Forensic Science International: Genetics 44 (2020) 102201

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsigen

Research paper

Reverse Complement PCR: A novel one-step PCR system for typing highly degraded DNA for human identification

Rachel E. Kieser[a,b,*], Magdalena M. Buś[a,b], Jonathan L. King[a], Walter van der Vliet[c], Joop Theelen[c], Bruce Budowle[a,b]

[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 Genetics, Graduate School of Biomedical Sciences, 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

ARTICLE INFO

Keywords:
Reverse Complement PCR
RC-PCR
SNPs
Identity SNPs
Degraded DNA

ABSTRACT

Reverse Complement PCR (RC-PCR) is an innovative, one-step PCR target enrichment technology adapted for the amplification of highly degraded (fragmented) DNA. It provides simultaneous amplification and tagging of a targeted sequence construct in a single, closed-tube assay. A human identification (HID) RC-PCR panel was designed targeting 27 identity single nucleotide polymorphisms (SNPs) generating targets only 50 base pairs in length. In a single reaction, the complete sequencing construct is produced which is essential for massively parallel sequencing (MPS) library preparation, thus reducing time and labor as well as minimizing the risk of sample carry-over or other forms of contamination. The RC-PCR system was evaluated and found to produce reliable and concordant variant calls. Also, the RC-PCR system demonstrated to have substantial sensitivity of detection with a majority of alleles detected at 60 pg of input DNA and robustness in tolerating known PCR inhibitors. The RC-PCR system may be an effective alternative to current forensic genetic methods in the analysis of highly degraded DNA.

1. Introduction

DNA analysis has proven to be one of the most reliable tools for human identification purposes. Typically, DNA profiling is grounded in the use of the polymerase chain reaction (PCR) for targeted amplification of specific regions within the human genome. These regions (i.e., loci) are selected due their high discriminatory power among individuals within the population(s). Primary analyses rely on polymorphic short tandem repeat (STR) loci that are size separated by the length of the amplicon via capillary electrophoresis (CE). These markers, however, have limitations, especially when examining highly damaged and degraded DNA. STR typing requires that the repeat and flanking regions of interest, defined by the primers, be intact for amplification. Degradation is inherent in DNA and a prominent issue to address motivating the development of alternative profiling techniques [1–3]. When DNA is exposed to damaging molecular influences, the integrity of the DNA molecular structure is affected. Nucleic acid modifications occur (e.g., depurination and base alteration) as well as structural changes such as crosslinking and strand breakage (fragmentation) [4,5]. Studies on highly degraded DNA from missing persons and ancient remains detail fragmentation of DNA to lengths of approximately 50 bases [2,6,7]. For example, the length of DNA in aged telogen hairs ranged from 49 to 88 bases [8]. The adverse influences limit the effective length (quality) and quantity of DNA fragments that can be amplified by PCR.

When forensic biological samples are compromised, it often becomes necessary to turn to alternative and/or additional DNA profiling methods for identification purposes. Single nucleotide polymorphisms (SNPs) may be pursued as the amplicons generated can be smaller than those of STR typing. SNPs consist of only a single base within a DNA molecule suggesting that they are more resistant to degradation and thus persist longer compared to STRs. SNP typing can be achieved via CE (e.g., minisequencing); however, the procedures available are laborious and non-quantitative [9]. With the advent of massively parallel sequencing (MPS), also known as next generation sequencing, SNP analysis can be quantitative and is a less demanding, more streamlined process. Current MPS protocols require library preparation involving multiple steps including the opening and transfer of tubes as well as sample movement which increase the chance of contamination, DNA loss, or even sample switches [10].

* Corresponding author.
E-mail address: Rachel.Kieser@my.unthsc.edu (R.E. Kieser).

https://doi.org/10.1016/j.fsigen.2019.102201
Received 30 August 2019; Received in revised form 22 October 2019; Accepted 3 November 2019
Available online 06 November 2019
1872-4973/

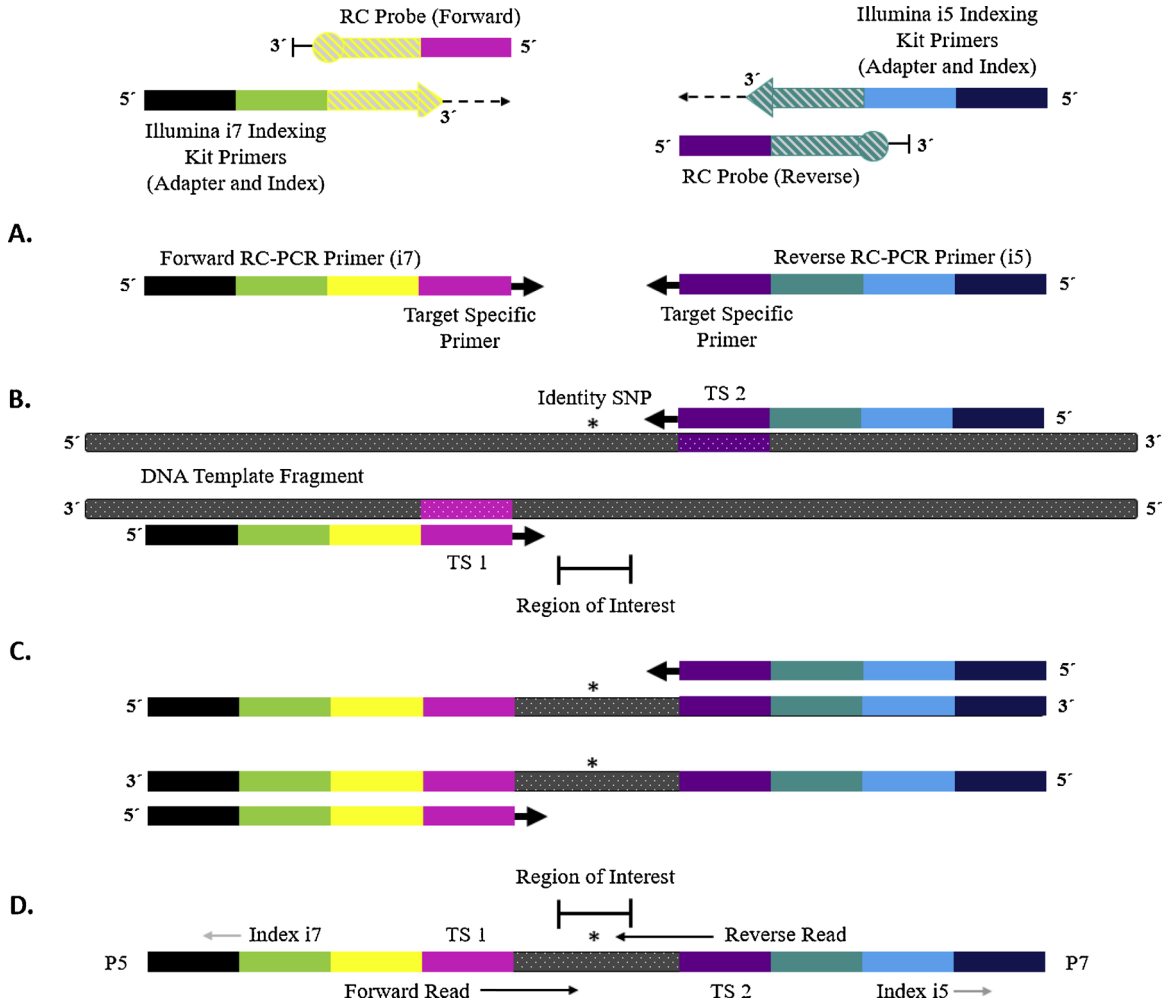

Fig. 1. Schematic depicting the unique capture and library generation of the RC-PCR technology for the enrichment of identity SNPs of interest for amplicons of only 50 base pairs in length. **A.** Target specific (TS) PCR primers are generated using the RC probes as template since they are the reverse complement of the genomic target sequence. This binding is facilitated by the initial annealing of the universal tails (slant patterned regions) to allow for elongation of the universal barcoding primers, thus generating the TS primer sequences. **B)** The initial copy of the tailed target sequence is produced after hybridization of the TS PCR primers to template DNA containing identity SNPs of interest. **C)** The library is amplified via the same RC-PCR primers generated in the primary step. **D)** The final Illumina library sequencing construct is formed [10].

