Benchmark

BioTechniques

Reverse complement-PCR, an innovative and effective method for multiplexing forensically relevant single nucleotide polymorphism marker systems

Magdalena M Bus*[,1,2], Erik AC de Jong[3], Jonathan L King[1], Walter van der Vliet[3], Joop Theelen[3] & Bruce Budowle[1,2]

*[1]Center for Human Identification, University of North Texas Health Science Center, 3500 Camp Bowie Blvd., Fort Worth, TX 76107, USA; [2]Department of Microbiology, Immunology & Genetics, Graduate School of Biomedical Sciences, University of North Texas Health Science Center, 3500 Camp Bowie Blvd., Fort Worth, TX 76107, USA; [3]NimaGen B.V., Lagelandseweg 56, 6545 CG Nijmegen, The Netherlands; *Author for correspondence: Magdalena.Bus@unthsc.edu*

BioTechniques 71: 00–00 (September 2021) 10.2144/btn-2021-0031

First draft submitted: 30 March 2021; Accepted for publication: 12 July 2021; Published online: 5 August 2021

ABSTRACT

DNA analyses from challenging samples such as touch evidence, hairs and skeletal remains push the limits of the current forensic DNA typing technologies. Reverse complement PCR (RC-PCR) is a novel, single-step PCR target enrichment method adapted to amplify degraded DNA. The sample preparation process involves a limited number of steps, decreasing the labor required for library preparation and reducing the possibility of contamination due to less sample manipulation. These features of the RC-PCR make the technology a unique application to successfully target single nucleotide polymorphisms (SNPs) in fragmented and low copy number DNA and yield results from samples in which no or limited data are obtained with standard DNA typing methods. The developed RC-PCR short amplicon 85 SNP-plex panel is a substantial improvement over the previously reported 27-plex RC-PCR multiplex that will provide higher discrimination power for challenging DNA sample analyses.

METHOD SUMMARY

The single-step PCR library preparation method employs two reverse complement target-specific primer probes (RC probes) with a universal tail, Illumina universal i5 and i7 indexes, and sequence adapters. Region-specific primers are attached to a universal tail and contain a blocker at the 3′-end to prevent extension. During RC-PCR, functional region-specific, tailed index PCR primers are generated and extended followed by multiplex amplification of the target regions.

KEYWORDS:

forensic DNA ● high sensitivity ● human identification ● MPS ● RC-PCR ● SNPs

For a little over three decades, forensic biological samples have been analyzed with highly sensitive and resolving DNA-based genetic markers. Advancements in technology have increased the array of sample types that can be subjected to DNA analysis. While several human identification marker systems are available and validated for forensic DNA typing purposes and offer amplification of short amplicons, several forensic samples (e.g., human remains and touch evidence) do not contain sufficient quantity and quality of DNA to be typed with current methodologies. Forensic samples are subject to environmental conditions, such as UV light, high humidity, oxidation and microbial degradation, which can modify DNA structural integrity [1,2]. Typical DNA damages in challenging samples are nucleotide modifications and strand fragmentation. The type and degree of DNA damage are usually not predictable, as they depend on which environmental factors affect a sample and the intensity and length of exposure. As a result, challenging samples with high DNA fragmentation and/or damage may no longer contain sufficient intact templates for successful PCR amplification. Several studies have reported that DNA fragments in highly degraded samples are very short and may be approximately 50 bases in length [3–6]. Only a limited number of DNA markers, if any, may be detected with such short fragments. Therefore, information from such forensic samples cannot be obtained if the target regions are longer than the DNA fragments available for amplification. Several methods have been devised to improve DNA recovery from compromised samples [7–10]; however, such efforts have had limited or no success. Researchers must continuously develop and improve DNA typing methods to address the challenges of highly degraded forensic samples. A more successful approach to typing highly fragmented DNA has been the reduction of amplicon sizes of short tandem repeats (i.e., miniSTRs) or single nucleotide polymorphisms (SNPs) [11,12]. Decreasing amplicon size has been successful in allowing the analysis of degraded DNA. Several currently available approaches include shorter amplicons but may still fall short of providing a typing success rate for highly fragmented DNA samples [13,14]. While SNPs are the most abundant genetic variation in the human genome, generating a robust

multiplex of short amplicons is challenging. Our first attempt to design reduced 50 bp-long SNP amplicons resulted in only 27 markers meeting the criterion to be included in the multiplex [15]. A requirement for short amplicons leaves limited flexibility to design primers for PCR, due to molecular and thermodynamic constraints. Therefore, while maintaining 25 of the initial 27 SNPs, additional SNPs were added by relaxing the amplicon size of 50 bp, and the multiplex was expanded to 85 SNPs.

