An innovative and efficient Reverse Complement PCR method for analysis of highly degraded DNA

Magdalena M. Bus [a,b], **Melissa Muenzler** [a,b], **Sammed Mandape** [a], **Jonathan L. King** [a], **Erik A.C. de Jong** [c], **Joop Theelen** [c]

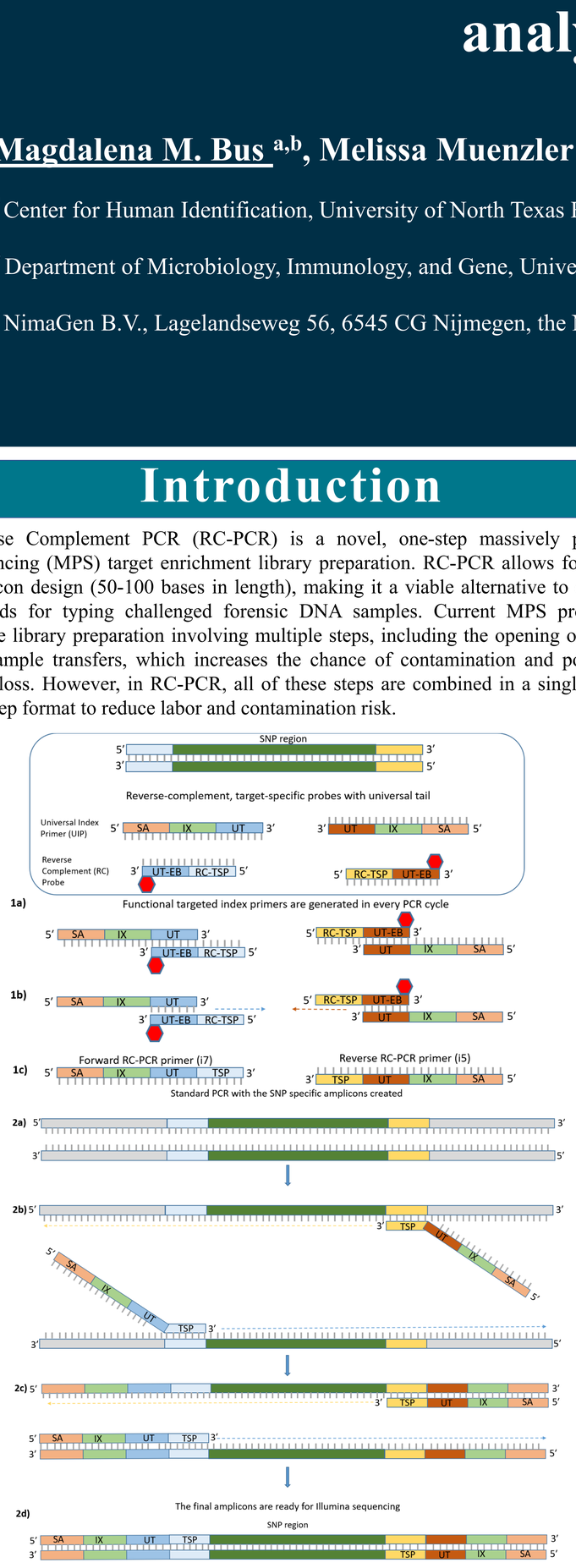
[a] Center for Human Identification, University of North Texas Health Science Center, 3500 Camp Bowie Blvd., Fort Worth, TX 76107, USA

[b] Department of Microbiology, Immunology, and Gene, University of North Texas Health Science Center, 3500 Camp Bowie Blvd., Fort Worth, TX 76107, USA

[c] NimaGen B.V., Lagelandseweg 56, 6545 CG Nijmegen, the Netherlands

Introduction

Reverse Complement PCR (RC-PCR) is a novel, one-step massively parallel sequencing (MPS) target enrichment library preparation. RC-PCR allows for short amplicon design (50-100 bases in length), making it a viable alternative to current methods for typing challenged forensic DNA samples. Current MPS protocols require library preparation involving multiple steps, including the opening of tubes and sample transfers, which increases the chance of contamination and potential DNA loss. However, in RC-PCR, all of these steps are combined in a single tube, one-step format to reduce labor and contamination risk.



