Chapter 3

Tri-allelic SNP markers enable analysis of mixed and degraded DNA samples

Forensic Science International: Genetics (2009) 3:233–241

Antoinette A. Westen
Anuska S. Matai
Jeroen F.J. Laros
Hugo C. Meiland
Mandy Jasper
Wiljo J.F. de Leeuw
Peter de Knijff
Titia Sijen

Abstract

For the analysis of degraded DNA in disaster victim identification (DVI) and criminal investigations, single nucleotide polymorphisms (SNPs) have been recognised as promising markers mainly because they can be analysed in short sized amplicons. Most SNPs are bi-allelic and are thereby ineffective to detect mixtures, which may lead to incorrect genotyping. We developed an algorithm to find non-binary (i.e. tri-allelic or tetra-allelic) SNPs in the NCBI dbSNP database. We selected 31 potential tri-allelic SNPs with a minor allele frequency of at least 10 %. The tri-allelic nature was confirmed for 15 SNPs residing on 14 different chromosomes. Multiplex SNaPshot™ assays were developed, and the allele frequencies of 16 SNPs were determined among 153 Dutch and 111 Netherlands Antilles reference samples. Using these multiplex SNP assays, the presence of a mixture of two DNA samples in a ratio up to 1:8 could be recognised reliably. Furthermore, we compared the genotyping efficiency of the tri-allelic SNP markers and short tandem repeat (STR) markers by analysing artificially degraded DNA and DNA from 30 approximately 500-year-old bone and molar samples. In both types of degraded DNA samples, the larger sized STR amplicons failed to amplify whereas the tri-allelic SNP markers still provided valuable information. In conclusion, tri-allelic SNP markers are suited for the analysis of degraded DNA and enable the detection of a second DNA source in a sample.

Introduction

DNA used in disaster victim identification (DVI) and forensic human genotyping is often degraded. In DNA profiling, this results in the loss of the higher molecular weight short tandem repeat (STR) markers and, consequently, in lower discrimination power of the obtained partial DNA profiles [1–4]. STR amplicons vary in length between 100 and 450 base pairs (bp). Two different strategies have been proposed to decrease the target region [5–7]: (1) the use of so-called mini-STRs for which the primer binding sites are moved closer to the repeat region resulting in amplicons usually <150 bp [8–10] and (2) single nucleotide polymorphism (SNP) markers that involve the analysis of only one nucleotide resulting in amplicons that can be designed to be as small as 50 bp [11,12]. These very small amplicons make SNPs particularly promising markers for forensic analysis of degraded DNA [13,14].

SNPs have several other advantageous characteristics. One of these advantages is a low mutation rate (10^{-8} versus 10^{-3} for STRs), which makes them useful for paternity testing and complex kinship analysis [15,16]. In addition, SNPs can be analysed using high throughput systems, and are not accompanied by the occurrence of stutter peaks, which simplifies the interpretation of the SNP based profiles [17–19]. The vast majority of SNPs are bi-allelic and these binary SNPs are unable to reliably detect the presence of a second DNA source in a sample [5,17]. However, Phillips et al. have described that non-binary SNPs can detect the presence of a DNA mixture [20]. This is important to recognise, for example, contamination by soft tissue or bodily fluids from other victims during a mass disaster. With computer simulations is estimated that 45–65 bi-allelic SNPs are needed to reach a discrimination power that is equal to 12–16 STRs [21–23]. In theory, less tri-allelic SNPs would be needed, since they have an increased discrimination power per SNP.

In this study we apply non-binary SNPs to forensic relevant samples. We developed an algorithm to search for non-binary SNPs in the NCBI SNP database (dbSNP). For a selection of the tri-allelic SNP candidates that were found, SNaPshot™ multiplex assays were set up, and over 250 reference samples from the Netherlands and the Netherlands Antilles were analysed. A web-based application was written to calculate allele frequencies from the SNP genotyping data. Furthermore, two-donor mixtures in various ratios were studied. Artificially degraded DNA and DNA from approximately 500-year-old bone samples were genotyped both by the tri-allelic SNP assays and standard STR profiling in order to compare the genotyping efficiency of both methods.

Materials and methods

Samples

The reference set for verification of the non-binary nature of the SNPs consisted of 153 Dutch and 111 Netherlands Antilles samples obtained from employees of the Netherlands Forensic Institute, anonymous Dutch blood donors and policemen from the Antilles. The Netherlands Antilles population has an admixed origin of Native Americans, Europeans and Africans with an undetermined mixture ratio. Y chromosomal research indicates that approximately half of the males from the reference population displays Y chromosomes of African origin (PdK, unpublished results). From the YCC panel that consists of cell lines from males representing worldwide populations, 59 samples were analysed: 5 European, 12 Russian/Siberian, 8 Asian/Pakistan, 9 African, 14 South African and 11 Native American [24]. The sensitivity of the SNPs was determined using pristine DNA (Quantifiler™ Human DNA standard denoted as hDNA) with a wide range of PCR inputs for both the SNP and STR analyses of 5 pg, 10 pg, 20 pg, 30 pg, 40 pg, 50 pg, 60 pg, 70 pg, 80 pg, 90 pg, 100 pg, 200 pg, 300 pg, 400 pg, 500 pg, 750 pg, 1 ng, 10 ng and 50 ng. For mixture analysis, DNA from several pairs of reference donors was mixed in various ratios: 1:8, 1:4, 1:2, 1:1, 2:1, 4:1 and 8:1.

In order to obtain artificially degraded DNA, pristine hDNA of 200 ng/μL was irradiated with 254 nm UV light in a CL-1000 UV CrossLinker (UVP, Inc.) at 0.9 J/cm2 for 0, 5, 10, 15, 30, 60, 90 and 120 min. Two series of hDNA were used: hDNA irradiated at room temperature or hDNA denatured for 5 min at 95 °C and placed and irradiated on ice. Furthermore, pristine hDNA samples were degraded by different concentrations of TURBO™ DNase (Ambion™ TURBO DNA-free™ Kit). DNA fragments of specific size ranges were isolated from agarose gel with the QIAquick Gel Extraction Kit (Qiagen), and diluted to 1 ng/μL after DNA concentration measurement with a NanoDrop™ 1000 spectrophotometer (Thermo Scientific). Genotyping efficiency was determined using the artificially degraded DNA samples and DNA of thirty 450–550-year-old bone and molar samples excavated in Delft (the Netherlands).

SNP selection

To find non-binary SNP candidates, the NCBI database dbSNP (build 126) was searched with a custom-made algorithm, which can be found on http://www.liacs.nl/rvjlaros/projects/snp/. This algorithm specifically searches for non-binary SNPs with variation allele: V (A, C or G), H (A, C or T), D (A, G or T), B (C, G or T) or N (G, A, T or C) and SNP class: snp. It filters out any unconfirmed allele calls from opposite strands (N) and non-existent data (-). SNPs with a minor allele frequency (i.e. the lowest frequency of the three alleles) above 10 % in at least one population in dbSNP were selected for further analysis. SNPs can erroneously be assigned non-binary due

to a lack of clarity regarding the direction of the sequence reads entered into dbSNP. Therefore, the non-binary character and allele frequencies of the SNP candidates were redetermined after manual entering in dbSNP. The test set of SNPs was selected on the following criteria: (1) a high minor allele frequency, (2) a high number of populations that showed three alleles for that SNP, (3) an equal distribution of the other two alleles, and (4) the opportunity to develop suitable primers. To diminish the chance of linkage between the SNPs, one SNP per chromosome was selected from the test set for the development of the SNaPshot™ multiplexes.

PCR

The web-based version of Primer3 was used to design PCR primers (supplementary table S1) resulting in amplicon sizes between 40 and 100 bp, with a primer length between 18 and 24 bases, a primer Tm between 55 and 61 °C and a primer GC percentage between 30 and 70 % [25]. All primers were checked for the absence of primer–dimer formation, hairpin structures and complementarity to other primers in the multiplex with the program Autodimer [26]. The primers were all HPLC purified after synthesis (Biolegio BV or Isogen Life Science).

A 12.5 µL PCR was set up using 1× PCR Gold buffer, 9 mM MgCl2, 2 mM dNTPs, 0.5 µL Taq Gold, 100 nM of each primer and 1 ng DNA. The PCR program consisted of an initial hot start of 10 min at 94 °C, followed by 35 cycles of 94 °C for 30 s, 60 °C for 30 s and 72 °C for 30 s and a final hold at 72 °C for 5 min. To remove unused primers and nucleotides 2.5 µL ExoSAP-IT™ (USB Corporation) was added to the PCR products. This enzyme mixture was incubated for 30 min at 37 °C and inactivated by incubation at 80 °C for 15 min.

All measurements were performed in dedicated laboratories (ISO 17025 accredited), while wearing protective clothing. For comparison of the SNP markers with STR markers, the AmpF*l*STR® SGM Plus™ kit from Applied Biosystems was used according to their protocol, but with half the volumes (25 µL PCR).

SNaPshot™ (single base extension)

Extension primers were designed immediately upstream of the SNP position [11,27]. Primer3 was used to design the primers with a primer size between 15 and 23 bases, a Tm between 48 and 52 °C, and a GC percentage between 25 and 60 % [25]. Since the SNPs are analysed in multiplex they need to be spatially separated during capillary electrophoresis; therefore the extension primers were tailed at the 5' end with a non-human DNA sequence resulting in primer sizes between 23 and 50 nucleotides.

Terminator ddNTPs, labeled with four different fluorescent dyes, were used to extend the primers at the SNP position. The extension reaction was set up with

2.5 µL SNaPshot™ Ready Reaction Mix (Applied Biosystems), extension primer concentrations between 5 and 75 nM (supplementary table S1), 1.0 µL ExoSAP-IT™ treated PCR product and added up to a total volume of 5 µL. The extension program has an initial denaturation step of 2 min at 96.0 °C, followed by 40 cycles of 96.0 °C for 10 s, 50.0 °C for 5 s and 60.0 °C for 30 s. To remove unincorporated nucleotides 1.5 µL Shrimp Alkaline Phosphatase (USB Corporation) was added to the extension products, incubated for 60 min at 37 °C and inactivated for 15 min at 72 °C.

SNP detection and analysis

The fluorescently labeled SNaPshot™ extension products were detected by capillary electrophoresis with an ABI Prism 3130xl Genetic Analyzer with a 36 cm capillary array and POP-4 polymer (Applied Biosystems). Data Collection software v3.0 with the default run module SNP36_POP4_1 and dye set E5 were used to analyse 1.0 µL SAP-treated extension product mixed with 8.75 µL Hi-Di™ formamide and 0.25 µL GeneScan-120LIZ™ size standard (Applied Biosystems) after 5 min of denaturation at 95 °C and 5 min of cooling on ice.

Alleles were automatically called with GeneMapper® ID v3.2.1. Since the ratio of the fluorescent signals for G, A, T and C differ per nucleotide dye and between SNPs, the allele balance cut-off value in the SNaPshot™ default analysis method was adjusted from 0.30 to 0.125 in order to call both G and C in a heterozygous locus. For the analysis of the dilution series and the artificially degraded DNA, reference DNA samples with a known SNP and STR profile were used. For these samples, the percentage of detected alleles could be calculated, in which homozygous alleles were counted as two alleles. Since the SNP and STR profiles of the analysed bone and molar samples were unknown, homozygous alleles could not be discriminated from a single heterozygous allele without the second allele. For these samples only the percentage of detected loci is determined.

Allele frequencies and statistics

To calculate the allele frequency distribution of the Dutch and Netherlands Antilles samples, genotype tables are exported from the GeneMapper® plot display to a .csv-file and copied into SNPstat, a custom-made program that can be found on http:// www.liacs.nl/rvhmeiland/projects/snpstat/. Expected and observed heterozygosity values and PIC values are calculated using the Excel Microsatellite Toolkit [28]. Genepop v4.0.7 is used to determine the p-value for Hardy–Weinberg (HW) equilibrium testing, for HW testing when H1 = heterozygote deficit, when H1 = heterozygote excess and for deviation from independence between or across loci [29]. The power of discrimination and the power of exclusion were calculated with the Excel spreadsheet Genetic Identity PowerStats v12 (Promega) [30].

Sequencing

Sanger sequencing was used to confirm the different alleles found with the SNaPshot™ method in the Dutch and Netherlands Antilles reference samples. Monoplex PCRs were performed under the same conditions and with the same primers as described above, and the PCR products were cleaned with 1.5 µL ExoSAP-IT™ (USB Corporation). The sequencing reaction was performed in a volume of 20 µL with 1x Sequencing Buffer, 1.0 µL BigDye™ Terminator v1.1 Ready Reaction Mix (Applied Biosystems), 0.16 mM SNP specific primer and 1.0 µL ExoSAP-IT™ treated PCR product. To remove unincorporated nucleotides and salts the 20 µL sequencing product was mixed with 20 µL XTerminator™ Solution and 90 µL SAM™ Solution in a MicroAmp™ Optical 96-Well Reaction Plate (Applied Biosystems). The plate was vortexed for 30 min at 2000 rpm, centrifuged for 2 min at 1000 × *g*, placed directly in an ABI Prism 3130xl Genetic Analyzer and analysed with the BDx_UltraSeq36_POP4_1 run module and dye set E-BigDyeV1. The sequences were analysed with Sequencing Analysis 5.2 (Applied Biosystems).