Two advances now may make it possible to analyze highly degraded DNA: reverse complement PCR (RC-PCR) and massively parallel sequencing (MPS). RC-PCR is a novel target enrichment and library preparation method for MPS [15]. Target enrichment and indexing are performed in a single closed-tube system. The method uses two reverse-complement, target-specific primer probes (RC probes) with a universal tail, Illumina universal i5 and i7 indexes (single indexing is also an option), and sequence adapters. Instead of targeting the flanking region of a DNA marker, the RC probe targets the Illumina index primer via the universal tail. During RC-PCR, a 3′-end blocker prevents extension of the RC probe while functional, target-specific index primers are generated and extended by the *Taq* polymerase that copies the sequence of the RC probe followed by multiplex amplification of target regions. The library is then ready for pooling, purification and sequencing (Graph 1, Supplementary File 1). With this system, the number of handling steps is substantially reduced, primarily because the PCR amplification and index and adapter ligation steps are contained in one RC-PCR step, which also reduces the possibility of contamination. The procedure from RC-PCR to loading pooled libraries for sequencing takes only a few hours, reducing turnaround time. With this design, fragments as short as approximately 50–100 nucleotides in length may be targeted, increasing the success of typing targets from highly degraded DNA samples. Lastly, MPS makes it possible to sequence these short-sequence libraries.

In a proof-of-concept study, Kieser and colleagues [15] demonstrated that RC-PCR technology could successfully amplify, with high fidelity, 27 SNPs with the targets contained within 50 bases in length. The method revealed a high sensitivity with a majority of SNP alleles detected with a limited DNA input of 60 pg (i.e., $<$10 cells of human genome equivalents) and successful analyses in the presence of PCR inhibitors (and, to a lesser extent, hematin and humic acid) and the potential for detecting mixed DNA profiles. The system showed typing success when evaluated using mock forensic samples as well as challenged casework bone samples. In this initial study, only 27 human-identity SNPs were tested as primer design and the constraining length of only 50 base long fragments did not enable the addition of other forensically relevant SNPs in the RC-PCR multiplex. However, by expanding the amplicon length to approximately 100 bp, it was possible to add 60 SNPs to produce a multiplex panel of 85 identity SNPs. Primer pairs were designed using Primer3Plus [16]. The enhanced panel includes SNPs contained in amplicons of lengths from 42 to 128 bp with 72 SNPs (~85%) less than 100 bp.

The preliminary tests of the RC-PCR 85-plex examined the sensitivity and concordance of the system. An input DNA dilution series was performed with a high-quality, single-source male human genomic 2800M DNA (Promega Corporation, 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. The serial dilution was prepared following the V1 = V2xC2/C1 formula (V1: volume of starting DNA solution; C1: concentration of starting solution; V2: final volume of new solution; C2: final concentration of new solution), preparing each desired concentration from the same starting concentration. A subset of diluted samples was quantified using a Qubit fluorometer (Thermo Fisher Scientific, MA, USA) to verify that the desired concentrations were obtained. 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 [17]. 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) [18]. To verify the success of library preparation, the length was checked on the Agilent 2200 TapeStation (Agilent Technologies, CA, USA) using the Agilent D1000 High Sensitivity ScreenTape system according to the manufacturer's recommendations. A Qubit fluorometer (Thermo Fisher Scientific) was used to determine library concentrations. The libraries were pooled and diluted to 9 pM. Sequencing was performed on the Illumina MiSeq FGx desktop sequencer (Illumina, CA, USA) using the MiSeq FGx™ Reagent Kit (Verogen, CA, USA) with a read length of 2 $\times$ 121. The FASTQ files were analyzed with STRait Razor Online v.0.1.7 and STRait Razor v3 [19,20] with a 10$\times$ read depth threshold (see Supplementary File 2 for the STRait Razor .config file). A set of negative controls was sequenced. To prepare the negative control samples, 4 µl of deionized molecular-grade water were added to the RC-PCR master mix instead of a DNA sample and subsequently processed in the same manner as all DNA samples. While eight SNP alleles were observed across the negative control data using a 10$\times$ read threshold, alleles were not reproducible (i.e., no SNP alleles were observed more than once across all negative controls) and, thus, were considered allele drop-in. To assess the level of noise observed in the sensitivity series, the total proportion of nontarget alleles (i.e., 1 - the proportion of reads consistent with a called allele) within each locus was calculated. These nontarget reads may be produced from PCR/sequencing error, somatic variation within an individual and/or by low-level background contamination. Regardless, the relative abundance or proportion of these reads may complicate allele calling of trace-level contributors to DNA mixtures and may vary across loci and/or DNA inputs. Future efforts should assess the noise level of the individual reads relative to a true call to better assess noise variance across a larger dataset.

