

OPEN

# An evaluation of the SureID 23comp Human Identification Kit for kinship testing

Hussain M. Alsafiah<sup>1,2\*</sup>, Ali A. Aljanabi<sup>1</sup>, Sibte Hadi<sup>1</sup>, Saleh S. Alturayef<sup>2</sup> & William Goodwin<sup>1\*</sup>

Short tandem repeat (STR) profiling has been routinely used in kinship testing since the introduction of commercial kits in the mid-1990s. While 15 to 23 STR loci normally give definitive results in simple kinship testing, additional loci are sometimes required to resolve complex cases. The SureID 23comp Human Identification Kit, recently released by Health Gene Technologies (China), multiplexes amelogenin and 22 autosomal STRs, 17 of which are non-CODIS STRs. This enables the profiling of 38–40 loci when used in conjunction with widely used commercial kits. In this study, the kit was evaluated for kinship applications as a supplementary STR kit following the minimum criteria for validation recommended by the European Network of Forensic Science Institutes (ENFSI) and the Scientific Working Group on DNA Analysis Methods (SWGAM) using 500 samples. Performance was comparable with other commercial kits demonstrating: repeatability and reproducibility; precision (maximum s.d. 0.1048 nt); accuracy, all alleles were within  $\pm 0.41$  nt compared to the actual sizes; heterozygous peak balances at all loci  $> 68\%$ ; stutter ratios ranged from 3.8% to 16.15%; full profiles were generated with 125 pg DNA (95.12% of alleles at 62 pg); and we found 100% concordance over 5 common STRs with the GlobalFiler kit.

The extended number of STR markers required for the Combined DNA Index System (CODIS) and for the European Standard Set (ESS)<sup>1,2</sup>, has led to the development of GlobalFiler PCR Amplification Kit, VeriFiler Plus PCR Amplification Kit (Applied Biosystems (AB), USA), PowerPlex Fusion 6C System (Promega Corporation, USA) and Investigator 24plex (Qiagen, Germany). The information obtained from these kits will be sufficient in most kinship cases; however, it is still possible to have inconclusive results in complex cases<sup>3</sup>. Kinship testing can be further complicated when the level of consanguinity in the target population is relatively high<sup>4</sup>, or when the family pedigree is deficient<sup>5</sup>.

It has been demonstrated that additional STRs can increase the power of genetic testing in determining the true relationship between parent-child, siblings or half siblings<sup>6</sup>. For example, Carboni *et al.*<sup>7</sup> described four complex cases, including incest, which were inconclusive using 13–15 STRs, but that could be resolved using 39–41 STRs.

As most loci are shared between the commonly used kits, the maximum number of STRs that can be tested when combining any two is 24 STRs (e. g. VeriFiler Plus and PowerPlex Fusion 6C), which necessitates the use of a supplementary STR kit when more loci need to be tested. A set of 25 supplementary STRs was suggested by the National Institute of Standards and Technology (NIST) to increase the certainty in kinship testing<sup>8</sup>; however, no multiplex combing these STRs is commercially available. Supplementary Kits: Microreader 23sp ID<sup>9</sup> (Suzhou Microread Genetics, China), Goldeneye DNA ID 22NC<sup>10</sup> (Goldeneye Technology Ltd., China), AGCU 21 + 1<sup>11</sup> (AGCU ScienTech Incorporation, China) have been developed, but they are only commercially available in China<sup>12</sup>.

Massively parallel systems (MPS) allow simultaneous sequencing of multiple DNA makers. For example, Precision ID GlobalFiler NGS STR<sup>13</sup> (20 CODIS STRs and nine non-CODIS STRs)(AB), Promega PowerSeq Auto/Y system<sup>14</sup> (20 CODIS STRs, Penta D, Penta E, and 23 Y-STRs) (Promega Corporation), and ForenSeq DNA Signature Prep<sup>15</sup> (20 CODIS STRs, seven non-CODIS STRs, 24 Y-STRs, 7X-STRs and 94 identity informative

<sup>1</sup>School of Forensic and Applied Sciences, University of Central Lancashire, Preston, United Kingdom. <sup>2</sup>Forensic Genetics Laboratory, General Administration of Criminal Evidences, Public Security, Ministry of Interior, Riyadh, Saudi Arabia. \*email: [hmhalsafiah@uclan.ac.uk](mailto:hmhalsafiah@uclan.ac.uk); [whgoodwin@uclan.ac.uk](mailto:whgoodwin@uclan.ac.uk)

Chr.	SureID 23comp			Investigator HDplex STRs <sup>(a,b)</sup>	PowerPlex CS7 STRs
	STRs <sup>(a)</sup>	Location (GRCh38)	Repeat structure*		
1	D1S1656 <sup>a</sup>	230769616–230769683	CCTA [TCTA] <sub>n</sub>		F13B
2	D2S441 <sup>a</sup>	68011947–68011994	[TCTA] <sub>n</sub>	D2S1360	
3	D3S1744	147374752–147374828	[ATAG] <sub>n</sub> <i>atg</i> [ATAG] <sub>n</sub> <i>at</i> [ATAG] <sub>n</sub>	D3S1744	
4	D4S2366	6483114–6483172	[GATA] <sub>n</sub> [GATT] <sub>n</sub> [GATA] <sub>n</sub> <i>gac</i> [GATA] <sub>n</sub>	D4S2366	
5	D5S2800	59403132–59403199	[GGTA] <sub>n</sub> [GACA] <sub>n</sub> [GATA] <sub>n</sub> [GATT] <sub>n</sub>	D5S2500	
6	D6S474	112557951–112558018	[AGAT] <sub>n</sub> [GATA] <sub>n</sub>	D6S474, SE33 <sup>b</sup>	F13A01
7	D7S3048	21227099–21227174	[TATC] <sub>n</sub> [TACC] <sub>n</sub> [CACCC] <sub>n</sub>	D7S1517	
8	D8S1132	106316692–106316774	[TCTA] <sub>n</sub> <i>tca</i> [TCTA] <sub>n</sub>	D8S1132	LPL
9	D9S1122	77073826–77073873	[TAGA] <sub>n</sub>		Penta C
10	D10S1248 <sup>a</sup>	129294244–129294295	[GGAA] <sub>n</sub>	D10S2325	
11	D11S2368	19259601–19259684	[TATC] <sub>n</sub> [TGTC] <sub>n</sub> [TATC] <sub>n</sub>		
12	D12S391 <sup>a</sup>	12297020–12297095	[AGAT] <sub>n</sub> [AGAC] <sub>n</sub> AGAT	D12S391 <sup>a</sup>	
13	D13S325	42599304–42599382	[TCTA] <sub>n</sub> <i>tca</i> [TCTA] <sub>n</sub>		
14	D14S1434	94842054–94842105	[CTGT] <sub>n</sub> [CTAT] <sub>n</sub>		
15	D15S659	46081911–46081966	[TATC] <sub>n</sub>		FESFPS, Penta E
16	D16S539 <sup>a</sup>	86352702–86352745	[GATA] <sub>n</sub>		
17	D17S1301	74684855–74684902	[AGAT] <sub>n</sub>		
18	D18S1364	65732998–65733056	[TAGA] <sub>n</sub> TACA [TAGA] <sub>n</sub>	D18S51 <sup>a</sup>	
19	D19S253	15617484–15617531	[ATCT] <sub>n</sub>		
20	D20S482	4525692–4525747	[AGAT] <sub>n</sub>		
21	D21S2055	39819508–39819649	[CTAT] <sub>n</sub> CTA [CTAT] <sub>n</sub> (30n) [TATC] <sub>n</sub>	D21S2055	Penta D
22	D22GATA198B05	17169811–17169882	CTCT [ATCT] <sub>n</sub> [ACCT] <sub>n</sub>		

**Table 1.** STR Markers included in the SureID 23comp Kit. This table shows the locations (GRCh38) and repeat structures of the 22 STRs included in the SureID 23comp Kit. Five loci are common with the CODIS and the ESS. Twelve loci are not included in other available supplementary kits (Investigator HDplex and PowerPlex CS7). All information was adopted from<sup>17,18,39</sup>. As proposed by Phillips<sup>19</sup>, the D5S2500 locus name is corrected to D5S2800 (Please see Note S1). <sup>a</sup>CODIS and ESS locus. <sup>b</sup>Germany core locus. \*Lowercase and italic nucleotides are not counted in allele nomenclature system<sup>39</sup>.

SNPs) (Verogen, USA). These can be utilised to increase the power of kinship testing<sup>15,16</sup>. However, the systems are expensive to establish and are not yet commonly used in many laboratories.