Results and discussion

SNP selection and multiplexing

Using our custom-made algorithm, dbSNP was searched for non-binary SNPs with a minor allele frequency above 10 % in at least one population. This search yielded 74 SNP candidates. After manual entry in dbSNP to correct for inconsistencies in direction of sequencing, 63 tri-allelic SNP candidates remained, which are listed per chromosome in Table 1 and supplementary table S2. To obtain a tri-allelic SNP for chromosome 3, the minor allele frequency had to be lowered to 6 % (Table 1). Since the first candidate on chromosome 22 (rs3859849) yielded no useful primers, a second tri-allelic SNP candidate was found by lowering the minor allele frequency to 8 % (Table 1). Thereby a total of 65 tri-allelic SNP candidates was obtained. For chromosome 13, 15, X and Y no suitable SNPs were found. In contrast, the number of tri-allelic SNPs found on chromosome 6 is considerable compared to the other chromosomes, probably due to extensive research to a particular part of this chromosome.

The 65 tri-allelic SNP candidates are distributed over 20 chromosomes. The chance of linkage between the SNPs is minimised when they reside on different chromosomes. In order to find the best SNP per chromosome, PCR and extension primers were tested in monoplex for the 31 SNPs that are shown in Table 1. To reduce the number of reactions and the amount of DNA needed, SNaPshot™ multiplex assays were designed. The primer sets that were examined for the SNPs on chromosome 16, 17, 19 and 21 proved unsuited for multiplexing, probably due to interactions with other PCR

Table 1
Examined potential non-binary SNPs found in dbSNP with a custom-made algorithm.

Chromosome	rs-number	dbSNP maximum minor allele frequency[a]	SNP name in multiplex	Allele frequencies for SNPs examined in multiplex							
				NL (*n* alleles = 306)				ANT (*n* alleles = 222)			
				A	C	T	G	A	C	T	G
1[b]	**3091244**	**26.1**	01a	31.37		6.21	62.42	31.08		25.68	43.24
1	1630312	20.3									
1	951416	19.3									
2	**727241**	**21.4**	02a		26.47	73.53			21.17	78.83	
2	1047883	11.5									
3	**35528968**	**6.2**	03a			100.00				100.00	
4	4540055	27.1									
4	**356167**	**17.5**	04b	17.65	11.11		71.24	4.05	24.77		71.17
5	**9329104**	**20.8**	05a	87.91			12.09	58.11	12.61		29.28
6	**9274701**	**28.6**	06a		+[d]	+	+		+	+	+
7	**2032582**	**15.2**	07a	0.98		43.79	55.23	1.35		10.36	88.29
8	**433342**	**18.8**	08a		70.59	27.78	1.63		36.04	35.59	28.38
8	4532634	14.4	[c]								
9	**1112534**	**12.5**	09c	22.22			77.78	26.13			73.87
9	3818367	10.1									
10	**17287498**	**24.0**	10e	23.86	51.63	24.51		43.69	46.40	9.91	
10	2803554	14.6									
10	9333212	11.9									
11	**5030240**	**14.8**	11a		26.80	7.84	65.36		29.28	32.88	37.84
11	11042874	12.5									
12	**2307223**	**26.8**	12a	81.05		16.67	2.29	28.38		9.91	61.71
12	7133606	15.3									
14	**1008686**	**10.0**	14a	43.79		56.21		63.51		36.49	
16	2278489	13.0									
16	11865501	10.0	[c]								
17	1050528	27.1	[c]								
18	**2853525**	**10.4**	18a		32.03	67.97		8.11	39.19	52.70	
19	385780	26.1	[c]								
20	**2069945**	**15.9**	20a		45.75	11.11	43.14		71.62	5.86	22.52
21	471010	10.0									
22	**34741930**	**8.3**	22b		100.00				100.00		

[a] The population with the highest minor allele frequency in dbSNP (when allele frequencies were determined for more than one population).
[b] SNPs shown in bold are tested in multiplex SNaPshot™ assays.
[c] Not suited for multiplex, but three alleles detected in monoplex SNaPshot™ pilot experiments.
[d] Allele frequencies could not be determined due to ambiguous SNaPshot™ results.

or extension primers. The SNPs that were chosen for the remaining 16 chromosomes are shown in bold in Table 1. These 16 SNPs were combined in multiplex A, B, and C with seven, four, and five SNP markers, respectively (supplementary table S1).

In order to test whether these 16 SNPs were non-binary, 153 Dutch and 111 Netherlands Antilles reference samples were analysed using the three multiplexes (supplementary table S3). Nine SNPs were found to be tri-allelic in both populations, and two SNPs were tri-allelic in the Antilles samples but appeared to be bi-allelic in the Dutch samples. Three SNPs were bi-allelic in both populations and two SNPs were fixed. The analyses were extended with 59 samples from the Y Chromosome Consortium dispersed over six genetically distinct populations, but no additional alleles were detected (supplementary table S3). Thereby, 11 of the 16 SNPs in the multiplexes were found to be truly tri-allelic. Monoplex assays on a limited number of samples from the reference set revealed four additional tri-allelic SNPs: a second one on chromosome 8 and three SNPs on chromosome 16, 17 and 19 for which the primers were unsuited for multiplexing (Table 1). Thus, in total 15 SNPs on 14 different chromosomes were confirmed to be tri-allelic.

For the 16 SNPs in the multiplex SNaPshot™ assays 41 different alleles were observed. To confirm the occurrence of these alleles, per SNP up to 8 samples were analysed by Sanger sequencing using the same primers as for the PCRs preceding the SNaPshot™ assays. The alleles that were observed by Sanger sequencing were consistent with the SNaPshot™ results. Eight of the 41 alleles could not be confirmed, which is probably due to ineffective sequencing within the very short sized amplicons. Sanger sequencing was very useful for the interpretation of ambivalent SNaPshot™ results obtained for SNP 06a. While in samples 1 and 2 (Fig. 1A and B, and C and D) a heterozygous GC and a homozygous T are detected by both methods, in sample 3 (Fig. 1E and F) Sanger sequencing clearly detects a homozygous G, while the SNaPshot™ shows a large G-peak and a small additional T-peak. This additional small T-peak is neither observed in the PCR and extension negative controls (data not shown) nor in Fig. 1B. Due to these ambiguous SNaPshot™ results, SNP 06a was left out of further analyses.

The SNaPshot™ was chosen as analysis platform since this method is not dependent on SNP specific probes and the possibility of a third allele does not complicate the analyses. Furthermore, multiplexing is possible, thereby reducing the amount of input DNA required. In addition, most forensic laboratories possess the instruments needed. A disadvantage of the SNaPshot™ is that the amount of fluorescent signal differs per nucleotide dye. The ratio G:A:T:C was estimated to be 3:2:1:1 after measuring the average allele peak heights per SNP in the 153 Dutch samples with SNPstat. This ratio showed some variation both between individual SNPs and between samples for the same SNP. For SNP 11a, the ratio G:C sometimes rose to 8:1, and the allele balance cut-off value in the analysis method was set to 1/8 = 0.125. Due to this difference in signal and to

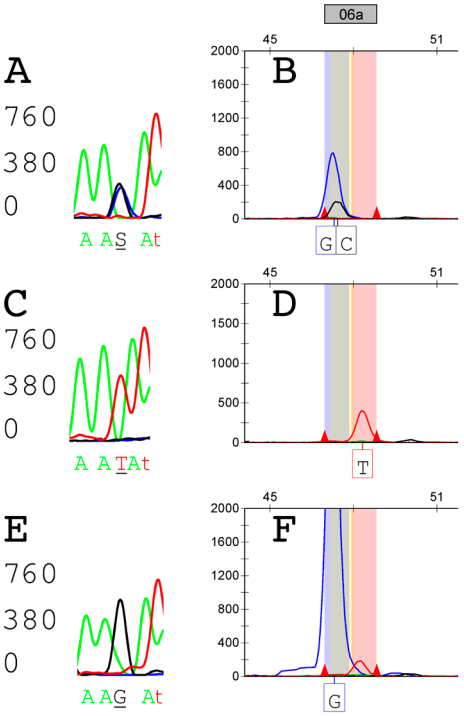


Fig. 1 Sanger sequencing (A, C, E) and SNaPshot™ (B, D, F) results for SNP 06a_rs9275142 for three individuals. (A and B) Heterozygous GC for both methods. (C and D) Homozygous T for both methods. (E and F) Sanger sequencing shows a clear homozygous G, but the SNaPshot™ results show a large G-peak and a small additional T-peak.

the interactions of the many PCR and extension primers present, SNaPshot™ multiplexes require several optimization steps. Phillips and co-workers also encountered these problems and compared four forensically relevant SNP typing techniques: SNaPshot™ genotyping, TaqMan™ real-time PCR assays, Sequenom™ iPLEX™ MALDI-TOF spectrometry and Genplex™ oligo-ligation assays (a modification of the SNPlex™ chemistry), of which the Genplex™ system seemed the most promising alternative [31].

Allele frequencies and statistics

Genotyping data, allele frequency distributions and a summary of the statistics for the SNP markers in the Dutch and Netherlands Antilles reference samples are shown in Table 1 and supplementary tables S3, S4 and S5. A few p-values are below the threshold of 0.05, but after Bonferroni correction, no significant deviation from Hardy–Weinberg equilibrium or linkage was observed. The number of tri-allelic SNP markers that we examined does not suffice to reach a discrimination power that equals 10–15 STR markers, and further research is needed (for discrimination and exclusion powers per SNP see supplementary table S4). Unfortunately, the amount of population data available in dbSNP is rather limited and does not enable an efficient pre-selection of non-binary SNPs with promising allele distributions at the moment. For the 15 SNP markers analysed, the allele distribution per population is visualised in Fig. 2. It is clear that the allele distribution of the tri-allelic SNPs can differ greatly between the two populations. For example, SNPs 08a and 12a both have an allele that is rare in the Dutch samples, while common in the Netherlands Antilles samples, and SNPs 05a and 18a show only two alleles in the Dutch while three alleles in the Netherlands

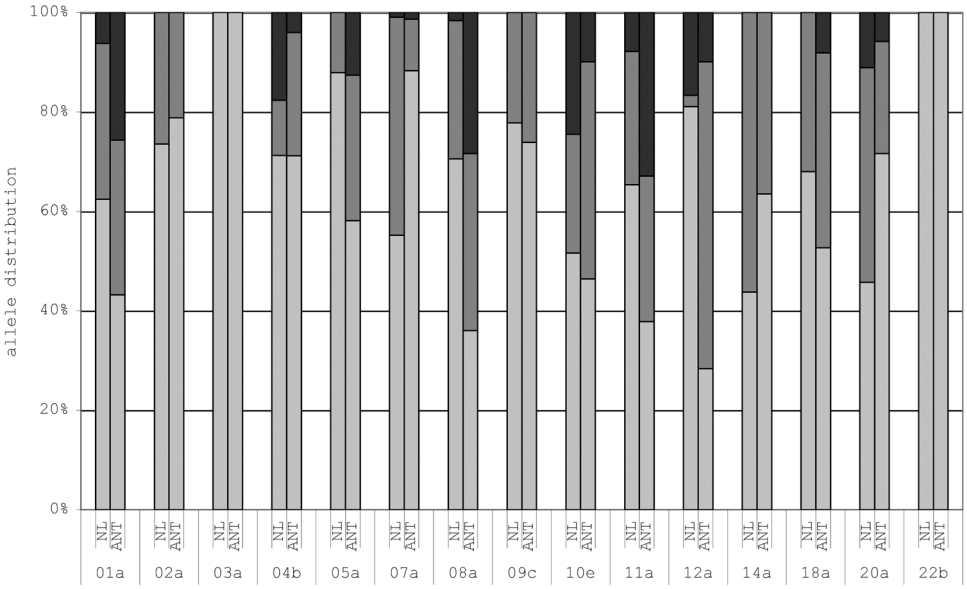


Fig. 2 Allele distributions for the SNPs that were analysed in the Dutch (NL, *n* = 306 alleles) and Netherlands Antilles (ANT, *n* = 222 alleles) samples.

Antilles samples (Fig. 2 and Table 1). As some of the alleles in these SNP markers seem to be determined geographically, these SNPs might not only be interesting for identification, but also as ancestry informative markers (AIMs) [32]. When the source individual of a DNA sample is unknown, AIMs can point out the most likely population of origin [32]. For this purpose, it is important to keep in mind that the examined Dutch DNA samples represent a cross-section of the Dutch population and that the donors may not all have a European background. The allele frequencies of the six YCC population groups are summarised in supplementary table S6 (notwithstanding the small sample sizes). These findings support the suggestion that SNPs 05a and 18a might be interesting AIMs, since the third allele is only detected in the African and South African populations.