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 1).

The noise expressed as the sum of all nonallelic signals for a locus was greater at lower DNA inputs, and the detected noise proportion ranged from 3.0% (SD = 5.5%) for 31 pg to 1.5% (SD = 1.5%) for 3 ng (Figure 2).

The percent of recovered SNP loci (i.e., showing at least one allele) ranged from 78.2% for 31 pg to over 99% for 125–500 pg (and 100% for greater input amounts). Allele dropout (ADO) was observed for inputs of 31 pg and 62 pg (8.4% and 5.5%, respectively), with 0.4%, 0.7% and 0.2% allele dropout detected for 125 pg, 250 pg and 500 pg, respectively. While the observation of greater dropout for the 250 pg sample compared with the 125 pg sample may seem unexpected, the proportional difference between the two input amounts

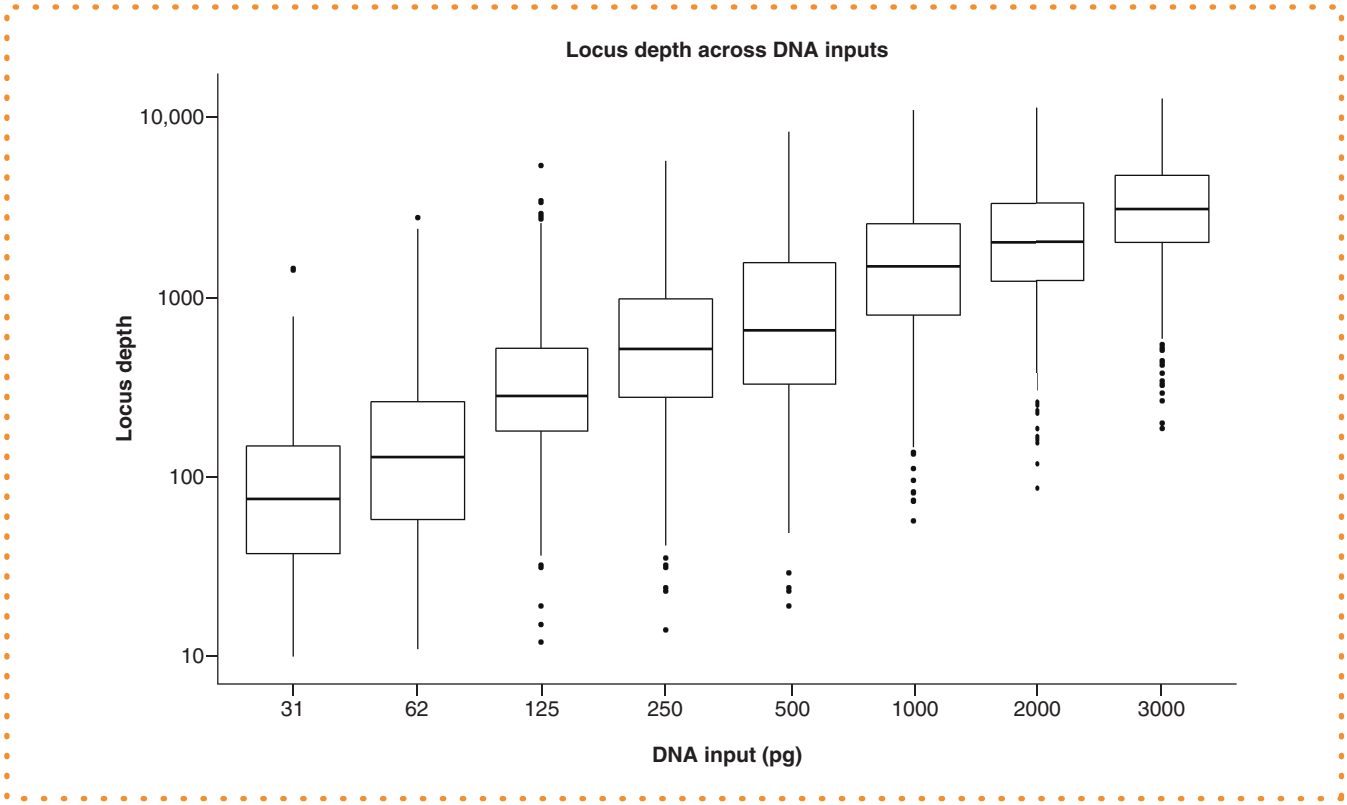

Figure 1. SNP locus read depth for eight DNA input amounts tested on high-quality samples (cell line 2800M and four blood samples) with hinges corresponding to first and third quartiles and outlying points representing values ± 1.5× the interquartile range.

