Int J Legal Med (2012) 126:781–789
DOI 10.1007/s00414-012-0724-4

ORIGINAL ARTICLE

Combining results of forensic STR kits: HDplex validation including allelic association and linkage testing with NGM and Identifiler loci

Antoinette A. Westen · Hinda Haned · Laurens J. W. Grol · Joyce Harteveld · Kristiaan J. van der Gaag · Peter de Knijff · Titia Sijen

Received: 25 April 2012 / Accepted: 12 June 2012 / Published online: 3 July 2012
© Springer-Verlag 2012

Abstract The autosomal short tandem repeat (STR) kits that are currently used in forensic science have a high discrimination power. However, this discrimination power is sometimes not sufficient for complex kinship analyses or decreases when alleles are missing due to degradation of the DNA. The Investigator HDplex kit contains nine STRs that are additional to the commonly used forensic markers, and we validated this kit to assist human identification. With the increasing number of markers it becomes inevitable that forensic and kinship analyses include two or more STRs present on the same chromosome. To examine whether such markers can be regarded as independent, we evaluated the 30 STRs present in NGM, Identifiler and HDplex. Among these 30 markers, 17 syntenic STR pairs can be formed. Allelic association between these pairs was examined using 335 Dutch reference samples and no linkage disequilibrium was detected, which makes it possible to use the product rule for profile probability calculations in unrelated individuals. Linkage between syntenic STRs was studied by determining the recombination fraction between them in five three-generation CEPH families. The recombination fractions were compared to the physical and genetic distances between the markers. For most types of pedigrees, the kinship analyses can be performed using the product rule, and for those cases that require an alternative calculation method (Gill et al., Forensic Sci Int Genet 6:477–486, 2011), the recombination fractions as determined in this study can be used. Finally, we calculated the (combined) match probabilities, for the supplementary genotyping results of HDplex, NGM and Identifiler.

Keywords Forensic science · Short tandem repeat (STR) · Investigator HDplex kit · Next Generation Multiplex (NGM) · Linkage disequilibrium (LD) · Kinship analysis

Electronic supplementary material The online version of this article (doi:10.1007/s00414-012-0724-4) contains supplementary material, which is available to authorized users.

A. A. Westen · H. Haned · L. J. W. Grol · J. Harteveld · T. Sijen (✉)
Department of Human Biological Traces (R&D),
Netherlands Forensic Institute,
P.O. Box 24044, 2490 AA, The Hague,
The Netherlands
e-mail: t.sijen@nfi.minvenj.nl

A. A. Westen
e-mail: t.westen@nfi.minvenj.nl

H. Haned
e-mail: h.haned@nfi.minvenj.nl

L. J. W. Grol
e-mail: l.grol@nfi.minvenj.nl

J. Harteveld
e-mail: j.harteveld@nfi.minvenj.nl

K. J. van der Gaag · P. de Knijff
Forensic Laboratory for DNA Research,
Leiden University Medical Centre,
P.O. Box 9503, 2300 RA, Leiden, The Netherlands

K. J. van der Gaag
e-mail: k.j.van_der_gaag@lumc.nl

P. de Knijff
e-mail: p.de_knijff@lumc.nl

Introduction

The expansion of the European standard set (ESS) of autosomal short tandem repeat (STR) markers [2, 3] has resulted in the development of new forensic STR kits such as the AmpF*l*STR® NGM™ (SElect) PCR Amplification Kit (Applied Biosystems (AB), Foster City, CA, USA), the PowerPlex® ESX and ESI Systems (Promega Corporation (Promega), Madison, WI, USA) and the Investigator® ESSplex Plus Kit (Qiagen Benelux B.V. (Qiagen), Venlo, the

Netherlands). These kits combine the gender-determining Amelogenin marker with the ten commonly used AmpF*l*STR® SGM Plus™ (AB) STRs and the five new ESS markers, with or without the addition of SE33 (ACTBP2) as a 16th STR. The five new ESS markers show a higher discrimination power than the five STRs in the AmpF*l*STR® Identifiler Kit (AB) that are additional to SGM Plus [4]. Nevertheless, in complex kinship analyses or in (missing person) cases in which the DNA has been severely degraded, the power of discrimination of the DNA profile may not be high enough to identify a person. In these cases, it is opportune to analyse additional highly discriminative STR markers.

In 2010, the Investigator® HDplex Kit (Qiagen; formerly known as Mentype® Chimera® PCR Amplification Kit, Biotype Diagnostic GmbH, Dresden, Germany) became available in the European forensic market. This kit contains nine highly discriminative STRs (D2S1360, D3S1744, D4S2366, D5S2500, D6S474, D7S1517, D8S1132, D10S2325 and D21S2055), next to the Amelogenin gender marker and three STRs that are included in the above-mentioned commercial forensic kits: D12S391, D18S51 and SE33. We validated the HDplex for human identification and generated allele frequencies based on 335 Dutch reference samples.

