Forensic Science International: Genetics 6 (2012) 341–349

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

A new SNP assay for identification of highly degraded human DNA

A. Freire-Aradas [a], M. Fondevila [a], A.-K. Kriegel [b], C. Phillips [a,*], P. Gill [c,d], L. Prieto [e], P.M. Schneider [b], Á. Carracedo [a], M.V. Lareu [a]

[a] Forensic Genetics Unit, Institute of Legal Medicine, University of Santiago de Compostela, Spain
[b] Institute of Legal Medicine, University Hospital, University of Cologne, Germany
[c] Institute of Legal Medicine, University of Oslo, Rikshospitalet, Norway
[d] Forensic Science Centre, University of Strathclyde, Glasgow, UK
[e] University Institute of Research Police Sciences (IUICP), DNA Laboratory, Comisaría General de Policía Científica, Madrid, Spain

ARTICLE INFO

Article history:
Received 14 April 2011
Received in revised form 7 July 2011
Accepted 8 July 2011

Keywords:
Nucleosomes
Degraded DNA
SNPs
Single nucleotide polymorphisms
Human identification

ABSTRACT

There is growing evidence that the histone–DNA complexes found in nucleosomes offer protection from DNA degradation processes, including apoptotic events in addition to bacterial and environmental degradation. We sought to locate human nucleosome regions and build a catalogue of SNPs sited near the middle of these genomic segments that could be combined into a single PCR multiplex specifically for use with extremely degraded human genomic DNA samples. Using recently optimized bio-informatics tools for the reliable identification of nucleosome sites based on sequence motifs and their positions relative to known promoters, 1395 candidate loci were collected to construct an 18-plex single base extension assay. Genotyping performance of the nucleosome SNPs was tested using artificially degraded DNA and 24 casework samples where the likely state of degradation of DNA was established by comparison to profile completeness in four other forensic assays: a standard 15-plex STR identification test, a miniaturized STR multiplex and two autosomal SNP multiplexes. The nucleosome SNP assay gave genotyping success rates 6% higher than the best existing forensic SNP assay: the SNP*for*ID Auto-2 29-plex and significantly higher than the mini-STR assay. The nucleosome SNPs we located and combined therefore provide a new type of marker set that can be used to supplement existing approaches when the analysed DNA is likely to be extremely degraded and may fail to give sufficient STR genotypes for a reliable identification.

1. Introduction

Highly degraded DNA presents a major challenge to the standard identification markers available for forensic analyses; though shortening the amplified fragments generated in PCR markedly improves genotyping success. The rate of DNA degradation is accelerated by the effect of environmental factors including temperature, humidity, ultraviolet radiation, pH, presence of microorganisms and the localized geochemical properties of the soil [1]. All these factors have a greater bearing on the condition of DNA than the time since deposition or death [2,3]. Chemical reactions affecting DNA stability and consequently PCR efficiency, can be categorized into three groups: hydrolysis leading to base loss [4], oxidation leading to base modification [5] and single/double strand breakage [6]. Post mortem, a corpse is subject to the action of a range of bacterial enzymes originating from the gastrointestinal tract and from the immediate environment. The principal catalytic activity of bacterial enzymes is to cleave DNA to generate a pool of small oligonucleotides where average fragment sizes and their range of ~80–200 base pairs (bp) fall within most forensic markers' inter-primer lengths and therefore compromise PCR amplification efficiency. Short Tandem Repeats (STRs) represent the first-choice markers for forensic identification due in large part to their high discrimination power [7,8]. However STR analysis of highly degraded samples is often inadequate in terms of profile completeness and this compromises the discrimination power that can be expected from genotyping of these markers alone [9,10]. The need to decrease amplicon sizes to the smallest possible amplifiable fragments has led to the development of several alternative marker sets specifically aimed at analyzing highly degraded DNA. These include: mini-STRs [11,12], indels (insertion/deletion polymorphisms) [13] and single nucleotide polymorphism (SNPs) [14–16]. SNPs offer ideal candidate loci for typing degraded DNA due to their simplified binary polymorphisms that allow large-scale multiplexing as well as their obvious potential for designing PCR amplicon sizes in a feasible range of 50–120 bp. SNPs have several additional advantageous characteristics: they are highly abundant in all regions of the human genome

* Corresponding author.
E-mail address: c.phillips@mac.com (C. Phillips).

1872-4973/$ – see front matter © 2011 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.fsigen.2011.07.010

and well characterized [17,18] - making multiplex development easier; their low mutation rate of 10^{-8} compared with 10^{-3} for STRs makes them informative and reliable supplements in relationship testing [19] and; they are adaptable to analysis using high throughput technologies [20].

Apoptosis or programmed cell death is a natural process involving a form of cell death in an active and controlled manner that deletes unwanted cells. Although the apoptotic process is complex it has been extensively studied and described [21]. Cells undergoing apoptosis show typical, well-defined morphological and biochemical changes [22,23]. Recently, it has become apparent that cellular necrosis is as equally controlled and programmed as apoptosis. Necrotic cell death is not the result of one signaling cascade but the consequence of extensive crosstalk between several biochemical and molecular events at different cellular levels [24]. The condensation of nuclei, in addition to the cleavage of chromosomal DNA, is one of the major indicators of apoptosis [25]. DNA in apoptotic cells is specifically targeted and degraded, resulting in a ladder of multiple fragments in ~200 bp steps. This ladder is a consequence of the digestion of chromatin by an endogenous endonuclease that targets the linker DNA between portions of the nucleosome [26]. The nucleosome is the basic structural unit of eukaryotic chromatin [27] comprising a repeating unit of eight histone molecules and approximately 200 nucleotides [28]. Detailed X-ray crystallography has refined this structural detail to further describe the nucleosome core particle as 146 bp of DNA wrapped around a histone octamer (two dimers H2A–H2B and a tetramer H3–H4) in 1.65 turns of a flat, left-handed superhelix [29]. In apoptotic cells DNA sequences interacting with histones escape enzymatic cleavage, therefore it has been widely suggested that this interaction has a protective effect localized at nucleosome positions and described as the theory of nucleosome protection. It can also be inferred that the same effect is likely to occur in cells undergoing necrosis.

The protective effect of histone–DNA interactions sited in nucleosomes provided the core basis for the study we report here, which sought to locate, characterize and combine SNPs sited in nucleosomic regions - identified as such with a high probability. Identification and cataloguing of nucleosome sited SNPs then enabled us to create a new multiplex of SNPs for forensic identification that was likely to benefit from greater resistance to degradation as well as from very short amplicons. The developed multiplex comprises 18 autosomal SNPs from nucleosomic regions genotyped in a single-tube assay followed by a mini-sequencing reaction based on SNaPshot™ primer extension. A comparison of performance typing highly degraded DNA was made between the nucleosome–SNP multiplex and other core forensic marker sets that included: AmpFℓSTR Identifiler™, mini-STRs of AmpFℓSTR® MiniFiler™ and the autosomal identification SNP (ID-SNP) sets of SNP*for*ID [14] using both artificially degraded DNA and real casework samples.

2. Material and methods

2.1. Samples, DNA extraction and quantification

A total of 27 artificially degraded exonuclease treated samples were made comprising three donors incubated for nine time intervals. Prior to enzymatic degradation the control samples were typed with each multiplex. At each time point the exonuclease was inactivated then DNA concentrations were determined using the Quantifiler™ Human DNA Quantification Kit with the AB 7300 real-time PCR systems (Applied Biosystems: AB).

