The author(s) shown below used Federal funding provided by the U.S. Department of Justice to prepare the following resource:

Document Title:	**Typing Highly Degraded DNA Using Circularized Molecules and Target Enrichment**
Author(s):	**Bruce Budowle**
Document Number:	255963
Date Received:	**January 2020**
Award Number:	**2016-DN-BX-0154**

| **Federal Agency:** | Office of Justice Programs, National Institute of Justice, Department of Justice |

Federal Agency: Office of Justice Programs, National Institute of Justice, Department of Justice

Federal Grant: Award Number 2016-DN-BX-0154

Project Title: Typing Highly Degraded DNA Using Circularized Molecules and Target Enrichment

PI: Bruce Budowle, University of North Texas Health Science Center at Fort Worth, Center for Human Identification, Fort Worth, TX 76107
Tel: 817-735-2979; Email: bruce.budowle@unthsc.edu

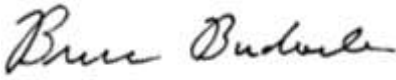**Submission Date:** June 04, 2019

DUNS Number:

EIN:

Recipient Organization: University of North Texas Health Science Center
3500 Camp Bowie Blvd., Fort Worth, TX 76107

Institutional Profile Number: 6108502

Project/Grant Period: 03-01-2017 – 06-30-2019

Reporting Period: 03-01-2017 – 06-30-2019

Report: Final Summary Report

Signature: *Bruce Budowle*
Bruce Budowle, Ph.D.

1

Summary

Greater demands have been placed on forensic laboratories to provide genetic results from a wide range of samples. These samples tend to come from high volume crime and touch evidence in addition to the remains of missing persons and victims of mass disasters. These samples often are limited in the quality and quantity of DNA and thus may not be sufficient for obtaining typable results. Molecular biology continues to improve the capability of obtaining DNA profiling results from severely compromised samples. Unfortunately, current methods fail to overcome the limitations of highly damaged or degraded DNA. Most short tandem repeat (STR) markers fail to amplify by PCR very degraded samples. Single nucleotide polymorphisms (SNPs) and insertions/deletions (InDels) should be targeted as their amplicon size can be substantially smaller than that of STRs. SNP typing, while desirable, was not readily amenable to capillary electrophoresis, the mainstay of genetic analyses. Since the allelic states of SNPs are similar in mass, laborious and often non-quantitative approaches had to be considered. Thus, SNPs residing in small size amplicons were not implemented into casework to allow for obtaining more genetic data from degraded samples.

However, with the advent of massively parallel sequencing (MPS), SNP typing is now feasible and indeed relatively straightforward. To exploit the added value of MPS and SNPs, methods need to be developed to prepare degraded samples for sequencing. The smaller the size of an amplicon, the better suited it is for analyzing degraded DNA. There are several technologies that have come available that may enable development of sample processing methods that can analyze highly degraded DNA. These methodologies include circularization of short single stranded DNA molecules, molecular capture and enrichment (by whole genome amplification) of targeted markers, molecular inversion probes (MIPs) that enable multiplex amplification under a single, universal PCR, and reverse complement PCR (RC-PCR). This project tested and evaluated these various molecular tools to determine if any were suitable for analysis of degraded DNA.

Most MPS systems require enrichment of targets of interest. One approach for enrichment is whole genome amplification (WGA), a promising technique for increasing viable template in trace or low-level DNA samples. Multiple Displacement Amplification (MDA) is one of the more robust and accurate WGA methodologies (1-7). However, MDA performs best on fragments of DNA greater than 2000 bases in length and DNA quantities of 10 ng or greater. Neither of these criteria are suitable for forensic analyses. In order to analyze highly degraded DNA, an alternative MDA method with accompanying molecular strategies is needed to type very short fragments of DNA at lower quantities.

