Forensic Science International 364 (2024) 112208

Contents lists available at ScienceDirect

Forensic Science International

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

It's all relative: A multi-generational study using ForenSeq™ Kintelligence

R. Daniel [a], J. Raymond [b], A. Sears [b], A. Stock [a], N. Scudder [c], G. Padmabandu [d], S.A. Kumar [d], J. Snedecor [d], J. Antunes [d], D. Hartman [a,e,*]

[a] Victorian Institute of Forensic Medicine, Victoria, Australia
[b] Forensic Evidence and Technical Services, New South Wales Police Force, Sydney, Australia
[c] Australian Federal Police, Canberra, Australian Capital Territory, Australia
[d] Qiagen HID LLC, Germantown, MD, United States
[e] Department of Forensic Medicine, Monash University, Victoria, Australia

ARTICLE INFO

Keywords:
Forensic DNA analysis
Forensic Investigative Genetic Genealogy
DNA Intelligence
Kintelligence
Kinship

ABSTRACT

The successful application of Forensic Investigative Genetic Genealogy (FIGG) to the identification of unidentified human remains and perpetrators of serious crime has led to a growing interest in its use internationally, including Australia. Routinely, FIGG has relied on the generation of high-density single nucleotide polymorphism (SNP) profiles from forensic samples using whole genome array (WGA) (~650,000 or more SNPs) or whole genome sequencing (WGS) (millions of SNPs) for DNA segment-based comparisons in commercially available genealogy databases. To date, this approach has required DNA of a quality and quantity that is often not compatible with forensic samples. Furthermore, it requires the management of large data sets that include SNPs of medical relevance. The ForenSeq™ Kintelligence kit, comprising of 10,230 SNPs including 9867 for kinship association, was designed to overcome these challenges using a targeted amplicon sequencing-based method developed for low DNA inputs, inhibited and/or degraded forensic samples. To assess the ability of the ForenSeq™ Kintelligence workflow to correctly predict biological relationships, a comparative study comprising of 12 individuals from a family (with varying degrees of relatedness from 1st to 6th degree relatives) was undertaken using ForenSeq™ Kintelligence and a WGA approach using the Illumina Global Screening Array-24 version 3.0 Beadchip. All expected 1st, 2nd, 3rd, 4th and 5th degree relationships were correctly predicted using ForenSeq™ Kintelligence, while the expected 6th degree relationships were not detected. Given the (often) limited availability of forensic samples, findings from this study will assist Australian Law enforcement and other agencies considering the use of FIGG, to determine if the ForenSeq™ Kintelligence is suitable for existing workflows and casework sample types considered for FIGG.

1. Introduction

The field of Forensic DNA Intelligence (FDI) relies on generating investigative leads from biological evidence in criminal or coronial investigations where forensic STR profiling has not resulted in the identification of the DNA donor. FDI, as a capability, was enabled by the rapid emergence of forensically relevant markers (beyond STRs) such as single nucleotide polymorphisms (SNPs) for identity/kinship analysis, and the inference biogeographical ancestry (BGA) and externally visible characteristics (EVCs) (known as Forensic DNA Phenotyping) of an individual. Increased resolution and accuracy of FDI has been due, in part, to the continuous and parallel evolution of genotyping technologies such as microarrays and sequencing which allow whole genome analysis of hundreds to millions of SNPs simultaneously [1]. Forensic Investigative Genetic Genealogy (FIGG) is the latest addition to the ever-expanding repertoire of FDI capabilities.

FIGG combines DNA testing and traditional genealogical methods to generate investigative leads from forensic samples to identify perpetrators of serious crime and for the identification of unidentified human remains (UHRs). To date, FIGG has relied on high-density SNP data generated from biological evidence uploaded to commercially available genealogy databases to perform familial searching to identify close (1st and 2nd degree) relatives [2–6] or more distant relatives (such as 3rd, 4th or 5th degree) of the DNA donor. In contrast, forensic STR profiling only allows direct, or familial, matching to close relatives (1st and occasionally 2nd degree) [7].

* Corresponding author at: Victorian Institute of Forensic Medicine, Victoria, Australia.
E-mail address: dadna.hartman@vifm.org (D. Hartman).

https://doi.org/10.1016/j.forsciint.2024.112208
Received 23 June 2024; Received in revised form 5 August 2024; Accepted 24 August 2024
Available online 27 August 2024

Potential relatives of the donor in the genealogy databases are determined using the DNA segment sharing, or segment matching, approach to estimate relatedness or kinship [8]. Segment matching is based on the principle that more DNA will be shared between closely related individuals compared to distant relatives and that SNPs that are identical-by-decent (IBD) are more likely to be inherited in contiguous blocks. The method analyses long stretches of identical shared alleles between two individuals, and the total distance covered by the matching segments (often in the hundreds) are the Total Shared centiMorgans (cMs) [9]. The possible familial matches are analysed to determine potential relationships using genealogical and research methods. Family trees are built to identify potential candidates, i.e., generate investigative leads, to assist with identifying the donor of the biological evidence. Hundreds of cases internationally [10–13] (predominantly in the USA) have utilised FIGG to generate leads to identity UHRs and the perpetrators of serious (or violent) crimes [2]. Given its demonstrated forensic and investigative utility, and the growing number of SNP profiles generated by the Direct-To-Consumer (DTC) companies (more than 26 million SNP profiles generated since 2019 [5]), FIGG is now a methodology to be considered when current forensic STR database matching and other established means of identification have been exhausted [4,6].

