosome short tandem repeats (X-STRs) for human identification particularly in resolving complex kinship cases has been advocated. Since, forensic statistical parameters vary among different populations, and because the X-STRs population data representing the diverse population of Peninsular Malaysia remain unavailable, the specific attempt reported here for the Malays ($n = 224$) and Chinese ($n = 201$) populations appears forensically relevant to support the evidential value of the 12 X-STRs markers for human identification in Malaysia. Results derived from the Qiagen Investigator® Argus X-12 kit revealed that DXS10135 as the most polymorphic locus with high genetic diversity, polymorphic information content, heterozygosity as well as power of exclusion. Based on allele frequencies, the combined power of discrimination as well as the mean exclusion chance ($MEC_{Krüger}$, $MEC_{Kishida}$, $MEC_{Desmarais}$ and $MEC_{DesmaraisDuo}$) values for the Malays and Chinese were individually ≥ 0.999995532964908 . As for the combined power of discrimination as well as the mean exclusion chance ($MEC_{Krüger}$,

* Corresponding author. Department of Chemistry, Faculty of Science, Universiti Teknologi Malaysia (UTM), 81310 Johor Bahru, Johor, Malaysia.

** Corresponding author. Centre for Research in Psychology and Human Well-Being, Faculty of Social Sciences and Humanities, Universiti Kebangsaan Malaysia (UKM), 43600 Bangi, Selangor, Malaysia

E-mail addresses: aedrianeer@kimia.gov.my (A.R. Alwi), naji.arafat@utm.my (N.A. Mahat), faezah@utm.my (F.M. Salleh), seri.ishar@ukm.edu.my (S.M. Ishar), rahimk@ukm.edu.my (M.R. Kamaluddin), mradzniwan@usim.edu.my (M.R.A. Rashid), karmillahassan@unisza.edu.my (S.N. Rahayu Karmilla Syed Hassan), Siyavoo@lincoln.ac.uk (S. Iyavoo).

<https://doi.org/10.1016/j.heliyon.2024.e38054>

Received 27 March 2024; Received in revised form 21 August 2024; Accepted 17 September 2024

Available online 18 September 2024

2405-8440/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

MEC_{Kishida}, MEC_{Desmarais} and MEC_{DesmaraisDuo}) calculated based on haplotype frequencies, the values were ≥ 0.9999986410567 for the Malays and Chinese populations. In addition, results from the genetic distance, neighbor-joining phylogenetic tree and principal component analysis revealed close biogeographical distributions of the studied populations with other South East Asian populations. Hence, the utilization of the X-STRs data for identifying individuals among the Malays and Chinese populations in Peninsular Malaysia for forensic application appears as highly supported.

1. Introduction

Malaysia is located in the South East Asia equatorial zone (Latitude: 1° and 7° North; Longitudes: 100° and 119° East) of which it covers an area of 330,534 square kilometres. The federation of Malaysia comprises of Peninsular Malaysia (11 states and 2 Federal Territories) as well as the East Malaysia (2 states and 1 Federal territory), separated by the South China Sea [1]. Being a multi-racial country, the Malaysian population has been estimated at 33.5 million, consisting of Malays and other Bumiputera groups (70.1 %), Chinese (22.6 %), Indians (6.6 %) and others (0.7 %). Out of 70.1 % of Bumiputera, the majority being the Malays (58.0 %) followed by other groups of Bumiputera (12.1 %), residing in the Peninsular and East Malaysia [2]. Previous researchers have described the Malays and Chinese in Peninsular Malaysia as the following. Malay is specifically defined in the Constitution of Malaysia (Article 160(2)) as “a person who professes the religion of Islam, habitually speaks the Malay language, conforms to Malay custom and - (a) was born before Merdeka Day in the Federation or in Singapore or born of parents one of whom was born in the Federation or in Singapore, or is on that day domiciled in the Federation or in Singapore; or (b) is the issue of such a person” [3]. The majority of the Malays and indigenous populations (Negrito, Senoi and Proto-Malay) reside in Peninsular Malaysia and they comprise of various sub-ethnicities namely Banjar, Java, Minangkabau, Bugis and other groups from Austronesian [4], emerged prominently during the Malacca Sultanate (15th century) through the trading activities and seafaring life [5].

Malaysian Chinese on the other hand make up the second-largest overseas Chinese population (after Thailand) and the second-largest ethnic group in Malaysia [6]. Chinese people first came to the Malacca Empire in the early 15th century, when the Sultan Mansur Syah married a Chinese princess named Hang Li Po [6]. This was the beginning of the first wave of migration, when Chinese merchants and young aristocrats from an imperial company started to settle in and around Melaka, assimilated with the local culture and custom through business and marriage [7]. Described as the second wave of migration, the Chinese in Peninsular Malaysia originating from the southern part of China arrived hugely in the 18th to the early 20th centuries due to open migration policy by the British to surplus the labour support for their colonial economic activities [8]. The majority of them are from the southeast coast of Mainland China's Fujian and Guangdong regions [7] and worked in tin mining which later developed into the Chinese civilization in Peninsular Malaysia [9].