SureID 23 comp Human Identification Kit (Health Gene Technologies, China), combines amelogenin and 22 autosomal STRs: D1S1656, D2S441, D10S1248, D12S391, D16S539 and 17 non-CODIS STRs (D3S1744, D4S2366, D5S2800, D6S474, D7S3048, D8S1132, D9S1122, D11S2368, D13S325, D14S1434, D15S659, D17S1301, D18S1364, D19S253, D20S482, D21S2055, and D22GATA198B05). Twelve of these STRs are not included in other supplementary kits, such as Investigator HDplex Kit<sup>17</sup> (Qiagen) and PowerPlex CS7 System<sup>18</sup> (Promega Corporation) (Table 1). The kit is now available in the UK and Europe.

It is important to note that the D5S2800 locus was named as D5S2500 in the panels and in the supporting documents of the SureID 23comp Kit that may cause some confusion<sup>19</sup> (Note S1).

Testing additional STRs increases the number of loci situated on the same chromosome (syntenic loci), but raises concerns regarding their independence during meiosis. Syntenic loci are regarded as independent (unlinked) if they are 50 centimorgans (cM) or more apart (at which point the probability of recombination between them is 0.5)<sup>20</sup>. As recombination rates vary along chromosomes, using the physical distance (per bp) may underestimate or overestimate the genetic distance between syntenic loci<sup>21</sup>. Therefore, family studies have been undertaken to estimate the recombination fraction (RF) between syntenic loci<sup>21–24</sup>. However, family studies are expensive and, sometimes, may not be informative enough due to the need of a large number of generations (meiosis) and high percentage of heterozygote genotypes<sup>4,22</sup>. An alternative approach employed by Phillips *et al.*<sup>4</sup> used the high-density multi-point SNP data of HapMap to approximate the genetic distance between syntenic loci to estimate the RFs, which showed RF values similar to those generated using the family studies.

The SureID 23 comp was used to generate population genetic data for three main populations European, South Asian and African<sup>25</sup>, but it is believed that the kit has not been validated as no publications currently exist, either independently or from the manufacturer. Therefore, this study aimed to evaluate the performance of the SureID 23comp Kit for kinship applications as a supplementary STR kit. The minimum criteria for validation recommended by the ENFSI and by the SWGDAM were followed; mixture studies were not included as the kit is specifically designed to be used in complex kinship testing. In addition, the 17 non-CODIS loci were evaluated using a population (500 individuals) from Saudi Arabia. The data of the 17 non-CODIS loci were merged with the data of 21 STRs generated from the same samples<sup>26</sup>, and were used for testing potential linkage between

syntenic loci (linkage disequilibrium, LD). RF values of syntenic pairs were reviewed from<sup>4,12,21,22,24</sup>, which were estimated using family studies and using HapMap data. RFs derived from HapMap data for D2S1338-D2S441, TPOX-D2S1338, FGA-D4S2366 and D5S2800-CSF1PO pairs, were calculated based on cumulative genetic map distance provided by Phillips<sup>12</sup> and using the Excel tool developed by Phillips *et al.*<sup>4</sup>.

## Methods

The study was performed in accordance with the ethical guidelines of the University of Central Lancashire (UK) and was approved by the Security Forces Hospitals Programme (SFHP, Saudi Arabia), the Medical Legal Directorate, Ministry of Health (MLD, Iraq) and by Ethics Committee of the University of Central Lancashire.

**DNA samples.** Initial tests of the SureID 23comp Kit were carried out using the 2800 M control DNA (Promega Corporation). The control DNA was also used for sensitivity and stochastic tests by preparing five serial dilutions of (500, 250, 125, 62, and 31) pg. In addition, 0.5 ng of the control were amplified with the presence of different concentrations (50, 75, 100, 120 and 150) ng/μl of common PCR inhibitors humic and tannic acids (Sigma-Aldrich, USA). The 9947 A control DNA provided with the kit was used with every run as a positive control.

The performance of the kit was further assessed using nine bone samples that were collected from a mass grave in Iraq after obtaining a permission from the MLD, Ministry of Health (Iraq). The bone samples were previously extracted using PrepFiler BTA Forensic DNA Extraction Kit (AB) and were quantified using Quantifiler Trio DNA Quantification Kit (AB). The concentrations of the small fragments of the Quantifiler Trio ranged from 0.0173 ng/μl to 0.3271 ng/μl and the Degradation Indexes (DI) were from 1.6758 to 57.666 (Table S1). These samples were previously profiled using one or more of the commonly used STR kits (Table S1).

The study of precision, accuracy, peak balance, concordance and stutter peak ratios were carried out using 500 samples from unrelated individuals from the population of Saudi Arabia. The samples were collected after obtaining informed consents from all participants. The extraction and the quantification were described in a previous study that evaluated the 21 STR loci included in the GlobalFiler PCR Amplification Kit<sup>26</sup>.

**Sample amplification.** The Initial tests of the SureID 23comp used two reaction volumes that were optimised by the manufacturer. A 25 μl volume that contained 12.5 μl master mix, 6.25 μl primer mix and up to 6.25 μl of DNA template; and a 10 μl volume that contained 5 μl master mix, 2.5 μl primer mix and up to 2.5 μl of DNA template. The range of recommended template DNA is 0.5–4 ng. To validate both volumes, two operators carried out the initial tests independently with 0.5 ng of control DNA in five replicates (20 tests in total).

Three DNA concentrations (0.5, 0.35, and 0.25) ng were used for the first 90 samples from the Saudi population to evaluate the performance of the 10 μl volume. Then, the rest of samples (410 samples) were genotyped using the 10 μl volume with 0.5 ng DNA input per reaction.

Microamp Optical 96-Well Reaction Plates and Microamp Optical Adhesive Films (AB), were used for amplification; 2.5 μl of the DNA and DNase/RNase-free water were added to 7.5 μl of the SureID 23comp mix (5 μl master mix and 2.5 μl primer mix).

A Veriti thermal cycler (AB) was employed to carry out the amplification reactions as follows: [95 °C (5 min)]/[94 °C (10 s) 61 °C (60 s) 70 °C (30 s)] 28–30 cycles/[60 °C (15 min)]. The 28 cycle protocol was used for the initial tests, stability tests and for the 500 samples. For sensitivity and stochastic study, the serial dilution samples were amplified in five replicates using both reaction volumes, each volume was tested with 28 and 30 PCR cycles (100 tests in total). For the bone samples, the 25 μl volume and 30 PCR cycles were used.

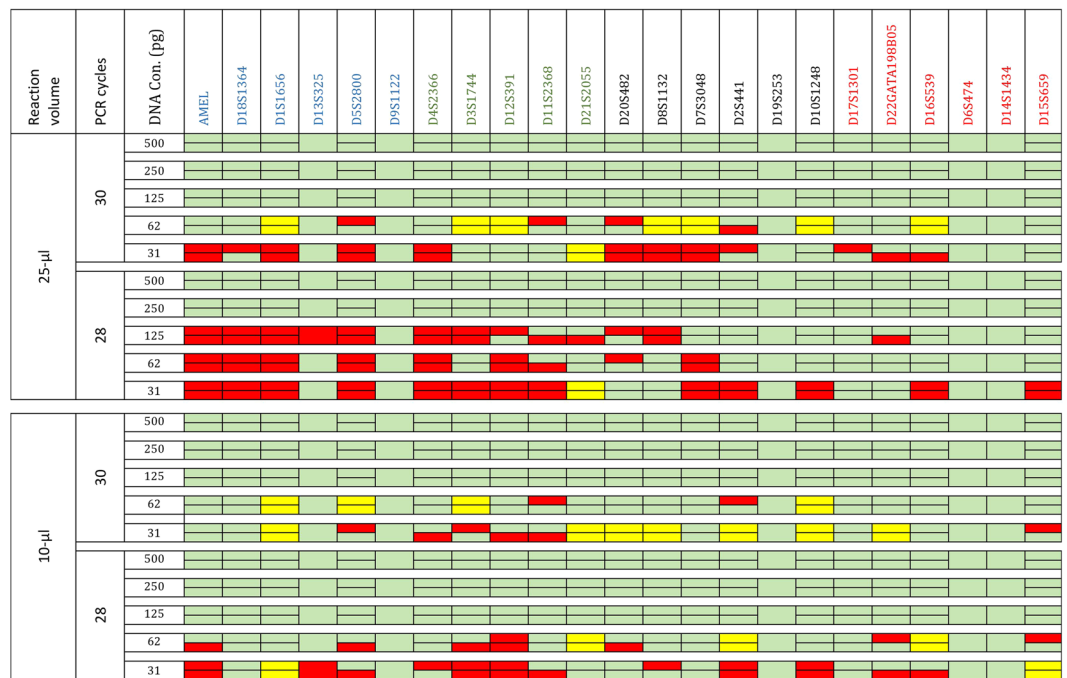
**DNA separation and analysis.** An ABI 3500 DNA Genetic Analyser with 50 cm capillaries and POP-6 polymer (AB) was used for the separation and detection. The spectral calibration mix was prepared by adding 8 μl HGT 5-Dye Matrix Standard (Health Gene Technologies) to 200 μl of Hi-Di Formamide (AB); 10 μl were dispensed to each well. In the data collection software (AB), the dye set of SureID 23comp was created based on the G5 template as recommended by manufacturer.