Until recently, generating SNP data for FIGG has been achieved through Whole Genome Array (WGA) [3] or Whole Genome Sequencing (WGS) [11]. However, limitations are encountered with FIGG due to the nature of biological evidence. Forensic samples retrieved from crime scenes or UHRs are problematic when the recovery of high yields of good quality DNA are needed for WGA and less so for WGS [11,14,15]. Furthermore, such samples are often limited in nature or contain environmental inhibitors, minimising the opportunity for repeat testing should one approach fail to generate results. Alternatives that aim to optimise SNP analysis for forensic samples have been developed. The FORensic Capture Enrichment (FORCE) panel is an all-in-one SNP panel (5422 SNP markers) covering a range of forensically relevant SNPs (identity, ancestry, phenotype, X- and Y-chromosomal SNPs) as well as kinship SNPs for the inference of distant relationships [16]. The ForenSeq™ Kintelligence kit (QIAGEN, Hilden, Germany) released by Verogen (CA, USA) in 2019/20, was developed as a fit-for-purpose, targeted massively parallel sequencing (MPS) FIGG assay to classify kinship from 1st to 4th and 5th degree relatives (such as Great-Great-Grandparent/child or first cousin once removed) with the highest accuracy for the detection of 1st to 4th degree relatives [9]. The kit is optimised for low quantity (1 ng DNA input requirement) and degraded forensic samples facilitated by the primer design (98 % of the amplicons are less than 150 bp in length) and a buffer system that tolerates a number of inhibitors, such as calcium, indigo and humic acid, commonly found in forensic samples [17]. Of the 10,230 SNPs in ForenSeq™ Kintelligence, 9867 (96 % of the total SNPs) are kinship-informative SNPs selected from commercial whole genome arrays commonly used by DTCs such as the Infinium CytoSNP-850 K BeadChip and Global Screening Array (GSA) (Illumina, Inc., San Diego, CA) [18]. The SNPs were selected to maximise the overlap with the SNPs utilised in profile comparisons in genealogy databases (to generate lists of related individuals) and were specifically curated for kinship analysis across global populations [9]. Maximal spacing of the kinship SNPs across the genome assists with minimising linkage effects. Medically relevant and minor allele frequency SNPs were excluded from ForenSeq™ Kintelligence to minimise privacy concerns. The remaining 363 SNPs (which include X (n=106) and Y (n=85) SNPs) enable BGA (aSNPs) (n=56) and EVC (pSNPs) (hair and eye colour) (n=22*) inferences, identity (n=94) analysis and determination of biological sex to be performed simultaneously (*two SNPs overlap the BGA and EVC categories but are only included in the EVC category) [9]. In addition to the ForenSeq™ Kintelligence BGA and EVC SNPs, the identity SNPs (iiSNPs) in ForenSeq™ Kintelligence and the ForenSeq DNA Signature Prep Kit are identical enabling cross checking of kinship between both kits. ForenSeq™ Kintelligence libraries occur on the MiSeq FGx® Sequencing System (QIAGEN, Hilden) which has been validated for operational use to generate investigative leads in addition to the ForenSeq Signature Prep kit (QIAGEN, Hilden) [19–22]. The ForenSeq Kintelligence Analysis Module, in the Universal Analysis Software (UAS), is used to analyse the sequencing data from Kintelligence libraries. In addition to converting sequence data into SNP calls, the UAS generates locus call rates and heterozygosity data with quality control indicators to assist the user with data review [18].

Kinship assessments are performed using all autosomal ForenSeq™ Kintelligence SNPs and the One-to-Many Kinship tool within GEDmatch PRO. Using SNPs with maximal discriminatory power and an algorithm designed for a sparser SNP set than from WGAs, performance similar to segment matching may be reached for challenging casework samples. The algorithm locates shared segments using kinship coefficients in 'windows' across the genome rather than stretches of identical SNP allele calls. The windowed kinship approach is a modification of the peer-reviewed PC-AiR and PC-Relate tools for genetic relatedness inference [9,23–25]. This approach controls for the presence of unknown or unspecified population substructure and background frequencies to generate genetic correlations without the need for BGA or reference population information thereby enhancing the discriminatory power of fewer SNPs [9]. The Generation Visual Chart, within GEDmatch PRO, is a schematic adapted from the commonly used Shared cM Project tool within DNA Painter [26]. As with the Shared cM Project tool, the Generation Visual Chart graphically maps potential degrees of relationships to a home person, or kit, and provides information on the ranges of shared cMs for each degree or relationship. The shared cMs are based on converting the kinship coefficient statistics described in Snedecor et al. [9], to cM statistics. An individual, or home, kit's shared cM value with its matches may then be used to predict potential degrees for relationships.

Several validation studies have been performed in accordance with SWGDAM Validation Guidelines for DNA Analysis Methods [27]. Peck *et al.* (2023) [28] performed an internal validation which included sensitivity, precision, accuracy, mixture and contamination studies as well as non-probative forensic-type samples such as bone, fired shell casing and adhesive tape. The results indicated that DNA input amounts as low as 0.05 ng generated accurate profiles and a minor contributor could be detected down to 0.02 ng of DNA. The non-probative samples with DNA inputs ranging from 0.05 to 1 ng generated uploadable profiles to GEDmatch PRO. In addition to international validation, Watson et al. optimised laboratory and bioinformatic analysis methods to operationalise Kintelligence for Australian unidentified and missing persons casework [29].

The ForenSeq™ Kintelligence workflow, including its associated analytical software, is a viable means of generating SNP profiles for FIGG use. At present, several Law Enforcement (LE) and government agencies in Australia are evaluating and applying FIGG in UHR and criminal investigations [6,30]. To inform this evaluation, a pilot study comprising of 12 individuals from one family with varying degrees of relatedness (1st to 6th degree relationships) was conducted to assess the ability of ForenSeq™ Kintelligence to accurately detect and predict the known relationships. In addition to ForenSeq™ Kintelligence data, SNP data was generated using a WGA commonly used for FIGG analysis, the Illumina GSA-24 version 3.0 Beadchip, to compare relationship predictions or estimations. A 'home' person was assigned in two evaluations to assess familial relationships ranging from parent-child to second cousin once removed. The kinship prediction assessments included within (and external to) family study samples using both SNP data sets in GEDmatch PRO and GEDmatch respectively. The focus of this study was on relationship prediction accuracy of ForenSeq™ Kintelligence; therefore, assay performance assessments or validation studies were not undertaken.

2. Methods

2.1. Sample selection

Samples were collected and analysed with approval from the Victorian Institute of Forensic Medicine Ethics Committee, Project 1151. The approval process included considerations regarding data storage and analysis of volunteer samples, as well as privacy and ethical matters. Written consent was sought and obtained from each participant. The participants self-declared their biological sex and biogeographical ancestry (BGA) (three generations). The family study comprised of 12 individuals (samples S1 to S12) whose relationships are depicted in Fig. 1. Participant S11 is not biologically related to any of the other members of the family and was included as negative control for the matches. Following collection, samples were transferred to the Forensic & Analytical Science Service (FASS), New South Wales Health, Australia for DNA extraction.

2.2. DNA extraction and quantification

Two buccal swabs were collected from each participant. DNA was extracted from the buccal swabs using the PrepFiler® Automated Forensic DNA Extraction Kit (ThermoFisher Scientific) by FASS, New South Wales Health, Australia, following the laboratory's standard procedures under conditions specified by the manufacturer, with an elution volume of 100 μL. All necessary controls were included to detect any laboratory contamination. Samples were stored at 4°C until further analysis.