The application of X-chromosome short tandem repeats (X-STRs) can be appealing for resolving kinship and deficiency paternity cases, and may prove useful in investigations involving missing persons and mass disasters, in particular when comparison profiles are not available [10]. However, the available of X-STRs population data for Malaysia are limited to Malays (in Kuala Lumpur alone) [11] as well as the minority sub-ethnics [12] with inadequate number of sample size and Kedayan (an ethnicity in the East Malaysia region) [13]. As such the available data may not represent the diverse population of Peninsular Malaysia, limiting their evidential values as evidence in court [10]. Not only that, the complete X-STRs population data for representing the admixture populations (e.g. Malay-Chinese) in Malaysia are also lacking. Considering that the forensic statistical parameters may differ among the different populations [14,15], the statistical support for the use of DNA profiles for one population may not be accurately representing other populations, especially among high diversity ethnic populations and genetic admixture [16,17]. In this context, genetic properties are evaluated based on the relevant populations to determine the rarity of the obtained profile [18]. Therefore, a thorough sampling of populations of interest is paramount for validating and documenting the forensic statistical support for the use of X-STRs analysis for complex kinship investigation as well as proving the evidence as ‘beyond any reasonable doubt’ in the court of law.

Hence, the aim of this research was to provide the empirically justified statistical support for the use of 12 X-STRs population data for the Malays and Chinese in Peninsular Malaysia for human identification in Malaysia, particularly for complex kinship investigation. Importantly, this would be the first attempt to provide such data for the Chinese population in Malaysia.

2. Materials and method

2.1. Ethical requirements and sample collection

The procedures and protocols used in this present research was approved by the Human Research Ethics Committee of Universiti Sains Islam Malaysia (USIM/JKEP/2022-225), dated July 21, 2022, in compliance with the ethical guidelines and requirements for the publication of genetic and genomic data. In this context, it is important to iterate that “Urban and cosmopolitan population studies should involve at least 200 individuals, whereas, regional datasets should contain a minimum of 100 samples. Exceptions are valid for extremely small population sizes.” [19]. The fact that this present research utilized 224 Malay and 201 Chinese participants (amounting to a total of 425 participants), it can be construed that the number of participants examined here is more than sufficient to represent the populations of Malay and Chinese in Peninsular Malaysia. The written informed consents were attained before the

finger-pricked blood samples on Flinder's Technology Associates (FTA) cards were collected from the unrelated healthy individuals consisting of (a) 224 Malays (113 females and 111 males) and (b) 201 Chinese (100 females and 101 males) residing in Peninsular Malaysia. For ensuring the representativeness, the stratified purposeful sampling method was used, stratifying the participants according to the different states and territory within the Peninsular Malaysia.

2.2. X-STRs typing

To facilitate international population data exchange, utilization of the X-chromosomal STR kit that has acquired popularity for forensic application must be chosen. The fact that Qiagen Investigator® Argus X-12 QS kit fulfils such a requirement, its utilization to generate suitable population data for the Malays and Chinese in the diverse Malaysian population proves relevant, although there are other kits (e.g. Microreader™ 19X) for X-chromosomal STR that offer more loci for analysis. Since the utilization of Microreader™ 19X Kit appears limited to specific populations and/or geographical location (e.g. Han and Sierra Leone), the ability to enable the international population data exchange may be limited. Furthermore, the lack of population and mutation data prevailed for the other X-STRs kits may impede forensic evaluation on their practical values.

The blood samples on FTA cards were punched (~1.2 mm each) and subsequently purified. The purified blood samples on FTA cards were then directly amplified using the reduced polymerase chain reaction (PCR) reaction volume of the Qiagen Investigator® Argus X-12 kit (Germany) as described by Alwi et al. [20], co-amplifying the 12 X-STRs loci namely DXS10103, DXS8378, DXS10101, DXS10134, DXS10074, DXS7132, DXS10135, DXS7423, DXS10146, DXS10079, HPRTB and DXS10148 [21]. These markers are grouped into four linkage groups: (LG1:DXS10148, DXS10135, DXS8378; LG2: DXS7132, DXS10079, DXS10074; LG3: DXS10103, HPRTB, DXS10101; LG4: DXS10146, DXS10134, DXS7423).

The PCR amplification was performed on the Applied Biosystems™ ProFlex™ PCR System (Thermo Fisher Scientific). Subsequently, the PCR products were separated via capillary electrophoresis on the Applied Biosystems 3500xL Genetic Analyzer (Thermo Fisher Scientific), and data were analyzed using the GeneMapper™ ID-X v1.5 software (Thermo Fisher Scientific). For confirming the presence of off-ladder alleles (as determined by size), re-injection of the samples into the capillary electrophoresis was performed.

2.3. Quality control

The research was performed in an ISO/IEC 17025:2017 accredited forensic DNA laboratory (Department of Chemistry Malaysia Johor State Laboratory; FT-0084) by the American National Standards Institute (ANSI) National Accreditation Board (ANAB) following the laboratory internal standards and kit controls. The female control DNA 9947A (Thermo Fisher Scientific, USA) was used as the positive control, while deionized water served as the negative ones.

