Reverse Complement PCR as a routine tool for Massively Parallel Sequencing of forensic STR loci; the IDseek® OmniSTR™ Global STR kit

K. J. van der Gaag[1], E.A.C. de Jong[2], N. Weiler[1], J. Hoogenboom[1], J.P.G. Theelen[2] and T. Sijen[1]

***1** Division of Biological Traces, Netherlands Forensic Institute, The Hague, The Netherlands **2** NimaGen B.V., Nijmegen, The Netherlands*

k.van.der.gaag@nfi.nl

Introduction

Forensic MPS assays tend to be applied foremost in specialty cases due to relatively high investment in both hands-on time and reagent costs. Reverse Complement PCR (RC-PCR) [1], a single-reaction closed-tube workflow, facilitates a routine MPS application (for the MiSeq™ instrument) due to a simplified workflow with reduced risk of contamination compared to other MPS methods hardly exceeding hands-on time of Capillary Electrophoresis (**Figure 1**).

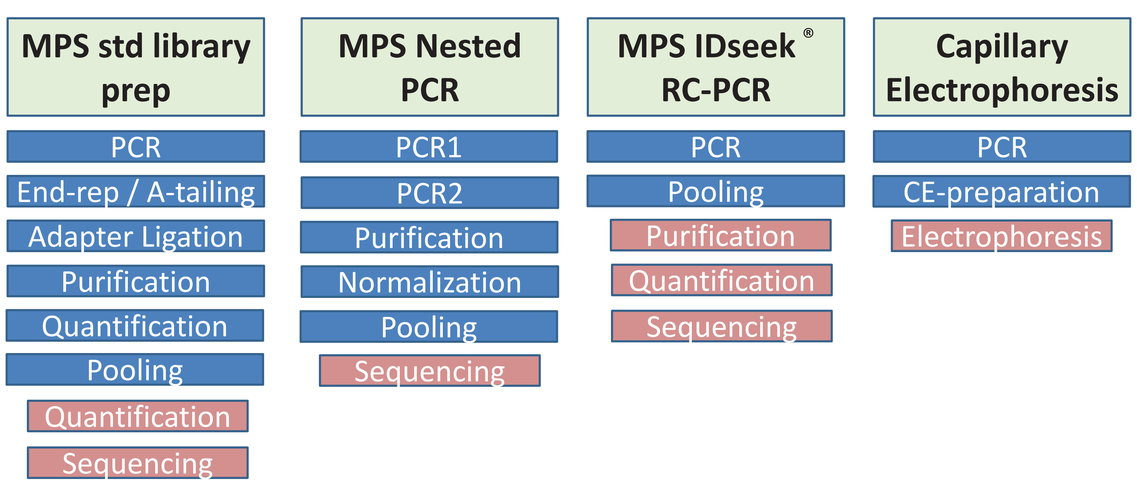

Figure 1. Different forensic lab workflows

Steps in blue are performed for each sample separately, steps in red are performed only once for each run

The IDseek® OmniSTR™ Global 30-plex includes all the US and European expanded core loci plus SE33, D4S2408, D6S1043, D9S1122, D17S1301, D20S482, PentaD, PentaE and DYS391 with short amplicons while maintaining the most variable sites in the flanking regions.

Methods

Performance characteristics and sequence variation for the IDseek® OmniSTR™ Global 30-plex were studied by analyzing:

- 284 reference samples (232 Dutch [2], 21 Himalayan [3], and 31 Pygmy [4])
- dilution series with PCR inputs ranging from 2ng-16 pg
- 20 direct PCRs on old swab fragments, also for Powerplex® Fusion 6C (PPF6C)
- two-person mixtures (1:2, 1:5, 1:10 and 1:20, two donor combinations in duplicate, 1ng input for 1:2, 1:5 and 1:10 and 1.5ng input for 1:20 mixtures)
- DNA-extracts of a sheep, goat, cow, pig, dog and cat

A noise correction database and analysis settings for references and traces were determined using FDSTools [5]. All analysis were performed using these settings without manual adjustments .

