ISFG 2019 - P595

EVALUATION OF A COSTUMISED PANEL OF 118 NOVEL MICROHAPLOTYPES SELECTED FOR FORENSIC IDENTIFICATION IN TWO MPS PLATFORMS

María de la Puente[1,2], Christopher Phillips[1], Catarina Xavier[2], Jorge Amigo[3], Ángel Carracedo[1,3], Walther Parson[2,4], María Victoria Lareu[1]

[1]Forensic Genetics Unit, Institute of Forensic Sciences, University of Santiago de Compostela, Spain
[2]Institute of Legal Medicine, Medical University of Innsbruck, Austria
[3]Fundación Pública Galega de Medicina Xenómica (FPGMX), Santiago de Compostela, Spain
[4]Forensic Science Program, The Pennsylvania State University, PA, USA

INTRODUCTION

Microhaplotypes (MHs) comprise two or more closely-sited SNPs in spans up to 200 nucleotides (nt)**.** MH alleles are defined by the **phased** combinations of the alleles of the component SNPs, that can be **obtained by Massive Parallel Sequencing (MPS)** if the SNPs are co-amplified in one sequence.

The advantages of MHs for forensics include a **higher level of polymorphism** when compared to single SNPs, while retaining the ability to be **amplified in short fragments** with higher probability of success for degraded DNA; and a **lack of stutter products** that can hinder mixture analysis when analysing STRs [1].

In this study, a set of **novel autosomal and X chromosome MHs** were combined in a single PCR capture based on **AmpliSeq** chemistry with **short amplicons** of 125-175 nt, including primers. This design was optimised and **evaluated in** both **Thermo Fisher Ion S5** and **Illumina MiSeq** sequencing platforms following a framework that allowed detailed assessments of sequencing quality, concordance of genotypes, sensitivity to low-level and degraded DNA and mixture detection capabilities.

MATERIALS AND METHODS

Microhaplotype discovery, screening and primer design

Publicly available **1KG Phase III data** [2] was queried for loci with **two or more polymorphic SNPs** with contrasting minor allele frequencies (MAF) higher than 0.1, in segments **shorter than 120 nt**.

Chromosomal regions with at least **10 Mb separation for autosomal and 5 Mb for X chromosome sites** were defined; candidate loci were placed into subsets accordingly and **ranked** by their overall **gene diversity** value (**GD**).

From each subset, the **most informative microhaplotype** fulfilling the quality requirements (including an absence of long poly-tracts, repetitive regions, Indels or structural variants) were incorporated into a **single-pool Hotspot panel targeting FFPE DNA** designed with Ion **AmpliSeq** Designer (TFS).

DNA samples, library construction and sequencing

Table 1: Samples selected for the evaluation of the MH panel on Ion S5 and MiSeq platforms. A total of **four sample sets** (I to IV) were used for assessment of the following:

I. **Sequencing quality and concordance** at two levels: inter-platform and with 1KG [2] and SGDP [3] databases.
II. Sensitivity to **low-level DNA**.
III. Sensitivity to **degraded DNA**.
IV. Capabilities for **mixture detection**.

* run in library duplicates
* * volume ratios of 2800M:9947A; only run on Ion S5

I. Coriell cell-line and control DNAS	II. Dilution series of 2800M*	III. Sonicated series of DNA 007 (40 kHz)	IV. Artificial mixtures**
NA18498 - 1KG	0.5 ng	0 min	1:1
NA06994 - 1KG	0.25 ng	90 min	1:3
NA07000 - 1KG	125 pg	180 min	1:7
NA11200 - SGDP	62.5 pg	240 min	1:15
NA07029	31.25 pg		
9947A			
007			
2800M			

	Ion S5	MiSeq
Library chemistry	Precision ID Library Kit	AmpliSeq Library PLUS
Barcodes	Ion Xpress Barcode Adapters	AmpliSeq CD Index Set A
Library quantification	Ion Library TaqMan Quantitation Kit	Agilent High Sensitivity DNA Kit
Library loading concentration	30 pM	7 pM
Libraries per templating reaction	32	32
Templating and sequencing chemistry	Ion S5 Precision ID Chef & Seq. Kit	MiSeq Reagent Kit v2 (300-cycles)
Modifications to manufacter's protocol	-	Increase library amp. cycles: 7 -> 10