Dilution series and mixtures

To assess the sensitivity of the three SNaPshot™ multiplex assays in relation to the STR profiling system AmpF*l*STR® SGM Plus™ a range of pristine hDNA PCR inputs between 5 pg and 50 ng was analysed. Using an input of 5 pg DNA, 43, 67, 40, and 9 % of the genotypes was obtained for multiplex A, B, C and SGM Plus™, respectively. With an input of 10 or 50 ng DNA, the SNaPshot™ multiplex assays were overloaded but still interpretable, while the SGM Plus™ resulted in strongly overloaded profiles or "no

sizing data". The minimal amount of input DNA with which full profiles were obtained are 300, 200, 100 and 50 pg for multiplex A, B, C and SGM Plus™, respectively (data not shown). Thus, although SGM Plus™ is better capable of generating full profiles, the SNaPshot™ multiplex assays provide a higher percentage of detected alleles using very minute amounts of DNA and give genotyping data when very high DNA inputs are used.

Next, we investigated whether the mixing of samples can be detected using the tri-allelic SNP assays. Two-donor mixtures in various ratios between 1:8 and 8:1 were analysed in which the total amount of input DNA was 2 ng per reaction. Two individuals were selected that differ for five of the seven SNP markers that are present in multiplex A (Fig. 3A and B). When the DNA of these two individuals is mixed, it is expected that (1) three alleles are visible for SNPs 05a and 08a, (2) altered peak height ratios are visible for SNPs with overlapping alleles like SNPs 07a and 02a and (3) SNPs with no overlapping alleles look like a heterozygous when mixed in a 1:1 ratio or have an altered peak height ratio with other mixture ratios like SNP 04b. In a 1:1 mixed sample (Fig. 3C), clearly three alleles are detected for SNPs 05a and 08a pointing to the presence of a second DNA source. This finding is supported by the detection of altered peak height ratios for SNPs 07a and 02a. Normally the ratio G:A:T:C is around 3:2:1:1. For SNP 02a the C:T ratio in the 1:1 mixed sample is around 1:7 and thereby distinct from the normal 1:1 ratio. For SNP 07a the G:T ratio is far above the normal 3:1 ratio resulting in an uncalled T-peak because this peak (that is clearly above the allele calling threshold of 50 rfu) falls below the allele balance cut-off value of 0.125 (corresponding to a G:T ratio of 8:1). In a 1:8 mixture (Fig. 3D), three alleles are detected for SNP 05a only. In addition, an altered peak height ratio is seen for SNP 04b: the normal G:C ratio of 3:1 has lowered to 1:4. A 8:1 mixture from the same donors and two-donor mixtures from other individuals show similar results (data not shown). In conclusion, in 1:8 to 8:1 mixed samples the presence of a second DNA source is recognised in the SNaPshot™ assays for the tri-allelic SNP markers. The indicators are the presence of three alleles on one locus, unexpected peak height ratios and uncalled peaks above the detection threshold.

The presence of three alleles on one locus is the clearest sign for the occurrence of a mixture and does not depend on quantification of the fluorescent signal. This quantification is complicated for the analysis of SNaPshot™ assays since the fluorescent signal differs per dye, but is possible in other SNP typing technologies such as pyrosequencing and mass spectrometry [33]. However, compared to SNaPshot™ assays these methods have other limitations such as less multiplexing capability or the need for a higher amount of input DNA [27,34].

In order to estimate the utility of the tri-allelic SNPs for mixture detection we determined the theoretical occurrence of a third allele on at least one locus by evaluating all possible two-person mixtures in the Dutch and Netherlands Antilles

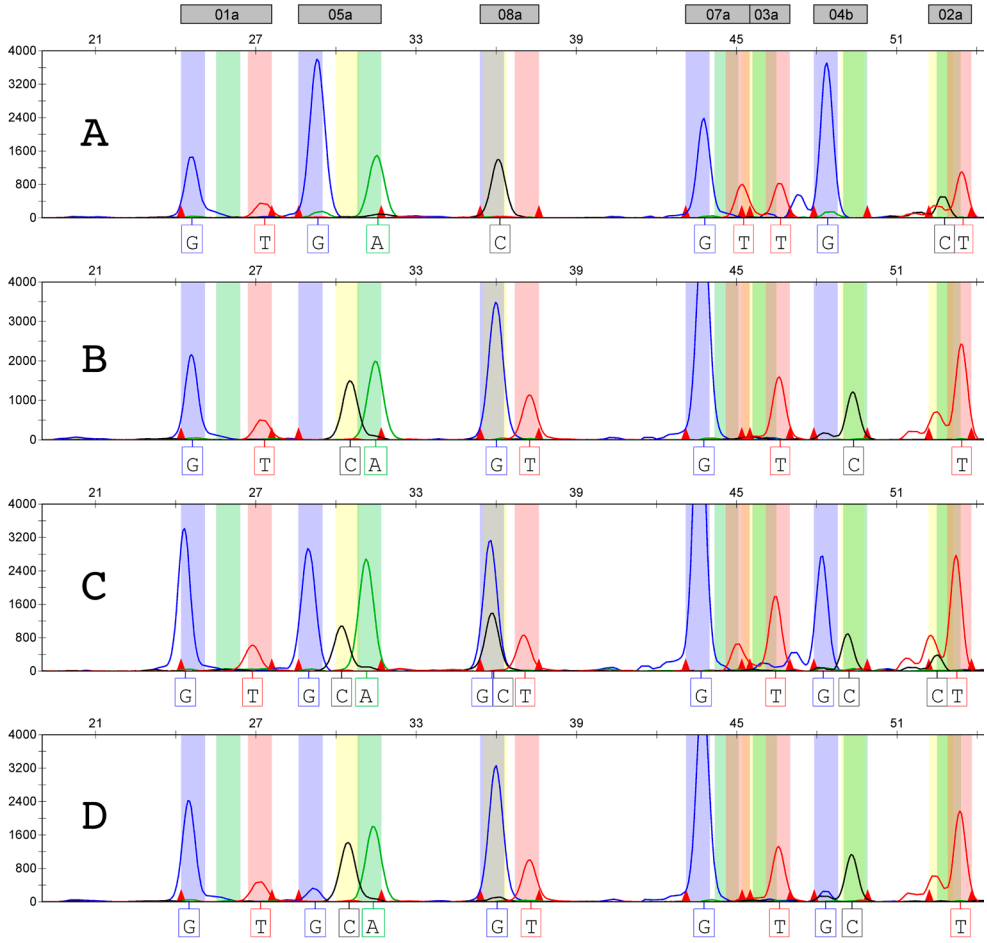


Fig. 3 Detection of a mixture of two individuals using tri-allelic SNP SNaPshot™ multiplex A that analyses 7 SNP markers. The horizontal grey bars on top label the SNP markers. (A) SNP profile for person 1. (B) SNP profile for person 2. (C) 1:1 mixture for person 1:person 2. (D) 1:8 mixture for person 1:person 2. These individuals differ for markers 05a, 08a, 07a, 04b and 02a.

reference samples. The percentage of detected mixtures was determined by two approaches: automated counting (Table 2) and a statistical approximation based on the allele frequencies (supplementary table S7). The Dutch and Netherlands Antilles populations have different allele frequencies, and therefore we determined the percentage of detected mixtures both separately and combined (Table 2). 75 % of the two-person mixtures within the Dutch population is detected (based on 8 tri-allelic SNPs), while 95 % of the mixtures is detected for the Netherlands Antilles samples

(based on 10 tri-allelic SNPs). The counting and the statistical approximation show similar results (Table 2). Even this limited number of tri-allelic SNP markers effectively detects the majority of the mixtures.

Degraded samples

In order to obtain information on the performance of the tri-allelic SNP assays to analyse degraded DNA, pristine DNA was artificially damaged. Native and denatured hDNA samples of 200 ng/μL were irradiated for increasing time with UV light in a cross-linker. By denaturing the DNA prior to UV irradiation, we intended to induce the formation of single-stranded breaks, which are the most common type of post-mortem DNA degradation [35]. Analysis of the samples on ethidium bromide stained, 0.8 % agarose gels showed that the UV treatment had resulted in DNA degradation rather than inter-strand cross-linking since reduced sized DNA smears were visible for both the native and the denatured DNA samples. In addition, longer UV treatment resulted in smears of reduced fragment length (results not shown). The denatured samples were selected to test the performance of the tri-allelic SNP assays on artificially degraded DNA. The samples of 200 ng/μL were diluted 200-fold and 1 μL was used as PCR input for the SNaPshot™ and SGM Plus™ analyses. Fig. 4A shows that SGM Plus™ STR profiling fails for the higher molecular weight STR markers when DNA is treated by 5 min of UV irradiation, and that only 14 % of the alleles is called when DNA is treated by UV irradiation for 120 min. In contrast, the SNaPshot™ multiplex assays show their first loss of marker detection when using DNA treated by UV irradiation for 60 min, and 73 % of the alleles are still called when using DNA treated for 120 min of UV irradiation.

Furthermore, pristine hDNA samples were degraded using increasing TURBO™ DNase concentrations. Reduced sized DNA smears were visible after running these samples on an ethidium bromide stained, 2 % agarose gel. DNA fragments were isolated in size ranges of approximately 400–350 bp, 300–250 bp, 200–150 bp and <100 bp by gel extraction. One nanogram DNA was used in both the SNaPshot™ and SGM Plus™ analyses. Fig. 4B shows that the percentage of detected STR alleles reduces with decreasing fragment length, and that no

Table 2
Two-person mixture detection by three alleles on at least one locus.

Population	*n* samples	*n* profile comparisons	*n* tri-allelic SNPs/population	% detected mixtures	Statistical approximation (%)
NL	153	11628	8	75.3	74.9
ANT	111	6105	10	95.7	94.7
NL and ANT	264	34716	10	87.5	n.a.[a]

[a] The different background of the populations does not allow the use of a combined allele frequency.

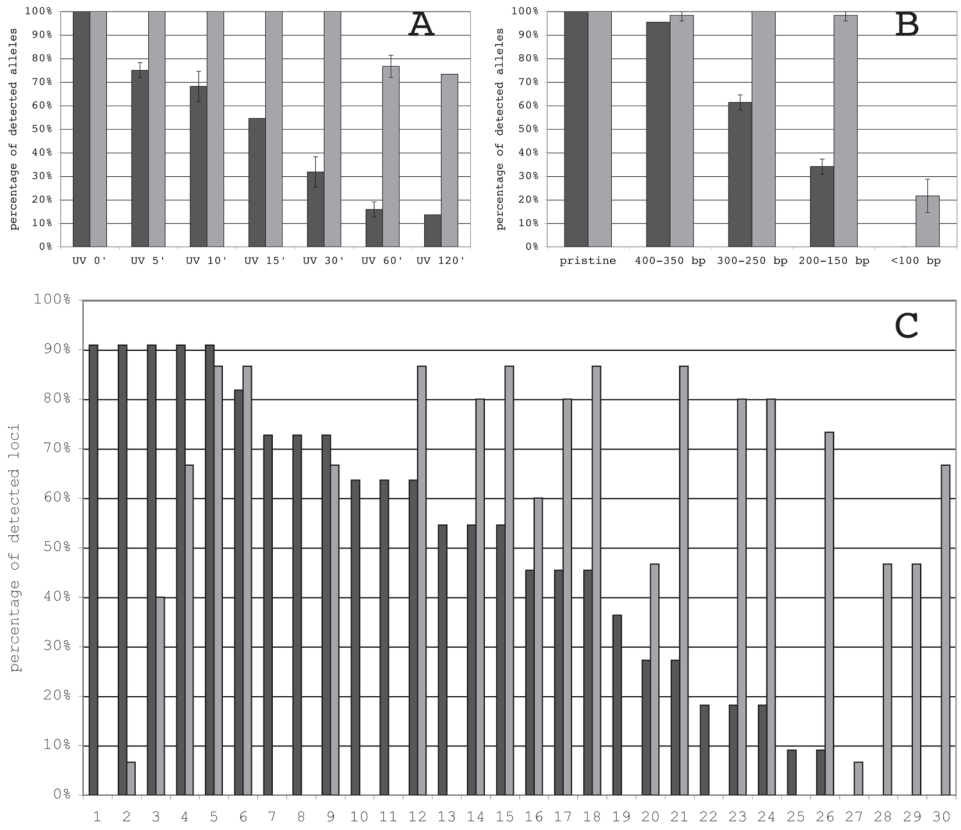


Fig. 4 Genotyping results for degraded DNA samples. Dark grey bars represent SGM Plus™ data (11 loci) and middle grey bars represent the SNaPshot™ results for multiplexes A, B, and C together (15 loci). The experiments shown in (A) and (B) were performed in duplo and the error bars represent the standard deviation; when no error bars are displayed, both measurements were equal. (A) UV irradiation time is plotted against the average percentage of detected alleles. (B) TURBO™ DNase degraded DNA fragments of decreasing length are plotted against the average percentage of detected alleles. (C) Results for 30 approximately 500-year-old bone and molar samples are plotted against the percentage of detected loci.

