Exploring the usefulness of microhaplotypes in forensic identification using massive parallel sequencing technology

Chiara TURCHI [1] *, Filomena MELCHIONDA [1], Mauro PESARESI [1], Eleonora CIARIMBOLI [1], Carla BINI [2], Paolo FATTORINI [3], Adriano TAGLIABRACCI [1]

[1]Section of Legal Medicine, Department of Excellence of Biomedical Sciences and Public Health, Polytechnic University of Marche, Ancona, Italy; [2]Section of Legal Medicine, Department of Medical and Surgical Sciences, University of Bologna, Bologna, Italy; [3]Section of Legal Medicine, Department of Medicine, Surgery and Health, University of Trieste, Trieste, Italy

*Corresponding author: Chiara Turchi, Section of Legal Medicine, Department of Excellence of Biomedical Sciences and Public Health, Polytechnic University of Marche, Via Tronto, 60126 Torrette (AN), Italy. E-mail: c.turchi@univpm.it

ABSTRACT

BACKGROUND: Microhaplotypes or microhaps (MH) were recently introduced in the landscape of forensic genetic and appear to be useful for identification purposes, genotyping of degraded DNA, reconstruction of family relationships, ancestry prediction and DNA mixtures deconvolution. In order to make inference about a set of microhaps useful in forensic casework with low amount of degraded DNA and useful in kinship analysis, several microhaps were tested by massive parallel sequencing (MPS) assay.

METHODS: We have investigated the effectiveness of 29 microhaps in a set of real forensic samples together with artificially degraded DNAs. Moreover, we explore the informativeness of 87 microhaplotypes in relationship analysis through a simulation of different kinship testing scenarios typically encountered in forensic identification.

RESULTS: The MPS coverage analysis showed a good performance of the designed panel. Full profiles could be obtained with 0.1 ng of input DNA even with highly degraded samples. The increment of the number of PCR cycles does not result in an improvement in genotyping results in samples with low amounts of DNA, as the increase of drop-in and drop-out events were observed at 25 number of PCR cycles. No correlation between amplicons size and occurrence of drop-outs and drop-ins was observed. Kinship simulations showed that full siblings and half siblings relationships would be readily distinguished respect unrelated condition using the 87 microhaps panel.

CONCLUSIONS: Results shown that microhaps could be a powerful tool for individual identification, relationship resolution and that they are sensitive and reliable in degraded DNA typing.

(*Cite this article as:* Turchi C, Melchionda F, Pesaresi M, Ciarimboli E, Bini C, Fattorini P, *et al*. Exploring the usefulness of microhaplotypes in forensic identification using massive parallel sequencing technology. Minerva Medicoleg 2020;140:26-36. DOI: 10.23736/S0026-4849.20.01790-3)

KEY WORDS: Haplotypes; Forensic genetics; Family relations; High-throughput nucleotide sequencing.

Microhaplotypes loci (microhaps or MHs) are a new type of genetic marker defined by at least two single nucleotide polymorphisms (SNPs) within a short distance from each other (<300 nucleotides *i.e.* 'micro').[1] The alleles at a microhaps locus, referred to as haplotypes, are defined by the allelic combinations of each individual SNPs. As results, microhaplotype locus can be considered as a single multiallelic marker: virtually there are 8 possible different haplotypes that can exist for a 3-SNPs site, 16 with a 4-SNPs site, 32 with 5-SNPs site, etc. Moreover, a total of 36 different genotypes can exist for a 3-SNPs site, 136 with a 4-SNPs site, 258 with 5-SNPs

site, etc. Therefore, the microhaps have higher levels of polymorphism than any of the individual SNPs that form it.

The short distance between the SNPs composing the microhap locus allows that a single small DNA sequence to cover the entire distance resolving the phase, *i.e.* the cis/trans relationship between the alleles at the individual SNPs.[1]

For this reason, the progress of massive parallel sequencing (MPS) turned out to be highly informative in microhaplotypes detection as a single sequence read can cover the expanse of the microhaplotype and these loci become phase-known (*i.e.* the allelic combination of multiple SNPs on each chromosome of an individual can be determined).[1] Moreover, MPS can analyze a large number and several types of markers in a single sequencing run, proving to be useful for addressing relevant forensic issues in a single assay.[2]

As well as for SNPs typed as individual loci, microhaplotypes present a number of advantages for forensic identification also compared to short tandem repeats (STRs) commonly used in forensics.[3]

One of the major issues regards the production of stutter that occurs when STR alleles are PCR amplified because of strand slippage. Stutter products were detected in both capillary electrophoresis both in MPS and their presence represent a big hindrance in mixture interpretation as they complicate the accurate identification of minor contributor alleles. Conversely, microhaps do not generate stutter peaks, and this characteristic makes them particularly attractive for mixtures interpretation. Another disadvantage of STRs over microhaps is that certain STR alleles, within a locus, do not have the same extent, but they may vary in length up to 100 nucleotides. This feature could cause a preferential amplification of the shorter alleles and an overrepresentation of them compared to the longer alleles in cases of degraded DNA samples, complicating both allele genotyping in single source samples and allele source attribution in mixtures. In addition, several STRs require long amplicons than microhaps to ensure the whole repeat region is amplified, with the result that many allelic or locus drop-out could be observed when degraded DNA is genotyped with STRs. Moreover, SNPs composing microhaps loci have lower mutation rates than STRs, approximately five to six orders of magnitude less, and therefore they are ideal loci for relationship testing.

Another important feature of microhaps is that they are potentially suitable for the analysis of highly degraded DNA samples as, due to the short distance between the SNPs within a microhap locus, the PCR amplicons size could be very small.

