

# Evaluation of parameters in mixed male DNA profiles for the Identifiler<sup>®</sup> multiplex system

NA HU<sup>1</sup>, BIN CONG<sup>1</sup>, TAO GAO<sup>2</sup>, RONG HU<sup>1</sup>, YI CHEN<sup>2</sup>, HUI TANG<sup>3</sup>,  
LUYAN XUE<sup>3</sup>, SHUJIN LI<sup>1</sup> and CHUNLING MA<sup>1</sup>

<sup>1</sup>Hebei Key Laboratory of Forensic Medicine, Department of Forensic Medicine, Hebei Medical University, Shijiazhuang, Hebei 050017; <sup>2</sup>Institute of Statistics, Renmin University of China, Beijing 100872; <sup>3</sup>DNA Laboratory, Forensic Science Service of Beijing Public Security Bureau, Beijing 100085, P.R. China

Received January 25, 2014; Accepted May 7, 2014

DOI: 10.3892/ijmm.2014.1779

**Abstract.** The analysis of complex DNA mixtures is challenging for forensic DNA testing. Accurate and sensitive methods for profiling these samples are urgently required. In this study, we developed 11 groups of mixed male DNA samples (n=297) with scientific validation of D-value [ $>95\%$  of D-values  $\leq 0.1$  with average peak height (APH) of the active alleles  $\leq 2,500$  rfu]. A strong linear correlation was detected between the peak height (PH) and peak area (PA) in the curve fit using the least squares method ( $P < 2e-16$ ). The Kruskal-Wallis rank-sum test revealed significant differences in the heterozygote balance ratio ( $H_b$ ) at 16 short tandem repeat (STR) loci ( $P=0.0063$ ) and 9 mixed gradients ( $P=0.02257$ ). Locally weighted regression fitting of APH and  $H_b$  (inflection point at APH = 1,250 rfu) showed 92.74% of  $H_b > 0.6$  with the APH  $\geq 1,250$ . The variation of  $H_b$  distribution in the different STR loci suggested the different forensic efficiencies of these loci. Allelic drop-out (ADO) correlated with the APH and mixed gradient. All ADOs had an APH of  $< 1,000$  rfu, and the number of ADO increased when the APH of mixed DNA profiles gradually decreased. These results strongly suggest that calibration parameters should be introduced to correct the deviation in the APH at each STR locus during the analysis of mixed DNA samples.

## Introduction

In forensic analysis, mixed DNA samples are composed of genetic material from more than one donor, and complex DNA

mixtures involve 3 or more individuals (1,2). During criminal investigations, specimens of blood, semen, secreted fluids, excretions and epithelial cell samples are often mixed, particularly in cases of vaginal rape, anal rape and oral sodomy. Complex mixtures of DNA, such as mixed semen collected in cases of gang rape or mixed blood samples in homicide cases, are the most challenging to analyze (3-7). The contemporary analysis of mixed DNA samples often yields low detection rates that are not useful in criminal investigations. Some results collected using these methods do not meet the legal standards of relevant court systems (2). Therefore, accurate and sensitive methods for profiling mixed DNA samples are urgently required in forensic DNA testing.

Mixed DNA evidence from semen and vaginal secretions are commonly submitted for laboratory analysis in sexual assault cases. The DNA Commission of the International Society of Forensic Genetics (ISFG) published standardized unrestricted and restricted combinatorial methods for interpreting mixed male-female DNA profiles in 2006 (2). In general, unrestricted combinatorial approaches involve probability analyses of allelic bands on a short tandem repeat (STR) locus based on the separation of allele peaks by size, discounting stochastic effects that lead to the substantial imbalance of two alleles at a given heterozygous locus (2). As a result, quantitative information, such as peak height (PH) and peak area (PA) are not considered in the calculation of likelihood ratios (LRs), which limits the accuracy of these methods.

In order to achieve improved accuracy, a number of studies have reported the use of restricted combinatorial methods that require quantitative parameters (1,2,8-16), such as the heterozygote balance ratio ( $H_b$ ) and mixture proportion ( $M_x$ ) (2) in the comparison of victim DNA profiles with filtered genotype combinations. In practice, however, the restricted combinatorial methods are complicated by the presence of stutter peaks (8,11,17), drop-outs and drop-ins (2,18-25), and low-template DNA (LT-DNA) (8,11,14,24-35). Preferential amplification may occur in low-molecular-weight (LMW) alleles during PCR when  $H_b > 1$  (2), resulting in lower accuracy. The mixed DNA profiles are further complicated by the effects of stutter-affected heterozygotes (SAH) and allele masking, which leads to considerably more complex  $H_b$  regularity. Mixed DNA samples involving very small or degenerated

---

*Correspondence to:* Dr Bin Cong, Hebei Key Laboratory of Forensic Medicine, Department of Forensic Medicine, Hebei Medical University, Zhongshan East Road No. 361, Shijiazhuang, Hebei 050017, P.R. China  
E-mail: hbybincong@126.com

**Key words:** forensic genetics, short tandem repeats, male-male mixed DNA, average peak height/area of the active alleles, heterozygote balance ratio, mixture proportion, allelic drop-out

samples are particularly prone to distortion by random effects associated with the mutual inhibition of PCR invariant DNA components, distortion of the peak balance in the PCR system, operational issues and equipment errors (10,36,37). These factors significantly increase the difficulty of scientific analysis and the interpretation of accurate results obtained from mixed DNA profiles.

For mixed male-female DNA, forensic sex-typing is generally conducted with commercial STR kits that apply the primers suggested in the study by Sullivan *et al* and the sex-typing marker, amelogenin (AMEL-) (38). Results produce characteristic male X and Y chromosome peaks that are easily distinguished from the single female X chromosome peak, although anomalous results have been reported due to abnormalities, such as primer binding-site mutations and chromosomal deletions (39). In this DNA profile, the intensity of the X peak is much higher than that of the Y peak, indicating heterozygote imbalance (i.e.,  $H_b < 0.6$ ) (2). A number of studies have applied this methodology to identify the  $M_x$  of male-female individuals (8-11,14,37). Notably, complex mixtures involving mixed male DNA, such as mixed semen, all present 2 X/Y peaks, causing heterozygote balance ( $H_b > 0.6$ ) that prohibits the estimation of  $M_x$  between DNA components using the AMEL locus. Therefore, the establishment of a more reliable and accurate method for the interpretation of mixed male DNA requires careful parameter selection and evaluation.

In the current study, the data distribution and statistical analysis of each parameter were carried out specifically for mixed male DNA. Using this experimental model, a relative fluorescence intensity (unit: rfu) range that provides the optimal distribution of D-value,  $H_b$  and allelic drop-out (ADO) according to average PHs (APHs) was determined. Mixed male DNA profiles should first meet these 'rfu' levels to ensure the accuracy and reliability interpretation. It should then be determined whether calibration parameters are required to correct the APH deviation during mixed DNA analysis.

## Materials and methods

**DNA sample collection.** Forty anti-coagulated blood samples collected from unrelated healthy males (5 ml each) were supplied by the Blood Center of Hebei Province, Shijiazhuang, China.