2.4. Forensic statistical analysis

Allele frequencies of male and female samples as well as the haplotype frequencies for each linkage group (LG) of male samples were estimated by simple counting method using Statistics for X-STR (StatsX) v2.0 [22]. The exact test of Hardy-Weinberg equilibrium (HWE) per locus in female samples and the linkage disequilibrium (LD) testing between all pairs of the 12 loci of male and female samples were computed using Arlequin v3.5.2.2 software [23]. The forensic statistical parameters that include the gene diversity (GD), haplotype diversity (HD), polymorphic information content (PIC), power of discrimination in males (PD_M) and in females (PD_F), mean exclusion chance in (a) deficiency cases ($MEC_{Krüger}$) [24], (b) standard trios involving daughters ($MEC_{Kishida}$) [25], (c) standard trios involving daughters ($MEC_{Desmarais}$) and (d) father/daughter or mother/son duos ($MEC_{DesmaraisDuo}$) [26] as well as the combined PD_M , PD_F , $MEC_{Krüger}$, $MEC_{Kishida}$, $MEC_{Desmarais}$ and $MEC_{Desmarais Duo}$ were calculated using StatsX v2.0 [22]. The heterozygosity (Het) and power of exclusion (PE) were calculated using the formula described by the previous researcher [27].

2.5. Population comparisons

The Nei's D_A genetic distances [28] among the studied populations (Malays and Chinese) with those of 21 other populations were measured and represented in the form of Neighbor-joining (NJ) tree, constructed using POPTREE2 software [29] with 1000 bootstrap replications. Visualization and interpretation of phylogenetic tree were attempted using Molecular Evolutionary Genetics Analysis version 11 [30]. Subsequently, the principal components analysis (PCA) was performed using the MultiVariate Statistical Package [31].

3. Results and discussion

3.1. Forensic statistical analysis

The allele and haplotype frequencies as well as the forensic parameters were calculated from the Malays and Chinese population in Peninsular Malaysia dataset (Supplementary Tables S1–S4) generated from this present research. Results from the independent sample *t*-test comparing the allele frequencies between male and female samples indicated no significant differences. However, to provide a clearer understanding of the genetic diversity and structure within and between the populations studied, the allele frequencies and forensic parameters for male and female subjects are separately represented prior to the integration (pooled) (Supplementary

Tables S5–S6). A total of 158 and 143 alleles were observed for the Malays and Chinese populations respectively. A total of 12 distinct off-ladder alleles were found at 6 loci in the Malays and Chinese populations. While the majority of the off-ladder alleles were observed at DXS10101 (4 distinct alleles), one allele each at loci DXS10103, DXS10134, DXS10135 and DXS10079 were found in the Malay population. As for the Chinese population, 5 distinct alleles at 4 loci (DXS10101, DXS10134, DXS10079 and DXS10148) were observed. Specific and elaborative discussions pertaining to the aspect of rare and off-ladder alleles in view of its application for forensic intelligence are reported in another research paper. In this present research, the most polymorphic locus was DXS10135 considering the high GD and PIC values, viz. Malays (PIC = 0.9135) and Chinese (PIC = 0.9157). In contrast, locus DXS7423 was found having the lowest PIC values for the Malays (0.4944) and Chinese (0.4322) since only 3–4 alleles were detected at that particular locus. The proportion of heterozygous individuals in the population (Het) as well as the PE values being the highest (DXS10135) as well as the lowest (DXS7423) at the similar loci. These findings were found similar with those reported for other populations [11,32,33] on the same 12 X-STRs markers.

No significant deviations from HWE were observed for the all the female samples (Supplementary Table S7), except in Chinese population (at locus HPRTB, $p = 0.0056$). However, after applying the Bonferroni's corrections ($p < 0.05/12 = 0.0042$), no HWE deviation was evident. The results of the LD analysis between all pairs of the 12 loci in both male and female samples are tabulated in Supplementary Tables S8a–b and S9a–b with a $p < 0.0008$ after an adjustment for multiple testing ($0.05/66$). Significant LD ($p = 0.0000$) was observed in both female and male samples of Chinese population within LG3 (between DXS10101–DXS10103) as well as in male samples of Malay population. As for the female samples of Malay population, significant LD was found between HPRTB–

