



## Authentication of forensic DNA samples

Dan Frumkin<sup>a,\*</sup>, Adam Wasserstrom<sup>a,\*</sup>, Ariane Davidson<sup>a</sup>, Arnon Grafit<sup>b</sup>

<sup>a</sup> Nucleix Ltd., 27 Habarzel St., Tel Aviv 69710, Israel

<sup>b</sup> Serious Crime Unit Mobile Lab., Division of Identification & Forensic Science, Israel Police, Israel

### ARTICLE INFO

#### Article history:

Received 3 April 2009

Received in revised form 11 June 2009

Accepted 16 June 2009

#### Keywords:

Artificial DNA

Methylation analysis

DNA authentication

Forensic science

### ABSTRACT

Over the past twenty years, DNA analysis has revolutionized forensic science, and has become a dominant tool in law enforcement. Today, DNA evidence is key to the conviction or exoneration of suspects of various types of crime, from theft to rape and murder. However, the disturbing possibility that DNA evidence can be faked has been overlooked. It turns out that standard molecular biology techniques such as PCR, molecular cloning, and recently developed whole genome amplification (WGA), enable anyone with basic equipment and know-how to produce practically unlimited amounts of *in vitro* synthesized (artificial) DNA with any desired genetic profile. This artificial DNA can then be applied to surfaces of objects or incorporated into genuine human tissues and planted in crime scenes. Here we show that the current forensic procedure fails to distinguish between such samples of blood, saliva, and touched surfaces with artificial DNA, and corresponding samples with *in vivo* generated (natural) DNA. Furthermore, genotyping of both artificial and natural samples with Profiler Plus<sup>®</sup> yielded full profiles with no anomalies. In order to effectively deal with this problem, we developed an authentication assay, which distinguishes between natural and artificial DNA based on methylation analysis of a set of genomic loci: in natural DNA, some loci are methylated and others are unmethylated, while in artificial DNA all loci are unmethylated. The assay was tested on natural and artificial samples of blood, saliva, and touched surfaces, with complete success. Adopting an authentication assay for casework samples as part of the forensic procedure is necessary for maintaining the high credibility of DNA evidence in the judiciary system.

© 2009 Elsevier Ireland Ltd. All rights reserved.

### 1. Introduction

The current forensic procedure that deals with DNA evidence starts at the crime scene where biological samples such as blood and saliva stains are detected, identified, documented, collected, and transferred to the forensic laboratory. In the laboratory, DNA is extracted and quantified, usually by real time PCR amplification of the hTERT locus (Quantifiler<sup>®</sup>) or similar targets [1]. Following quantification, about 1 ng of the DNA is used for a profiling reaction, in which 9–15 highly polymorphic short tandem repeat (STR) loci and the sex-typing marker amelogenin are genotyped. The loci are usually chosen from a standard set of core loci such as the 13 Combined DNA Index System (CODIS) loci. A detailed description of the forensic procedure is provided in Text S1.

The DNA profile of every person is considered unique (except for identical twins) [2], and consequently, this “DNA fingerprint” is used in police investigations to link between a crime scene and a specific individual, who is either a suspect in the case, or identified

by an automatic search of the database (e.g. CODIS). In recent years, DNA evidence has become the “gold standard” of forensic testing, and is an invaluable tool for the criminal justice community [3–7]. The high credibility of DNA evidence in court stems from the fact that it uses a statistical approach based on population genetics and empirical testing [8], in contrast to other types of forensic evidence, such as ballistics, blood-spatter analysis, and fiber analysis, which rely on expert judgment and have limited connection to established science [7]. It is even considered to be more reliable than eyewitness evidence, which is known to suffer from a relatively high rate of errors [8].

The use of DNA recovered at crime scenes as evidence in court relies on the implicit assumption that the DNA is genuine—originating from natural biological material. However, as we show here, this assumption may not necessarily be true: DNA with any desired genetic profile can easily be synthesized *in vitro* using common [9,10], and recently developed [11,12] biological techniques, integrated into genuine human tissues or applied to surfaces of objects, and then planted in crime scenes. When the current forensic procedure is applied to objects or human tissues that contain synthesized DNA, it fails to recognize the artificial origin of the sample, and the resulting profile is indistinguishable from a genuine DNA profile. Nevertheless, we demonstrate

\* Corresponding authors. Tel.: +972 3 768 4935; fax: +972 3 768 4945.

E-mail addresses: [dan@nucleix.com](mailto:dan@nucleix.com) (D. Frumkin), [adam@nucleix.com](mailto:adam@nucleix.com) (A. Wasserstrom).

that natural and artificial samples can be differentiated based on differential methylation patterns. Methylation is an epigenetic chemical modification of DNA, occurring in mammals in the form of a methyl group ( $-CH_3$ ) that is enzymatically added to the C5 position of cytosine in some CpG dinucleotides [13]. DNA methylation is believed to inhibit gene expression in animal cells, probably by affecting chromatin structure [14]. In the human genome 70–80% of all CpGs are methylated, while unmethylated CpGs are grouped in clusters called “CpG islands” [15].

## 2. Materials and methods

### 2.1. Collection of biological tissues

Samples of blood, dry saliva stains on absorbent paper, skin scrapings, hair, and smoked cigarette butts were collected from volunteers. Informed consent was obtained from all participants recruited into the study. DNA from these samples was extracted and quantified as described in Section 2.6.

### 2.2. CODIS allele library

For construction of the library, individual alleles of CODIS STRs and the hTERT locus were amplified from pooled DNA (Control Human Genomic DNA of the GenomePlex WGA2 kit, Sigma–Aldrich) by separate PCR reactions (primers and conditions as described in Section 2.9). Amplified fragments were purified (QIAquick PCR purification kit, QIAGEN), and cloned into the pGEM-T-Easy vector (Promega). Plasmid DNA was purified by the QIAprep Spin Miniprep kit (QIAGEN) and quantified (Nanodrop 1000, Thermo Scientific). For genotyping of cloned alleles, the PowerPlex16 (Promega) kit was used. Genotyping was performed in a high throughput manner by simultaneously genotyping 10–15 clones (from different CODIS loci) in a single PowerPlex16 reaction. In the resulting library each element is a microcentrifuge tube with trillions of copies of a single allele (for example, one element is allele 11 of locus D8S1179, while

another is allele 12 of D8S1179, and likewise for the other CODIS loci). We note that 1 fg of plasmid in the library contains  $\sim 160$  copies of its cloned allele—the same copy number that is present in  $\sim 1$  ng of a haploid genome.