STR alleles are found when using DNA fragments smaller than 100 bp. For the latter DNA fragments, 22 % of the SNP alleles are still detected, and the SNP profiles are nearly complete when using DNA fragmented to 150–400 bp. These results show that the tri-allelic SNP markers are better capable of analysing artificially degraded DNA than SGM Plus™ STR profiling, which is most likely due to the use of smaller sized amplicons in the SNaPshot™ assays.

In addition to the artificially degraded DNA samples, thirty 450–550-year-old bone and molar samples were analysed using both the SNaPshot™ multiplex assays and SGM Plus™ STR profiling with a constant input of 3 µL DNA extract. Sixteen out of the 30 samples show an increase in the percentage of loci that were called for the SNP markers compared to the STR markers (Fig. 4C). The finding that for some samples STR data but no SNP data are obtained may have various reasons: (1) limiting sensitivity of SNP assays with low quantities of DNA, (2) differences between the two PCR assays in susceptibility for PCR inhibitors, and (3) level of optimization of the multiplex PCR. The individual SNP markers vary in robustness and the assays could be improved by further balancing of the multiplexes or development of a more sensitive assay, which would aid the analysis of both degraded and low quantities of DNA. STR locus drop-out mainly occurs for the larger sized amplicons, which is in accordance with DNA degradation in the samples. Four samples provided SNaPshot™ results, while no SGM Plus™ data are obtained. This is probably due to a high level of DNA degradation in these samples. Thus, when the higher molecular weight STR markers fail to amplify, tri-allelic SNP markers may provide additional information.

Conclusion

In this study 15 tri-allelic SNPs on 14 different chromosomes are detected in DNA samples from Dutch and Netherlands Antilles donors. We showed that such non-binary SNPs have the ability to reveal the presence of a second DNA donor in mixed samples with a ratio up to 1:8. Indications for a mixture are the presence of a third allele on one locus, unexpected peak height ratios and uncalled peaks above the detection threshold. Several of the tri-allelic SNP markers may not only be interesting for identification purposes, but also as ancestry informative markers. Furthermore, degraded (UV irradiated, TURBO™ DNase treated and 500-year-old bone and molar) DNA samples show that when the higher molecular weight STR markers fail to amplify, tri-allelic SNP markers can still provide valuable information.

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, the technical advice from Kristiaan van der Gaag and the statistical advice from Klaas Slooten.

Supplementary data

Table S1
Primers for the non-binary SNP candidates in multiplex A, B, and C

multiplex	SNP name	F-primer	R-primer	F&R conc (nM)	amplicon size	B-primer	B conc (nM)	direction	SNP
A	01a_rs3091244	GCGAAAATAATGGGAAATGG	GGCTGAAGTAGGTGTTGGAG	100	86	aaGGAAATGGTAACATATTAAAC	30	F	D
A	02a_rs727241	TGCTTTCAATCATTCCTTGTC	ACTCACAGCTCTGCTCCTTAC	100	100	actaggtgccacgtctgtaaagtcttgacaaTGCTCCTTACTAATTCAAAG	40	R	D
A	03a_rs35528968	TCCATTTGATGGTGCTGAC	TTCAATAGCCAGATCCTTTG	100	86	ccacgtcgtaaagtcttgacaaTGGTGCTGACTCTTTATTAT	30	F	H
A	04b_rs356167	AAAACCATCACCACAGTTCC	CCAAATGTGCCAGATACAAAC	100	91	gtgccacgtcgtgaaagtcttgacaaCTGTTGTTGTTCTTCTTCTT	5	F	V
A	05a_rs9329104	AATCTCTAACGTAAGCAACTGC	GCTGGAAACACAGCCTGTAG	100	96	ctgacaaCTTTTAGAACTATCTCCTCCT	10	F	V
A	07a_rs2032582	CATATTTAGTTTGACTCACCTTCC	TGTTGTCTGGACAAGCACTG	100	72	cgtcgtgaaagtcttgacaaAAGATTAAGAAAGAACTAGAAGGT	15	R	H
A	08a_rs433342	AATCTGTGAGCATCCACCTC	AGAACAGGCAAAAATGATGG	100	58	tgaaagtcttgacaaGATGGGAAAAGAGATGTG	20	R	V
B	06a_rs9275142	AATCCCAGAAGGGTTTATGC	GACCTCAAGACCAAATTAGATG	100	97	taggtgccacgtcgtgaaagtcttgacaaTACTCCATCAAGGAGAA	75	R	V
B	09c_rs1112534	GTTTGCTGATCCTTTGGGTC	TTCAGACCCCTTCCTAAGTTC	100	97	agtcttgacaaTTTTAACTCTAGAATGGGTAGTA	5	F	V
B	11a_rs5030240	CAAAGTGCCAGGATCACAG	TTCAGCCCTAGAAAATGTGG	100	67	ccacgtcgtgaaagtcttgacaaGGAAAGTAGGCCAGG	15	R	V
B	12a_rs2307223	CACCTGTAGCAGATAGGAACTG	GAGCGGAACGAGAAGATG	100	96	tgacaaCAATTACACAACAGAACACTAT	5	R	H
C	10e_rs17287498	CAGGGAGGGTGCAAAGTG	GGCTGGACAGGAGAACAAC	100	69	cacgtcgtgaaagtcttgacaaTTAACTGCATTTGTTTTTAC	20	F	H
C	14a_rs1008686	AGCTTCCTGCAACAAGAAC	GCCCTCAAAACACTATCAAGC	100	73	caaATCTCTGGTTTCTCTGAGTT	10	R	D
C	18a_rs2853525	GCTGTTGGATTTTAGTAGGAAAG	TTCCTGTGAATCTGTAAAACAA	100	68	tgccacgtcgtgaaagtcttgacaaTTGGATTTTAGTAGGAAAGTT	15	F	H
C	20a_rs2069945	CCTACGTGCAAACCTTGG	TTTCCCCAGTGGCTTAATG	100	100	caaTATTAACCCCAGTCTACATG	20	R	V
C	22b_rs34741930	ACCAGAACCTCCCTAAGCAC	TCCTTGAAGAAGACGACAGG	100	68	actaggtgccacgtcgtgaaagtcttgacaaTTAGCTTTCAAGTGGACAG	40	R	D

Table S2

Unexamined potential non-binary SNPs found in dbSNP with a custom-made algorithm

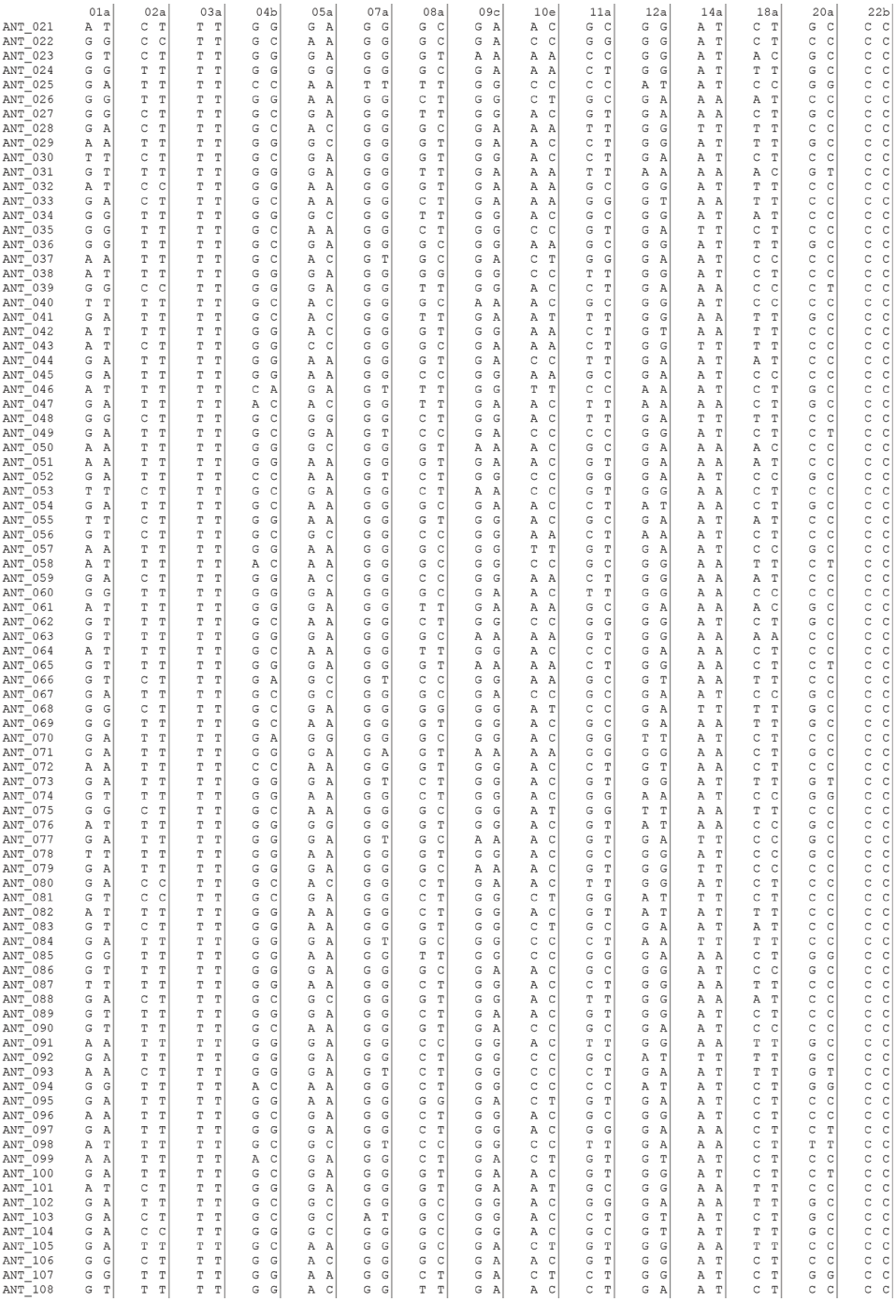
chromosome	rs-number	dbSNP maximum minor allele frequency*
1	2184030	11,9
6	668871	11,8
6	865577	10,4
6	956297	14,3
6	1059553	21,4
6	1694115	10,7
6	2072899	15,1
6	2246068	19,2
6	2523610	10,7
6	2621363	12,5
6	2647086	21,4
6	3095299	12,5
6	3104368	10,3
6	3129158	12,9
6	4993691	17,9
6	6925893	17,9
6	9264962	15,9
6	9274703	28,6
6	9274723	21,1
6	9275142	27,1
6	9276016	14,3
6	17203067	10,4
6	17203741	12,0
6	17210062	12,0
8	353721	12,5
8	36031660	10,0
9	6560007	18,2
10	1047111	16,7
12	720578	20,0
14	28909974	12,5
17	2642157	27,6
17	5819132	16,7
19	6508976	16,7
22	3859849	19,6

* The population with the highest minor allele frequency in dbSNP (when allele frequencies were determined for more than one population).

Table S3
Genotyping data of Dutch, Netherlands Antilles and YCC samples analyzed with multiplex A, B and C