With the increasing number of forensically available STRs, it becomes inevitable that two or more markers are present in the same chromosome (a.k.a. syntenic markers). To ensure independent inheritance, syntenic markers are ideally situated on different arms of the chromosome or at least 50 centiMorgans (cM) apart. The latter means that there is at least a 50 % chance of recombination between the two syntenic markers, which are therefore regarded as unlinked [5]. The vWA and D12S391 markers are only 6.36 Mb apart, and several studies have addressed the possible linkage or allelic association (a.k.a. linkage disequilibrium) between them [1, 4, 6–8]. From their results, the authors expect no interpretation problems at the population level, but they do express their concerns for the interpretation of data from closely related individuals. In our study, we evaluated the 30 STRs present in NGM (AB), Identifiler (AB) and HDplex (Qiagen), from which 17 pairs of syntenic loci can be formed. For these pairs, we determined the recombination fraction in five three-generation CEPH families and compared our results with those of Phillips et al. [7], who studied the recombination landscape around a broad spectrum of forensically relevant STRs based on HapMap data. In addition, we evaluated whether linkage disequilibrium was detectable at the population level. Finally, we determined the random match probability when combining two of the three (NGM, Identifiler or HDplex) or all three kits.

Material and methods

DNA samples

Validation tests for the HDplex were performed using the pristine DNA samples DNA XY5 (positive control HDplex (Qiagen)), DNA007 (positive control NGM (AB)) and hDNA (Quantifiler™ Human DNA standard (AB)). 2085 DNA samples, representative for the Dutch population, were used with informed consent of the donors. Five three-generation CEPH families (Coriell Institute, Camden, NJ, USA) were analysed: one French family with pedigree number 0066 and four Utah families with pedigree numbers 1362, 1423, 1454 and 1463. Each family consisted of four grandparents, two parents and seven to 11 children. To prepare artificially degraded DNA, hDNA (200 ng/μL) was treated with UV–light for 0, 10, 30 and 60 min, following the protocol described in Westen et al. [9].

PCR amplification, capillary electrophoresis and DNA profile analysis

All 2,085 population samples were amplified with the NGM (AB) and Identifiler (AB) PCR amplification kits (de Knijff and Sijen, in preparation). Three hundred thirty-five of the 2,085 Dutch population samples were amplified with the HDplex kit (Qiagen) using a 750-pg PCR input. The CEPH family samples were amplified with all three kits using PCR inputs of 500 pg for NGM, 1 ng for Identifiler and 750 pg for HDplex. During the HDplex validation, study various amounts of template DNA were used: a series from 8 to 750 pg during sensitivity assays, 750 pg for mixture studies, 1 ng when assessing resistance to PCR inhibitors and 1 ng artificially degraded hDNA (1 μL of 200-fold diluted UV-treated hDNA). PCR products were detected by capillary electrophoresis (CE) on an ABI Prism 3130xl Genetic Analyzer (AB). PCR amplifications and CE detection were performed according to the manufacturer's instructions. DNA profiles were analysed using GeneMapper® ID-X v. 1.1.1 (AB). When analysing HDplex profiles, we found that all D10S2325 peaks were detected on the left-hand side within or adjacent to the bin for the amplified samples, but not for the allelic ladder. We solved this issue by diluting the allelic ladder 1,000-fold and re-amplifying it for 15 additional PCR cycles [10]. We suspect that the D10S2325 primers used to amplify the allelic ladder and those provided in the kit originate from different synthesis batches, resulting in a shift of the amplified alleles of approximately 0.5 nt compared to the allelic ladder.

Statistical analyses

For the 2,085 population samples, allele frequencies were calculated using the Excel Microsatellite Toolkit [11]. Departure from the Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) testing was performed using Arlequin v. 3.5 [12]. HWE exact tests were performed with 100,000 dememorisation steps and 1,000,000 steps in the Markov chain. LD between pairs of loci was calculated with three initial conditions for the Expectation–Maximisation algorithm and 10,000 permutations. Power of discrimination, power of exclusion, polymorphic information content and match probabilities were calculated using the Genetic Identity PowerStats v. 12 Excel spreadsheet (Promega) [13]. We estimated the genotypic linkage disequilibrium correlation coefficients (r^2) using the LINKDIS program [14] implemented in GENETIX v. 4.05.2 [15]. Allele frequencies and descriptive statistics for NGM and Identifiler will be published elsewhere (de Knijff et al., in preparation). In order to determine the recombination fraction between pairs of syntenic loci, the three-generation CEPH family data were analysed with LINKAGE [16, 17] (see supplementary note 1 for additional information).

Results and discussion

Validation of the Investigator HDplex kit

In order to test the sensitivity of the HDplex, a dilution series between 750 and 8 pg DNA XY5 was genotyped in threefold. Full profiles were detected down to 63 pg DNA (for one of the three replicates, Table 1). In our laboratory, the optimal input was found to be 750 pg template DNA, which resulted in heterozygous peak balances between 0.79 and 0.93 (calculated by dividing the height of the lower peak by that of the higher peak) compared to 0.64 to 0.92 for 500 pg template DNA, and the average peak heights for heterozygous alleles were around 2,300 rfu (relative fluorescence unit) for 750 pg and around 1,376 rfu for 500 pg template DNA.

For the analysis of low template DNA, a comparison was made between the standard protocol (30 PCR cycles with 3 kV/10 s CE injection settings), two additional PCR cycles (32 cycles with CE at 3 kV/10 s, as recommended by the manufacturer for DNA inputs <100 pg) and increasing the CE injection voltage to 9 kV (30 PCR cycles with CE at 9 kV/10 s) [18]. A dilution series of 63, 31, 16 and 8 pg DNA007 was amplified in threefold. Supplementary Fig. 1a shows that both increased cycling and increased CE injection voltage are functional to obtain a higher percentage of detected alleles, with 9 kV injection voltage giving a slightly higher percentage of detected alleles for most profiles. This method is easily performed, without the use of additional DNA extract, to increase the sensitivity of STR typing [18]. In our laboratory (NFI), for low template DNA samples, multiple independent amplifications are performed, and the results are interpreted in combination with the consensus method as described by Benschop et al. [19], irrespective of the sensitising technique used.