To assess the sensitivity of the nucleosome SNP multiplex, serial dilutions were made of control DNA from a single donor individual in duplicate in a dilution series to give: 10 ng/μl, 5, 2.5, 1.25 and 625 pg/μl, 312, 156, 78, 36 and 18.

Twenty-four casework samples were assessed across all genotyping systems appropriate for challenging DNA typing. Cases comprised: eight post-mortem muscle portions; nine telogen hairs, mainly short fragments; DNA from four toothbrushes; one razor blade; one cigarette butt and; one contact lens. Casework samples were chosen initially on the basis of the observed Identifiler STR profile quality. Some additional regard was made to the quantifications obtained and the normal expectations for certain samples such as toothbrushes that usually can be expected to provide good results, but in the chosen cases did not. Overall, a wide range of casework material was originally examined before selecting appropriate test samples. Degraded DNA controls were extracted with the QIAamp DNA Micro Kit (Qiagen), and casework samples with the Qiagen EZ-1 robot and EZ-1 DNA Investigator kit, while 5 of 9 hairs were extracted by standard Chelex bead protocols. In all cases negative controls were extracted in parallel. Three separate amplifications were made of each sample with each multiplex apart from the hairs (single analyses due to shortage of target DNA) and typing with Identifiler (single initial analysis). Genotyping performance was assessed by recording locus dropout rates and allele dropout rates (the latter by reference to the consensus genotypes from three runs). At the time of the casework tests SNP set Auto-1 comprised 20 of 23 SNPs (markers rs1886510, rs722098 and rs2016276 gave inconsistent results with a range of positive controls so were excluded) and Auto-2 comprised 28/29 SNPs (SNP rs1024116 excluded). Analytical thresholds for STRs were 50 RFU while no heterozygote peaks were imbalanced beyond a 60% limit. For SNP typing standard blue:-green:yellow/red peak height ratios of 4:2:1:1 were used to detect and assign SNP alleles [14]. We recorded the locus and allele dropout rates for each component marker and these were used to gauge individual performance of markers then make a collective assessment of each multiplex across 24 typical cases representative of challenging DNA.

2.2. Location of potential nucleosomic regions in human promoter sequences

From a set of 465 promoter regions that presented 3 or more well positioned nucleosomes in at least four different cell types [30], human promoter sequences were obtained by using the *Homo sapiens Promoter Database* [31]. Potential nucleosomic regions were recognized by using the bioinformatics software RECON [32–34]: a program designed for constructing profiles of nucleosome forming potential in the human genome by characterizing the probability of nucleosome formation along the DNA sequence analysed. From sequence searches spanning 1500 bp segments at each promoter region locations corresponding to the three highest nucleosome forming probabilities in any one region were selected as potential nucleosomic sites.

2.3. SNP selection

The NCBI dbSNP database was used to locate and scrutinize SNPs positioned within the selected nucleosomic regions [35]. Selection criteria for ideal forensic candidate SNPs comprised: location inside a potential nucleosomic region with preference given to closest proximity to the middle of the nucleosome; a lower limit of heterozygosity of 0.2 in at least two of the three major population groups of Africa, Europe and East Asia; proper validation status and, as a set, a broad genomic distribution that did not position candidates too closely to previous identification SNP or STR locations. With these criteria applied, twenty nucleosomic SNPs were selected to create a working developmental multiplex to study genotyping performance with highly degraded DNA.

2.4. Development of PCR and extension primers

PCR and extension (SBE) primer design used *Primer 3* software [36] and *NetPrimer* [37] following the guidelines from Sanchez and Endicott [38]. The amplicon lengths ranged between 56 and 118 bp and the theoretical melting temperature (T_m) around 60 °C ± 2 °C (SNP rs6763138 was the only T_m outlier with a predicted melting temperature of 62.7 °C). Primers were checked for primer–dimer formation and hairpin structures using *Autodimer* [39]. Supplementary Table S1 shows the sequences and the ratio of concentrations of the amplification primers in the final multiplex PCR. The lengths of the SBE primers ranged between 17 and 76 nucleotides and specific lengths were tailored using poly-CT tails. Supplementary Table S1 lists the sequences and the ratio of concentrations of the SBE primers in the extension reaction.

2.5. PCR conditions and purification of PCR products

PCR optimization was carried out following the guidelines proposed by Sanchez and Endicott [38] plus those of Henegariu et al. [40]. PCR amplification conditions were: 2 μl of DNA (1–10 ng/μl DNA) in a 9.8 μl reaction volume containing 1.25 μl of 10× PCR buffer without MgCl2, 1.25 μl of 1.6 μg/ml bovine serum albumin, 3.25 μl 25 mM MgCl2, 0.875 μl of 10 mM dNTPs, 0.2 μl of 5 U AmpliTaq Gold® (all AB), and 3 μl of the PCR primer mix. Amplification was performed in an AB GeneAmp® 9700 thermal cycler with the following cycle program: denaturation at 95 °C for 10 min followed by 35 cycles of 95 °C for 30 s, 61 °C for 50 s, 65 °C for 30 s, then a final extension at 65 °C for 6 min. Excess primers and dNTPs were removed by adding 1 μl ExoSAP-IT (1 U/μl Exonuclease I and Shrimp Alkaline Phosphatase, GE Healthcare) to 2.5 μl PCR product and incubation at 37 °C for 45 min and 85 °C for 15 min.

2.6. Single base extension and SNP allele detection

Single base extension reactions were performed in a final volume of 6 μl containing 2.5 μl of SNaPshot™ reaction mix (AB), 1.5 μl of SBE primer mix and 2 μl of purified PCR product. The SBE primer mix was diluted in 160 mM ammonium sulphate to avoid non-specific hybridizations amongst the primers. The SBE reaction was performed in an AB 9700 thermal cycler with the following cycle program: 30 cycles of 96 °C for 10 s, 55 °C for 5 s and 60 °C for 30 s. Excess nucleotides were removed by addition of 1 μl SAP (1 U/μl Shrimp Alkaline Phosphatase, GE Healthcare) to the total volume of the extension products and incubation at 37 °C for 80 min and 85 °C for 15 min.

A combination of 3 μl of sample, 9.5 μl LIZ 120 size standard plus HiDi formamide at a ratio of 1:33.3 (both AB) was analysed by capillary electrophoresis using an AB 3130 Genetic Analyzer with POP4 or POP6 polymer and analysed with GeneMapper v4.0. Predefined size windows for each allele were determined from prior analysis of a minimum of 20 samples for both polymers.

2.7. Preparation of artificially degraded DNA

The protocol of Timken et al. [41] was modified to create a series of progressively increasing levels of degraded DNA in a set of control samples. Firstly 200 μl of deionized sterile water was added to 75 μl of whole blood samples from three individuals and kept overnight at room temperature. Then 234 μl of the resulting cell lysates were taken and combined with 26 μl of reaction buffer (400 mM Tris–HCl pH 8, 100 mM $MgSO_4$ and 10 mM $CaCl_2$). Aliquots of 30 μl were made and 5.6 μl of 0.1 U/μl micrococcal nuclease (MNase, GE Healthcare) added. This enzyme specifically targets regions of DNA linkers and is therefore able to cleave DNA segments between nucleosomes. Samples were removed from incubation at room temperature at intervals: 3 h, 10 h, 24 h, 3 days, 7 days, 2 weeks, 9 weeks, 4 months and 7 months. MNase activity was quenched by adding 6 μl of 20 mM EDTA and heating at 85 °C for 15 min. After quenching to inactivate the exonuclease, samples were run on standard agarose check gels to confirm a smoothly graded smear of multiple sized fragments had been achieved in each case.