Reverse Complement PCR (RC-PCR) is a one-step PCR technology (NimaGen B.V.; The Netherlands – exclusive license). While RC-PCR kits are currently in development for other purposes such as medical sample tracking and authentication, the system's first application for which it has been validated is for human identification purposes described herein [11]. Also, RC-PCR allows for small amplicon design making it a viable alternative to current methods for typing challenged forensic DNA samples [12]. The RC-PCR technology uses four probes: two target-specific primer probes (RC probes) with a universal tail and two universal barcoding primers. These barcoding primers contain sample-specific barcodes, the Illumina i5 or i7 index and adaptor sequences, and the complement sequence to the tail of the RC probes. The target-specific primers do not directly target the flanking region of interest, but rather are the reverse complement and are blocked at the 3′ end to halt extension (Fig. 1). In a single PCR step, a complete sequencing construct is generated that is ready for MPS library preparation [10,11,13].

This study describes the design and evaluation of the RC-PCR method as a forensic application. It is capable of targeting human identity informative SNPs contained within short amplicons (50 base pairs (bps) in length) thus enabling enrichment of targets in highly degraded (fragmented) DNA templates. Amplicon herein is defined throughout the evaluation as the 50 bp fragment produced upon successful targeting of the SNP of interest to include the primers and region of interest (excluding sequencing-specific construct components such as adapters).

2. Materials and methods

2.1. HID SNP panel design

A candidate list was compiled consisting of informative identity SNPs shared between commercially available forensic sequencing kits, the ForenSeq™ DNA Signature Prep Kit (Verogen, Inc.; San Diego, CA) and the HID-Ion AmpliSeq™ Identity Panel (Thermo Fisher Scientific; Waltham, MA) as well as well-defined in peer-reviewed literature [14,15]. SNPs demonstrating minimal allele frequency variation across major populations (low F_{ST}, usually below 0.06) and high heterozygosity increase the efficiency of a panel for human identification purposes [14,15].

Stringent requirements were applied to an in-house RC-PCR primer design pipeline (NimaGen B.V.) in order to achieve the short (only 50 bps in length) amplicons for the enrichment of highly fragmented DNA template. Due to the restricted size of the amplicon available for designing primers flanking the identity SNPs of interest, only 27 identity SNPs from the original candidate list (82 SNPs) were selected for

incorporation into a human identification (HID) RC-PCR panel for evaluation (Supplemental Table 1). These SNPs offered a high success rate probability and compatibility for multiplexing *in-silico* with the RC-PCR methodology.

2.2. Sample preparation

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, all samples were collected and used to validate the performance of the RC-PCR methodology. Each study is outlined below including sample preparation and analysis.

2.2.1. Population study

Fifty (N = 50) unrelated Caucasian individuals, previously genotyped using the ForenSeq™ DNA Signature Prep Kit, were sequenced using the RC-PCR technology with 1 ng of total DNA input. Concordance was evaluated between the genotypes reported with the ForenSeq™ DNA Signature Prep Kit (1 ng total DNA input) and the genotypes observed with the RC-PCR system for each corresponding population sample [17].

The reference samples used were collected from anonymous donors and were extracted using the QIAamp® DNA Blood Mini kit (Qiagen; Valencia, CA) according to the manufacturer's protocol and stored at −20 °C until needed [18]. Prior to use in the RC-PCR validation, all samples were quantified using the Qubit® 2.0 Fluorometer (Thermo Fisher Scientific) via the Qubit® dsDNA High Sensitivity Assay (Thermo Fisher Scientific) according to manufacturer's recommendations [19].

2.2.2. Analytical threshold determination

A subset of the population samples (N = 45) and three negative controls were used for determining the system's analytical threshold. This subset was selected because these samples were included on a single sequencing run, thus, removing potential influences of run-to-run variation. Forward reads were examined independently of the reverse reads at each SNP position.

2.2.3. Sensitivity study

A dilution series was performed with DNA samples of two individuals. Samples were prepared at total DNA inputs of 1 ng, 250 pg, and 60 pg. All DNA input quantities were confirmed using the Qubit® dsDNA High Sensitivity Assay according the manufacturer's protocol [19] and amplified by the RC-PCR system in duplicate.

2.2.4. Mixture detection

The same two individuals used in the sensitivity study were used for evaluating the system's capacity to detect and quantify mixtures. A total of 14 mixtures were prepared and analyzed. The mixtures were produced in varying ratios of individual 1 to individual 2 at 1:1, 1:9, 1:19, 1:49, 9:1, 19:1, and 49:1. Supplemental Table 2 lists the expected allelic proportions based on the known genotypes of both contributors.

2.2.5. Inhibitor tolerance

Common inhibitors known to co-extract with DNA were evaluated to determine their effects on the performance of the RC-PCR technology using the DNA from one of the previously studied contributors (from the sensitivity and mixture studies). The inhibitors examined were calcium, humic acid, collagen, and hematin. Inhibitor stock solutions were prepared, and all subsequent concentrations were diluted from the stock in molecular biology grade water (Phenix Research Products; Candler, NC). Calcium phosphate dibasic (100 mM) (Sigma Aldrich; Milwaukee, WI) was prepared in 0.5 N hydrochloric acid (VWR International; Radnor, PA) in a total volume of 10 mL. Humic acid (Sigma Aldrich) was prepared to a concentration of 1 mg/mL with molecular biology grade water in a total volume of 10 mL. Collagen from calf skin (Sigma Aldrich) was made in 0.1 N acetic acid (Fisher Scientific; Hampton, NH) to a total concentration of 1 mg/mL in 10 mL. Hematin (100 mM) (Sigma Aldrich) was prepared in 0.1 N sodium hydroxide (Fisher Scientific) in a total volume of 10 mL [20].

Serial dilutions of the stock inhibitors were prepared as described collectively in peer-reviewed literature specifically examining the inhibitory effects of these co-extracted materials [20–22]. All inhibitors were prepared at concentrations where 1 µL input achieved the final concentration for the reaction volume. These preliminary inhibitor concentrations were initially assessed using the GlobalFiler™ PCR Amplification Kit (Thermo Fisher Scientific) for conventional STR analysis to determine at what inhibitor concentration allele dropout (if any) was observed. Supplemental Table 3 lists the final inhibitor concentrations used for STR analysis. STR typing was performed with 1 ng of total DNA input and 1 µL of inhibitor for each reaction following the manufacturer's protocols [23]. PCR amplification was performed on the GeneAmp® PCR System 9700 Thermal Cycler (Thermo Fisher Scientific) and fragment separation was carried out on the Applied Biosystems ABI 3500xL Genetic Analyzer (Thermo Fisher Scientific). The collection data were analyzed with the GeneMapper® ID software v3.2.1 (Thermo Fisher Scientific) using a detection threshold of 50 Relative Fluorescence Units (RFUs) for allele calling [20].