### 2.3. In vitro synthesis of DNA

Artificial DNA was synthesized by one of the following methods:

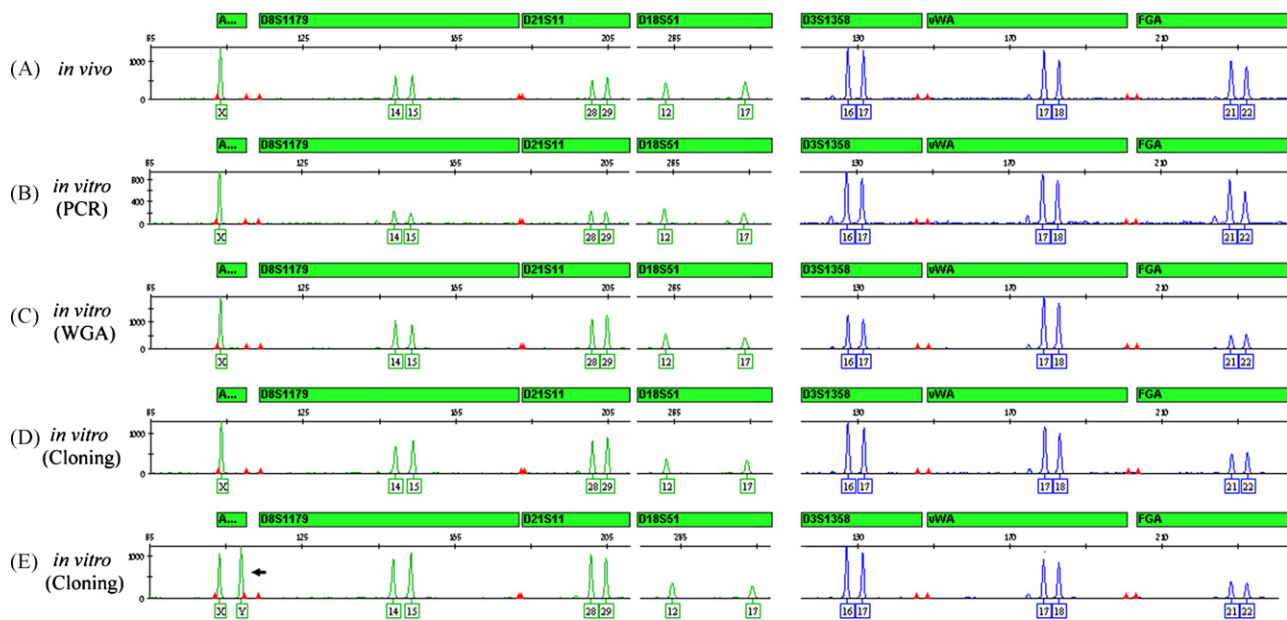
**PCR:** For the sample whose profile is shown in Fig. 1, the 10 loci included in the Profiler Plus<sup>®</sup> kit (Applied Biosystems) were amplified separately from 1 ng of DNA extracted from a cigarette butt smoked by ‘N400’ (PCR conditions were as described in Section 2.9; primer sequences are in Text S3). Individual amplified fragments were purified (QIAquick PCR purification kit, QIAGEN), quantified (Nanodrop 1000, Thermo Scientific), diluted about a million fold (depending on the concentration of the specific amplicon), and combined in a single test tube. For the sample whose profile is shown in Fig. 2, 1 ng of ‘N222’ DNA (extracted from a saliva stain on absorbent paper) was used as template in a single PCR reaction using the Profiler Plus<sup>®</sup> primer mix. A 1:1000 dilution of the PCR reaction was used for generating the artificial sample.

**WGA:** Whole genome amplification was performed by multiple displacement amplification [16] with the Repli-g Midi kit (QIAGEN) using 10 ng of natural DNA as template.

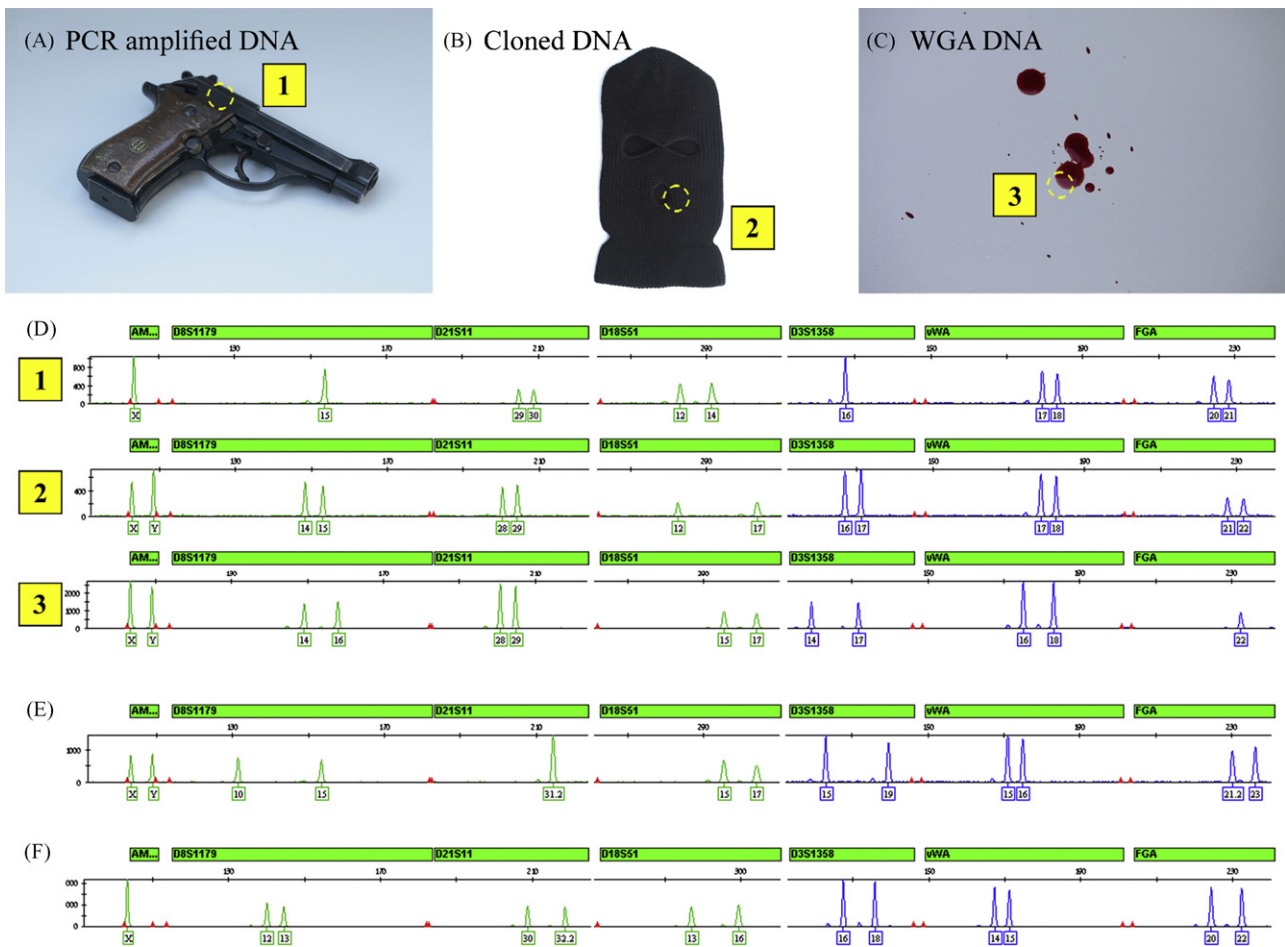
**Assembly from CODIS allele library:** For assembling profiles using the CODIS allele library, equal quantities of alleles (cloned into plasmids) in the desired profile were picked from the library and combined in a single tube.