**Experimental design.** For our study purposes, DNA was extracted from the 40 whole blood samples, and quantified by the Applied Biosystems (Foster City, CA, USA) 7500 real-time PCR system (Life Technologies Inc., Carlsbad, CA, USA). Single DNA samples with concentration differences of  $< 0.5$  ng/ $\mu$ l were paired to construct a simulated mixed male DNA experimental model of 2 individuals. Gradient ratios between 2 individual DNA samples were achieved by adjusting the dilution and volume of each sample that was added to the model. In addition, the concentration of the simulated mixed DNA stock solutions were all adjusted to the desired levels within the working solution concentration range of 0.5-1.25 ng/ $\mu$ l, so as to meet the requirements for the DNA template of the testing kit.

## Establishment of mixed male DNA model

**DNA extraction.** Genomic DNA was extracted from the 40 male whole blood samples using the Invitrogen® PureLink™ Genomic DNA Mini kit (Life Technologies Inc.). A total of 20  $\mu$ l of each DNA stock solution (numbered 1-40) was diluted with nuclease-free water (Ambion, Carlsbad, CA, USA) to a final volume of 200  $\mu$ l. The stock solutions of Promega® 9948 Male DNA and 2800M Control DNA standards (Promega Corp., Madison, WI, USA; concentration of 10 ng/ $\mu$ l; volume, 25  $\mu$ l) were diluted with Ambion® nuclease-free water, to a final volume of 250  $\mu$ l.

**DNA quantification.** DNA quantification was performed using the Quantifiler® Human DNA Quantification kit (Life Technologies Inc.) containing DNA standard solution (200 ng/ $\mu$ l), Quantifiler Human Primer mix, and Quantifiler PCR Reaction Mix. Human Primer Mix (10.5  $\mu$ l/sample) and PCR Reaction Mix (12.5  $\mu$ l/sample) were mixed and then dispensed into reaction wells (23  $\mu$ l each) followed by the addition of 2  $\mu$ l of sample or standard to each well, to obtain a 25- $\mu$ l PCR reaction system. DNA quantification was repeated 3 times for each sample, and the mean result served as the final DNA concentration.

**Identifiler® PCR and electrophoresis.** With 25  $\mu$ l PCR system (containing 10.5  $\mu$ l of PCR reaction mix, 5.5  $\mu$ l of Identifiler® Primer Set, 0.5  $\mu$ l of Gold® DNA polymerase, 9.0  $\mu$ l of nuclease-free water and 1  $\mu$ l of DNA template), Identifiler® PCR amplification was performed as follows: pre-denaturation at 95°C for 11 min, followed by 28 cycles of denaturation at 94°C for 1 min, annealing at 59°C for 1 min and extension at 72°C for 1 min, and a final extension step at 60°C for 60 min. The PCR products were then examined using a 10- $\mu$ l electrophoresis system containing 0.25  $\mu$ l of GeneScan™ 500 LIZ® Size Standard, 9.25  $\mu$ l of Hi-Di™ formamide and 0.5  $\mu$ l of PCR product or the AmpFISTR® Identifiler® allelic ladder. Capillary electrophoresis was performed on an ABI3130xl Genetic Analyzer (Applied Biosystems).

## Parameters of mixed DNA profiles

**APH/average PA (APA) of the active alleles.** In a mixed DNA profile obtained by STR analysis, the height of the y-axis corresponding to the band of an allele is termed as the PH. The area surrounded by the x-axis and the peak outline is termed as the PA. Both allele PH and PA are expressed as relative fluorescence intensity (unit: rfu). APH or average PA (APA) is defined as the mean of PHs or PAs from all loci (excluding drop-out) in a DNA profile. The parameters, such as  $M_x$ ,  $H_b$  and stutter ratio fluctuate with alterations in APH or APA. APH or APA often serves as a quantitative parameter to evaluate the distribution regularity in mixed DNA analysis.

**Analysis of  $M_x$ .**  $M_x$  is the proportion of a single DNA component in mixed DNA. The original proportion (theoretical  $M_x$ ) is usually not unknown in an undefined sample. Instead, PHs or PAs of the allelic bands measured in a STR locus are used for the calculation ( $M_x^l$ ). The mean  $M_x$  is calculated as the average of the  $M_x^l$  from all loci.

**Analysis of D-value.** The D-value indicates the error between the  $M_x$  values calculated using the PH or PA of a locus ( $M_x^l$ ) in a mixed-DNA profile and the mean  $M_x$  value of all loci in the mixed DNA profile, which is calculated using the following formula:  $D = |M_x^l - \bar{M}_x|$

Table I. DNA concentration of the 11 groups of male-male mixed samples.

Sample no.	No. 1	Concentration (ng/ $\mu$ l)	No. 2	Concentration (ng/ $\mu$ l)	Difference (ng/ $\mu$ l)
NAN1	1	5.80	3	5.75	0.05
NAN2	11	7.20	40	7.21	0.01
NAN3	11	7.20	14	7.22	0.02
NAN4	20	9.24	26	9.35	0.11
NAN5	12	6.58	19	6.68	0.10
NAN6	24	5.46	27	5.48	0.02
NAN7	4	6.92	15	6.98	0.06
AN8	7	6.13	29	6.22	0.09
NAN9	11	7.20	38	7.14	0.06
NAN10	37	6.40	39	6.39	0.01
AN11	9948	0.223	2800M	0.243	0.02

Table II. Mixed gradient of simulated male-male mixed DNA.

Two-male mixed DNA	Person	Mixed gradient								
		1:1	1:2	1:3	1:4	1:5	1:6	1:7	1:8	1:9
Volume ( $\mu$ l)	No. 1	5	4	3	2	2	2	2	2	2
	No. 2	5	8	9	8	10	12	14	16	18

For the defined mixed DNA experimental model, the mean  $M_x$  can be replaced by the theoretical  $M_x$  (hybrid gradient designed in the model), and the following formula is used to estimate the D-value:  $D = |M_x^t - M_x|$

The D-value reflects the difference between the actual and theoretical  $M_x$  value of a locus in a mixed DNA profile. A low D-value indicates an accurate quantitative result of a mixed DNA experimental model.

**Analysis of  $H_b$ .** The  $H_b$  is calculated as the ratio between PHs (measured as fluorescent intensities  $\phi$ , units: rfu) of the lower peak and higher peak in the same locus,  $H_b = \phi_a/\phi_b$ . In the case that all DNA template amounts are >500 pg and not degraded,  $H_b > 0.6$ , defined as heterozygote balance, usually indicates that both alleles originate from one the heterozygote of one individual.  $H_b < 0.6$  is defined as heterozygote imbalance, indicating that the alleles are from different individuals.

Given that the  $H_b$  cannot be estimated in the locus with ADO, the  $H_b$  is only estimated in loci without drop-out and allele sharing. The  $H_b$  of the mixed DNA profile is only calculated with genotypes of AB:CD or AB:CC without allele sharing. Two  $H_b$  values ( $H_b$  of AB and CD) for type AB:CD, and one  $H_b$  value ( $H_b$  of AB) for type AB:CC, as well as the corresponding numbers of APH values can be calculated.

**Analysis of ADO.** The low level of a specific DNA content may cause relative fluorescence intensity which is too low, and which cannot be separated from the background, and therefore results in the loss of an allelic peak, presenting a false homozygote. ADO can be of a single allele, two alleles, and low-copy-number mixed DNA allele.