2.8. Autosomal STR and alternative ID-SNP genotyping

Standard protocols were followed to type the core STR sets of AmpFℓSTR® Identifiler™ and AmpFℓSTR® MiniFiler™ (AB). The 52 SNP*for*ID autosomal ID-SNPs were typed following the protocols for a combined 23 SNP set (herein Auto-1) and 29 SNP set (herein Auto-2) amplified in a single initial 52-plex PCR followed by two tandem extension reactions as outlined by Sanchez et al. [14]. Previously we did not record significant performance differences between a single PCR compared to a split 23-plex and 29-plex PCR [15,42]. Therefore we decided to amplify 52 SNPs in one PCR as this most closely matches the approach dictated by scarcity of casework material where input of a small extract volume to a single amplification is better than dividing the extract across two reactions.

2.9. Forensic statistical informativeness metrics

Cumulative random match probabilities and discrimination indices were calculated using in-house calculators for each SNP combination based on complete genotype profiles obtained from 1000 Genomes SNP data (Phase I interim data release, December 2010) of 90 Europeans (CEU), 78 Africans (YRI) and 68 East Asians (CHB). Nucleosome SNP: rs2316213 does not have genotype data in 1000 Genomes so HapMap data from the same populations was used.

3. Results

3.1. Final selection of nucleosome SNPs and development of an optimized multiplex

A total of 465 promoter regions each with 3 potential nucleosomic sites were screened for SNP content. From the resulting catalogue of 1395 candidate SNPs twenty loci matched the strict criteria used for marker selection. All candidate loci were successfully amplified in singleplex PCR to allow checks for single product peaks corresponding to predicted sizes and to ensure an absence of artefact peaks and self-extension prior to multiplex optimization. Two candidate SNPs: rs2277121 and rs2071457 were removed from the multiplex reaction due to repeated failure to amplify efficiently in combination with the others. Fig. 1 shows an example of the resulting optimized 18-plex nucleosome SNP typing assay.

Allele frequencies for the 18 nucleosome SNPs are shown in Fig. 2 based on 1000 Genomes genotype data for African, European and East Asian populations, with a single marker not characterized in this database: rs2316213, based on HapMap data.

3.2. Sensitivity of nucleosome SNP multiplex

All samples in the dilution series down to an estimated DNA concentration of 78 pg/μl gave full profiles for the nucleosome assay while the lowest concentrations of 36 and 18 pg/μl showed allele dropout (6/36 and 14/36 = 16.7% and 38.9% respectively) and to a lesser extent, allele dropin (1/36 and 3/36 = 2.8% and 8.3% respectively).

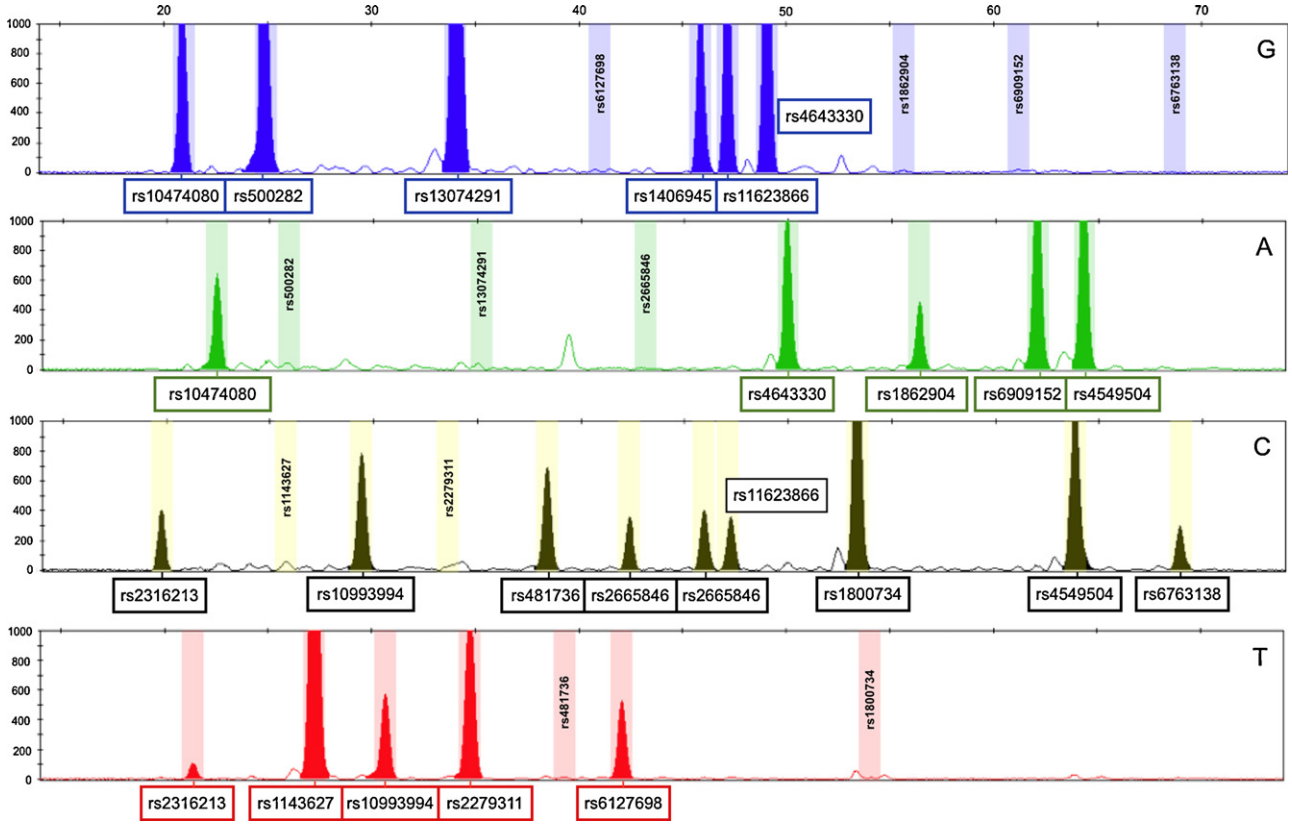

Fig. 1. Typical SNaPshot™ electropherogram of the nucleosome SNP multiplex typing 1 ng of DNA. Missing peak positions are marked as coloured size panels with vertical rs-number SNP identifiers superimposed.