### 2.4. Generation of mock forensic samples

For generating artificial touch DNA samples, *in vitro* synthesized DNA was applied directly to the surface of the object and allowed to dry. For generating artificial blood samples, red blood cells were isolated from whole blood by centrifugation ( $1500 \times g$ , 10 min), and mixed with *in vitro* synthesized DNA. Drops of the red



**Fig. 1.** Profiles of *in vivo*- and *in vitro*-synthesized DNA are indistinguishable. (A) Profile of natural DNA obtained from the saliva of female donor ‘N400’. (B–D) Artificial profiles of ‘N400’ obtained from DNA that was synthesized *in vitro* by three different methods: PCR (B), WGA (C), and assembly from a library of cloned CODIS alleles (D). (E) Artificial profile of ‘male-N400’, which is identical to the profile of ‘N400’ at all loci, except for the Amelogenin locus. This profile was created by adding a cloned Y allele (indicated by arrow) to the mix used to generate the profile in (D). In A–E partial profiles are depicted; full profiles are provided in Text S2.



**Fig. 2.** Mock forensic samples with artificial DNA. (A) Handgun with PCR amplified DNA with the profile of 'N222' applied to the external surface of its action. (B) Ski-mask with artificial saliva applied to its inner surface. The artificial saliva contained an extract of natural saliva from 'N270' (without DNA) and DNA fragments with the profile of 'male N400' assembled from the CODIS allele library. (C) Artificial bloodstains containing red blood cells from natural blood of 'N227' and artificial 'N283' DNA generated by WGA. In A–C, yellow circles depict the areas from which samples were taken for analysis. (D) Profiles of the three artificial samples. All three profiles received a "perfect" GeneMapper ID-X score, and are identical to the genotypes of the artificial DNA that was used in their production. No traces of DNA from the saliva extract and red blood cells are visible in the profiles from the ski-mask and bloodstains (see E and F). (E) Profile of donor 'N270', whose saliva extract was used for manufacturing the ski-mask sample. (F) Profile of donor 'N227', whose red blood cells were used for manufacturing the bloodstain sample. In D–F partial profiles are depicted; full profiles are provided in Text S2.

blood cell–DNA mix were dripped from a height of 1 m and allowed to dry. For generating artificial saliva samples, saliva extract (containing no cells) was isolated from the top phase of centrifuged natural saliva (1500 × g, 10 min), and mixed with *in vitro* synthesized DNA. The saliva extract–DNA mix was applied directly to the surface of the object and allowed to dry. A detailed description of all samples is provided in Text S4.

### 2.5. Identification and collection of mock forensic samples

Stains were identified as human blood using the HEXAGON OBTI kit (BLUESTAR), and as saliva using Phadebas<sup>®</sup> Amylase test (Phadebas). Samples of blood and touch DNA were collected with a sterile cotton swab, dampened with distilled water. Saliva samples were composed of cut-out portions of the ski-mask fabric around the mouth orifice.

### 2.6. DNA extraction and quantification

DNA extraction from all samples was performed according to an organic extraction protocol [17]. DNA quantification was performed using the Quantifiler<sup>®</sup> Human DNA quantification kit (Applied Biosystems). Real time PCR was performed on a StepOne<sup>™</sup> system (Applied Biosystems).

### 2.7. DNA profiling, capillary electrophoresis and signal analysis

STR loci were amplified using the Profiler Plus<sup>®</sup> (Applied Biosystems) and PowerPlex16 (for preparing the CODIS allele library; Promega) kits using a GeneAmp<sup>®</sup> PCR System 9700 (Applied Biosystems). Amplification products were run on an ABI 310 Genetic Analyzer (Applied Biosystems) according to the manufacturer's instructions. The resulting electropherograms were analyzed using GeneMapper ID-X analysis software (Applied Biosystems).

### 2.8. Bisulfite conversion and methylation analysis

Bisulfite conversion was performed with the EpiTect<sup>™</sup> kit (Qiagen). Converted DNA was amplified by PCR at the set of loci used for authentication. In each PCR, 1/10 of the EpiTect<sup>™</sup> products were used as template and the reaction was performed as described in Section 2.9. Amplified fragments were purified using the QIAquick PCR purification kit (QIAGEN) and sequenced.

### 2.9. PCR

All non-profiling PCRs were performed in a total volume of 50 µl with 0.2 µM each primer, 0.2 mM each dNTP, 5 U AmpliTaq

Gold (Applied Biosystems), and 5  $\mu$ l 10 $\times$  PCR Buffer containing 15 mM MgCl<sub>2</sub> (Applied Biosystems). Amplification was performed in a GeneAmp<sup>®</sup> PCR System 9700 (Applied Biosystems). The PCR program used was: 95 °C for 11 min, followed by 35 cycles of 94 °C for 1 min, 59 °C for 1 min, 72 °C for 1 min, and followed by a final extension step of 60 °C for 45 min. PCRs for profiling reactions were performed according to the manufacturer's instructions (with 28 cycles).

### 2.10. Probability of “non-existent” profile

The probability that a random unrelated male has the Profiler Plus<sup>®</sup> profile of ‘male-N400’ (a profile identical to that of ‘N400’ with the exception of the Amelogenin locus, in which its genotype is XY instead of XX) was calculated based on allele frequencies in the US Caucasian population [18]. This probability was multiplied by  $3.5 \times 10^9$  (approximate male population) to yield the approximate probability that there exists a person with the ‘male N400’ profile (excluding close relatives of ‘N400’).

### 2.11. DNA mixtures

DNA mixtures were created by combining natural ‘N217’ DNA (extracted from blood with the FlexiGene DNA kit, QIAGEN) with artificial ‘N226’ DNA (amplified by Repli-g Midi kit, QIAGEN from DNA extracted by organic extraction from a single hair).

## 3. Results

### 3.1. Profiles of *in vivo*- and *in vitro*-synthesized DNA are indistinguishable

To demonstrate that DNA can be synthesized *in vitro* such that its profile will be indistinguishable from that of DNA of *in vivo* origin, we profiled a natural DNA sample and compared it to corresponding profiles from DNA that was synthesized *in vitro* by three different methods. Natural DNA was extracted from a saliva sample of female donor ‘N400’ and genotyped using the Profiler Plus<sup>®</sup> and GeneMapper ID-X (Applied Biosystems); (Fig. 1A). The profile obtained from the saliva of donor ‘N400’ was perfect, as indicated by the green bars above all loci.