Samples were prepared for separation and detection by adding 1 μl of PCR products or of an allelic ladder (Health Gene Technologies) to 9 μl of Hi-Di Formamide/Size-500-Plus Mix. This mix was prepared by 9 μl of Hi-Di Formamide (AB) and 0.25 μl Size-500-Plus (Health Gene Technologies), for each sample. Based on the manufacturer's guidelines, the run time in the module HID36\_POP4 should be 1,210–1,500 s when using a 36 cm capillary. In this study, the run time was increased to 3900 s to accommodate the use of the 50 cm capillaries.

Alleles from the 23 markers were called using GeneMapper ID-X Software v1.2 (AB) with an allelic ladder mix supported by panels and bins provided by the manufacturer. For the sensitivity and stability tests, the minimum relative fluorescent units (RFU) was 50 RFU for heterozygous genotypes and was 150 RFU for homozygous genotypes.

**Population study.** PowerStat v 1.2 (Promega Corporation) was utilised to calculate allele frequencies, match probability, power of discrimination, power of exclusion, typical paternity index and polymorphic information contents. RStudio platform v1.2.1335<sup>27</sup> and DNA tools package v0.1-22<sup>28</sup>, were used to identify the maximum number of matched loci within the 500 samples. The Exact test for detecting deviation from the Hardy-Weinberg equilibrium (HWE) was carried out by Arlequin v3.5.2.1 Software<sup>29</sup>, using 1,000,000 steps for the Markov chain and 100,000 for the dememorization steps.

**LD and RF study.** The Arlequin v3.5.2.1 Software<sup>29</sup> was also used to test LD between potentially 18 syntenic loci resulting from combining 38 loci available in GlobalFiler (21 STRs) and SureID 23comp (17 non-CODIS STRs) kits. This was carried out by applying 1000 in the permutations and 2 in the Expectation-Maximisation (EM) algorithm. RF values of syntenic pairs were reviewed from<sup>4,12,21,22,24</sup>, which were estimated using



**Figure 1.** Sensitivity and stochastic tests for the SureID 23comp Kit. Serial dilutions (500, 250, 125, 62, and 31) pg were prepared from the 2800 M control DNA (Promega Corporation). Each test was done on five replicates and the sample with the highest number of detected alleles are shown. Each cell represents an allele and merged cells represent homozygote loci in the 2800 M. Green cells identify detected alleles with  $\geq 60\%$  peak balance ratios. Yellow cells identify detected alleles with  $< 60\%$  peak balance ratios. Red cells represent undetected alleles with threshold of 50 RFU/150 RFU for heterozygotes/homozygotes.

family studies and using HapMap data. RFs derived from HapMap data for D2S1338-D2S441, TPOX-D2S1338, FGA-D4S2366 and D5S2800-CSF1PO pairs, were calculated based on cumulative genetic map distance provided by Phillips<sup>[12]</sup> and using the Excel tool developed by Phillips *et al.*<sup>4</sup>.

## Results and Discussion

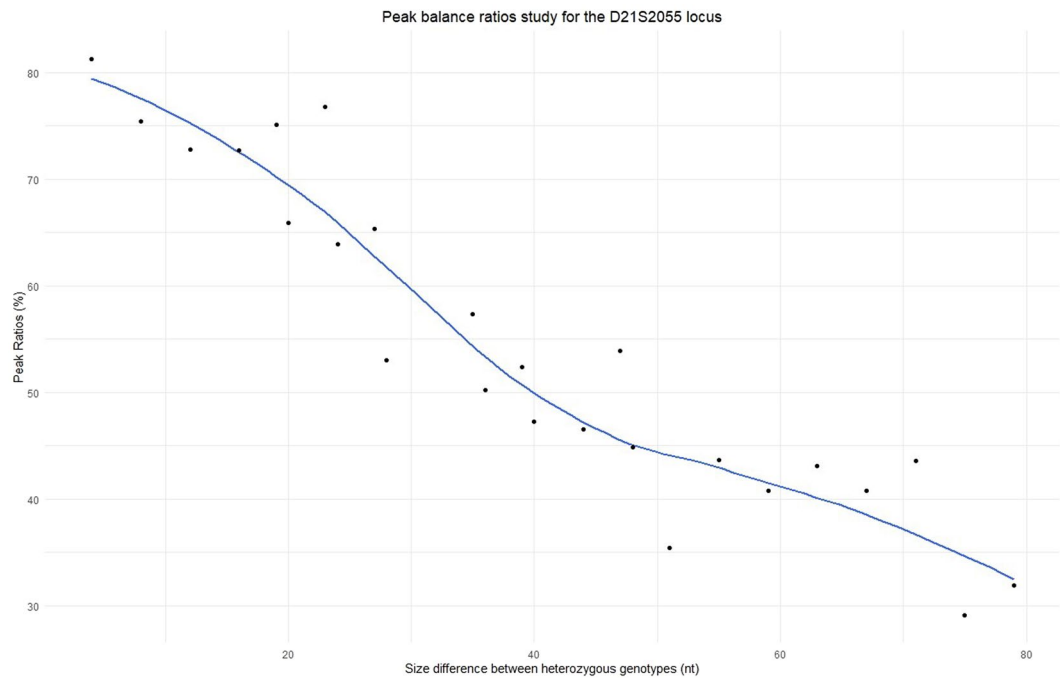
**Evaluation of the SureID 23comp Kit.** The first step in the validation was a confirmation of the identity of the D5 locus included in the multiplex by testing the 9947 A control DNA provided in the kit as a positive control (Note 1). The 9947 A control DNA showed alleles 14, 23 at the D5 locus confirming that the locus is D5S2800 and not D5S2500<sup>19</sup>.

In the initial tests of the SureID 23comp, the replicates of the 2800 M control DNA, were successfully profiled by two independent operators and 20 replicates were fully concordant confirming the repeatability and reproducibility.

Each of the dilution series samples was replicated five times using the 25  $\mu\text{l}$  and 10  $\mu\text{l}$  volumes with 28 and 30 cycles (100 reactions in total). Full profiles were generated from the 125 pg samples when using the 10  $\mu\text{l}$  volume (28 and 30 cycles), while 25  $\mu\text{l}$  volume was able to generate a full profile with 30 PCR cycles only. For the 62 pg samples, the 28 cycle protocol allowed detection of 60.9% (25  $\mu\text{l}$ ) to 80.48% (10  $\mu\text{l}$ ) and the 30 cycle protocol allowed detection of 90.2% (25  $\mu\text{l}$ ) to 95.12% (10  $\mu\text{l}$ ) where the rest of the alleles were visible and could be detected with a reduced RFU threshold of 30. With 31 pg, the profile percentage was from 53.6% (25  $\mu\text{l}$ ) to 85.3% (10  $\mu\text{l}$ ) using the 30 cycle protocol, while allele dropout was observed with 28 cycle protocol (Fig. 1). The sensitivity results are comparable to other commonly used kits, for example, Identifiler Kit<sup>30</sup>, Investigator HDplex Kit<sup>21</sup> and PowerPlex Fusion 6 C System<sup>31</sup> where the profile percentage ranged from 82% to 94% for the 62 pg and from 37% to 72% for the 31 pg.

The performance of the SureID 23comp Kit with different concentrations of two common PCR inhibitors was tested. Full profiles were generated in the presence of  $\leq 120$  ng/ $\mu\text{l}$  of tannic acid and of  $\leq 75$  ng/ $\mu\text{l}$  of humic acid (Figs S1 and S2). Although these levels are similar to those reported for the SureID PanGlobal Kit (Health Gene Technologies)<sup>32</sup>, other commonly used kits are more robust in the presence of higher concentrations of inhibitors<sup>33</sup> (Fig. S3).