3.3. Typing performance of artificially degraded DNA amongst marker sets

Artificially degraded micrococcal nuclease treated samples from three individuals were amplified with standard marker choices for degraded DNA: Identifiler™; MiniFiler™; and; SNP*for*ID Auto1/2, and compared to the nucleosome SNP multiplex. Since these represent a range of marker numbers, from 8 to 52, it was important to arrange a suitable framework for comparing performance that took account of overall success per multiplex as well as differences that could occur between the three DNA samples used for assessment of sensitivity to degradation. We decided to create heatmap plots based on the proportion of locus dropout observed in each case. This same method of assessment could then be extended to measuring performance with casework material, since a common problem in such comparative frameworks is to properly gauge the actual degree of degradation of the target DNA obtained from a casework sample of uncertain history. If the summary dropout rate is measured – i.e. across all genotype assays used – the casework DNA quality can then be ranked in the

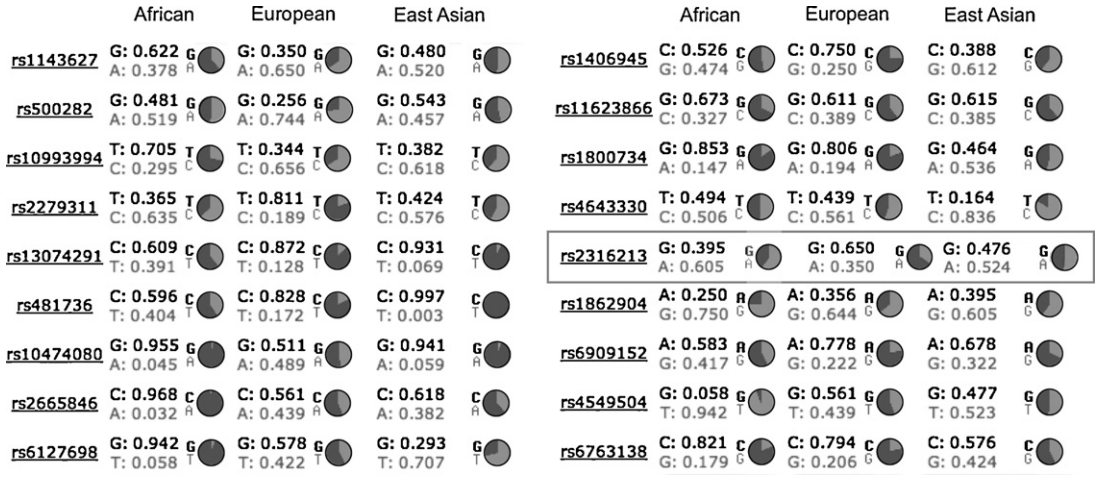

Fig. 2. Allele frequency distributions in three population groups, collected from combined population data of 1000 Genomes, Africans: Luhya in Webuye, Kenya and Yoruba in Ibadan, Nigeria; Europeans: Finnish in Finland, British in England and Scotland, CEPH Utah residents with N & W European ancestry and Toscans in Italy; East Asians: Han Chinese in Beijing, Han from Southern China and Japanese in Tokyo. The pie-charts of rs2316213 marked with a grey box represent data collected for a single SNP without 1000 genomes data, taken from HapMap equivalent populations.

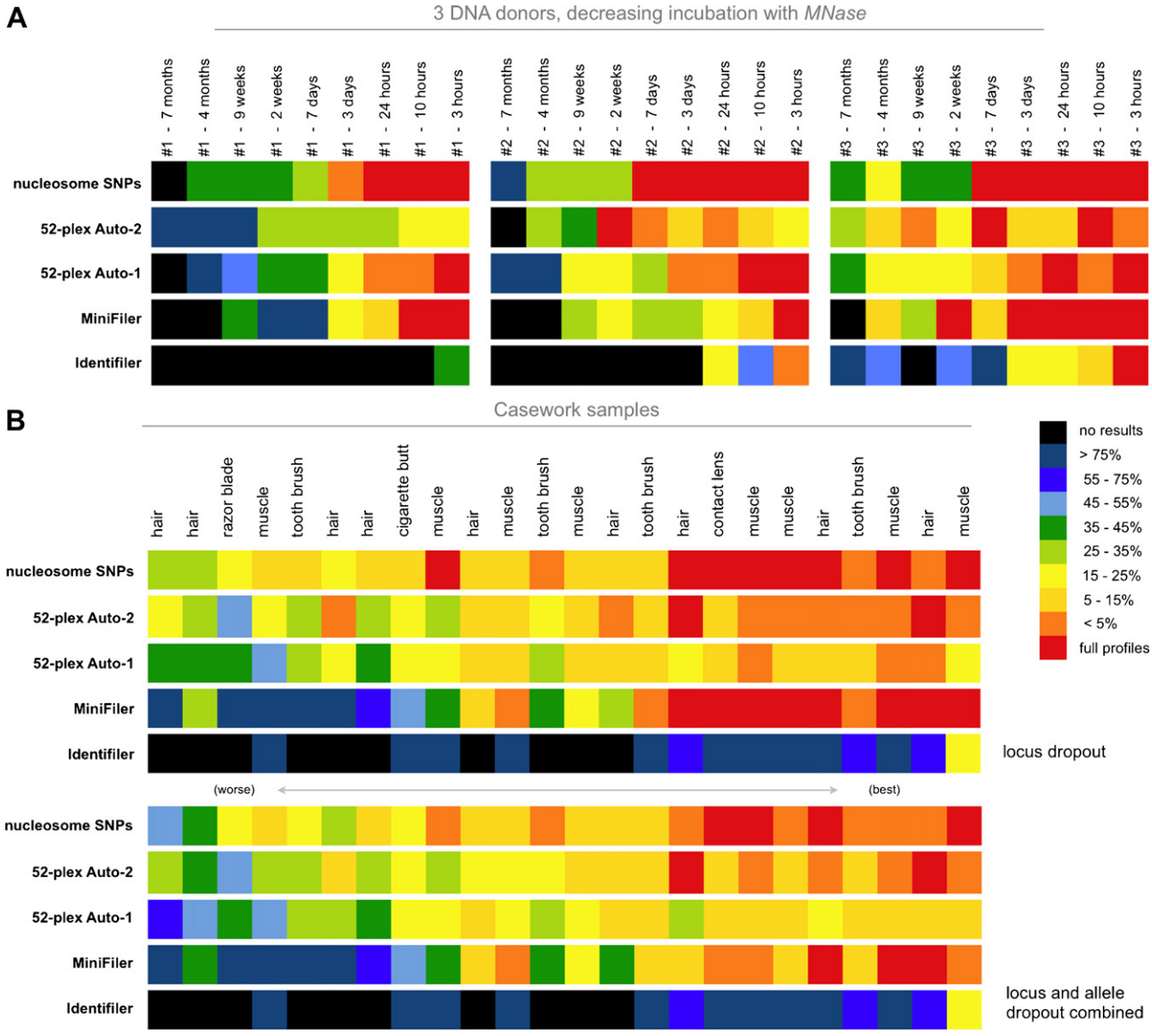

Fig. 3. Heatmap representations of genotyping success using five multiplexes. (A) MNase degraded controls arranged by donor and in decreasing incubation time left to right. 52-plex describes the normal SNP numbers of the combined Auto1/2 PCR and does not equate to the 20 + 28 SNP data shown. (B) 24 casework samples arranged in descending order of likely state of degradation in the extracted DNA from worst: least successful genotyping, most degraded DNA on the left, and best: most successful, least degraded on the right. Locus dropout rates compared in the top chart, allele and locus dropout rates combined in the bottom chart.

same way, in an order that properly reflects likely DNA quality: from most severely degraded to least degraded. For simplicity, the amelogenin component of Identifiler™ and MiniFiler™ was not included in success counts. Once a suitable ranked order for the degree of degradation is established, comparisons between alternative genotyping approaches are both more accurate and better reflect the final genotype information that could be obtained. However typing performance measured in this way still only represents total loci successfully typed, not information per locus. This latter characteristic is a common problem when comparing the total number of binary SNPs successfully typed with the total multi-locus STRs typed. Therefore for assessment of casework DNA we applied a simple guideline of information content for 'SNPs per STR' in forensic identification use suggested by Charles Brenner [43]. For identification (not paternity) applications this equates to a SNP to STR ratio of approximately 2.5 to 1, assuming perfect 0.5:0.5 allele frequency SNPs, therefore a partial profile of five SNPs would be roughly twice as informative as a partial profile of one STR.