Based on the STR inhibition results, three concentrations were selected for additional testing with the RC-PCR system. If there was no inhibition detected in the GlobalFiler™ STR profiles, then the highest three concentrations evaluated were selected for continued testing. If inhibition was observed, then the concentration yielding the last full STR profile, and the two subsequent concentrations (producing partial GlobalFiler™ STR profiles or complete dropout) were used for examining inhibitor tolerance with the RC-PCR methodology. All inhibitors (1 µL to achieve desired final concentration) were included in the 20 µL total RC-PCR volume during the initial library preparation setup.

2.2.6. Mock forensic casework samples

Mock forensic DNA samples were prepared from a variety of samples and substrate types. High quality and touch samples were examined in order to assess the bounds of the samples potentially encountered in forensic casework. Two samples were prepared by depositing 10 µL of the biological fluid (saliva or semen) on 2 cm by 2 cm cotton swatches. Additionally, 10 µL of blood, saliva, and semen were placed on clean glass slides. These samples were allowed to dry overnight in a hood at room temperature. Cuttings (approximately 1 cm by 1 cm) were taken from the biological stains. A sample cutting (about 1 cm by 1 cm) of a bloodstain on a paper towel from an anonymous donor previously collected also was tested. Stains deposited on glass slides were collected following the wet/dry swabbing method where 30 µL of molecular biology grade water were used to moisten one side of a cotton swab head (Puritan®; Guilford, ME) in order to dislodge and suspend the stain from the slide followed by absorption with the dry side of the swab. In addition, the mouth region of a used coffee cup and touch samples were collected from various objects employing the wet/dry swabbing method described. Touch samples included the ear pad of set of headphones, the grip of a phone, and a worn necklace (the gem and the necklace clasp swabbed separately). All samples were extracted following the "Pretreatment for Various Casework and Reference Samples" protocol using the EZ1® DNA Investigator® kit (Qiagen) on the EZ1 Advanced XL (Qiagen) instrument following the manufacturer's recommendations [24]. These samples were quantified using the Qubit® dsDNA High Sensitivity Assay kit as well [19]. High quality samples with known contributors were examined for concordance. These individuals were previously genotyped by either the ForenSeq™ DNA Signature Prep Kit or the Precision ID Identity Panel (Thermo Fisher Scientific) on the Ion GeneStudio S5 (Thermo Fisher Scientific) with 1 ng total DNA input [17,25].

2.2.7. Challenged samples

Four problematic reference samples from a previous population study were examined. DNA was extracted with the QIAamp® DNA Investigator Kit following the protocol "Isolation of Total DNA from FTA and Guthrie Cards" as per the manufacturer's suggestion [26]. These samples were quantified on the ABI 7500 Real-Time PCR System (Thermo Fisher Scientific) using the Quantifiler® Trio DNA Quantification Kit (Thermo Fisher Scientific) [27] and genotyped using the ForenSeq™ DNA Signature Prep Kit for library preparation according to the manufacturer´s protocol [17]. Sequencing results indicated considerable heterozygous imbalance for some SNPs as well as allelic dropout suggesting some level of degradation or inhibition occurred. DNA extracts from these same four samples were amplified using the RC-PCR technology with a total DNA input of 1 ng or 4 µL if the samples contained a lower amount of DNA.

Additionally, DNA extracts from the remains of six challenged bone samples, submitted to the University of North Texas Center for Human Identification (UNTCHI) for genetic testing, were used to evaluate the utility of the RC-PCR system in casework. These bone samples were previously extracted following the phenol-chloroform (organic) extraction protocol described in Ambers et al. [28] and undergo an extensive cleaning process including multiple sanding, bleaching, and washing steps to reduce contamination with modern, exogenous DNA. The extracts were quantified subsequently by UNTCHI laboratory personnel following the manufacturer's recommendations for the Quantifiler™ Duo DNA Quantification Kit (Thermo Fisher Scientific) [29]. Aliquots (5 µL volume) of each DNA sample extract were provided by UNTCHI for RC-PCR testing. The total allowable sample input of 4 µL was used for system evaluation as none of the challenged bone sample extracts contained sufficient DNA template quantity for 1 ng total input.

Random Match Probability (RMP) calculations were determined for both the consensus STR profile results reported by UNTCHI as well as the SNP profiles generated via the RC-PCR methodology assuming that these samples were single source and taking into account the potential of allele dropout using $2p\text{-}p^2$ for loci displaying a single allele above the run specific analytical threshold [30]. Allele frequencies were derived from the Federal Bureau of Investigation (FBI) 2015 expanded STR database and those reported by the 1000 Genomes Project in the National Center for Biotechnology Information (NCBI) dbSNP database for the STR and SNP RMP calculations, respectfully [16,31]. RMPs were determined using the allele frequencies for the major US populations (STRs) and five major global populations (SNPs) [16,31].

2.3. Library preparation

The protocol was performed according to the EasySeq™ NGS Reverse Complement PCR (RC-PCR) Quick Reference Guide (NimaGen B.V.) where the complete workflow for efficient MPS library preparation is detailed [13]. All assay components are combined in a single tube where the genetic marker(s) of interest are targeted, amplified, and indexed/tagged without the traditional multiple PCR step approach of other MPS protocols. The sample specific universal barcoding primers were pre-aliquoted in PCR strip tubes to maintain proper indexing to sample position assignment. The RC-probe HID multiplex mix was 10X concentrated in a separate tube for master mix preparation dependent on the number of reactions (or samples) examined. Once thawed, a Probe-Polymerase master mix was prepared. For each sample, 0.2 µL of the RC-probe mix (10X), 10 µL of the RC-PCR HiFi Mastermix, and 1.8 µL of molecular biology grade water were combined and mixed in a separate centrifuge tube. Into each well, 12 µL of the Probe-Polymerase master mix were added. The reaction requires 4 µL of input DNA to be added to the system for a total reaction volume of 20 µL. The RC-PCR samples were placed on the Veriti Thermal Cycler (Thermo Fisher Scientific) and amplified according to the conditions detailed in Supplemental Table 4.

The RC-PCR libraries were pooled in equal volume amounts by run and taken through a bead-based purification as described in the manufacturer's recommended protocols [12]. An aliquot of each run's purified library was taken and diluted 1:9 with molecular biology grade water. The diluted library was verified on the Agilent 2200 TapeStation (Agilent Technologies; Santa Clara, CA) using the Agilent D1000 High Sensitivity ScreenTape system (Agilent Technologies) according to the manufacturer's recommendations [32]. The expected fragment length for the SNPs is between 150–160 bps (including the target 50 bps and sequences for library preparation) [13]. The concentration of the 1:9 diluted library was determined using the Qubit® dsDNA High Sensitivity Assay [19]. The library was diluted to 2 nM in elution buffer for denaturation and subsequently loaded onto the sequencing cartridge according to the recommended protocol [13]. For libraries having low diversity, PhiX Control v3 was spiked into the final purified library at 15 % prior to loading into the sequencing cartridge [33]. Sequencing was performed on the Illumina MiSeq desktop sequencer (Illumina; San Diego, CA) using the MiSeq FGx™ Reagent Kit (v3 600 cycles) (Verogen).

2.4. Sequence data analysis

Generated FASTQ files were aligned to the hg38 build reference genome using the BWA-MEM algorithm for the generation of *bam* files. BAM files were evaluated and sequence read depth for each SNP position was visually examined using the Integrative Genomics Viewer (IGV, https://software.broadinstitute.org/software/igv/). SNP allele calling (i.e., genotyping) was performed by using a custom configuration file of STRait Razor v3, a freely available software package [34]. The configuration file for STRait Razor v3 to generate allele calls for the 27 identity SNPs is listed in Supplemental Table 5. For instances where the number of alleles observed was reported, apparent homozygous loci are counted as a single allele and heterozygous loci are counted as two alleles.