It may be possible to create the effect of a long linear molecule with short DNA fragments if the molecules can be circularized. A circle in effect is an infinite linear molecule as it has not end point. MDA could in theory continue primer elongation and generate multiple tandem copies of the circular template. The enzyme, CircLigase™ II ssDNA Ligase (Lucigen®; Middelton, WI) (8), catalyzes the intra-strand ligation of the 5'-monophosphate group and the 3'-hydroxyl group of a single stranded DNA molecule to form a circular template. By use of the highly processive polymerase (Φ29) (3), MDA is converted to a rolling circle amplification (RCA) process.

The SNP targets do not reside in all of the DNA in a sample. Indeed, most of the DNA is not of interest, and the SNP targets being at very low concentration in any DNA sample may in essence be lost amid the greater DNA background. Thus, to be efficient a capture assay is needed

to select for those fragments that contain SNPs of interest. To enrich for molecules before or after circularization, two capture methods, described by Templeton et al. (9) and Carpenter et al (10), were assessed based on the capture of targets and percent enrichment.

Padlock probes or MIPs provide an alternative circular molecule approach to typing degraded DNA (11-14). A MIP is a single-stranded oligonucleotide containing target-complementary regions that flank a SNP of interest. The two target-complementary arms of the MIP hybridize to the regions flanking a specific SNP position. The short gap (ideally only one base) including the SNP will be filled in and the probe arms ligated. The result is a circular molecule containing the SNP state complement. With proper design MIPs could recognize SNPs in fragments in theory at a minimum of approximately 50 bases in length. Once circularized, probes can be released (denatured) from their targets. The MIP is linearized and amplified by PCR with the MIPs possessing already integrated primer binding sites. This MIP approach would allow for the use of the same primers for PCR thus amplifying the target markers more efficiently than a multiplex with different primers for each SNP. MIPs are highly specific since the probe ends must be in immediate proximity to one another or the probe will not ligate and circularize. This constraint promotes efficient ligation and reduces mismatches.

Another approach that could enrich degraded DNA is a novel one step PCR target enrichment technology (Reverse Complement PCR (RC-PCR)) originally developed for medical sample tracking and authentication in conjunction with the Salisbury NHS Foundation Trust (Salisbury; United Kingdom). RC-PCR allows for the combined generation and tagging of a desired sequence construct in a single, closed tube amplification and indexing reaction (15). Moreover, the primers for this reaction can be designed such that fragments as small as 50 bases could be amplified; thus, degraded samples may be amplified and typed with a reduction in sample manipulation. RC-PCR has additional advantages over the techniques above as it requires less sample manipulation and as a closed single tube assay, chances for contamination are reduced. However, the assay has never been tested on limited quality and quantity samples such as those encountered in forensic biological evidence.

The primary goal of this research project was to develop a method that can analyze degraded DNA. RC-PCR turned out to be the most successful of the three strategies regarding efficiency, sensitivity of detection, and robustness. Thus, this technique appears to be a viable option for forensic applications and combined within a single process makes it highly desirable for casework workflow.

Circularization of short fragments was sought, but was not successful. The testing involved selection of candidate SNPs for human identification, design of synthetic oligonucleotides containing the selected SNP targets, and development of methods to visualize and quantify circular products. While these aspects of the project were met, circularization using CircLigase was inefficient. Relatively large quantities of DNA were required and only a few molecules were effectively circularized. Furthermore, the RCA method also required substantial template and was highly inefficient. These limitations encountered with circularization and RCA could not be overcome, and it was deemed that this approach with current knowledge and technology would not be viable for typing degraded DNA.