Fig. 3A shows the heatmap genotyping summary indicating the enzymatic effect of micrococcal nuclease, ordered by time of degradation. Each column represents a time period of nuclease incubation, arranged in three samples sets. The colours are skewed into cold blue-green for high dropout and hot orange-red for low or zero locus dropout. Rows have been arranged in descending order of multiplex performance such that a trend of blue lower left to red upper right is discernable. The order of marker sets being: nucleosome SNP multiplex, 29-plex Auto-2, 23-plex Auto-1, MiniFiler™ and Identifiler™. Better performance is observed for the nucleosome SNP multiplex while the first markers to fail are STRs. The performance of Auto-1 and Auto-2 SNPs can be interpreted as comparable, though previous experience suggests that Auto-2 performs slightly better when typing degraded DNA than Auto-1, so this same order of multiplex sets was kept for arranging casework genotyping success. Some difference is also discernable between DNA donors: sample 1 is more affected by enzymatic cleavage while sample 3 is more resistant. This could be interpreted as indicating some variation in resistance to DNA

degradation exists between individuals, although qualitative differences between the three blood samples cannot be ruled out.

3.4. Typing of degraded DNA in casework samples

Genotyping success from the analysis of the 24 degraded casework samples are summarized in Fig. 3B using the same blue to red heat-map approach outlined in Section 3.3 to denote a ranked order of overall success, with the underlying data summarized in the top half of Table 1. All amplifications apart from those made from hair samples were triplicated, so as well as locus dropouts, allele dropouts could be detected by comparison to reference and consensus genotypes. We recorded overall allele and locus dropout rates per case and per multiplex and used these values to establish an order of likely state of degradation in the casework DNA comparable to the trends shown in Fig. 3A. Therefore both Fig. 3A and B show a more strongly differentiated overall success rate discernable in the nucleosome SNPs and Auto-2 – particularly in the most degraded casework DNA analyses on the left of Fig. 3B. All SNP multiplexes demonstrate greater success than use of MiniFiler™ while this shortened-amplicon STR set offers better chances of success than the comparable conventional STR multiplex of Identifiler™. This applies across a broad range of likely states of degradation, as summarized by the average percentage of markers successfully typed in each multiplex, shown on the right of Table 1, in descending order: nucleosome SNPs 87%; Auto-2 81.4%; Auto-1 73.8%; MiniFiler™ 63.3%.

In order to gauge individual component SNP performance each marker was assessed for total number of allele and locus dropouts across the 24 casework analyses. The ranked locus dropout rates for each short amplicon marker set are shown in Fig. 4 with allele dropout rates (for the same order of component loci) in the inverted plot below. The underlying data for the plots of Fig. 4 is given in Supplementary Data Table S2. Fig. 4 indicates that locus dropout rates vary much more than allele dropout rates and are almost one order of magnitude higher for the worse performing loci. There were also differences in performance amongst the SNP sets. Nucleosome SNPs have an average dropout rate of 7%: half the 14% of Auto-2 and a third of the 18.7% of Auto-1. Outlier SNPs with higher than average dropout rates are evident in all SNP sets on the left-most side of the ranked SNP lists underneath the plot. These indicate rs2665846 and rs11623866 of nucleosomes; rs1357617, rs719366 and rs917118 of Auto 1 and; rs914165 of Auto 2 fail most readily when typing very degraded DNA. Since these values are averaged across 24 challenging casework samples they provide indications of the component markers most likely to fail when typing highly degraded DNA.

One important qualification that should be made here is that, since this study was made, STR kit formulations have changed to incorporate miniaturized amplicon primer sets and anti-inhibition components. This will improve the performance of standard forensic identification loci in such cases. However SNP analysis is likely to remain a realistic option for cases with extremely degraded DNA where all genotypes obtained provide valuable data to help achieve an unequivocal identification of the contributor.

As well as demonstrating a trend in overall success between marker sets we summarized the forensic identification information content that can be expected from each multiplex using the 2.5 SNPs-per-STR ratio. The information provided by any one result as 'STR equivalents' is shown in the lower half of Table 1. The average 'STR equivalent' values across all casework results are also shown and despite the range of DNA quality amongst cases, these average values best summarize the final discrimination power from an average case profile that could be expected when opting for a particular multiplex strategy. Here the 29 loci of Auto-2 provides the obvious best combination of information and success,

Table 1
Genotyping success of 24 casework samples measured as detected genotypes in each of five multiplexes. Equivalent informativeness for each recorded profile is given in the lower half of the table. 'STR equivalents' equate the informativeness of 2.5 SNP genotypes to one STR genotype in identification applications.

Success: total loci typed	Loci in multiplex	Hair	Hair	Razor blade	Razor blade	Muscle	Tooth brush	Hair	Hair	Cigarette butt	Muscle	Hair	Muscle	Tooth brush	Muscle	Hair	Tooth brush	Hair	Contact lens	Muscle	Muscle	Hair	Tooth brush	Muscle	Hair	Muscle	Average % success rate	Average total loci successfully typed
Nucleosome SNPs	18	7	10	13	16	15	14	16	14	17	15	15	17	17	15	16	17	18	18	17	18	17	18	18	18	18	87.0	16
Auto-2	28	18	16	14	21	19	17	21	21	20	23	23	24	23	25	24	28	25	28	25	26	25	26	28	28	28	81.4	23
Auto-1	20	4	8	13	11	13	13	13	16	14	17	16	16	14	16	17	15	17	18	16	16	16	16	17	18	18	73.8	15
AB	8	0	3	1	2	0	0	3	5	5	8	9	7	5	5	8	8	8	8	8	8	8	9	8	9	9	63.3	6
MiniFiler™																												
AB Identifiler™	15	0	0	0	2	0	0	0	2	1	0	1	0	0	0	2	4	2	2	3	3	3	7	2	4	12	12.5	2
Informativeness: total 'STR equivalents' typed	Loci in multiplex	Hair	Hair	Razor blade	Muscle	Tooth brush	Hair	Hair	Cigarette butt	Muscle	Hair	Muscle	Tooth brush	Muscle	Hair	Tooth brush	Hair	Contact lens	Muscle	Muscle	Hair	Tooth brush	Muscle	Hair	Muscle	Average		
Nucleosome SNPs	7	2.8	4.0	5.3	6.4	5.9	5.6	6.3	5.7	6.8	6.1	6.0	6.8	6.8	6.8	6.0	6.3	6.8	7.2	7.2	6.9	7.2	6.9	7.1	7.0	7.2	6.3	
Auto-2	12	7.2	6.4	5.6	8.3	7.5	6.8	8.3	8.5	8.0	9.1	9.2	9.2	9.5	9.5	10	9.7	11.2	10.0	11	10.1	10.4	10.1	10.5	11.2	11.1	9.1	
Auto-1	9	1.6	3.2	5.1	4.5	5.2	5.2	5.1	6.5	5.7	6.8	6.5	5.5	6.4	6.4	6.4	6.8	6.0	6.8	7.3	6.5	6.5	6.7	7.1	7.2	7.1	5.9	
AB	8	0	3	1	2	0	0	3	5	5	8	9	5	7	5	5	8	8	8	8	8	9	8	9	9	9	5.7	
MiniFiler™																												
AB Identifiler™	15	0	0	0	2	0	0	0	2	1	0	1	0	0	0	0	2	4	2	2	3	3	7	2	4	12	1.9	