Table 2: Chemistry, conditions and modifications to the standard protocols for library construction, templating and sequencing with the Ion S5 and MiSeq platforms

Data analysis and MH allele construction

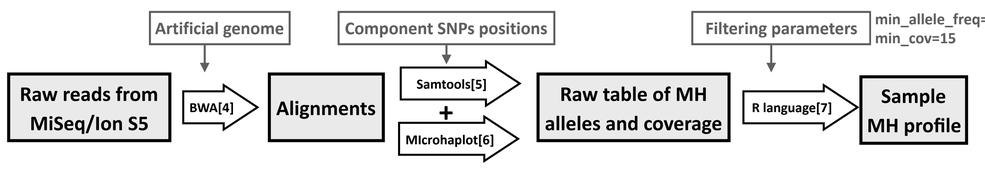


Figure 1: Flowchart representing the **custom analysis pipeline** based on open-source software used to obtain the **haplotypes of each MH locus** and their levels of sequence coverage from raw sequences generated on both MPS platforms.

RESULTS AND DISCUSSION

Marker selection and panel construction

A total of **107 autosomal and 11 X chromosome** highly polymorphic MHs were incorporated into the panel. **Mean GD values reached levels of 0.741 and 0.655, for autosomal and X chromosome, respectively**.

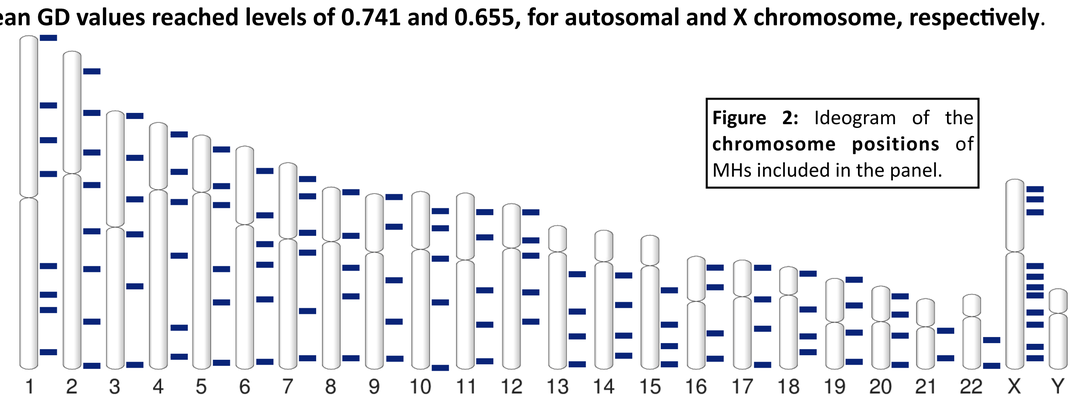


Figure 2: Ideogram of the **chromosome positions** of MHs included in the panel.

Concordance and underperforming SNPs

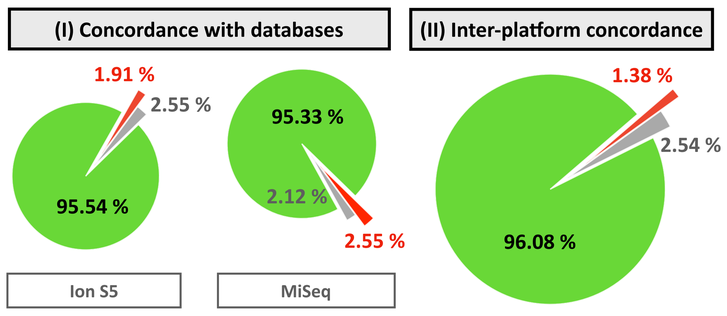


Figure 3: Pie charts representing the **concordance results** from the MH haploypes constructed through the custom pipeline, evaluated **at two levels**: (I) with online databases and (II) between platforms.

Results from **concordance** and **sequencing quality** (in terms of **coverage, strand bias, base misincorporation and allele balance**) were taken into account to assess the performance of the MHs included in the panel and classify them into three categories as detailed in Figure 4.

