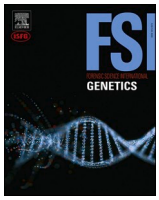
Forensic Science International: Genetics 74 (2025) 103169

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsigen



Characterisation of identity-informative genetic markers in the Australian population with European ancestry

Jessica L. Watson [a,b,c,*], Kaymann Cho [c], Kelly Grisedale [a,c], Jodie Ward [a,b], Dennis McNevin [a,b]

[a] National DNA Program for Unidentified and Missing Persons, Australian Federal Police, Australia
[b] Centre for Forensic Science, School of Mathematical & Physical Science, Faculty of Science, University of Technology Sydney, Australia
[c] Biology, AFP Forensics, Australian Federal Police, Australia

ARTICLE INFO

Keywords:
Single nucleotide polymorphism
Allele frequency
Australia
Human identification
Massively parallel sequencing
Short tandem repeat

ABSTRACT

Identity-informative single nucleotide polymorphisms (iiSNPs) are valuable genetic markers for human identification and kinship testing in forensic casework, especially when the quality and quantity of DNA evidence is not suitable for routine short tandem repeat (STR) profiling. This study analysed 105 buccal samples representing the Australian population with European ancestry in order to assign allele frequencies and conduct population genetic analyses for 94 iiSNPs and 20 STRs. The markers were assessed by calculating relevant forensic statistics and testing for deviations from Hardy-Weinberg and linkage equilibrium. No linkage of statistical significance was observed between any of the pair-wise combinations of the combined 114 identity-informative markers and only one STR exhibited deviation from Hardy-Weinberg equilibrium (D8S1179). The probability of matching genotypes being observed within this population was of the order of 10^{-23} for STRs, 10^{-38} for iiSNPs and 10^{-60} for the combined identity-informative marker panel, improving the ability to discriminate between individuals when calculating likelihood ratios in direct or indirect matching scenarios. Further, the addition of iiSNPs will facilitate identifications when suboptimal STR profiles are recovered from compromised or challenging samples and aid comparisons to genetic relatives for familial or kinship testing.

1. Introduction

First introduced in the 1990s, short tandem repeats (STRs) are segments of repeated DNA motifs consisting of two to six bases dispersed throughout the genome [1]. STRs located on the autosomal chromosomes are the most common genetic marker currently targeted for forensic DNA profiling applications [2]. Differentiation between individuals is made possible by the combination of alleles inherited from each biological parent, with each allele defined by the number of times the DNA motif is repeated within an STR [3,4]. The combination of multiple STRs within one profile increases the discrimination power and uniqueness of the profile. Their highly polymorphic nature is due in some part to their high mutation rates of the order of 10^{-3} per meiosis [3].

STR profiling is the gold standard forensic genetic method for human identification and is typically used in criminal and coronial investigations to identify a person of interest [5]. In the majority of jurisdictions, STRs have been the only genetic marker acknowledged by the Court as a sound method of DNA profiling and able to be used as evidence [6]. As a result, law enforcement databases have been populated with evidentiary and reference STR profiles [7]. STRs are also used routinely for the identification of human remains in coronial investigations, incorporating missing persons and disaster victim identification (DVI) efforts [8,9].

Single nucleotide polymorphisms (SNPs) are single base genetic variants [10]. In the human genome, the average person will have approximately 5 million SNPs [2,11]. For the last 20 years, SNPs have been investigated as an alternative marker to STRs, but SNP genotyping is yet to become common practice in forensic genetic laboratories [12–14]. As biallelic genetic markers such as SNPs are less polymorphic, their discrimination power is significantly lower than that for STRs [9,15]. In order to produce profiles with a similar discrimination power, larger SNP panels are required [16]. Kidd et al. proposed that at least 45 SNPs would have the equivalent discrimination power of the 13 CODIS STRs [15]. However, the mutation rates for autosomal SNPs are approximately 100,000 times lower than in STRs (10^{-8}, compared to 10^{-3}), making SNP genotypes more stable than STRs across multiple generations and thereby reducing the risk of mutations confounding

* Corresponding author at: National DNA Program for Unidentified and Missing Persons, Australian Federal Police, Australia.
E-mail address: Jessica.Watson@student.uts.edu.au (J.L. Watson).

https://doi.org/10.1016/j.fsigen.2024.103169
Received 5 May 2024; Received in revised form 10 October 2024; Accepted 27 October 2024
Available online 28 October 2024

typing [13,15]. SNPs are less prone to degradation due to their shorter amplicon sizes compared to STRs.

The most common genotyping technology currently available for STRs involves fragment length analysis. The targeted DNA regions are amplified with primers containing fluorescent dye labels. When the DNA fragments are separated by size during capillary electrophoresis (CE), fluorescent imaging generates an electropherogram consisting of fluorescent signals (peaks) representing the alleles within each dye channel [17,18]. Whereas CE only provides information on the size of DNA fragments, massively parallel sequencing (MPS) is capable of determining the actual DNA sequence [19]. MPS is a genotyping capability that sequences millions of DNA fragments from multiple samples in one sequencing run. This technology can be applied to both STR and SNP genotyping.

Identity-informative SNPs (iiSNPs) are a category of SNP that have characteristics most beneficial for individualisation of a genotype. iiSNPs require high heterozygosity and should also ideally have low allele frequency heterogeneity to minimise the difference in allele frequencies between populations [20]. SNP allele frequency databases have currently been developed for population and subpopulation groups in Europe, Asia, North America and South America [21–29].

Allele frequency databases are important to develop as these frequencies will tend to vary in different populations as a result of genetic drift [30]. The allele frequencies, reflecting the genetic diversity within a population, can be used to determine the random match probability (RMP) of a particular genotype in a forensic investigation [4]. This can be used to calculate a likelihood ratio (LR), the most commonly employed statistical method to compare an unknown and known DNA profile [31]. This is a ratio of two conditional probabilities for the same observations under alternative hypotheses [32]. It is also possible to calculate a combined LR by combining multiple DNA marker panels [33,34].

In order to be suitable for forensic use, identity-informative markers should: 1) be in Hardy-Weinberg equilibrium (HWE) to ensure that genotype frequencies can be inferred from allele frequencies (within locus independence); 2) be in linkage equilibrium (LE) to ensure that locus genotypes are independently inherited and that LRs from individual loci can be multiplied together (between locus independence); 3) have high heterozygosity to increase the discrimination power of the panel; and 4) the first three conditions should apply across subpopulations [35]. If a locus has significant deviation from HWE, it means a process is influencing the distribution of alleles and genotype frequencies within a population (e.g. inbreeding, hidden population structures, natural selection) [36]. LE tests assess the probability that the alleles of any two loci are inherited independently as a result of recombination and are usually influenced by the physical proximity between the pair [37].

Taylor et al. (2017) published STR allele frequencies in Australian and New Zealand populations for the GlobalFiler™ PCR Amplification Kit (Thermo Fisher Scientific, Waltham, MA, USA), including those whose members declared Aboriginal, European and Asian ancestries [30]. However, there has been no published study on the suitability of a SNP panel for population groups relevant to Australia or establishment of a SNP allele frequency database for this region to date. Furthermore, studies have primarily examined each class of identity marker separately and not assessed the LE between SNPs and STRs or the power of discrimination for a profile containing both marker types [21,23,33,38]. In this study, we examine the suitability of the 94 iiSNPs included in the ForenSeq® DNA Signature Prep Kit (Verogen, Inc., San Diego, CA, USA; now a QIAGEN company) and the ForenSeq® Kintelligence Kit (Verogen, Inc.) for use in the Australian population with European ancestry.