[1] Buś et al. 2021, BioTechniques

Legend:
UT – universal tail
IX – index (single or unique dual index possible)
SA – sequence adapter
UT-EB – extension blocker
RC-TSP – reverse complement target specific primer
TSP – target specific primer

Graph 1. The graphical presentation of the RC-PCR technology. The universal index primer (UIP) contains a unique dual index i7/i5 (single indexing also possible), sequence adapter, and universal tail. The RC probe includes an extension blocker with a universal sequence and the reverse complement of the SNP target-specific region (F/R). The indexing and multiplex PCR amplification are performed at the same time in one closed tube. The RC-PCR consists of 2 major steps. **1a – 1c)** The universal tail sequences of UIPs hybridize with the corresponding forward/reverse RC probe, and the target specific index primers are generated. In each PCR cycle, new target specific index primers are generated by the *Taq* polymerase that copies the sequence of the RC probe. **2a – 2d)** The PCR amplification of SNP-specific amplicons. DNA samples are now tagged with a sample-specific index and Illumina sequence adapter. Samples then can be pooled, purified, and sequenced.

RC-PCR has been used to develop a short amplicon 85 SNP-plex panel for human identification purposes in fragmented and low copy number DNA. The new RC-PCR SNP panel includes 85 human identity SNPs. The majority of SNPs in the new panel are contained within amplicons that are < 100 bp (74 SNPs, 85%). See Figure 1 for SNP amplicon length distribution of the 85-plex RC-PCR panel.

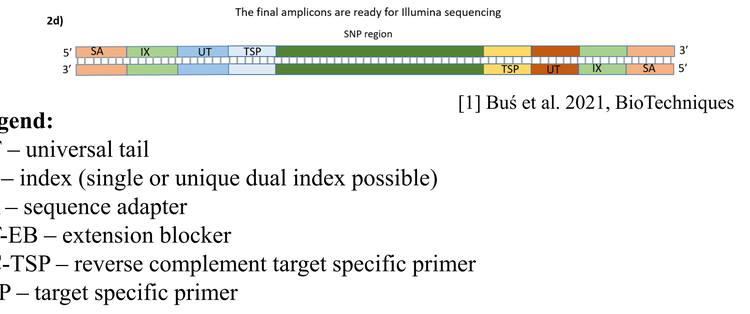


Figure 1. SNP amplicon length distribution in the 85-plex RC-PCR panel. The numbers represent the base pair length of an amplicon.

Materials and Methods

Samples The preliminary tests of the RC-PCR 85-plex system were focused on sensitivity and concordance of the multiplex. A DNA dilution series was performed with a high-quality, single-source, male, human, genomic 2800M DNA (Promega Corporation, Madison, WI, USA). The sample was amplified independently at total DNA inputs of 3 ng, 2 ng, 1 ng, 500 pg, 250 pg, 125 pg, 62 pg, and 31 pg. A similar dilution series was tested with four DNA samples extracted from blood that displayed high heterozygosity detected in a previous study using the ForenSeq™ DNA Signature Prep Kit (ForenSeq) [2].

Subsequently, SNP markers included in the 85-plex panel were amplified at total inputs of 1 ng and sequenced in DNA samples from Caucasian (N=48), Hispanic (N=48), and African-American (N=48) populations.

For testing the whole mtDNA RC-PCR panel a dilution series was performed on seven samples – one DNA 2800 reference sample and six population samples (African-American = 2, Caucasian = 2, Hispanic = 2). DNA samples were diluted to 400, 300, 200, 100, 50, 25, and 12.5 pg (100 pg = approximately 2900 mtDNA copies).