Artificially degraded samples were genotyped in duplicate for both HDplex and NGM. Supplementary Fig. 1b shows that NGM profiling is less sensitive to DNA degradation than HDplex analysis, as full NGM profiles are found up to 30 min of UV irradiation, while the average percentage of detected alleles for HDplex starts to decrease with 10 min of irradiation. Both for HDplex and NGM, allele dropout is most prominent in larger sized markers, which is in agreement with earlier findings [20–22]. HDplex carries relatively more large-sized amplicons, ranging from 70 to 475 bp, while NGM spans 76 to 352 bp. This is probably due to the fact that HDplex makes use of a four-dye chemistry, instead of a five-dye chemistry as used with NGM and Identifiler, thereby providing less room for markers with small amplicon sizes.

Further characteristics, for which the performance of the HDplex was tested, were resistance to PCR inhibitors and DNA mixture analysis. Also for these aspects, the HDplex performed within the boundaries we had set (results not shown); HDplex tolerated 50 μM hematin and correctly analysed two- and three-person mixtures that were within the sensitivity range of the kit (Table 1).

DNA profile characteristics

Several aspects of the HDplex, such as the inter-locus balance, the intra-locus peak height ratio and the stutter ratio

Table 1 HDplex sensitivity characteristics based on a dilution series of pristine XY5 DNA (n=3)

	750 pg	500 pg	250 pg	125 pg	63 pg	31 pg	16 pg	8 pg
Average % detected alleles	100±0.00	100±0.00	100±0.00	100±0.00	94±0.06	72±0.21	51±0.24	35±0.10
Average peak height (rfu)	2,293±309	1,376±309	1,009±233	731±94	245±118	185±63	108±25	83±10
Minimum peak height (rfu)	1,182	1,022	682	632	193	130	58	30
Maximum peak height (rfu)	2,658	1,493	967	1,460	489	547	190	146

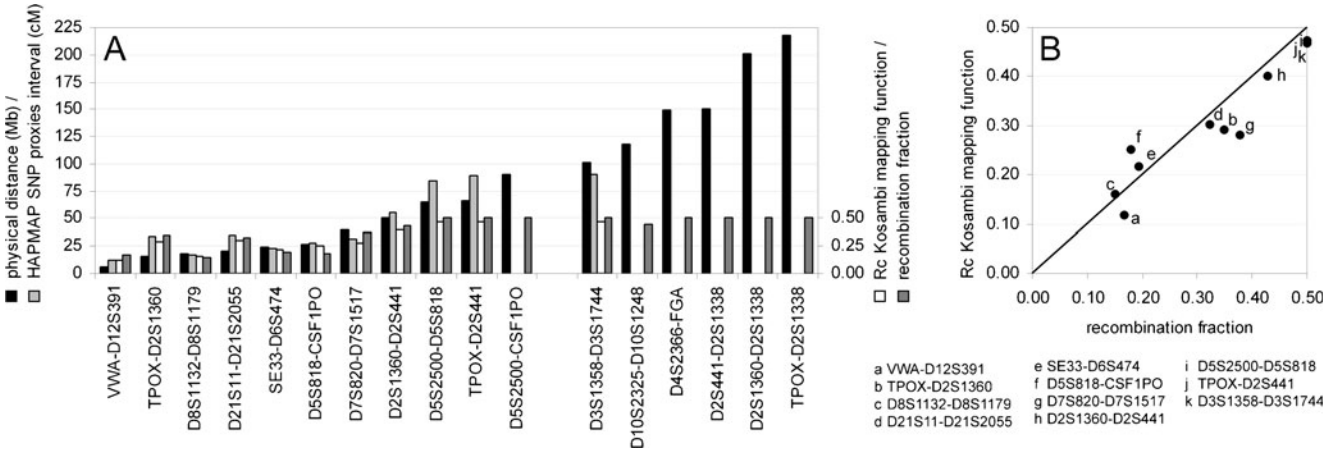

Fig. 1 **a** Four aspects of the 17 syntenic STR pairs: (1) the physical distance in Mb (*black bars*), (2) the cM interval of the closest HAPMAP SNP proxies [7] (*light gray bars*) together with (3) the associated recombination fraction (Rc) as derived from the Kosambi mapping function by Phillips et al. [7] (*white bars*) and (4) the recombination fraction as determined from CEPH families (*dark grey bars*) for which the values >0.5 are presented as 0.5. Syntenic STR pairs on the same chromosome arm are shown left and those on different chromosome arms are shown right from the dividing space. STR pairs are ordered by physical distance. The physical and genetic distances are shown on the left *Y*-axis, while the Kosambi-derived Rc and the CEPH-based recombination fraction are shown on the right *Y*-axis. **b** The Rc derived from the Kosambi mapping function (for syntenic STR pairs for which the Kosambi-derived Rc was presented by Phillips et al. [7]) is plotted against the CEPH-based recombination fraction. The *diagonal line* represents the theoretical situation in which the Kosambi-derived Rc is equal to the estimated recombination fraction. Exact values for the plotted data can be found in Supplementary Table 3

thresholds, were evaluated based on the genotyping results for 335 Dutch reference DNA samples. The inter-locus balance is calculated by dividing the average peak height on a locus by the average peak height of the complete profile. As apparent from Supplementary Fig. 2a, the shorter loci have an inter-locus balance that is above one (one is a perfect balance), while the longer loci have a balance below one. Overall, a general decreasing trend is visible with increasing amplicon length, although D8S1132 shows slightly higher values than the other short loci and SE33 shows slightly lower values than the other markers.