2. Methods

2.1. Ethics approval and sample procurement

Ethics approval for this research was granted by the University of Technology Sydney (UTS) Human Research Ethics Committee (HREC) (UTS HREC NO. ETH21-5821 and amendments ETH21–6606 and ETH23-8117 relate). All volunteers provided a buccal swab with informed consent. A questionnaire was completed to provide self-declared biogeographical ancestry (BGA) for each participant, as well as their parents and grandparents. A total of 105 volunteers with self-declared Australian European ancestry provided self-administered buccal swabs.

Additional casework-type samples (two teeth and eight bones) were sourced from the Australian Facility for Taphonomic Experimental Research with ethics approval (UTS HREC NO. ETH18–2999) and approved research samples submitted to the Australian Federal Police National DNA Program for Unidentified and Missing Persons.

2.2. Sample preparation

DNA from the buccal swabs was manually extracted using the EZ1® DNA Investigator Kit (QIAGEN, Hilden, Germany) [39]. For the bone and tooth samples, 500 mg of pulverised powder underwent total demineralisation lysis, concentration using the Amicon® 30 K Ultra Centrifugal Filter (Sigma-Aldrich, St. Louis, MO, USA) and extraction with the MinElute® PCR Purification Kit (QIAGEN). Samples were quantified with the Quantifiler™ Trio DNA Quantification Kit (Thermo Fisher Scientific) [40] on a QuantStudio™ 5 Real-time PCR System (Thermo Fisher Scientific) [41]. All protocols were performed according to the manufacturers' recommended protocols unless otherwise specified.

2.3. Library preparation and sequencing

There were two MPS panels utilised in this study; the ForenSeq® DNA Signature Prep Kit and the ForenSeq® Kintelligence Kit, each with the same 94 iiSNPs [42,43]. For both kits, the recommended DNA input for library preparation is 1.0 ng. For samples that required dilution and were degraded (with a degradation index (DI) greater than 1), the large autosomal (LA) target concentration was used to determine DNA input. If the samples were not degraded (with a DI equal to or less than 1), the small autosomal (SA) target concentration was used to avoid over-diluting the DNA extract.

For 99 buccal swab samples, libraries were prepared using the ForenSeq® DNA Signature Prep Kit with primer mix B according to either the manufacturer's recommended protocol (n = 42) [42] or an automated library preparation method utilising a quantitative polymerase chain reaction (qPCR) normalisation protocol (n = 57) [44]. Sequencing was performed on the MiSeq® FGx Sequencing System (Verogen, Inc.) using the MiSeq® FGx Reagent Kit (Verogen, Inc.) and standard flow cell according to the manufacturer's recommended protocol [45]. Each ForenSeq® DNA Signature Prep sequencing batch consisted of a positive control (2800 M), negative control and 14 samples. Different index combinations were used on successive sequencing runs to limit sample cross-contamination between batches. The profiles were analysed on the Universal Analysis Software v1.3 (UAS; Verogen, Inc.) using the default analytical and interpretation thresholds and the STR and SNP genotypes exported in Sample Details Reports [46].

For the remaining six buccal swabs and 10 casework-type samples, libraries were prepared following a modified protocol for the ForenSeq® Kintelligence Kit [47]. Sequencing was performed on the MiSeq® FGx Sequencing System using the MiSeq® FGx Reagent Kit and standard flow cell according to the manufacturer's recommended protocol [45]. The ForenSeq® Kintelligence Kit sequencing batches consisted of three samples per flow cell, with a positive control (NA24385) and negative control for each library preparation batch of 12 libraries. The SNP profiles were exported in Sample Reports from the UAS v2.5 and analysed according to the optimised thresholds published by Watson *et al.* (2023) [47,48].

2.4. STR profiling

For buccal swab samples (n = 6) and casework-type samples (n = 10) sequenced with the ForenSeq® Kintelligence Kit, STR profiles were generated using the GlobalFiler™ PCR Amplification Kit [49]. The buccal swab samples were amplified on the Veriti™ 96-Well Fast Thermal Cycler (Thermo Fisher Scientific) with a 29 cycle amplification using the LA target for DNA input. The casework-type samples were amplified with a 30 cycle method. Capillary electrophoresis was performed on the 3500xL Genetic Analyser (Thermo Fisher Scientific) according to the manufacturer's recommended protocols [50]. The resulting electropherogram profiles were analysed using the GeneMapper™ *ID-X* v1.6 software. To distinguish between background noise and the detected peaks in the DNA profile, the analytical threshold was set to 250 relative fluorescence units (RFU) for the buccal swab samples amplified at 29 cycles and 60 RFU for the casework-type samples amplified at 30 cycles. For both amplification methods, the homozygous threshold was set to 1000 RFU.

2.5. Statistical analysis

Statistical analysis was performed on the iiSNPs and a subset of STRs, consisting of the 20 STRs in common between the GlobalFiler™ and ForenSeq® DNA Signature Prep panels. This included D3S1358, vWA, D16S539, CSF1PO, TPOX, D8S1179, D21S11, D18S51, D2S441, D19S433, TH01, FGA, D22S1045, D5S818, D13S317, D7S820, D10S1248, D1S1656, D12S391 and D2S1338. Fig. 1 shows the centimorgan (cM) positions of the STRs and SNPs on each chromosome.

Forensic efficiency parameters and allele frequencies for the iiSNP genotypes were generated using the STR Analysis for Forensics (STRAF) v2.1.5 program [51]. This included calculating the probability of matching (PM), polymorphism information content (PIC), expected heterozygosity (H_{exp}), observed heterozygosity (H_{obs}), power of exclusion (PE), typical paternity index (TPI) and power of discrimination (PD).

The number of possible genotypes (PG) for each locus was calculated based on the number of observed alleles (N) following:

$$PG = \frac{N(N+1)}{2}$$

LE and HWE tests were conducted for all iiSNPs and STRs with the Arlequin v3.5.2.2 software [52]. For both tests, a sequential Bonferroni correction was applied in order to account for false positives that would arise as a result of multiple comparisons [53].

In large populations with minimum inbreeding, the fixation index (F) is approximately equivalent to the coancestry coefficient (θ). θ is the probability that two homologous alleles, one drawn from each of the two individuals, are identical by descent (IBD) [55]. This was calculated for each locus following:

$$\theta \approx F = 1 - \frac{H_{obs}}{H_{exp}}$$

For each of the 10 casework-type samples, an LR was calculated comparing the single-source casework profile and a theoretical matching single source reference profile using the following propositions:

H_1: The DNA originated from the person of interest (POI).
H_2: The DNA originated from someone other than, and unrelated to, the POI in the Australian population with European ancestry.