	01a	02a	03a	04b	05a	07a	08a	09c	10e	11a	12a	14a	18a	20a	22b
NL_001	G G	C C	T T	G C	A A	G T	C C	G G	A C	G G	A A	T T	C C	T T	C C
NL_002	G G	T T	T T	G G	A A	G T	C T	G G	C C	G G	A A	A A	T T	G G	C C
NL_003	G A	T T	T T	G G	A A	G T	C C	G A	C C	G C	A A	A T	T T	G C	C C
NL_004	G A	C T	T T	G G	A A	G T	C C	G G	C T	G C	A T	A T	T T	G C	C C
NL_005	G A	C T	T T	G G	A A	G T	T T	G A	C T	G C	A T	T T	C C	G G	C C
NL_006	G G	C T	T T	G G	A A	G T	C T	A A	C T	G G	A T	A T	C T	C T	C C
NL_007	A T	T T	T T	G G	G A	G T	C T	G G	A C	C C	A A	A T	T T	G T	C C
NL_008	G G	T T	T T	G G	A A	G T	C T	G A	C C	G T	G A	A A	T T	G T	C C
NL_009	G A	T T	T T	G G	A A	G T	C C	G G	C C	G G	A A	A T	T T	G T	C C
NL_010	G A	C T	T T	G A	G A	G T	C C	G A	C C	C C	A A	A T	T T	G T	C C
NL_011	A A	T T	T T	G A	A A	G T	C C	G A	C T	A A	T T	C T	C T	G C	C C
NL_012	G A	C C	T T	G G	A A	G T	C T	G G	A T	G G	A A	A T	T T	G G	C C
NL_013	G A	C C	T T	G G	G A	G T	C T	G G	C T	G T	A A	A T	T T	G G	C C
NL_014	G A	C T	T T	G G	A A	G T	T T	G G	C C	G C	A A	T T	T T	G G	C C
NL_015	G G	T T	T T	G G	A A	G T	C T	G G	C C	G G	A T	T T	C T	G G	C C
NL_016	G T	T T	T T	G A	A A	G T	C T	G A	C C	G C	A A	A T	T T	G G	C C
NL_017	G A	C T	T T	G G	A A	G T	G G	G A	C C	G C	A A	T T	T T	G C	C C
NL_018	G G	C T	T T	G C	A A	G G	C T	G A	C T	G C	A A	A T	T T	G G	C C
NL_019	A A	C T	T T	G A	A A	G G	C T	G A	C C	G C	A A	T T	C T	G C	C C
NL_020	A A	T T	T T	G A	G A	G T	C T	G A	A A	G G	A A	A T	C T	G G	C C
NL_021	G G	C C	T T	G G	G A	G T	C T	G A	A A	G G	A A	T T	C T	C T	C C
NL_022	G G	T T	T T	A A	A A	G T	T C	G A	C C	C C	A A	A T	C T	C T	C C
NL_023	G G	T T	T T	G A	A A	G T	C T	A A	A T	G G	A A	A T	T T	G C	C C
NL_024	G G	C C	T T	G C	A A	T G	T C	G G	C T	G G	A A	T T	T T	C C	C C
NL_025	G A	C T	T T	G C	A A	G T	C T	G G	C C	G G	A A	T T	T T	G C	C C
NL_026	G G	C T	T T	G C	G A	G T	C T	G G	C C	G C	A T	T T	T T	C C	C C
NL_027	G G	C T	T T	G C	A A	G T	C T	G G	C T	G G	A A	A T	C T	G G	C C
NL_028	G A	C T	T T	G C	A A	G G	C T	G G	A A	C C	A A	A A	T T	C C	C C
NL_029	G A	T T	T T	G G	A A	G T	C T	G A	A T	C T	A A	A T	C T	C C	C C
NL_030	G G	T T	T T	G A	A A	G G	C T	G A	C C	G C	A T	A T	C T	C C	C C
NL_031	G G	T T	T T	G C	A A	G T	C T	G A	C C	G C	A T	T T	C T	G C	C C
NL_032	G G	T T	T T	G C	G A	G T	C T	G G	C C	G C	A A	A A	C T	G G	C C
NL_033	G A	T T	T T	G G	A A	G T	C T	G G	C C	G G	A A	T T	C T	G G	C C
NL_034	A A	T T	T T	G C	A A	G T	C T	G G	A C	G T	A A	A T	C T	G G	C C
NL_035	G A	T T	T T	G A	A A	T G	T G	G G	A C	G C	A A	A A	C T	T T	C C
NL_036	A A	C C	T T	G G	A A	T G	C T	G G	A C	G T	A A	A T	C T	G G	C C
NL_037	A A	C C	T T	G A	A A	T T	C T	G A	C T	C G	A A	A T	T T	G C	C C
NL_038	G G	C T	T T	G A	A A	T G	C T	G A	C C	G G	A A	T T	C C	G C	C C
NL_039	G A	C T	T T	G C	A A	T G	C C	G A	C T	G G	A A	T T	C C	C C	C C
NL_040	G G	C T	T T	G G	A A	G T	C T	G G	C T	G T	A A	T T	C T	C C	C C
NL_041	G A	C C	T T	G G	A A	G T	C T	G G	T C	G T	A A	A T	C T	G G	C C
NL_042	G G	C T	T T	G A	G A	T T	C T	G G	A C	G G	A A	A T	C T	T G	C C
NL_043	G A	T T	T T	G G	A A	T G	C C	G G	A C	G G	A A	A T	C T	G G	C C
NL_044	G G	C T	T T	G A	A A	G G	C C	G G	A C	G T	A A	A T	C T	T C	C C
NL_045	G G	C T	T T	G G	A A	G G	C T	G G	A C	G G	A T	A T	C T	G C	C C
NL_046	A A	T T	T T	G G	G A	G A	C T	G G	A T	G C	A A	A T	T T	G C	C C
NL_047	G A	T T	T T	A A	A A	G T	C T	G A	A C	G G	A A	A T	C T	G C	C C
NL_048	G G	T T	T T	C C	A A	G T	C C	G A	C T	G G	A A	T T	C T	G C	C C
NL_049	G A	T T	T T	G G	A A	G G	C C	G A	C T	G T	A T	T T	C T	G C	C C
NL_050	A A	T T	T T	G G	A A	G G	C C	G G	A C	G C	A T	A T	C T	G C	C C
NL_051	G T	T T	T T	G A	A A	G G	C C	G G	A C	G G	A T	A A	T C	T C	C C
NL_052	G A	T T	T T	G A	A A	G T	C C	G G	A T	G C	A A	A T	C T	C T	C C
NL_053	G A	T T	T T	G G	G A	G T	C C	G G	C T	G C	A A	A T	C T	T T	C C
NL_054	G G	T T	T T	G G	A A	G T	C C	G G	A C	G C	A A	A T	C T	G T	C C
NL_055	G G	C T	T T	G G	A A	T G	C T	G G	T C	G C	A T	A T	T T	C C	C C
NL_056	G G	C T	T T	G G	G A	A G	C T	G A	C C	G C	A A	A T	C T	T T	C C
NL_057	G A	T T	T T	G A	A A	T G	C C	G G	A C	G C	A A	T T	C T	G C	C C
NL_058	G G	T T	T T	G A	A A	T G	C T	G G	C A	G C	A A	A T	T T	G C	C C
NL_059	G A	C T	T T	G G	A A	G T	C T	G A	C T	G T	A A	A T	T T	C T	C C
NL_060	G A	C T	T T	G G	A A	G T	A C	G G	A A	G G	A T	A T	C T	C C	C C
NL_061	G T	C T	T T	G G	G A	G A	C C	G A	C C	G C	A T	T T	C T	G C	C C
NL_062	G G	C T	T T	G A	A A	G G	C C	G A	C C	G T	A A	A T	C T	T C	C C
NL_063	G G	C T	T T	G A	A A	G T	C C	A G	C T	G G	A A	A T	C T	G G	C C
NL_064	G A	T T	T T	G G	G A	G T	C C	G G	A A	G G	A T	A T	T T	G C	C C
NL_065	A A	C C	T T	G A	A A	T G	C C	G G	A T	G G	A A	T T	C T	G C	C C
NL_066	G T	C T	T T	G A	A A	G T	T C	G G	C C	G G	A A	T T	C T	G C	C C
NL_067	A A	C T	T T	C C	A A	G T	C T	G G	C C	G G	A T	T T	C T	C T	C C
NL_068	G G	T T	T T	G C	A A	G T	C T	G G	A C	G C	T T	A T	T T	C C	C C
NL_069	G A	C T	T T	G C	A A	G G	C T	G A	A C	G C	G T	A T	C T	C C	C C
NL_070	G A	C C	T T	A A	A A	G G	C T	G A	C T	G C	A T	T T	C T	G C	C C
NL_071	G G	C T	T T	G C	G A	G T	C T	G A	C T	G C	A A	T T	C T	C C	C C
NL_072	G G	T T	T T	G G	A A	G T	C T	G G	A C	G C	A A	T T	C T	G C	C C
NL_073	G A	T T	T T	G G	G A	G A	C T	G A	C T	G C	A A	A T	C T	G C	C C
NL_074	G G	T T	T T	G G	A A	G G	C T	G A	A C	G G	A T	T T	C C	T C	C C
NL_075	G A	C T	T T	G G	A A	G G	C T	G G	T C	G G	A A	A T	C C	G C	C C
NL_076	G T	C T	T T	G G	A A	G G	C C	G G	C C	G G	A A	A T	C T	T C	C C
NL_077	G A	C T	T T	G G	A A	G T	C C	G G	A C	C T	A A	A A	T T	G G	C C
NL_078	G A	C T	T T	G C	A A	G T	C T	G G	C C	C T	A A	A T	T T	G G	C C
NL_079	A A	C T	T T	G A	A A	G T	T C	G G	C T	G C	A A	A A	T T	G C	C C
NL_080	G G	C T	T T	A A	A A	G T	C C	G G	C C	G G	A A	A T	C T	G C	C C
NL_081	G A	C T	T T	G G	A A	G T	C T	G G	A C	G C	A T	A T	T T	C C	C C
NL_082	G G	T T	T T	G A	G A	G T	T T	G G	C C	G C	A A	A A	T T	G C	C C
NL_083	G G	C T	T T	G G	A A	G G	C T	G G	C C	G C	A A	A T	T T	G C	C C
NL_084	G T	C T	T T	G C	A A	G G	C T	G G	A A	G C	A A	A A	C C	G G	C C
NL_085	G G	T T	T T	G G	A A	G T	C T	G G	C T	G C	T T	A T	C T	G C	C C