Next, we produced artificial DNA with the same genotype as ‘N400’ using three different types of *in vitro* synthesis: PCR (Fig. 1B), WGA (Fig. 1C), and molecular cloning (Fig. 1D). The genotypes of all *in vitro* synthesized ‘N400’ samples were perfect according to GeneMapper ID-X analysis and identical to the genotype of natural ‘N400’ DNA. Template DNA for PCR was 1 ng of DNA extracted from a cigarette butt smoked by ‘N400’, and the template for WGA was 10 ng of ‘N400’ DNA extracted from a dry saliva stain on absorbent paper. The sample created by molecular cloning did not require any ‘N400’ DNA as template (only a priori knowledge of her profile) and was assembled using the “CODIS allele library” that we created beforehand. The library consists of an array of single CODIS alleles cloned into plasmids, and can be used to generate different desired profiles by assembly of their constituent alleles. In order to demonstrate the possibility to create any desired profile from such a library, we also assembled a profile of a non-existent person, which we term ‘male N400’. This profile is identical to that of ‘N400’, with the exception of the Amelogenin locus, in which its genotype is XY instead of XX (Fig. 1E). We calculated that the probability that a male unrelated to ‘N400’ has a profile identical to that of ‘male N400’ is  $7.95 \times 10^{-12}$ , and consequently the probability that there does not exist in the world population an unrelated male with an identical profile is greater than 99.99%.

### 3.2. The current forensic procedure fails to distinguish between natural and artificial DNA evidence

#### 3.2.1. Generation of artificial DNA evidence

We created 10 mock forensic samples with artificial DNA, of types that may be found in crime scenes, and subjected three of these samples to analysis through the complete forensic procedure (the rest of the samples are discussed in Section 3.4). These three samples contained artificial DNA that was synthesized using different methods: a handgun sample with PCR amplified DNA, a ski-mask with saliva containing DNA fragments from the CODIS allele library, and bloodstains containing DNA synthesized by WGA (Fig. 2A–C; see detailed description in Text S4).

#### 3.2.2. Analysis of artificial DNA evidence

The three samples were processed according to the routine forensic procedure performed in crime scenes. Samples were collected from the external surface of the handgun action, from the ski-mask fabric, and from the bloodstains. A portion of the ski-mask sample was tested for presence of saliva using the Phadebas<sup>®</sup> assay, and the results were positive (data not shown), due to the presence of amylase in the supernatant of the natural saliva extract. A portion of the bloodstain sample was tested for the presence of human blood DNA using the HEXAGON OBTI assay, and the results were positive (data not shown), due to the presence of hemoglobin in the red blood cells. DNA was extracted from the samples and profiled (Fig. 2D). The genotypes of all three samples were identical to the genotypes of the artificial DNA that was used in their production (‘N222’, ‘male N400’, and ‘N283’, respectively). Furthermore, in the artificial saliva and blood samples there were no observable traces of natural DNA from the saliva and blood donors (Fig. 2E and F), and all artificial profiles received a perfect GeneMapper ID-X score, consistent with a single contributor.

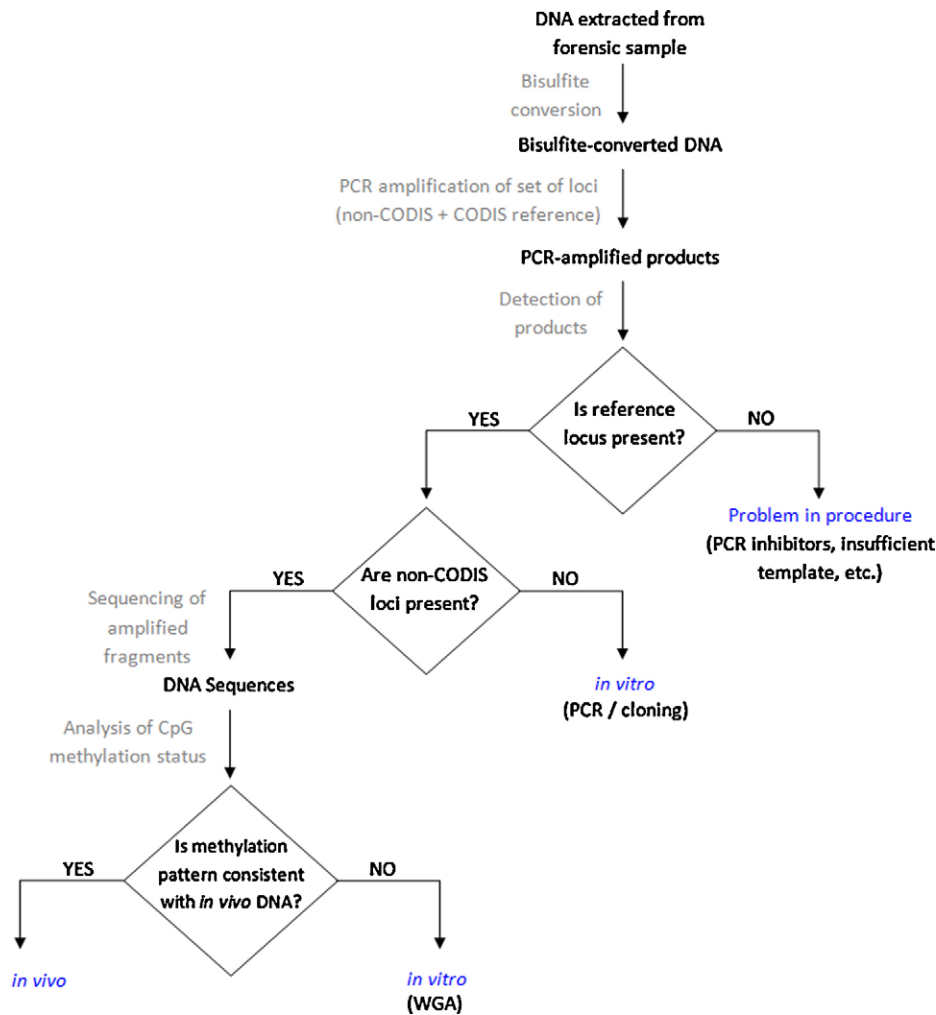
#### 3.2.3. Independent analysis of artificial blood evidence

In order to check whether the profiling results obtained in our laboratory were dependant on our specific setup, we sent a duplicate swab of the artificial blood sample to a leading forensic DNA laboratory for analysis. The procedures employed by this laboratory have been validated according to standards established by the Scientific Working Group on DNA Analysis Methods (SWGDM) and adopted as US Federal Standards. DNA was extracted from the sample in the laboratory using the EZ1 DNA Investigator Kit (QIAGEN), and quantified using a proprietary real time PCR assay (both extraction and quantification methods were different than those employed in our lab). Genotyping was performed with Profiler Plus<sup>®</sup> and COfiler<sup>®</sup> (Applied Biosystems). The report received from the laboratory states that “The DNA profile obtained from sample 2S09-002-001 [the artificial blood swab] is consistent with a male contributor”, and the profiling results were identical to the genotype of the artificial DNA of donor ‘N283’, with “No Edits” (i.e. no anomalies found in any of the analyzed loci; see report in Text S5).