The Balding-Nichols formulae (National Research Council (NRC) Recommendation 4.2 Equation 4.10a and 4.10b) were used to calculate the match probabilities [35]. The iiSNP allele frequencies determined in this study and the STR frequencies in the Australian European population published by Taylor et al. (2017) were used in a bespoke Excel-based workbook [30]. Single-source LRs were produced for the 20 STRs, 94 iiSNPs and the product of the two calculated for the combined LR for the 114 combined identity markers. A θ of 0.02 was used as recommended by Buckleton et al. (2016) and Taylor et al. (2017) to be consistent with current practice for STRs in the Australian population

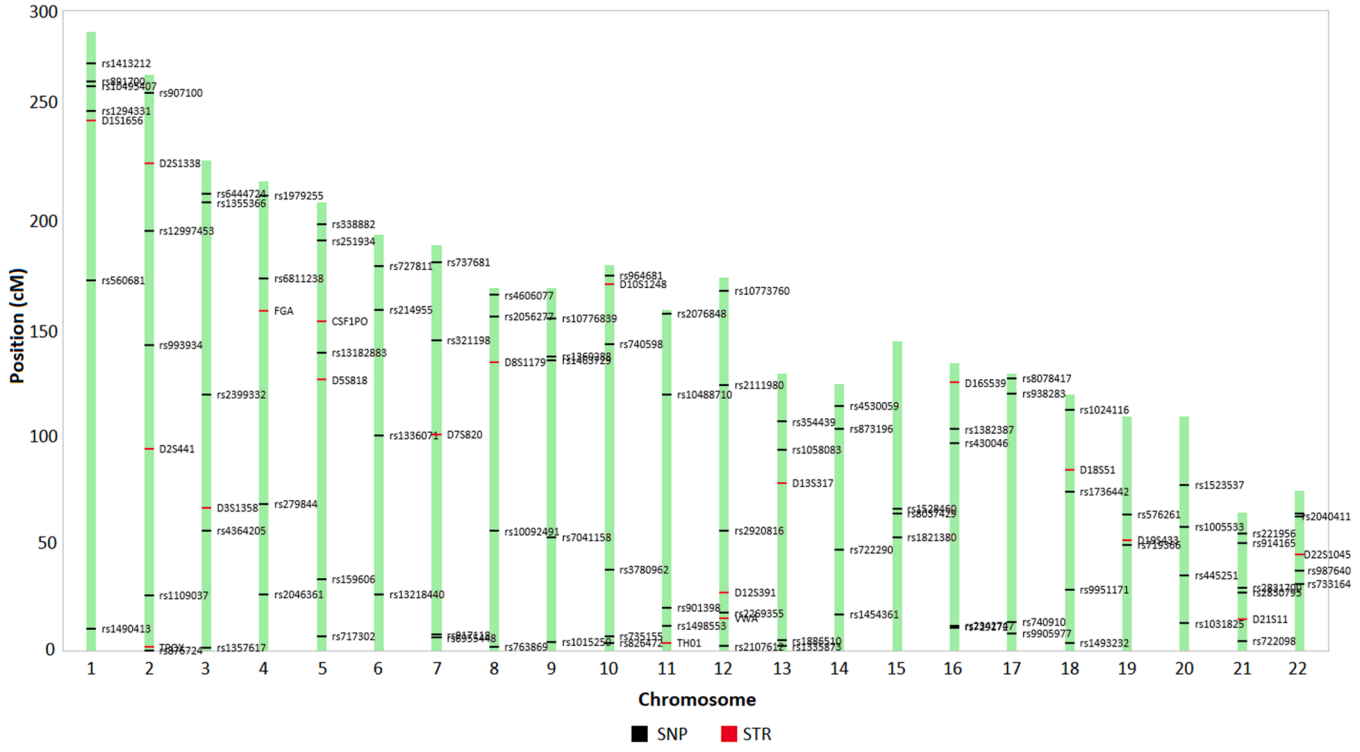


Fig. 1. Chromosome map with the positions of single nucleotide polymorphisms (SNPs; black) and short tandem repeats (STRs; red) in centimorgans (cM), as derived from HapMap [54].

with European ancestry [30,56].

3. Results

3.1. Observed alleles and possible genotypes

For the 105 individuals with self-declared European ancestry that were genotyped, no off-ladder microvariant alleles were observed in the STR profiles or tri-allelic genotypes in the STR and iiSNP profiles. Full STR and iiSNP profiles were obtained for the majority (90 %) of samples; locus dropout of one or two iiSNPs or STRs was observed for nine samples (9 %), and one sample produced a partial profile with combined call rate of 93 %.

As all iiSNPs targeted by the ForenSeq® Kintelligence and ForenSeq® DNA Signature Prep Kits are biallelic (N = 2), there are only three possible genotypes for each locus which indicates a low degree of polymorphism. When all 94 iiSNPs are taken into account, the number of possible genotypes is 7.07×10^{44} (assuming no linkage), enhancing the potential for individualising DNA profiles. However, STRs are highly polymorphic and will have multiple alleles observed for each locus. The total number of observed alleles across the 105 STR profiles was 184 corresponding with a total of 4.30×10^{30} possible genotypes (Table 1). The alleles observed ranged from five alleles for TH01 to 16 alleles for D1S1656.

3.2. Forensic efficiency parameters

The allele frequencies for the iiSNPs in the Australian population with European ancestry are in Table S1. The allele population frequencies for the STRs are in Table S2. The forensic efficiency parameters for all STRs and iiSNPs are detailed in Table S3 and Table S4, respectively. Summaries of the forensic efficiency parameters are provided in Fig. 2 and Table 2.

Although for individual loci the PM was always higher for iiSNPs than for STRs (Fig. 2a), the combined probability of matching (CPM) was calculated to be 1.71×10^{-23} for the 20 STRs and 7.77×10^{-38} for the 94 iiSNPs. Therefore, the probability that any two genotypes will match at all 114 loci within the population is negligible. For the PIC, PE, TPI and PD parameters calculated for all markers, SNPs produced lower values on average when compared to STRs (Fig. 2).

Table 1
Number of observed alleles and possible genotypes for short tandem repeats (STRs) in common between the GlobalFiler™ PCR Amplification Kit and the ForenSeq® DNA Signature Prep Kit.

STR Locus	Observed Alleles (N)	Possible Genotypes (PG)
D3S1358	8	28
vWA	7	21
D16S539	7	21
CSF1PO	8	28
TPOX	6	15
D8S1179	10	45
D21S11	11	55
D18S51	14	91
D2S441	10	45
D19S433	9	36
TH01	5	10
FGA	10	45
D22S1045	7	21
D5S818	7	21
D13S317	7	21
D7S820	9	36
D10S1248	7	21
D1S1656	16	120
D12S391	14	91
D2S1338	12	66
Total	**184**	**4.30×10^{30}**

3.3. Within locus (HWE) and between locus (LE) independence tests

To assess locus independence, HWE tests were conducted for all STRs and iiSNPs. Of the 94 iiSNPs, four loci (rs1357617, rs4374205, rs6955448 and rs1335873) returned p-values below 0.05 that indicated deviation from HWE. However, it is likely that approximately 5 % ($0.05 \times 94 \approx 4.7$) of the total tests will deviate from HWE by chance due to multiple comparisons. After a sequential Bonferroni correction was applied, no significant deviation was observed at any locus. For the STRs, only D8S1179 showed deviation from HWE ($p < 0.05$) and it remained statistically significant after applying the sequential Bonferroni correction, implying that the locus should be excluded when using the STR frequencies derived from this dataset.