A buccal swab of each participant was subsampled into three sections, and all extracted. These extracts were quantified and then combined and vacuum concentrated (SpeedVac vacuum concentrator, ThermoFisther Scientific), re-quantified and profiled to confirm single source. If sufficient DNA was not obtained, the second buccal swab was subjected to the same procedure. The total six extracts both two swabs were then combined and concentrated after confirming single source

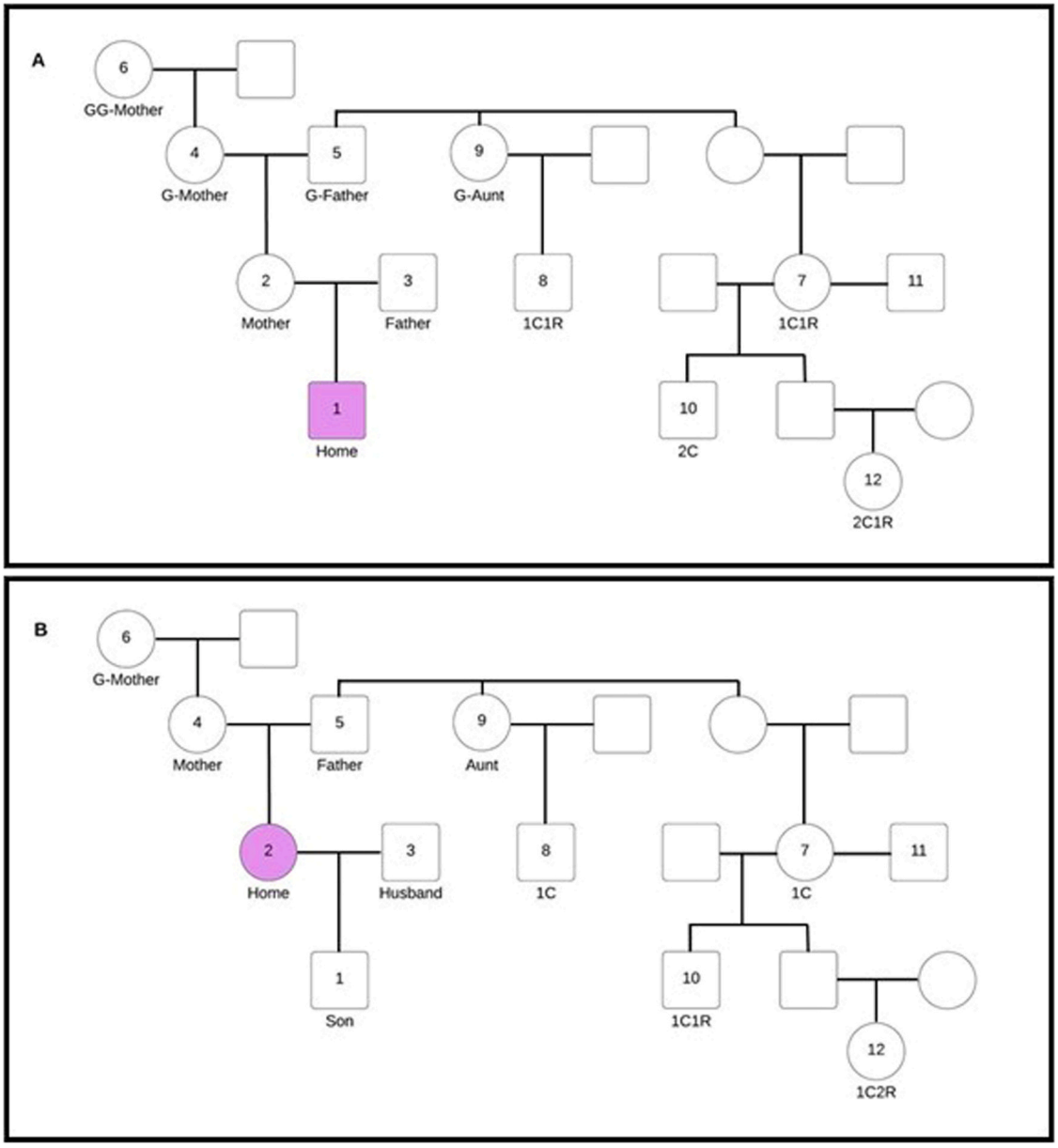

Fig. 1. Family study – Relationships between the 12 members of the family study (Samples 1–12). *Panel A*: The home person is Sample 1 (S1). *Panel B*: The home person is Sample 2 (S2). Abbreviations**:** GG-Mother: great grandmother; G-Mother: grandmother; G-Father: grandfather; G-Aunt: great aunt; 1 C: first cousin; 1C1R: first cousin once removed; 1C2R: first cousin twice removed; 2 C: second cousin; and 2C1R: second cousin once removed.

status. The final elution volume for the combined extracts was ~64 μL. Nuclear DNA concentration was determined using Quantifiler® Trio kit (ThermoFisher Scientific) on a 7500 real-time PCR instrument (ThermoFisher Scientific) by FASS, New South Wales Health, Australia, under conditions specified by the manufacturer.

Aliquots of each of the 12 DNA samples were simultaneously prepared for STR profiling, targeted next generation sequencing and WGA analysis.

2.3. STR amplification and profiling

Aliquots of the 12 DNA samples were amplified by FASS, New South Wales Health, Australia, using the PowerPlex® 21 System (Promega) under conditions specified by the manufacturer, to confirm the contributor status of all samples (i.e., single source) and to demonstrate full STR profiles were obtained.

2.4. Generating SNP data for FIGG

SNP data was generated from the 12 samples using targeted MPS and whole genome array; the ForenSeq® Kintelligence Kit (Verogen) [17], and the Infinium Global Screening Array (GSA)-24 version 3.0 Beadchip (Illumina) [31], respectively, for comparison purposes.

2.4.1. ForenSeq® Kintelligence Kit

Aliquots of each of the 12 extracted DNA samples were provided to Verogen, CA, USA (now QIAGEN), for analysis using the ForenSeq® Kintelligence kit (QIAGEN, Hilden). The DNA extracts were quantified using Quantifluor ONE dsDNA (Promega, WI, USA) and Quantus (Promega) fluorometer.

The samples were diluted to allow 1 ng of DNA to be added to PCR1 of the ForenSeq® Kintelligence kit (QIAGEN, Hilden), according to manufacturer's instructions. Three technical replicates were prepared for each sample, except for sample S6 where due to low concentration, only one replicate was prepared. One nanogram (ng) of high-quality genomic DNA control sample NA24385 (HG002) (Coriell Institute for Medical Research, Camden NJ) (provided in the ForenSeq® Kintelligence kit) was included as Positive Amplification Control, and DNase-RNase free water (MPBiomedicals, CA, USA) used as Negative Amplification Control. The PCR1 master mix for the ForenSeq® Kintelligence kit was prepared according to the manufacturer's recommendations. Twenty-five μL of master mix were added to the wells of a 96-well PCR plate. Twenty-five μL of DNA samples, NA24385 DNA or water were added to the corresponding wells of the PCR plate.