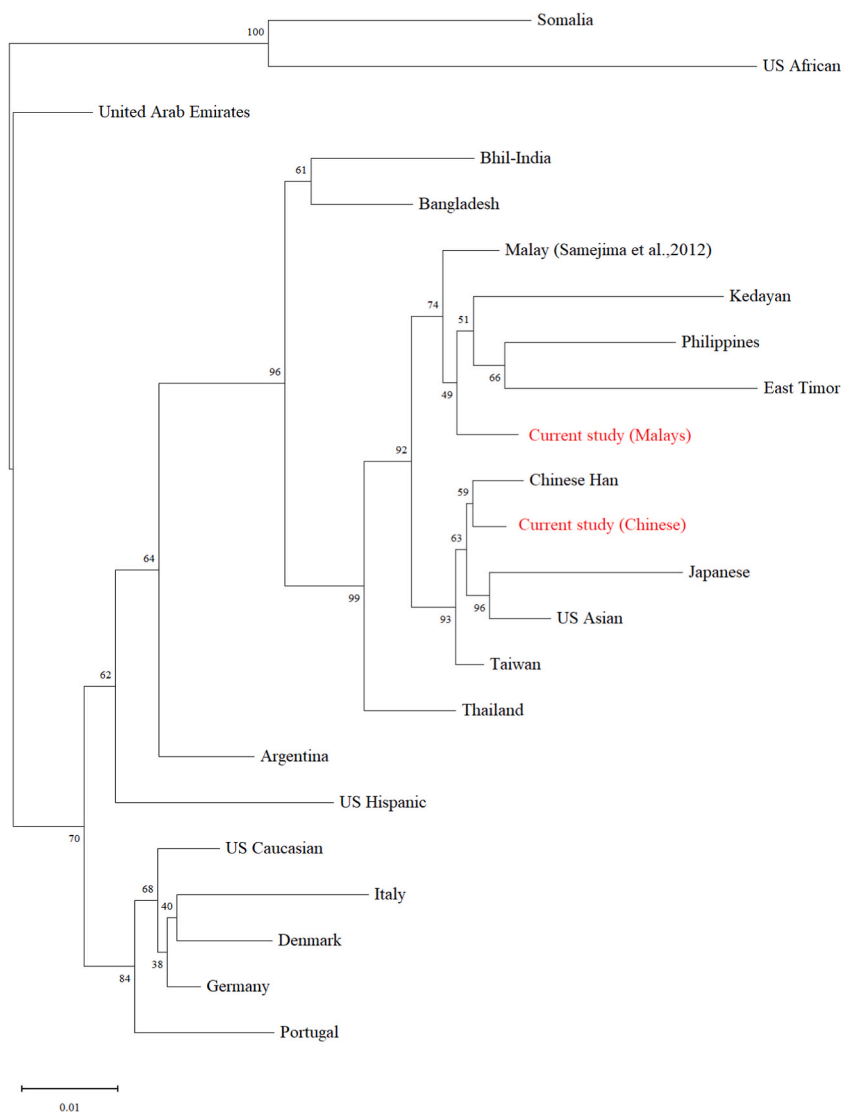


Fig. 1. The Neighbor-Joining tree constructed based on the genetic distance (allele frequencies data) among the Malays and Chinese against other populations (as in Supplementary Table S13). Bootstrap values indicate the support for each node. The bar indicates 0.01 substitutions per site.

DXS10103 ($p = 0.0000$), comparable with the findings by Samejima et al. [11].

The haplotypes frequencies and forensic statistical parameters corresponding to the four LGs (LG1-DXS10148, DXS10135, DXS8378; LG2- DXS7132, DXS10079, DXS10074; LG3- DXS10103, HPRTB, DXS10101; LG4- DXS10146, DXS10134, DXS7423) are shown in [Supplementary Tables S10–S11](#). The number of unique haplotypes in LG1-LG4 observed were 95, 68, 72 and 81 for the Malays respectively. As for the Chinese population, the same for LG1-LG4 were 90, 68, 75 and 70, correspondingly. For both the Malays ($HD = 0.9971$) and Chinese ($HD = 0.9976$), LG1 was found as the most informative LG, while the lowest being LG3 for the Malays ($HD = 0.9876$) and LG4 for the Chinese ($HD = 0.9905$).

In this present research, the combined PD_m , PD_f , MEC_{Kruger} , $MEC_{Kishida}$, $MEC_{Desmarais}$ and $MEC_{Desmarais Duo}$ were calculated based on allele (pooled) and haplotype frequencies. The combined PD_m , PD_f , MEC_{Kruger} , $MEC_{Kishida}$, $MEC_{Desmarais}$ and $MEC_{Desmarais Duo}$ calculated based on allele frequencies revealed the values ≥ 0.999995532964908 for the Malays and Chinese population ([Supplementary Table S12](#)). When calculated using the haplotype frequencies, the combined PD_m , PD_f , MEC_{Kruger} , $MEC_{Kishida}$, $MEC_{Desmarais}$ and $MEC_{Desmarais Duo}$ were ≥ 0.9999986410567 for the Malays and Chinese population ([Supplementary Tables S10–S11](#)).

3.2. Population comparisons

The genetic distance (D_A) of all the pairs calculated among the Malaysian Malays and Chinese as well as the 21 other published populations are shown in [Supplementary Table S13](#). The reasons for choosing the 21 other published populations are bipartite in nature. Firstly, all these populations data were derived from the same X-STRs markers and therefore would enable the appropriate discussion to be made with the present data. Secondly, the population data included would represent the different continents, as well as encompassing all the available population data for the South East Asian region, justifying the suitability of the markers for DNA profiling across the different populations. The D_A values ranged from 0.0120 to 0.1340 for the studied populations against that of the other populations. A higher D_A value was observed when the populations were geographically distant, comparable with the finding reported by previous researchers [34]. [Fig. 1](#) depicts the NJ tree constructed based on the genetic distance (allele frequencies data) among the populations. The NJ tree revealed that the populations of Malays and Chinese studied here are clustered within a separate larger group (bootstrap support: 99 %) than those of the other group that comprised of United Arab Emirates, Somalia, US African, US Caucasian, Italy, Denmark, Germany, Portugal, US Hispanic and Argentina (bootstrap support: 70 %). Furthermore, it can be seen that the Malay and Chinese populations are grouped separately with high bootstrap support (92 %). While the current Malay population is grouped together (bootstrap support: 74 %) with the Malays from the previous study [11], Kedayan [13], Philippines [33] and East Timor [35], the current Chinese population, on the other hand, is grouped together (bootstrap support: 59 %) with the Chinese Han [36], Taiwanese [37], Japanese [38] and US Asian [39].