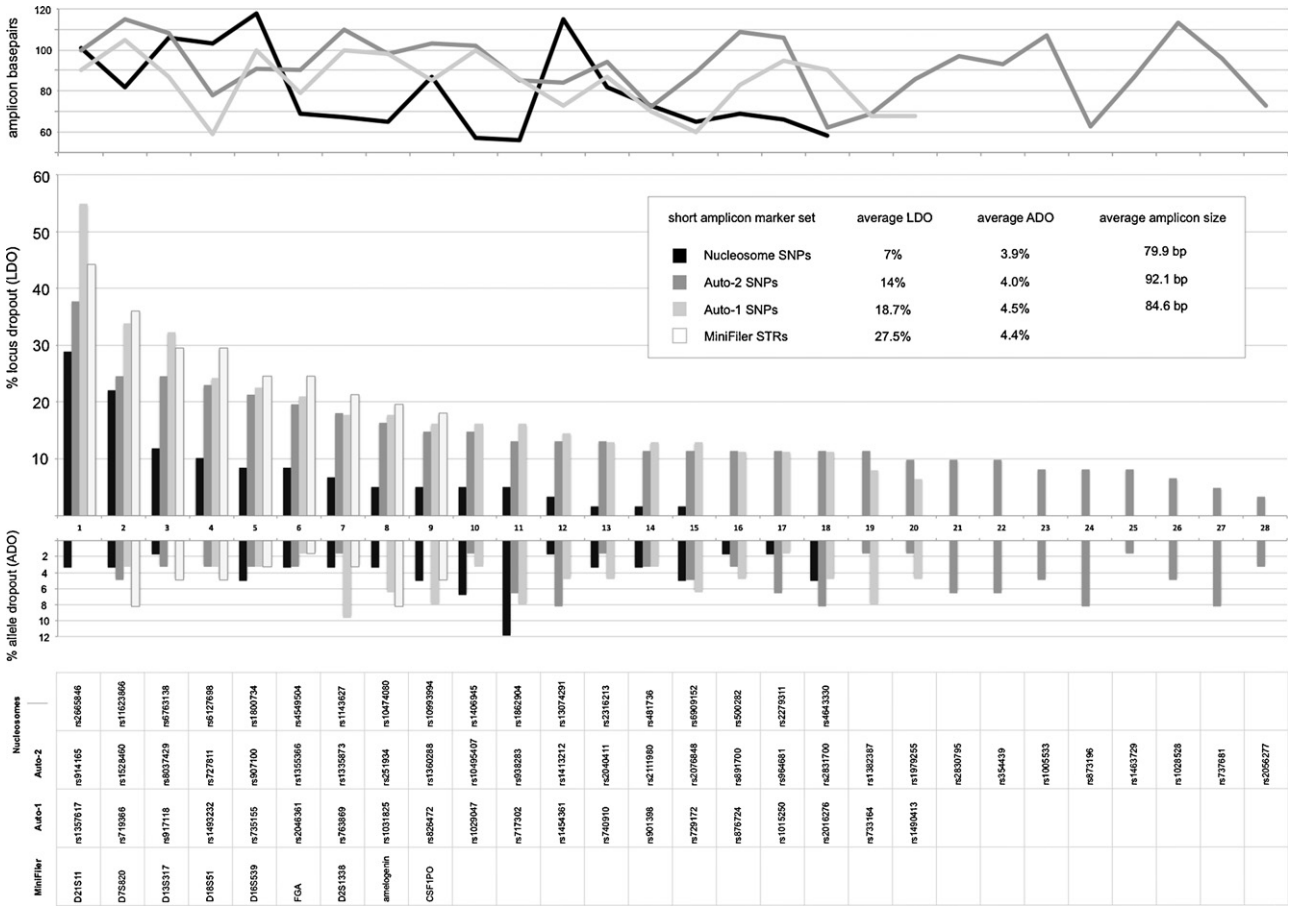

Fig. 4. Locus and allele dropout rates from 24 casework analyses showing individual performance of component markers of the four short amplicon sets. Markers are listed below both charts ordered by locus dropout. For reference the amplified fragment sizes of the SNP components are charted above the dropout plots indicating there is no correlation between amplicon size and dropout rates in the three SNP multiplexes.

Table 2
Summary forensic informativeness parameters for three SNP sets and two multiplex combinations (plus Identifiler STRs) in ascending order of total marker numbers and discrimination power using allele frequency data from three 1000 Genomes population groups. Dp, discrimination index; RMP, random match probability.

SNP No.	Multiplex	Dp	Dp expressed as '1 in value'	RMP
AFRICAN allele frequencies				
18	Nucleosome SNPs	5.E+05	490,983	2.0367E−06
23	Auto-1	1.E+07	14,180,796	7.0517E−08
29	Auto-2	4.E+08	400,412,938	2.4974E−09
40	Auto-2 + NUC	3.E+12	2,674,293,233,100	3.7393E−13
52	52plex	6.E+15	5,678,174,350,263,740	1.7611E−16
69	All SNPs	1.E+21	1,070,821,611,188,000,000,000	9.3386E−22
	Identifiler	2.E+17	216,991,347,078,572,000	4.6084E−18
EUROPEAN allele frequencies				
18	Nucleosome SNPs	3.E+06	2,552,449	3.9178E−07
23	Auto-1	7.E+08	666,346,741	1.5007E−09
29	Auto-2	3.E+11	264,735,495,225	3.7773E−12
40	Auto-2 + NUC	7.E+14	684,558,302,797,446	1.4608E−15
52	52plex	2.E+20	176,405,634,544,315,000,000	5.6687E−21
69	All SNPs	2.E+26	181,226,881,301,332,000,000,000,000	5.5179E−27
	Identifiler	9.E+16	94,564,025,650,079,600	1.0574E−17
EAST ASIAN allele frequencies				
18	Nucleosome SNPs	2.E+06	1,836,771	5.4443E−07
23	Auto-1	4.E+07	37,173,219	2.6901E−08
29	Auto-2	3.E+10	34,099,499,653	2.9325E−11
40	Auto-2 + NUC	3.E+13	25,643,968,133,970	1.4645E−14
52	52plex	1.E+18	1,267,588,185,243,330,000	7.889E−19
69	All SNPs	9.E+23	874,446,482,480,001,000,000,000	1.1435E−24
	Identifiler	6.E+16	61,663,500,000,000,000	1.6217E−17

with the smallest SNP multiplex of nucleosomes matching the information value of Auto-1 with an average six STR equivalents, both one and a half times as informative as MiniFiler™.

The above results obtained from the range of challenging DNA typed to assess nucleosome SNPs and alternatives, show that in cases involving degraded DNA miniaturized amplicon approaches will be between two to four times more informative than using conventional STRs. Our suggested approach for an optimum balance between expected genotyping success and information content from use of SNP multiplexes, either alongside MiniFiler™ or as a first strike strategy, would be to combine Auto-2 and nucleosome SNP multiplexes as this can be expected to provide ~80% genotyping success and approaches the same discrimination power of full Identifiler™ profiles: an average 37 SNPs typed per case (or 14.7 'STR equivalents').