LE tests were performed for inter-locus independence on all 6441 pairwise combinations of STRs and iiSNPs, of which 256 were syntenic pairs. There were 379 pairs (5.88 %) that were in disequilibrium ($p < 0.05$) consisting of 243 SNP/SNP pairs, 132 STR/SNP pairs and 4 STR/STR pairs. Of these, 14 pairs were located on the same chromosome with the distance between loci ranging from 7.59 cM (D22S1045/rs987640) to 212.76 cM (rs1355366/rs6444724). However, after a sequential Bonferroni correction was applied, all pairs were found to be in LE.

3.4. Accounting for population sub-structure

In a population with no inbreeding, F would equal 0, indicating that the number of observed and expected heterozygous genotypes are the same. In this study, F was calculated for each STR and iiSNP locus. The overall population had an average F of −0.002 across all loci, suggesting that there is little evidence of population sub-structure [55]. Some variation were observed, with F values ranging from −0.21 (rs1024116) to 0.35 (rs1357617; Fig. 3). However, this variation, with some values falling slightly above or below 0, is likely a result of sampling error. $F \approx \theta$ for large populations with minimal inbreeding, but nevertheless, we recommend a conservative θ correction factor of 0.02 for an Australian population with European ancestry to be consistent with the recommendations of Buckleton et al. (2016) and Taylor et al. (2017) for STRs [30,56].

3.5. Likelihood ratio calculations

The 10 casework samples yielded combined call rates ranging from 96.49 % to 100 %, with locus dropout only occurring for iiSNPs (Table 3). The logarithm of the LRs had an average of 23.70 ± 1.29 for the STR panel and 38.79 ± 1.43 for the iiSNP panel (Fig. 4). As these markers are in LE, inter-locus independence is ensured and the LRs can be combined into an overall LR. The average logarithm of the combined LR was 62.49 ± 1.65, meaning the likelihood of the combined genotype being shared with a randomly selected member of the Australian population with European ancestry is negligible.

4. Discussion

STR allele frequency data for the Australian population has previously been studied using the AmpFlSTR® Profiler Plus and GlobalFiler™ assays [30,57,58]. When compared to the latter and more widely adopted GlobalFiler™ STR study, Taylor *et al.* (2017) collected DNA from seven subpopulation groups in Australia and New Zealand with 2274 samples, of which 528 were identified as “Australian Caucasian” (assumed to be congruent with our DNA donors who declared European ancestry) [30]. Guidelines for genetic population data suggest a minimum of 500 individuals are required to generate reliable allele frequencies due to the high degree of polymorphism in STRs [59]. However, far fewer are required for bi-allelic SNPs which are far less polymorphic. There are dissimilarities between the STR analyses in this study and that by Taylor et al. (2017), likely due to the smaller samples size employed in this study [30]. As such, the Australian Caucasian STR

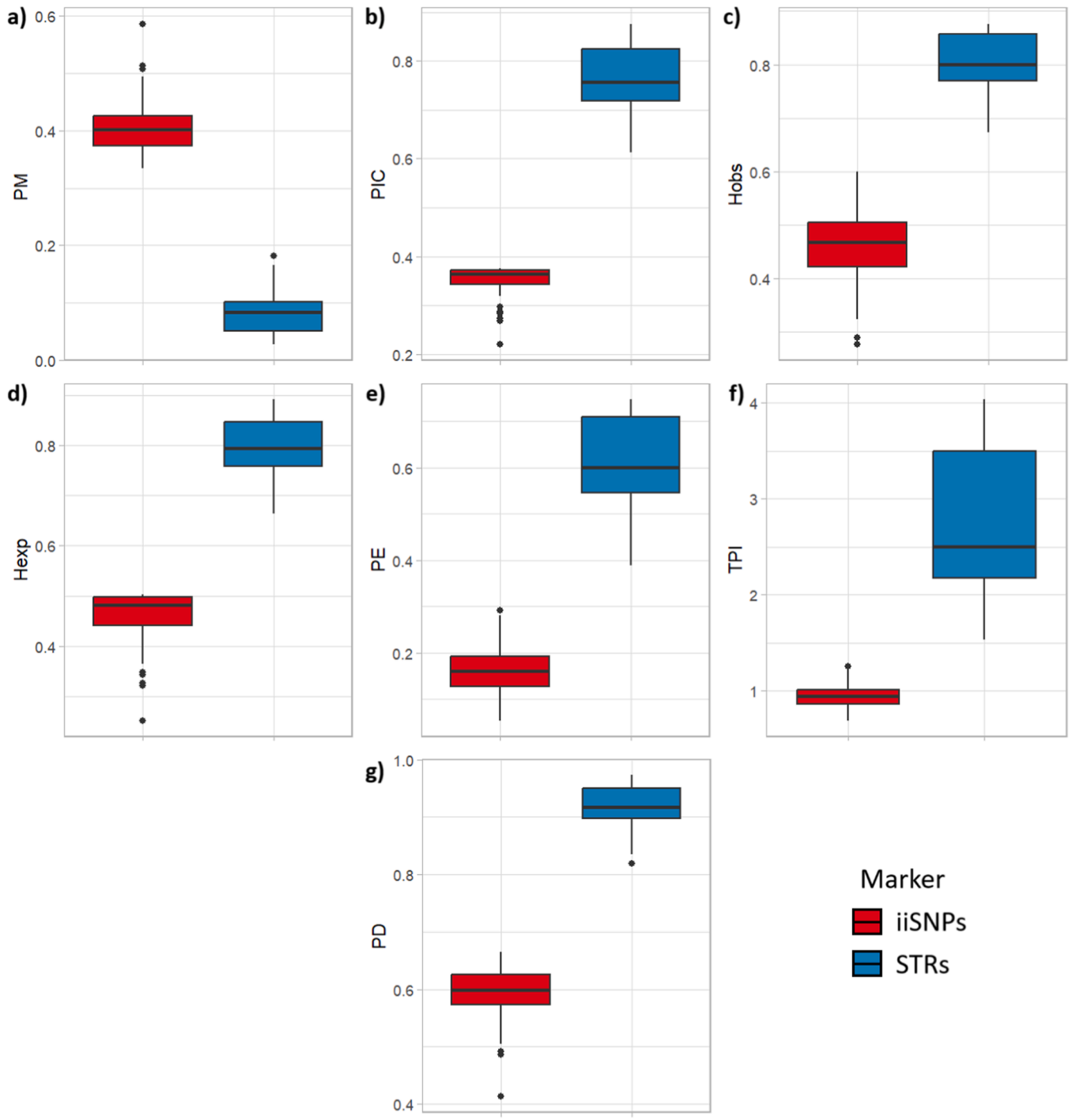


Fig. 2. Forensic efficiency parameters for all individual identity-informative single nucleotide polymorphism (iiSNP) and short tandem repeat (STR) markers: a) probability of matching (PM); b) polymorphism information content (PIC); c) observed heterozygosity (H_{obs}); d) expected heterozygosity (H_{exp}); e) power of exclusion (PE); f) typical paternity index (TPI); and g) power of discrimination (PD).

frequencies published by Taylor et al. (2017) were used for the calculation of LRs in this study (because of the larger sample size) and are recommended for forensic use for the Australian population with European ancestry [30].