In summary, microhaplotypes loci are characterized by higher levels of polymorphism, absence of stutter production, low mutation rate and short amplicons. All these features make microhaps exploitable in forensic genetics for identification purposes, reconstruction of family relationships,[4] biogeographic ancestry prediction and can be useful for both detecting and deconvoluting DNA mixtures.[5, 6] Microhaps loci useful for forensic identification should be highly heterozygous with multiple alleles and the main relevant metric to be considered is the effective number of alleles (A_e). A_e represents the number of equally frequent alleles that would generate the same heterozygosity as the locus with multiple alleles at very different frequencies.[7, 8] This parameter relates to the level of intra-population variation and is associated with the random match probability (RMP), the marker's ability to distinguish individuals. The A_e statistic relates also to the usefulness of the locus in resolution of relationships and resolving DNA mixture as the higher the effective number of alleles, the more probable a mixture could be detected. For biogeographic ancestry inference, microhaps must show different allele frequencies among various populations. For this purpose, the Informativeness (I_n) statistic should be considered and microhaps with high I_n values show more differentiation between different population and should be candidates for ancestry inference.[8]

In the past years, different studies have explored the potential of microhaplotypes in forensic genetics, by using different SNP typing methods. One of the first studies on microhaps included a panel of 31 loci analyzed by using TaqMan assay.[5] The study that has been subsequently extended to 130 microhaps,[1] with

haplotype frequencies evaluated in 83 different population samples and then by testing 65 of the microhaps in 13 additional populations.[9] Overall, the results showed that these loci provide some useful information for distinguishing up to 10 clusters of populations.

Anyway, the majority of the studies have been performed by MPS technology on a large number of different loci and have been addressed to different applications. The potential of microhaps for individual identification and kinship analysis was assessed in several papers,[10-14] as well as their usefulness in ancestry inference.[15-19] Finally, the potential utility of microhaps in mixture resolution was also recently evaluated.[10, 14, 20-24]

In previous studies, we selected 87 MH loci annotated in the ALlele FREquency Database (https://alfred.med.yale.edu) and their genetic variation was evaluated in 100 Italian individuals using MPS, in order to make inference about the usefulness if the microhaps panel in forensic genetic.[25, 26] Overall, the A_e values for the 87 microhaps range from 1.010 to 8.344, with about 80% showing values greater than 2.0. Noteworthy, 32 microhaps display A_e values greater than 3.0 and 18 loci A_e above 4.0. The individual matching probabilities (PI) of the 87 microhaps ranged from 0.032 to 0.9802. Considering the 32 microhap loci with Ae greater than 3.0, the cumulative PI value was 1.6×10^{-33}. To explore the suitability of this MH panel in mixture deconvolution, the probability of detecting a mixture, as a function of A_e, was inferred and, combining only the 32 MH loci with A_e above three, the theoretical probability of detecting a mixture was 0.999999999999973. These results make the subset of 32 loci informative for mixture resolution. In order to better understanding the performance of microhaplotypes to challenging samples as low amounts of degraded samples, we selected a subset of 29 MH among the 87 MH previously explored. MPS panel were designed keeping the amplicons size below 180 bp, to investigate their usefulness with degraded DNA samples and preliminary results of this assay was reported in Turchi *et al.*[27]

In the present study, we have investigated more thoroughly the effectiveness of the 29 microhaps with low amounts of degraded samples, in terms of sequence coverage and genotyping results in a set of real forensic samples together with artificially degraded DNAs. Moreover, we explore the informativeness of the 87 and 29 microhaplotypes in relationship analysis through a simulation of different kinship testing scenarios typically encountered in forensic identification.

Materials and methods

Two microhaplotypes panel were evaluated in this study composed of 87 and 29 microhaps respectively, and they will be referred to as 87 MH panel and 29MH panel.

87 MH panel description and analysis

Sample preparation, selection of 87 microhaps loci, primer design for massive parallel sequencing, MPS libraries preparation and sequencing, data analysis and statistical analysis are reported and extensively explained in Turchi *et al.*[25]

29 MH panel description and analysis

Selection of 29 microhaps loci, primer design for massive parallel sequencing, MPS libraries preparation and sequencing, data analysis and statistical analysis are reported in Turchi *et al.*[27] Briefly, PCR primers for MPS library were designed keeping the amplicons size between 125 and 175 bp, making sure that the SNPs included in the MH locus were amplified in a single amplicon Different types of samples were analyzed: six blood samples, two bone remains, two FFPE tissues together with a set of artificially degraded DNAs. Sensitivity study was performed by using the 2800M DNA (Promega, WI, USA) as control, diluted to concentrations of 5 ng/μL, 1 ng/μL, 500 pg/μL, 100 pg/μL, 50 pg/μL and 25 pg/μL. The DNA quantity and DNA degradation index (DI) were assessed by Quantifiler™ Trio DNA Quantification Kit. Library amplifications were performed with DNA input ranging from 5 ng to 25 pg and testing different number of PCR cycles (21 and 25). Libraries for MPS analysis were prepared with Precision ID Library kit according to the user guide (MAN0015830) and sequenced on Ion PGM™ System. A total of 28 barcoded libraries were sequenced on one chip.

Raw data was processed by the Torrent Suite v. 5.0.4 and the reads aligned against human reference genome (GRCh37/hg19). A minimum coverage of 50× was used as thresholds for reliable genotyping.

The relative Depth of Coverage (rDoC) across all loci sequence was calculated as the ratio of Depth of Coverage (DoC) at single locus to total DoC of the sample. Locus strand balance was measured with the ratio of forward strand coverage to total coverage. For data elaboration, the analyzed samples were pooled in two groups that were named, according to the DNA degradation index (DI), as follows: non-degraded DNA samples: DI<5; degraded DNA samples: DI>5.

Kinship analysis simulations

In order to assess the statistical power of the two microhaplotypes panels for analyzing complex kinship testing scenarios, simulations of different pedigrees were performed in Familas v. 3.2.6.[28] Deficient pedigrees involving pairwise relationship and included full-sibling, half-siblings and first-cousins were tested, using the haplotype frequency estimates published in Turchi *et al.*[25] The same simulations were also performed for the STR markers included in the PowerPlex® Fusion 6C System (Promega), using the allele frequency estimates reported in Hill *et al.*[29] The number of simulations for each pedigree was set to one hundred thousand and the resulting distributions of likelihood ratios (LRs) for each kinship hypothesis *versus* the values for the unrelated hypothesis were plotting by using R v. 3.6.0.[30]

Results

The 29 MH panel designed for degraded DNA consists of 15 2-SNPs, 12 3-SNPs and 2 4-SNPs microhaps, comprised of 74 SNPs spread across 15 of 22 human autosomes. The effective number of alleles at each locus (A_e) ranged from 1.800 to 5.089 and 9 microhaps have values greater than 3.0. The description and the summary statistics of the 29 microhaplotypes are reported in Table I.