The sensitivity and the stability of the SureID 23comp were further evaluated using nine bone samples. The bone samples were profiled using the 25  $\mu\text{l}$  volume to increase the volume of the DNA input in the PCRs to 6.25  $\mu\text{l}$ . Seven samples, where the total DNA input ranged from 0.2575 ng to 2.0444 ng/reaction, showed similar percentage of detected alleles to other kits previously used (Table S1). However, two samples that had lower concentrations and higher DIs of 0.0173 ng/ $\mu\text{l}$  (DI: 57.7) and 0.0194 ng/ $\mu\text{l}$  (DI: 16.2) (total DNA input 0.1081 ng and 0.1213 ng), the performance deteriorated, both in absolute terms, and in comparison to other kits. It is notable that the capacity of DNA quantity in the other kits (15  $\mu\text{l}$ ) allowed 2.4 fold more DNA to be added to the reaction



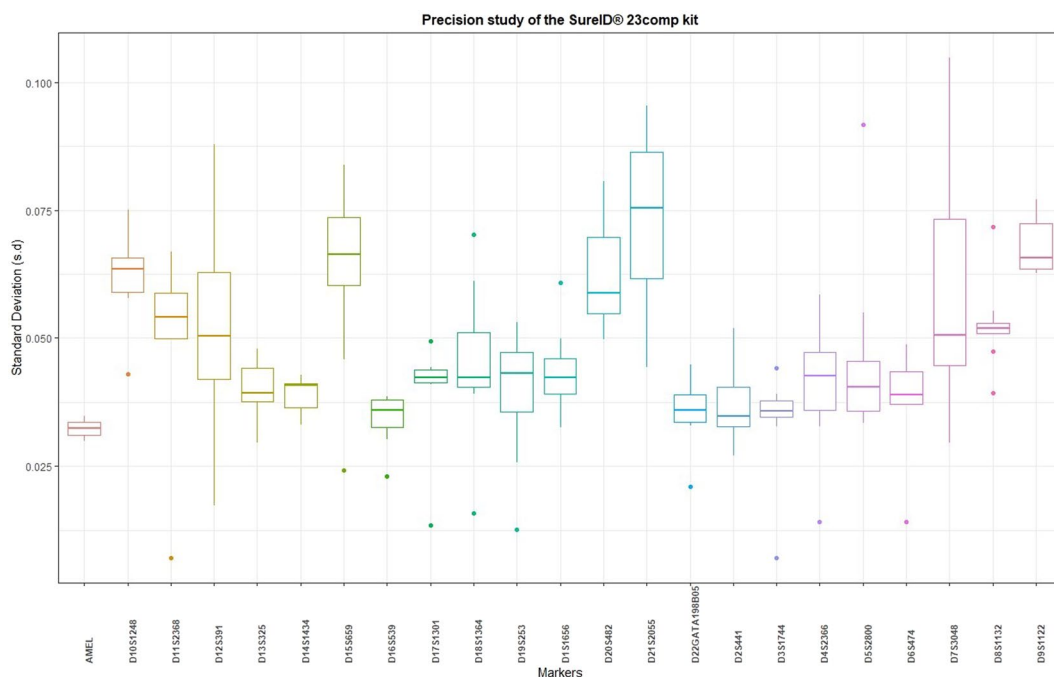
**Figure 2.** Peak balance ratios study for the D21S2055 locus. This figure shows a study of the correlation between the size differences between heterozygous alleles and the peak balance ratio for the D21S2055 locus using data of 500 samples. The peak ratios of all genotypes that have the same size difference (nt) (e.g. the genotypes 13, 17; 14, 18; and 15, 19 have the same size difference of 4 nt) were averaged and are represented by the black dots. The blue line shows the smoothed mean of the peak ratios. Heterozygote alleles with >50 nucleotides differences showed peak ratios <45%.

compared to the SureID 23comp (6.25 µl). Overall, the sensitivity tests when using the DNA control, demonstrated the robustness of generating full profiles even below the recommended DNA concentrations and showed similar sensitivity to other commonly used STR kits. However, this kit demonstrated reduced sensitivity with the bone samples, which is most likely due to the limited capacity of DNA input compared to other kits (Table S2). Although this kit was designed as a supplementary kit for forensic genetics laboratories, some cases may involve human remains (e.g. disaster victim identification (DVI)). Therefore, increasing the concentrations of the master and primer mixes (e.g. to 2X) would permit additional space for more DNA input especially for highly degraded samples.

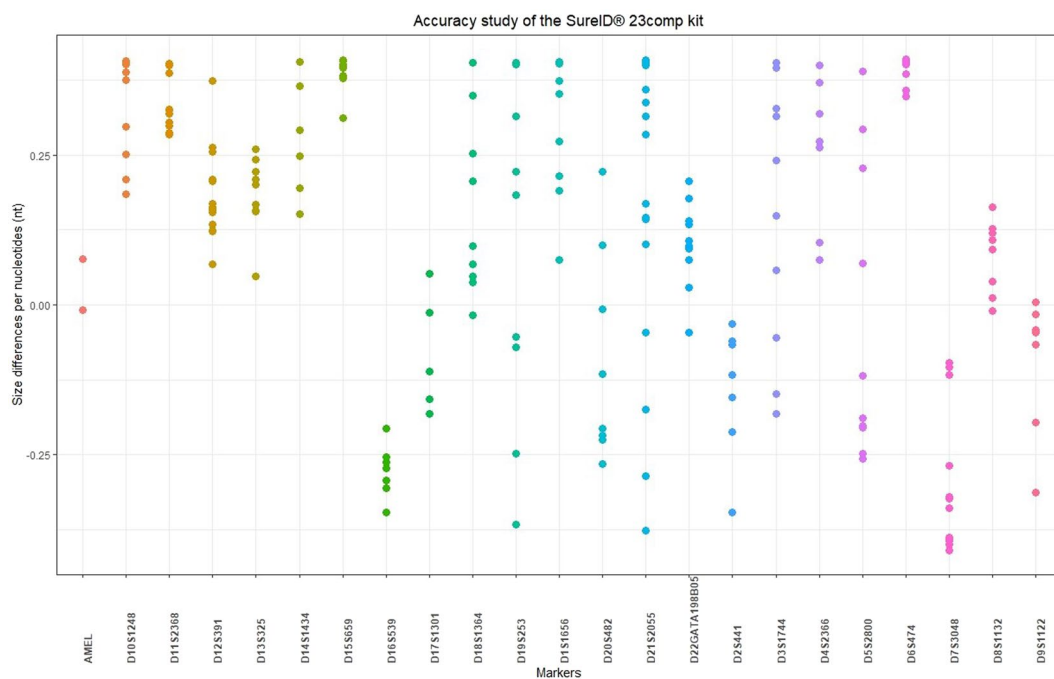
Peak balances evaluation started with measuring of the optimal DNA quantity for the 10 µl volume using the first 90 samples that were tested using three different DNA quantities (0.5, 0.35, and 0.25) ng. With all template amounts the minimum peak balance ratios were >68%, which meets the criteria set out in the ENFSI guidelines (>60%). The DNA input of 0.5 ng achieved the most balanced heterozygous peaks with an average of 88.31% (Table S3), the values were which are similar to ratios observed when testing other kits, for example Investigator HDplex Kit<sup>21</sup>. The D21S2055 showed the lowest degree of balance at all template concentrations with ratios of 73.11% at 0.5 ng, 79.75% at 0.35 ng and 68.12% at 0.25 ng (Table S3). The remaining 410 samples were successfully profiled using the 10 µl volume and 0.5 ng of DNA input. Overall, the intra-locus balances were 81.8% (D21S2055) - 96.9% (D16S539), the intra-dye balances 71.9% (TAMRA) - 82.6% (JOE), and the inter-dye balances >43%. These figures are consistent with the recommended levels that are >70% for intra-locus balance, 50% for intra-dye balance, and >30% for inter-dye balance<sup>32</sup>. The peak imbalances of the D21S2055 became less than 50% when the size difference between heterozygous alleles was more than ten repeats (40 nt) (Fig. 2). The peak balances further decreased to <45% when the size difference became >50 nt (>12 repeats). For example, an average of 43.5% for the genotypes (16.1, 34) (two samples), and 31.8% for the genotype (16.1, 36) (one sample) (Fig. 2). This locus is the longest marker in this kit (332 bp to 420 bp) and has the highest number of possible alleles (23 alleles: 16.1 to 38).