The first difference between this study and the Australian Caucasian data by Taylor et al. (2017) was highlighted in the number of observed alleles; an additional 40 alleles were observed by Taylor et al. (2017) across 16 loci [30]. The largest differences were at each of the loci D21S11, D19S433 and FGA, where Taylor *et al.* found seven more alleles than in this study. An additional allele was also observed at D1S1656 (allele 19) in this study that was not observed in the larger published dataset by Taylor et al. (2017) [30]. The added alleles increased the number of possible genotypes to 2.57×10^{40} from 274 observed alleles, compared to the 184 alleles that were observed in this study.

This study demonstrated that established STR allele frequency data generated with CE could be combined with iiSNP data to improve the discriminatory power of a DNA profile. The majority of STR profiles produced in this study were sequenced using MPS technology, for which sequence-based allelic frequencies could have higher discrimination power than the length-based allelic frequencies determined with CE

Table 2
Summary of forensic efficiency parameters calculated for the individual identity-informative single nucleotide polymorphism (iiSNP) and short tandem repeat (STR) markers.

Forensic Efficiency Parameter	STR		iiSNP	
	Minimum	Maximum	Minimum	Maximum
PM	0.030 (D12S391)	0.181 (TPOX)	0.335 (rs1357617)	0.587 (rs938283)
PIC	0.613 (D1S1656)	0.876 (D1S1656 and D12S391)	0.220 (rs938283)	0.375 (13 iiSNPs)
H_{obs}	0.673 (D5S818)	0.876 (D18S51 and D12S391)	0.276 (rs938283)	0.600 (rs1024116)
H_{exp}	0.664 (TPOX)	0.891 (D1S1656 and D12S391)	0.253 (rs938283)	0.500 (13 iiSNPs)
PE	0.388 (D5S818)	0.747 (D18S51 and D12S391)	0.054 (rs938283)	0.291 (rs1024116)
TPI	1.529 (D5S818)	4.038 (D18S51 and D12S391)	0.691 (rs938283)	1.250 (rs1024116)
PD	0.819 (TPOX)	0.973 (D12S391)	0.413 (rs938283)	0.665 (rs1357617)

[60]. Using a panel such as the ForenSeq® DNA Signature Prep Kit allows for the generation of iiSNPs and STRs with sequence-based variation in alleles. However, the use of sequence-based allele frequencies for STRs is not commonly practiced in Australian laboratories at the time of publication.

SNP studies have been published for a number of population and subpopulation groups, but there has not yet been a study on the Australian population [21–29]. While a minimum of 500 individuals is recommended to adequately assess STR allele frequencies, substantially fewer are required for biallelic SNPs [23,27,28,59]. In this study, the most informative loci were rs717302 and rs1498553. The locus rs938283 was the lowest performing in the panel, producing the highest PM and the lowest PIC, H_{obs}, H_{exp}, PE, PD and TPI; these results were congruent with the findings for populations with European ancestry in studies in the United States, France and the United Kingdom [21,22,26].

When assessing intra- and inter-locus independence, all SNPs in this study were in HWE and LE for the Australian population with European ancestry. Furthermore, when combining these identity-informative SNP markers with the existing suite of STRs, pairwise tests showed all 114 loci were in LE. Only one STR, D8S1179, was found to be out of HWE. However, when compared to Taylor *et al.* (2017), there were variances in the p-values produced by up to 0.78 that were likely due to sampling error and D8S1179 was not out of equilibrium [30]. SE33 was not assessed in this study as only a few samples were profiled with the GlobalFiler™ PCR Amplification Kit and SE33 is not included in the ForenSeq® DNA Signature Prep panel. Similarly, D4S2408, D6S1043, D9S1122, D17S1301, D20S482, PentaD and PentaE were also not assessed as these markers were not included in the GlobalFiler™ panel. These STRs would require additional analysis to determine whether they are in LE with each other and with the 94 iiSNPs.

Table 3
Locus call rates (%) of casework-type samples for short tandem repeats (STRs) with the GlobalFiler™ PCR Amplification Kit and identity-informative single nucleotide polymorphisms (iiSNPs) with the ForenSeq® Kintelligence Kit. The call rates for the combined identity markers (20 STRs plus 94 iiSNPs) are also reported.

Sample	STRs (20)	iiSNPs (94)	Combined Identity Markers (114)
Tooth 1	100.0 %	95.7 %	96.5 %
Tooth 2	100.0 %	100.0 %	100.0 %
Bone 1	100.0 %	97.9 %	98.3 %
Bone 2	100.0 %	97.9 %	98.3 %
Bone 3	100.0 %	98.9 %	99.1 %
Bone 4	100.0 %	100.0 %	100.0 %
Bone 5	100.0 %	100.0 %	100.0 %
Bone 6	100.0 %	100.0 %	100.0 %
Bone 7	100.0 %	98.9 %	99.1 %
Bone 8	100.0 %	95.7 %	96.5 %

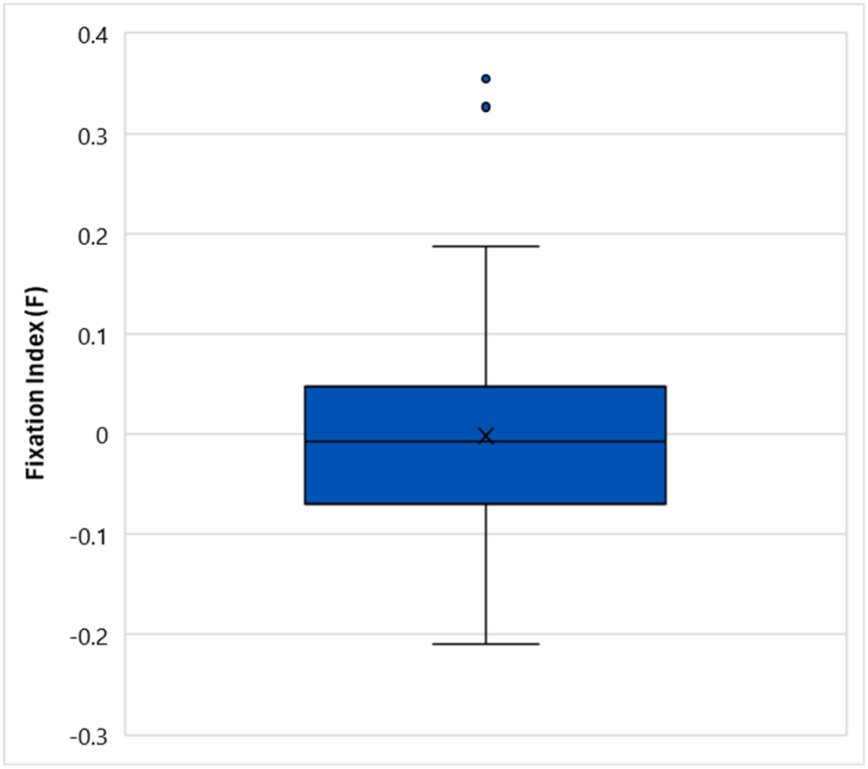


Fig. 3. The fixation index (F) for all short tandem repeat (STR) and identity-informative single nucleotide polymorphism (iiSNP) loci.