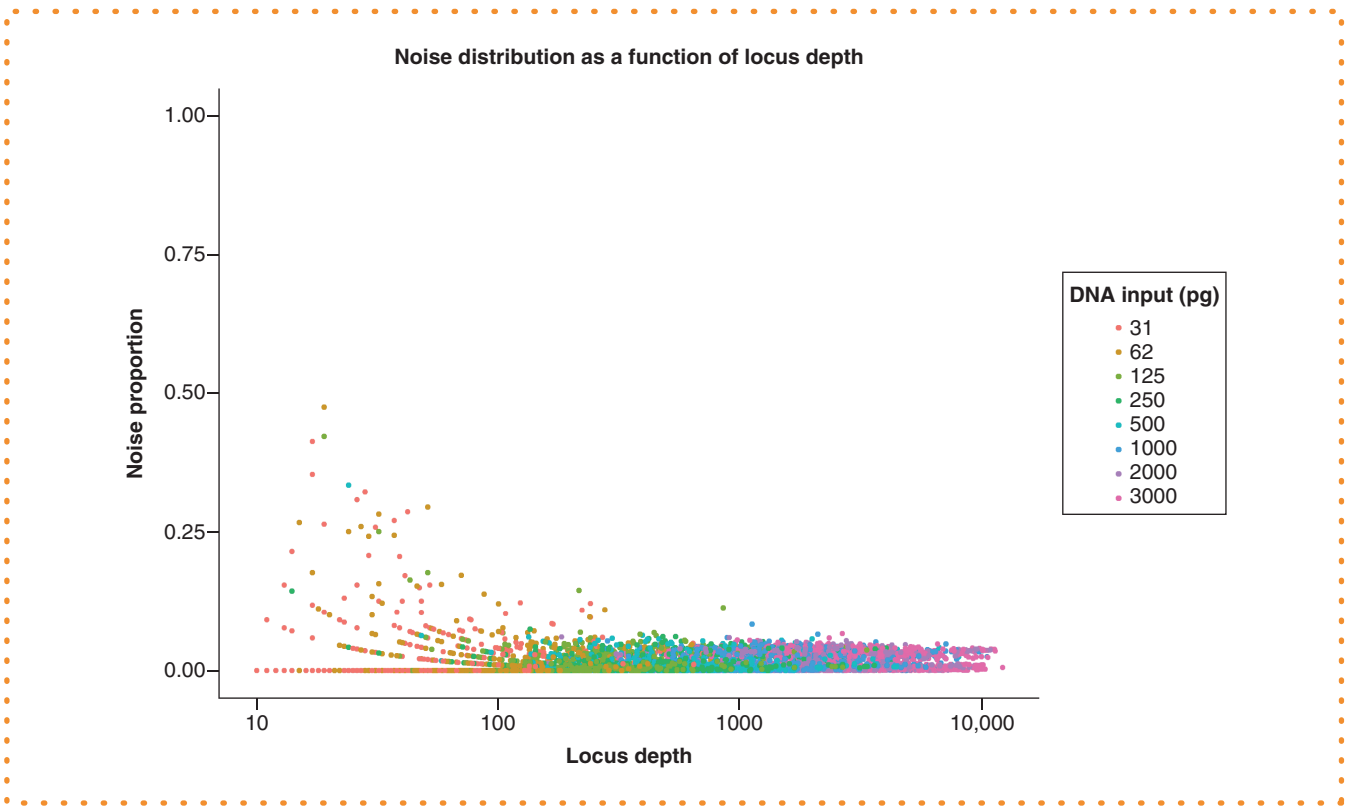

Figure 2. Proportion of nontarget reads within each locus as a function of total locus read depth across eight DNA input amounts.

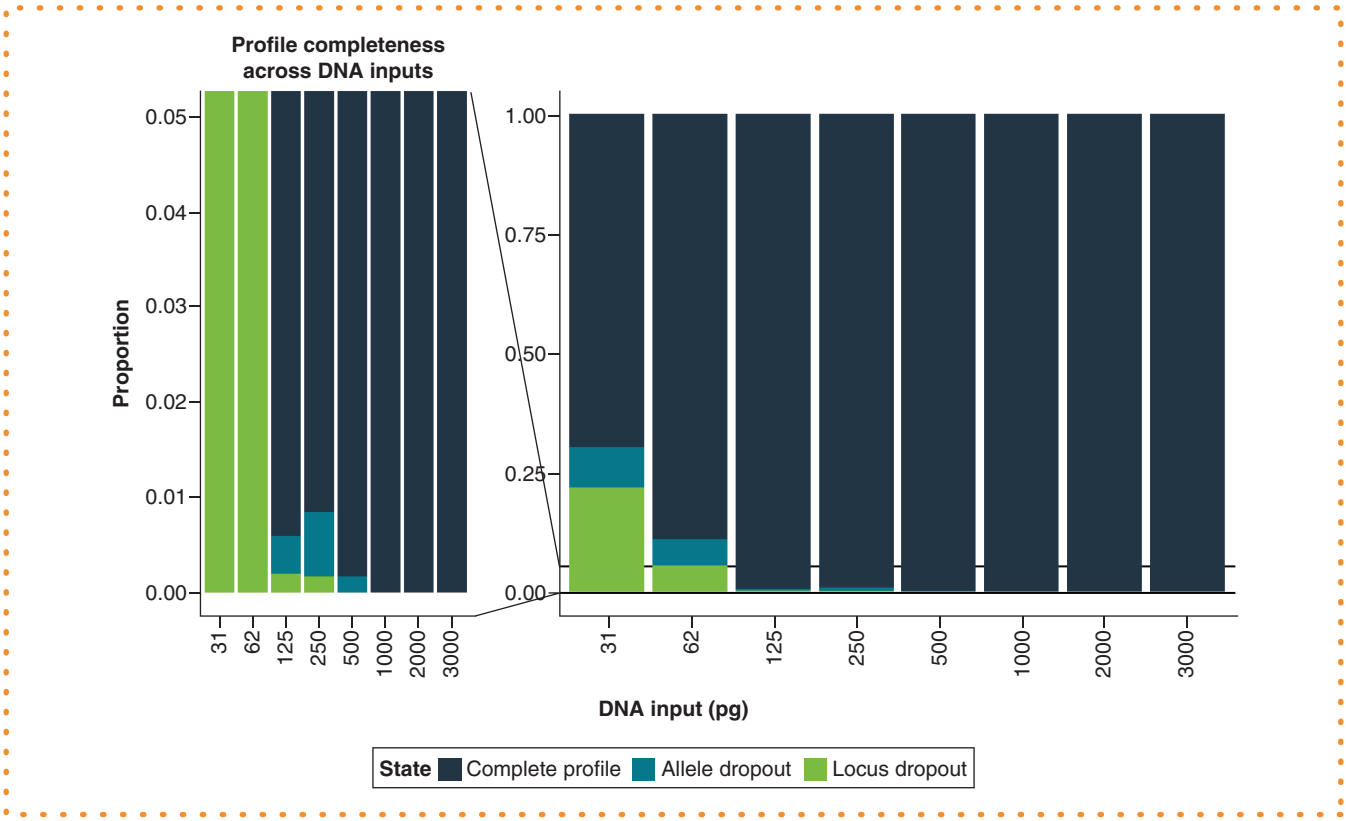

Figure 3. SNP profile completeness across eight DNA inputs, shown as the proportion of loci within each DNA input with all alleles observed (complete profile), at least one allele observed (allele dropout), and no alleles observed (locus dropout).