Similar findings on the categorical separation of the populations studied *via* NJ were also observed in the two-dimensional PCA case

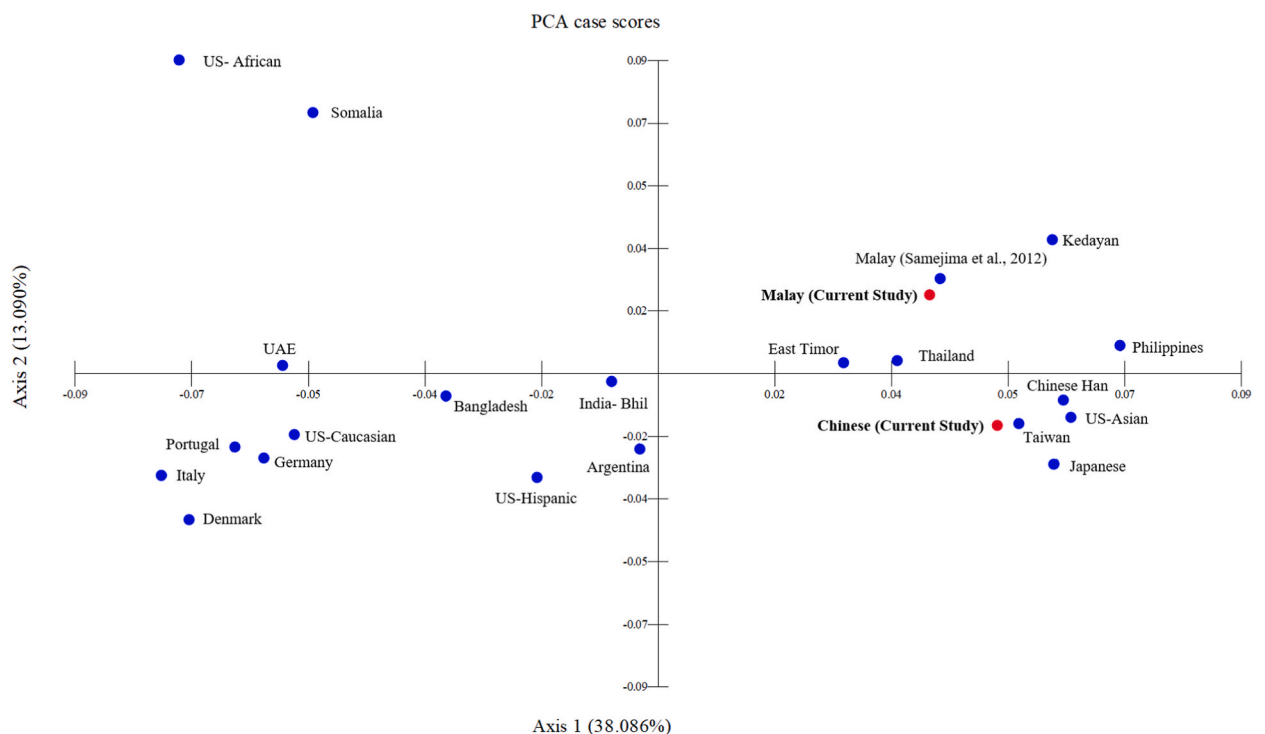


Fig. 2. The two-dimensional principal components analysis case scores constructed using the data of allele frequencies across the 12 X-STRs loci among the Malays and Chinese against other populations (as in [Supplementary Table S13](#)).

scores (Fig. 2). The close proximity observable via NJ and PCA between the Chinese observed in this current research with Chinese Han, Tawian, Japanese and US-Asian populations was within expectation, and it can be explained by several theories. The first theory is “Out of Taiwan” and “Express Train to Polynesian” models, suggesting that Austronesian populations (like the Malays) are originally from the provinces in China namely Taiwan and Yunnan, before arriving in Peninsular Malaysia in about 4500 years ago [40]. The second theory, revolves around the fact that transportation and the globalization especially in Southeast and South Asian populations have resulted in the frequent occurrence of gene flow among the different populations in the surrounding neighbourhoods [41]. The Malays were generally maritime traders, and as such indirectly promoting mixed marriages among the populations of the islands around the peninsular [40]. Deng et al. [42] reported that the population of Malaysia is affected by the genetic admixture from neighbouring East Asian populations. In fact, it was reported that between 4 % and 16 % of Malay genomes are descended from East Asians, with the remainder coming from South Asians, Austronesian, and aboriginal South Asians who interacted with Malays between 175 and 1500 years ago [41]. In point of fact, the assessments of molecular variance and average coefficient of correlation that were carried out by earlier researchers have revealed that the Malay and Chinese populations have the strongest correlation between them, in comparison to those of other populations [43,44].

Others studies on genetic markers also demonstrated similar findings attributable to the genetic variations of the Peninsular Malaysia population [5,45]. The fact that Malaysia is a multi-racial country which composed of different genetic lineages, current DNA study should be cautiously interpreted to avoid contradiction with archaeological findings as historical events [45]. It is duly noted that integrating several genetic markers will strengthen the links between ethnic groups in terms of genetic distance, improving the understanding of the genetic makeup of the populations [46].

4. Conclusion

In conclusion, the allele frequencies and forensic statistical parameters of the studied Malaysian Malays and Chinese populations demonstrated that the 12 X-STRs markers were highly informative and discriminating. The population data reported here for the Chinese population in Malaysia appear to be the first of its kind to represent such population in the South East Asian region, substantiating the evidential value of the 12 X-STRs markers for forensic human identification. Therefore, the incorporation of X-STRs analysis for routine DNA casework appears pertinent, especially for complex kinship investigations in multi racial countries like Malaysia. The fact that little emphasis has been given on the admixture populations, suitable effort to provide reliable forensic population data for such populations shall be considered as an interesting future study.