A human identification (HID) MIP panel was developed based on 24 HID SNPs. The design of the MIPs included a specific criterion that the insert region (the region to be filled by the polymerase including the target-complementary MIP arms had annealed to the DNA template) not exceed 50 basepairs (bps) in length. Results of tests of the MIPs with high quality data were unsuccessful. The vast majority of MIPs self-ligated and thus did not show any target SNPs. In a

3

very few MIPs was there any evidence of SNP incorporation. There is no explanation for self-ligation of MIP arms as the ligase should work predominately on double stranded DNA. No solution was obtained to resolve the MIP problem. One fundamental issue for the HID MIP panel may be in the design of the hybridization primers for the target DNA templates. Perhaps the target-specific hybridization arms require additional distance around the SNP of interest for optimal capture performance, and if so then the MIP assay would not be suitable for the intended purpose of analyzing degraded DNA.

In contrast to the methods above, the RC-PCR methodology was a viable approach to analyze degraded DNA. While the design of this technology was for large quantity samples, the design suggested that it could be applicable to analysis of degraded and low quantity DNA. This one step PCR technology employs reverse complement probes. Reverse Complement PCR (RC-PCR) allows for amplification and tagging of a targeted sequence construct in a single, closed tube (15). Moreover, the primers for this reaction were designed such that fragments as small as 50 bps could be generated; thus, degraded samples may be amplified and typed with a reduction in sample manipulation. For this application, 27 human identity SNPs were found suitable for capture targets with the fragment length restriction of approximately 50 bps.

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 region of interest, but rather are the reverse complement and are blocked at the 3′ end to halt extension. In a single reaction, a complete oligonucleotide (sequencing construct) is generated that includes everything necessary for MPS library preparation.

The recommended DNA input for this methodology was 80 (nanograms) ng (well above an amount desired for forensic applications) (16). Therefore, studies included a sensitivity study to determine the lower limits of detection for the system, a population study (N=50) to test robustness of allele calling, inhibitor tolerance; and challenging samples both from actual forensic casework or mock samples.

A sensitivity study of the HID RC-PCR system was performed. Samples from two contributors were tested with DNA inputs of 1 ng, 250 (picogram) pg, and 60 pg. The RC-PCR system was able to detect greater than 83% of SNP alleles with a DNA input of 60 pg. Based on these results, an input of 1 ng was identified as a sufficient target quantity to use for the population/concordance study. This optimum DNA input is comparable to other forensic MPS and capillary electrophoresis systems. A population/concordance study with 50 Caucasian samples was performed and all allele calls were concordant to those produced by the ForenSeq™ DNA Signature Prep kit (Illumina; San Diego, CA).

Although the RC-PCR assay was not optimized to be resistant to the inhibitors that may be encountered in forensic biological evidence, an inhibitor study was conducted. The inhibitors tested were calcium, humic acid, collagen, and hematin. Inhibitor concentrations were selected based on a small inhibition study using the GlobalFiler PCR Amplification Kit (Thermo Fisher Scientific; Waltham, MA). The concentrations of inhibitors where inhibition was observed with the GlobalFiler kit were used for evaluating performance of the RC-PCR method. Hematin was the only inhibitor that had a negative impact on PCR yield. The other inhibitors were not effective until the highest concentrations were introduced into the samples.

Lastly, samples that were highly degraded (problematic population samples and old casework samples, 11 total) were analyzed and compared with ForenSeq MPS and capillary

4

electrophoresis STR analyses, respectively. A higher percentage of alleles were obtained for all samples using the RC-PCR technique. The success is likely due to the small amplicon size of approximately 50 bps which favors analysis of degraded DNA and a more efficient amplification than larger amplicons by other approaches.

In conclusion, the RC-PCR methodology was the only approach in which successful enrichment of short amplicons containing SNPs was achieved. The methodology has several features that make it appealing for human identification purposes: 1) applicable to degraded samples; 2) high sensitivity of detection; 3) single tube assay that reduces hands on time, workflow and chance of contamination; and 4) appears to be relatively robust. Given these results, the UNT Center for Human Identification intends to continue to validate RC-PCR for missing persons identification purposes.