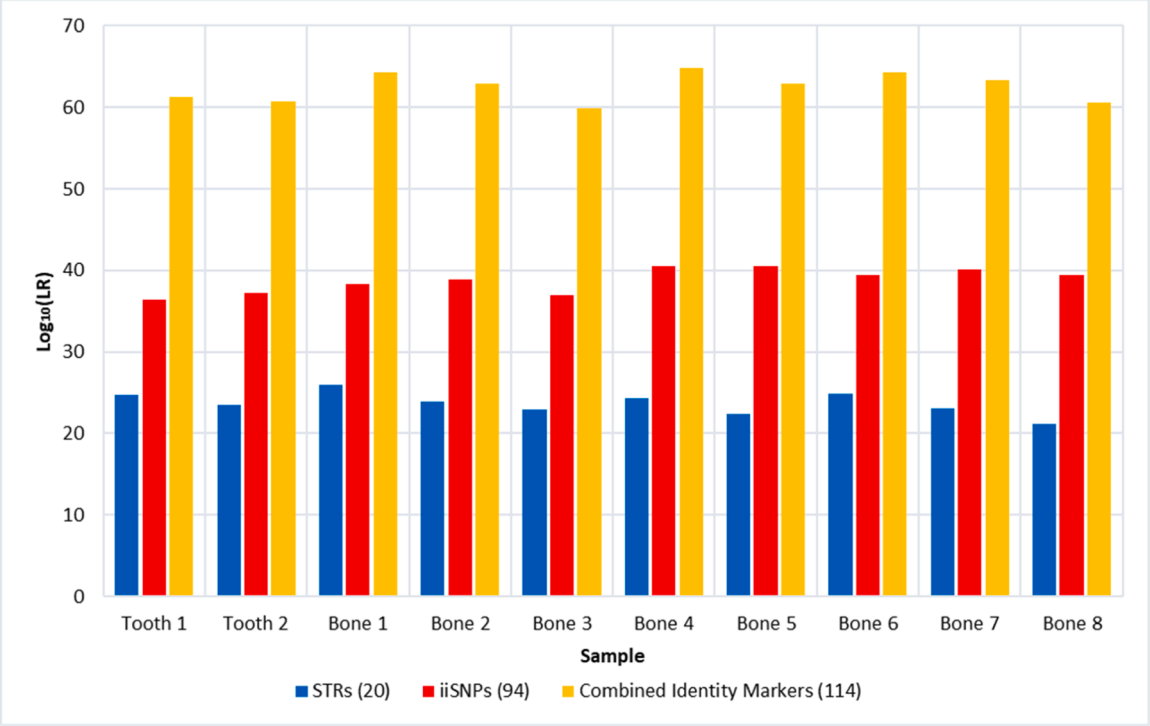


Fig. 4. Logarithm of the likelihood ratios (LR) generated for casework-type samples using short tandem repeats (STRs) with the GlobalFiler™ PCR Amplification Kit, identity-informative single nucleotide polymorphisms (iiSNPs) with the ForenSeq® Kintelligence Kit and the combined identity markers (20 STRs plus 94 iiSNPs).

The overall F for the Australian population with European ancestry was −0.002, indicating there is little evidence of population sub-structure [55]. Regardless, the NRC II recommendations specify the importance of accounting for population sub-structure by applying a θ correction factor between 0.01 (minimal inbreeding) and 0.03 (excess inbreeding) [35]. For STRs, Buckleton *et al.* (2016) and Taylor *et al.* (2017) recommended using a θ correction factor of 0.02 for an Australian population with European ancestry as a conservative measure when calculating the RMP [30,56]. In accounting for possible inbreeding, the θ correction factor raises the RMP and lowers the subsequent LR so as not to over-estimate the weight of the evidence [35]. The application of a conservative θ correction factor of 0.02 for the 94 iiSNPs is consistent with the use of this value for STRs.

The LRs calculated with the iiSNPs for single source profiles were orders of magnitude larger than those calculated with STRs, with a CPM of 10^{-38} for SNPs, compared to 10^{-23} for STRs. This value is similar to that observed by Kiesler *et al.* (2023), who produced a CPM of 10^{-39} for the population with European ancestry in the United States, and Davenport *et al.* (2023), who produced a CPM of 10^{-38} for the "White British" subpopulation in the United Kingdom [26]. Due to their independence, a combined LR can be calculated from the combination of STRs and iiSNPs with a CPM of 10^{-60} which would produce astronomical LRs for complete, matching single source profiles, far beyond the maximum LR reported in Australian forensic laboratories currently (100 billion) [61]. Similar CPMs for combined iiSNP and STR profiles were seen in population studies for French (10^{-69}) and Northeastern Peruvian Andes (10^{-66}) populations [21,23].

The combined LR may be beneficial for samples that produce partial STR profiles where iiSNPs could provide supplementary information to improve discrimination between individuals. The more powerful LRs could also impact kinship calculations, potentially extending the applicability of STRs beyond first order relationships (i.e. parent/offspring and full siblings) if their relatively high mutation rates can be accounted for [33,38].

5. Conclusions

This study has confirmed the forensic applicability of the 94 ForenSeq® iiSNPs in the Australian population with European ancestry, as well as the combined power of identity markers consisting of both iiSNPs and STRs for improved discrimination between individuals for forensic casework. By themselves, iiSNPs can produce LRs that exceed those produced with the established STRs, and their combined power may aid in identifying persons of interest through indirect matching to their genetic relatives or from challenging or compromised samples that have produced suboptimal partial profiles. This study has facilitated the creation of a SNP allele frequency database in Australia, starting with individuals of European ancestry. In order to expand the potential uses of iiSNP markers in routine casework, these loci should be evaluated in other Australian subpopulations including those with Aboriginal and Torres Strait Islander and Asian self-declared ancestries.

CRediT authorship contribution statement

Jodie Ward: Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Dennis McNevin:** Writing – review & editing, Supervision, Methodology, Formal analysis, Conceptualization. **Kaymann Cho:** Writing – review & editing, Methodology. **Jessica L. Watson:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Kelly Grisedale:** Writing – review & editing, Methodology, Formal analysis.

Informed Consent statement

Written informed consent was obtained from all volunteer sample donors involved in this study.

Institutional Review Board Statement

The study was conducted according to the guidelines of the

Declaration of Helsinki and approved by the Human Research Ethics Committee (HREC) of the University of Technology Sydney (UTS) (UTS HREC NO. ETH21–5821, ETH21–6606 and ETH23–8117).

Funding

This research was funded by the Australian Federal Police (AFP) National DNA Program for Unidentified and Missing Persons, AFP Innovation Fund and AFP Forensics. Jessica Watson is supported by an Australian Government Research Training Program (RTP) Scholarship.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank the volunteer sample donors from the Australian Federal Police and the University of Technology Sydney for providing their DNA and associated meta data for this study.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.fsigen.2024.103169.

Data availability

Data are stored at the Australian Federal Police and may be made available to approved entities upon written request and subject to ethics and consent provisions.