2.5. Population data analysis

Each SNP locus was reviewed manually using an in-house Excel allele calling workbook. Allele frequencies were calculated employing the counting method for the Caucasian population data (N = 50). Population genetics analyses were performed including heterozygosity (expected, H_{exp}, and observed, H_{obs}) and tests for departures from Hardy-Weinberg equilibrium (HWE) using Genetic Data Analysis (GDA) [35].

3. Results and discussion

3.1. HID SNP panel selection

The primer design process relies on a number of factors in order to achieve successful targeting and multiplexing. Beyond the restrictions to design primers generating amplicons only 50 bps in length, other design characteristics must be taken into consideration such as primer annealing temperature and their compatibility with the overall system's salt and deoxyribonucleotide triphosphate (dNTP) concentrations for sequence generation. With stringent requirements applied to an in-house RC-PCR primer design pipeline (NimaGen B.V.), only 27 identity SNPs of the total 82 candidate SNPs met the design criteria to be included in the HID RC-PCR panel. An efficient primer pairing design to yield amplicons 50 bps in length and/or a favorable success rate probability *in-silico* for inclusion into the HID RC-PCR panel were not possible for the remaining candidate SNPs.

3.2. MiSeq run metrics

Sequencing was performed on the Illumina MiSeq desktop sequencer and the respective RC-PCR sequencing run metrics (i.e., cluster

density, clusters percentage passing filter) are detailed in Supplemental Table 6. A total of five sequencing runs were performed. The cluster densities ranged from 671 k/mm^2 to 1050 k/mm^2 and the lowest percentage of clusters passing filter was 88.4 %. All run metrics were within acceptable ranges [36].

3.3. Population study

Fifty (N = 50) unrelated Caucasian individuals, previously typed using the ForenSeq™ DNA Signature Prep Kit, were carried through RC-PCR library preparation and sequenced for genotype accuracy based on concordance. All 50 samples genetically profiled using the RC-PCR technology were 100 % concordant with the allele calls produced by the ForenSeq™ DNA Signature Prep Kit.

3.3.1. Population genetic analyses

Allele frequencies for the Caucasian population (N = 50) were determined per locus and are reported in Supplemental Table 7. Expected and observed heterozygosities were computed for the population samples studied. No significant departures from HWE expectations were detected (Supplemental Table 8).

3.3.2. Analysis of data quality

A subset (N = 45) of the population samples was used to calculate a series of system quality metrics. These samples were all contained within a single sequencing run and therefore would not introduce inter-run variation into the metrics. The average forward read depth (N = 45) is shown in Fig. 2. Supplemental Fig. 1 displays the average read depth of the reverse read. Both reads performed similarly with the reverse read yielding slightly lower read depth. Fig. 3 depicts the forward read's average heterozygous balance for each identity SNP. Supplemental Fig. 2 shows the average heterozygous balance for the reverse read. A majority of the loci yield relatively balanced heterozygous loci with the heterozygosity ratio above 0.6. Imbalance is observed in both reads specifically at SNPs: rs1015250, rs1493232, and rs1736442. These SNPs contain documented allele specific sequence variation in the flanking regions surrounding the SNP of interest that may have interfered with primer annealing and amplification [37]. The addition of a degenerate primer in these instances could improve performance at these SNP loci and potentially result in more balanced heterozygosity.

Read error was determined by evaluating the percentage of reads yielding alternate haplotypes (or alternative sequences between the established STRait Razor anchors) not contributing to allele calling over the total read depth at each SNP position. These reads are usually attributed to unincorporated primers (not containing the SNP insert), PCR artifacts, and/or sequencing error. Fig. 4 depicts the average percent error for each identity SNP targeted by the RC-PCR system separated by read directionality. The majority of the error is observed in the reverse read; thus, the findings for the forward read only are reported for the rest of the study herein.

3.3.3. Analytical threshold determination

A population sample subset (N = 45) was utilized to establish a RC-PCR system-based analytical threshold for variant calling to be used herein for analysis of study performance. This subset was used because all samples (N = 45, and three negative controls) were processed together on a single run eliminating potential cofounding run-specific sequencing variables. Because the true genotypes are known for these population samples, the corresponding RC-PCR data generated could be analyzed based on that ground truth.

The negative control samples were initially analyzed for background signal. This background signal could be the result of low-level contamination either in the reagents or consumables. For each SNP position, an average was obtained for all background signal across the three negatives for any reads resulting in an allele call. For PCR or sequence related noise, only those SNP calls in the samples that were homozygotes were assessed. The population sample subset (N = 45) was filtered for true homozygotes based on the concordance data. The alternate allele state was documented for each homozygote at every SNP. The read depth count of the alternate allele state (incorrect allele call) was counted; other SNP state noise was ignored. The mean count of each alternate allele state was calculated for each SNP. This mean (plus three standard deviations (SDs)) was set as a baseline count of the system's fixed sequence noise. Each SNP's fixed sequence noise (count) was then added to each specific run's average background read depth per SNP, thus, establishing specific analytical thresholds in a dynamic (run-specific) process.

Any allele call with a read depth below each SNP's analytical threshold was not used for variant calling. Additionally, if the read depth for an allele was above the analytical threshold, but less than 10 % of the corresponding allele, this allele was not called. The 10 percent criterion was based on the heterozygosity balance calculated for the

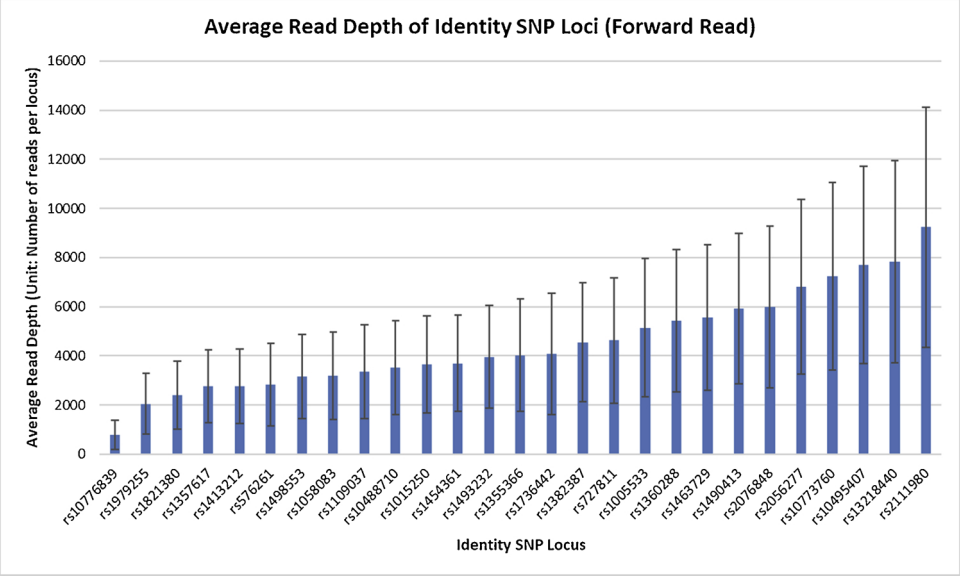

Fig. 2. Average read depth for each of the identity SNPs across the population study subset (N = 45) for the forward read of the HID RC-PCR panel. These samples were sequenced in a single run with total DNA input of 1 ng per sample for library preparation. Error bars represent one standard deviation.