Account of the Activities

The primary goals of this research were to develop a methodology that could analyze highly degraded DNA. Three approaches were considered. The methodologies included 1) circularization of short single stranded DNA molecules, molecular capture and enrichment (by whole genome amplification) of targeted markers, 2) molecular inversion probes (MIPs) that could target 50 nucleotide fragments and enable multiplex amplification under a single, universal PCR, and 3) reverse complement PCR (RC-PCR) that also could target 50 nucleotide long fragments and multiplex markers. The target markers for all three were SNPs as they allow for smaller size amplicons to be generated which would facilitate analysis of degraded DNA. The detection platform/system was MPS to enable SNP states to be determined. This project tested and evaluated these various molecular tools to determine if any were suitable for analysis of degraded DNA.

Accomplishments

The project was successful in that a methodology was developed and tested that is capable of analyzing degraded DNA.

RC-PCR methodology was the only one of the three approaches in which successful enrichment of short amplicons containing SNPs was achieved. The methodology was shown to be able analyze fragments as little 50 nucleotides in length, had sensitivity of detection to the limits of input DNA tested of 60 pg, provided concordant typing with data from an orthogonal system, was fairly resistant to the effects of known inhibitors, and was able to obtain better results (i.e. percentage of typable alleles) with actual and mock casework. Since the assay is designed as a single, closed system process (i.e., amplification and library preparation), the workflow is streamlined and chance for contamination is reduced.

Although the circularization of DNA using CircLigase and use of MIPs was not successful, the research provides data on approaches that are not viable with current capabilities. Both systems were highly inefficient (i.e., requiring large quantity of input DNA) and suffered from various issues.

References

Zhang L, Cui X, Schmitt K, Hubert R, Navidit W, Arnheim N. 1992. Whole genome amplification from a single cell: Implications for genetic analysis. PNAS 89:5847-5851.

2. Hughes S, Lasken R. 2005. Whole Genome Amplification. Oxfordshire, England: Scion Publishing Ltd.

3. Nelson JR, Cai YC, Giesler TL, Farchaus JW, Sundaram ST, Ortiz-Rivera M, Hosta LP, Hewitt PL. 2002. TempliPhi, phi29 DNA polymerase based rolling circle amplification of templates for DNA sequencing. BioTechniques Supplement:44-47.

4. Dean FB, Hosono S, Fang L, Wu X, Faruqi AF, Bray-Ward P, Sun Z, Zong Q, Du Y, Du J, Driscoll M, Song W, Kingsmore SF, Egholm M, Lasken RS. 2002. Comprehensive whole genome amplification using multiple displacement amplification. PNAS USA 99:5261-5266.

5. Giardina E, Pietrangeli I, Martone C, Zampatti S, Marsala P, Gabriele L, Ricci O, Solla G, Asili P, Arcudi G, Spinella A, Novelli G. 2009. Whole genome amplification and real-time PCR in forensic casework, BMC Genomics 10:159.

6. Maragh S, Jakupciak JP, Wagner PD, Rom WN, Sidransky D, Srivastava S, O'Connell CD. 2008. Multiple strand displacement amplification of mitochondrial DNA from clinical samples. BMC Med. Genet. 9:7.

7. Nara A, Harihara S, Iwadate K, Uemura K. 2009. Sequence analysis for HV I region of mitochondrial DNA using WGA (Whole Genome Amplification) method, Leg. Med. 11:S115–S118.

8. Tate CM, Nunez AN, Goldstein CA, Gomes I, Robertson JM, Kavlick MF, Budowle B. 2012. Evaluation of circular DNA substrates for whole genome amplification prior to forensic analysis. Forensic Sci. Int. Genetics 6:185-190.

9. Templeton JEL, Brotherton PM, Llamas B, Soubrier J, Haak W, Cooper A, Austin JJ. 2013. DNA capture and next-generation sequencing can recover whole mitochondrial genomes from highly degraded samples for human identification. Investig Genet 4:26.