These results demonstrate that artificial DNA can easily be applied to surfaces of objects or incorporated into genuine human tissues, thereby creating artificial forensic evidence that, after undergoing the entire forensic casework procedure, yields perfect profiles.

### 3.3. Description of the DNA authentication assay

We developed an authentication assay capable of differentiating between natural and all types of artificial DNA. The assay is based on the fact that unlike *in vitro* synthesized DNA which is completely unmethylated, *in vivo* generated DNA contains loci that are completely and consistently methylated and other loci that are



**Fig. 3.** Scheme of the DNA authentication assay. The assay accepts as input DNA that was extracted from the forensic sample in question and outputs whether the DNA is authentic (*in vivo* generated) or non-authentic (*in vitro* synthesized).

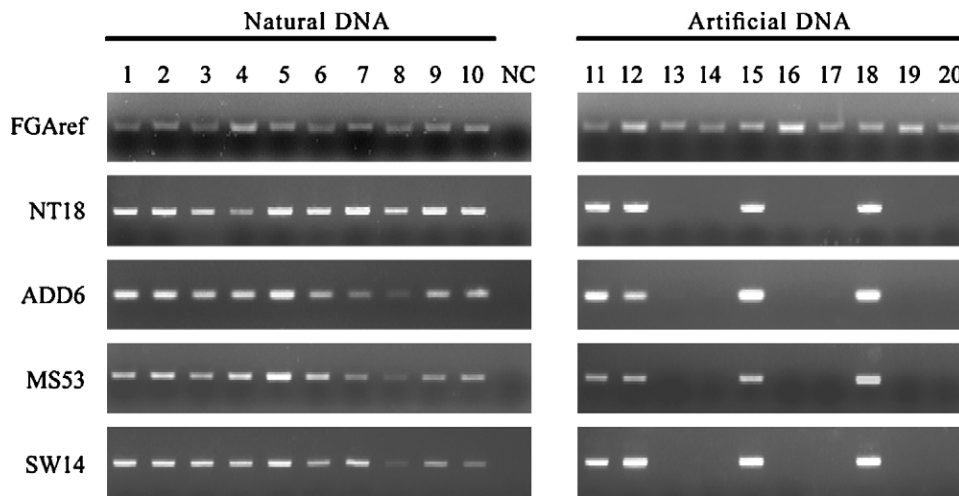
completely and consistently unmethylated. A scheme of the assay is presented in Fig. 3. DNA from a forensic sample in question is treated with sodium bisulfite, which converts all unmethylated cytosines to uracils, while leaving the methylated cytosines unaffected [19]. Following bisulfite conversion, the DNA is amplified by PCR at a set of loci, containing one reference CODIS locus (FGAref), and four non-CODIS loci (NT18, ADD6, MS53, SW14; Text S6). The set of loci consist of high-complexity, non-repetitive DNA sequences (FGAref consists of the non-repetitive part of the FGA locus). They were chosen because NT18 and ADD6 are consistently methylated, while MS53 and SW14 are consistently unmethylated in human tissues such as blood, saliva and epidermis (the source of touch DNA). For increasing the reliability of the assay, the primers for these loci were designed to amplify with equal efficiency both converted and unconverted DNA, thus enabling detection of incomplete bisulfite conversion. Following PCR, the presence or absence of amplicons is determined by electrophoresis (alternatively, real time PCR can be used). Complete absence of amplicons (including FGAref) indicates a problem in the procedure due to PCR inhibitors, insufficient template, etc. Successful amplification of FGAref with concomitant failure of amplification of the non-CODIS loci indicate that the DNA is artificial and was synthesized by one of the methods that generate only a subset of genomic loci (e.g. PCR or cloning of CODIS loci). Successful amplification of all loci indicates that the DNA contains a full representation of the genome and is either natural DNA or artificial DNA synthesized by WGA. Differentiation

between these two types of DNA is achieved by sequencing the four non-CODIS amplicons and analyzing their methylation pattern. The DNA is determined to be of *in vivo* origin if its methylation pattern is consistent with that of *in vivo* generated DNA (i.e. complete methylation of all CpGs in NT18 and ADD6 alongside with complete non-methylation of all CpGs in MS53 and SW14), otherwise it is determined to be of *in vitro* origin.

#### 3.4. Demonstration of the DNA authentication assay

We applied the DNA authentication assay to 20 mock forensic samples, 10 with natural DNA, 10 with artificial DNA (three of these samples were described in Section 3.2), and a negative control sample without DNA (detailed description in Text S4). All samples with natural DNA showed successful amplification of all loci, and the FGAref amplicon was present in all samples, both natural and artificial (Fig. 4). Samples 13,14,16,17,19,20 which contain artificial DNA synthesized by PCR or molecular cloning, failed to amplify the four non-CODIS loci, since the DNA in these samples contains only CODIS loci. These samples were therefore determined to be non-authentic and were not processed further. The remaining artificial DNA samples (11,12,15,18) contained WGA-synthesized DNA and in these samples all loci amplified successfully, similar to natural DNA.

The natural and WGA-synthesized DNA samples were processed further by sequencing the four non-CODIS loci and analyzing the methylation status at all CpG positions (Table 1). All natural DNA



**Fig. 4.** Amplification products in natural and artificial mock forensic samples. Aliquots of PCR products were run on a 2% Agarose gel. The FGaref locus is amplified in all samples (both natural and artificial), but not in the negative control sample. Non-CODIS loci are amplified in all natural (1–10) and in WGA-based artificial samples (11,12,15,18), but are absent in PCR- and cloning-based artificial samples (13,14,16,17,20).

samples showed complete methylation of all CpG positions in NT18 and ADD6, and no methylation in any of the CpG positions in MS53 and SW14. In contrast, all WGA-synthesized samples showed complete lack of methylation in all loci (Fig. 5). Based on this analysis, the 10 natural samples were determined to be authentic, and the four WGA-synthesized samples were determined to be non-authentic. Therefore the assay was successful in determining the correct status of all 20 samples (Table 1).

### 3.5. Natural–artificial DNA mixtures

We checked the authentication assay on mixtures of natural ‘N217’ DNA and artificial ‘N226’ DNA with various ratios of DNA (the percentage of artificial DNA ranged from 10 to 67%). The mixtures were profiled with GeneMapper ID-X software using the mixture analysis mode and all mixtures were correctly identified as two-contributor mixtures, except for the 10% artificial DNA

mixture, which was identified as a single contributor sample with a profile identical to that of the natural DNA (Text S7). In methylation analysis of the mixtures, natural and artificial sequence signals are superimposed and signals from both types of DNA are observed even in the 10% artificial DNA mixture (Text S7). These results indicate that artificial DNA can be detected in DNA mixtures, even when it constitutes a minor component.