Three libraries including one positive amplification control (DNA NA24385), one negative amplification control and one sample were analysed on each MiSeq FGx® standard flow cell. Libraries for all 12 DNA samples and controls were prepared simultaneously using the ForenSeq® Kintelligence Kit as described in the manufacturer's reference guide [32]. Each purified library was quantified using the Quantifluor ONE dsDNA (Promega) and normalized by diluting to 0.75 ng/μL with RSB. Five μL of each one of three normalized libraries were pooled to a total of five pools. Five sequencing runs were prepared and sequenced on the MiSeq FGx Sequencing System (QIAGEN, Hilden) according to manufacturer's MiSeq FGx Sequencing System reference guide [33]. Sequenced libraries were visualised and analysed using the ForenSeq Kintelligence Analysis Module in the ForenSeq Universal Analysis Software (UAS) (QIAGEN, Hilden). Kintelligence GEDmatch PRO reports were generated for all libraries directly from the UAS.

Library preparation, sequencing, and allele calling were performed as described in the ForenSeq® Kintelligence Kit Reference Guide and the Universal Analysis Software – Kintelligence Module Reference Guide [32,34]. All analysis within the UAS were performed with the default analytical and interpretation thresholds at 3 % in the ForenSeq® Kintelligence analysis method. This represents a minimum read count of 20 reads [18].

2.4.2. Global Screening Array-24 version 3.0 BeadChip

Aliquots of the 12 samples (DNA extracts) were provided to the Australian Genome Research Facility (AGRF) (Melbourne, Australia) and stored at −20°C until analysis. For each sample, 4 μL (range of 45–536 ng) were processed using the Illumina Infinium HTS Assay [35] for analysis with the Infinium Global Screening Array-24 BeadChip version 3.0 [36] and the iScan system array scanner (Illumina) [37] following the conditions specified by the manufacturer. Genotype normalisation, genotype calling and data analysis were conducted using GenomeStudio 2.0.4 with Genotyping module 2.0.4 (Illumina, San Diego, CA, USA) [38], using the default Illumina settings and Illumina GSA-24v3–0_A1 manifest and GSA-24v3–0_A1_ClusterFIle cluster files. GSA SNP data was prepared for upload to GEDmatch using Illumina's GenomeStudio 2.0.

For ease of discussion, the ForenSeq® Kintelligence Kit and the Global Screening Array (GSA)-24 version 3.0 BeadChip will herein be referred to as Kintelligence and GSA.

2.5. GEDmatch and GEDmatch PRO Uploads

An account was created under GEDmatch and GEDmatch PRO, respectively, with GSA data uploaded to GEDmatch, and Kintelligence data uploaded to GEDmatch PRO (by Verogen). Uploads and comparisons within GEDmatch were conducted with the consent of the donors, and prior to changes to GEDmatch and/or GEDmatch PRO terms and conditions (updated in April 2022) preventing LE agencies from uploading reference samples to GEDmatch and limiting the search of reference samples to the One-to-One tools. Each data file uploaded to GEDmatch was assigned a unique kit identifier for subsequent comparisons, with all kits having a status of 'research' to ensure that the kit's DNA data would not be included in relationship results of other users. Uploads to GEDmatch PRO were similarly assigned a unique kit identifier for subsequent comparisons.

2.6. GEDmatch and GEDmatch PRO comparisons

2.6.1. Kintelligence data

Kintelligence kits for each of the 12 samples were queried against all kits that had opted-in for law-enforcement searches in GEDmatch PRO. The One-to-Many Kinship tool was applied to all samples and the 'High Confidence Matches' and an 'Expanded Match List' were generated and downloaded. For all matches observed on GEDmatch PRO, a shared cMs value is provided (which is comparable but not identical to the shared cMs values provided by GEDmatch).

The Generation Visual Chart (based on the shared cM values), available within GEDmatch PRO, was used to assess possible relationship degrees of each individual such as 1st, 2nd, 3rd etc to the matches. The Generation Visual Chart is a schematic adapted from the Shared cM Project tool [26].

2.6.2. GSA data

GSA data for each sample was uploaded to GEDmatch and queried (in research mode) against all GEDmatch profiles using the One-to-Many comparison tool. The Tier 1 subscription was obtained to enable access to the advanced analysis tools. Population admixture proportions were estimated using the Eurogenes K13 model [39] in GEDmatch.

The GEDmatch shared cMs values of each individual related to each of the 12 donor samples were used to assess likely relationship using the using the Shared cM Project 4.0 tool v4 in DNA Painter [26].

3. Results

3.1. Samples

DNA (from combined extractions) for all 12 samples were quantified, with quantification and yield ranges of 11.17–1374.40 ng/μL and

748.57–8793.87 ng observed respectively, furthermore, the Degradation Index ranged from 0.618 to 1.334 (data not shown).

Full STR profiles obtained for all 12 samples indicated the contributor status of each sample was single source (data not shown). SNP data was generated for each sample using Kintelligence and GSA by external service providers Verogen (now QIAGEN) and AGRF respectively. SNP call rates, genotypes and assay/marker performance specific to each method were detailed in a report from each provider.

3.1.1. Kintelligence

The ForenSeq Universal Analysis Software (UAS) was used to analyse the quality of data from each family study sample sequenced across five sequencing runs. All analysis within the UAS were performed with default thresholds resulting in the average number of reads of 15,236,917 (range: 12,708409 – 17,963,427) and the average number of SNPs (above the analytical threshold) of 10,185 (range: 10,130–10,229) per sample.

The locus call rate indicated the number of SNPs typed of the total possible number present. The total number of expected loci depends on the biological sex of the sample donor where 10,230 SNPs are expected for males and 10,145 SNPs for females. Call rates were determined by the number of SNPs typed vs the number of SNPs expected with call rates for all samples >99 % (av. 99.90 %) (Table S1). A consistent loss of typed kinship SNPs (kiSNPs) compared to other SNP types was observed in all samples. The lowest observed call rate was for sample S5 (99.75 %). Of the 26 SNPs that were not typed, 24 were kiSNPs (9843/9867) with 23 of these kiSNPs having amplicons less than 150 bps in length. However, the number of kiSNPs in Kintelligence is disproportionately high compared to other SNPs in the multiplex. The loss of kiSNPs in this and other samples with similar call rates did not impact on the kinship predictions in this study.