Results

Allele variation

Sequence variation for the STRs of 284 samples was determined and subdivided by CE-length and added variation from repeat sequences or flanking regions. Not surprisingly, SE33 exceeded the sequence variation of all markers.

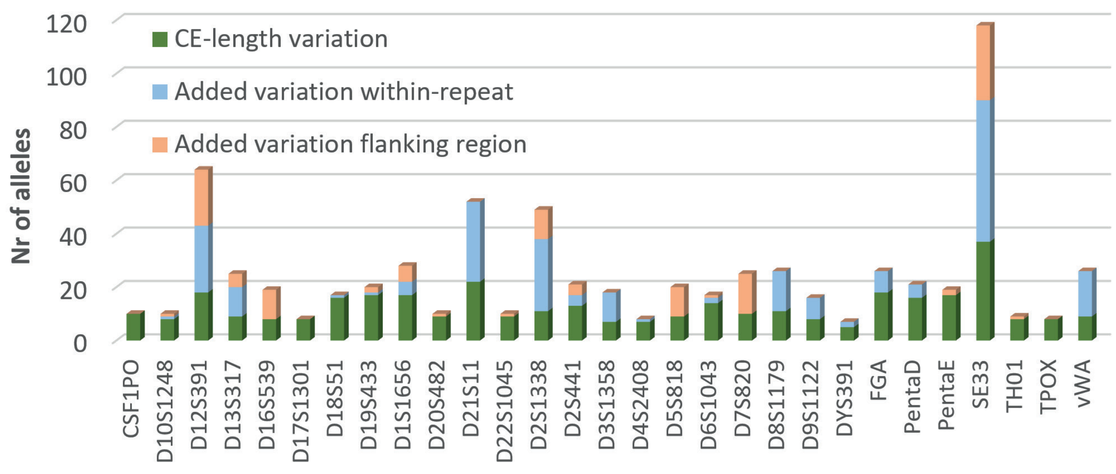

Figure 2. Allele variation observed in 284 samples

Sensitivity and direct PCR

Complete profiles were obtained down to 63pg DNA in the PCR while between-marker balance was only impacted to limited extend by the reduced input. Direct PCR however does negatively impact the marker balance for a few markers, namely D16S539, D2S1338, D21S11 and SE33 (**Figure 3**). PCR efficiency remains strong with about 90% reads on-target (excluding primer dimers) for 500pg PCR input as well as for direct PCR and about 79% for 63pg input.

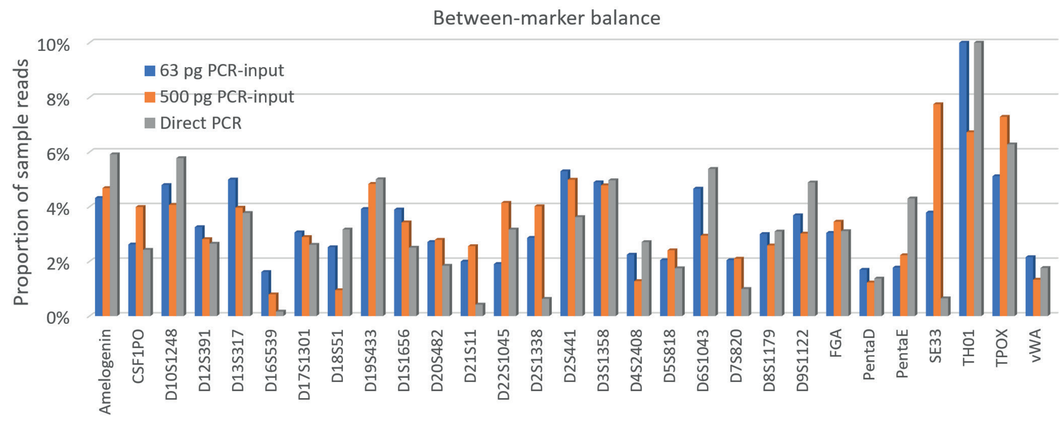

Figure 3. Marker balance for 500pg and 63pg PCR-input and direct PCR

Marker balance (DYS391 not displayed as the sample is female)