References

[1] P. Gill, P.L. Ivanov, C. Kimpton, R. Piercy, N. Benson, G. Tully, et al., Identification of the remains of the Romanov family by DNA analysis, Nat. Genet. 6 (2) (1994) 130–135.
[2] Angers A., Kagkli D., Oliva L., Petrillo M., Raffael B. Study on DNA Profilng Technology for Its Implementation in Central Schengen Information System: EUR29766 Publications Office of the EU, Luxembourg; 2019. Available from: ⟨https://op.europa.eu/en/publication-detail/-/publication/7ae614b0-b274-11e9-9d01-01aa75ed71a1/language-en⟩.
[3] J.L. Weber, C. Wong, Mutation of human short tandem repeats, Hum. Mol. Genet. 2 (8) (1993) 1123–1128, https://doi.org/10.1093/hmg/2.8.1123.
[4] J.M. Butler, Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers, Elsevier, 2005.
[5] R. Puch-Solis, S. Pope, Interpretation of DNA data within the context of UK forensic science—evaluation, Emerg. Top. Life Sci. (2021), https://doi.org/10.1042/ETLS20200340.
[6] Australian Law Reform Commission. Report 96: Essentially yours - the protection of human genetic information in Australia. 2003.
[7] J.W. Hazel, E.W. Clayton, B.A. Malin, C. Slobogin, Is it time for a universal genetic forensic database? Science 362 (6417) (2018) 898–900, https://doi.org/10.1126/science.aav5475.
[8] C.H. Brenner, B.S. Weir, Issues and strategies in the DNA identification of World Trade Center victims, Theor. Popul Biol. 63 (3) (2003) 173–178, https://doi.org/10.1016/S0040-5809(03)00008-X.
[9] E.S.M. Iwamura, M.A. Guimarães, M.P. Evison, DNA methods to identify missing persons, in: S.J. Morewitz, C. Sturdy Colls (Eds.), Handbook of Missing Persons, Springer International Publishing, Cham, 2016, pp. 337–352.
[10] B. Budowle, A. Van Daal, Forensically relevant SNP classes, Biotechniques 44 (5) (2008) 603–610, https://doi.org/10.2144/000112806.
[11] R. Decorte, Genetic identification in the 21st century—current status and future developments, Forensic Sci. Int. 201 (1-3) (2010) 160–164.
[12] P. Gill, An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes, Int. J. Leg. Med. 114 (4-5) (2001) 204–210, https://doi.org/10.1007/s004149900117.
[13] J.M. Butler, M.D. Coble, P.M. Vallone, STRs vs. SNPs: thoughts on the future of forensic DNA testing, Forensic Sci., Med. Pathol. 3 (3) (2007) 200–205, https://doi.org/10.1007/s12024-007-0018-1.
[14] K.M. Canturk, R. Emre, B. Buşpinar, M. Ozen, Current status of the use of single-nucleotide polymorphisms in forensic practices, Genet. Test. Mol. Biomark. 18 (7) (2014) 455–460, https://doi.org/10.1089/gtmb.2013.0466.
[15] K.K. Kidd, A.J. Pakstis, W.C. Speed, E.L. Grigorenko, S.L.B. Kajuna, N.J. Karoma, et al., Developing a SNP panel for forensic identification of individuals, Forensic Sci. Int. 164 (1) (2006) 20–32, https://doi.org/10.1016/j.forsciint.2005.11.017.
[16] S.B. Seo, J.L. King, D.H. Warshauer, C.P. Davis, J. Ge, B. Budowle, Single nucleotide polymorphism typing with massively parallel sequencing for human identification, Int J. Leg. Med 127 (6) (2013) 1079–1086, https://doi.org/10.1007/s00414-013-0879-7.
[17] B.C. Durney, C.L. Crihfield, L.A. Holland, Capillary electrophoresis applied to DNA: determining and harnessing sequence and structure to advance bioanalyses (2009–2014), Anal. Bioanal. Chem. 407 (23) (2015) 6923–6938, https://doi.org/10.1007/s00216-015-8703-5.
[18] J.G. Shewale, L. Qi, L.M. Calandro, Principles, practice, and evolution of capillary electrophoresis as a tool for forensic DNA analysis, in: J.G. Shewale, Q. Liu (Eds.), Forensic DNA Analysis: Current Practices and Emerging Technologis, CRC Press, 2012, p. 131.
[19] Illumina. An Introduction to Next-Generation Sequencing Technology 2017. [Last accessed: 27 May]. Available from: ⟨https://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf⟩.
[20] Y.-L. Wei, C.-X. Li, J. Jia, L. Hu, Y. Liu, Forensic Identification using a multiplex assay of 47 SNPs*, J. Forensic Sci. 57 (6) (2012) 1448–1456, https://doi.org/10.1111/j.1556-4029.2012.02154.x.
[21] A. Delest, D. Godfrin, Y. Chantrel, A. Ulus, J. Vannier, M. Faivre, et al., Sequenced-based French population data from 169 unrelated individuals with Verogen's ForenSeq DNA signature prep kit, Forensic Sci. Int. Genet. 47 (2020) 102304, https://doi.org/10.1016/j.fsigen.2020.102304.
[22] L. Davenport, L. Devesse, D.S. Court, D. Ballard, Forensic identity SNPs: Characterisation of flanking region variation using massively parallel sequencing, Forensic Sci. Int. Genet. 64 (2023) 102847, https://doi.org/10.1016/j.fsigen.2023.102847.
[23] E.K. Guevara, J.U. Palo, J.L. King, M.M. Buś, S. Guillen, B. Budowle, et al., Autosomal STR and SNP characterization of populations from the Northeastern Peruvian Andes with the ForenSeq™ DNA Signature Prep Kit, Forensic Sci. Int. Genet. 52 (2021) 102487, https://doi.org/10.1016/j.fsigen.2021.102487.
[24] R. Tao, X. Dong, X. Zhen, R. Xia, Y. Qu, S. Liu, et al., Population genetic analyses of Eastern Chinese Han nationality using ForenSeq™ DNA Signature Prep Kit, Mol. Genet. Genom. 299 (1) (2024) 9, https://doi.org/10.1007/s00438-024-02121-w.
[25] J.A. Aguilar-Velázquez, M.Á. Duran-Salazar, M.F. Córdoba-Mercado, C. E. Coronado-Avila, O. Salas-Salas, G. Martinez-Cortés, et al., Characterization of 58 STRs and 94 SNPs with the ForenSeq™ DNA signature prep kit in Mexican-Mestizos from the Monterrey city (Northeast, Mexico), Mol. Biol. Rep. 49 (8) (2022) 7601–7609, https://doi.org/10.1007/s11033-022-07575-y.
[26] K.M. Kiesler, L.A. Borsuk, C.R. Steffen, P.M. Vallone, K.B. Gettings, US population data for 94 identity-informative SNP loci, Genes 14 (2023) 1071, https://doi.org/10.3390/genes14051071.
[27] I. Grandell, R. Samara, A. Tillmar, A SNP panel for identity and kinship testing using massive parallel sequencing, Int. J. Leg. Med. 130 (2016) 905–914, https://doi.org/10.1007/s00414-016-1341-4.
[28] D. Augustinus, M.E. Gahan, D. McNevin, Development of a forensic identity SNP panel for Indonesia, Int J. Leg. Med. 129 (2015) 681–691, https://doi.org/10.1007/s00414-014-1055-4.
[29] F. Guo, Z. Liu, G. Long, B. Zhang, D. Liu, S. Yu, Performance and characterization of 94 identity-informative SNPs in Northern Han Chinese using ForenSeq™ DNA Signature Prep Kit, J. Forensic Leg. Med 103 (2024) 102678, https://doi.org/10.1016/j.flm.2024.102678.
[30] D. Taylor, J.-A. Bright, C. McGovern, S. Neville, D. Grover, Allele frequency database for GlobalFiler™ STR loci in Australian and New Zealand populations, Forensic Sci. Int. Genet. 28 (2017) e38–e40, https://doi.org/10.1016/j.fsigen.2017.02.012.
[31] P. Gill, T. Hicks, J.M. Butler, E. Connolly, L. Gusmão, B. Kokshoorn, et al., DNA commission of the International society for forensic genetics: Assessing the value of forensic biological evidence-Guidelines highlighting the importance of propositions: Part I: evaluation of DNA profiling comparisons given (sub-) source propositions, Forensic Sci. Int. Genet. 36 (2018) 189–202, https://doi.org/10.1016/j.fsigen.2018.07.003.
[32] B.S. Weir, A.D. Anderson, A.B. Hepler, Genetic relatedness analysis: modern data and new challenges, Nat. Rev. Genet. 7 (10) (2006) 771–780, https://doi.org/10.1038/nrg1960.
[33] I.Z. Pajnič, T. Leskovar, M. Črešnar, Improving kinship probability in analysis of ancient skeletons using identity SNPs and MPS technology, Int. J. Leg. Med. 137 (2023) 1007–1015, https://doi.org/10.1007/s00414-023-03003-3.
[34] A. Angers, J. Drabek, M. Fabbri, M. Petrillo, M. Querci, whole genome sequencing and forensics genomics, Jt. Res. Cent. (2021), https://doi.org/10.2760/864087.
[35] National Research Council Committee on DNA Forensic Science, The Evaluation of Forensic DNA Evidence, National Academies Press (US), Washington (DC), 1996, https://doi.org/10.17226/5141.
[36] J.F. Crow, M. Kimura, An Introduction to Population Genetics Theory, Harper and Row, New York, 1970.
[37] A. Tillmar, C. Phillips, Evaluation of the impact of genetic linkage in forensic identity and relationship testing for expanded DNA marker sets, Forensic Sci. Int. Genet. 26 (2017) 58–65, https://doi.org/10.1016/j.fsigen.2016.10.007.
[38] C. Phillips, M. Fondevila, M. García-Magariños, A. Rodriguez, A. Salas, Á. Carracedo, et al., Resolving relationship tests that show ambiguous STR results using autosomal SNPs as supplementary markers, Forensic Sci. Int. Genet. 2 (3) (2008) 198–204, https://doi.org/10.1016/j.fsigen.2008.02.002.
[39] QIAGEN. EZ1&2 DNA Investigator Kit. 2022.