For the precision study, the data of 22,975 alleles (23,000 alleles from 500 samples excluding 25 alleles with a single observation) were used to calculate the standard deviation (s.d.) of the fragment sizes of each allele at a locus. Overall, the maximum s.d. was 0.1048 nucleotide (nt) observed in allele 21 at D7S3048 and the minimum was 0.0071 nt observed in allele 22 at D3S1744 (Fig. 3). To measure the accuracy of the kit, the average sizes of each allele in the data of the 500 samples and in 21 allelic ladders were compared to the actual size values of the corresponding allele (actual sizes provided by the manufacturer). All alleles fell within the range of  $\pm 0.41$  nt, where allele 17 at D6S474 (0.4096 nt) and allele 26 at D7S3048 (-0.4084 nt) recorded the highest difference comparing to the actual sizes (Fig. 4). The precision and the accuracy tests demonstrated the capability of detecting heterozygous alleles that differ by a single nucleotide and demonstrate that it is unlikely for any allele to be sized





**Figure 3.** Precision study of the SureID 23comp Kit. The figure shows standard deviation (s.d) values of the fragment sizes of 22,981 alleles generated from 500 samples tested by the SureID 23comp. The highest s.d. was observed in allele 21 at D7S3048 (0.1048 nt).



**Figure 4.** Accuracy study of the SureID 23comp Kit. The average of the size values of each allele in the data of the 500 samples and in 21 allelic ladders were compared to the actual sizes of the corresponding allele (actual sizes provided by the manufacturer). The size differences per nucleotides were calculated and are represented by the coloured dots. All alleles fell within the range of  $\pm 0.41$  nt of the allelic window; the largest differences were seen at D6S474 allele 17 (0.4096 nt) and D7S3048 allele 26 ( $-0.4084$  nt).

out of the designated window ( $\pm 0.5$  nt). The SureID 23comp was reliably able to detect genotypes where the difference between the alleles was a single nucleotide, for example (11.3, 12) at D2S441, (15.3, 16) and (16.3, 17) at D1S1656 (100% concordant with GlobalFiler genotypes).

STRs	Allele	frequency		STRs	Allele	frequency	
		Saudi	Ningbo			Saudi	Ningbo
D18S1364	11	0.001	0.002	D13S325	26.3	0.002	0
D1S1656	7	0.001	0		27.3	0.001	0
	8	0.001	0	D8S1132	13.1	0.001	0
	10	0.004	0.001		15	0.003	0
	14.3	0.002	0	D7S3048	30	0.001	0
	15.3	0.040	0	D2S441	8.3	0.001	0
	16.3	0.061	0.007		9	0.005	0
	18	0.003	0.011		11.3	0.066	0
	19.3	0.006	0.003		13.3	0.001	0
	20.3	0.001	0.002	D19S253	6	0.004	0
D9S1122	7	0.001	0		16	0.001	0
D4S2366	16	0.002	0	D22GATA198B05	11.2	0.001	0
D3S1744	12	0.001	0		12	0.004	0
D12S391	18.3	0.005	0	D6S474	10	0.001	0
	19.1	0.001	0	D14S1434	16	0.004	0.004
	19.3	0.004	0	D15S659	6	0.001	0
	27	0.003	0.003		7	0.003	0

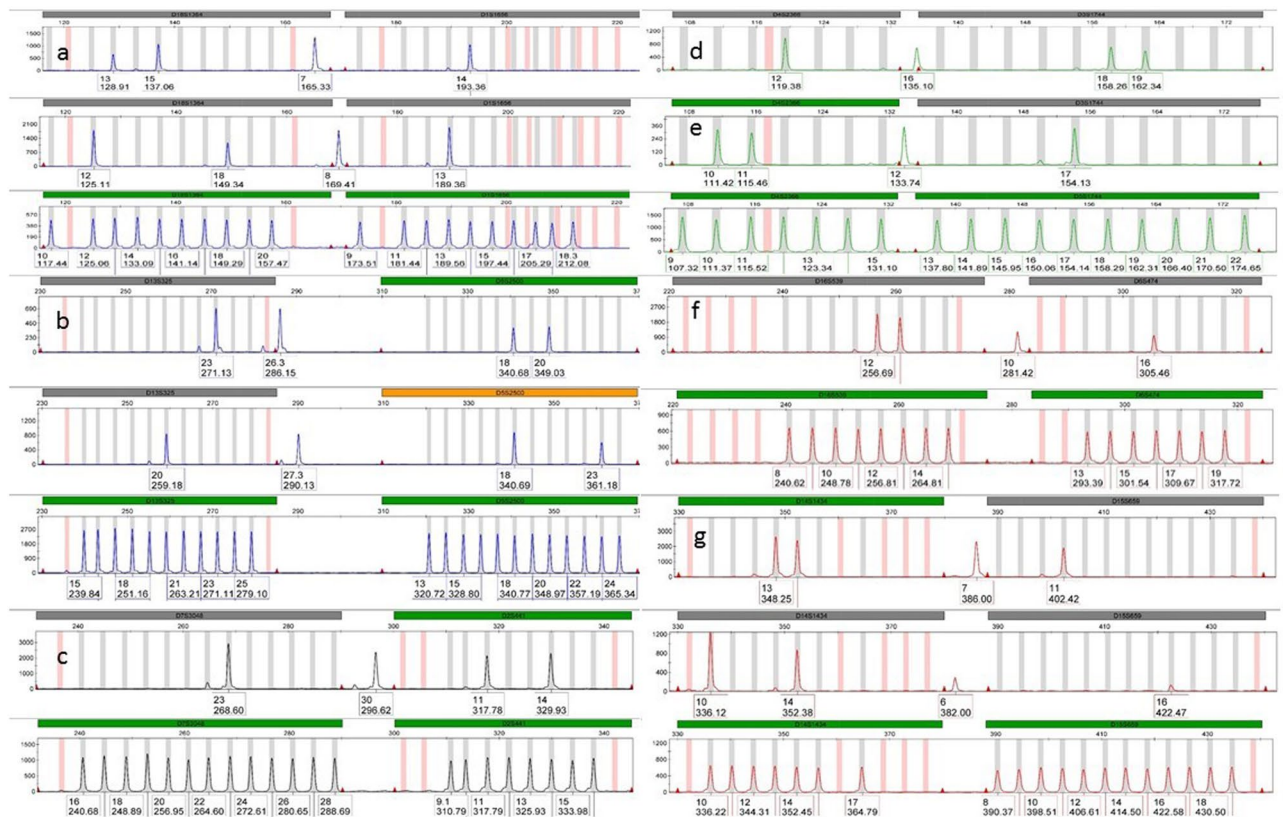
**Table 2.** Alleles not represented by the allelic ladder of SureID 23comp Kit detected in the population of Saudi Arabia. Thirty-four alleles were detected at 15 STRs that are not represented by the allelic ladder of the kit. It shows also the frequency of these alleles in Ningbo population (data provided by the Health Gene Technologies) and in the population of Saudi Arabia. The frequencies of detected alleles in the population of Saudi Arabia ranged from 0.001 (one observation) to 0.066 (66 observations).

Stutter artefacts are common to all PCR-based STR analysis and the most common type of stutter is a peak with one repeat smaller than the true allele<sup>34</sup>. In this study, the average of the stutter peak ratios was 9.18% and the range was from 3.8% for D2S441 to 16.15% for D12S391 (Fig. S4).

This kit provides an allelic ladder representing 232 alleles that are supported by 53 additional bins for variant alleles (Fig. S5). After analysing the 500 samples, 34 alleles in 15 loci were not represented by the allelic ladder, three of which had been observed  $\geq 40$  times (Table 2). In addition, ten of these alleles were situated outside the designated window of their loci: alleles 7 and 8 at D1S1656, 26.3 and 27.3 at D13S325, allele 16 at D4S2366, allele 12 at D3S1744, allele 30 at D7S3048, allele 10 at D6S474 and alleles 6 and 7 at D15S659 (Fig. 5a–g). One allele (7 at D1S1656) was situated under the designated area of D18S1364 locus (Fig. 5a). Although this allele could, in principle, belong to D18S1364 forming triplet allele genotype, it was confirmed by sequencing that it belongs to D1S1656<sup>35</sup>. It is not necessary for an allelic ladder to represent all rare alleles, however, alleles outside the designated window of a locus may be misinterpreted especially when adjacent loci are homozygous. Examining data of 256 samples collected from the population of Ningbo, China (data provided by the Health Gene Technologies) (Table 2), most alleles present in the Saudi Arabian population but not present in the allelic ladder were absent in the Ningbo population.

The concordance study was also carried out by comparing data of the 500 samples obtained from this study and that generated using the GlobalFiler kit<sup>26</sup>. The five common loci (D1S1656, D2S441, D10S1248, D12S391 and D16S539) showed 100% concordance. In addition, alleles generated from the bone samples using the SureID 23comp Kit at the common loci were concordant with alleles generated using the other kits. In addition, the amelogenin showed concordant genotypes to those generated by the GlobalFiler kit.