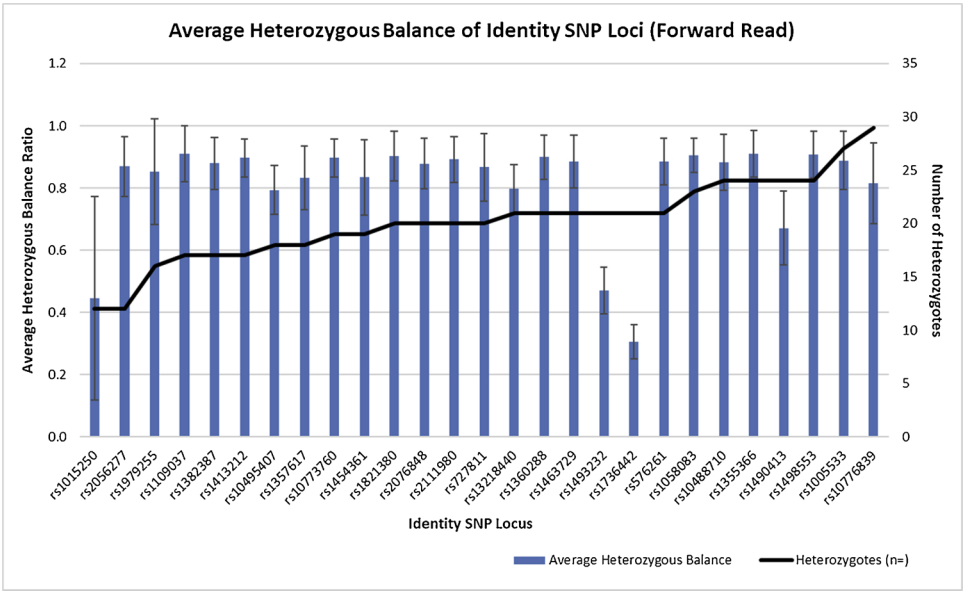

Fig. 3. Average heterozygous balance for the identity SNPs based on the forward read of the HID RC-PCR panel using the population sample subset (N = 45). A total DNA input of 1 ng per sample was used for RC-PCR library preparation, and all samples were sequenced in a single run. Average heterozygous balance is calculated by dividing the read depth of the lower performing allele by the read depth of the higher allele at a particular locus. A value of 1.0 indicates equal or balanced allelic read depth at a specific heterozygous locus. The horizontal black line represents the number of heterozygotes present within the population sample subset for each SNP. One standard deviation is represented by the error bars.

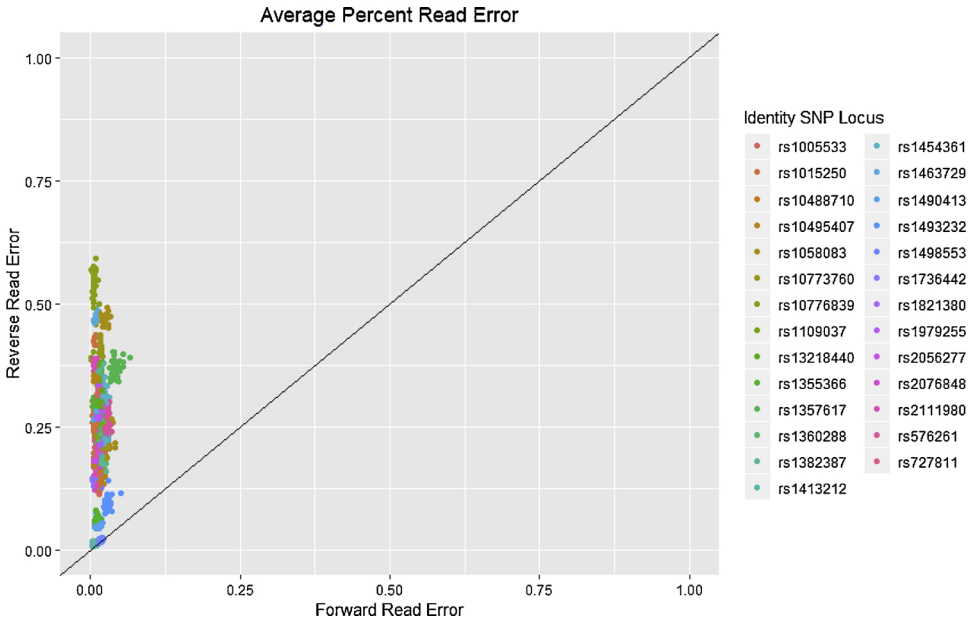

Fig. 4. Average percent read error per identity SNP locus yielding alternate haplotypes. The percent error is separated by sequence read directionality for the population sample subset (N = 45). The diagonal line depicts equal error between the forward and reverse read.

population samples (N = 45) and the lowest performing identity SNP's (rs1015250) average heterozygous balance minus one standard deviation (approximately 12 %) (as shown in Fig. 3).

3.4. Sensitivity study

As forensic samples may not yield sufficient DNA for an optimum template input, a sensitivity of detection study was conducted to evaluate the performance of the HID SNP panel with the RC-PCR system with less DNA. DNA extracts from two individuals were amplified at 1 ng (optimum), 250 pg, and 60 pg. The RC-PCR system was able to detect concordant alleles down to a total DNA input of 60 pg where a majority of the alleles (at a minimum of 83.7 %) was still detected (Supplemental Table 9).

3.5. Mixture detection study

Forensic biological samples are of unknown source and at times may yield mixtures. A total of 14 mixtures were prepared and analyzed in this study to assess the ability of the RC-PCR system to detect and quantify mixtures. The mixed samples were prepared in varying ratios of individual 1 to individual 2 at 1:1, 1:9, 1:19, 1:49, 49:1, 19:1, and 9:1. These ratios were evaluated at total quantities of 1 ng and 250 pg of input DNA. When comparing the known genotypes of the 27 identity SNPs for the two individuals, two SNPs were identified to display as the contributors were homozygous for the alternate allele states (rs1979255 and rs727811). The read depth at these two loci was further investigated to examine the method's sensitivity and ability to detect the appropriate contributor sample proportions. The majority of the mixtures examined was consistent with the expected proportions of two contributors for both 1 ng and 250 pg template amounts (Figs. 5 and 6). However, the read depth ratios are skewed slightly for the mixtures