The intra-locus peak height ratio (PHR, also known as heterozygous peak balance) should, for standard PCR conditions, be between 0.6/0.7 and 1 (perfect balance) [23, 24] to enable correct interpretation of the DNA profile. Supplementary Fig. 2b illustrates that the shorter loci all have median values >0.80 and, thus, better PHRs (i.e. closer to one) than the longer loci that show median values down to 0.74. Nevertheless, all loci show PHRs that meet the requirements for correct interpretation of the DNA profile.

The HDplex consists of 11 tetra- and one penta-nucleotide (D10S2325) repeat markers and does not contain trinucleotide STRs (such as D22S1045 in NGM) that are prone to increased stuttering at both −1 and +1 position (one repeat unit shorter or longer than the parent allele, respectively) [25, 26]. Accordingly, the (pristine) DNA profiles did not invoke inference of locus-specific +1 stutter ratio filters. However, several −1 repeat stutters were called, indicating that the locus-specific −1 stutter ratio filters provided by the manufacturer do not suffice. Therefore, we determined the −1 stutter ratio thresholds empirically using the same method as in Westen et al. [26]. These stutter ratio thresholds comprise 99 % of the −1 stutters and are based on 139 to 639 observations per locus (Table 2). Table 2 shows the empirically determined −1 stutter ratio thresholds, which are compared to the ones provided by Qiagen. For all 12 loci, the empirically determined −1 stutter ratio thresholds were higher than the thresholds provided by Qiagen, and we elevated the stutter ratio filters with 2.25 % to 6.11 % in the profile analysis software. The large differences between our thresholds and those suggested by Qiagen may have several reasons. It could result from our relatively small number of observations (when compared to our NGM validation [26]), the method by which stutter thresholds are calculated (empirically by us, not known for Qiagen), whether stutters at −1 and +1 position (heterozygous pair with two repeat lengths size difference) are included (included by us, not known for Qiagen) or how alleles not showing detectable stutters are regarded (excluded by us, not known for Qiagen). Nevertheless, we feel that the stutter ratios as provided by Qiagen are too low, as we observed several called stutter peaks when using their stutter ratio filters, already when analysing pristine DNA and optimal inputs (increased stutters are well-known for low template samples). Therefore, we recommend determining the −1 stutter ratio thresholds in-house when working with the HDplex.

Table 2 Locus-specific −1 repeat stutter ratio filters for HDplex

Locus	D2S1360	D3S1744	D4S2366	D5S2500	SE33	D6S474	D7S1517	D8S1132	D10S2325	D12S391	D18S51	D21S2055
Number[a]	448	594	139	505	504	356	587	639	519	611	546	394
emp[b,c] (%)	14.61	15.96	9.76	8.91	16.11	10.28	13.38	17.28	10.67	18.13	13.25	20.58
Qia[d] (%)	9.00	11.00	6.00	6.00	10.00	8.00	10.00	13.00	5.00	14.00	11.00	15.00
emp-Qia (%)	+5.61	+4.96	+3.76	+2.91	+6.11	+2.28	+3.38	+4.28	+5.67	+4.13	+2.25	+5.58

[a] Number of −1 stutters >25 rfu; this number differs between loci from variation in the numbers of stutters above detection threshold, homozygous alleles (which give only one stutter at a locus) and overlapping alleles at heterozygous loci (e.g., for an 11/12 genotype, the −1 stutter of allele 12 overlaps with allele 11)

[b] Empirically determined −1 stutter ratio threshold comprising 99 % of the stutters

[c] Presented with two decimals to fit GeneMapper ID-X entry

[d] Stutter ratio filter as provided by Qiagen

Statistical analysis

Hardy–Weinberg equilibrium, linkage disequilibrium and linkage

The genotyping results of the 335 Dutch reference samples were used to determine the allele frequencies and summary statistics for the HDplex (Supplementary Table 1). One important aspect of the summary statistics is the Hardy–Weinberg equilibrium (which refers to the independent association of alleles within one locus [27]). The data should not deviate significantly from HWE to enable assessment of gametic disequilibrium or linkage between syntenic STR pairs. Gametic disequilibrium is also known as linkage disequilibrium (LD) and refers to the non-random association of alleles at different loci into gametes [28]. Since we aim to assess these aspects for all syntenic STR pairs residing in HDplex, NGM and Identifiler, the p value for HWE testing was also determined for the syntenic markers in NGM and Identifiler (based on the complete DNA reference set of 2,085 samples; results not shown). For the syntenic markers in our population data, no significant deviation from HWE was detected after Bonferroni correction (Supplementary note 2).