is not substantial and may reflect variance in amplification (i.e., many of the ADO events were observed at low-performing loci) or in the stochastic level (i.e., near-threshold read depths). Of the nine dropout events observed in profiles with greater than 62 pg of DNA, more than half (five of nine) occurred at the locus rs4606077, which somewhat underperformed concerning number of reads at all DNA inputs (data not shown). The genotypes typed with the RC-PCR method were 100% complete at input amounts >500 pg and genotypes were concordant with those previously typed with ForenSeq™ DNA Signature Prep Kit. The percent of concordant profile SNP loci ranged among low DNA inputs, with 69.8% loci for 31 pg, to over 99.0% loci concordant for 125–500 pg and 100% for greater input amounts (Figure 3).

Heterozygote balance (HB), calculated as second-most abundant allele divided by most abundant allele, was generally good at higher input levels (Figure 4); however, imbalance (i.e., ≤60% HB) was observed predominantly (i.e., ~50% of observations) at the lower (31 pg and 62 pg) DNA inputs. Additionally, major imbalance (<50%) was detected for all DNA input amounts in three loci, rs1490413, rs430046 and rs907100 which may be due to primer-binding site variants; however, a more comprehensive population study is needed to confirm the observation.

The preliminary sensitivity and concordance tests suggest that the 85-SNP RC-PCR panel is characterized by good performance. Locus dropout and allele dropout were observed predominantly for 31 and 62 pg input DNA. These dropouts are likely due to stochastic effects, as they correspond to the approximate quantity of nuclear DNA in five and ten cells, respectively, and are considered low template DNA samples [21,22]. 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 missing persons and criminal cases. While there is support for the utility of analysis for the RC-PCR system on forensic-type samples, the newly developed panel should be tested on challenging DNA samples in future studies. The next steps of testing will include analysis of the IDseek® SNP85 (formerly EasySeq 85-Plex) panel in different human populations, mixture detection, inhibition resistance and performance on degraded forensic samples and human remains. The technology may also be beneficial for typing formalin-fixed paraffin-embedded tissue samples. Given the diversity and throughput of MPS, other types of SNPs (e.g., ancestry informative SNPs) and other markers (e.g., short STRs and mitochondrial DNA) can be incorporated into the same multiplex system. The entire protocol from RC-PCR through library preparation can be accomplished within a few hours. Therefore, the method is a substantial improvement compared with most methods currently used in forensic MPS library preparation protocols, but the usefulness of the panel for casework samples typing requires further assessment. Lastly, the RC-PCR process reduces library preparation manipulation steps, including the opening and changing of the tubes and sample transfer,