A total of four SNPs in the positive control were flagged by the UAS as discordant with the known genotypes for NA24385. Further analysis of the discordant genotypes indicated that this observation was due to allelic dropout (either both or one allele) rather than incorrect/genotype error. Ten SNPs were detected in the negative control with an average of 36x coverage (range: 24–62) and a total read count of 902. In comparison to the positive control DNA (18,914,177 reads) and the study sample reads (ranging from 12,708,409 (S5) to 17,963,427 (S3)), the average coverage in the negative control is significantly lower. Using the same analytical threshold (3 %), on average, 13 SNPs (+/- 10) were typed in 21 negative controls in a recent developmental validation study with the manufacturer which stated that negative amplification controls are rarely free of any SNP detection [40].

Biological sex was correctly determined by the UAS for all samples when compared to self-declared information.

3.1.2. GSA

The performance of the samples using the GSA was reported by the service provider, with all the required controls including the non-specific binding controls that target bacterial sequences passing thresholds established by the service provider (data not shown). DNA input varied between 45 and 536 ng depending on the concentration of the sample (DNA extract). The estimated gender was as expected based on self-declared donor information and the call rates were > 0.99 for all samples (Table S2).

3.2. Kintelligence – assessment of kinship prediction: within family study samples

An assessment of Kintelligence performance in inferring kinship was performed using S1 or S2 as the 'home' person with the corresponding Kintelligence kit searched in GEDmatch PRO using the One-to-Many Kinship tool. The Generation Visual Chart within GEDmatch PRO was used to predict the relationships. Although additional matches (other than to the family study samples) were noted, only the matching GSA family study samples were included in this analysis. The kinship assessment using S1 as the home person resulted in the expected relationships to the other samples in the family study are shown in Fig. 1A. Both match lists (High Confidence Matches and Expanded Match List) returned the same relationship outcomes (Table 1). All, but one, of the relationships were predicted as expected, with S12 (a 2C1R – 6th degree relationship to S1) not identified as a match. The kinship assessment using S2 as the home person and the expected relationships to the remaining samples in the family study are shown in Fig. 1B. Both match lists (High Confidence Matches and Expanded Match List) returned the same relationship outcomes (Table 1). All of the relationships were predicted as expected.

3.3. Kintelligence – assessment of kinship prediction: external to family study samples

Exclusion of the family samples from the One-to-Many Kinship tool match list generated above where S1 is the home person resulted in one match (3rd degree) obtained in the 'High Confidence Matches' list; with an additional match (5th degree) noted when the 'Expanded Match List' was included. No additional matches were obtained in the 'High Confidence Matches' list when S2 is the home person, however, two matches (both 5th degree) were noted when the 'expanded Match List' was included (Table 2).

3.4. GSA – assessment of kinship prediction: within family study samples

To evaluate how GSA derived kits performed in inferring kinship, a comparison was performed using either S1 or S2 as the home person, with the corresponding GSA kit searched in GEDmatch using the One-to-Many segment-based search tool. Although additional matches were noted, only the family study GSA samples were analysed. The expected relationships to the other samples in the family study when S1 was the home person are shown in Fig. 1A, with DNA Painter Shared cM tool used to predict the relationships. For S12 (a 2C1R – 6th degree relationship) a match was made with a lower shared cMs value than expected (38.32 cMs) predicted with a 3 % probability (Table 3). The expected relationships to the other samples in the family study with S2 as the home person are shown in Fig. 1B. However, for S12 (a 1C2R – 4th degree relationship) a match was observed with a lower shared cMs value than expected (156.649 cMs) predicted with a 27 % probability (Table 3).

3.5. GSA – assessment of kinship prediction - external to family study samples

Exclusion of the family samples from the One-to-Many match list generated above with S1 as the home person, 46 additional matches were observed with a shared cMs value ≥ 30 (Table S3). These included the matches observed when using the S1-Kintelligence kit on GEDmatch PRO. For S2 as the home person, there were 40 additional matches observed with a shared cMs value ≥ 30 (Table S3). These included the matches observed when using the S2 Kintelligence kit on GEDmatch PRO.

3.6. Comparison of All relationships within the family

Based on the family tree, all the predicted relationships (1st, 2nd, 3rd, 4th, 5th and 6th degree) within the family study were assessed. Using the One-to-Many comparison for each sample (S1-S12) for the Kintelligence and GSA kits, the predicted relationships using the Generation Visual Chart or Shared cM Project tool [26] respectively were noted in comparison to the expected relationships (Table 4). This included eight 1st degree, nine 2nd degree, and seven 3rd degree relationships, respectively, and all were predicted as expected by both Kintelligence and GSA. Of the eight 4th degree relationships, all 8 were

Table 1
Kintelligence – Assessment of kinship prediction: within family study samples. The corresponding match list for high and expanded searches when either S1 or S2 is the home person; the relationships predicted corresponds to the known relationship (green); or the known relationships is not predicted (orange). Excluding all ½ relationships, the relationships refer to *1st degree*: parent, child or sibling; *2nd degree*: grandparent/child, aunt/uncle, or niece/nephew; *3rd degree*: Great-grandparent/child, great-aunt/uncle/niece/nephew or 1 C; *4th degree*: GG-grandparent/child, GG-aunt/uncle/niece/nephew, 1C1R or 1C1R; *5th degree*: GGG-grandparent/child, GGG-aunt/uncle/niece/nephew or 2 C; and *6th degree*: 1C3R or 2C1R [11].

			GSA Samples											
			S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
Kintelligence	S1	High	Self	1st	1st	2nd	2nd	3rd	4th	4th	3rd	5th		ND
		Expanded	Self	1st	1st	2nd	2nd	3rd	4th	4th	3rd	5th		ND
	S2	High	1st	Self		1st	1st	2nd	3rd	3rd	2nd	4th		4th
		Expanded	1st	Self		1st	1st	2nd	3rd	3rd	2nd	4th		4th

ND: not detected.

Table 2
Kintelligence - umber of matches. Observed number of matches for each Relationship Degree when S1 or S2 is the home person, for the high and expanded searches, when searched against family members only (int.) or external to family members (ext.).

			Observed # Matches for each Relationship Degree					
			1st	2nd	3rd	4th	5th	6th
S1	High	Int.	3	2	2	2	1	-
		Ext.	-	-	1	-	-	-
	Expanded	Int.	3	2	2	2	1	-
		Ext.	-	-	1	-	1	-
S2	High	Int.	4	2	2	1	-	-
		Ext.	-	-	-	-	-	-
	Expanded	Int.	4	2	2	1	-	-
		Ext.	-	-	-	-	2	-

predicted as potential 4th degree relationships using Kintelligence and the Generation Visual Chart. Of the GSA kits, all 8 were predicted as 4th degree relationships, (where one was matched but the shared cMs value was less than expected for a 4th degree relationship). One 5th degree relationship was predicted by both Kintelligence (only in the expanded match list) and GSA, and one 6th degree relationship was only detected by GSA (although it was predicted to be more distantly related then a 6th degree relationship).