ADO is primarily caused by a very low PH value. The total number of drop-out alleles in a DNA profile correlates with the APH of a DNA sample. The APH is defined as the mean APH of loci without drop-out in a DNA profile.

**Statistical analysis.** Locally weighted polynomial regression is a non-parametric regression method. It does not require the hypothesis of data distribution. Instead it describes the relationship of variables according to the morphology of the data. This method is more robust than the conventional least squares regression model.

Kernel density estimation is a non-parametric method to estimate a density curve driven by data distribution. The Kruskal-Wallis rank sum test is also a non-parametric test that does not depend on normal distribution of data. These provide more reliable results when analyzing non-normal distributed data compared to variance analysis.

All graphics were made using the R (version 3.0.1) software package ggplot2 (version 0.9.3).

## Results

**Establishment of simulated mixed male DNA experimental model.** Based on the difference in DNA concentration ( $\leq 0.5$  ng/ $\mu$ l), 40 male DNA samples (No. 1-40) and the Promega male DNA standard were paired. A total of 22 single eligible DNA were prepared into 11 groups of 2 mixed male DNA samples. The Promega 2800M male DNA standard working sample was further diluted to a final concentration of 0.243 ng/ $\mu$ l for pairing. The DNA concentrations of the 11 groups are presented in Table I. A total of 9 hybrid gradients were designed in each group of mixed male DNA samples, in triplicate for each gradient, resulting in a total of 297 samples (Table II).

**DNA quantity of mixed male DNA.** The DNA quantity of the 297 simulated mixed male DNA samples was examined by assessing the selected samples using the ABI 7500 real-time PCR system. The quantification of each sample was repeated 3 times, and the mean values were taken as the DNA concentration (Table III).

Given that the concentration of the DNA template recommended by the Identifiler kit used in this study was 0.5-1.25 ng/ $\mu$ l, 99 mixed male DNA working solutions (11 groups of mixed DNA at 9 hybrid gradients) were diluted accordingly. Each 2- $\mu$ l aliquot was diluted by 10- or 15-fold with nuclease-free water (Ambion). The 9948 and 2800M DNA standards with a concentration of <0.5 ng/ $\mu$ l were left undiluted. Additionally, the volume of DNA template for the NAN11 mixed DNA samples (n=27) corresponding to the male-male DNA standards was 2 and 1  $\mu$ l for the other groups, including single DNA samples used for mixed male DNA PCR system (Table IV).

**D-value analysis of the experimental model.** Fig. 1A illustrates the correlation between the APH and D-value of each locus in the mixed DNA profiles estimated using the 2 formulas:  $D = |M_x^t - \bar{M}_x|$  and  $D = |M_x^t - M_x|$ . Of the total D-values estimated, 99.88% of the D-values estimated using the first formula were  $\leq 0.2$ , and 98.25% of them were  $\leq 0.1$ . While 99.88% of the D-values estimated using the 2nd

Table III. DNA quantity of the male-male mixed samples at gradient 1:9.

Sample no.	Mixed DNA quantity (ng/ $\mu$ l)	No. 1	Concentration (ng/ $\mu$ l)	No. 2	Concentration (ng/ $\mu$ l)	Difference (ng/ $\mu$ l)
NAN1	8.32	1	5.80	3	5.75	2.57
NAN2	10.94	11	7.20	40	7.21	3.74
NAN3	10.51	11	7.20	14	7.22	3.31
NAN4	13.38	20	9.24	26	9.35	4.14
NAN5	9.60	12	6.58	19	6.68	3.02
NAN6	7.01	24	5.46	27	5.48	1.55
NAN7	9.62	4	6.92	15	6.98	2.70
NAN8	8.68	7	6.13	29	6.22	2.55
NAN9	10.38	11	7.20	38	7.14	3.24
NAN10	9.31	37	6.40	39	6.39	2.92
NAN11	0.337	9948	0.223	2800M	0.243	0.114

Table IV. Dilution of male-male mixed DNA working solutions.<sup>a</sup>

Sample no.	DNA quantity (ng/ $\mu$ l)	Dilution factor	PCR template ( $\mu$ l)
NAN1	8.32	10-fold	1
NAN2	10.94	10-fold	1
NAN3	10.51	10-fold	1
NAN4	13.38	15-fold	1
NAN5	9.60	10-fold	1
NAN6	7.01	10-fold	1
NAN7	9.62	10-fold	1
NAN8	8.68	10-fold	1
NAN9	10.38	10-fold	1
NAN10	9.31	10-fold	1
NAN11	0.337	1-fold	2

<sup>a</sup>Desired concentration of DNA template, 0.5-1.25 ng/ $\mu$ l.

formula were  $\leq 0.2$ , 95.7% of them were  $\leq 0.1$ . Both methods achieved 95% D  $\leq 0.1$ , indicating that the error between the measured and theoretical  $M_x$  value of each locus in the mixed DNA profile was  $\leq 10\%$ . PCR amplification did not result in great alterations in the  $M_x$  value, indicating the reliability of PCR-based mixed DNA analysis.

The correlation between the APH and D-value of the mixed DNA samples is presented in Fig. 1B. A similar tendency in D-value distribution was observed in the 9 mixed gradients. A D-value  $> 0.2$  was found only in the gradients of 1:2, 1:5, 1:6 and 1:9, while most D-values were  $\leq 0.1$  with an APH  $\leq 2,500$  rfu. These results demonstrate a minor error between the measured and theoretical  $M_x$  value of each locus, suggesting that the mixed male DNA experimental models meet the requirements of scientific and rational mixed DNA analysis.

**Correlation between PH and PA.** The correlation between PH and PA in the mixed DNA profile is shown in Fig. 2A.

The point of inflection in the curve (blue line) fitted with the generalized additive model was present at PH of 6,250 rfu, PH increase accelerated at  $\geq 6,250$  rfu. The following regression equation was obtained from the curve (red line) fitted with the least squares method:

$$\text{Area} = \begin{matrix} 10.12 \text{ height} & 439.48 \\ (0.01825) & + (39,04147) \\ (<2e-16^{***}) & (<2e-16^{***}) \end{matrix}$$

For which  $R^2 = 0.9588$  and the P-value was  $< 2e-16$ , indicating a strong linear correlation between the PH and PA.

For the 16 STR loci analyzed (Fig. 2B), loci D19S433, D3S1358, D5S818 and D8S1179 presented a relatively weaker linear correlation between PH and PA. The other 12 STR loci showed strong linear correlation between PH and PA. These results are basically consistent with the conclusion drawn in the study by Tvedebrink *et al.* (40) that there is a strong linear correlation between PH and PA. Generally, for all 16 STR loci, linear correlations between PH and PA were detected. Therefore both PH and PA may be used for the quantitative analysis of mixed DNA samples, without apparent differences in efficacy.

**Correlation between APH and  $H_b$ .** After the locus with drop-out was excluded, 2,535  $H_b$  values and the corresponding APH values were calculated in the 297 mixed male DNA profiles. Both the  $H_b$  and APH showed skewed distribution (Fig. 3).