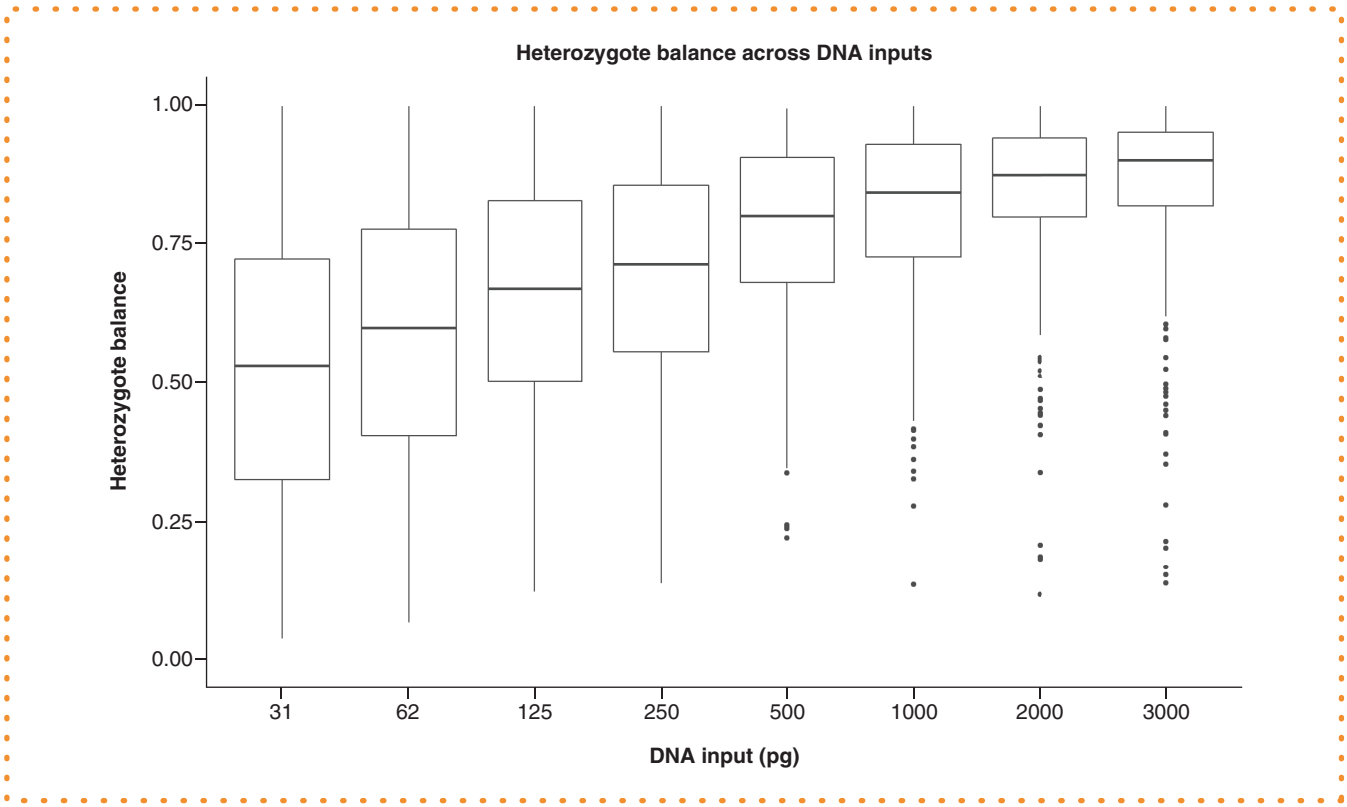

Figure 4. Heterozygous balance across eight DNA input amounts tested on high-quality samples (cell line 2800M and four blood samples) with hinges corresponding to first and third quartiles and outlying points representing values ± 1.5 x interquartile range.

which reduces the chance of contamination or DNA loss. The RC-PCR system may be an effective alternative to the current forensic genetic methods used to analyze highly degraded DNA, as well as high-quality DNA samples.

Author contributions

MMB: study design, library preparation, sequencing, data analysis and writing of the manuscript; JLK: data analysis and writing of the manuscript; BB: study design and writing of the manuscript; EACdJ, WvdV and JT: primer design, RC-PCR beta kit design and writing of the manuscript.

Financial & competing interests disclosure

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. The authors affiliated with NimaGen BV disclose financial competing interests, as they are employees of NimaGen BV. The authors affiliated with the University of North Texas Center for Human Identification (UNTCHI) disclose no competing interests. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Ethical conduct of research

The research and sample collection were performed in accordance with the approved policies and procedures set forth by the Institutional Review Board at the University of North Texas Health Science Center (UNTHSC) in Fort Worth, Texas, USA. The data are available on request.