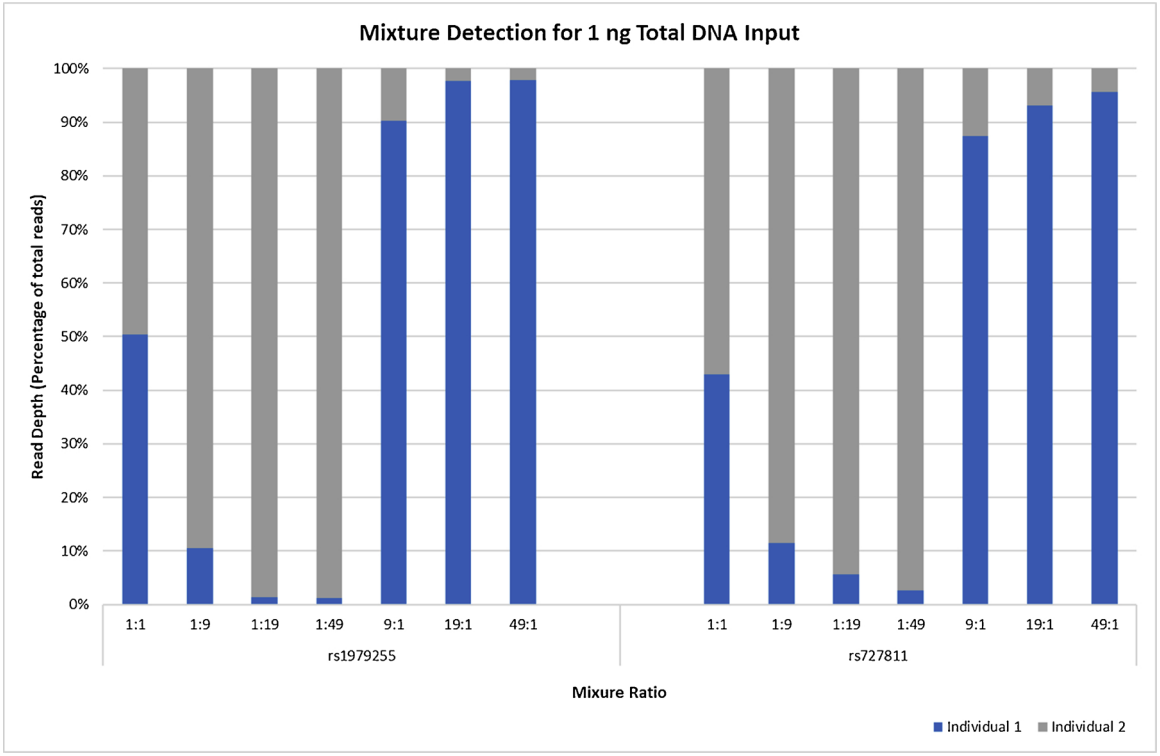

Fig. 5. Percentage-based read depth for a mixture of two individuals with a total DNA input of 1 ng at various mixture ratios. At SNP positions rs1979255 and rs727811, the contributors are homozygous for the alternate allele states. The expected allele proportions for this scenario are identified by an asterisk in Supplemental Table 2.

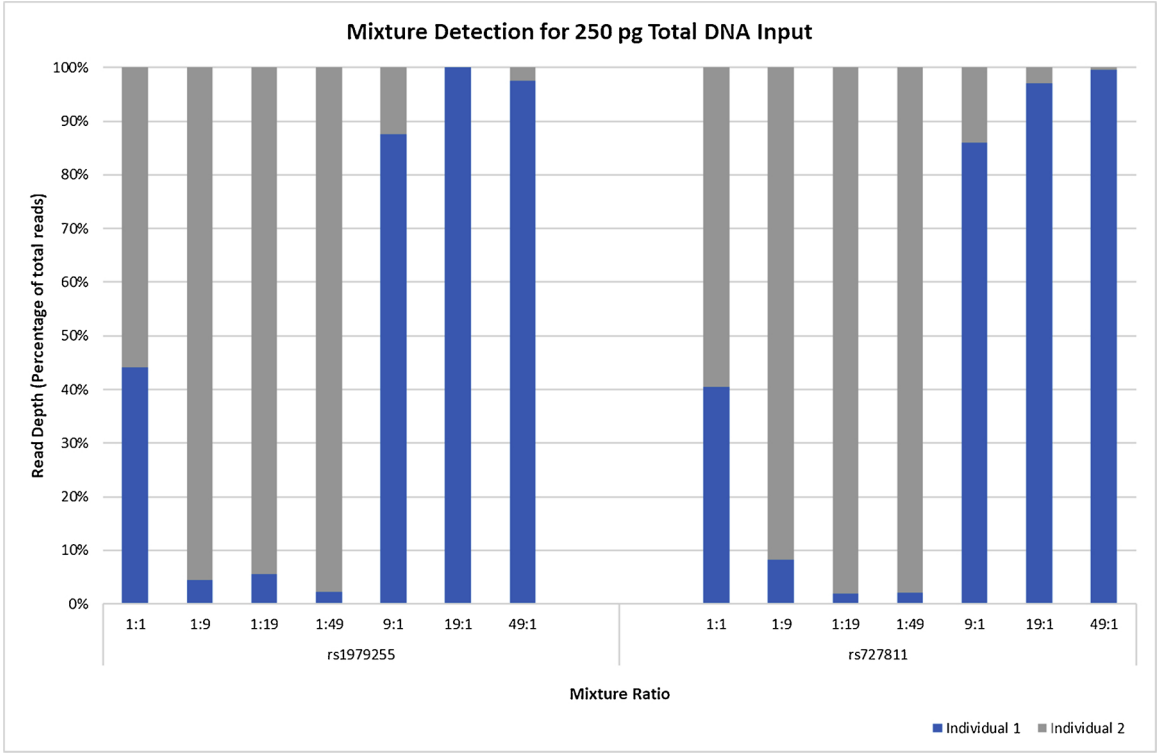

Fig. 6. Percentage-based read depth detection for a mixture of two individuals with a total DNA input of 250 pg at various mixture ratios. The two contributors are homozygous for the alternate allele at SNP positions rs1979255 and rs727811. The expected allele proportions for this scenario are identified by an asterisk in Supplemental Table 2.

with a total DNA input of 250 pg as would be expected with lower template quantities.

For all mixtures at each SNP, the allele read depth proportions (expected and observed) were computed based on read depth and plotted against each other for the various mixture ratios studied with a total DNA input of 1 ng (Supplemental Fig. 3). Supplemental Fig. 4 depicts the same comparison separated by mixture ratio. Quantitatively, the 1:1, 1:9, and 9:1 mixtures performed comparably to the expected allele proportions (Supplemental Fig. 4); however, the mixtures at 1:19, 19:1, 1:49, and 49:1 show increased variation in the read depth and stochastic effects compared with predicted proportions. No allele dropout was observed for mixtures consisting of 1 ng of total DNA input. Supplemental Figs. 5 and 6 show the results of the expected and observed allele proportions for all mixture ratios with a total DNA input of 250 pg. The lower template quantity mixtures depict increased variability most likely due to increased stochastic effects.

3.6. Inhibitor tolerance study

DNA profiling of forensic samples can be negatively affected by PCR inhibitors. Commonly known inhibitors are reported to either interfere with the DNA molecule or with the DNA polymerase consequently inhibiting PCR amplification. To evaluate the effects of commonly co-extracted inhibitors in forensic samples and the system's capacity to tolerate these effects, stock solutions were made of calcium, humic acid, collagen, and hematin. Inhibitor concentrations were prepared according to those reported in peer-reviewed literature [20–22]. Three concentrations for each inhibitor investigated were examined using the RC-PCR methodology. Supplemental Table 10 details the number of alleles observed for each inhibitor concentration for both STR profiling (GlobalFiler™) and SNP genotyping (RC-PCR).

Calcium inhibited STR typing as its concentration increased; however, for SNP profiling (RC-PCR) there was only a reduction in allele detection observed at the highest concentration of calcium (Supplemental Table 10). Calcium is a commonly found inhibitor co-extracted with DNA from skeletal remains [20]. It is classified as a Taq polymerase inhibitor. When oxidized, calcium is a competitive cofactor with magnesium for polymerase binding sites [20]. Perhaps the high-fidelity polymerase included in the RC-PCR HiFi Mastermix can better accommodate the inhibitory effects of calcium as the higher concentrations of calcium were tolerated. Additionally, Ralec et al. [38] reports that some polymerases (e.g., hyperthermophilic archael family B polymerases) seem to be able to utilize calcium as a cofactor for DNA polymerization although at a slower rate than magnesium [38].