4. Discussion

The use of methods to generate whole genome SNP data required for upload to genealogy databases, has greatly improved the utility of FIGG for forensic (poor quality and quantity) samples [41]. The use of targeted amplicon sequencing or WGS, however, presents its own set of challenges. In particular, the creation of large DNA data sets (e.g., >10 million) compared with WGA (e.g., < 1 million) for each sample analysed. WGS generates a large amount of surplus data when used for FIGG, as most of the profiles in the major genealogy databases were genotyped using microarrays. Therefore, only a proportion of the information generated using WGS is required for upload and subsequent analysis. Furthermore, similar to WGA, the data includes medically relevant SNPs which require appropriate management of the data/-information generated.

Kintelligence was designed specifically for FIGG, as a targeted amplicon sequencing based method that generates data from kinship SNPs (9867 of the total 10,230) enabling the detection of 1st, 2nd, 3rd and 4th degree relationships with high sensitivity, and to a lesser extent 5th degree relationships with 6th degree relationships not detectable in this study [9]. Importantly, the Kintelligence SNPs are spaced across the genome to minimise linkage effects and (to date) have no known medical relevance [9]. The automatable lab workflow and software analysis within the UAS and the companion windowed kinship algorithm (within GEDmatch PRO) supports the utilisation of Kintelligence to address issues currently experienced when generating SNP data from forensic samples for FIGG using WGA/WGS. Kintelligence can be applied to improve recovery from compromised DNA samples (poor quality and quantity) while minimising access to potential personal health

Table 4
Family members relationship predictions using Kintelligence and GSA. The predicted and expected number of 1st, 2nd, 3rd 4th, 5th and 6th degree family members within the family study using either Kintelligence or GSA comparisons. The number of relationships that were not correctly called although they were matched are shown in brackets. The DNA Painter Shared cM tool was used to predict the relationships of the GSA matches based on the total shared cM values.

	Predicted # / Expected #	
Relationship Degree	Kintelligence	GSA
1st	8/8	8/8
2nd	9/9	9/9
3rd	7/7	7/7
4th	4/8 (2/8)	7/8 (1/8)
5th	1/1	1/1
6th	0/1	0/1 (1/1)

Table 3
GSA – Assessment of kinship prediction: within family study samples. Shared cMs values from the One-to-Many comparison when either S1 or S2 is the home person. The % probability values (shown in brackets) obtained using the Shared cM Project tool (in DNA Painter) for the known relationship; when the top relationships predicted by the Shared cM Project tool includes the known relationship (green), or the top relationships predicted does not include the known relationship (orange) – in these instances the known relationship is observed as a lower probability (shown in brackets).

		GSA Samples											
		Shared cMs Match Values											
		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
GSA	S1	Self	3570.433 (100%)	3569.778 (100%)	1569.671 (100%)	2007.682 (98%)	1007.728 (100%)	548.16 (87%)	506.074 (90%)	1033.025 (100%)	350.93 (43%)		38.321 (3%)
	S2	3569.778 (100%)	Self		3569.027 (100%)	3569.827 (100%)	1874.09 (100%)	952.883 (100%)	863.097 (97%)	1858.723 (100%)	592.544 (77%)		156.649 (27%)

information and reducing costs and time associated WGA and WGS methods [9,17].

To evaluate the kinship prediction accuracy of Kintelligence, a family study was undertaken consisting of twelve individuals with known familial relationship spanning 1st to 6th degree relationships. Given the well-established use of WGAs such as the GSA by DTCs for FIGG [42], the twelve samples were also analysed using the GSA for comparison purposes. The SNP data generated by Kintelligence and GSA were uploaded into GEDmatch PRO and GEDmatch respectively, as per consent policies for each database effective during the study. The uploaded data was removed upon the completion of the study. Due to the limited number of samples in this assessment, this study focused on kinship prediction accuracy using both above-mentioned methods and not assay/method performance of Kintelligence which is better informed by current published studies [28,29,40,43,44] or expanded future studies.

Kintelligence and GSA generated 99.75 % and 99.16 % or higher SNP call rates respectively as well as correct predictions of biological sex. Selection of the family study donors included a requirement for European BGA to increase the likelihood of database matches outside of relatives in the study given the representation of individuals with European BGA in LE permitted databases such as GEDmatch PRO and FamilyTreeDNA. While BGA predictions using GSA and Kintelligence data were obtained, they were not presented as part of this study as the focus was on kinship prediction. However, it was noted that BGA prediction using Kintelligence was limited by the 2D visualisation of PCA resulting in overlapping population clusters which prevented reliable interpretation of the results. Further investigation of BGA prediction using Kintelligence would involve analysis of the aSNP genotypes in 3rd party tools/software such as FROG-KB, Snipper and Structure [45–48] to provide additional resolution.

Kintelligence generated sufficient data for kinship, biological sex, BGA and EVC predictions using only 1 ng of DNA input [17]. Conversely, an input of 100 ng, or higher, of good quality DNA is required for the WGA assays such as GSA [31]. DNA inputs ranging from ~ 45–536 ng were used for the GSA assay in this study with no obvious impact on the call rates or prediction accuracy. The authors of this study have previously demonstrated that as little as 0.1 ng of a good quality sample (with a low degradation index) is sufficient to generate data on a WGA that yields the expected match results when uploaded to a genealogy database [14], it is accepted that greater amounts are required to achieve an uploadable SNP profile for forensic samples using WGA [14,15].

The first step in evaluating the ability to call the expected relationship, searches on the genealogy databases were anchored using two of the 12 participants, S1 or S2 respectively. For Kintelligence, the expected matches for S1 or S2 to the remaining family members were observed for all but one family member, S12, considered to have the most distant relationship to both S1 and S2. This family member was either a 1 C2R (4th degree) or 2 C1R (6th degree) depending on which kit anchored the search and was matched to the 4th degree relationship, but not the 6th degree relationship in either the high confidence or expanded match lists generated using Kintelligence.

Similar to the Shared cM Tool, degrees of relationships may have overlapping cM ranges and relationships between two kits may be captured across multiple degrees. For example, a true 5th degree relationship would lie within a cM range of 0–319, while a true 4th degree relationship would lie within a cM range of 78–655. Given the overlap between the ranges, a potential genetic relation with shared cM between 78 and 319 can be either 5th or 4th degree relationships. For purposes of this study, if the shared cM was within the range of two degrees of relatedness, and one of them was the true degree of relatedness, it was considered an accurate prediction given the equivalent study using The Shared cM Project tool with a segment-based approach would similarly include probabilities associated with different degrees of relatedness.