	01a	02a	03a	04b	05a	07a	08a	09c	10e	11a	12a	14a	18a	20a	22b
NL_086	A T	T T	T T	G G	A A	G T	C T	G A	A C	G C	A A	A T	T T	G C	C C
NL_087	G G	C T	T T	G A	A A	G T	G T	G A	A C	G C	A T	A A	T C	G C	C C
NL_088	G G	C T	T T	G G	A A	G T	C T	G G	A T	G C	A A	A T	C T	C T	C C
NL_089	G A	C T	T T	G G	A A	G T	C T	G G	C T	G G	A A	A T	C C	T C	C C
NL_090	G A	C T	T T	G G	A A	G T	C T	G G	C C	G C	A T	A T	C T	C T	C C
NL_091	G A	C T	T T	G G	G A	G T	C T	G A	C C	G G	A A	T T	C C	G C	C C
NL_092	G G	C T	T T	G G	A A	G T	C T	G A	C C	G G	A T	A A	C T	G C	C C
NL_093	G A	C T	T T	G C	A A	G T	C C	G A	C C	G G	T T	A A	C T	C G	C C
NL_094	G T	C T	T T	G A	G A	G T	C C	A A	C C	G T	A A	T T	C T	G G	C C
NL_095	G A	T T	T T	G A	A A	G T	C C	G A	C A	G C	A T	A T	C T	G G	C C
NL_096	A A	T T	T T	G C	A A	G T	C C	G A	C C	G C	A A	A A	C T	C C	C C
NL_097	G A	C T	T T	G A	A A	G T	C C	G G	A C	G C	A T	A T	C T	G C	C C
NL_098	G A	C T	T T	G A	A A	G T	C C	G G	C T	G T	A A	A T	C T	C C	C C
NL_099	G A	T T	T T	G C	A A	G G	C C	G G	C T	G G	A A	A A	C T	G G	C C
NL_100	G G	T T	T T	G A	A A	G G	C T	G G	A T	G G	A A	A A	C T	G C	C C
NL_101	G A	T T	T T	G G	A A	G G	C C	G A	A T	C T	A T	A A	T C	C C	C C
NL_102	G T	T T	T T	G C	G A	G T	C C	G A	A C	T T	G A	A T	C C	C C	C C
NL_103	G G	C T	T T	G A	A A	G T	C C	G G	A T	G G	A A	T T	C T	C C	C C
NL_104	G G	T T	T T	G G	G A	G T	C C	G G	A T	G C	A A	T T	C T	G T	C C
NL_105	G A	C T	T T	G G	G A	A A	C C	G G	C C	G G	G A	T T	C T	T C	C C
NL_106	A T	C T	T T	G G	A A	G T	C T	G G	A C	G C	A A	A T	C T	C T	C C
NL_107	G G	C T	T T	G A	A A	G T	C T	G G	A T	G C	A A	A T	C T	C T	C C
NL_108	G G	C T	T T	G G	A A	T T	C C	G G	A T	G G	G T	A T	C T	C G	C C
NL_109	G A	C T	T T	G G	A A	G G	C C	G G	A C	G C	A A	T T	C T	G G	C C
NL_110	G T	C T	T T	G G	A A	G T	C T	G G	A C	G C	A T	A T	C T	C T	C C
NL_111	G G	C T	T T	G G	A A	G T	C T	G G	A T	G C	A A	A A	C T	G G	C C
NL_112	G G	C T	T T	G A	G A	A T	C C	G G	A T	C C	A A	A T	C C	T G	C C
NL_113	A A	C T	T T	G G	A A	G T	C C	G G	A T	C C	A T	A A	C T	G G	C C
NL_114	G A	C T	T T	A C	G A	G T	C C	G G	C A	G C	A T	A T	C T	T G	C C
NL_115	G G	C T	T T	G C	A A	G T	C C	G G	C A	G C	A A	T T	C T	C T	C C
NL_116	G G	C T	T T	G C	A A	G T	C C	G G	C A	C C	A A	T T	C C	C T	C C
NL_117	G T	C T	T T	G C	G A	G T	C C	G G	A C	T C	G T	A A	C C	G C	C C
NL_118	A A	C T	T T	G C	A A	T T	G C	G G	A T	G C	A A	A T	C C	C C	C C
NL_119	G A	C T	T T	G A	G A	G T	C C	G A	C C	G C	A A	A T	C C	C C	C C
NL_120	G A	C T	T T	G G	A A	G T	C T	G G	A T	G C	A A	A T	C C	G T	C C
NL_121	G G	C T	T T	G G	G A	A T	C T	G G	A T	G G	A A	A T	C C	C T	C C
NL_122	A T	C T	T T	G A	A A	G T	C T	G G	A A	G G	A T	A T	C C	G C	C C
NL_123	G G	C T	T T	G G	A A	G T	C T	G G	A C	G G	A T	A T	C T	C C	C C
NL_124	G T	T T	T T	G G	A A	G T	C T	G G	A T	G G	A A	T T	C T	G G	C C
NL_125	G A	C T	T T	G G	A A	G G	C C	G G	C C	G C	A T	A T	C T	C C	C C
NL_126	G A	C T	T T	G A	A A	G G	C C	G G	C C	G G	A A	A T	C C	G C	T C
NL_127	G T	T T	T T	G G	A A	G T	C C	G G	C T	G C	A T	T T	C T	C T	C C
NL_128	G A	C T	T T	G A	A A	G T	C C	G A	C T	G G	A A	T T	C T	G C	C C
NL_129	G G	T T	T T	G A	G A	G T	C T	G G	C C	G G	A T	A T	T T	C G	C C
NL_130	G A	T T	T T	G G	A A	G T	C T	G G	C C	G G	A A	A T	T T	G G	C C
NL_131	G G	T T	T T	G A	A A	G T	C C	G G	C C	G G	A A	A T	T T	G G	C C
NL_132	A A	C T	T T	G A	A A	G G	C C	G G	C C	G C	A A	A T	T T	G G	C C
NL_133	G G	C T	T T	G A	A A	G T	C C	G G	C A	G C	A T	A T	C T	G G	C C
NL_134	G G	C T	T T	G G	G A	A T	C C	G G	C A	G T	A A	A T	C T	T G	C C
NL_135	G A	C T	T T	G C	A A	G G	C T	G G	A A	G G	A T	A T	C T	G C	C C
NL_136	G A	C T	T T	G A	A A	G G	C C	G G	A A	G G	A A	A T	T T	G C	C C
NL_137	G A	C T	T T	G A	G A	G T	C C	G A	A C	G G	A A	T A	C T	G C	C C
NL_138	G G	C T	C T	G C	A A	G T	G C	A G	C C	G G	A T	A A	C T	G C	C C
NL_139	G G	C T	T T	G G	A A	G G	C C	G G	C C	G C	G T	T T	C T	C C	C C
NL_140	A T	T T	T T	G G	A A	G T	C T	G G	A C	G C	A T	A T	C T	G G	C C
NL_141	A A	T T	T T	G C	A A	G T	C T	G G	C C	G T	A A	A A	C T	T G	C C
NL_142	G G	C T	T T	G C	A A	G T	G C	G G	C C	G G	A A	A T	C T	G G	C C
NL_143	G G	C T	T T	G A	A A	G T	C C	G G	A C	G G	A T	A A	C T	G G	C C
NL_144	G A	C T	T T	G G	A A	G T	C C	G G	A C	G T	A A	A T	C T	G C	C C
NL_145	G A	C T	T T	A C	A A	G T	C C	G G	C T	G G	C A	A A	C C	T C	C C
NL_146	G G	C T	T T	G G	A A	G T	C T	G G	A T	G G	A A	T T	C C	T C	C C
NL_147	G G	C T	T T	G A	A A	G T	G C	G A	C C	G G	C A	A T	C T	C T	C C
NL_148	A A	C T	T T	G A	A A	G T	C T	G A	C C	G G	A T	A A	T T	C C	C C
NL_149	G A	C T	T T	G G	G A	G T	C T	A G	A C	G G	T A	A T	C T	G C	C C
NL_150	G A	C T	T T	G G	G A	G T	C T	A G	A C	G G	A A	T T	C T	T C	C C
NL_151	G A	C T	T T	G A	A G	T T	C T	G G	A C	G G	A A	A T	C T	G C	C C
NL_152	G T	C T	T T	G A	A A	G T	C T	G G	C T	G G	A A	A T	C C	C C	C C
NL_153	G A	C T	T T	G G	A A	G T	C C	G G	C A	G G	A T	T T	C T	G T	C C
ANT_001	G G	C T	T T	G G	A A	G T	C T	G G	A C	G C	A A	A T	T T	C C	C C
ANT_002	G A	C T	T T	G G	G A	G T	G T	G A	A C	G C	A A	A T	A C	C C	C C
ANT_003	G G	C T	T T	G G	G A	G T	C T	G A	A C	G C	A A	A T	G C	C C	C C
ANT_004	G T	C T	T T	C C	G A	C T	G T	G A	C C	C T	G G	A A	C T	C C	C C
ANT_005	G T	T T	T T	G G	A C	G G	G C	G A	C T	C G	G G	A T	C T	C C	C C
ANT_006	A T	C T	T T	G G	G A	G T	G C	G A	C A	G G	A A	A A	C T	C C	C C
ANT_007	G A	C T	T T	G C	G A	G T	G C	G G	A T	G T	G G	A A	C C	C T	C C
ANT_008	G G	C T	T T	G G	G A	G T	G C	G G	C C	G C	A A	A T	C A	T C	C C
ANT_009	G T	C C	T T	G G	G A	G T	G C	G G	C T	G C	A A	A T	C A	T C	C C
ANT_010	T T	C C	T T	G C	A A	G G	G C	G G	A A	C T	G A	A A	C T	C C	C C
ANT_011	A T	C T	T T	G C	A A	G G	G C	G G	A C	G T	A A	A A	C T	C C	C C
ANT_012	A T	C T	T T	G C	G A	G G	C T	G G	A C	G T	G A	A A	C T	C C	C C
ANT_013	T T	C T	T T	G C	G A	G T	G C	G A	A C	G T	A A	T T	T T	G C	C C
ANT_014	G T	C T	T T	G C	A A	G T	C C	G A	A T	G G	T A	A T	T T	G C	C C
ANT_015	T T	C T	T T	G C	G C	A T	G C	G G	A C	G G	G A	A A	T T	G C	C C
ANT_016	A T	C T	T T	G C	G A	G T	C T	G G	A C	G G	G A	A T	A C	C C	C C
ANT_017	T T	C T	T T	G G	G A	G T	G C	A G	A A	G G	G A	A T	C T	C C	C C
ANT_018	A T	T T	T T	C C	A C	G G	C T	G G	A C	G G	G A	T T	C T	C C	C C
ANT_019	G A	C T	T T	G C	G A	G T	C C	G A	C C	C G	G A	A T	T T	G G	C C
ANT_020	G G	C T	T T	G G	G A	G A	G C	G G	A C	G T	T T	A T	C C	G G	C C


01a
02a
03a
04b
05a
07a
08a
09c
10e
11a
12a
14a
18a
20a
22b
ANT_021
ANT_022
ANT_023
ANT_024
ANT_025
ANT_026
ANT_027
ANT_028
ANT_029
ANT_030
ANT_031
ANT_032
ANT_033
ANT_034
ANT_035
ANT_036
ANT_037
ANT_038
ANT_039
ANT_040
ANT_041
ANT_042
ANT_043
ANT_044
ANT_045
ANT_046
ANT_047
ANT_048
ANT_049
ANT_050
ANT_051
ANT_052
ANT_053
ANT_054
ANT_055
ANT_056
ANT_057
ANT_058
ANT_059
ANT_060
ANT_061
ANT_062
ANT_063
ANT_064
ANT_065
ANT_066
ANT_067
ANT_068
ANT_069
ANT_070
ANT_071
ANT_072
ANT_073
ANT_074
ANT_075
ANT_076
ANT_077
ANT_078
ANT_079
ANT_080
ANT_081
ANT_082
ANT_083
ANT_084
ANT_085
ANT_086
ANT_087
ANT_088
ANT_089
ANT_090
ANT_091
ANT_092
ANT_093
ANT_094
ANT_095
ANT_096
ANT_097
ANT_098
ANT_099
ANT_100
ANT_101
ANT_102
ANT_103
ANT_104
ANT_105
ANT_106
ANT_107
ANT_108

	01a	02a	03a	04b	05a	07a	08a	09c	10e	11a	12a	14a	18a	20a	22b
ANT_109	G G	T T	T T	G A	A A	G G	C C	G G	A A	G C	G A	A A	C C	C C	C C
ANT_110	G G	T T	T T	G A	A A	G G	G T	G A	A T	C T	G G	A T	C C	G C	C C
ANT_111	G T	T T	T T	C C	A C	G G	G T	G A	A C	G T	G G	T T	A C	C C	C C
AFR_YCC01	G A	T T	T T	G C	A A	G G	G G	A A	C C	G G	G G	A A	A C	C C	C C
AFR_YCC02	G T	T T	T T	G C	A A	G G	T T	G A	A C	T T	G A	A A	A T	C C	C C
AFR_YCC03	G G	C T	T T	G G	G A	G G	C T	A A	A A	G C	G A	A A	T T	C C	C C
AFR_YCC04	G G	C T	T T	G G	G C	G G	G T	G G	A C	G T	A A	A A	T T	C T	C C
AFR_YCC05	T T	T T	T T	G C	A A	G G	G G	G A	A A	C C	G G	A T	A C	G C	C C
AFR_YCC52	G G	C T	T T	G G	A A	G G	G T	G A	A C	G G	G G	A A	C T	G C	C C
AFR_YCC54	G T	C T	T T	G C	A A	G G	T T	G G	A C	G C	G G	A T	C C	C T	C C
AFR_YCC55	G A	C T	T T	G G	G A	G G	G T	G G	A A	G T	G G	A A	A T	C C	C C
AFR_YCC57	G A	T T	T T	G C	G A	G G	C T	G A	C C	T T	G G	A T	T T	C T	C C
NAM_YCC06	G G	T T	T T	G C	A A	G T	T T	G A	A C	G G	A T	A A	T T	C C	C C
NAM_YCC07	G G	T T	T T	G C	A A	G G	T T	A A	C C	G C	G A	A A	C C	G G	C C
NAM_YCC09	G A	C C	T T	G G	A A	G G	C T	G A	A C	G G	A A	T T	T T	G T	C C
NAM_YCC21	G A	T T	T T	G G	A A	G A	T T	G G	C C	C C	T T	A A	C T	G C	C C
NAM_YCC22	G A	T T	T T	G C	G A	G A	T T	G G	C C	C C	A A	A A	C T	G C	C C
NAM_YCC23	G G	T T	T T	G C	G A	G G	T T	A A	C C	C C	A A	A A	C T	G G	C C
NAM_YCC24	G A	T T	T T	G C	A A	G T	T T	G A	C C	C C	A A	A A	C T	G G	C C
NAM_YCC25	G G	T T	T T	G C	G A	G G	T T	A A	C C	C C	G A	A A	C T	G G	C C
NAM_YCC26	G A	T T	T T	G A	A A	G T	T T	A A	C T	G G	A A	A T	C T	G G	C C
NAM_YCC27	G A	T T	T T	C C	G G	G T	T T	A A	C C	G G	A A	A A	C T	G C	C C
NAM_YCC59	G A	T T	T T	G G	A A	G A	T T	G G	C C	C C	G A	A A	C T	G G	C C
EUR_YCC08	G A	C T	T T	G G	A A	G T	C T	G G	C T	C T	A A	A T	T T	G C	C C
EUR_YCC64	G A	T T	T T	G G	A A	G T	C C	G G	C C	G G	A A	A T	C T	C C	C C
EUR_YCC66	G G	C T	T T	G A	A A	G T	G C	G G	A C	G T	A A	A T	C T	G C	C C
EUR_YCC67	G G	T T	T T	G G	G A	G T	C C	G G	C C	G G	A T	A T	C C	G C	C C
EUR_YCC75	A A	C T	T T	G A	A A	G T	T T	G G	C C	G G	A A	A T	T T	C C	C C
ASI_YCC10	G G	T T	T T	G A	G A	G T	C T	G G	C T	G C	G A	A T	C T	G G	C C
ASI_YCC11	G T	T T	T T	C C	G A	G G	C T	G G	C T	G C	G G	A T	C T	G G	C C
ASI_YCC12	G G	T T	T T	C C	A A	G T	T T	G A	C C	G C	A T	A A	C T	G C	C C
ASI_YCC16	G A	T T	T T	G G	G A	G A	C C	A A	C C	C T	G T	A T	C T	C C	C C
ASI_YCC17	G A	T T	T T	C C	G A	G A	C C	G A	C C	G T	A A	A A	C T	C T	C C
ASI_YCC65	G G	C T	T T	G G	G A	G T	C T	G G	A C	G G	G A	A T	C C	C C	C C
ASI_YCC72	G T	T T	T T	G C	A A	T T	T T	G A	C C	G C	A A	A A	C C	G C	C C
ASI_YCC73	G G	T T	T T	G C	A A	A T	C T	G A	A C	C C	T T	A T	C C	G G	C C
SAF_YCC13	G G	T T	T T	G C	G A	G G	T T	G G	A A	G C	G G	A A	T T	C C	C C
SAF_YCC14	G T	T T	T T	G C	G A	G G	G C	G G	A A	G G	G A	A A	C T	C C	C C
SAF_YCC15	T T	T T	T T	C C	G C	G G	G T	G G	A C	G T	G G	A A	A T	C T	C C
SAF_YCC18	T T	C T	T T	G G	G A	G G	G T	G G	C C	G G	G G	A A	T T	C T	C C
SAF_YCC19	G A	T T	T T	G G	C C	G G	G T	G A	C C	G G	G G	T T	C C	C T	C C
SAF_YCC20	G T	T T	T T	G G	A C	G G	T T	G G	A C	G G	G G	T T	C C	C C	C C
SAF_YCC42	G T	T T	T T	G G	G C	G G	T T	G G	A C	G C	G G	A T	C T	G C	C C
SAF_YCC43	G T	C T	T T	G G	G C	G G	T T	G A	C C	T T	G A	A T	C T	C C	C C
SAF_YCC46	G G	C T	T T	G G	A A	G G	T T	G A	C C	G C	G G	A A	C T	C T	C C
SAF_YCC47	G T	C T	T T	G G	A A	G G	C T	G G	C C	C T	G G	A A	C T	C T	C C
SAF_YCC48	G A	C C	T T	G G	A A	G G	C T	G A	A C	G G	A A	T T	T T	G T	C C
SAF_YCC49	G G	T T	T T	G C	G A	G G	C T	G A	A C	T T	G A	A A	T T	C C	C C
SAF_YCC50	G G	T T	T T	C C	G A	G G	C T	G G	A A	G T	G G	A A	T T	C C	C C
SAF_YCC51	G G	T T	T T	G G	A A	G G	G T	G G	C C	G T	G G	A A	T T	C C	C C
RUS_YCC28	G G	T T	T T	G C	A A	A T	C C	G G	C T	G C	G A	T T	C C	C C	C C
RUS_YCC29	G G	T T	T T	G C	A A	G T	C T	G A	C C	G G	A A	T T	T T	G T	C C
RUS_YCC30	G G	T T	T T	G G	G A	G T	C T	G A	C C	G C	A T	A T	T T	G C	C C
RUS_YCC31	G A	T T	T T	G C	A A	G T	T T	A A	A T	G C	G A	A A	T T	G C	C C
RUS_YCC32	G G	T T	T T	G C	A A	T T	C T	G G	A C	G C	G A	A T	C T	G G	C C
RUS_YCC33	G G	T T	T T	G G	A A	G T	C T	G G	A T	G C	A A	A T	C T	G T	C C
RUS_YCC35	G T	T T	T T	G G	G A	G T	C T	G G	C T	G T	A A	T T	C T	G C	C C
RUS_YCC37	G G	T T	T T	A C	G A	G T	C T	G A	A C	G C	A A	A T	C T	G C	C C
RUS_YCC38	G G	T T	T T	G G	A A	T T	C T	G G	C T	G C	A A	A T	C T	G C	C C
RUS_YCC39	G G	C C	T T	G G	G A	G T	C T	G G	C C	G G	A A	A A	C T	G C	C C
RUS_YCC40	G G	C T	T T	G A	A A	G G	C C	G A	C T	G G	A A	T T	C T	G C	C C
RUS_YCC41	A A	T T	T T	G C	A A	G G	C T	G G	C C	G T	A A	A T	T T	G G	C C