Ethics and consent statement

The ethical approval for this research was awarded by the Human Research Ethics Committee of Universiti Sains Islam Malaysia (USIM/JKEP/2022-225), dated on July 21, 2022. Every participant provided a written informed consent to participate in the research and for their data to be published. The Human Participant Declaration form was provided during submission.

Funding

This research was supported by Ministry of Higher Education Malaysia Fundamental Research Grant Scheme (FRGS/1/2023/STG01/UTM/02/1) awarded to Universiti Teknologi Malaysia, as well as the funding from Faculty of Social Sciences and Humanities, Universiti Kebangsaan Malaysia (PP-FSSK-2024).

Data availability statement

The data are presented in the supplementary materials.

CRediT authorship contribution statement

Aedrianeer Reeza Alwi: Writing – original draft, Visualization, Investigation, Formal analysis, Data curation. **Naji Arafat Mahat:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization. **Faezah Mohd Salleh:** Writing – review & editing, Supervision, Methodology. **Seri Mirianti Ishar:** Writing – review & editing, Supervision, Methodology. **Mohammad Rahim Kamaluddin:** Writing – review & editing. **Mohd Radzniwan A. Rashid:** Writing – review & editing. **Sharifah Nany Rahayu Karmilla Syed Hassan:** Writing – review & editing. **Sasitaran Iyavoo:** Writing – review & editing, Formal analysis.

Declaration of competing interest

The authors declare no conflicts of interest.

Acknowledgments

The authors are thankful to the Public Service Department, Malaysia for awarding the Hadiah Latihan Persekutuan (HLP)

scholarship to Aedriane Reeza Alwi, and the Department of Chemistry Malaysia for all the necessary support for this research. The sampling assistance provided by Ahmad Sabri Mohamed, Fadilah Mohd Hassan, Hazwani Hapiz, Nor Wajihan Muda, Nik Ihtisyam Majdah Nik Razi, Salina Mohd Isa and Varohthini K Maharaja are highly appreciated.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.heliyon.2024.e38054>.