The loci ranged from 18 bp to 115 bp in length between the outermost SNPs and the sizes of the targeted PCR amplicons range from 126 to 174 bp. To keep the amplicons length below 180 bp, the extent of six microhaplotypes was reduced by excluding, respect the original configuration reported in ALFRED, the following SNPs: rs72623112 in mh02KK-134; rs3775867 and rs17088476 in mh04KK-013; rs74865590 in mh05KK-170; rs9536430 in mh13KK-218; rs2838081 in mh21KK-320 and rs6518223 in mh21KK-324.

The average locus strand balance across all 29 loci was estimated and results indicated that most of microhaps in the panel were balanced with average ratio of forward strand coverage to total coverage within 0.5±0.20. Overall, fluctuation of strand balance at each locus was observed, but values were within the range considered (Figure 1). Only the mh04kk-074 locus has been identified with imbalance, with most of the values above the upper threshold of 0.7.

The coverage analysis of MPS data displayed a good uniformity. The rDoC was assessed in degraded and non-degraded DNA samples, amplified by using above 1 ng of input DNA. A uniform rDoC distribution across all 29 loci in non-degraded samples, regardless of the amplicons' size, was found. The locus that showed the lowest rDoC value was mh09KK-034, with an amplicon size of 174 bp (Figure 2). In degraded samples we observed a slight decrease of the rDoC values at the high molecular weight loci, together with an increase of the rDoC values at the low molecular weight loci. Nevertheless, the trend of rDoC distribution in degraded samples remains quite homogeneous in all loci, except the microhap mh02KK-134, which displayed a remarkable decrease of rDoC value respect to non-degraded samples (Figure 2).

The rDoC was also assessed in samples of sensitivity study, using dilution of 2800M for DNA input. Figure 3 shows the distribution of rDoC per microhap obtained with 5 ng, 1 ng, 500 pg, 100 pg, 50 pg, and 25 pg of 2800M DNA as input. Overall, these distributions share a common pattern, despite different quantities of DNA amplified. Noteworthy, mh002KK-134 and mh-09KK-034 loci displayed low level of coverage respect to all other loci, regardless of input DNA quantities.

Genotyping results of the diluted control DNA samples showed that full genotypes concordant

TABLE I.—*Summary statistics of the 29 microhaplotypes analyzed in this study.*

#Chr	Microhap	#SNP	SNPs	Amplicon size	# genotype	#Haplotype	FMFH	A_e	Het	PI
2	mh02KK-073	2	rs1374748/rs7583554	174	6	3	0.4511	2.740	0.6351	0.220
	mh02KK-134*	3	rs12469721/rs3101043/rs3111398	170	26	8	0.2071	6.129	0.8369	0.054
	mh02KK-136	3	rs6714835/rs6756898/rs12617010	173	15	6	0.2680	4.538	0.7796	0.092
	mh02KK-213	3	rs7568519/rs7577785/rs1519654	162	6	3	0.5500	2.462	0.5938	0.230
3	mh03KK-007	2	rs4513489/rs6441961	174	6	3	0.4750	2.587	0.6135	0.217
	mh03KK-009	2	rs3732783/rs6280	164	5	3	0.6450	1.983	0.4957	0.347
4	mh04KK-010	2	rs3135123/rs495367	126	9	4	0.4250	3.117	0.6792	0.159
	mh04KK-011	2	rs6855439/rs6531591	171	6	3	0.5000	2.579	0.6122	0.227
	mh04KK-013*	3	rs13131164/rs3775866/rs11725922	173	10	4	0.4400	3.104	0.6778	0.168
	mh04KK-074	2	rs11932595/rs17085763	170	4	3	0.5663	2.049	0.5120	0.357
5	mh05KK-022	2	rs41461/rs41462	173	6	3	0.3550	2.988	0.6654	0.177
	mh05KK-062	2	rs870348/rs870347	174	5	3	0.4800	2.224	0.5504	0.317
	mh05KK-170*	3	rs438055/rs370672/rs6555108	162	29	8	0.3250	5.381	0.8142	0.066
6	mh06KK-026	3	rs4565296/rs4431439/rs179939	173	10	5	0.5100	2.596	0.6148	0.247
7	mh07KK-031	2	rs17168174/rs10246622	173	4	3	0.6717	1.800	0.4443	0.394
9	mh09KK-034	2	rs1408800/rs1408801	174	6	3	0.6100	2.219	0.5494	0.262
	mh09KK-153	3	rs10125791/rs2987741/rs7047561	163	10	6	0.5867	2.404	0.5840	0.236
11	mh11KK-039	2	rs2288159/rs10891537	160	6	3	0.5550	2.347	0.5740	0.256
12	mh12KK-043	3	rs11613749/rs11062734/rs17780102	174	9	4	0.6150	2.127	0.5299	0.302
	mh12KK-046	2	rs1503767/rs11068953	170	8	4	0.5150	2.666	0.6249	0.219
13	mh13KK-218*	3	rs1927847/rs9536429/rs7492234	157	23	8	0.3000	5.089	0.8035	0.075
	mh13KK-225	3	rs4884651/rs9529023/rs7329287	151	13	5	0.3700	3.774	0.7351	0.122
16	mh16KK-302	4	rs1395579/rs1395580/rs1395582/rs9939248	174	11	5	0.7050	1.921	0.4794	0.318
17	mh17KK-053	2	rs3760370/rs3760371	174	5	3	0.5350	2.253	0.5561	0.278
	mh17KK-054	2	rs2233362/rs634370	151	6	3	0.4800	2.551	0.6080	0.242
18	mh18KK-293	4	rs621320/rs621340/rs678179/rs621766	166	11	6	0.6450	2.051	0.5124	0.319
21	mh21KK-320*	3	rs2838082/rs78902658/rs2838083	173	14	5	0.3724	3.777	0.7353	0.124
	mh21KK-324*	3	rs2838868/rs7279250/rs8133697	174	17	7	0.3150	4.081	0.7550	0.111
22	mh22KK-060	2	rs4818/rs4680	152	5	3	0.4800	2.381	0.5800	0.256

#Chr: chromosome number; FMFH: frequency of most frequent haplotype; A_e: effective number of alleles; Het: heterozygosity; PI: matching probability.
*Microhaplotypes modified — compared to the original configuration reported in ALFRED — by excluding one or two outermost SNPs.