## 4. Discussion

### 4.1. Producing artificial DNA evidence requires only basic equipment and know-how

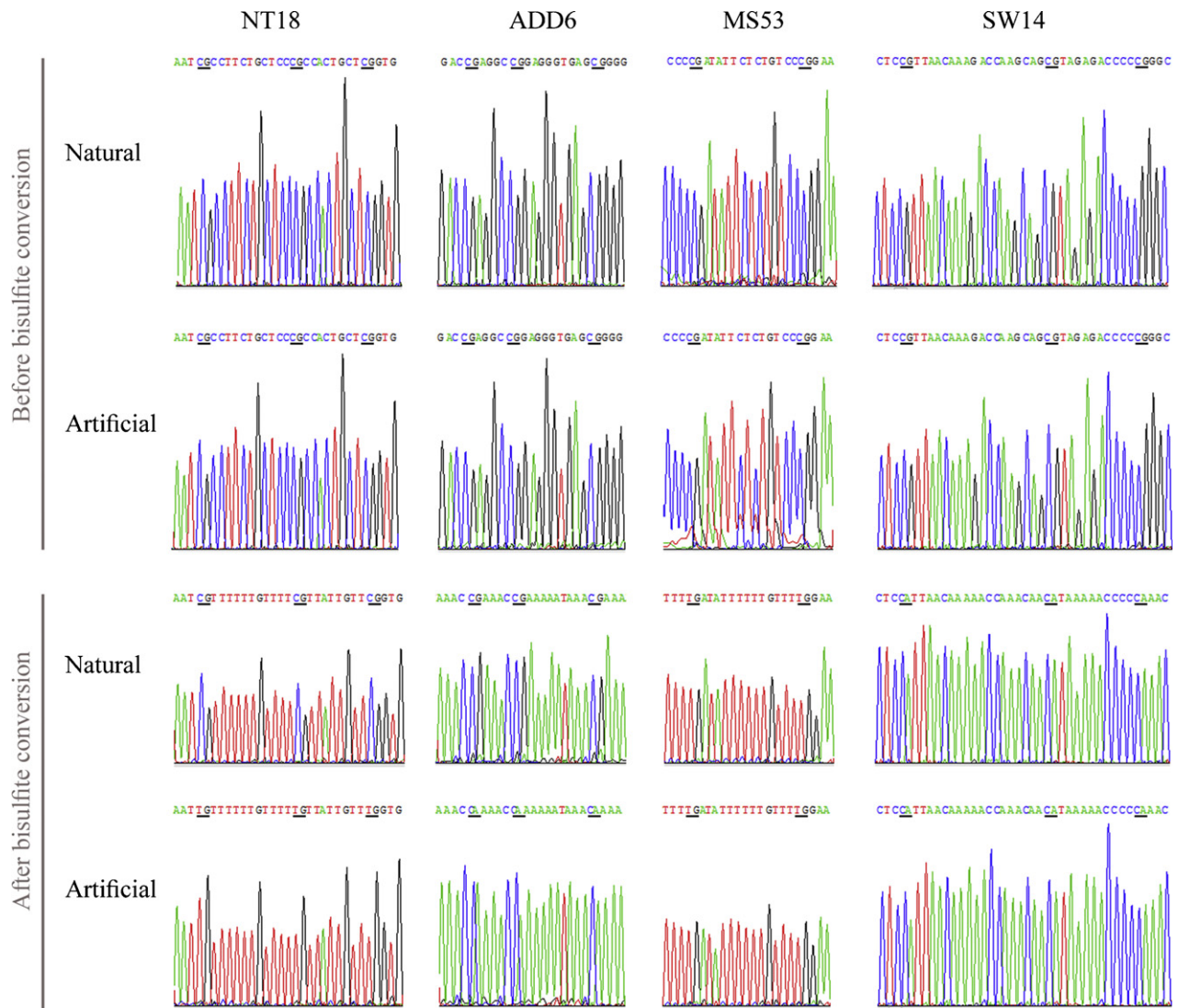
We demonstrated the ease at which artificial DNA evidence can be produced, and that such evidence “passes” the current forensic procedure as genuine. The fact that an independent forensic laboratory, which provides services to United States law enforce-

**Table 1**  
DNA authentication results on natural and artificial mock forensic samples.

#	Sample Source of DNA	FGaref amplified	Methylated CpG positions <sup>a</sup>				Decision
			NT18	ADD6	MS53	SW14	
1	<i>In vivo</i> (blood)	Yes	12/12	11/11	0/6	0/17	Authentic
2	<i>In vivo</i> (blood)	Yes	12/12	11/11	0/6	0/17	Authentic
3	<i>In vivo</i> (blood)	Yes	12/12	11/11	0/6	0/17	Authentic
4	<i>In vivo</i> (blood)	Yes	12/12	11/11	0/6	0/17	Authentic
5	<i>In vivo</i> (saliva)	Yes	12/12	11/11	0/6	0/17	Authentic
6	<i>In vivo</i> (saliva)	Yes	12/12	11/11	0/6	0/17	Authentic
7	<i>In vivo</i> (saliva)	Yes	12/12	11/11	0/6	0/17	Authentic
8	<i>In vivo</i> (skin)	Yes	12/12	11/11	0/6	0/17	Authentic
9	<i>In vivo</i> (skin)	Yes	12/12	11/11	0/6	0/17	Authentic
10	<i>In vivo</i> (skin)	Yes	12/12	11/11	0/6	0/17	Authentic
11	<i>In vitro</i> (WGA)	Yes	<b>0/12</b>	<b>0/11</b>	0/6	0/17	Non-authentic
12	<i>In vitro</i> (WGA)	Yes	<b>0/12</b>	<b>0/11</b>	0/6	0/17	Non-authentic
13	<i>In vitro</i> (PCR)	Yes	<b>No amp.</b>	<b>No amp.</b>	<b>No amp.</b>	<b>No amp.</b>	Non-authentic
14	<i>In vitro</i> (Cloning)	Yes	<b>No amp.</b>	<b>No amp.</b>	<b>No amp.</b>	<b>No amp.</b>	Non-authentic
15	<i>In vitro</i> (WGA)	Yes	<b>0/12</b>	<b>0/11</b>	0/6	0/17	Non-authentic
16	<i>In vitro</i> (PCR)	Yes	<b>No amp.</b>	<b>No amp.</b>	<b>No amp.</b>	<b>No amp.</b>	Non-authentic
17	<i>In vitro</i> (Cloning)	Yes	<b>No amp.</b>	<b>No amp.</b>	<b>No amp.</b>	<b>No amp.</b>	Non-authentic
18	<i>In vitro</i> (WGA)	Yes	<b>0/12</b>	<b>0/11</b>	0/6	0/17	Non-authentic
19	<i>In vitro</i> (PCR)	Yes	<b>No amp.</b>	<b>No amp.</b>	<b>No amp.</b>	<b>No amp.</b>	Non-authentic
20	<i>In vitro</i> (Cloning)	Yes	<b>No amp.</b>	<b>No amp.</b>	<b>No amp.</b>	<b>No amp.</b>	Non-authentic
21	Negative Control	No	No amp.	No amp.	No amp.	No amp.	No decision <sup>b</sup>

<sup>a</sup> Number of methylated CpG positions out of total number of CpG positions in each locus. No amp. = No amplicon observed; bold indicates results inconsistent with DNA of *in vivo* origin.