References

- [1] Department of Information Malaysia, Official year book 2021. <https://dbook.penerangan.gov.my/dbook/?product=official-year-book-2021>, 2022. (Accessed 28 October 2023).
- [2] Department of Statistics Malaysia, Media statement for the report of demographic statistics Malaysia, third quarter 2023. https://www.dosm.gov.my/uploads/release-content/file_20231107114523.pdf, 2023. (Accessed 7 November 2023).
- [3] N.M. Ahmad, The definition of Malay according to the federal constitution of Malaysia and its impact on the understanding and islamic practices in Malaysia, *Sains Insani* 5 (2) (2020) 9–16.
- [4] P. Yahya, S. Sulong, A. Harun, H. Wan Isa, N.S. Ab Rajab, P. Wangkumhang, A. Wilantho, C. Ngamphiw, S. Tongsim, Z. Alwi, Analysis of the genetic structure of the Malay population: ancestry-informative marker SNPs in the Malay of Peninsular Malaysia, *Forensic Sci Int Genet* 30 (2017) 152–159, <https://doi.org/10.1016/j.fsigen.2017.07.005>.
- [5] S. Parthipan, S.M. Ishar, Perspectives on ancestral lineages and genetic markers of Malay population in Peninsular Malaysia, *Jurnal Sains Kesihatan Malaysia* 20 (2022) 83–95, <https://doi.org/10.17576/jskm-2022-2001-08>.
- [6] R. Vollmann, T. Wooi Soon, Chinese identities in multilingual Malaysia. <https://doi.org/10.25364/04.45:2018.89.3>, 2018.
- [7] M.F. Zakaria, A. Ibrahim, Chinese civilization in Malaysia: history and contribution, *Journal of Social Science and Humanities* 5 (2022) 1–6, <https://doi.org/10.26666/rmp.jssh.2022.1.1>.
- [8] K. Ismail, Imperialism, colonialism and their contribution to the formation of Malay and Chinese ethnicity: an historical analysis. <http://journals.iium.edu.my/intdiscourse/index.php/islam>, 2020.
- [9] M.F. Zakaria, A. Ibrahim, Development of Chinese civilisation and education in Malaysia, *Journal of Management & Science* 19 (2021) 6, <https://doi.org/10.57002/jms.v19i1.223>.
- [10] A.R. Alwi, N.A. Mahat, F.M. Salleh, S.M. Ishar, M.R. Kamaluddin, M.R.A. Rashid, Applications of X-chromosome short tandem repeats for human identification: a review, *J Trop Life Sci* 13 (2023) 187–212, <https://doi.org/10.11594/jtls.13.01.19>.
- [11] M. Samejima, Y. Nakamura, P. Nambiar, K. Minaguchi, Genetic study of 12 X-STRs in Malay population living in and around Kuala Lumpur using Investigator Argus X-12 kit, *Int. J. Leg. Med.* 126 (2012) 677–683, <https://doi.org/10.1007/s00414-012-0705-7>.
- [12] N.A. Abdullah, Genetic Variations of Six Orang Asli Sub-groups in Peninsular Malaysia Using Autosomal and X-Chromosome Short Tandem Repeats (STRS) Analysis, *Universiti Sains Malaysia*, 2018. Master thesis.
- [13] H.M. Hakim, H.O. Khan, S.A. Ismail, J. Lalung, A.E. Kofi, M.Y. Aziz, S. Pati, B.R. Nelson, G.K. Chambers, H.A. Edinur, Population data and genetic characteristics of 12 X-STR loci using the Investigator® Argus X-12 Quality Sensor kit for the Kedayan population of Borneo in Malaysia, *Int. J. Leg. Med.* (2021), <https://doi.org/10.1007/s00414-021-02577-0>.
- [14] M.N.A. Rashid, N.A. Mahat, H.O. Khan, R.A. Wahab, H. Maarof, D. Ismail, A.R. Alwi, S.N.R.K. SyedHassan, Population data of 21 autosomal STR loci in Malaysian populations for human identification, *Int. J. Leg. Med.* 134 (2020) 1675–1678, <https://doi.org/10.1007/s00414-020-02279-z>.
- [15] D. Takić Miladinov, P. Vasiljević, D. Šorgić, E. Podovšovnik Axelsson, A. Stefanović, Allele frequencies and forensic parameters of 22 autosomal STR loci in a population of 983 individuals from Serbia and comparison with 24 other populations, *Ann. Hum. Biol.* 47 (2020) 632–641, <https://doi.org/10.1080/03014460.2020.1846784>.
- [16] M. Baeta, E. Prieto-Fernández, C. Núñez, T. Kleinbielen, P. Villacausa, L. Palencia-Madrid, O. Alvarez-Gila, Begoña Martínez-Jarreta, M.M. De Pancorbo, Study of 17 X-STRs in Native American and Mestizo populations of Central America for forensic and population purposes, *Int. J. Leg. Med.* 135 (2021) 1773–1776, <https://doi.org/10.1007/s00414-021-02536-9>.
- [17] G. He, Y. Li, X. Zou, Y. Zhang, H. Li, M. Wang, J. Wu, X-chromosomal STR-based genetic structure of Sichuan Tibetan minority ethnicity group and its relationships to various groups, *Int. J. Leg. Med.* 132 (2018) 409–413, <https://doi.org/10.1007/s00414-017-1672-9>.
- [18] J.S. Buckleton, J.-A. Bright, J.M. Curran, D. Taylor, Validating databases, in: J.S. Buckleton, J.-A. Bright, D. Taylor (Eds.), *Forensic DNA Evidence Interpretation*, second ed., CRC Press, 2016, pp. 133–181.
- [19] W. Parson, L. Roewer, Publication of population data of linearly inherited DNA markers in the International Journal of Legal Medicine, *Int. J. Leg. Med.* 124 (2010) 505–509, <https://doi.org/10.1007/s00414-010-0492-y>.
- [20] A.R. Alwi, N.A. Mahat, F. Mohd Salleh, S.M. Ishar, M.R. Kamaluddin, M.R.A. Rashid, Internal validation of reduced PCR reaction volume of the Qiagen Investigator® Argus X-12 QS Kit from blood samples on FTA cards, *J. Forensic Sci.* (2023), <https://doi.org/10.1111/1556-4029.15370>.
- [21] Qiagen, Investigator® Argus X-12 QS Handbook, Qiagen, Germany, 2022.
- [22] Y. Lang, F. Guo, Q. Niu, StatsX v2.0: the interactive graphical software for population statistics on X-STR, *Int. J. Leg. Med.* 133 (2019) 39–44, <https://doi.org/10.1007/s00414-018-1824-6>.
- [23] L. Excoffier, H.E.L. Lischer, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows, *Mol Ecol Resour* 10 (2010) 564–567, <https://doi.org/10.1111/j.1755-0998.2010.02847.x>.
- [24] J. Krüger, W. Fuhrmann, K. Lichte, C. Steffens, On the utilization of erythrocyte acid phosphatase polymorphism in paternity evaluation, *Dtsch. Z. Gesamte Gerichtl. Med.* 64 (1968) 127–146.
- [25] T. Kishida, W. Wang, M. Fukuda, Y. Tamaki, Duplex PCR of the Y-27H39 and HPRT loci with reference to Japanese population data from the HPRT locus, *Nihon Hoigaku Zasshi* 51 (1997) 67–69.
- [26] D. Desmarais, Y. Zhong, R. Chakraborty, C. Perreault, L. Busque, Development of a highly polymorphic STR marker for identity testing purposes at the human androgen receptor gene (HUMARA), *J. Forensic Sci.* 43 (1998) 1046–1049.
- [27] J.M. Butler, *Advanced Topics in Forensic DNA Typing: Interpretation*, third ed., Academic Press, 2015.
- [28] M. Nei, F. Tajima, Y. Tateno, Accuracy of estimated phylogenetic trees from molecular data II. Gene frequency data, *J. Mol. Evol.* 19 (1983) 153–170.
- [29] N. Takezaki, M. Nei, K. Tamura, POPTREEW: web version of POPTREE for constructing population trees from allele frequency data and computing some other quantities, *Mol. Biol. Evol.* 31 (2014) 1622–1624, <https://doi.org/10.1093/molbev/msu093>.
- [30] K. Tamura, G. Stecher, S. Kumar, MEGA11: molecular evolutionary genetics analysis version 11, *Mol. Biol. Evol.* 38 (2021) 3022–3027, <https://doi.org/10.1093/molbev/msab120>.
- [31] W.L. Kovach, MVSP-A Multi-Variate Statistical Package for Windows (2007), Kovach Computing Services, Pentraeth, Wales, UK, 1999. Version 3.1, 137(3).
- [32] J. Xing, A. Adnan, A. Rakha, K. Kasim, A. Noor, J. Xuan, X. Zhang, J. Yao, D. McNeven, B. Wang, Genetic analysis of 12 X-STRs for forensic purposes in Liaoning Manchu population from China, *Gene* 683 (2019) 153–158, <https://doi.org/10.1016/j.gene.2018.10.020>.