Open access

References

Papers of special note have been highlighted as: • of interest

1. Allentoft ME, Collins M, Harker D *et al.* The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proc. R. Soc. B: Biol. Sci.* 279(1748), 4724–4733 (2012).
- **The publication presents that samples subjected to environmental factors may be degraded.**
2. Nandineni MR, Vendanayagem JP. Selective enrichment of human DNA from non-human DNAs for DNA typing of decomposed skeletal remains. *Forensic Sci. Int. Gen. Supplement Series* 2, 520–521 (2009).
3. Brandhagen MD, Loreille O, Irwin JA. Fragmented nuclear DNA is the predominant genetic material in human hair shafts. *Genes* 9(12), 640 (2018).
4. Damgaard PB, Margaryan A, Schroeder H, Orlando L, Willerslev E, Allentoft ME. Improving access to endogenous DNA in ancient bones and teeth. *Sci. Rep.* 5, 11184 (2015).
5. Gettings KB, Kiesler KM, Vallone PM. Performance of a next generation sequencing SNP assay on degraded DNA. *Forensic Sci. Int. Gen.* 19, 1–9 (2015).
6. Loreille O, Ratnayake S, Bazinet AL *et al.* Biological sexing of a 4000-year-old Egyptian mummy head to assess the potential of nuclear DNA recovery from the most damaged and limited forensic specimens. *Genes* 9(3), 135 (2018).
- **The publication supports the statement that forensic samples subjected to environmental factors may be highly fragmented.**
7. Ambers A, Turnbough M, Benjamin R *et al.* Modified DOP-PCR for improved STR typing of degraded DNA from human skeletal remains and bloodstains. *Leg. Med. (Tokyo)* 18, 7–12 (2016).
- **An example of research focusing on developing a new method for highly degraded DNA typing.**
8. Ambers A, Turnbough M, Benjamin R, King J, Budowle B. Assessment of the role of DNA repair in damaged forensic samples. *Int. J. Legal. Med.* 128(6), 913–921 (2014).
9. Bonnette MD, Pavlova VR, Rodier DN *et al.* dcDegenerate oligonucleotide primed-PCR for multilocus, genome-wide analysis from limited quantities of DNA. *Diagn. Mol. Pathol.* 18(3), 165–175 (2009).
10. Nelson JR, Cai YC, Giesler TL *et al.* TempliPhi, phi29 DNA polymerase based rolling circle amplification of templates for DNA sequencing. *BioTechniques* Suppl., S44–S47 (2002).
11. Butler JM, Shen Y, McCord BR. The development of reduced size STR amplicons as tools for analysis of degraded DNA. *J. Forensic Sci.* 48(5), 1054–1064 (2003).
- **An example of the reduction of the amplicon sizes for typing highly fragmented DNA.**
12. Freire-Aradas A, Fondevila M, Kriegel AK *et al.* A new SNP assay for identification of highly degraded human DNA. *Forensic Sci. Int. Genet.* 6(3), 341–349 (2012).
13. Elwick K, Bus MM, King JL, Chang J, Hughes-Stamm S, Budowle B. Utility of the Ion S5 and MiSeq FGx sequencing platforms to characterize challenging human remains. *Leg. Med. (Tokyo)* 41, 101623 (2019).
- **An example that there are methods with reduced amplicon sizes, but they may still be inefficient for some challenging samples.**
14. Xavier C, Eduardoff M, Bertoglio B *et al.* Evaluation of DNA extraction methods developed for forensic and ancient DNA applications using bone samples of different age. *Genes (Basel)* 12(2), 146; 1–15 (2021).
15. Kieser RE, Bus MM, King JL, van der Vliet W, Theelen J, Budowle B. Reverse complement PCR: a novel one-step PCR system for typing highly degraded DNA for human identification. *Forensic Sci. Int.Gen.* 44, 102201 (2020).
- **The first study using RC-PCR for typing forensic samples. The study was a basis for further improvements of the panel and the current research.**
16. Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JA. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.* 35(Web Server issue), W71–74 (2007).
- **The software used for primer design in the present study.**
17. 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).
- **The study which was the source of genotype data for testing concordance of results in the current study.**
18. 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%20SNP85v80-83.pdf**.
- **The RC-PCR manual.**
19. 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).
- **The software used for data analysis in the current study.**
20. Woerner AE, King JL, Budowle B. Fast STR allele identification with STRait Razor 3.0. *Forensic Sci. Int. Genet.* 30, 18–23 (2017).
21. Budowle B, Hobson DL, Smerick JB, Smith JAL. Low copy number – consideration and caution. In: *Twelfth International Symposium on Human Identification 2001* (2001). **http://www.promega.com/ussymp12proc/default.htm**.
22. Gill P. Application of low copy number DNA profiling. *Croat. Med. J.* 42(3), 229–232 (2001).
- **This study presents challenges associated with typing of low copy number DNA.**