**Population study.** Assuming independence, the combined match probability (CMP) for the 22 STRs was  $7.4 \times 10^{-27}$ , and the combined power of exclusion (CPE) was 0.999998692. D21S2055 was the most informative locus with a match probability (MP) of 0.016, and D17S1301 was the least informative locus with a MP of 0.162. Heterozygosity ranged from 0.624 (D20S482) to 0.89 (D21S2055). The number of observed alleles per locus varied from 7 alleles in D17S1301 to 20 alleles in D21S2055. Three alleles, allele 14 in D20S482, allele 12 in D17S1301 and allele 12 in D9S1122; showed very high frequencies of 0.477, 0.450 and 0.405 respectively. The frequency of the theoretical most common SureID DNA profile, generated based on the frequencies of the 22 STRs (and assuming heterozygosity) was  $3 \times 10^{-21}$  that equates to 1 in  $3.3 \times 10^{20}$  (Table S4). The CMP of the 22 STRs ( $7.4 \times 10^{-27}$ ) is greater than CMP calculated when using the 21 loci of GlobalFiler kit ( $1.421 \times 10^{-26}$ )<sup>26</sup>. Apart of SE33, the SureID 23comp Kit includes the four most informative loci that have been studied for the population of Saudi Arabia (D21S2055, D12S391, D7S3048, and D1S1656), two of which are included in the GlobalFiler kit (Table S4). Data of the 17 non-CODIS loci was merged to the data of 21 loci (total 38 loci) and was statistically evaluated for the population of Saudi Arabia. Assuming independence, the 38 loci combined provided  $1.7 \times 10^{-46}$  CMP, and 0.99999999934248 CPE, and the theoretically most common DNA profile (assuming heterozygosity at all loci) was  $2.7 \times 10^{-36}$ , which equates to 1 in  $3.7 \times 10^{35}$  (Table S4).



**Figure 5.** Alleles outside the allelic ladder windows of the SureID 23comp Kit. This figure shows ten alleles observed in the population of Saudi Arabia that are not represented in the allelic ladder and were situated outside the designated window of their loci: (a) Alleles 7 and 8 at D1S1656; (b) Alleles 26.3 and 27.3 at D13S325; (c) Allele 30 at D7S3048; (d) Allele 16 at D4S2366; (e) Allele 12 at D3S1744; (f) Allele 10 at D6S474; (g) Alleles 6 and 7 at D15S659. Allele 7 at D1S1656 (a) was situated under the designated area of D18S1364 locus.

The 22 loci were in HWE ( $P$ -value  $> 0.05$ ) except D20S482 which showed significant deviation ( $P$ -value = 0). For the Saudi population, therefore, this locus should not be included in the product rule to calculate the probability of a DNA profile.

The data of the 22 loci of the SureID 23comp was uploaded to the R studio to find out the maximum number of matched loci between any two DNA profiles using the DNA Tools package. In the 500 samples, the maximum number of loci matching between any two samples was 9 out of 22 loci (40% of the 22 loci), which was observed in two sample pairs. One pair of profiles showed partial matching (i.e. one of the two alleles) at 20 out of 22 loci. This illustrates the power of the additional loci for human identification and kinship testing (Table S5).

To assess the SureID 23comp for kinship testing based on the typical paternity index, the combined typical paternity index (CPI) was used to calculate the paternity probabilities with different prior probabilities (Pr): 0.90, 0.50 and 0.10. Assuming that all loci are independent, the probabilities of paternity were 99.9999988% (Pr = 0.90), 99.9999893% (Pr = 0.50) and 99.9999041% (Pr = 0.10), which are higher than those probabilities calculated when using the GlobalFiler kit (Table S6). Under the same assumptions, the 38 loci increased the probabilities of paternity to 99.9999999999972% (Pr = 0.90), 99.99999999999750% (Pr = 0.50) and to 99.999999999997750% (Pr = 0.10) (Table S6).

The ForenSeq DNA Signature Prep (Verogen), when combining the 94 SNPs and the 27 STRs, has shown much higher CMP of  $10^{-67}$ – $10^{-69}$  (length-based STRs calls) and of  $(10^{71}$ – $10^{74})$  (sequence-based STRs calls), where the CMP of the 94 SNPs alone were  $(10^{-38}$ – $10^{-35})$ <sup>36</sup>. In addition, using MPS systems in kinship testing could help in tracking mismatches between tested individuals that have occurred due to mutation in the binding sites of primers<sup>15</sup>. However, this requires additional technology to be implement and is currently not available in many countries.

Five loci included in the SureID 23comp (D18S1364, D4S2366, D3S1744, D20S482, and D17S1301) were designed to have amplicons  $< 200$  bp facilitating characterisation of partial degraded samples. In addition, the primer pairs of the D2S441 and D10S1248 mini-STRs, in the SureID 23comp Kit, were designed to generate longer amplicons (310–446 bp) than primer pairs used in the GlobalFiler kit ( $< 125$  bp), which enabled more space for other mini-STRs, e.g. D17S1301 and D4S2366 ( $< 150$  bp) in the SureID 23comp. As a result, combining both kits will increase the number of STRs  $< 200$  bp to 12 STRs and thus increases the CMP to  $7.9 \times 10^{-14}$  (Table S7).



Chr.	Syntenic Pair	LD test P-value	HapMap proxy SNP position (GRCh37) <sup>b</sup>	Cumulative genetic map distance in cM <sup>b</sup>	Genetic map distance in cM	RFs from Kosambi mapping function <sup>a-c</sup>	RFs estimated based on family studies <sup>d-f</sup>
Chr.2	TPOX	0.97764	1493487	1.6661	88.8129	0.4721 <sup>a</sup>	0.53 <sup>d</sup>
	D2S441		68239020	90.47903			
Chr.2(p-q)	D2S1338	0.99164	218879435	223.4832	133.0042	0.4951 <sup>c</sup>	0.58 <sup>d</sup>
	D2S441		68239020	90.47903			
Chr.2(p-q)	TPOX	0.79338	1493487	1.6661	221.8171	0.4999 <sup>c</sup>	0.51 <sup>d</sup>
	D2S1338		218879435	223.4832			
Chr.3(p-q)	D3S1358	0.99193	45582627	67.1789	90.0624	0.4735 <sup>a</sup>	0.64 <sup>d</sup>
	D3S1744		147092143	157.24131			
Chr.4(p-q)	FGA	0.99771	155508100	156.81293	143.8662	0.4968 <sup>c</sup>	0.51 <sup>d</sup>
	D4S2366		6484806	12.9467			
Chr.5	D5S2800	0.15198	58698677	70.3208	56.3520	0.4050 <sup>b</sup>	N/A
	D5S818		123111652	126.67284			
Chr.5	D5S2800	0.85646	58698677	70.3208	84.1132	0.4666 <sup>c</sup>	N/A
	CSF1PO		149455757	154.43395			
Chr.5	D5S818	0.69008	123111652	126.67284	27.7611	0.2522 <sup>a</sup>	0.18 <sup>d</sup>
	CSF1PO		149455757	154.43395			
Chr.6	SE33	0.99963	88986609	95.44921	23.2133	0.2168 <sup>a</sup>	0.19 <sup>d</sup>
	D6S474		112879893	118.66248			
Chr.7(p-q)	D7S3048	0.98684	21266723	36.14071	64.0605	0.4284 <sup>b</sup>	0.4997 <sup>c</sup>
	D7S820		83789257	100.2012			
Chr.8	D8S1132	0.23577	107330479	119.96228	16.4809	0.1591 <sup>a</sup>	0.1443 <sup>c</sup>
	D8S1179		125907927	136.44313			
Chr.11	TH01	0.20421	2192166	4.48933	28.3996	0.2569 <sup>b</sup>	0.2152 <sup>c</sup>
	D11S2368		19281171	32.88891			
Chr.12	vWA	0.89307	6093924	15.63031	11.9410	0.1172 <sup>a</sup>	0.1259 <sup>c</sup>
	D12S391		12450501	27.57129			
Chr.13	D13S325	0.97422	43173444	44.90825	34.9225	0.3016 <sup>b</sup>	0.2533 <sup>c</sup>
	D13S317		82721723	79.83074			
Chr.18	D18S51	0.04312	60949983	88.92051	2.2970	0.0229 <sup>b</sup>	0.0327 <sup>c</sup>
	D18S1364		63400151	91.21746			
Chr.19(p-q)	D19S433	0.78742	30417603	51.72618	12.4538	0.1228 <sup>b</sup>	0.1101 <sup>f</sup>
	D19S253		15728103	39.27234			
Chr.21	D21S11	1	20554558	14.64555	34.8192	0.3010 <sup>a</sup>	0.32 <sup>d</sup>
	D21S2055		41191871	49.46478			
Chr.22	D22S1045	0.98893	37535663	46.21362	38.8178	0.3253 <sup>b</sup>	N/A
	D22GATA198B05		17651831	7.39585			
Chr.15*	Penta E	—	97377441	124.05054	74.5331	0.4517 <sup>b</sup>	N/A
	D15S659		46371620	49.51748			
Chr.21*	Penta D	—	45056178	59.37591	9.9111	0.0978 <sup>b</sup>	N/A
	D21S2055		41191871	49.46478			