In contrast, humic acid showed no inhibitory effect on STR profiling while there was considerable dropout for RC-PCR at the highest inhibitor concentration (Supplemental Table 10). Humic acid is a major component of soil often comprised of decomposed animal and plant material. It is an inhibitor that is known to intercalate within the DNA template [20], consequently limiting the availability of the DNA template for PCR amplification. Current inhibition models for humic acid suggest that the smaller amplicons should be more efficient in amplification; however, the RC-PCR system data do not fit this model entirely. This might suggest that there are other mechanisms causing the humic acid inhibition effect in this system. Future studies should be done to verify this observation.

Collagen strongly inhibited STR typing as allele dropout was observed with increasing inhibitor concentration. RC-PCR SNP genotyping however showed no inhibition with the concentrations of collagen tested (Supplemental Table 10). Collagen is a protein and a large component of the organic material in bone and a major portion of connective tissue [20,21]. It is a mixed mode inhibitor with its main effect preventing denaturation of double stranded DNA [20]. Additionally, collagen is reported to bind to DNA inhibiting polymerase binding activity for subsequent amplification. Perhaps the small size of the RC-PCR template (only 50 bps in length) provides an advantage that may allow amplification as the short size of targets may not all have collagen bound to the template.

Hematin showed no inhibitory effect in STR typing, but complete allele dropout was observed for the RC-PCR system except for the highest concentration in which only 14 % of the SNP (RC-PCR) alleles were detected (Supplemental Table 10). Hematin is a metal-chelating (binding) molecule found in red blood cells with its specialized structure (i.e., porphyrin) binding magnesium ions that are necessary for PCR amplification [20,21]. Perhaps different polymerases and magnesium ion concentrations may have considerable effects on overcoming hematin inhibition in both systems.

3.7. Mock forensic casework study

Mock forensic samples were prepared to test the HID RC-PCR panel. Supplemental Table 11 describes the sample types evaluated and the number of alleles observed. Because these samples were prepared to simulate casework samples and some are characterized as touch samples or may contain mixtures, observed alleles were reported. The percentage of total possible alleles for each sample is described for high quality samples with known contributors. Homozygous loci were counted as a single allele and heterozygotes as two alleles. For the prepared stains complete and concordant profiles with those of the known donor were observed. All but one of the touch samples yielded a high number of SNP alleles.

3.8. Challenged samples

Four problematic (i.e., low signal) population samples, identified when previously sequenced using the ForenSeq™ DNA Signature Prep Kit, were genotyped and analyzed with the HID RC-PCR panel. Typing success was assessed by comparing the 27 identity SNPs amplified by both systems as well as the corresponding RMP values. The RC-PCR system was able to yield complete SNP profiles, while the ForenSeq™ DNA Signature Prep Kit was only able to identify between 59.6 and 96.8 % of possible SNP loci for these samples (Supplemental Table 12). The difference in RMP values is not a good indicator of the better method for analyzing challenged samples. The ForenSeq™ DNA Signature Prep Kit has more SNPs than the HID RC-PCR panel and overall would be expected to yield a lower RMP. However, the use of the RC-PCR panel could be quite productive as can be seen for sample IBN017 in which only 59.6 % of the identity SNPs yielded results with the ForenSeq™ DNA Signature Prep Kit compared with 100% success with the RC-PCR panel. Clearly, the RMP would increase with more SNPs which could be added except that some amplicons would have to be greater than 50 bps in length. The SNPs in the HID RC-PCR panel were selected because their chemistry was compatible with a 50 bp (small) amplicon length. The percent marker success rate is a better indicator of the value of the RC-PCR method; the success rate was 100 % for the RC-PCR panel. It can be anticipated simply based on amplicon length that the RC-PCR panel will more likely yield results from highly degraded samples than other current routine DNA profiling methods.

The six challenged forensic bone samples were genotyped with the RC-PCR system. The sample donors are unknown; therefore, locus detection was assessed. Partial consensus STR profiles were achieved when genotyped by UNTCHI by traditional methods ranging from 27.3 % to 64.3 %. For the SNPs (RC-PCR), locus success ranged from 66.6 % to 100 %. Supplemental Fig. 7 illustrates the condition of one of the forensic bone samples, specifically CHI-0112. The condition suggests a high probability of DNA damage and/or degradation. Table 1 lists the total number of loci observed, STR and SNP (RC-PCR), for the six challenged bone samples that were genotyped. All the SNP loci were observed for three of the six challenged casework samples, and for the other three bone samples, the percent marker typing success was higher than that with the STR markers.

Table 1
The total number of STR (CE based methods) and SNP (RC-PCR system) loci observed for six challenged casework bone samples. The percentage included in parentheses is the total number of observed loci out of the possible loci for multiple STR systems attempted and SNPs included in the HID RC-PCR panel.

	Loci Observed	
Sample	CE STR Loci	RC-PCR Identity SNP Loci
CHI-0107	9/14[a] (64.3%)	27/27 (100%)
CHI-0108	9/14[a] (64.3%)	27/27 (100%)
CHI-0109	6/22[b] (27.3%)	23/27 (85.2%)
CHI-0110	5/13[c] (38.5%)	18/27 (66.6%)
CHI-0111	8/15[d] (53.3%)	18/27 (66.6%)
CHI-0112	7/15[d] (46.7%)	27/27 (100%)

[a] AmpFLSTR® Profiler Plus® PCR Amplification Kit, AmpFLSTR® COfiler® PCR Amplification Kit, and AmpFLSTR® MiniFiler PCR Amplification Kit used for STR typing.

[b] PowerPlex® Fusion 6C System and AmpFLSTR® MiniFiler PCR Amplification Kit used for STR typing.

[c] AmpFLSTR® Profiler Plus® PCR Amplification Kit and AmpFLSTR® MiniFiler PCR Amplification Kit used for STR typing.

[d] AmpFLSTR® Profiler Plus® PCR Amplification Kit, AmpFLSTR® MiniFiler PCR Amplification Kit, and AmpFLSTR® Identifiler Plus PCR Amplification Kit used for STR typing.

RMPs were computed for the four major US populations (i.e., African American, Caucasian, Southeast Hispanic, and Southwest Hispanic) using the allele frequencies reported in the FBI 2015 expanded database [31]. Supplemental Table 13 shows the RMPs for each of the casework bone samples based on their respective STR profiles. Following the same interpretation guidelines, RMP was calculated for the HID SNP loci observed by the RC-PCR system for the bone samples (Supplemental Table 14). The SNPs provided higher RMPs compared to the STR RMPs due to their bi-allelic nature.

4. Conclusion

A one-step system for the simultaneous targeting and amplification of short DNA templates was evaluated for human identification purposes. The RC-PCR technology appears to successfully amplify with high fidelity identity informative SNP makers on DNA target fragments of only 50 bases in length. The method demonstrates to have a substantial sensitivity of detection with a majority of alleles detected with as little as 60 pg of input DNA (approximately the quantity of nuclear DNA contained within 10 cells) and robustness in the presence of known co-extracted inhibitors (except for hematin and humic acid). Additionally, RC-PCR appears to be quantitative enabling deconvolution of contributors to mixtures. The system's success in typing artificial (mock samples) or challenged casework bone samples indicates that it may be an effective adjunct to current methods for analysis of damaged and degraded DNA. For samples that contain no DNA or limited quality DNA such as highly degraded bone and hair samples, the higher typing success of the RC-PCR method may yield results where no or very limited data are obtained with standard DNA typing methods; thus, this small amplicon method, RC-PCR, may add value for identification typing purposes for a wider range of challenging cases. Future SNP panel designs could include additional identity SNPs to increase the discrimination power as well as Y-chromosomal SNPs for sex and lineage determination. Only 27 identity SNPs were included in this preliminary panel due to chemistry and amplicon size constraints of this study; however, by increasing the size of the allowable amplicon, more identity SNPs could be added to the HID RC-PCR panel which would increase the overall power of discrimination for more robust samples and yet maintain the single tube assay format.