An overview of all 30 markers that are present in HDplex, NGM and Identifiler and their chromosome location is presented in Table 3. The 17 syntenic STR pairs that can be formed out of these 30 markers are shown in Table 4. For these pairs, we tested for departure from linkage equilibrium using the Arlequin software. The results are presented in Table 4, both for the subset of 335 Dutch reference samples (meaning 335 HDplex, 335 NGM and 335 Identifiler DNA profiles) and for the full set extending to 2,085 samples (which means 2,085 NGM, 2,085 Identifiler and 335 HDplex DNA profiles). No significant departure from linkage equilibrium was detected after Bonferroni correction (Supplementary note 2). Using the GENETIX software, we found genotypic correlation coefficients between 0.014 and 0.051. As a comparison, r^2 values as high as 0.35 [29] or 0.45 [30, 31] have been found for Dutch or European populations, albeit for much smaller physical distances. Thus, the correlation coefficients that we found seem to indicate low genotypic LD for all 17 STR pairs tested in our population samples.

To determine the recombination fraction between the 17 above-mentioned syntenic STR pairs, 5 three-generation CEPH pedigrees were profiled for HDplex, NGM and Identifiler. The genotypes of the 78 individuals are provided in Supplementary Table 2. Figure 1a visualises four aspects of the 17 syntenic STR pairs (the corresponding numeric values are shown in Supplementary Table 3): (1) the physical distance between the markers as derived from the NCBI UniSTS database [32], (2) the genetic distance between the

Table 3 Overview of the STR markers present in HDplex (HD), NGM and Identifiler (Id)

chr	STR	kit[a]	Location	Location (Mb)
01	D1S1656	NGM	1q42.2	230.91
02	TPOX	Id	2p25.3	1.49
02	D2S1360	HD	2p24-p22	17.49
02	D2S441	NGM	2p14	68.24
02	D2S1338	NGM, Id	2q35	218.88
03	D3S1358	NGM, Id	3p21.3	45.58
03	D3S1744	HD	3q24	147.09
04	D4S2366	HD	4p16-p15.2	6.48
04	FGA	NGM, Id	4q28	155.51
05	D5S2500	HD	5q11.2	58.70
05	D5S818	Id	5q23.2	123.11
05	CSF1PO	Id	5q33.1	149.46
06	SE33	HD	6q14	88.99
06	D6S474	HD	6q21-22	112.88
07	D7S820	Id	7q21.1	83.79
07	D7S1517	HD	7q31.3	123.50
08	D8S1132	HD	8q23.1	107.33
08	D8S1179	NGM, Id	8q24.1	125.91
10	D10S2325	HD	10p12	12.71
10	D10S1248	NGM	10q26.3	131.09
11	TH01	NGM, Id	11p15.5	2.19
12	vWA	NGM, Id	12p13.3	6.09
12	D12S391	NGM, HD	12p13.2	12.45
13	D13S317	Id	13q31.1	82.72
16	D16S539	NGM, Id	16q24.1	86.39
18	D18S51	NGM, Id, HD	18q21.3	60.95
19	D19S433	NGM, Id	19q12	30.42
21	D21S11	NGM, Id	21q11.2-q21	20.55
21	D21S2055	HD	21q22	41.19
22	D22S1045	NGM	22q12.3	37.54

HD HDplex, *Id* Identifiler

[a] 2,085 genotypes were assessed for all loci present in NGM, Identifiler and locus SE33; 335 genotypes were determined for the nine (remaining) loci in HDplex

nearest HapMap SNP proxies as determined by Phillips et al. [7] together with (3) the Kosambi-derived recombination fractions (Rc) that were converted from the genetic distance using the Kosambi mapping function [7] and (4) the recombination fractions based on the CEPH pedigree data that were determined using the LINKAGE program (for which estimates greater than 0.5 are interpreted as being equal to 0.5 (Prof. J. Ott, personal communication)). As apparent from Fig. 1a and Supplementary Table 3, the physical distance can present both an overestimation and an underestimation of the genetic distance. The largest underestimations are made for the STR pairs vWA-D12S391 and TPOX-D2S1360, which might be due to the fact that both these pairs are situated near the telomeres, where crossover rates are generally higher [7, 33]. The recombination fractions that we found based on the CEPH pedigree data are comparable to the Kosambi-derived Rc values (that were generated based on the cM interval of the HAPMAP SNP proxies [7]), as shown in Fig. 1a and Table 4. Another representation of these data is given in Fig. 1b, from which it is clear that the data points reside around the diagonal line, indicating similar results for both methods.

The pair with the smallest physical distance and the second smallest recombination fraction is vWA and D12S391 (Supplementary Table 3). Since these STRs are both present in the current generation STR kits, several studies have assessed their possible allelic association [1, 4, 6–8]. As confirmed by our results, none of these studies have found indications for linkage disequilibrium between these markers at the population level, and it is inferred that it is legitimate to use the product rule for DNA profile probability calculations involving unrelated individuals. Regarding the assessment of linkage, a different approach is needed. In our and one of the other studies [6], three-generation CEPH families are used to determine the recombination fraction between vWA and D12S391. We find a recombination fraction of 0.17; Budowle et al. [6] estimate a value of 0.11. The Kosambi-derived Rc for vWA-D12S391 is 0.12 [1, 7], which is in the same range. Taken together, all values indicate the presence of (loose) physical linkage, which may influence the interpretation of genotyping data from (closely) related individuals.