Tables V and VI show the percentages of  $H_b > 0.6$  and  $> 0.7$  at 16 STR loci in the 9 mixed gradients. Fig. 4 illustrates the data distribution of  $H_b$  at each locus and mixed gradient, in which the red dotted lines indicate  $H_b = 0.7$  and  $= 0.9$ . Allele sharing at loci D3S1358 and AMEL resulted in the lack of  $H_b$  estimation; therefore, the statistical analysis of the  $H_b$  values was only carried out for 14 STR loci. It was found that the median  $H_b$  was higher at loci TPOX, TH01 and D21S11, with most  $H_b \geq 0.9$ . The median  $H_b$  was lowest at locus D5S818.  $H_b$  distribution fluctuated greatly at loci D5S818 and D2S1338. There was little difference in the median  $H_b$  distribution for

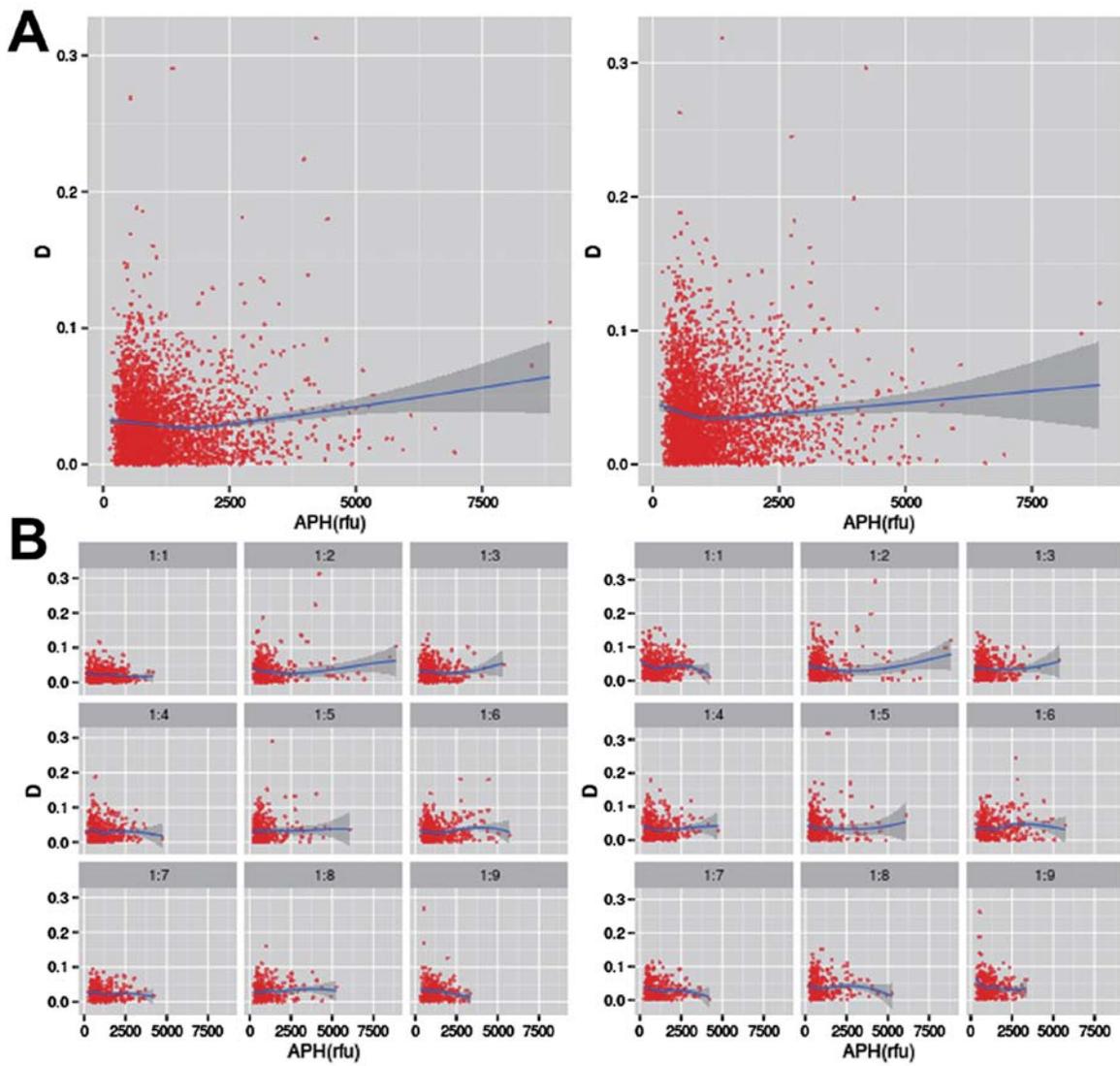


Figure 1. (A) Correlation between the average peak height (APH) of the active alleles and D-value; and (B) separated analysis for the 9 mixed gradients. Left panel,  $D = |M'_x - \bar{M}_x|$ ; right panel,  $D = |M'_x - M_x|$ .

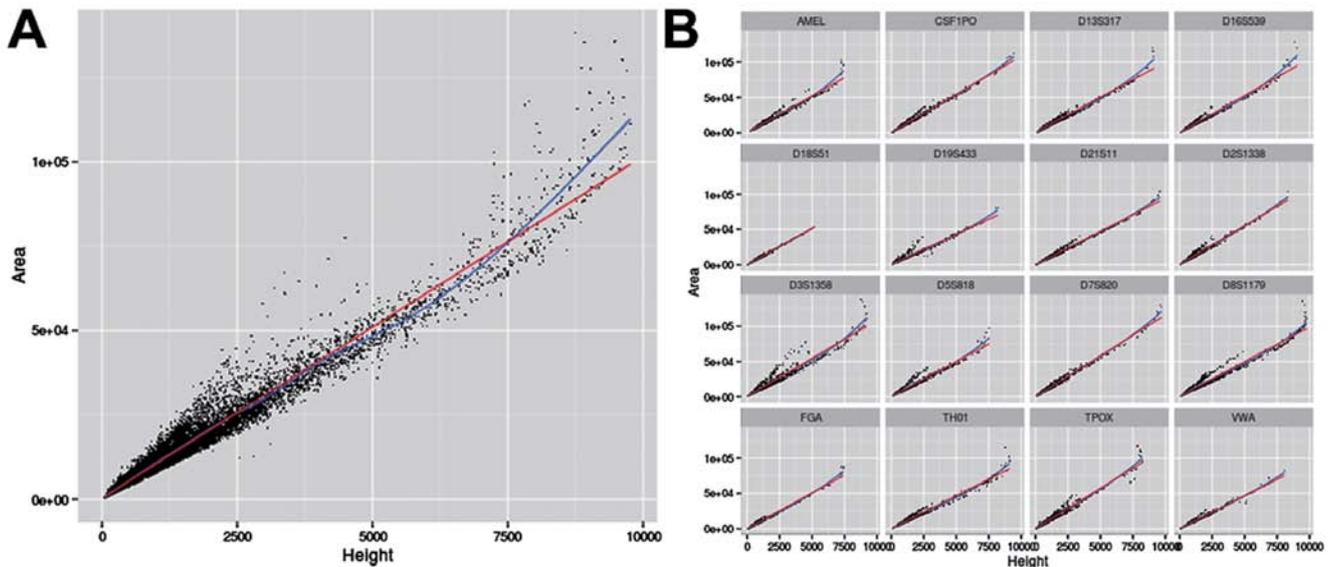


Figure 2. (A) Linear correlation between peak height and peak area and (B) separate analysis for the 16 short tandem repeat (STR) loci. The blue line indicates the curve fitted with the generalized additive model, and the red line indicates the curve fitted with the least squares method.