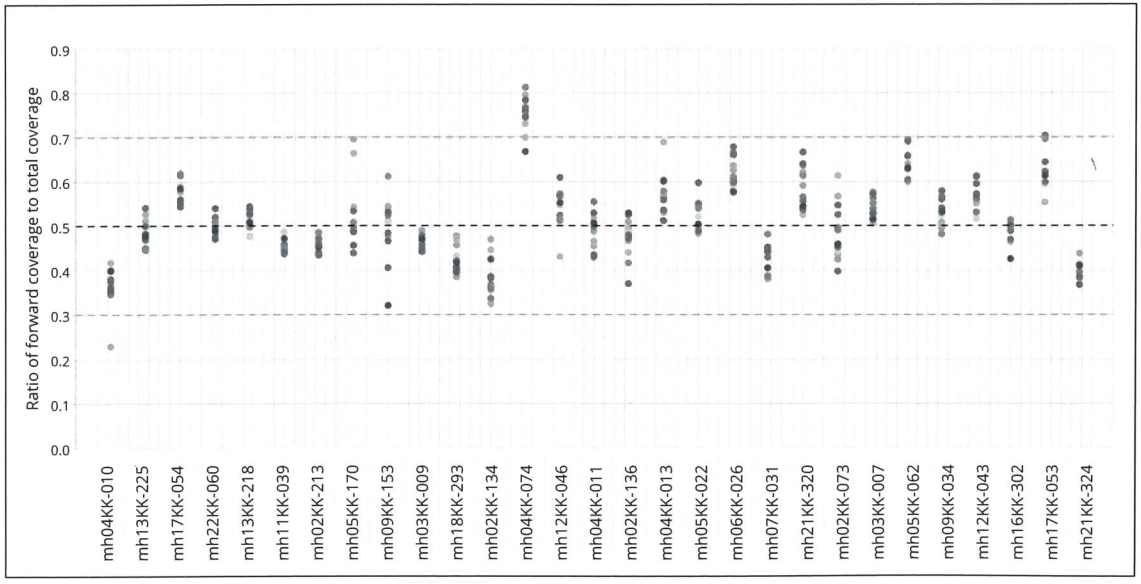

Figure 1.—Locus strand balance observed in 29 microhaplotypes. Ratio equal to 0.5 indicate that forward and revers strands have the same coverage. The lines at 0.3 and 0.7 are the lower and the upper thresholds defined in this study.

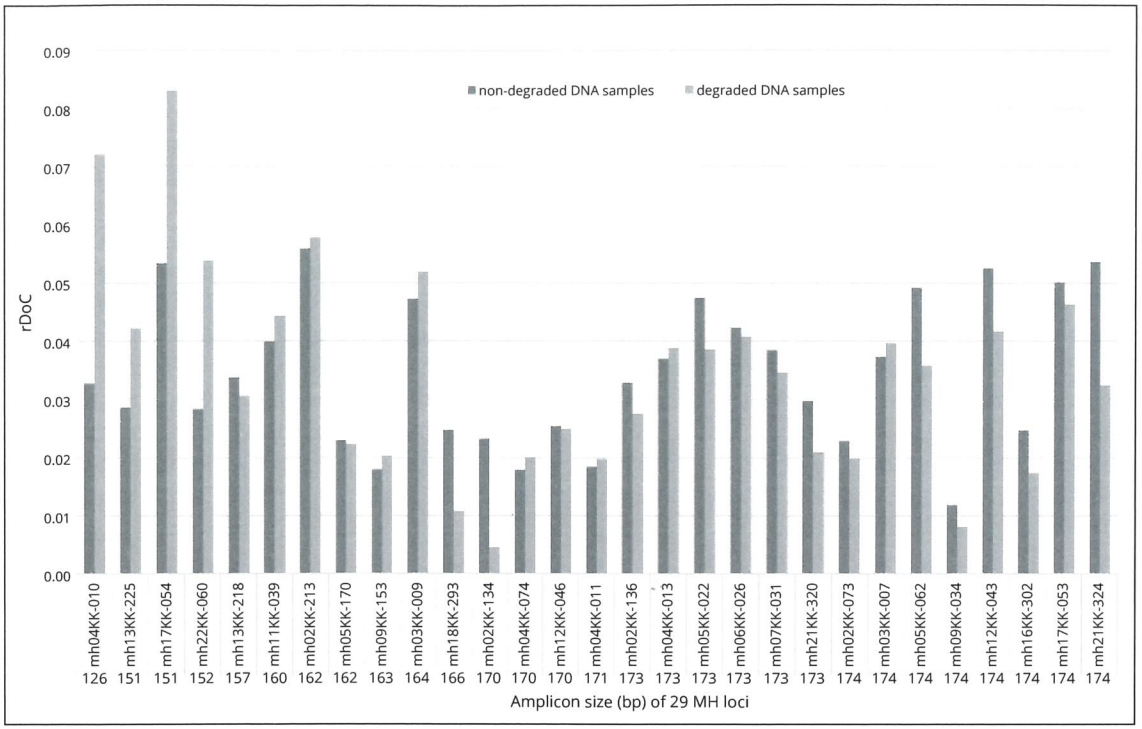

Figure 2.—Relative depth of coverage (rDoC) in non-degraded and degraded samples observed in the 29 microhaplotypes. The markers are sorted by their molecular weight (from left to right, 126 bp to 174 bp).

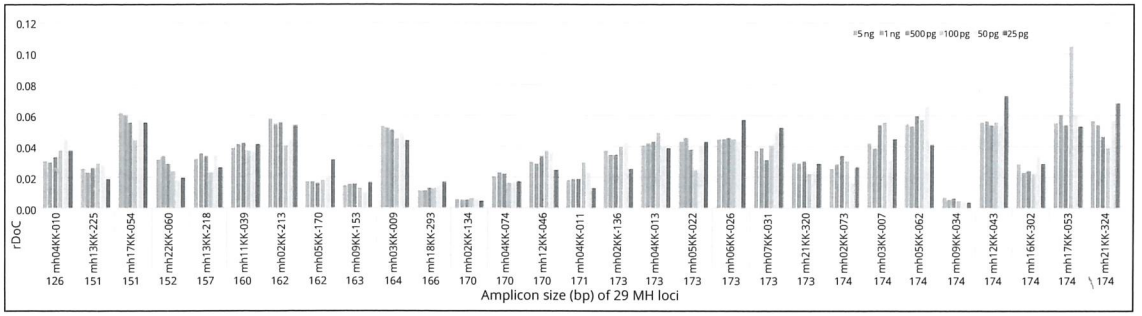

Figure 3.—Distribution of rDoC per microhap obtained with 5 ng, 1 ng, 500 pg, 100 pg, 50 pg and 25 pg of 2800M as input DNA for libraries amplification.

with the expected profiles were obtained whit input DNA above 0.1 ng. The percentage of correct genotypes at different PCR cycles were previously reported in Turchi *et al.*[27] and are displayed in Figure 4A. As expected, stochastic events such as allele drop-outs and drop-ins occur at a higher frequency at the lowest input DNA amounts. Drop-ins were mainly observed with 25 pg and were more frequent than allelic drop-outs (Figure 5). Locus drop-outs were observed only with 25 pg of DNA as input.