3.5. Forensic statistics

Table 2 lists the discrimination index and random match probability estimates using the data of Fig. 2 for expanding sets of SNPs going from just using Auto-1 to the 52-plex plus nucleosome SNPs. Equivalent Identifiler™ values are also listed as a reference point. Combining nucleosomic SNPs with Auto-2, our suggested optimum approach for extremely degraded material, gives random match probabilities of 1E−13 in Africans, 6E−16 in Europeans and 1E−14 in East Asians, comparable values between each group that exceed any single SNP multiplex. Values for the other SNP combinations in Table 2 show that choosing the Auto-2/nucleosome multiplex combination provides levels of discriminatory power falling between Auto-1 alone and the 52-plex (Auto-1 and 2 together). Therefore, although combining nucleosome SNPs with Auto-2, rather than Auto-1, is potentially less informative, the expected increase in genotyping success would compensate for the reduction in power. Furthermore, although use of these nucleosome and Auto-2 SNPs corresponds to a reduction in discrimination power of between five (African) and one order of magnitude compared to a full Identifiler™ profile, all SNP multiplex combinations exceed the minimum value required to describe the profile obtained as globally unique (greater than a random match probability of 1E−10, discrimination index of 7E+9). Lastly use of all three SNP sets will always exceed the power from a single STR multiplex and this represents a realistic strategy for extremely degraded DNA where quantities are not limited, even when a proportion of SNPs may fail.

4. Discussion

In the study reported here we aimed to test the theory of nucleosome protection by selecting SNPs with a high probability to be within nucleosome forming regions and a potential benefit of resistance to several common degradation processes provided by the persistence of histone–DNA complexes. The initial idea was based on known properties of the apoptotic process where the specific DNA degradation pathway maintains intact, uncleaved nucleosomic regions with an inferred protective effect from histone binding within the nucleosome structure as previously suggested by Foran [44] and more recently by Thanakiatkrai et al. in a study specifically focused on potential protection of forensic STRs [45].

Location of nucleosomes to build a candidate SNP list was based on a double hit approach we developed to exploit growing knowledge of the position of nucleosome sites combined with the location of diagnostic sequence motifs recognized in the histone binding nucleotide segments found in nearly all human nucleosomes to date. Eukaryotic gene promoter chromatin generally presents a recognizable architecture characterized by a nucleosome-free region (NFR) flanked by at least one H2A.Z variant nucleosome [46]. In humans there appear to be multiple H2A.Z nucleosomes found both upstream and downstream of NFRs. NFR-adjacent nucleosomes are the most precisely positioned in the genome, with neighbouring nucleosomes becoming less precise in their locations as distance from NFRs increases. By acting as anchor points, the tight positioning of NFR-flanking nucleosomes may be the dominant pattern of nucleosome positioning genome-wide [46] and gave us the highest certainty of position. We therefore targeted these particular promoter sequences to analyze segments showing characteristic dinucleotide periodicity patterns and/or other sequence patterns using the RECON program to recognize these diagnostic motifs. This helped to ensure identified sites and their SNPs were highly likely to be positioned within the nucleosome structure but ultimately restricted the total number of SNPs available to select loci that satisfied the strict criteria required for forensic use.

Once we had collected sufficient numbers of candidates with high probability to be sited in the middle of nucleosomes we went on to audit the SNP*for*ID 52-plex SNPs to analyze their positions relative to likely nucleosome regions. In fact this audit confirmed all 52 SNPs chosen for identification applications are located randomly in non-coding positions across the human genome with no indications of nucleosome region characteristics. This is not unexpected, and follows from the original selection criteria of a minimum 100 kilobase distance from genes, considered sufficient separation to ensure all component SNPs were neutral and more likely to be in Hardy–Weinberg equilibrium compared to candidate pools that did not exclude coding region SNPs. The opposite applies to the nucleosome region SNPs that were collected in this study specifically focusing on loci within, or close to, promoters. Therefore the possibility of association between nucleosome component markers and the coding regions they are close to cannot be completely discounted. However we agree with the assessment of the ability to infer gene variation from closely sited single SNPs made by Budowle and van Daal [47], i.e. high heterozygosity SNPs by themselves effectively have zero predictive value for gene variation in close linkage.

When the nucleosome SNP multiplex was developed and tested against the performance of established forensic multiplexes the possible effect of nucleosome protection could be properly assessed. Our results show this potential protective effect is evident in the success rates observed, but an improvement in success from 81.4% to 87% is relatively small in scale. This suggests SNPs already enjoy a greater benefit from very short amplicon sizes and well-optimized PCR multiplexes so any additional protective effect from the nucleosome structures is marginal. In fact it can be argued that Auto-2 is only slightly less successful at typing highly degraded DNA than nucleosome SNPs and it remains the best performing multiplex in terms of total informative genotypes delivered amplifying challenging DNA. Therefore we believe the addition of nucleosome SNPs creates an improved system for analyzing challenging DNA compared to use of Auto-1 and Auto-2 alone. In comparison it is noteworthy that STR performance has been enhanced in the last two years by a concerted effort by the manufacturers to develop miniaturized primer sets enabling all STRs except SE33 and the longer FGA alleles to achieve amplicon sizes below 200 bp: significantly lower in most cases to the equivalent sizes of previous SGM™, Profiler Plus™ and Powerplex™ assays. Furthermore the reported performance of SE33 and FGA compared to shorter component STRs analyzing challenging DNA suggests new buffer formulations, increasingly used in revamped forensic kits, also contribute to improvements in performance. However it is also the case that MiniFiler™ was the first kit to incorporate these new buffer formulations [11]

and this STR multiplex is still noticeably less successful than SNP analysis in our casework analyses.

In many cases very degraded DNA sources are not particularly limited in the quantity of material available for amplification. Therefore based on our observations of consistently robust SNP performance across a range of real casework samples it is appropriate to recommend the use of the nucleosome multiplex we have developed alongside Auto-1 and Auto-2. Short amplicon SNP multiplexes with or without protection from DNA–histone complexes continue to provide informative and reliable supplementary genotype data when a strategy of STR typing alone may fail to give sufficiently useful profiles for the investigation.

Acknowledgments

The use of heatmaps to assess the likely state of DNA degradation in casework samples comparing different marker sets was developed by Walther Parson, Institute of Legal Medicine, Innsbruck Medical University, and we are grateful for the opportunity he provided for us to adapt these for comparisons of SNP and STR multiplex performance. MVL was supported by funding from Xunta de Galicia INCITE 09 208163PR and Ministerio de Educación y Ciencia BIO2006-06178. AFA was supported by a María Barbeito grant from Xunta de Galicia, MF by a fellowship from Fundación Pedro Barrié de la Maza.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.fsigen.2011.07.010.