According to Buckleton and Triggs [34], recombination fractions of 0.197 and 0.316 (derived using the Haldane mapping function for the STR pairs CSF1PO–D5S818 (25 cM) and Penta D–D21S11 (50 cM), respectively) are sufficiently small to affect match probability calculations for relatives and some pedigree analyses. An influence of physical linkage was also found by Nothnagel et al. [35], who simulated pairwise kinship analyses with or without taking linkage between STR markers into account. The overall results for both strategies were very similar, although the assessment of certain kinships (such as full siblings versus half siblings) could be affected by ignoring linkage [35]. The three kits that are assessed in our study each contain one syntenic STR pair residing on one chromosomal arm. When using genotyping data from one kit, the influence of one loosely linked pair among the set of 12 or 15 markers might not be substantial. However, when combining kits, the number of paired loci on the same chromosomal arm having recombination fractions <0.50 increases substantially as NGM with Identifiler presents two, NGM with HDplex five, Identifiler with HDplex seven and all three kits together eight of these pairs. Such numbers of loosely linked syntenic pairs may affect kinship analyses. Gill et al. [1] elaborated on this kind of kinship analyses using the vWA-D12S391

Table 4 Overview of pairwise linkage disequilibrium test results for syntenic STR loci in HDplex, NGM, and Identifiler

chr	STR pair	position[a]	Arlequin LD p-value 335NL	Arlequin LD p-value 2085NL	GENETIX corr.coef. 335NL	GENETIX p-value 335NL	GENETIX corr.coef. 2085NL	GENETIX p-value 2085NL
02	TPOX-D2S1360		0.038	0.076	0.045	0.040	idem[b]	idem
02	TPOX-D2S441		0.221	0.314	0.046	0.340	0.016	0.482
02	TPOX-D2S1338	*	0.953	0.449	0.035	0.994	0.016	0.894
02	D2S1360-D2S441		0.796	0.915	0.041	0.456	idem	idem
02	D2S1360-D2S1338	*	0.702	0.498	0.040	0.114	idem	idem
02	D2S441-D2S1338	*	0.418	0.507	0.042	0.462	0.015	0.968
03	D3S1358-D3S1744	*	0.402	0.580	0.042	0.562	idem	idem
04	D4S2366-FGA	*	0.191	0.135	0.050	0.000	idem	idem
05	D5S2500-D5S818		0.771	0.913	0.040	0.878	idem	idem
05	D5S2500-CSF1PO		0.037	0.138	0.051	0.023	idem	idem
05	D5S818-CSF1PO		0.334	0.822	0.043	0.658	0.014	0.956
06	SE33-D6S474		0.188	0.265	0.040	0.661	idem	idem
07	D7S820-D7S1517		0.946	0.957	0.041	0.921	idem	idem
08	D8S1132-D8S1179		0.493	0.211	0.043	0.020	idem	idem
10	D10S2325-D10S1248	*	0.845	0.889	0.039	0.860	idem	idem
12	vWA-D12S391		0.821	0.591	0.038	0.659	0.015	0.963
21	D21S11-D21S2055		0.731	0.332	0.034	0.999	idem	idem

For the Arlequin and GENETIX results, two datasets were used: "335NL" referring to a set of 335 samples profiled with NGM, Identifiler and HDplex, and "2085NL" which refers to 2,085 samples profiled with NGM and Identifiler, complemented with the genotyping data for HDplex for a subset of 335 of the samples

[a] Syntenic STR pairs on different chromosomal arms are marked by an asterisk in this column and shown in grey

[b] Equal to the value for 335NL

pair as an example (and they say their methods can be extended to evaluate linkage effects between any pair of loci with known recombination rate). Under the assumption of linkage equilibrium at the population level, they state that linkage has no effect and should not be considered in a pedigree unless at least one individual is involved in at least two transmissions of genetic material, as a parent and/or a child, and that individual is a double heterozygote at the loci involved. When the pedigree is informative of phase and the recombination rate between the markers is known, linkage can be accounted for statistically with the equations given in their paper [1]. Otherwise (under the assumption of linkage

Table 5 Match probability (when assuming independence) and number of available loci for NGM, Identifiler, HDplex and combinations thereof, based on genotyping data of 335 Dutch reference samples

	Complete profile		<200 bp[a]		<150 bp[b]	
	MP	# STRs	MP	# STRs	MP	# STRs
NGM	1.8E-19	15	4.9E-07	6	1.3E-03	3
Identifiler	5.2E-18	15	4.7E-06	5	6.5E-03	2
HDplex	6.1E-18	12	2.6E-06	4	3.3E-02	1
NGM+Id	3.3E-24	20	6.2E-09	8	8.4E-06	5
NGM+HD	1.6E-33	25	1.3E-12	10	4.3E-05	4
Id+HD	9.9E-34	27	1.2E-11	9	2.1E-04	3
NGM+Id+HD	2.9E-38	30	1.6E-14	12	2.8E-07	6

MP match probability, *Id* Identifiler, *HD* HDplex

[a] Loci <200 bp: the loci described for "b" (loci <150 bp) and the following loci: NGM: D3S1358, D8S1179, D19S433; Id: D5S818, D8S1179, TH01; HD: D3S1744, D8S1132, D10S2325

[b] Loci <150 bp: NGM: D2S441, D10S1248, D22S1045; Id: D3S1358, D19S433; HD: D7S1517

equilibrium at the population level), both loci can be used for kinship analysis employing the product rule. An overview and more details are given in Supplementary textbox 1.