In comparison, WGA was able to detect and match all family members including the 6th degree relationship between S1 and S12. However, the relationship was not predicted within the highest % probability value, suggesting a more distant relationship than expected or a limitation of the method to detect distant relationships of this degree. Furthermore, the searches conducted with Kintelligence only provided few additional matches outside the family members with either the high confidence or expanded match lists. GSA searches resulted in more than 40 additional matches of $\geq$ 30 total cMs shared (including those observed with Kintelligence) for either S1 or S2. The value of having the additional match information, from a genealogy perspective for the development of investigative leads, and the time associated with investigative work may be debatable and context dependent.

The genealogy component of FIGG is resource intensive depending on the degree of relatedness and quality of information available in public records to develop an investigative lead. Given the resource intensive nature of the genealogy search, these tend to focus on 3rd degree relatives or closer to reduce the effort associated with building trees and finding associations with more distant relatives [1]. Kintelligence has been reported to have a high degree of sensitivity for 1st, 2nd, 3rd and 4th degree relationships, and to a lesser extent beyond the 5th degree [9].This is similar to Snedecor *et al.* [9], who reported the (robust) detection of 3rd degree relationships in a simulated and real data set. In comparison, the GSA approach was able to detect all 1st to 6th degree relationships in the family study, however, the 6th degree relationship was predicted as being more distant.

Kintelligence is a suitably reliable tool for use in a FIGG casework approach focusing on 4th degree relatives or closer. However, for those cases where only 5th-6th degree relationships or higher are available for the genealogy component, operational considerations to apply Kintelligence would need to include an understanding of the resource intensive approach that would be required to ensure reliability of results. Based on our current experience, the absence of 1st to 3rd degree matches is a common occurrence in cases outside of the US, particularly in Australia. A clear advantage of Kintelligence compared to GSA and other WGAs is the low DNA input requirement amenable to the analysis of degraded and low DNA quantity samples. In addition, the compatibility of Kintelligence with The ForenSeq DNA Signature Prep kit using the MiSeq FGx MPS platform (both forensically validated for casework [19–22]) and the UAS analysis modules, expands the utility of Kintelligence to generate reliable data from identity, BGA and EVC SNPs within operational forensic laboratories.

Declaration of Competing Interest

All authors declare that they have no conflicts of interest.

Acknowledgements

The authors greatly acknowledge the support of the Victorian Institute of Forensic Medicine, New South Wales Police Force, Australian Federal Police and Qiagen, and the collaborative nature in which agencies work together to advance the evaluation of new capabilities for the betterment of forensics. The authors also thank Victoria Police Forensic Services Department, for their early contributions to project planning. Furthermore, the authors acknowledge the support and expertise provided by Verogen for the analysis using Kintelligence of the family study samples. The technical expertise of staff at Forensic Biology, Forensic & Analytical Science Service, and the Molecular Biology Laboratory at the Victorian Institute of Forensic Medicine was instrumental in the preparation of samples 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.forsciint.2024.112208.

References

[1] R. Daniel, S. Harbison, Sequencing Technology in Forensic Science: Next-Generation Sequencing, in: A.B. Elena Pilli (Ed.), Forensic DNA Analysis, Apple Academic Press, New York, 2021.

[2] J. Ge, B. Budowle, Forensic investigation approaches of searching relatives in DNA databases, J. Forensic Sci. 66 (2) (2021) 430–443.

[3] E.M. Greytak, C. Moore, S.L. Armentrout, Genetic genealogy for cold case and active investigations, Forensic Sci. Int 299 (2019) 103–113.

[4] United States Department of Justice, Interim Policy Forensic Genetic Genealogical DNA analysis and Searching. 2019 [cited 2023 08 February]; Available from: ⟨https://www.justice.gov/olp/page/file/1204386/dl⟩.

[5] A.Regalado, More than 26 million people have taken an at-home ancestry test. 2019 [cited 2023 7 March]; Available from: ⟨https://www.technologyreview.com/2019/02/11/103446/more-than-26-million-people-have-taken-an-at-home-ancestry-test/⟩.

[6] N. Scudder, et al., Operationalising forensic genetic genealogy in an Australian context, Forensic Sci. Int 316 (2020) 110543.

[7] C.L. Glynn, Bridging disciplines to form a new one: the emergence of forensic genetic genealogy, Genes (Basel) 13 (8) (2022).

[8] D. Kling, et al., Investigative genetic genealogy: Current methods, knowledge and practice, Forensic Sci. Int. Genet. 52 (2021) 102474.

[9] J. Snedecor, et al., Fast and accurate kinship estimation using sparse SNPs in relatively large database searches, Forensic Sci. Int. Genet. 61 (2022) 102769.

[10] M. Vogen, Victorian Institute of Forensic Medicine, Australian Federal Police & NSW Police Team with Othram to Identify Sandy Point John Doe. 2023 [cited 2024 11 June]; Available from: ⟨https://dnasolves.com/articles/vifm-afp-nsw-christopher-luke-moore/⟩.

[11] A. Tillmar, et al., Getting the conclusive lead with investigative genetic genealogy - A successful case study of a 16 year old double murder in Sweden, Forensic Sci. Int. Genet. 53 (2021) 102525.

[12] R.Armitage, The Golden State Killer: How it took four decades to catch the serial predator who terrorised California. 2020 [cited 2024 11 June]; Available from: ⟨https://www.abc.net.au/news/2020-08-22/the-forty-year-hunt-for-the-golden-state-killer/12579638⟩.

[13] T.L. Dowdeswell, Forensic genetic genealogy: a profile of cases solved, Forensic Sci. Int. Genet. 58 (2022) 102679.

[14] A. Davawala, et al., Forensic genetic genealogy using microarrays for the identification of human remains: the need for good quality samples - a pilot study, Forensic Sci. Int 334 (2022) 111242.

[15] J.H. de Vries, et al., Impact of SNP microarray analysis of compromised DNA on kinship classification success in the context of investigative genetic genealogy, Forensic Sci. Int. Genet. 56 (2022) 102625.

[16] A. Tillmar, et al., The FORCE Panel: An All-in-One SNP marker set for confirming investigative genetic genealogy leads and for general forensic applications, Genes (Basel) 12 (12) (2021).

[17] Verogen. ForenSeq Kintelligence Kit. [cited 2022 09 December]; Available from: ⟨https://verogen.com/products/forenseq-kintelligence-kit/⟩.