Genotyping results were then evaluated both in artificial degraded DNA samples both in real forensic samples. Two artificially degraded DNA (depurinated DNA) were analyzed: the first one (#1) showed a DNA concentration of 0.462 ng/μL and a degradation index (DI) not measurable (n.m.) due to the failure to amplify or to a very low amplification result (<0.001 ng/μL) of the large amplicon in the qPCR assay; the second one (#2) showed a very low DNA content (not quantifiable, DNA quantity below the limit

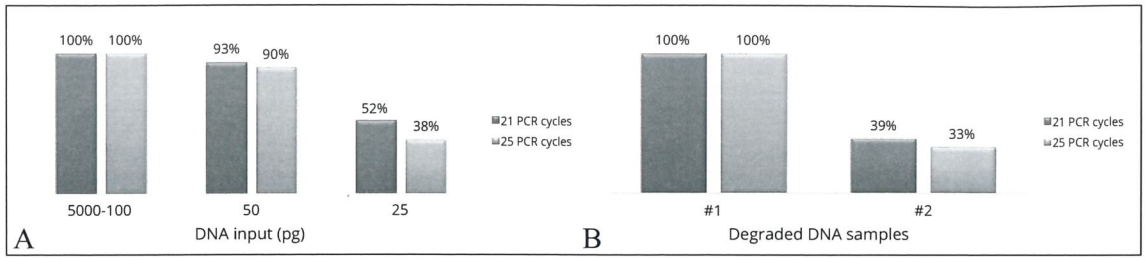

Figure 4.—Genotyping results of the diluted control DNA samples (A) and of the artificially degraded DNA (B). The plots indicated the percentage of correct genotypes obtained with different amounts of input DNA, at 21 and 25 PCR cycles.

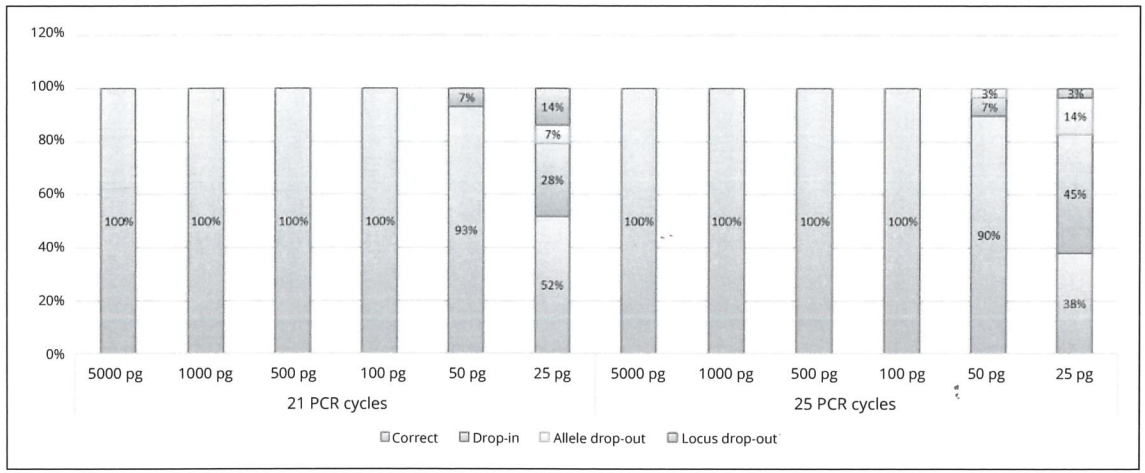

Figure 5.—Genotyping results of series dilutions in sensitivity study. Allele and locus drop-outs and drop-ins occur at a higher frequency at the lowest input DNA amounts (50 pg and 25 pg).

of quantification of the qPCR assay) and DI not measurable (n.m.). A reference sample (#3) for both the artificially degraded DNA samples was also analyzed.

The two artificially degraded DNA samples gave different genotyping results. The percentage of correct genotypes at different PCR cycles were previously reported in Turchi *et al.*[27] and were displayed here in Figure 4B. In sample #2 drop-ins events were observed more frequent than allelic drop-outs (Figure 6) and locus drop-outs occur at a higher frequency at 25 PCR cycles.

The forensic specimens tested consisted of two FFPE tissue samples and two bone remains. MPS of the two FFPE tissue, which displayed a good DNA concentration (8 ng/µL and 3.5 ng/µL), but high level of DNA degradation (DI=40 and DI=n.m.), showed high amplicons mean depth values (5530 and 3470) and full profiles with all microhaps loci typed. One bone samples,

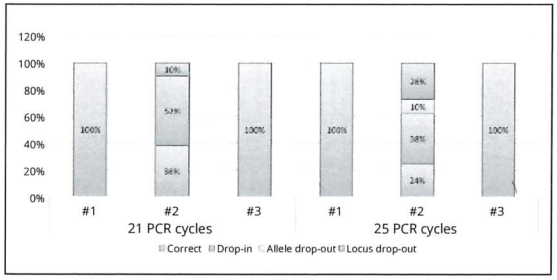

Figure 6.—Genotyping results of two artificially degraded DNA samples (#1 and #2) compared to reference sample (#3). Full genotypes concordant with the reference were obtained in sample #1 (1 ng of input DNA; DI=n.m.), regardless the number of PCR cycles. Sample #2 (input DNA not quantifiable; DI=n.m.) showed partial results, both using 21 and 25 PCR cycles.

which displayed a DNA concentration of 0.7 ng/µL and a DI=5, showed high amplicons mean depth values (4946) and full profiles. The other bone sample, which contained very low DNA amount (not quantifiable, DNA quantity below

the limit of quantification of the qPCR assay) and very degraded DNA (DI=n.m.), despite quite amplicons mean depth values (528.6) displayed a partial profile, with only 75.8% of microhaps loci typed.