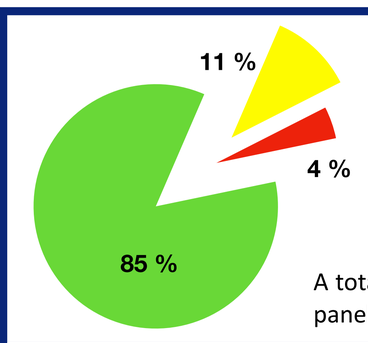


Figure 4: Pie chart represents results of evaluation of the **performance** of the MHs included in the panel, classified in the following categories:

- **Cannot** be **reliably** genotyped
- Require a **manual inspection** of the genotypes as they carry a higher risk of drop-in / drop-out when analysing challenging DNA
- Can be **reliably** genotyped

A total of **5 MHs** could not be reliably genotyped and were **excluded** from the panel and further analysis.

Sensitivity to low level DNA

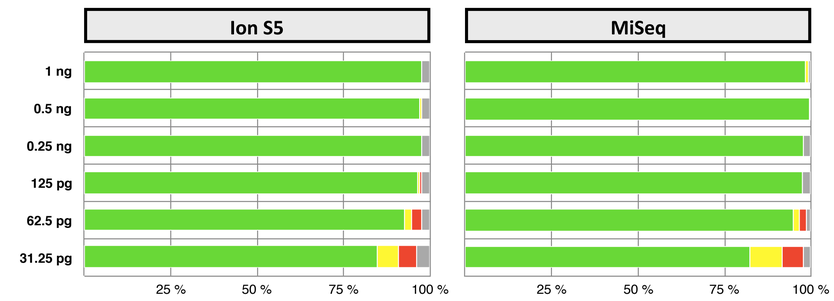


Figure 5: Bar chart showing genotyping performance for the **dilution series** in both platforms.

Profiles with a **95% completeness were obtained with 31.25 pg of DNA**. However, drop-out and drop-in rates increased at the lower DNA input amounts.

Sensitivity to degraded DNA

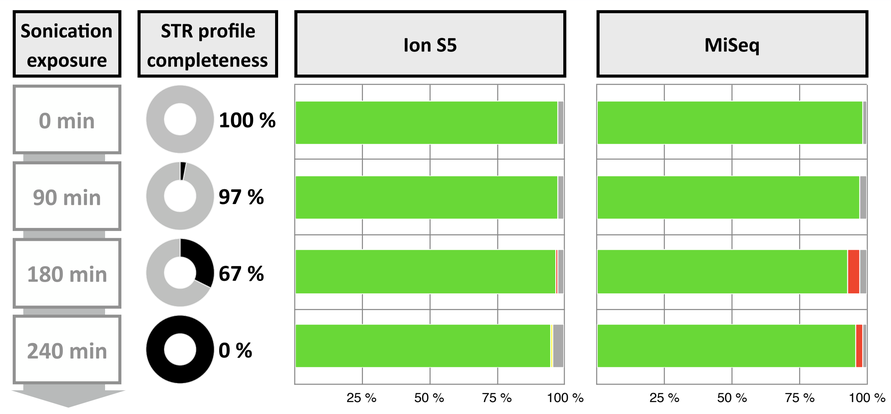
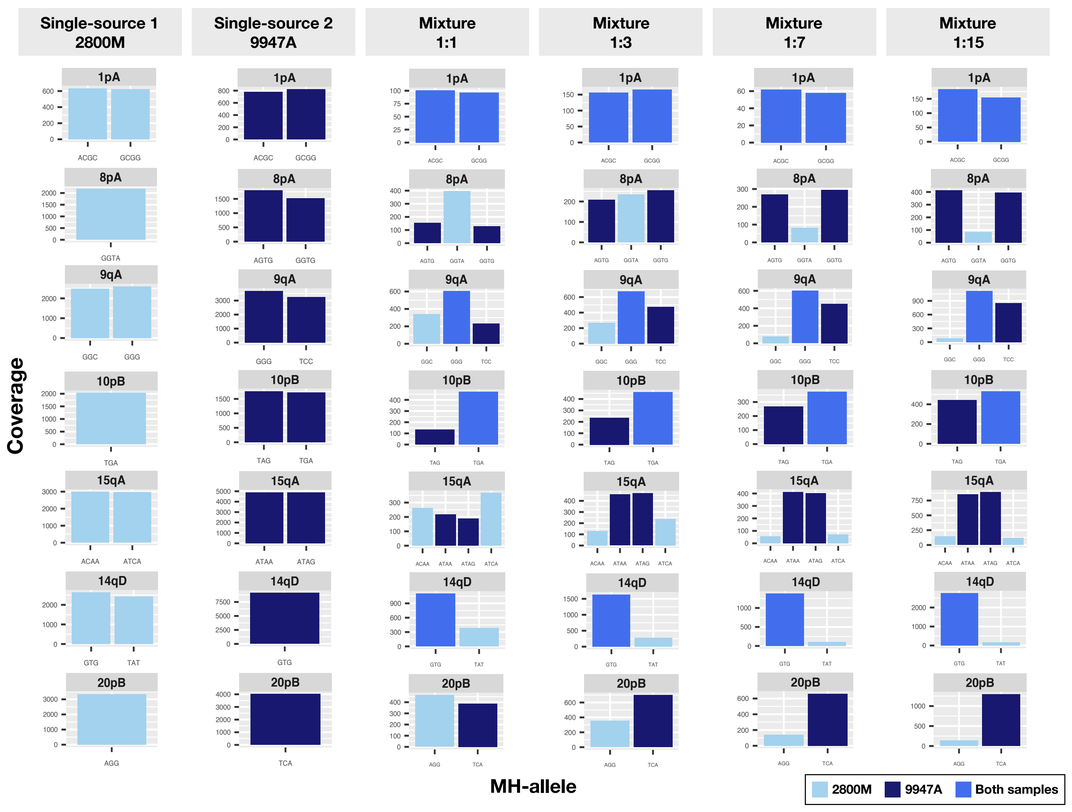