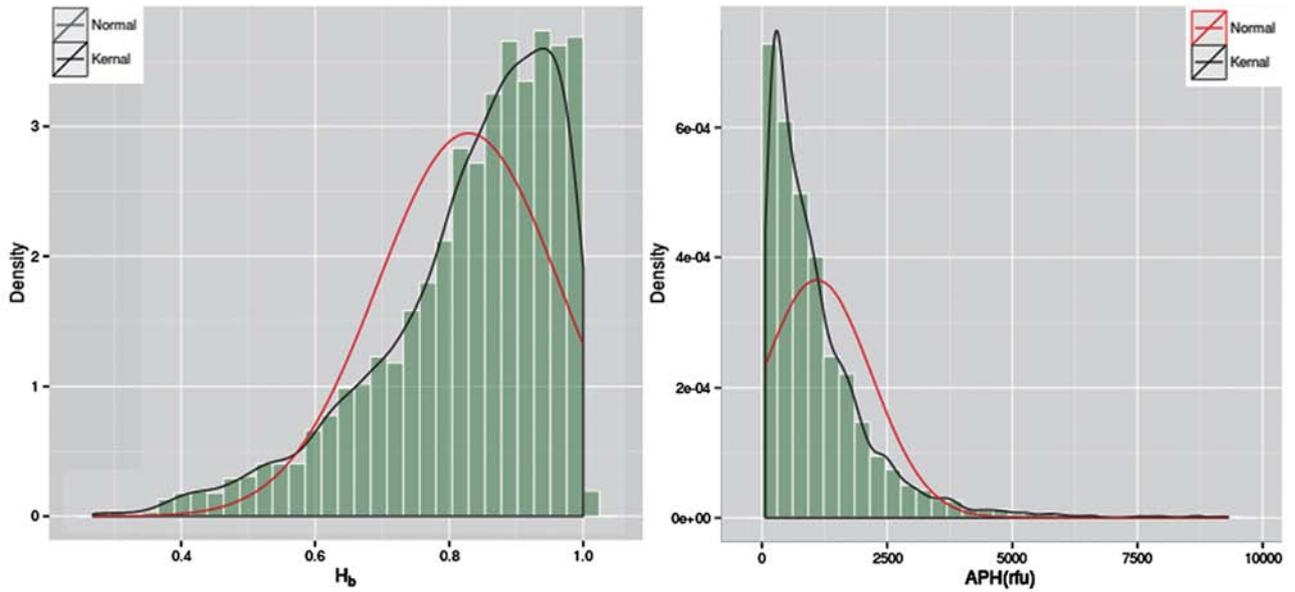


Figure 3. Histogram of the heterozygote balance ratio ( $H_b$ ) and corresponding average peak height (APH) at 16 short tandem repeat (STR) loci. The black lines indicate the curves fitted using the kernel density estimation, and the red lines indicate the curves fitted with normal distribution.

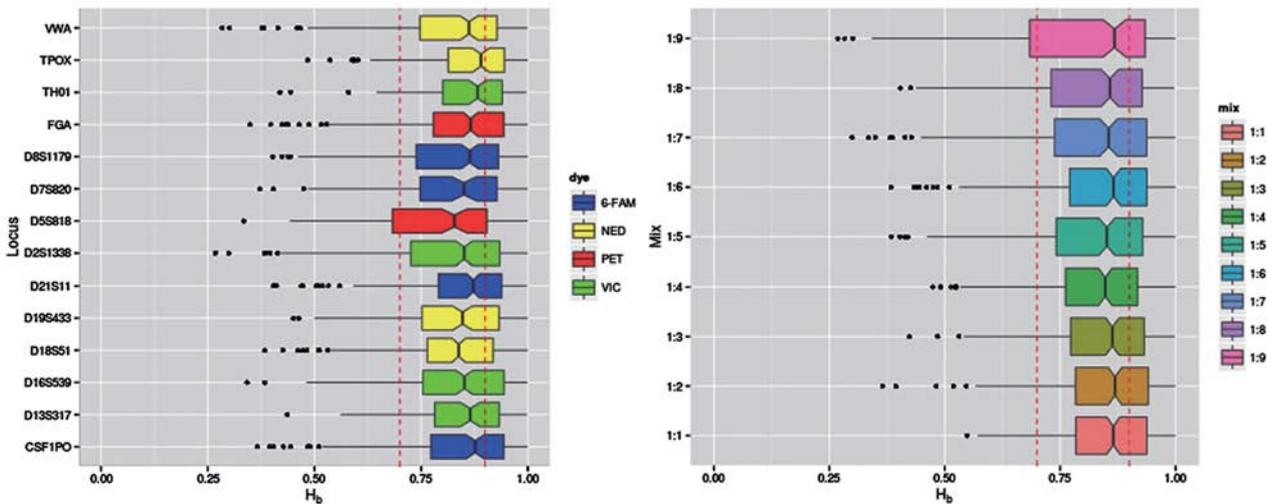


Figure 4. Distribution of the histogram of the heterozygote balance ratio ( $H_b$ ) at 16 short tandem repeat (STR) loci (left panel) and at 9 mixed gradients (right panel). The 4 fluorescence-labeled loci are differentiated by using 4 different colors; the red dotted lines indicate  $H_b = 0.7$  and  $0.9$ ; the groove interval in the right box-whisker plot indicates the 95% confidence interval of the median.

various mixed gradients (apart from gradient 1:8 and 1:9). The Kruskal-Wallis rank sum test revealed significant differences in the distribution of  $H_b$  at 16 STR loci ( $P=0.0063$ ) and at 9 mixed gradients ( $P=0.02257$ ).

Fig. 5A presents the correlation between the 2,535  $H_b$  values and the corresponding APH in the mixed male DNA profiles. The blue solid line was plotted by using the locally weighted regression, while the grey region indicates the corresponding confidential interval. An inflection point was presented at APH = 1,250 rfus in the curve fitted by non-parametric regression. When the APH was <1,250 rfus, the  $H_b$  value varied between 0.75 and 0.87. In cases when the APH was  $\geq 1,250$  rfus, the  $H_b$  value was almost stable. The green dotted line indicates the mean  $H_b$  value of 0.878 corresponding to APH >1,250 rfus. It was found that 92.66% of the  $H_b$  values were >0.6 and 83.04%

of the  $H_b$  values were >0.7. When the APH was  $\geq 1,250$  rfus, 92.74% of the  $H_b$  values were >0.6.

Fig. 5B and C show the changing tendency of the APH and  $H_b$  at the 9 mixed gradients and 16 loci fitted by the locally weighted regression, in which the 2 red dotted lines indicate  $H_b = 0.6$  and  $0.7$ . Fig. 5B illustrates that >90% of the  $H_b$  values were >0.6 at gradients from 1:1 to 1:8, while 82% of the  $H_b$  values were >0.6 at gradient 1:9. The percentage of high  $H_b$  value and high APH was greater at mixed gradients of 1:2, 1:6, 1:7 and 1:8 than the other gradients. Fig. 5C illustrates that >90% of the  $H_b$  values were >0.6 at loci apart from D2S1338 and D5S818; moreover, the percentage of high  $H_b$  value and high APH was greater at loci CSF1PO, D19S433, D21S11, D2S1338 and vWA than the other loci, while APH was almost concentrated at <2,500 rfus at other loci.