<sup>b</sup> “No decision” is outputted when there is no amplification in any of the loci. Possible reasons may be insufficient/degraded template DNA, PCR inhibitors, etc.



**Fig. 5.** Methylation analysis of natural and artificial samples. Partial sequences of DNA from natural and artificial blood samples (samples 2 and 11, respectively) at non-CODIS loci (CpG dinucleotides are underlined). The sequences of unconverted DNA are identical at all loci, demonstrating that natural and artificial samples cannot be distinguished on the basis of sequence alone. Following bisulfite conversion, the differential methylation pattern of natural vs. artificial DNA is exposed: natural DNA is methylated at NT18 and ADD6, and unmethylated at MS53 and SW14, while artificial DNA is unmethylated at all four loci.

ment agencies, analyzed our artificial blood sample yielding a perfectly normal, single contributor DNA profile—attests to the problem.

In this case the artificial DNA was designed to have the profile of donor 'N283', and was amplified from a minute amount of DNA extracted from a single hair of this donor. Similarly, we produced artificial samples of DNA amplified from a cigarette butt and a dry saliva stain on absorbent paper. Such common everyday objects, which can be used to obtain source DNA for producing artificial samples, can be obtained from practically anyone. Even this constraint is removed when considering the possibility to produce artificial evidence using the "CODIS allele library", since any profile can be assembled without the need for source DNA, only requiring knowledge of the desired profile. A library containing 425 clones corresponding to all known CODIS alleles (including all rare microvariants) is sufficient to generate any desired profile, while a much smaller library is sufficient to generate the profiles of the vast majority of the human population.

Once source DNA from a person or knowledge of his/her profile is obtained, the actual manufacturing of the artificial sample is simple and straightforward. Generating large amounts of artificial

DNA can be performed overnight, using basic laboratory equipment and commercial kits, requires only basic knowledge in molecular biology, and little financial expense. There is a very large and growing number of people with the necessary expertise and access to the required equipment, such as scientists, research students, lab technicians in hospitals, pharmaceutical or biotech companies, etc. Such people might manufacture artificial DNA and use it maliciously themselves, or transfer it to other people who do not have the ability to manufacture the DNA. Moreover, since commercial molecular biology services are becoming widespread and DNA with any sequence can be ordered online, manufacturing an artificial DNA sample does not require much more than a personal computer and link to the internet.

#### 4.2. Authentication is necessary for preventing false DNA matches

The DNA profiles of millions of people are registered in rapidly growing national databases, and the current trend around the world is to include more and more profiles in them, not only of convicted offenders, but also of arrestees. Profiles from casework samples are routinely searched against these databases (e.g. by

automatic software such as CODIS), and when an identical profile is found, a DNA “match” is made, making the identified person a suspect in the case and usually leading to his arrest [20]. In some jurisdictions, DNA evidence alone can lead to conviction without the requirement of any corroborating evidence [21]. However, even when supporting evidence is required by law, there is little doubt that the presence of DNA evidence from a crime scene against a defendant places him/her at a dire position.

The combination of the ease at which artificial DNA samples can be manufactured, with the fact that a registered DNA profile found at a crime scene will automatically lead to a database “match”, and the heavy weight of DNA evidence in the courtroom, creates a problematic situation which we believe should be addressed by the forensic community by adopting a DNA authentication assay for casework samples.

Artificial DNA with unregistered profile can also be produced, and in such a case it will not result in a match. Nevertheless, this artificial DNA evidence might hamper the investigation and authentication of this DNA could aid in focusing the investigation to relevant directions.

#### 4.3. SNP based profiling approaches are also susceptible to fabrication

Recently, alternatives to STR based profiling have been proposed, primarily single nucleotide polymorphism (SNP) based approaches [22], which may be advantageous over STR profiling [23–25]. Similar to STR based profiling, SNP based approaches are also susceptible to fabrication by the methods described here. Even if a very large number of SNPs are to be used in profiling, this will not effectively deal with the problem of WGA-based fabrication, since WGA produces a full representation of the genome, and therefore is expected to produce a perfect “SNP profile”.

#### 4.4. Integrating DNA authentication into the forensic procedure

The assay described here employs bisulfite sequencing, a procedure that is relatively labor intensive, time consuming, and requiring specific expertise, and therefore may be best suited as a service provided by dedicated labs to the forensic community. However, in order to reduce costs and possible backlogs, and to reduce the risks of errors related to lengthening of the chain of custody, it may be advantageous to develop an integrated DNA authentication assay that will be performed in existing forensic laboratories, as part of the regular forensic procedure. The question of integrating DNA authentication into the forensic procedure also has legal aspects, and therefore we hope this work will invoke a discussion in legal as well as in scientific circles.

#### 4.5. DNA mixtures

Artificial DNA evidence can contain “pure” artificial DNA or a mixture of artificial and natural DNA. For example, such a mixture may contain artificial DNA incorporated in or applied onto genuine tissues from the victim (e.g. blood, fingernails). As we demonstrated, mixture samples can be authenticated in the same manner as single source samples, and the artificial DNA is detected even when it is a minor component of the mixture. The automatic software used for sequencing assigns a nucleotide at a certain position when the template is pure or contains a major component, and outputs ‘N’ when there is ambiguity. Therefore, artificial DNA can be detected automatically in samples when it is the major component, using existing sequencing software. However, the interpretation of mixtures in which the artificial DNA is a minor component is more complex and may require the development of guidelines, similar to those that have been suggested for profiling [26].