[18] Verogen. SNP Typing in Universal Analysis Software and Kinship Estimation with GEDmatch PRO. 2021 [cited 2023 28 September]; Available from: ⟨https://verogen.com/wp-content/uploads/2021/03/snp-typing-uas-kinship-estimation-gedmatch-pro-tech-note-vd2020058-a.pdf⟩.

[19] J.D. Churchill, et al., Population and performance analyses of four major populations with Illumina's FGx Forensic Genomics System, Forensic Sci. Int. Genet. 30 (2017) 81–92.

[20] S. Kocher, et al., Inter-laboratory validation study of the ForenSeq DNA Signature Prep Kit, Forensic Sci. Int. Genet. 36 (2018) 77–85.

[21] A. Senst, et al., Validation and beyond: Next generation sequencing of forensic casework samples including challenging tissue samples from altered human corpses using the MiSeq FGx system, J. Forensic Sci. 67 (4) (2022) 1382–1398.

[22] C. Xavier, W. Parson, Evaluation of the Illumina ForenSeq DNA signature prep Kit - MPS forensic application for the MiSeq FGx benchtop sequencer, Forensic Sci. Int. Genet. 28 (2017) 188–194.

[23] A. Manichaikul, et al., Robust relationship inference in genome-wide association studies, Bioinformatics 26 (22) (2010) 2867–2873.

[24] M.P. Conomos, et al., Model-free estimation of recent genetic relatedness, Am. J. Hum. Genet 98 (1) (2016) 127–148.

[25] M.P. Conomos, M.B. Miller, T.A. Thornton, Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness, Genet Epidemiol. 39 (4) (2015) 276–293.

[26] B.T.Bettinger, The Shared cM Project 4.0 tool v4. 2020 [cited 2024 11 June]; Available from: ⟨https://dnapainter.com/tools/sharedcmv4⟩.

[27] Methods, S.W.Go.D.A. Validation Guidelines for DNA Analysis Methods. 2016 [cited 2024 11 June]; Available from: ⟨https://www.swgdam.org/_files/ugd/4344b0_813b241e8944497e99b9c45b163b76bd.pdf⟩.

[28] M.A. Peck, et al., Internal validation of the forenseq kintelligence kit for application to forensic genetic genealogy, Forensic Genom. (December) (2022) 103–114.

[29] J. Watson, et al., Operationalisation of the ForenSeq(R) Kintelligence Kit for Australian unidentified and missing persons casework, Forensic Sci. Int. Genet. 68 (2024) 102972.

[30] J. Ward, et al., Operationalization of the national DNA program for unidentified and missing persons' forensic investigative genetic genealogy capability for human remains identification in Australia, Forensic Genom. 4 (1) (2024).

[31] Illumina. Infinium Global Screening Array-24 Kit. [cited 2022 09 December]; Available from: ⟨https://sapac.illumina.com/products/by-type/microarray-kits/infinium-global-screening.html⟩.

[32] Verogen. ForenSeq Kintelligence Kit: Reference Guide. 2021 [cited 2023 3 October]; Available from: ⟨https://verogen.com/wp-content/uploads/2021/03/forenseq-kintelligence-reference-guide-vd2020053-b.pdf⟩.

[33] Verogen. MiSeq FGx Sequencing System: Reference Guide. 2021 [cited 2023 3 October]; Available from: ⟨https://verogen.com/wp-content/uploads/2021/02/miseq-fgx-system-reference-guide-vd2018006-f.pdf⟩.

[34] Verogen. Universal Analysis Software Kintelligence Module - Version 2. 2022 [cited 2023 3 October]; Available from: ⟨https://verogen.com/wp-content/uploads/2022/08/universal-analysis-software-v2-reference-guide-Kintelligence-VD2022002-RevA.pdf⟩.

[35] Illumina. Infinium HTS Assay Reference Guide. 2019 [cited 2024 11 June]; Available from: ⟨https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/infinium_assays/infinium-hts/infinium-hts-assay-reference-guide-15045738-04.pdf⟩.

[36] Illumina. Infinium™ Global Screening Array-24 v3.0 BeadChip. [cited 2024 11 June]; Available from: ⟨https://www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/infinium-global-screening-array-data-sheet-370-2016-016.pdf⟩.

[37] Illumina. iScan System Support Resources. [cited 2024 11 June]; Available from: ⟨https://sapac.support.illumina.com/array/array_instruments/iscan.html⟩.

[38] Illumina. GenomeStudio® Genotyping Module v2.0 Software Guide. 2016 [cited 2024 11 June]; Available from: ⟨https://support.illumina.com/content/dam/illumina-support/documents/documentation/software_documentation/genomestudio/genomestudio-2-0/genomestudio-genotyping-module-v2-user-guide-11319113-01.pdf⟩.

[39] Eurogenes Genetic Ancestry Project, Updated Eurogenes K13 now at GEDmatch. 2021.

[40] J. Antunes, et al., Developmental validation of the ForenSeq(R) Kintelligence kit, MiSeq FGx(R) sequencing system and ForenSeq Universal Analysis Software, Forensic Sci. Int. Genet. 71 (2024) 103055.

[41] A. Tillmar, et al., Whole-genome sequencing of human remains to enable genealogy DNA database searches - a case report, Forensic Sci. Int. Genet. 46 (2020) 102233.

[42] C. Lu, B. Greshake Tzovaras, J. Gough, A survey of direct-to-consumer genotype data, and quality control tool (GenomePrep) for research, Comput. Struct. Biotechnol. J. 19 (2021) 3747–3754.

[43] R. Mitchell, E.S, K. Eskey, T. Ferguson, R. Just, A method to enable forensic genetic genealogy investigations from DNA mixtures, Forensic Sci. Int.: Genet. Suppl. Ser. 8 (2022) 159–161.

[44] R.H. Oefelein, Seven years of SNPs: an assessment of methods utilized for generating profiles for forensic genetic genealogy, Forensic Sci. Int.: Genet. Suppl. Ser. 8 (2022) 347–348.

[45] H. Rajeevan, et al., Introducing the Forensic Research/Reference on Genetics knowledge base, FROG-kb, Invest. Genet. 3 (1) (2012) 18.

[46] K.K. Kidd, et al., The redesigned forensic research/reference on genetics-knowledge base, FROG-kb, Forensic Sci. Int. Genet. 33 (2018) 33–37.

[47] C. Santos, et al., Inference of ancestry in forensic analysis II: analysis of genetic data, Methods Mol. Biol. 1420 (2016) 255–285.

[48] J.K. Pritchard, M. Stephens, P. Donnelly, Inference of population structure using multilocus genotype data, Genetics 155 (2) (2000) 945–959.