The unique features of the RC-PCR technology lend itself to a seamless application to successfully target SNPs on templates confined to 50 bases in length. The RC-PCR protocol in its entirety can be performed in a single day. The one-step, closed tube process is conducive to automation and decreases the labor required to perform library preparation. Use of RC-PCR may substantially reduce the probability of contamination due to less sample manipulation.

Declaration of Competing Interest

The authors affiliated with NimaGen B.V. disclose financial competing interests as they are employees of NimaGen B.V.

The authors affiliated with the University of North Texas Center for Human Identification (UNTCHI) disclose no competing interests.

Acknowledgments

The authors would like to thank Dixie Peters from the University of North Texas Center for Human Identification (UNTCHI) for her guidance and support in the identification and procurement of the challenged forensic casework samples for evaluation of the RC-PCR technology. This work was funded by Award No. 2016-DN-BX-0154, awarded by the National Institute of Justice (NIJ), Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions expressed within this publication are those of the author(s) and do not necessarily reflect those of the U.S. Department of Justice.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.fsigen.2019.102201.

References

[1] E.M. Gorden, K. Sturk-Andreaggi, C. Marshall, Repair of DNA damage caused by cytosine deamination in mitochondrial DNA of forensic case samples, Forensic Sci. Int. Genet. 34 (2018) 257–264.
[2] O. Loreille, et al., Biological sexing of a 4,000-Year-Old egyptian mummy head to assess the potential of nuclear DNA recovery from the most damaged and limited forensic specimens, Genes 9 (3) (2018) 135.
[3] M.D. Brandhagen, R.S. Just, J.A. Irwin, Validation of NGS for mitochondrial DNA casework at the FBI laboratory, Forensic Sci. Int. Genet. 6 (2019) 102151.
[4] R. Alaeddini, S.J. Walsh, A. Abbas, Forensic implications of genetic analyses from degraded DNA—a review, Forensic Sci. Int. Genet. 4 (3) (2010) 148–157.
[5] A. Ambers, M. Turnbough, R. Benjamin, J. King, B. Budowle, Assessment of the role of DNA repair in damaged forensic samples, Int. J. Legal Med. 128 (6) (2014) 913–921.
[6] D. Gaudio, et al., Genome-Wide DNA from Degraded petrous Bones and the Assessment of sex and probable Geographic origins of Forensic Cases, Sci. Rep. 9 (1) (2019) 8226.
[7] P.B. Damgaard, et al., Improving access to endogenous DNA in ancient bones and teeth, Sci. Rep. 5 (2015) 11184.
[8] M.D. Brandhagen, O. Loreille, J.A. Irwin, Fragmented nuclear DNA is the predominant genetic material in human hair shafts, Genes 9 (12) (2018) 640.
[9] A.C. Syvänen, From gels to chips:"minisequencing" primer extension for analysis of point mutations and single nucleotide polymorphisms, Hum. Mutat. 13 (1) (1999) 1–10.
[10] NimaGen B.V. EasySeq™ NGS Reverse Complement PCR Brochure, (2019).
[11] NimaGen, EasySeq NGS Reverse Complement PCR, (2019) (Accessed 31 July 2019), https://www.nimagen.com/products/EasySeq-RC-PCR/EasySeq-RC-PCR/RC-PCR/.
[12] K.B. Gettings, K.M. Kiesler, P.M. Vallone, Performance of a next generation sequencing SNP assay on degraded DNA, Forensic Sci. Int. Genet. 19 (2015) 1–9.
[13] NimaGen B.V., The EasySeq™ NGS Reverse Complement PCR (RC-PCR) Quick Reference Guide, (2019).
[14] A.J. Pakstis, W.C. Speed, J.R. Kidd, K.K. Kidd, Candidate SNPs for a universal individual identification panel, Hum. Genet. 121 (2007) 305–317.
[15] I. Grandell, R. Samara, A.O. Tillmar, A SNP panel for identity and kinship testing using massive parallel sequencing, Int. J. Legal Med. 130 (4) (2016) 905–914.
[16] National Center for Biotechnology Information (NCBI) dbSNP Database. dbSNP Database, (2019) https://www.ncbi.nlm.nih.gov/snp/.
[17] Verogen, ForenSeq™ DNA Signature Prep Reference Guide. (2015).
[18] Qiagen, QIAamp® DNA Mini and Blood Mini Handbook, (2016).
[19] Thermo Fisher Scientific, Qubit® dsDNA HS Assay Kit, (2015).
[20] R.E. Thompson, G. Duncan, B.R. McCord, An investigation of PCR inhibition using Plexor®-Based quantitative PCR and short tandem repeat amplification, J. Forensic Sci. 59 (6) (2014) 1517–1529.
[21] K.L. Opel, D. Chung, B.R. McCord, A study of PCR inhibition mechanisms using real time PCR, J. Forensic Sci. 55 (1) (2010) 25–33.

[22] K. Elwick, X. Zeng, J. King, B. Budowle, S. Hughes-Stamm, Comparative tolerance of two massively parallel sequencing systems to common PCR inhibitors, Int. J. Legal Med. 132 (4) (2018) 983–995.
[23] Thermo Fisher Scientific, GlobalFiler™ PCR Amplification Kit User Guide, (2016).
[24] Qiagen, EZ1® DNA Investigator® Handbook, (2014).
[25] Thermo Fisher Scientific, Precision ID SNP Panels with the HID Ion S5™/HID Ion GeneStudio™ S5 System Application Guide, (2019).
[26] Qiagen, QIAamp® DNA Investigator Handbook, (2012).
[27] Thermo Fisher Scientific, Quantifiler HP and Trio DNA Quantification Kit User Guide, (2017).
[28] A. Ambers, et al., Autosomal and Y-STR analysis of degraded DNA from the 120-year-old skeletal remains of Ezekiel Harper, Forensic Sci. Int. Genet. 9 (2014) 33–41.
[29] Thermo Fisher Scientific, Quantifiler® Duo DNA Quantification Kit User's Manual, (2008).
[30] Scientific Working Group on DNA Analysis Methods (SWGDAM), SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories, (2017).
[31] T.R. Moretti, et al., Population data on the expanded CODIS core STR loci for eleven populations of significance for forensic DNA analyses in the United States, Forensic Sci. Int. Genet. 25 (2016) 175–181.
[32] Agilent Technologies, Agilent High Sensitivity D1000 ScreenTape System Quick Guide, (2015).
[33] Illumina, MiSeq System: Denature and Dilute Libraries Guide, (2018).
[34] A.E. Woerner, J.L. King, B. Budowle, Fast STR allele identification with STRait Razor 3.0, Forensic Sci. Int. Genet. 30 (2017) 18–23.
[35] Genetic Data Analysis Software, (1996) http://en.bio-soft.net/dna/gda.html.
[36] Illumina, MiSeq FGx™ Instrument Reference Guide, (2015).
[37] J.L. King, 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 (2018) 60–76.
[38] C. Ralec, E. Henry, M. Lemor, T. Killelea, G. Henneke, Calcium-driven DNA synthesis by a high-fidelity DNA polymerase, Nucleic Acids Res. 45 (21) (2017) 12425–12440.