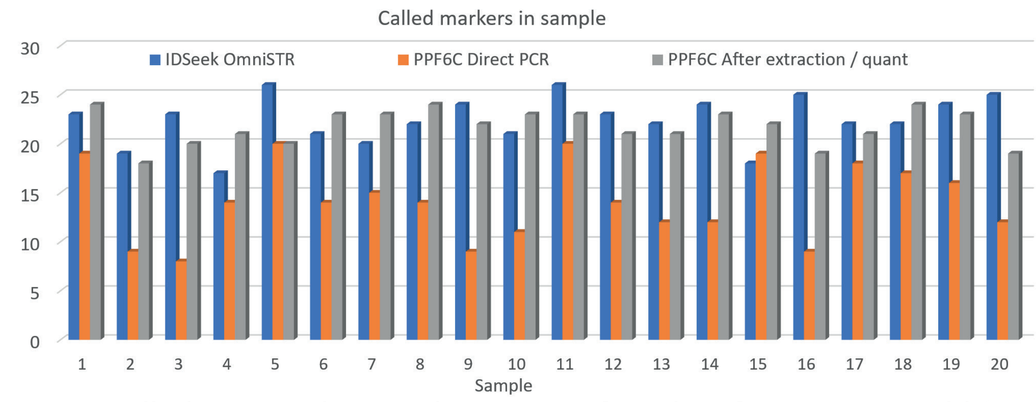

Figure 4. Called STR markers in direct PCR of swabs of 10-15 years old (partly degraded)

Comparison of the number of called markers in direct PCR for the IDseek® OmniSTR™, CE PPF6C and PPF6C after extraction and quantification

Mixture analysis

Using automatic analysis settings, very few alleles dropped out for 1:2 and 1:5 mixture ratios. For 1:10 and 1:20 mixture ratios the number of drop-outs increase **(Table 1),** mostly due to minor alleles overlapping with stutter positions of the major, but for the 1:20 mixtures occasionally due to complete drop-out of an allele. All drop-ins arise either from increased stutter or by 'switched repeats' (*repeat units that change one nt changing them to a different adjacent repeat sequence).*

Mixture ratio	**Succesfully typed minors** (excluding overlapping major alleles)	**% drop-ins** (as perc of all alleles)
1:2	99.2%	1.7%
1:5	97.6%	1.4%
1:10	87.3%	2.9%
1:20	74.6%	0.3%

Table 1. Called STR of minors in mixtures of different ratios

Comparison of the proportion of minor alleles called in the mixtures. No drop-outs were observed for the majors, drop-out percentages are calculated with respect to the number of alleles of the minor that are not shared with the major

Human specificity

No alleles were called for any of the tested animals

Discussion

The IDseek® STR OmniSTR™ kit provides an efficient MPS workflow suitable for high-throughput approaching the simple CE-workflow for hands-on time and suitable for a highly automated analysis.

Due to the short amplicon sizes (mostly 100-220bp except for SE33 that exceeds 300bp for long alleles), the performance for degraded samples is very good, even when applying direct PCR which can be very useful for upgrades of profiles from old samples that were typed using an old STR-assay with much fewer markers.

The RC-PCR is not only sensitive, but remains relatively stable for marker balance when low inputs are used, even with limited increase of primer-dimers thereby retaining an efficient use of sequencing reads which is not common for most MPS assays.

References

[1] M.M. Bus et al. Reverse complement-PCR, an innovative and effective method for multiplexing forensically relevant single nucleotide polymorphism marker systems. BioTechniques 2021 71:3, 484-489

[2] A. Westen et al. Comparing six commercial autosomal STR kits in a large Dutch population sample. Forensic Sci Int Genet. 10 (2014), 55–63.

[3] T. Kraaijenbrink et al. A linguistically informed autosomal STR survey of human populations residing in the greater himalayan region. PLoS One 9 (3) (2014), e91534.

[4] P. de Knijff and J. Pijpe. Population Genetics of African Pygmies, 2015. Unpublished work.

[5] J. Hoogenboom, K.J. van der Gaag et al. FDSTools: A software package for analysis of massively parallel sequencing data with the ability to recognise and correct STR stutter and other PCR or sequencing noise. Forensic Sci Int Genet. 27.

NFI, www.forensicinstitute.nl