RC-PCR Library Preparation and Illumina Sequencing The PCR amplification and library preparation were performed according to the EasySeq Beta Human ID 85-plex SNP genotyping by NGS sequencing kit Quick Reference Guide v. RC-HID85-v0 (NimaGen BV, Nijmegen, The Netherlands) [3]. The libraries were pooled equimolar and diluted to 9 pM. Sequencing was performed on the Illumina MiSeq FGx desktop sequencer (Illumina, San Diego, CA, USA) using the MiSeq FGx™ Reagent Kit (Verogen, San Diego, CA, USA) with a read length of 2 × 121.

Data Analysis The FASTQ files were analyzed with STRait Razor Online v.0.1.7 and STRait Razor v3 [4,5] and visualized with Integrative Genomics Viewer, IGV v.2.5.3 [6].

Results

The initial tests of the 85-plex SNP RC-PCR panel were focused on the sensitivity of detection based on SNP data recovery. The preliminary tests of 85-plex RC-PCR panel on DNA extracted from four blood samples confirmed concordance between genotypes obtained by EasySeq and ForenSeq. The average locus read depth ranged from 99 (SD=151) for 31 pg to 3530 (SD=2211) for 3 ng of total DNA input (Figure 2).

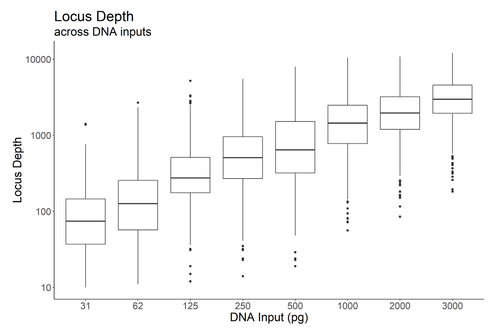


Figure 2. Boxplot of SNP locus read depth for eight DNA input amounts tested on high-quality samples (cell line 2800M and four blood samples) with hinges corresponding to first and third quartiles and outlying points representing values ± 1.5 x interquartile range.

The RC-PCR method revealed high data recovery with a majority of SNP alleles detected with a limited DNA input of 31 and 62 pg. Locus and allele dropout were primarily observed at low quantity inputs (i.e., 31 and 62 pg DNA), likely due to stochastic effects (Figure 3). However, the data obtained for low quantity DNA inputs ≤ 125 pg are promising since highly degraded DNA samples tend to be of low copy number.

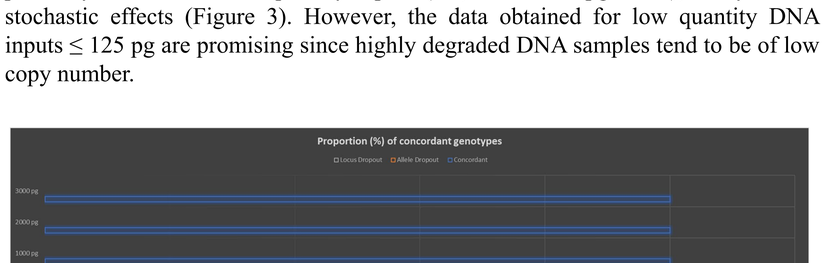


Figure 3. The proportion (%) of complete and concordant RC-PCR SNP genotypes, and detected locus and allele drop out received in preliminary sensitivity and concordance tests.

Data were obtained from over 50% of SNP loci among a set of 31 pg samples; in addition, the results were concordant with results from a previous study performed with the ForenSeq that includes the same identity SNPs. Recovery of the data from over 50% loci from as little as 31 pg DNA supports a high sensitivity of detection.

The preliminary results of sequence data analyses in population samples suggest that 81 SNPs display robust heterozygote balance (Figure 4) and four SNP loci - rs430046, rs1490413, rs1015250, and rs9071100 showed allele imbalance (see Figure 5). Heterozygote imbalance was operationally defined as HB < 0.5 or > 2.0.