**Table 3.** LD test results for 18 pairs of syntenic loci formed when combining loci included in the SureID 23comp and GlobalFiler kits. This test was performed for data generated from the 500 samples from Saudi population for the 38 loci. No significant deviation from linkage equilibrium (LE) was detected ( $P$ -value = 0.003). RFs that previously derived from HapMap data and from family studies were reviewed from literature<sup>4,12,21,22,24</sup>. RFs derived from HapMap data for D2S1338-D2S441, TPOX-D2S1338, FGA-D4S2366 and D5S2800-CSF1PO pairs, were calculated based on cumulative genetic map distance provided in Phillips<sup>12</sup> and using Excel tool developed by Phillips *et al.*<sup>4</sup>. <sup>a</sup>Data reviewed from Phillips *et al.*<sup>4</sup>. <sup>b</sup>Data reviewed from Phillips<sup>12</sup>. <sup>c</sup>RFs were calculated based on cumulative genetic map distance provided in Phillips<sup>12</sup> and using an excel tool developed by Phillips *et al.*<sup>4</sup>. <sup>d</sup>Data reviewed from Westen *et al.*<sup>21</sup>. <sup>e</sup>Data reviewed from Liu *et al.*<sup>22</sup>. <sup>f</sup>Data reviewed from Wu *et al.*<sup>24</sup>. \*Potential syntenic pairs when combining the SureID 23comp with kits that include the Penta E and Penta D loci (LD was not tested). N/A: No data available from family studies.

**LD and RF study.** Family studies have previously been carried out to estimate recombination fraction (RF) for four syntenic loci residing on the same chromosome's arm vWA-D12S391, D5S818-CSF1PO, TPOX-D2S441 and D21S11-PentaD typed when using any of the commonly used STR kits<sup>21–23</sup>. RF values were 0.17<sup>21</sup>, 0.13<sup>22</sup>, and 0.11<sup>23</sup> for vWA-D12S391; 0.197<sup>37</sup> for D5S818-CSF1PO; 0.53<sup>21</sup> for TPOX-D2S441 and 0.316<sup>37</sup> D21S11-PentaD. The

high-density multi-point SNP data of HapMap was also used to approximate the genetic distance between these syntenic loci and gave RF values of 0.12 for vWA-D12S391, 0.25 for D5S818-CSF1PO, 0.36 for D21S11-PentaD, and of 0.47 for TPOX-D2S441<sup>4</sup>; these values are similar to those generated using family studies.

It was concluded that, for most pedigrees, the RF value of ~0.12 has almost zero effect in any population as long as no linkage disequilibrium was detected<sup>38</sup>. For some pedigrees, where at least one individual has a heterozygote genotype in the both syntenic loci and is involved in at least two transmissions of genetic components, linkage can have a significant effect in the product rule calculation in kinship testing<sup>38</sup>.

However, using the SureID 23comp Kit in conjunction with any of commonly used kits, will increase the number of potentially linked syntenic loci located on the same arm to 12–15 pairs (D5S818-D5S2800, CSF1PO-D5S2800, SE33-D6S474, D8S1179-D8S1132, TH01-D11S2368, D13S317-D13S325, Penta E-D15S659, D18S51-D18S1364, D21S2055-D21S11, D21S2055-Penta D, D22S1045- D22GATA198B05, vWA-D12S391, D5S818-CSF1PO, TPOX-D2S441 and D21S11-PentaD).

The potential linkage between 18 syntenic loci (12 on the same arm and 6 on different arms), which resulted from combining the 38 STRs, were tested using the likelihood ratio test (because the gametic phase of the genotypes is unknown)<sup>29</sup>. Although, significant departure from HWE could invalidate the test, none of the 18 pairs is on chromosome 20 and thus the significant departure of D20S482 from HWE is not relevant. After Bonferroni correction ( $P$ -value = 0.003), no significant deviation from linkage equilibrium (LE) was detected between any of the 18 syntenic loci (Table 3).

Table 3 also reviews RFs that previously derived from HapMap data and from family studies<sup>4,12,21,22,24</sup>. In addition, RFs derived from HapMap data for D2S1338-D2S441, TPOX-D2S1338, FGA-D4S2366 and D5S2800-CSF1PO pairs, were calculated based on cumulative genetic map distance provided in Phillips<sup>12</sup> and using the Excel tool developed by Phillips *et al.*<sup>4</sup>.

The genetic distance (cM) estimated based on the HapMap data ranged from 3.49 cM for D18S51-D18S1364 pair to 226.5 cM for TPOX-D2S1338 pair. Apart from the D18S51-D18S1364 and PentaD-D21S2055 pairs, all potential linked loci showed RF values > 0.12 and thus they can be included in the product rule calculation for kinship testing for most pedigrees. For pairs showing RF values < 0.12, excluding the less informative locus from the probability estimation<sup>23</sup>, which is the D18S1364 for the Saudi population, is an option. However, due to concerns that this may lead to an overestimation or to an underestimation of the strength of an evidence, Gill *et al.*<sup>38</sup> illustrated a methodology by which the RFs can be included in the probability estimation.

## Conclusion

The SureID 23comp was validated following the minimum criteria for validation recommended by the ENFSI and the SWGDAM for forensic applications as a supplementary STR kit. The kit is reproducible, precise, accurate and reliable for forensic application as a supplementary kit and for databasing. The identity of the D5 locus was confirmed and has now been updated by the manufacturer to D5S2800.

The sensitivity tests demonstrate the capability of generating a full profile below the recommended DNA input but showed that the kit was less sensitive compared to other commonly with degraded samples, which was at least in part because of the lower volume of template that can be added. Therefore, increasing the concentration of the reaction mix will allow more space for DNA input to 15  $\mu$ l rather than 6.25  $\mu$ l, which will increase the sensitivity of the kit.

Including additional alleles and allele variations in the available spaces of the allelic ladder will allow specific allele designation and will minimise the need to re-run undesignated alleles. In addition, caution should be taken when using the kit with potentially degraded samples (e.g. DVI cases) due to the peak imbalance of D21S2055.

For kinship testing, typically, the kit achieved a CPI of 93835307.21 that is two times higher the CPI recorded for the GlobalFiler kit and allowed higher paternity probabilities 99.99999988%, 99.99999893% and 99.99999041% in different prior probabilities, demonstrating the informativeness of the STRs included in the kit. In addition, utilising the 38 loci will increase the certainty of the paternity probabilities, using different prior probabilities, to 99.99999999999972%, 99.999999999999750% and to 99.9999999999997750%, which is useful for complex cases. However, caution should be considered when estimating likelihoods if the D18S51-D18S1364 and/or PentaD-D21S2055 pairs are included.

Overall, this study evaluates the utility of the SureID 23comp as a supplementary kit for kinship testing and determined that the kit met the criteria commonly used in forensic genetics laboratories. The kit allows the analysis of 17 non-CODIS loci and increases likelihood ratios, and thereby has the potential to increase the level of confidence in conclusions in complex kinship tests using already established technology.

## Data availability

Data can be requested via: whgoodwin@uclan.ac.uk or hmhalsafiah@uclan.ac.uk.

Received: 29 November 2018; Accepted: 21 October 2019;

Published online: 14 November 2019

## References

- Gill, P., Fereday, L., Morling, N. & Schneider, P. M. New multiplexes for Europe-amendments and clarification of strategic development. *Forensic Sci. Int.* **163**, 155–157 (2006).
- Hares, D. R. Selection and implementation of expanded CODIS core loci in the United States. *Forensic. Sci. Int. Genet.* **17**, 33–34 (2015).
- Goodwin, W. *et al.* Case study: paternity testing when 21 loci are not enough. *Int. Congr. Ser.* **1261**, 460–462 (2004).
- Phillips, C. *et al.* The recombination landscape around forensic STRs: Accurate measurement of genetic distances between syntenic STR pairs using HapMap high density SNP data. *Forensic. Sci. Int. Genet.* **6**, 354–365 (2012).