References

[1] J. Burger, S. Hummel, B. Hermann, W. Henke, DNA preservation: a microsatellite-DNA study on ancient skeletal remains, Electrophoresis 8 (1999) 1722–1728.
[2] T. Schultes, S. Hummel, B. Herrmann, Amplification of Y-chromosomal STRs from ancient skeletal material, Hum. Genet. 104 (1999) 164–166.
[3] C. Kaiser, B. Bachmeier, C. Conrad, A. Nerlich, H. Bratzke, W. Eisenmenger, et al., Molecular study of time dependent changes in DNA stability in soil buried skeletal residues, Forensic Sci. Int. 177 (2008) 32–36.
[4] T. Lindahl, Instability and decay of the primary structure of DNA, Nature 362 (1993) 709–715.
[5] V.I. Bruskov, L.V. Malakhov, Z.K. Masalimov, A.V. Chernikov, Heat-induced formation of reactive oxygen species and 8-oxoguanine, a biomarker of damage to DNA, Nucleic Acids Res. 30 (2002) 1354–1363.
[6] J.H. Sadowski, Enzymatic breakage of deoxyribonucleic acid. I. Purification and properties of endonuclease II from T4 phage-infected *Escherichia coli*, J. Biol. Chem. 244 (1969) 6182–6191.
[7] J.M. Butler, Forensic DNA Typing: Biology and Technology behind STR markers, Academic Press, 2001.
[8] 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.
[9] S.R. Hughes-Stamm, K.J.K. Ashton, A. van Daal, Assessment of DNA degradation and the genotyping success of highly degraded samples, Int. J. Leg. Med. 25 (2010) 341–348.
[10] T. Senge, B. Madea, A. Junge, M. Rothschild, P.M. Schneider, STRs, mini STRs and SNPs - a comparative study for typing degraded DNA, Leg. Med. 13 (2011) 68–74.
[11] J.J. Mulero, C.W. Chang, R.E. Lagace, D.Y. Wang, J.L. Bas, T.P. McMahon, L.K. Hennessy, Development and validation of the AmpFlSTR MiniFiler PCR Amplification Kit: a MiniSTR multiplex for the analysis of degraded and/or PCR inhibited DNA, J. Forensic Sci. 53 (2008) 838–852.
[12] C. Luce, S. Montpetit, D. Gangitano, P.O. Donnell, Validation of the AmpFlSTR MiniFiler PCR Amplification Kit for use in forensic casework, J. Forensic Sci. 54 (2009) 1046–1054.
[13] R. Pereira, C. Phillips, C. Alves, A. Amorim, Á. Carracedo, L. Gusmão, A new multiplex for human identification using insertion/deletion polymorphisms, Electrophoresis 30 (2009) 3682–3690.
[14] J.J. Sanchez, C. Phillips, C. Børsting, K. Balogh, M. Bogus, M. Fondevila, et al., A multiplex assay with 52 single nucleotide polymorphisms for human identification, Electrophoresis 27 (2006) 1713–1724.
[15] E. Musgrave-Brown, D. Ballard, K. Balogh, K. Bender, B. Berger, M. Bogus, et al., Forensic validation of the SNPforID 52-plex assay, Forensic Sci. Int. Genet. 1 (2007) 186–190.
[16] C. Lou, B. Cong, S. Li, L. Fu, X. Zhang, T. Feng, S. Su, C. Ma, F. Yu, J. Ye, L. Pei, A SNaPshot assay for genotyping 44 individual identification single nucleotide polymorphisms, Electrophoresis 32 (2011) 368–378.
[17] D.G. Wang, et al., Large-scale identification, mapping and genotyping of single-nucleotide polymorphisms in the human genome, Science 280 (1998) 1077–1082.
[18] R. Sachidanandam, et al., A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms, Nature 409 (2001) 928–933.
[19] C. Phillips, M. Fondevila, M. Garcia-Magarinños, A. Rodriguez, A. Salas, Á. 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.
[20] A.M. Dearlove, High throughput genotyping technologies, Brief. Funct. Genomics Proteomics 1 (2002) 139–150.
[21] R.A. Lockshin, Z. Zakeri, Programmed cell death and apoptosis: origins of the theory, Nat. Rev. Mol. Cell Biol. 2 (2001) 545–550.
[22] R.C. Taylor, S.P. Cullen, S.J. Martin, Apoptosis: controlled demolition at the cellular level, Nat. Rev. Mol. Cell Biol. 9 (2008) 231–241.
[23] D.R. Green, Apoptotic pathways: ten minutes to dead, Cell 121 (2005) 671–674 (Review).
[24] N. Festjens, T. Vanden Berghe, P. Vandenabeele, Necrosis, a well-orchestrated form of cell demise: signalling cascades, important mediators and concomitant immune response, Biochim. Biophys. Acta 1757 (2006) 1371–1387.
[25] N.N. Danial, S.J. Korsmeyer, Cell death: critical control points, Cell 116 (2004) 205–219.
[26] A.H. Wyllie, Glucocorticoid-induced thymocyte apoptosis is associated with endogenous endonuclease activation, Nature 284 (1980) 555–556.
[27] J.T. Finch, L.C. Lutter, D. Rhodes, R.S. Brown, B. Rushton, M. Levitt, A. Klug, Structure of nucleosome core particles of chromatin, Nature 269 (1977) 29–36.
[28] R.D. Kornberg, Chromatin structure: a repeating unit of histones and DNA, Science 184 (1974) 868–871.
[29] R.K. Richmond, D.F. Sargent, T.J. Richmond, K. Luger, A.W. Ma, Crystal structure of the nucleosome resolution core particle at 2.8 A, Nature 389 (1997) 251–260.
[30] F. Ozsolak, J.S. Song, X.S. Liu, D.E. Fisher, High-throughput mapping of the chromatin structure of human promoters, Nat. Biotechnol. 25 (2007) 244–248.
[31] http://rulai.cshl.edu/cgi-bin/CSHLmpd2/promExtract.pl?species=Human.
[32] V.G. Levitsky, O.A. Podkolodnaya, N.A. Kolchanov, N.L. Podkolodny, Nucleosome formation potential of eukaryotic DNA: calculation and promoters analysis, Bioinformatics 17 (2001) 998–1010.
[33] V.G. Levitsky, RECON: a program for prediction of nucleosome formation potential, Nucleic Acids Res. 32 (2004) W346–W349.
[34] http://wwwmgs.bionet.nsc.ru/mgs/programs/recon/.
[35] http://www.ncbi.nlm.nih.gov/snp/.
[36] http://frodo.wi.mit.edu/primer3/.
[37] ttp://www.premierbiosoft.com/jsp/marketing/FreeToolLogin.jsp?PID=3.
[38] J.J. Sanchez, P. Endicott, Developing multiplexed SNP assays with special reference to degraded DNA templates, Nat. Protoc. 1 (2006) 1370–1378.
[39] P.M. Vallone, J.M. Butler, AutoDimer: a screening tool for primer–dimer and hairpin structures, Biotechniques 37 (2004) 226–231.
[40] O. Henegariu, N.A. Heerema, S.R. Dlouhy, G.H. Vance, P.H. Vogt, Multiplex PCR: critical parameters and step-by-step protocol, Biotechniques 23 (1997) 504–511.
[41] M.D. Timken, K.L. Swango, C. Orrego, M.R. Buoncristiani, A duplex real-time qPCR assay for the quantification of human nuclear and mitochondrial DNA in forensic samples: implications for quantifying DNA in degraded samples, J. Forensic Sci. 50 (2005) 1044–1060.
[42] M. Fondevila, C. Phillips, N. Naveran, L. Fernandez, M. Cerezo, A. Salas, A. Carracedo, M.V. Lareu, Case report: identification of skeletal remains using short-amplicon marker analysis of severely degraded DNA extracted from a decomposed and charred femur, Forensic Sci. Int. Genet. 2 (2008) 212–218.
[43] Charles Brenner (http://dna-view.com/SNPpost.htm).
[44] D.R. Foran, Relative degradation of nuclear and mitochondrial DNA: an experimental approach, J. Forensic Sci. 51 (2006) 766–770.
[45] P. Thanakiatkrai, L. Welch, Evaluation of nucleosome forming potentials (NFPs) of forensically important STRs, Forensic Sci. Int. Genet. (2010), doi:10.1016/j.fsigen.2010.05.002.
[46] P.D. Hartley, H.D. Madhani, Mechanisms that specify promoter nucleosome location and identity, Cell 137 (2009) 445–458.
[47] B. Budowle, A. van Daal, Forensically relevant SNP classes, Biotechniques 44 (2008) 603–610.