Table S4
Summary statistics

						NL									
SNP name	01a	02a	03a	04b	05a	07a	08a	09c	10e	11a	12a	14a	18a	20a	22b
frequency allele A	31,37			17,65	87,91	0,98		22,22	23,86		81,05	43,79			
frequency allele C		26,47		11,11			70,59		51,63	26,80			32,03	45,75	100,00
frequency allele T	6,21	73,53	100,00			43,79	27,78		24,51	7,84	16,67	56,21	67,97	11,11	
frequency allele G	62,42			71,24	12,09	55,23	1,63	77,78		65,36	2,29			43,14	
n alleles	306	306	306	306	306	306	306	306	306	306	306	306	306	306	306
expected heterozygosity*	0,5098	0,3905	0,0000	0,4504	0,2133	0,5048	0,4257	0,3468	0,6184	0,4965	0,3159	0,4939	0,4368	0,5942	0,0000
observed heterozygosity*	0,4967	0,3856	0,0000	0,4706	0,2288	0,4641	0,4314	0,3660	0,5752	0,4902	0,3072	0,5490	0,3529	0,5817	0,0000
PIC-value*	0,4277	0,3135	0,0000	0,4041	0,1900	0,3860	0,3471	0,2859	0,5472	0,4274	0,2776	0,3711	0,3406	0,5046	0,0000
Hardy-Weinberg equilibrium*	0,8188	1,0000	n.a.	0,1586	0,6992	0,2334	0,7885	0,6397	0,6216	0,4770	0,0068	0,1899	0,0249	0,8872	n.a.
HW H1=heterozygote deficit*	0,5259	0,3525	n.a.	0,4453	0,9162	0,1411	0,5644	0,8215	0,1030	0,2333	0,6643	0,9399	0,0146	0,3605	n.a.
HW H1=heterozygote excess*	0,4854	0,6475	n.a.	0,5547	0,0838	0,8589	0,4356	0,3321	0,8989	0,7667	0,3362	0,0601	0,9854	0,6449	n.a.
power of discrimination	0,686	0,552	0,000	0,653	0,363	0,653	0,589	0,511	0,789	0,671	0,477	0,589	0,601	0,749	0,000
power of exclusion	0,185	0,105	0,000	0,163	0,038	0,158	0,134	0,095	0,262	0,179	0,067	0,234	0,088	0,269	0,000

						ANT									
SNP name	01a	02a	03a	04b	05a	07a	08a	09c	10e	11a	12a	14a	18a	20a	22b
frequency allele A	31,08			4,05	58,11	1,35		26,13	43,69		28,38	63,51	8,11		
frequency allele C		21,17		24,77	12,61		36,04		46,40	29,28			39,19	71,62	100,00
frequency allele T	25,68	78,83	100,00			10,36	35,59		9,91	32,88	9,91	36,49	52,70	5,86	
frequency allele G	43,24			71,17	29,28	88,29	28,38	73,87		37,84	61,71			22,52	
n alleles	222	222	222	222	222	222	222	222	222	222	222	222	222	222	222
expected heterozygosity*	0,6534	0,3353	0,0000	0,4324	0,5632	0,2105	0,6660	0,3878	0,5866	0,6660	0,5312	0,4656	0,5646	0,4348	0,0000
observed heterozygosity*	0,6126	0,2793	0,0000	0,4234	0,6126	0,1712	0,7658	0,3604	0,5946	0,6306	0,4505	0,4955	0,4955	0,3874	0,0000
PIC-value*	0,5770	0,2781	0,0000	0,3664	0,4893	0,1926	0,5888	0,3115	0,4938	0,5889	0,4584	0,3561	0,4711	0,3770	0,0000
Hardy-Weinberg equilibrium*	0,4185	0,0897	n.a.	0,2138	0,1022	0,0733	0,0218	0,4665	0,4731	0,5550	0,0659	0,5428	0,3485	0,2332	n.a.
HW H1=heterozygote deficit*	0,1705	0,0714	n.a.	0,5680	0,9284	0,0456	0,9925	0,3023	0,3366	0,2237	0,0145	0,8122	0,0925	0,0766	n.a.
HW H1=heterozygote excess*	0,8310	0,9770	n.a.	0,4395	0,0716	0,9544	0,0076	0,8411	0,6677	0,7783	0,9861	0,1878	0,9113	0,9234	n.a.
power of discrimination	0,807	0,496	0,000	0,604	0,716	0,335	0,780	0,552	0,724	0,817	0,706	0,591	0,736	0,615	0,000
power of exclusion	0,306	0,055	0,000	0,129	0,306	0,022	0,537	0,092	0,284	0,329	0,148	0,184	0,184	0,106	0,000

* *p*-values
p-value < 0.05

Table S5
Genotypic linkage disequilibrium *p*-values

ANT \ NL	01a	02a	03a	04b	05a	07a	08a	09c	10e	11a	12a	14a	18a	20a	22b
01a	x	0.776864	n.a.	0.768094	0.554408	0.449084	0.906388	0.386674	0.666754	0.522774	0.910890	0.686982	0.525956	0.604220	n.a.
02a	0.693724	x	n.a.	0.541810	0.717092	0.584350	0.062954	0.390606	0.434530	0.318558	0.932134	0.516262	0.122974	0.617360	n.a.
03a	n.a.	n.a.	x	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
04b	0.809268	0.055774	n.a.	x	0.582870	0.307594	0.051810	0.102298	0.659548	0.336222	0.468554	0.371268	0.085544	0.018580	n.a.
05a	0.986294	0.640120	n.a.	0.200956	x	0.009324	0.620296	0.539510	0.962066	0.463736	0.177510	0.380892	0.948088	0.970650	n.a.
07a	0.230586	0.911522	n.a.	0.506320	0.202782	x	0.668608	0.874236	0.006062	0.474902	0.333948	0.209784	0.728714	0.473842	n.a.
08a	0.907792	0.761834	n.a.	0.452190	0.416768	0.561336	x	0.034968	0.843162	0.903742	0.999132	0.926062	0.208814	0.640604	n.a.
09c	0.036390	0.800150	n.a.	0.868634	0.012470	0.646698	0.463700	x	0.974860	0.036878	0.324318	0.600836	0.811638	0.959972	n.a.
10e	0.470148	0.419254	n.a.	0.065620	0.945624	0.369076	0.176734	0.378790	x	0.759494	0.870230	0.863910	0.090670	0.442952	n.a.
11a	0.924560	0.878178	n.a.	0.347226	0.427664	0.577676	0.323374	0.509954	0.443796	x	0.237430	0.844922	0.884320	0.880672	n.a.
12a	0.607886	0.740100	n.a.	0.462906	0.287044	0.161766	0.604688	0.071070	0.417078	0.962820	x	0.804282	0.888624	0.565182	n.a.
14a	0.790824	0.853852	n.a.	0.790394	0.163306	0.870160	0.049378	0.415234	0.624794	0.127598	0.688134	x	0.002640	0.991248	n.a.
18a	0.745114	0.542452	n.a.	0.320042	0.804058	0.731918	0.086346	0.037428	0.105320	0.155318	0.680152	0.116036	x	0.212866	n.a.
20a	0.800878	0.873784	n.a.	0.698248	0.029354	0.176908	0.388732	0.984974	0.050990	0.572468	0.370824	0.426766	0.121558	x	n.a.
22b	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	x

p-value < 0.05

Table S6
YCC allele frequencies

European

SNP name	01a	02a	03a	04b	05a	07a	08a	09c	10e	11a	12a	14a	18a	20a	22b
frequency allele A	40,00			20,00	90,00				10,00		90,00	50,00			
frequency allele C		30,00					60,00		80,00	10,00			30,00	70,00	100,00
frequency allele T		70,00	100,00			50,00	30,00		10,00	20,00	10,00	50,00	70,00		
frequency allele G	60,00			80,00	10,00	50,00	10,00	100,00		70,00				30,00	
n alleles	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10

Russian / Siberian

SNP name	01a	02a	03a	04b	05a	07a	08a	09c	10e	11a	12a	14a	18a	20a	22b
frequency allele A	12,50			8,33	83,33	4,17		25,00	16,67		83,33	41,67			
frequency allele C		12,50		25,00			62,50		58,33	33,33			37,50	41,67	100,00
frequency allele T	4,17	87,50	100,00			50,00	37,50		25,00	8,33	4,17	58,33	62,50	8,33	
frequency allele G	83,33			66,67	16,67	45,83		75,00		58,33	12,50			50,00	
n alleles	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24

Asian / Pakistan

SNP name	01a	02a	03a	04b	05a	07a	08a	09c	10e	11a	12a	14a	18a	20a	22b
frequency allele A	12,50			6,25	68,75	18,75		37,50	12,50		43,75	68,75			
frequency allele C		6,25		50,00			50,00		75,00	43,75			62,50	43,75	100,00
frequency allele T	12,50	93,75	100,00			37,50	50,00		12,50	12,50	25,00	31,25	37,50	6,25	
frequency allele G	75,00			43,75	31,25	43,75		62,50		43,75	31,25			50,00	
n alleles	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16

African

SNP name	01a	02a	03a	04b	05a	07a	08a	09c	10e	11a	12a	14a	18a	20a	22b
frequency allele A	16,67				72,22			44,44	55,56		22,22	83,33	22,22		
frequency allele C		27,78		27,78	5,56		11,11		44,44	22,22			27,78	72,22	100,00
frequency allele T	22,22	72,22	100,00				50,00			33,33		16,67	50,00	16,67	
frequency allele G	61,11			72,22	22,22	100,00	38,89	55,56		44,44	77,78			11,11	
n alleles	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18

South African

SNP name	01a	02a	03a	04b	05a	07a	08a	09c	10e	11a	12a	14a	18a	20a	22b
frequency allele A	7,14				50,00			17,86	39,29		17,86	71,43	3,57		
frequency allele C		17,86		25,00	21,43		21,43		60,71	14,29			28,57	78,57	100,00
frequency allele T	32,14	82,14	100,00				60,71			28,57		28,57	67,86	14,29	
frequency allele G	60,71			75,00	28,57	100,00	17,86	82,14		57,14	82,14			7,14	
n alleles	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28