#### 4.6. Other approaches to DNA authentication

Analysis of methylation patterns represents only one of several possible approaches that can be used for DNA authentication. Alternative methods may be based on analysis of stutter products, representation bias, distribution of DNA fragment sizes, and presence of non-genomic sequences. Stutter products are artifacts caused by slippage of the DNA polymerase on repeated sequences [27]. Upon profiling, artificial DNA that was pre-amplified by PCR undergoes more amplification cycles than natural DNA (pre-amplification also includes several cycles of PCR) and because a higher number of cycles is associated with elevated stutter levels [28], such pre-amplified DNA may be distinguished from natural DNA by higher stutter percentages. Representation bias refers to differences in copy number between different genomic loci that are an inherent consequence of *in vitro* amplification of DNA [11]. Since this bias might not necessarily be apparent in the small set of CODIS loci used for profiling, it may be necessary to analyze a wider set of loci. Analysis of the distribution of fragment sizes can also reveal the origin of the DNA: in natural DNA, the distribution has an expected stereotypical pattern (which is a function of the extraction method used and the extent of degradation), different from the patterns observed in various types of *in vitro* synthesized DNA. Non-genomic sequences such as primer dimers, plasmid sequences, artificial oligonucleotide linkers, etc., are not expected to be found in natural DNA (with the possible exception of bacterial sequences), but are expected to be found in various types of *in vitro* synthesized DNA.

### 5. Conclusion

In this work we address the disturbing possibility that DNA evidence can be faked and planted in crime scenes, and the current inability of the forensic procedure to detect such artificial evidence. We present a solution to this problem in the form of a DNA authentication assay that can distinguish between natural and artificial DNA evidence. In order to preserve the high credibility of DNA evidence in the courtroom, if there is a concern regarding the possible authenticity of DNA evidence, an approach such as presented here may be taken to test for the presence of artificial DNA.

#### Acknowledgements

We thank Howard Cedar for his insights on methylation analysis, Eran Eldar for performing blood extraction, Shira Grafit for help with the figures, and Avraham Levy, Yuval Dor, Yaakov Benenson, Eliahu Wasserstrom, Elon Ganor, and Meirav Chovav for critical review of this manuscript.

#### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.fsigen.2009.06.009.

#### References

- [1] J.M. Butler, Forensic DNA Typing. Biology, Technology and Genetics of STR Markers, 2nd ed., Elsevier Academic Press, Burlington, 2005.
- [2] A.J. Jeffreys, V. Wilson, S.L. Thein, Individual-specific ‘fingerprints’ of human DNA, *Nature* 316 (1985) 76–79.
- [3] M.A. Jobling, P. Gill, Encoded evidence: DNA in forensic analysis, *Nat. Rev. Genet.* 5 (2004) 739–751.
- [4] N. Morling, Forensic genetics, *Lancet* 364 (Suppl. 1) (2004) s10–s11.
- [5] P. Reilly, Legal and public policy issues in DNA forensics, *Nat. Rev. Genet.* 2 (2001) 313–317.
- [6] A. Opar, Crime and punishment, *Nat. Med.* 12 (2006) 1110–1111.



- [7] M. Lynch, God's signature: DNA profiling, the new gold standard in forensic science, *Endeavor* 27 (2003) 93–97.
- [8] M.J. Saks, J.J. Koehler, The coming paradigm shift in forensic identification science, *Science* 309 (2005) 892–895.
- [9] D. Stirling, J.M.S. Bartlett (Eds.), *PCR Protocols (Methods in Molecular Biology)*, 2nd edition, Humana Press, New Jersey, 2003.
- [10] J. Sambrook, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York, 2001.
- [11] R.S. Lasken, M. Egholm, Whole genome amplification: abundant supplies of DNA from precious samples or clinical specimens, *Trends Biotechnol.* 21 (2003) 531–535.
- [12] S. Panelli, G. Damiani, L. Espen, G. Micheli, V. Sgarabella, Towards the analysis of the genomes of single cells: further characterisation of the multiple displacement amplification, *Gene* 372 (2006) 1–7.
- [13] T.B. Miranda, P.A. Jones, DNA methylation: the nuts and bolts of repression, *J. Cell Physiol.* 213 (2007) 384–390.
- [14] T. Hashimshony, J. Zhang, I. Keshet, M. Bustin, H. Cedar, The role of DNA methylation in setting up chromatin structure during development, *Nat. Genet.* 34 (2003) 187–192.
- [15] A. Bird, DNA methylation patterns and epigenetic memory, *Genes Dev.* 16 (2002) 6–21.
- [16] F.B. Dean, S. Hosono, L. Fang, X. Wu, A.F. Faruqi, P. Bray-Ward, Z. Sun, Q. Zong, Y. Du, J. Du, M. Driscoll, W. Song, S.F. Kingsmore, M. Egholm, R.S. Lasken, Comprehensive human genome amplification using multiple displacement amplification, *Proc. Natl. Acad. Sci. U.S.A.* 99 (2002) 5261–5266.
- [17] E.F. Fritsch, T. Maniatis, J. Sambrook, *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, New York, 1989.
- [18] J.M. Butler, R. Schoske, P.M. Valone, J.W. Redman, M.C. Kline, Allele frequencies for 15 autosomal STR loci on U.S. Caucasian, African American, and Hispanic populations, *J. Forensic Sci.* 48 (2003) 908–911.
- [19] M. Frommer, L.E. McDonald, D.S. Millar, C.M. Collis, F. Watt, G.W. Grigg, P.L. Molloy, C.L. Paul, A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands, *Proc. Natl. Acad. Sci. U.S.A.* 89 (1992) 1827–1831.
- [20] J.W. Bond, C. Hammond, The value of DNA material recovered from crime scenes, *J. Forensic Sci.* 53 (2008) 797–801.
- [21] M. Levitt, Forensic databases: benefits and ethical and social costs, *Br. Med. Bull.* 83 (2007) 235–248.
- [22] B. Sobrino, M. Brion, A. Carracedo, SNPs in forensic genetics: a review on SNP typing methodologies, *Forensic Sci. Int.* 154 (2005) 181–194.
- [23] 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.
- [24] K. Babol-Pokora, J. Berent, SNP-minisequencing as an excellent tool for analysing degraded DNA recovered from archival tissues, *Acta Biochim. Pol.* 55 (2008) 815–819.
- [25] H. Nakahara, N. Hosono, T. Kitayama, K. Sekiguchi, M. Kubo, A. Takahashi, Y. Nakamura, Y. Yamano, K. Kai, Automated SNPs typing system based on the invader assay, *Leg. Med.* 11 (2009) S111–S114.
- [26] P. Gill, C.H. Brenner, J.S. Buckleton, A. Carracedo, M. Krawczak, W.R. Mayr, N. Morling, M. Prinz, P.M. Schneider, B.S. Weir, DNA commission of the International Society of Forensic Genetics: recommendations on the interpretation of mixtures, *Forensic Sci. Int.* 160 (2006) 90–101.
- [27] D. Shinde, Y. Lai, F. Sun, N. Arnheim, Taq DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: (CA/GT)<sub>n</sub> and (A/T)<sub>n</sub> microsatellites, *Nucleic Acids Res.* 31 (2003) 974–980.
- [28] L. Forster, J. Thomson, S. Kutranov, Direct comparison of post-28-cycle PCR purification and modified capillary electrophoresis methods with the 34-cycle “low copy number” (LCN) method for analysis of trace forensic DNA samples, *Forensic Sci. Int. Genet.* 2 (2008) 318–328.