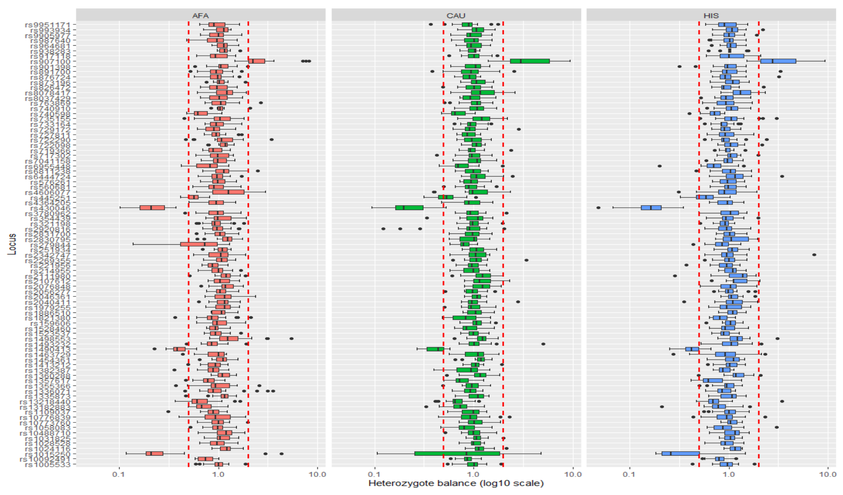


Figure 4. The distribution of mean heterozygote balance across 85 SNP loci included in the RC-PCR 85-plex panel. The heterozygote balance is calculated by dividing the largest allele (lexicographically) by the second largest allele (e.g., G/A).

Next, the four imbalanced SNP loci - rs430046, rs1490413, rs1015250, and rs9071100 – were highlighted (see Figure 5). Heterozygote balance/imbalance will be further evaluated with the applied stochastic and analytical thresholds.

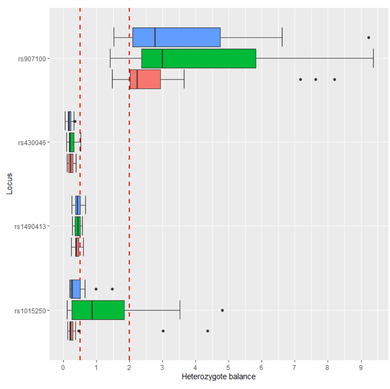


Figure 5. Mean heterozygote balance in the four SNP loci included in the RC-PCR 85-plex panel that demonstrated notable imbalance in three populations.

The preliminary results of the mixture detection revealed that the majority of the mixture components were consistent with the expected SNP proportions of two contributors. See Figure 6 for an example of a 1:1 mixture at 500 pg input.

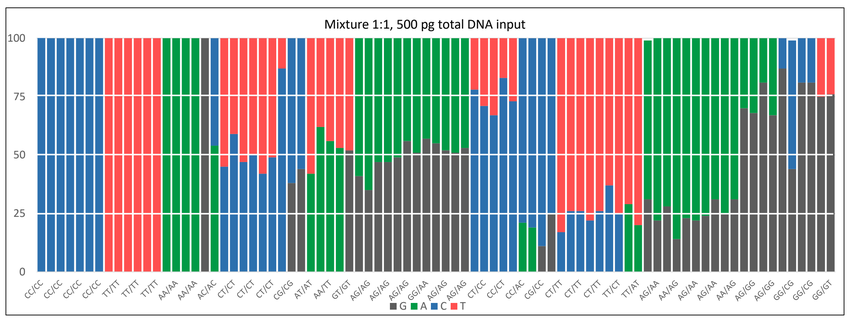


Figure 6. The proportion (%) of recovered RC-PCR SNP alleles in a two person 1:1 Hispanic population mixture. The total DNA input was 500 pg.