- [33] J.M. Salvador, D.L.T. Apaga, F.C. Delfin, G.C. Calacal, S.E. Dennis, M.C.A. De Ungria, Filipino DNA variation at 12 X-chromosome short tandem repeat markers, *Forensic Sci Int Genet* 36 (2018) e8–e12, <https://doi.org/10.1016/j.fsigen.2018.06.008>.
- [34] I. Veselinović, G. Petrić, D. Vapa, M. Djan, N. Veličković, T. Veljović, Genetic analysis of 12 X-STR loci in the Serbian population from Vojvodina province, *Int. J. Leg. Med.* 132 (2018) 405–408, <https://doi.org/10.1007/s00414-017-1677-4>.
- [35] H. Moreira, H. Costa, F. Tavares, L. Souto, Genetic variation of 12 X-chromosomal STR loci in an East Timor sample, *Int. J. Leg. Med.* 129 (2015) 257–258, <https://doi.org/10.1007/s00414-014-1126-6>.
- [36] X.P. Zeng, Z. Ren, J.D. Chen, D.J. Lv, D.Y. Tong, H. Chen, H.Y. Sun, Genetic polymorphisms of twelve X-chromosomal STR loci in Chinese Han population from Guangdong Province, *Forensic Sci Int Genet* 5 (2011), <https://doi.org/10.1016/j.fsigen.2011.03.005>.
- [37] M.Y. Chen, C.W. Ho, C.E. Pu, F.C. Wu, Genetic polymorphisms of 12 X-chromosomal STR loci in Taiwanese individuals and likelihood ratio calculations applied to case studies of blood relationships, *Electrophoresis* 35 (2014) 1912–1920, <https://doi.org/10.1002/elps.201300645>.
- [38] S. Uchigasaki, J. Tie, D. Takahashi, Genetic analysis of twelve X-chromosomal STRs in Japanese and Chinese populations, *Mol. Biol. Rep.* 40 (2013) 3193–3196, <https://doi.org/10.1007/s11033-012-2394-1>.
- [39] T.M. Diegoli, A. Linacre, P.M. Vallone, J.M. Butler, M.D. Coble, Allele frequency distribution of twelve X-chromosomal short tandem repeat markers in four U.S. population groups, *Forensic Sci Int Genet Suppl Ser* 3 (2011), <https://doi.org/10.1016/j.figss.2011.09.102>.
- [40] N.W. Hanis Zainal Abidin, N. Mohd Nor, P. Sundararajulu, Z. Zafarina, Understanding the genetic history of Malay populations in Peninsular Malaysia via KIR genes diversity, *Am. J. Hum. Biol.* 33 (2021), <https://doi.org/10.1002/ajhb.23545>.
- [41] L. Deng, B.P. Hoh, D. Lu, W.Y. Saw, R. Twee-Hee Ong, A. Kasturiratne, H. Janaka De Silva, B.A. Zilfalil, N. Kato, A.R. Wickremasinghe, Y.Y. Teo, S. Xu, Dissecting the genetic structure and admixture of four geographical Malay populations, *Sci. Rep.* 5 (2015), <https://doi.org/10.1038/srep14375>.
- [42] L. Deng, B.P. Hoh, D. Lu, R. Fu, M.E. Phipps, S. Li, A.R. Nur-Shafawati, W.I. Hatin, E. Ismail, S.S. Mokhtar, L. Jin, B.A. Zilfalil, C.R. Marshall, S.W. Scherer, F. Al-Mulla, S. Xu, The population genomic landscape of human genetic structure, admixture history and local adaptation in Peninsular Malaysia, *Hum. Genet.* 133 (2014) 1169–1185, <https://doi.org/10.1007/s00439-014-1459-8>.
- [43] Y.M. Chang, R. Perumal, P.Y. Keat, D.L.C. Kuehn, Haplotype diversity of 16 Y-chromosomal STRs in three main ethnic populations (Malays, Chinese and Indians) in Malaysia, *Forensic Sci. Int.* 167 (2007) 70–76, <https://doi.org/10.1016/j.forsciint.2006.01.002>.
- [44] Y. Nakamura, M. Samejima, K. Minaguchi, P. Nambiar, Population genetics of identifier System in Malaysia, *Bull. Tokyo Dent. Coll.* 57 (2016) 233–239, <https://doi.org/10.2209/tdcpublication.2016-1400>.
- [45] H.K. Norhalifah, F.H. Syaza, G.K. Chambers, H.A. Edinur, The genetic history of Peninsular Malaysia, *Gene* 586 (2016) 129–135, <https://doi.org/10.1016/j.gene.2016.04.008>.
- [46] Z. Wang, M. Song, Q. Lyu, J. Ying, Q. Wu, F. Song, L. Jiang, X. Wei, S. Wang, F. Wang, Y. Zhou, X. Song, H. Luo, A forensic population database of autosomal STR and X-STR markers in the Qiang ethnic minority of China, *Heliyon* 9 (2023) e21823, <https://doi.org/10.1016/j.heliyon.2023.e21823>.