Table V. Percentages of  $H_b > 0.6$  and  $> 0.7$  at 14 STR loci.

Locus	CSF1PO	D13S317	D16S539	D18S51	D19S433	D21S11	D2S1338
$H_b > 0.6$	0.91	0.97	0.94	0.92	0.96	0.95	0.88
$H_b > 0.7$	0.82	0.87	0.86	0.84	0.83	0.86	0.77
Locus	D5S818	D7S820	D8S1179	FGA	TH01	TPOX	vWA
$H_b > 0.6$	0.88	0.91	0.95	0.93	0.96	0.94	0.90
$H_b > 0.7$	0.73	0.80	0.82	0.85	0.89	0.89	0.83

$H_b$ , heterozygote balance ratio; STR, short tandem repeat.

Table VI. Percentages of  $H_b > 0.6$  and  $> 0.7$  at the 9 mixed gradients.

	Mixed gradient								
	1:1	1:2	1:3	1:4	1:5	1:6	1:7	1:8	1:9
$H_b > 0.6$	0.98	0.97	0.97	0.96	0.92	0.91	0.90	0.89	0.82
$H_b > 0.7$	0.93	0.89	0.87	0.84	0.80	0.82	0.80	0.77	0.73

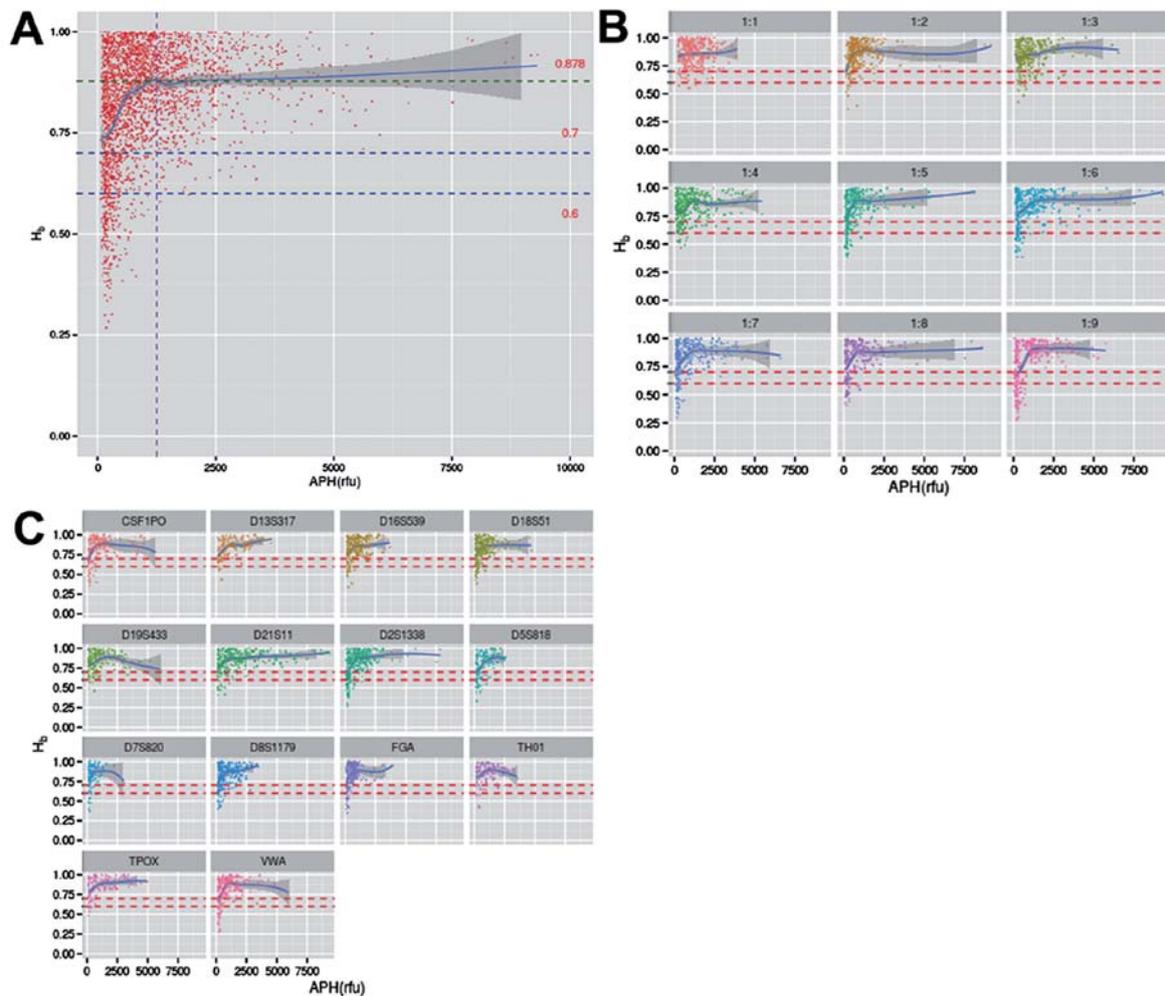


Figure 5. (A) Correlation between average peak height (APH) of the active alleles and 2,535 histogram of heterozygote balance ratio ( $H_b$ ) values in male-male mixed DNA profiles and (B) separated analysis of the 9 mixed gradients or (C) for the 16 short tandem repeat (STR) loci. (A) The horizontal blue dotted lines indicate  $H_b = 0.6$  and  $0.7$ ; the vertical blue dotted line indicates  $APH = 1,250$  rfu; the horizontal green dotted line indicates the mean  $H_b$  of  $0.878$  corresponding to  $APH > 1,250$  rfu; (B and C) The red dotted lines indicate  $H_b = 0.6$  and  $0.7$ .

Table VII. Number of drop-out alleles in the 297 male-male mixed DNA profiles.

No. of drop-out alleles	0	1	2	3	4	5	6	7	8	9	10	12	13
No. of samples	245	19	8	5	1	2	2	3	3	3	2	2	2