Kinship simulations were performed on the 87 microhaps loci included in the 87 MH panel and on the 29 loci included in the 29 MH panel. Moreover, simulations were also carried out for the autosomal STRs included in the PowerPlex® Fusion 6C System (Promega), to compare the informativeness of microhaps with those of STR markers commonly used in relationship tests.

The overall results of kinship testing simulations are reported in Figure 7. Considering the 87 MH panel, the distribution of LR of full siblings and unrelated individuals do not overlap, with average $\log_{10}$ LRs for related individuals of 13.74 and for unrelated individuals of -12.18. Even for the half siblings scenario, the LR distribution clearly separate related and unrelated individuals, with average $\log_{10}$ LRs of 3.51 for related pairs and -3.339 for unrelated pairs. For first cousin simulation the LR distribution plot displayed some degree of overlap, with average $\log_{10}$ LRs of 0.9253 for related pairs and -0.8398 for unrelated pairs.

Similar results were observed when testing the same pedigrees with the 29 MH panel, even if a decrease in the LR values has been observed in all simulations, as expected. Only the full sibling test results in a clear separation between related and unrelated individual, with average $\log_{10}$ LRs of 4.332 for related pairs and -3.965 for unrelated pairs. For the half sibling and first cousin simulation, distributions overlapped to some extent.

The LR distributions from kinship test simulations of the STRs included in the PowerPlex® Fusion 6C System showed a separation between full siblings and unrelated individuals, with average log10 LRs for related individuals of 7.043 and for unrelated individuals of -5.5. For the half sibling simulation, LR distributions overlapped to some extent, with average log10 LRs for related individuals of 1.944 and for unrelated individuals of -1.731. The first cousin scenario was not evaluated for STR markers, due to the low informativeness obtained from half sibling test.

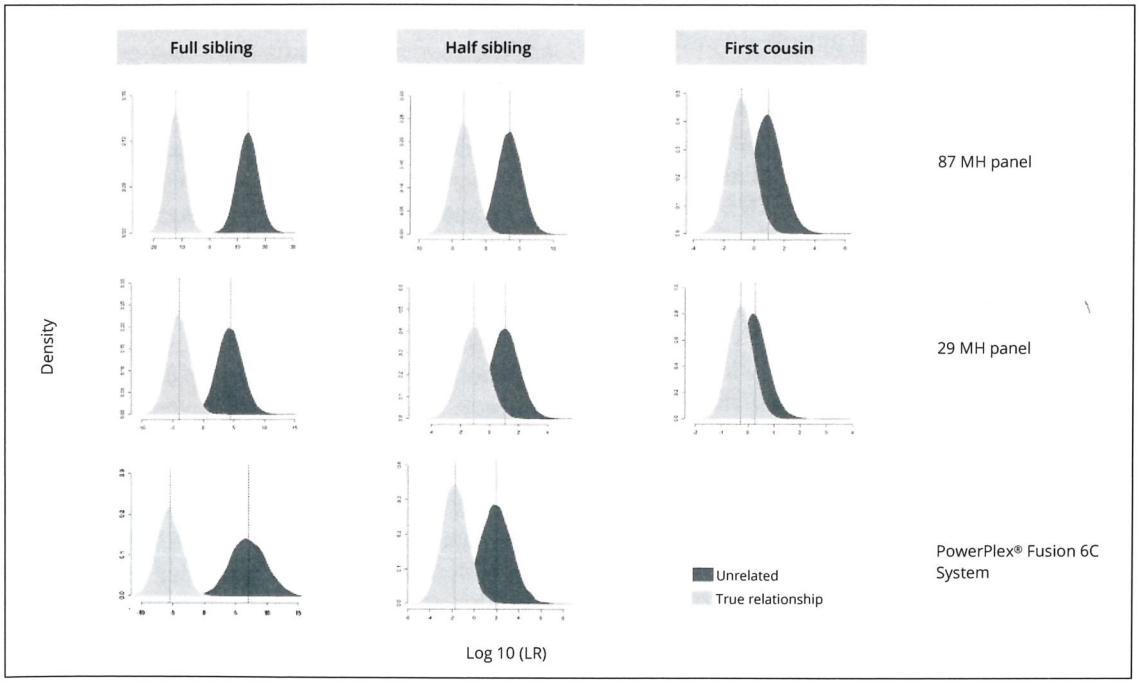

Figure 7.—Log_{10} LR distribution plots obtained from pedigree simulations tests for the supposed true relationships plotted against unrelated condition. Relationship tests comprised full siblings *versus* unrelated, half siblings *versus* unrelated and first cousins *versus* unrelated. Simulations were performed for the 87 MH panel, the 29 MH panel and the PowerPlex® Fusion 6C System (Promega).

Discussion

The introduction of MPS within the forensic community makes it possible to explore a new type of genetic marker, consisting of a combination of several SNPs, mapped in a short DNA sequence, known as microhaplotypes. These microhaps were proposed as alternative markers for STR typing of mixtures, degraded DNA, identification purposes and reconstruction of close biological relationships.[3]

In a previous study, 87 MH loci annotated in ALlele FREquency Database were selected and evaluated for their genetic variation in 100 Italian individuals using MPS, in order to make inference about their useful in forensic genetic.

More recently, to better understand the performance of microhaplotypes with challenging samples as low amounts of degraded samples, we selected a subset of 29 MH among the 87 MH previously explored. MPS panel were designed keeping the amplicons size below 180 bp, to investigate their usefulness in a set of real forensic samples together with artificially degraded DNAs. Indeed, starting from the 87 MH panel, we first selected 32 microhaps, with A_e values above 3, which are found to be highly efficient in detecting mixture and individual identification. Anyway, these 32 microhaps displayed too large extent between the outermost composite SNPs and results unusable to design small amplicons. As a result, only nine of the 32 MH with A_e values above 3 were included in the final 29 MH panel. The 29 MH panel consists of 15 2-SNPs, 12 3-SNPs and 2 4-SNPs microhaps. The 15 microhaps with two composite SNPs were, as expected, the least informative: only mh04KK-010 and mh12KK-046, displayed all the four possible alleles and mh04KK-010 displayed nine out of the 10 possible genotypes and the higher A_e (3.117) than all other 2-SNPs microhaps analyzed. Among the 12 microhaps with three composite SNPs, only mh13KK-218, mh05KK-170 and mh02KK-134 displayed all the eight possible alleles and the highest A_e values in the category of 3-SNPs microhaps. Noteworthy, the 4-SNPs microhaps mh16KK-302 and mh18KK-293 displayed only five and six of the 16 possible alleles, respectively, and A_e values around 2. The