5. Poetsch, M., Preusse-Prange, A., Schwark, T. & von Wurmb-Schwark, N. The new guidelines for paternity analysis in Germany-how many STR loci are necessary when investigating duo cases? *Int. J. Legal Med.* **127**, 731–734 (2013).
6. O'Connor, K. L., Butts, E., Hill, C. R., Butler, J. M. & Vallone P. M. *Evaluating the effect of additional forensic loci on likelihood ratio values for complex kinship analysis* (The 21st International Symposium on Human Identification, 2010).
7. Carboni, I., Iozzi, S., Nutini, A. L., Torricelli, F. & Ricci, U. Improving complex kinship analyses with additional STR loci. *Electrophoresis* **35**, 3145–3151 (2014).
8. Hill, C. R., Butler, J. M. & Vallone, P. M. A 26plex autosomal STR assay to aid human identity testing\*. *J. Forensic Sci.* **54**, 1008–1015 (2009).
9. Li, J. *et al.* Validation of the Microreader™ 23sp ID system: A new STR 23-plex system for forensic application. *Forensic. Sci. Int. Genet.* **27**, 67–73 (2017).
10. Fu, X. *et al.* The validation of Goldeneye™ DNA ID 22NC kit and the genetic polymorphism of 21 short tandem repeat loci in the Chinese Hunan Han population. *J. Forensic Sci. Med.* **4**, 122–128 (2018).
11. Zhu, B. *et al.* Developmental validation of the AGCU 21 + 1 STR kit: A novel multiplex assay for forensic application. *Electrophoresis* **36**, 271–276 (2015).
12. Phillips, C. A genomic audit of newly-adopted autosomal STRs for forensic identification. *Forensic. Sci. Int. Genet.* **29**, 193–204 (2017).
13. Li, H. *et al.* Applying massively parallel sequencing to paternity testing on the Ion Torrent Personal Genome Machine. *Forensic. Sci. Int. Genet.* **31**, 155–159 (2017).
14. Montano, E. A. *et al.* Optimization of the Promega PowerSeq™ Auto/Y system for efficient integration within a forensic DNA laboratory. *Forensic. Sci. Int. Genet.* **32**, 26–32 (2018).
15. Li, R. *et al.* Improved pairwise kinship analysis using massively parallel sequencing. *Forensic. Sci. Int. Genet.* **38**, 77–85 (2019).
16. Ma, Y., Kuang, J., Nie, T., Zhu, W. & Yang, Z. Next generation sequencing: Improved resolution for paternal/maternal duos analysis. *Forensic. Sci. Int. Genet.* **24**, 83–85 (2016).
17. Qiagen. Investigator HDplex Handbook. Preprint at, <https://www.qiagen.com/gb/resources/resourcedetail?id=7d1661bd-a47b-4b19-a882-357a61b48c64&lang=en> (2012).
18. Promega Corporation. PowerPlex® CS7 System Technical Manual. Preprint at, <https://www.promega.co.uk/products/genetic-identity/genetic-identity-workflow/str-amplification/powerplex-cs7-system/?catNum=DC6613> (2016).
19. Phillips, C. *et al.* D5S2500 is an ambiguously characterized STR: Identification and description of forensic microsatellites in the genomics age. *Forensic. Sci. Int. Genet.* **23**, 19–24 (2016).
20. Lobo, I. & Shaw, K. Thomas Hunt Morgan, genetic recombination, and gene mapping. *Nature Education* **1**(1):205, <https://www.nature.com/scitable/topicpage/thomas-hunt-morgan-genetic-recombination-and-gene-496> (2008).
21. Westen, A. A. *et al.* Combining results of forensic STR kits: HDplex validation including allelic association and linkage testing with NGM and Identifiler loci. *Int. J. Legal Med.* **126**, 781–789 (2012).
22. Liu, Q. L. *et al.* Recombination analysis of autosomal short tandem repeats in Chinese Han families. *Electrophoresis* **35**, 883–887 (2014).
23. Budowle, B. *et al.* Population genetic analyses of the NGM STR loci. *Int. J. Legal Med.* **125**, 101–109 (2011).
24. Wu, W. *et al.* Analysis of linkage and linkage disequilibrium for syntenic STRs on 12 chromosomes. *Int. J. Legal Med.* **128**, 735–739 (2014).
25. Iyavoo, S., Afolabi, O., Boggi, B., Bernotaite, A. & Haizel, T. Population genetics data for 22 autosomal STR loci in European, South Asian and African populations using SureID® 23comp Human DNA Identification Kit. *Forensic Sci. Int.* **301**, 174–181 (2019).
26. Alsafiah, H. M., Goodwin, W. H., Hadi, S., Alshaiqi, M. A. & Wepeba, P. Population genetic data for 21 autosomal STR loci for the Saudi Arabian population using the GlobalFiler® PCR amplification kit. *Forensic. Sci. Int. Genet.* **31**, e59–e61 (2017).
27. RStudio Team. Integrated Development for R. RStudio, Inc., Boston, MA, <http://www.rstudio.com> (2016).
28. James, M. Curran Torben Tvedebrink. DNAtools: Tools for empirical testing of DNA match probabilities, <https://CRAN.R-project.org/package=DNAtools> (2017).
29. Excoffier, L., Laval, G. & Schneider, S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinform Online* **1**, 47–50 (2007).
30. Collins, P. J. *et al.* Developmental validation of a single-tube amplification of the 13 CODIS STR loci, D2S1338, D19S433, and amelogenin: The AmpFISTR® identifiler® PCR amplification kit. *J. Forensic Sci.* **49**, 1265–1277 (2004).
31. Ensenberger, M. G. *et al.* Developmental validation of the PowerPlex® Fusion 6C System. *Forensic. Sci. Int. Genet.* **21**, 134–144 (2016).
32. Liu, Y. *et al.* Developmental validation of a 6-dye typing system with 27 loci and application in Han population of China. *Sci. Rep.* **7**, 4706-017–04548-1 (2017).
33. Lin, S. W., Li, C. & Ip, S. C. Y. A selection guide for the new generation 6-dye DNA profiling systems. *Forensic. Sci. Int. Genet.* **30**, 34–42 (2017).
34. Brookes, C., Bright, J. A., Harbison, S. & Buckleton, J. Characterising stutter in forensic STR multiplexes. *Forensic. Sci. Int. Genet.* **6**, 58–63 (2012).
35. Alsafiah, H. M., Iyengar, A., Hadi, S., Alshlash, W. M. & Goodwin, W. Sequence data of six unusual alleles at SE33 and D1S1656 STR Loci. *Electrophoresis* **39**, 2471–2476 (2018).
36. Churchill, J. D., Novroski, N. M. M., King, J. L., Seah, L. H. & Budowle, B. Population and performance analyses of four major populations with Illumina's FGx Forensic Genomics System. *Forensic. Sci. Int. Genet.* **30**, 81–92 (2017).
37. Buckleton, J. & Triggs, C. The effect of linkage on the calculation of DNA match probabilities for siblings and half siblings. *Forensic Sci. Int.* **160**, 193–199 (2006).
38. Gill, P., Phillips, C., McGovern, C., Bright, J. & Buckleton, J. An evaluation of potential allelic association between the STRs vWA and D12S391: Implications in criminal casework and applications to short pedigrees. *Forensic. Sci. Int. Genet.* **6**, 477–486 (2012).
39. Phillips, C. *et al.* "The devil's in the detail": Release of an expanded, enhanced and dynamically revised forensic STR Sequence Guide. *Forensic. Sci. Int. Genet.* **34**, 162–169 (2018).

## Acknowledgements

Alsafiah, Hussain M. and this work were supported by the Saudi Arabian cultural bureau in London (UKSACB). The UKSACB is a part of the ministry of education in Saudi Arabia and supports Saudi students in the UK. We would like to thank Ms Ausma Bernotaite (Health Gene Technologies) for the technical support.

## Author contributions

Hussain M. Alsafiah Collection of the reference samples, DNA extraction from the reference samples, DNA Quantification, Genotyping the reference samples and the bone samples, reproducibility tests, sensitivity and stochastic tests, stability tests, peak ratios, stutter ratios, statistical evaluation, precision and accuracy tests, LR study, RFs calculation, and writing. Ali Aljanabi Collection of the bone samples, DNA extraction from the bone

samples, DNA Quantification, Genotyping the bone samples, reproducibility tests, sensitivity and stochastic tests, stability tests. Sibte Hadi co-writer of paper. Saleh. S. Alturayef co-writer of paper. Will Goodwin Supervisor of Studies and co-writer of paper.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-019-52838-7>.

**Correspondence** and requests for materials should be addressed to H.M.A. or W.G.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019