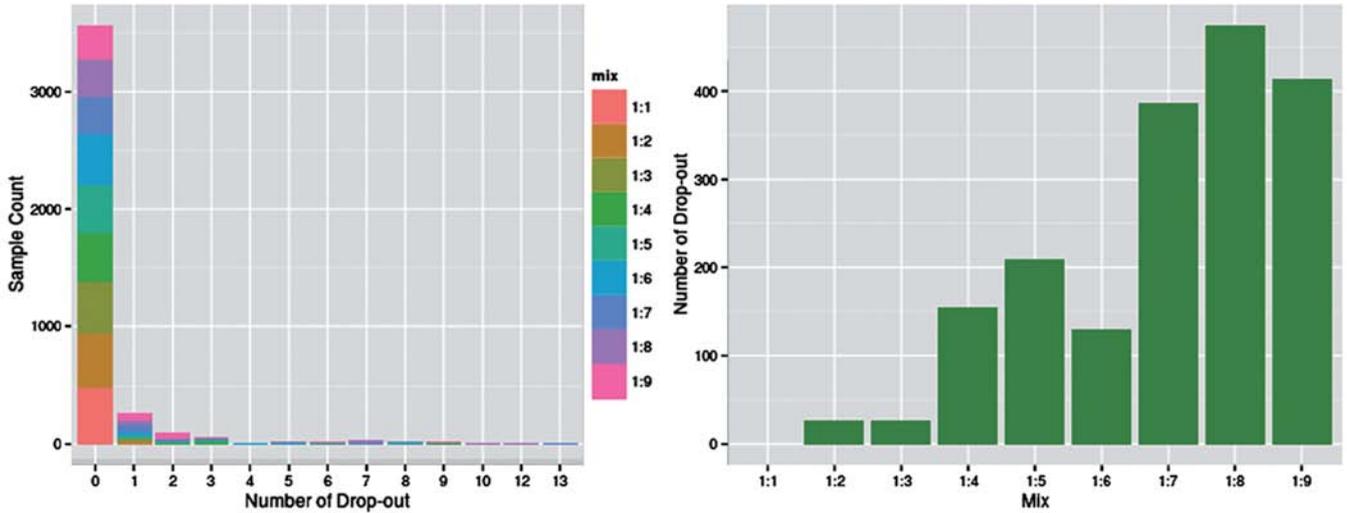


Figure 6. Data distribution and number of drop-out alleles at the 9 mixed gradients. The 9 gradients are marked with different colors.

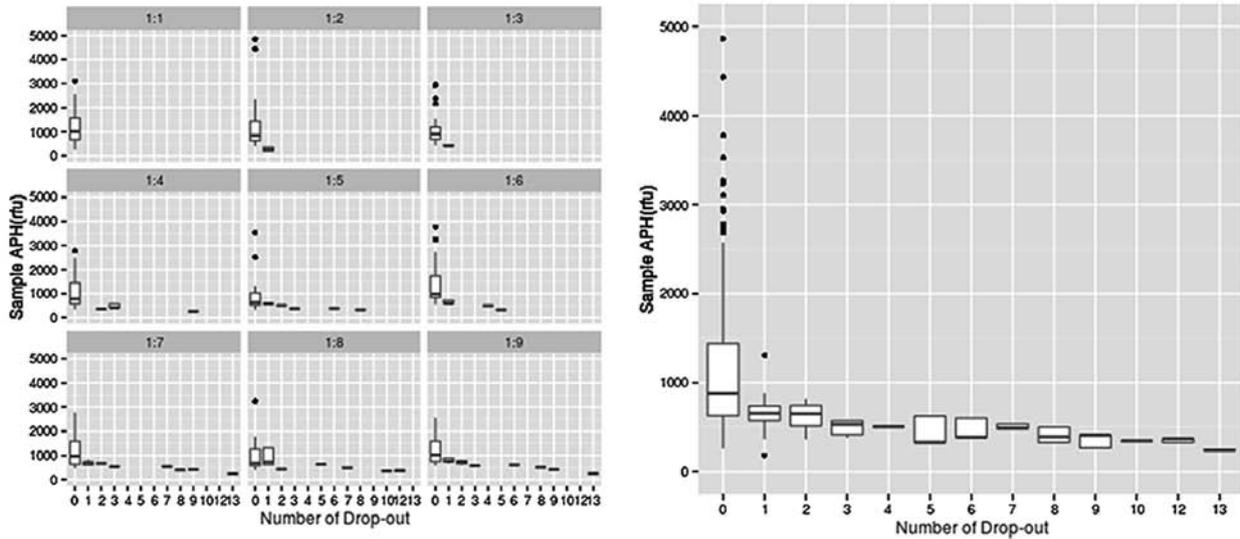


Figure 7. Number of drop-out alleles and distribution of average peak height (APH) in the 9 mixed gradients (left panel) and in all the mixed DNA samples (right panel).

*Correlation between APH and ADO.* Table VII presents 52 drop-out DNA samples in the 297 mixed male DNA profiles, with corresponding 218 ADOs. The statistical analysis of the drop-out samples distribution at the 9 mixed gradients showed that few ADOs were observed at gradients from 1:1 to 1:3, while the number of ADOs increased sharply at gradients from 1:7 to 1:9 (Fig. 6). These results demonstrated that the number of ADO correlated with the  $M_x$  value in the mixed DNA profiles. The incidence of drop-out would be greatly increased in an extremely imbalanced gradient (e.g., 1:7-1:9). Fig. 7 shows that many ADOs were

present at gradients from 1:5 to 1:9, with a corresponding APH of <1,000 rfu (Fig. 7, left panel), while Fig. 7 (right panel) shows a wide coverage of the APH without drop-out, with the highest median detected. The gross tendency appeared to be a gradual drop in APH is accompanied by an increase in the number of ADOs.

**Discussion**

The parameters analysis of this experimental model revealed a close linear correlation between PH and PA, 2 quantita-

tive parameters of mixed DNA analysis. These results are in agreement with the conclusion drawn in the study by Tvedebrink *et al* (40). The Kruskal-Wallis rank sum test revealed differences in the  $H_b$  distribution at 16 STR loci and the 9 mixed gradients. The changing tendency in APH and  $H_b$  fitted by locally weighted regression showed a difference in the  $H_b$  distribution at various STR loci, suggesting different efficiencies of these loci in the mixed DNA analysis. ADO correlated with both APH and mixed gradient, and all APH drop-out values were <1,000 rfu. All results indicated that APH affects  $H_b$  and drop-out distribution, and  $H_b$  correlates with the STR locus and mixed gradients. Further studies are required to investigate the causes responsible for the variation in the forensic efficiency of Identifiler® STR loci, including the different fluorescence sensitivity of genetic analyzer, which may cause APH distortion; and parameter analysis of the inter-locus balance ( $C_i$ ) on various STR loci, to reduce the bias in mixed DNA analysis.

Although stringent criteria are not generally necessary for single-DNA testing, our findings suggest that mixed STR profiles analyses should meet certain rfu levels in order to ensure the accuracy and reliability of the interpretation. The results from our study suggest that the forensic efficiency of the STR multiplex we used should be firstly evaluated in mixed DNA analysis, and calibration parameters should be introduced to correct the APH deviation of the STR loci during mixed DNA analysis.

### Acknowledgements

This study was supported by grants from the National Natural Science Foundation of China (grant no. 81273348), the National Key Technology R&D Program of China (grant no. 2012BAK02B01), and the Hebei Provincial Science and Technology Program of China (grant no. 12275648D).

### References

1. Budowle B, Onorato AJ, Callaghan TF, *et al*: Mixture interpretation: defining the relevant features for guidelines for the assessment of mixed DNA profiles in forensic casework. *J Forensic Sci* 54: 810-821, 2009.
2. Gill P, Brenner CH, Buckleton JS, *et al*: DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Sci Int* 160: 90-101, 2006.
3. Li CX, Han JP, Ren WY, Ji AQ, Xu XL and Hu L: DNA profiling of spermatozoa by laser capture microdissection and low volume-PCR. *PLoS One* 6: e22316, 2011.
4. Finkelhor D, Ormrod R, Turner H and Hamby S: School, police, and medical authority involvement with children who have experienced victimization. *Arch Pediatr Adolesc Med* 165: 9-15, 2011.
5. Longombe AO, Claude KM and Ruminjo J: Fistula and traumatic genital injury from sexual violence in a conflict setting in Eastern Congo: case studies. *Reprod Health Matters* 16: 132-141, 2008.
6. Tell K: African women struggling against female circumcision and sexual violence. *Reprod Freedom News* 8: 4-5, 1999.
7. Greene D, Maas CS, Carvalho G and Raven R: Epidemiology of facial injury in female blunt assault trauma cases. *Arch Facial Plast Surg* 1: 288-291, 1999.
8. Bright JA, Turkington J and Buckleton J: Examination of the variability in mixed DNA profile parameters for the Identifiler multiplex. *Forensic Sci Int Genet* 4: 111-114, 2010.
9. Bright JA, McManus K, Harbison S, Gill P and Buckleton J: A comparison of stochastic variation in mixed and unmixed casework and synthetic samples. *Forensic Sci Int Genet* 6: 180-184, 2012.