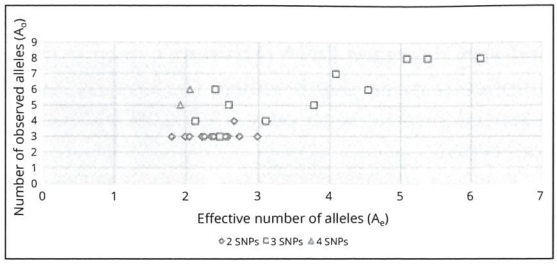

Figure 8.—Scatterplot of 29 microhaplotypes by their number of alleles observed (A_o) and effective number of alleles (A_e). In the figure appears that 4-SNPs loci tested in this study have an equal informativeness in terms of A_e than 2-SNPs loci.

calculated A_e values were compared with the number of alleles observed at each locus (A_o), related to the number of SNPs that comprise the microhap (Figure 8) and a correlation between A_o and A_e values was not always detected. Moreover, the 4-SNPs loci tested in this study have an equal informativeness in terms of A_e than 2-SNPs loci. Overall, these findings corroborate the importance of extensive genotyping study of microhaps in different population to evaluate haplotypes and genotypes frequency and to make inference about their informativeness in forensics.

The coverage analysis showed a good performance of the 29 MH panel, with all loci uniformly amplified without substantial differences between degraded and non-degraded DNAs, except for microhap mh02KK-134 which displayed a remarkable decrease of coverage in degraded samples. The choice to keep the PCR target sizes below the 180 bp was proved to be sufficient for allow an uniform rDoC at all loci, with no substantial differences between small and long amplicons. Even the quantity of DNA amplified did nott affect the depth of coverage and similar rDoC distribution were observed for different quantities of input DNA, even if mh002KK-134 and mh09KK-034 loci displayed low level of coverage respect to all other loci. Noteworthy, mh002KK-134 and mh09KK-034 loci, despite their low level of coverage, were correctly genotyped in all diluted samples, even with the lowest input DNA amounts.

Analysis of drop-ins and drop-outs events showed that accurate and reliable microhap profiles could be obtained with 0.1 ng of input

DNA, even with highly degraded samples, and that these events occur mainly at 25 number of PCR cycles. Moreover, no correlation between amplicons size and occurrence of drop-outs and drop-ins was observed. These results confirmed and consolidated the previous findings[27] which showed that the increment of the number of PCR cycles does not result in an improvement in genotyping results in samples with low amounts of DNA and that the critical parameter that affect the result, both in terms of coverage both in terms of correct genotyping, is the quantity of input DNA. Therefore, the 29 MH panel results suitable for degraded DNA typing, even if the performance decreases with low values of input DNA.

To test the forensic ability of the 87 MH panel and 29 MH panel to distinguish individuals, the matching probability (PI) for all 87 microhaplotype loci and for the subset of 29 loci was calculated. The combined matching probability values of are equal to 5.7×10^{-63} and 3.3×10^{-21} for 87 loci and 29 loci respectively. The matching probability calculated for the 29 microhaps was similar to those observed for other microhaplotype panels with comparable number of loci[5] and for STRs panels commonly used in forensics.[25] This feature, together with small amplicons size and absence of stutter artefacts, makes these loci a useful alternative method in cases of unbalanced mixed sample or degraded DNA.

Kinship test simulations were performed to gauge the informativeness of the microhap panels when applied to diffcult kinship testing scenarios. When 87 microhaps were considered in the simulations, full siblings and half siblings relationships would be readily distinguished respect unrelated condition using such a microhaps panel. For first cousin simulations distributions overlapped to some extent, indicating that on this scenario the 87 MH panel is unable to distinguish related and unrelated hypotheses Comparison between LR distributions of these microhaps and the 24 STRs included in the PowerPlex® Fusion 6C System (Promega) showed that the 87 MH panel resulted more informative than the STRs kit in resolving full and half sibling. The 29 MH panel resulted less informative in resolution of the same two kinship scenarios than STRs (Figure 7). As results, the 87 microhaps appear to be very usefulness both for individual identification, both for reconstruction of relationships, also in complex kinship testing scenarios.

Conclusions

The introduction of MPS technologies allows other types of loci to be included in the set of markers currently used in forensic genetics. Microhaplotypes have several features that make them exploitable in forensic genetics for identification purposes, genotyping of degraded DNA, reconstruction of family relationship, biogeographic ancestry prediction and can be useful for both detecting and deconvoluting DNA mixtures. Indeed, they are characterized by higher levels of polymorphism, absence of stutter production, low mutation rate and short amplicons. Several microhaplotypes loci to increase our knowledge about their usefulness in forensics was explored in this study. The combined matching probability and the kinship test simulations carried out shown that microhaps could be a powerful tool for individual identification, relationship resolution and that they are sensitive and reliable in degraded DNA typing.

Overall, the results confirm the utility of microhaps in forensics, even if before these markers could be used in forensic practice many technical and interpretation issues must be addressed and solved. First, the potential value of each microhaplotype locus for different forensic purposes is related to specific characteristics and it seems to be difficult to identify a set of microhaps suitable for all possible forensic applications. Careful loci selection and evaluation of relevant metrics must be performed before addressed the use of microhaps in forensic routine. A very important issue will be to improve MPS data analysis by validated bioinformatics pipeline to allow reliable haplotypes resolution. Finally, comprehensive microhaplotype frequency data should be obtained for different populations to allow their use in kinship tests and individual identification.

References

1. Kidd KK, Speed WC, Pakstis AJ, Podini DS, Lagacé R, Chang J, *et al.* Evaluating 130 microhaplotypes across a global set of 83 populations. Forensic Sci Int Genet 2017;29:29–37.

2. Børsting C, Morling N. Next generation sequencing and its applications in forensic genetics. Forensic Sci Int Genet 2015;18:78–89.

3. Oldoni F, Kidd KK, Podini D. Microhaplotypes in forensic genetics. Forensic Sci Int Genet 2019;38:54–69.

4. Kidd KK, Pakstis AJ, Speed WC, Lagace R, Chang J, Wootton S, *et al.* Microhaplotype loci are a powerful new type of forensic marker. Forensic Sci International Genet Suppl Ser 2013;4:e123–4.