Match probabilities for combined kits

In the previous section, it is shown that for unrelated individuals and for kinship analyses in which linkage has no effect, all 30 loci residing in NGM, Identifiler and HDplex can be employed in profile probability calculations using the product rule. Therefore, we evaluate the combined power of the three STR kits. Table 5 shows the match probabilities for NGM, Identifiler, HDplex and combinations thereof. NGM has the best match probability of these three kits; Identifiler and HDplex have a comparable match probability even though Identifiler contains three STRs more. Thus, the match probability per locus is more favourable for HDplex than for Identifiler. Combining the results of two kits gives the most informative match probability for Identifiler together with HDplex (9.9×10^{-34}), which is based on genotyping data from 27 unique STRs. The combination of the 25 different STRs of NGM and HDplex gives a very informative match probability (1.6×10^{-33}) as well. Table 5 also shows the match probabilities and number of available loci for amplicon sizes <200 bp and <150 bp that apply to DNA of different degradation levels. The most informative combination for fragments <200 bp is NGM with HDplex (1.3×10^{-12}), and for fragments <150 bp NGM with Identifiler (8.4×10^{-6}). Combining all three kits further improves the match probability up to 2.8×10^{-7} for degraded DNA (<150 bp) and up to 2.9×10^{-38} for non-degraded DNA.

Conclusion

The Investigator HDplex is suitable for forensic DNA analysis (when used with a re-amplified ladder to prevent binning problems for D10S2325). No linkage disequilibrium was detected between the syntenic STRs of HDplex and those of NGM or Identifiler, and we infer that the product rule can be applied for profile probability calculations in unrelated individuals. In kinship analyses, the product rule cannot always be applied (depending on the pedigree), and readers are referred to the paper by Gill et al. [1] for a description of these cases and for methods to implement the recombination rate between markers (like determined in this study) into the calculations. HDplex has many non-overlapping markers with NGM and Identifiler, and the power of discrimination per marker is, on average, higher than for the other kits. In conclusion, the HDplex is a good complementary STR kit that can be used for complex kinship analyses and may aid the analysis of degraded DNA.

Acknowledgements This study was supported by a grant from the Netherlands Genomics Initiative/Netherlands Organization for Scientific Research (NWO) within the framework of the Forensic Genomics Consortium Netherlands. The authors are grateful for the voluntary cooperation of the DNA donors for the Dutch reference set. We would like to thank Klaas Slooten and Ate Kloosterman for valuable discussions and Arnoud Kal for critically reading the manuscript.

References

1. Gill P, Phillips C, McGovern C, Bright JA, Buckleton J (2011) 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
2. Gill P, Fereday L, Morling N, Schneider PM (2006) New multiplexes for Europe-amendments and clarification of strategic development. Forensic Sci Int 163:155–157
3. Gill P, Fereday L, Morling N, Schneider PM (2006) The evolution of DNA databases - recommendations for new European STR loci. Forensic Sci Int 156:242–244
4. Phillips C, Fernandez-Formoso L, Garcia-Magarinos M, Porras L, Tvedebrink T, Amigo J, Fondevila M, Gomez-Tato A, varez-Dios J, Freire-Aradas A, Gomez-Carballa A, Mosquera-Miguel A, Carracedo A, Lareu MV (2011) Analysis of global variability in 15 established and 5 new European Standard Set (ESS) STRs using the CEPH human genome diversity panel. Forensic Sci Int Genet 5:155–169
5. Lobo I and Shaw K (2008) Thomas Hunt Morgan, Genetic Recombination, and Gene Mapping. Nature Education 1. http://www.nature.com/scitable/topicpage/thomas-hunt-morgan-genetic-recombination-and-gene-496
6. Budowle B, Ge J, Chakraborty R, Eisenberg AJ, Green R, Mulero J, Lagace R, Hennessy L (2011) Population genetic analyses of the NGM STR loci. Int J Legal Med 125:101–109
7. Phillips C, Ballard D, Gill P, Court DS, Carracedo A, Lareu MV (2012) 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
8. O'Connor KL, Tillmar AO (2012) Effect of linkage between vWA and D12S391 in kinship analysis. Forensic Sci Int Genet. doi:10.1016/j.fsigen.2012.03.008
9. Westen AA, Matai AS, Laros JF, Meiland HC, Jasper M, de Leeuw WJ, de Knijff P, Sijen T (2009) Tri-allelic SNP markers enable analysis of mixed and degraded DNA samples. Forensic Sci Int Genet 3:233–241
10. Butler JM, Shen Y, McCord BR (2003) The development of reduced size STR amplicons as tools for analysis of degraded DNA. J Forensic Sci 48:1054–1064
11. Park SDE (2001) Trypanotolerance in West African Cattle and the population genetic effects of selection. Dissertation, University of Dublin
12. Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10:564–567
13. (2000) Genetic Identity PowerStats v12. Promega. http://www.promega.com/geneticidtools/powerstats/
14. Black WC, Krafsur ES (1985) A FORTRAN program for the calculation and analysis of two-locus linkage disequilibrium coefficients. Theor Appl Genet 70:491–496