Native American

SNP name	01a	02a	03a	04b	05a	07a	08a	09c	10e	11a	12a	14a	18a	20a	22b
frequency allele A	31,82			4,55	77,27	13,64		59,09	9,09		72,73	86,36			
frequency allele C		9,09		36,36			4,55		86,36	59,09			36,36	22,73	100,00
frequency allele T		90,91	100,00			18,18	95,45		4,55		13,64	13,64	63,64	4,55	
frequency allele G	68,18			59,09	22,73	68,18		40,91		40,91	13,64			72,73	
n alleles	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22

Table S7
Statistical approximation of mixture detection probability by three alleles on at least one locus

If all three allele variants are present in a random sample of four alleles, then one variant must be realized twice and the other two once. The number of distinct permutations of four objects, two of which are indistinguishable from each other, is 4!/(2!1!1!)=12. Hence, the probability of obtaining all three allele variants in a random sample of four alleles equals 12*p1*p2*p3(p1+p2+p3)

For $p_{L,1}$ = estimated allele frequency of allele 1 on locus L
$p_{L,2}$ = estimated allele frequency of allele 2 on locus L
$p_{L,3}$ = estimated allele frequency of allele 3 on locus L

The probability of detecting three alleles on locus L in a two-person mixture = $12*p_{L,1}*p_{L,2}*p_{L,3}(p_{L,1}+p_{L,2}+p_{L,3}) = 12*p_{L,1}*p_{L,2}*p_{L,3}(1) = 12*p_{L,1}*p_{L,2}*p_{L,3}$

The probability of not detecting three alleles (i.e. one or two) on locus L in a two-person mixture =
$1 - 12*p_{L,1}*p_{L,2}*p_{L,3}$

The probability of not detecting three alleles on all loci in a two-person mixture =
$\Pi_L(1 - 12*p_{L,1}*p_{L,2}*p_{L,3})$

The probability of detecting three alleles on at least one of all loci in a two-person mixture =
$1 - \Pi_L(1 - 12*p_{L,1}*p_{L,2}*p_{L,3})$

NL (n alleles = 306)				
L	$p_{L,1}$	$p_{L,2}$	$p_{L,3}$	$1 - 12*p_{L,1}*p_{L,2}*p_{L,3}$
01a	0,3137	0,0621	0,6242	0,8541
04b	0,1765	0,1111	0,7124	0,8324
05a	0,8791	0,1209	0	1
07a	0,0098	0,4379	0,5523	0,9716
08a	0,7059	0,2778	0,0163	0,9616
10e	0,2386	0,5163	0,2451	0,6377
11a	0,2680	0,0784	0,6536	0,8352
12a	0,8105	0,1667	0,0229	0,9629
18a	0,3203	0,6797	0	1
20a	0,4575	0,1111	0,4314	0,7369
$1 - \Pi_L(1 - 12*p_{L,1}*p_{L,2}*p_{L,3})$				0,7490

ANT (n alleles = 222)				
L	$p_{L,1}$	$p_{L,2}$	$p_{L,3}$	$1 - 12*p_{L,1}*p_{L,2}*p_{L,3}$
01a	0,3108	0,2568	0,4324	0,5859
04b	0,0405	0,2477	0,7117	0,9143
05a	0,5811	0,1261	0,2928	0,7425
07a	0,0135	0,1036	0,8829	0,9852
08a	0,3604	0,3559	0,2838	0,5632
10e	0,4369	0,464	0,0991	0,7589
11a	0,2928	0,3288	0,3784	0,5628
12a	0,2838	0,0991	0,6171	0,7917
18a	0,0811	0,3919	0,5270	0,7990
20a	0,7162	0,0586	0,2252	0,8866
$1 - \Pi_L(1 - 12*p_{L,1}*p_{L,2}*p_{L,3})$				0,9471

References

1. M.D. Coble, J.M. Butler, Characterization of new miniSTR loci to aid analysis of degraded DNA, J. Forensic Sci. 50 (2005) 43–53.
2. L.A. Dixon, C.M. Murray, E.J. Archer, A.E. Dobbins, P. Koumi, P. Gill, Validation of a 21-locus autosomal SNP multiplex for forensic identification purposes, Forensic. Sci. Int. 154 (2005) 62–77.
3. P. Grubwieser, R. Muhlmann, B. Berger, H. Niederstatter, M. Pavlic, W. Parson, A new "miniSTR-multiplex" displaying reduced amplicon lengths for the analysis of degraded DNA, Int. J. Legal Med. 120 (2006) 115–120.
4. A.A. Westen, R.R.R. Gerretsen, G.J.R. Maat, Femur, rib, and tooth sample collection for DNA analysis in disaster victim identification, Forensic Sci. Med. Pathol. 4 (2008) 15–21.
5. J.M. Butler, M.D. Coble, P.M. Vallone, STRs vs. SNPs: thoughts on the future of forensic DNA testing, Forensic Sci. Med. Pathol. 3 (2007) 200–205.
6. L.A. Dixon, A.E. Dobbins, H.K. Pulker, J.M. Butler, P.M. Vallone, M.D. Coble, W. Parson, B. Berger, P. Grubwieser, H.S. Mogensen, N. Morling, K. Nielsen, J.J. Sanchez, E. Petkovski, A. Carracedo, P. Sanchez-Diz, E. Ramos-Luis, M. Brion, J.A. Irwin, R.S. Just, O. Loreille, T.J. Parsons, D. Syndercombe-Court, H. Schmitter, B. Stradmann-Bellinghausen, K. Bender, P. Gill, Analysis of artificially degraded DNA using STRs and SNPs — results of a collaborative European (EDNAP) exercise, Forensic Sci. Int. 164 (2006) 33–44.
7. P. Gill, L. Fereday, N. Morling, P.M. Schneider, The evolution of DNA databases — recommendations for new European STR loci, Forensic Sci. Int. 156 (2006) 242–244.
8. J.M. Butler, Y. Shen, B.R. McCord, The development of reduced size STR amplicons as tools for analysis of degraded DNA, J. Forensic Sci. 48 (2003) 1054–1064.
9. C.R. Hill, M.C. Kline, M.D. Coble, J.M. Butler, Characterization of 26 miniSTR loci for improved analysis of degraded DNA samples, J. Forensic Sci. 53 (2008) 73–80.
10. T.J. Parsons, R. Huel, J. Davoren, C. Katzmarzyk, A. Milos, A. Selmanovic, L. Smajlovic, M.D. Coble, A. Rizvic, Application of novel "mini-amplicon" STR multiplexes to high volume casework on degraded skeletal remains, Forensic Sci. Int. Genet. 1 (2007) 175–179.
11. B. Budowle, SNP typing strategies, Forensic Sci. Int. 146 (2004) S139–S142.
12. J.J. Sanchez, C. Phillips, C. Borsting, K. Balogh, M. Bogus, M. Fondevila, C.D. Harrison, E. Musgrave-Brown, A. Salas, D. Syndercombe-Court, P.M. Schneider, A. Carracedo, N. Morling, A multiplex assay with 52 single nucleotide polymorphisms for human identification, Electrophoresis 27 (2006) 1713–1724.
13. B. Budowle, A. van Daal, Forensically relevant SNP classes, Biotechniques 44 (2008) 603–610.
14. M. Fondevila, C. Phillips, N. Naveran, M. Cerezo, A. Rodriguez, R. Calvo, L.M. Fernandez, A. Carracedo, M.V. Lareu, Challenging DNA: assessment of a range of genotyping approaches for highly degraded forensic samples, Forensic Sci. Int. Genet. Supplement Series 1 (2008) 26–28.
15. C. Borsting, J.J. Sanchez, H.E. Hansen, A.J. Hansen, H.Q. Bruun, N. Morling, Performance of the SNPforID 52 SNP-plex assay in paternity testing, Forensic Sci. Int. Genet. 2 (2008) 292–300.
16. C. Phillips, M. Fondevila, M. Garcia-Magarinos, A. Rodriguez, A. Salas, A. Carracedo, M.V. Lareu, Resolving relationship tests that show ambiguous STR results using autosomal SNPs as supplementary markers, Forensic Sci. Int. Genet. 2 (2008) 198–204.
17. K.K. Kidd, A.J. Pakstis, W.C. Speed, E.L. Grigorenko, S.L. Kajuna, N.J. Karoma, S. Kungulilo, J.J. Kim, R.B. Lu, A. Odunsi, F. Okonofua, J. Parnas, L.O. Schulz, O.V. Zhukova, J.R. Kidd, Developing a SNP panel for forensic identification of individuals, Forensic Sci. Int. 164 (2006) 20–32.

18. E. Musgrave-Brown, D. Ballard, K. Balogh, K. Bender, B. Berger, M. Bogus, C. Borsting, M. Brion, M. Fondevila, C. Harrison, C. Oguzturun, W. Parson, C. Phillips, C. Proff, E. Ramos-Luis, J.J. Sanchez, P. Sanchez Diz, B. Sobrino Rey, B. Stradmann-Bellinghausen, C. Thacker, A. Carracedo, N. Morling, R. Scheithauer, P.M. Schneider, D. Syndercombe-Court, Forensic validation of the SNPforID 52-plex assay, Forensic Sci. Int. Genet. 1 (2007) 186–190.
19. J.J. Sanchez, C. Borsting, K. Balogh, B. Berger, M. Bogus, J.M. Butler, A. Carracedo, D. Syndercombe-Court, L.A. Dixon, B. Filipovic, M. Fondevila, P. Gill, C. Harrison, C. Hohoff, R. Huel, B. Ludes, W. Parson, T.J. Parsons, E. Petkovski, C. Phillips, H. Schmitter, P.M. Schneider, P.M. Vallone, N. Morling, Forensic typing of autosomal SNPs with a 29 SNP-multiplex — Results of a collaborative EDNAP exercise, Forensic Sci. Int. Genet. 2 (2008) 176–183.
20. C. Phillips, V. Lareu, A. Salas, A. Carracedo, Nonbinary single-nucleotide polymorphism markers, Int. Congress Ser. 1261 (2004) 27–29.
21. R. Chakraborty, D.N. Stivers, B. Su, Y. Zhong, B. Budowle, The utility of short tandem repeat loci beyond human identification: implications for development of new DNA typing systems, Electrophoresis 20 (1999) 1682–1696.
22. P. Gill, An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes, Int. J. Legal Med. 114 (2001) 204–210.
23. A. Amorim, L. Pereira, Pros and cons in the use of SNPs in forensic kinship investigation: a comparative analysis with STRs, Forensic Sci. Int. 150 (2005) 17–21.
24. The Y Chromosome Consortium, A nomenclature system for the tree of human Y-chromosomal binary haplogroups, Genome Res. 12 (2002) 339–348.
25. S. Rozen, H.J. Skaletsky, Primer3 on the WWW for general users and for biologist programmers, in: S. Krawetz, S. Misener (Eds.), Bioinformatics Methods and Protocols: Methods in Molecular Biology, Humana Press, Totowa, NJ, 2000, pp. 365–386.
26. P.M. Vallone, J.M. Butler, AutoDimer: a screening tool for primer-dimer and hairpin structures, Biotechniques 37 (2004) 226–231.
27. B. Sobrino, M. Brion, A. Carracedo, SNPs in forensic genetics: a review on SNP typing methodologies, Forensic Sci. Int. 154 (2005) 181–194.
28. S.D.E. Park, Trypanotolerance in West African Cattle and the Population Genetic Effects of Selection, University of Dublin, 2001.
29. F. Rousset, GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux, Mol. Ecol. Resour. 8 (2008) 103–106.
30. Promega, http://www.promega.com/geneticidtools/powerstats/, 2000.
31. C. Phillips, R. Fang, D. Ballard, M. Fondevila, C. Harrison, F. Hyland, E. Musgrave-Brown, C. Proff, E. Ramos-Luis, B. Sobrino, A. Carracedo, M.R. Furtado, D. Syndercombe-Court, P.M. Schneider, Evaluation of the Genplex SNP typing system and a 49plex forensic marker panel, Forensic Sci. Int. Genet. 1 (2007) 180–185.
32. C. Phillips, A. Salas, J.J. Sanchez, M. Fondevila, A. Gomez-Tato, J. Alvarez-Dios, M. Calaza, M. Casares de Cal, D. Ballard, M.V. Lareu, A. Carracedo, Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs, Forensic Sci. Int. Genet. 1 (2007) 273–280.
33. C. Harrison, E. Musgrave-Brown, K. Bender, A. Carracedo, N. Morling, P. Schneider, D. Syndercombe-Court, A sensitive issue: pyrosequencing as a valuable forensic SNP typing platform, Int. Congress Ser. 1288 (2006) 52–54.
34. P. Oeth, M. Beaulieu, C. Park, D. Kosman, G. del Mistro, D. van den Boom, C. Jurinke, iPLEX Assay: increased plexing efficiency and flexibility for MassARRAY system through single base primer extension with mass-modified terminators, Sequenom Application Note (2006).
35. S. Paabo, H. Poinar, D. Serre, V. Jaenicke-Despres, J. Hebler, N. Rohland, M. Kuch, J. Krause, L. Vigilant, M. Hofreiter, Genetic analyses from ancient DNA, Annu. Rev. Genet. 38 (2004) 645–679.