5. Kidd KK, Pakstis AJ, Speed WC, Lagacé R, Chang J, Wootton S, *et al.* Current sequencing technology makes microhaplotypes a powerful new type of genetic marker for forensics. Forensic Sci Int Genet 2014;12:215–24.

6. Pakstis AJ, Fang R, Furtado MR, Kidd JR, Kidd KK. Minihaplotypes as lineage informative SNPs and ancestry inference SNPs. Eur J Hum Genet 2012;20:1148–54.

7. Kidd KK, Speed WC. Criteria for selecting microhaplotypes: mixture detection and deconvolution. Investig Genet 2015;6:1.

8. Kidd KK, Speed WC, Wootton S, Lagace R, Langit R, Haigh E, *et al.* Genetic markers for massively parallel sequencing in forensics. Forensic Sci International Genet Suppl Ser 2015;5:e677–9.

9. Bulbul O, Pakstis AJ, Soundararajan U, Gurkan C, Brissenden JE, Roscoe JM, *et al.* Ancestry inference of 96 population samples using microhaplotypes. Int J Legal Med 2018;132:703–11.

10. de la Puente M, Phillips C, Xavier C, Amigo J, Carracedo A, Parson W, *et al.* Building a custom large-scale panel of novel microhaplotypes for forensic identification using MiSeq and Ion S5 massively parallel sequencing systems. Forensic Sci Int Genet 2020;45:102213.

11. Khubrani YM, Hallast P, Jobling MA, Wetton JH. Massively parallel sequencing of autosomal STRs and identity-informative SNPs highlights consanguinity in Saudi Arabia. Forensic Sci Int Genet 2019;43:102164.

12. Phillips C, Amigo J, Tillmar AO, Peck MA, de la Puente M, Ruiz-Ramírez J, *et al.* A compilation of tri-allelic SNPs from 1000 Genomes and use of the most polymorphic loci for a large-scale human identification panel. Forensic Sci Int Genet 2020;46:102232.

13. Sun S, Liu Y, Li J, Yang Z, Wen D, Liang W, *et al.* Development and application of a nonbinary SNP-based microhaplotype panel for paternity testing involving close relatives. Forensic Sci Int Genet 2020;46:102255.

14. Pang JB, Rao M, Chen QF, Ji AQ, Zhang C, Kang KL, *et al.* A 124-plex Microhaplotype Panel Based on Next-generation Sequencing Developed for Forensic Applications. Sci Rep 2020;10:1945.

15. Chen P, Zhu W, Tong F, Pu Y, Yu Y, Huang S, *et al.* Identifying novel microhaplotypes for ancestry inference. Int J Legal Med 2019;133:983–8.

16. King JL, Churchill JD, Novroski NM, Zeng X, Warshauer DH, Seah LH, *et al.* Increasing the discrimination power of ancestry- and identity-informative SNP loci within the ForenSeq™ DNA Signature Prep Kit. Forensic Sci Int Genet 2018;36:60–76.

17. Phillips C, McNevin D, Kidd KK, Lagacé R, Wootton S, de la Puente M, *et al.* MAPlex - A massively parallel sequencing ancestry analysis multiplex for Asia-Pacific populations. Forensic Sci Int Genet 2019;42:213–26.

18. van der Gaag KJ, de Leeuw RH, Laros JF, den Dunnen JT, de Knijff P. Short hypervariable microhaplotypes: A novel set of very short high discriminating power loci without stutter artefacts. Forensic Sci Int Genet 2018;35:169–75.

19. Zhu J, Lv M, Zhou N, Chen D, Jiang Y, Wang L, *et al.* Genotyping polymorphic microhaplotype markers through the Illumina® MiSeq platform for forensics. Forensic Sci Int Genet 2019;39:1–7.

20. Chen P, Yin C, Li Z, Pu Y, Yu Y, Zhao P, *et al.* Evaluation of the Microhaplotypes panel for DNA mixture analyses. Forensic Sci Int Genet 2018;35:149–55.

21. Oldoni F, Podini D. Forensic molecular biomarkers for mixture analysis. Forensic Sci Int Genet 2019;41:107–19.

22. Bennett L, Oldoni F, Long K, Cisana S, Madella K, Wootton S, *et al.* Mixture deconvolution by massively parallel sequencing of microhaplotypes. Int J Legal Med 2019;133:719–29.

23. Chen P, Deng C, Li Z, Pu Y, Yang J, Yu Y, *et al.* A microhaplotypes panel for massively parallel sequencing analysis of DNA mixtures. Forensic Sci Int Genet 2019;40:140–9.

24. Yang J, Lin D, Deng C, Li Z, Pu Y, Yu Y, *et al.* The advances in DNA mixture interpretation. Forensic Sci Int 2019;301:101–6.

25. Turchi C, Melchionda F, Pesaresi M, Tagliabracci A. Evaluation of a microhaplotypes panel for forensic genetics using massive parallel sequencing technology. Forensic Sci Int Genet 2019;41:120–7.

26. Turchi C, Pesaresi M, Tagliabracci A. A microhaplotypes panel for forensic genetics using massive parallel sequencing. Forensic Sci International Genet Suppl Ser 2017;6:e117–8.

27. Turchi C, Melchionda F, Pesaresi M, Fattorini P, Tagliabracci A. Performance of a massive parallel sequencing microhaplotypes assay on degraded DNA. Forensic Sci International Genet Suppl Ser 2019;7:782–3.

28. Kling D, Tillmar AO, Egeland T. Familias 3 - Extensions and new functionality. Forensic Sci Int Genet 2014;13:121–7.

29. Hill CR, Duewer DL, Kline MC, Sprecher CJ, McLaren RS, Rabbach DR, *et al.* Concordance and population studies along with stutter and peak height ratio analysis for the PowerPlex ® ESX 17 and ESI 17 Systems. Forensic Sci Int Genet 2011;5:269–75.

30. R: a Language and Environment for Statistical Computing; 2020 [Internet]. Available from: http://www.rproject.org/2019 [cited 2020, May 11].

Conflicts of interest.—The authors certify that there is no conflict of interest with any financial organization regarding the material discussed in the manuscript.

History.—Manuscript accepted: May 11, 2020. - Manuscript received: February 24, 2020.