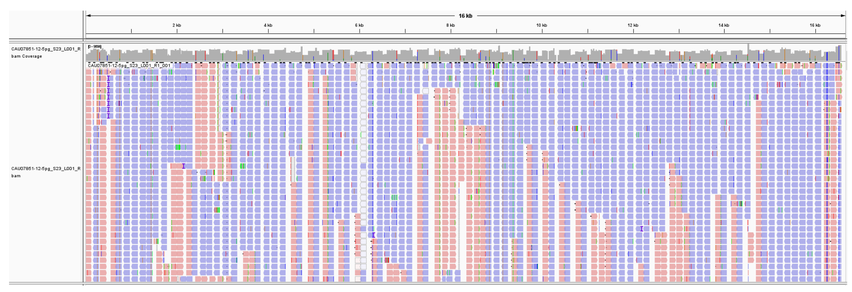


Figure 7. The whole mitochondrial DNA genome coverage for a sample from the Caucasian population. The total DNA input was 12.5 pg. The total tested DNA input was 12.5 pg.

Conclusions

The preliminary tests of the RC-PCR 85 SNP-plex panel suggest that the multiplex may be an efficient alternative to type low copy number samples and, ultimately, highly degraded/fragmented DNA. The data obtained for low DNA inputs ≤ 125 pg are promising since highly degraded samples tend to be of low copy number; even partial SNP profiles can provide critical information for, e.g., missing persons cases. The next steps of testing will include analysis of the IDseek® OmniSNP85™ Identity Informative SNP Typing Kit(formerly EasySeq 85-Plex) panel for mixture detection, inhibition resistance, and performance on degraded forensic samples and human remains. The first attempt to design and develop the RC-PCR whole mitochondrial DNA panel resulted in substantial sequencing coverage (Figure 7) of low copy number DNA samples and concordant mtDNA variant calls. The entire protocol from RC-PCR through library preparation can be accomplished within a few hours. The RC-PCR process reduces chances for sample loss and contamination by minimizing the number of library preparation steps, eliminating sample transfers between tubes, and reducing tube handling.

References

[1] Bus MM, de Jong EAC, van der Vliet W, Theelen J and Budowle B. Reverse complement-PCR, an innovative and effective method for multiplexing forensically relevant single nucleotide polymorphism marker systems. *BioTechniques* doi:10.2144/btn-2021-0031 (2021) (Epub ahead of print).
[2] King JL, Churchill JD, Novroski NMM *et al.* Increasing the discrimination power of ancestry- and identity-informative SNP loci within the ForenSeq DNA Signature Prep Kit. *Forensic Sci.Int. Genet.* 36, 60–76 (2018).
[3] NimaGen. The EasySeq Beta Human ID 85-plex SNP genotyping by NGS sequencing kit Quick Reference Guide.http://www.nimagen.com/gfx/Forensics/SNP85/protocol%20IDseek%20 SNP85v80-83.pdf.
[4] Woerner AE, King JL, Budowle B. Fast STR allele identification with STRait Razor 3.0. *Forensic Sci. Int. Genet.* 30, 18–23 (2017).
[5] King JL, Woerner AE, Mandape SN *et al.* STRait Razor Online: an enhanced user interface to facilitate interpretation of MPS data. *Forensic Sci. Int. Genet.* 52, 102463 (2021).
[6] Robinson JT, Thorvaldsdóttir H, Wenger AM, Zehir A, Mesirov JP. Variant Review with the Integrative Genomics Viewer. Cancer Res. 2017 Nov 1;77(21):e31-e34.

Acknowledgements

This work was funded in part by award no. 2020-DQ-BX-0005, awarded by the National Institute of Justice (NIJ), Office of Justice Programs, US Department of Justice.

The opinions, findings, and conclusions expressed within this publication are those of the authors and do not necessarily reflect those of the US Department of Justice.