[40] Thermo Fisher Scientific. Quantifiler™ HP and Trio DNA Quantification Kits User Guide [User Guide]. 2017.
[41] Thermo Fisher Scientific, QuantStudio™ 5 Real-Time PCR Instrument(for Human Identification) [User Guide], Life Technol. (2017).
[42] Verogen. ForenSeq DNA Signature Prep Reference Guide. ForenSeq DNA Signature Prep Kit. 2020.
[43] Verogen. ForenSeq Kintelligence Kit Reference Guide. ForenSeq Kintelligence Kit. 2021.
[44] B. Mehta, S. Venables, P. Roffey, Comparison between magnetic bead and qPCR library normalisation methods for forensic MPS genotyping, Int. J. Leg. Med. 132 (2018) 125–132, https://doi.org/10.1007/s00414-017-1591-9.
[45] Verogen, MiSeq FGx Sequencing System Reference Guide, MiSeq FGx Seq. Syst. (2021).
[46] Verogen, Universal Analysis Software v1.3, Reference Guide, Universal Analysis Software, 2021.
[47] J.L. Watson, D. McNevin, K. Grisedale, M. Spiden, S. Seddon, J. Ward, Operationalisation of the ForenSeq® Kintelligence Kit for Australian Unidentified and Missing Persons Casework, Forensic Sci. Int. Genet. (2023) 102972, https://doi.org/10.1016/j.fsigen.2023.102972.
[48] Verogen. Universal Analysis Software v2.0 and above, Reference Guide. Universal Analysis Software. 2021.
[49] Thermo Fisher Scientific. GlobalFiler PCR Amplification Kits. 2019.
[50] Thermo Fisher Scientific. Applied Biosystems 3500/3500xL Genetic Analyzer User Guide. 2010.
[51] A. Gouy, M. Zieger, STRAF—a convenient online tool for STR data evaluation in forensic genetics, Forensic Sci. Int. Genet. 30 (2017) 148–151, https://doi.org/10.1016/j.fsigen.2017.07.007.
[52] L. Excoffier, H.E. 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.
[53] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. R. Stat. Soc.: Ser. B (Methodol. 57 (1) (1995) 289–300, https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.
[54] The International HapMap Consortium, The International HapMap project, Nature 426 (2003) 789–796, https://doi.org/10.1038/nature02168.
[55] B.S. Weir, C.C. Cockerham, Estimating F-statistics for the analysis of population structure, Evolution 38 (6) (1984) 1358–1370, https://doi.org/10.2307/2408641.
[56] J. Buckleton, J. Curran, J. Goudet, D. Taylor, A. Thiery, B.S. Weir, F.S. T. Population-specific, values for the forensic STR markers: a worldwide survey, Forensic Sci. Int. Genet. 23 (2016) 91–100, https://doi.org/10.1016/j.fsigen.2016.03.004.
[57] S.J. Walsh, J.S. Buckleton, Autosomal microsatellite allele frequencies for a nationwide dataset from Australian Caucasian sub-population, Forensic Sci. Int. 168 (2-3) (2007) e47–e50, https://doi.org/10.1016/j.forsciint.2007.01.030.
[58] S.J. Walsh, J.S. Buckleton, Autosomal microsatellite allele frequencies for 15 regionally defined Aboriginal Australian population datasets, Forensic Sci. Int. 168 (2-3) (2007) e29–e42, https://doi.org/10.1016/j.forsciint.2007.01.029.
[59] Á. Carracedo, J.M. Butler, L. Gusmão, A. Linacre, W. Parson, L. Roewer, et al., New guidelines for the publication of genetic population data, Forensic Sci. Int. Genet. 7 (2) (2013) 217–220, https://doi.org/10.1016/j.fsigen.2013.01.001.
[60] A. Alonso, P.A. Barrio, P. Müller, S. Köcher, B. Berger, P. Martin, et al., Current state-of-art of STR sequencing in forensic genetics, Electrophoresis 39 (21) (2018) 2655–2668, https://doi.org/10.1002/elps.201800030.
[61] Biology Specialist Advisory Group. Biology Specialist Advisory Group (BSAG) Statistics Scientific Working Group (StatSWG) recommendations for the interpretation of DNA, Version 3. BSAG StatSWG. 2016.