Figure 6: Genotyping performance for the **artificially degraded** DNA with both MPS platforms and common STRs.

Haplotypes with **95% completeness were obtained from an artificially degraded sample** that gave no results with mainstream STR profiling.

Mixture detection

Figure 7: Summary of **profiles** obtained from 1 ng input of **single-source samples** 2800M and **9947A** and **volume mixture** ratios of 1:1, 1:3, 1:7 and 1:15 with the Ion S5 platform. Bar colours indicate the source of each allele in the mixture according to the legend.



Mixtures were **easily detected** and, taking into account relative sequence coverage, **alleles could be assigned** to a major or minor component **in imbalanced mixtures**, in a similar way to mainstream STR profiles.

CONCLUSIONS

- A total of **118 novel and highly polymorphic MHs** - 107 autosomal and 11 X chromosome loci - were combined in a single-multiplex capture PCR using Ampliseq chemistry and **implemented in** two MPS platforms: **Ion S5 and MiSeq**
- **Efficiently phased** MH-haplotypes were obtained through a **custom analysis pipeline**. **High data concordance rates** (>95%) were obtained **between platforms** and **with online genome databases**.
- Discordances were mainly restricted to **5 MHs** that were consequently **excluded** from the panel. A total of **13 MHs** require **manual correction** of genotypes when analysing challenging DNA.
- **Forensic sensitivity** to low-level and degraded DNA showed **promising results**, as ~95% data completeness was obtained from 31.25 pg of input DNA or a sonicated sample that gave no profile with STRs. Stochastic effects such as **drop-ins and drop-outs** should be expected when analysing **challenging DNA.**
- **Mixtures** were **detected** in a straightforward way and **deconvolution achieved** in imbalanced mixture ratios.

References

[1] F. Oldoni, K.K. Kidd, D. Podini, Microhaplotypes in forensic genetics. Forensic Sci. Int. Genet. 38 (2018) 54-69.
[2] The Genomes Project Consortium, A global reference for human genetic variation. Nature 526 (2015) 68-74.
[3] S. Mallick, H. Li, M. Lipson, I. Mathieson, M. Gymrek, F. Racimo, M. Zhao, N. Chennagiri, S. Nordenfelt, A. Tandon, et al., The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature 538 (2016) 201-206.
[4] H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25 (2009) 1754-1760.
[5] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, The Sequence Alignment/Map format and SAMtools. Bioinformatics 25 (2009) 2078-2079.
[6] N. Thomas, R package - microhaplot. https://github.com/ngthomas/microhaplot.
[7] R: A language and environment for statistical computing. http://www.r-project.org/

Acknowledgements

This research is part of the MAPA project, funded by grant BIO2016-78525 (AEI, Spain). MdlP is supported by postdoctoral fellowship ED481B 2017/088 (Xunta de Galicia).