15. Belkhir K, Borsa P, Chikhi L, Raufaste N, and Bonhomme F (2011) Genetix. Université Montpellier II, France. http://kimura.univ-montp2.fr/genetix/. Accessed 1 Nov 2011
16. Lathrop GM, Lalouel JM, Julier C, Ott J (1984) Strategies for multilocus linkage analysis in humans. Proc Natl Acad Sci U S A 81:3443–3446
17. Li W, Haghighi VFG, Matise TC, Ott J (1996) Using web browser as the graphical user interface for linkage analysis. Am J Hum Genet supplement to vol 59:A307
18. Westen AA, Nagel JH, Benschop CC, Weiler NE, de Jong BJ, Sijen T (2009) Higher capillary electrophoresis injection settings as an efficient approach to increase the sensitivity of STR typing. J Forensic Sci 54:591–598
19. Benschop CC, van der Beek CP, Meiland HC, van Gorp AG, Westen AA, Sijen T (2011) Low template STR typing: effect of replicate number and consensus method on genotyping reliability and DNA database search results. Forensic Sci Int Genet 5:316–328
20. Dixon LA, Dobbins AE, Pulker HK, Butler JM, Vallone PM, Coble MD, Parson W, Berger B, Grubwieser P, Mogensen HS, Morling N, Nielsen K, Sanchez JJ, Petkovski E, Carracedo A, Sanchez-Diz P, Ramos-Luis E, Brion M, Irwin JA, Just RS, Loreille O, Parsons TJ, Syndercombe-Court SH, Stradmann-Bellinghausen B, Bender K, Gill P (2006) Analysis of artificially degraded DNA using STRs and SNPs - results of a collaborative European (EDNAP) exercise. Forensic Sci Int 164:33–44
21. Westen AA, Sijen T (2009) Degraded DNA sample analysis using DNA repair enzymes, mini-STRs and (tri-allelic) SNPs. Forensic Sci Int Genet Suppl Ser 2:505–509
22. Fondevila M, Phillips C, Naveran N, Cerezo M, Rodriguez A, Calvo R, Fernandez LM, Carracedo A, Lareu MV (2008) Challenging DNA: assessment of a range of genotyping approaches for highly degraded forensic samples. Forensic Sci Int Genet Suppl Ser 1:26–28
23. Gill P, Sparkes R, Kimpton C (1997) Development of guidelines to designate alleles using an STR multiplex system. Forensic Sci Int 89:185–197
24. Butler JM (2005) Forensic issues: degraded DNA, PCR inhibition, contamination, mixed samples and low copy number. Forensic DNA typing, 2nd edn. Elsevier, Amsterdam, pp 145–180
25. Gibb AJ, Huell AL, Simmons MC, Brown RM (2009) Characterisation of forward stutter in the AmpFlSTR SGM Plus PCR. Sci Justice 49:24–31
26. Westen AA, Grol LJW, Harteveld J, Matai AS, de Knijff P, Sijen T (2012) Assessment of the stochastic threshold, back- and forward stutter filters and low template techniques for NGM. Forensic Sci Int Genet. doi:10.1016/j.fsigen.2012.05.001
27. Butler JM (2005) STR population database analyses. Forensic DNA typing, 2nd edn. Elsevier, Amsterdam, pp 473–496
28. Hedrick PW (2011) Linkage disequilibrium and recombination. Genetics of populations, 4th edn. Jones and Bartlett publishers, Sadbury, MA, pp 523–597
29. McEvoy BP, Powell JE, Goddard ME, Visscher PM (2011) Human population dispersal "Out of Africa" estimated from linkage disequilibrium and allele frequencies of SNPs. Genome Res 21:821–829
30. Jakobsson M, Scholz SW, Scheet P, Gibbs JR, VanLiere JM, Fung HC, Szpiech ZA, Degnan JH, Wang K, Guerreiro R, Bras JM, Schymick JC, Hernandez DG, Traynor BJ, Simon-Sanchez J, Matarin M, Britton A, van de Leemput J, Rafferty I, Bucan M, Cann HM, Hardy JA, Rosenberg NA, Singleton AB (2008) Genotype, haplotype and copy-number variation in worldwide human populations. Nature 451:998–1003
31. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Shen Y, Sun W, Wang H, Wang Y, Wang Y, Xiong X, Xu L, Waye MM, Tsui SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallee C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PK, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PI, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Sham PC, Varilly P, Altshuler D, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Qin ZS, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Tsunoda T, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Zeng C, Zhao H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CD, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Watkin J, Gibbs RA, Belmont JW, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Wheeler DA, Yakub I, Gabriel SB, Onofrio RC, Richter DJ, Ziaugra L, Birren BW, Daly MJ, Altshuler D, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archeveque P, Bellemare G, Saeki K, Wang H, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL, Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R, Stewart J (2007) A second generation human haplotype map of over 3.1 million SNPs. Nature 449:851–861
32. UniSTS database (2011) NCBI. http://www.ncbi.nlm.nih.gov/sites/entrez?db=unists. Accessed 24 Feb 2011
33. Buard J, de Massy B (2007) Playing hide and seek with mammalian meiotic crossover hotspots. Trends Genet 23:301–309
34. Buckleton J, Triggs C (2006) The effect of linkage on the calculation of DNA match probabilities for siblings and half siblings. Forensic Sci Int 160:193–199
35. Nothnagel M, Schmidtke J, Krawczak M (2010) Potentials and limits of pairwise kinship analysis using autosomal short tandem repeat loci. Int J Legal Med 124:205–215