10. Debernardi A, Suzanne E, Formant A, Pène L, Dufour AB and Lobry JR: One year variability of peak heights, heterozygous balance and inter-locus balance for the DNA positive control of AmpFISTR® Identifiler® STR kit. *Forensic Sci Int Genet* 5: 43-49, 2011.
11. Wetton JH, Lee-Edghill J, Archer E, *et al*: Analysis and interpretation of mixed profiles generated by 34 cycle SGM Plus® amplification. *Forensic Sci Int Genet* 5: 376-380, 2011.
12. González-Andrade F, Bolea M, Martínez-Jarreta B and Sánchez D: DNA mixtures in forensic casework resolved with autosomic STRs. *Int Congr Ser* 1288: 580-582, 2006.
13. Noël J, Lavergne L, Mailly F, Roberge D and Jolicoeur C: Searching a DNA databank with complex mixtures from two individuals. *Forensic Sci Int Genet Suppl Ser* 2: 464-465, 2009.
14. Bright JA, Huizing E, Melia L and Buckleton J: Determination of the variables affecting mixed MiniFiler™ DNA profiles. *Forensic Sci Int Genet* 5: 381-385, 2011.
15. Perlin MW, Legler MM, Spencer CE, *et al*: Validating TrueAllele® DNA mixture interpretation. *J Forensic Sci* 56: 1430-1447, 2011.
16. Ellard S, Bulman MP, Frayling TM, *et al*: Allelic drop-out in exon 2 of the hepatocyte nuclear factor-1alpha gene hinders the identification of mutations in three families with maturity-onset diabetes of the young. *Diabetes* 48: 921-923, 1999.
17. Hill CR, Duewer DL, Kline MC, *et al*: Concordance and population studies along with stutter and peak height ratio analysis for the PowerPlex® ESX 17 and ESI 17 Systems. *Forensic Sci Int Genet* 5: 269-275, 2011.
18. Coletti A, Merigioli S, Severini S, *et al*: Statistical analysis of DNA mixtures using peak area information and allelic drop out. *Forensic Sci Int Genet Suppl Ser* 2: 202-203, 2009.
19. Tvedebrink T, Eriksen PS, Mogensen HS and Morling N: Estimating the probability of allelic drop-out of STR alleles in forensic genetics. *Forensic Sci Int Genet* 3: 222-226, 2009.
20. Van Nieuwerburgh F, Goetghebeur E, Vandewoestyne M and Deforce D: RMNE probability of forensic DNA profiles with allelic drop-out. *Forensic Sci Int Genet Suppl Ser* 2: 462-463, 2009.
21. Post WJ, Buijs C, Stolk RP, de Vries EG and le Cessie S: The analysis of longitudinal quality of life measures with informative drop-out: a pattern mixture approach. *Qual Life Res* 19: 137-148, 2010.
22. Haned H, Egeland T, Pontier D, Pene L and Gill P: Estimating drop-out probabilities in forensic DNA samples: a simulation approach to evaluate different models. *Forensic Sci Int Genet* 5: 525-531, 2011.
23. Tvedebrink T, Eriksen PS, Asplund M, Mogensen HS and Morling N: Allelic drop-out probabilities estimated by logistic regression-further considerations and practical implementation. *Forensic Sci Int Genet* 6: 263-267, 2012.
24. Tvedebrink T, Eriksen PS, Mogensen HS and Morling N: Statistical model for degraded DNA samples and adjusted probabilities for allelic drop-out. *Forensic Sci Int Genet* 6: 97-101, 2012.
25. Weiler NE, Matai AS and Sijen T: Extended PCR conditions to reduce drop-out frequencies in low template STR typing including unequal mixtures. *Forensic Sci Int Genet* 6: 102-107, 2012.
26. Kloosterman AD and Kersbergen P: Efficacy and limits of genotyping low copy number (LCN) DNA samples by multiplex PCR of STR loci. *J Soc Biol* 197: 351-359, 2003.
27. Ballantyne KN, van Oorschot RA and Mitchell RJ: Comparison of two whole genome amplification methods for STR genotyping of LCN and degraded DNA samples. *Forensic Sci Int* 166: 35-41, 2007.
28. McCartney C: LCN DNA: proof beyond reasonable doubt? *Nat Rev Genet* 9: 325, 2008.
29. Gill P: LCN DNA: proof beyond reasonable doubt? - a response. *Nat Rev Genet* 9: 726; author reply 726, 2008.
30. Gu LH, Dong Y, Zhang C, *et al*: Forensic analysis of LCN DNA using sample concentration methods followed by miniSTR genotyping. *Fa Yi Xue Za Zhi* 26: 361-363, 2010 (In Chinese).
31. Petricevic S, Whitaker J, Buckleton J, *et al*: Validation and development of interpretation guidelines for low copy number (LCN) DNA profiling in New Zealand using the AmpFISTR SGM Plus multiplex. *Forensic Sci Int Genet* 4: 305-310, 2010.
32. Tobe SS and Linacre A: Identifying endangered species from degraded mixtures at low levels. *Forensic Sci Int Genet Suppl Ser* 2: 304-305, 2009.

33. Tvedebrink T, Eriksen PS, Mogensen HS and Morling N: Amplification of DNA mixtures-missing data approach. *Forensic Sci Int Genet Suppl Ser 1*: 664-666, 2008.
34. Voskoboinik L and Darvasi A: Forensic identification of an individual in complex DNA mixtures. *Forensic Sci Int Genet* 5: 428-435, 2011.
35. Benschop CC, van der Beek CP, Meiland HC, van Gorp AG, Westen AA and Sijen T: 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, 2011.
36. Gilder JR, Inman K, Shields W and Krane DE: Magnitude-dependent variation in peak height balance at heterozygous STR loci. *Int J Legal Med* 125: 87-94, 2011.
37. Simoneau M and Corbeil P: The effect of time to peak ankle torque on balance stability boundary: experimental validation of a biomechanical model. *Exp Brain Res* 165: 217-228, 2005.
38. Sullivan KM, Mannucci A, Kimpton CP and Gill P: A rapid and quantitative DNA sex test: fluorescence-based PCR analysis of X-Y homologous gene amelogenin. *Biotechniques* 15: 636-638, 640-641, 1993.
39. Brinkmann B: Is the amelogenin sex test valid? *Int J Legal Med* 116: 63, 2002.
40. Tvedebrink T, Eriksen PS, Mogensen HS and Morling N: Evaluating the weight of evidence by using quantitative short tandem repeat data in DNA mixtures. *J R Stat Soc: Ser C (Appl Stat)* 59: 855-874, 2010.