10. Carpenter ML, Buenrostro JD, Valdiosera C, Schroeder H, Allentoft ME, Sikora M, Rasmussen M, Gravel S, Guillén S, Nekhrizov G, Leshtakov K, Dimitrova D, Theodossiev N, Pettener D, Luiselli D, Sandoval K, Moreno-Estrada A, Li Y, Wang J, Gilbert MT, Willerslev E, Greenleaf WJ, Bustamante CD. 2013. Pulling out the 1%: whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. Am J Hum Genet 93(5):852-864.

11. Hardenbol P, Baner J, Jain M, Nilsson M, Namsaraev EA, Karlin-Neumann GA, Frakhrai-Rad H, Ronaghi M, Willis TD, Landegren U, Davis RW. 2003. Multiplexed genotyping with

sequence-tagged molecular inversion probes, Nat. Biotechnol. 21:673-678.

12. Hardenbol P, Yu F, Belmont J, Mackenzie J, Bruckner C, Brundage T, Boudreau A, Chow S, Eberle J, Erbilgin A, Falkowski M, Fitzgerald R, Ghose S, Iartchouk O, Jain M, Karlin-Neumann G, Lu X, Miao X, Moore B, Moorhead M, Namsaraev E, Pasternak S, Prakash E, Tran K, Wang Z, Jones HB, Davis RW, Willis TD, Gibbs RA. 2005. Highly multiplexed molecular inversion probe genotyping: over 10,000 targeted SNPs genotyped in a single tube assay. Genome Res. 15(2):269-275.

13. Turner EH, Lee C, Ng SB, Nickerson DA, Shendure J. 2009. Massively parallel exon capture and library-free resequencing across 16 genomes. Nat Methods 6(5):315-316.

14. Li JB, Gao Y, Aach J, Zhang K, Kryukov GV, Xie B, Ahlford A, Yoon JK, Rosenbaum AM, Zaranek AW, LeProust E, Sunyaev SR, Church GM. 2009. Multiplex padlock targeted sequencing reveals human hypermutable CpG variations. Genome Res.19(9):1606-1615.

15. NimaGen B.V. EasySeq™ NGS Reverse Complement PCR Brochure. 2019. https://www.nimagen.com/media/207096/NimaGen-Easyseq-NGS_RCPCR.pdf

16. NimaGen B.V. EasySeq™ NGS Reverse Complement-PCR Kit: Quick Reference Guide. 2019. https://www.nimagen.com/media/207096/NimaGen-Easyseq-NGS_RCPCR.pdf

Products Produced

In addition to the research results obtained, presentations and draft manuscripts have been produced that document the work.

Presentations at National and International Meetings that were supported by this work

Wiley, R., Bus, M., King, J., and Budowle, B.: Typing highly degraded DNA using circularized molecules and target enrichment, American Academy of Forensic Sciences (AAFS) 71st Annual Scientific Meeting. Baltimore, Maryland, 2019.

Publications

Kieser, R., Bus, M., King, J., Van der Liet, W., Thelen, J., Budowle, B., Reverse Complement-PCR: A novel one-step system for the human identification of highly degraded DNA. Forensic Science International: Genetics. https://doi.org/10.1016/J.FSIGEN.2019.102201

Kieser, R. and Budowle, B. Rolling Circle Amplification: A (random) primer on the enrichment of an infinite linear DNA template. WIREs Forensic Science. https://doi.org/10.1002/wfs2.1359

Invention Report
No inventions. No patents were submitted related to this project

Participating Scientists and Collaborators

What individuals have worked on the project?

Name: Bruce Budowle
Project role: PI
Nearest person month worked: 2 years
Contribution to Project: Dr. Budowle provided overall direction and management of the project as well as participating in the technical research and data analysis. He will ensure that results from this project are prepared for reports and publication.

Name: Rachel Kieser
Nearest person month worked: 2 years
Contribution to Project: Rachel Kieser is a graduate student performing the main effort of the work on development of the CircLigase, MIP, and RC-PCR methods. She performed DNA extraction, DNA quantitation, standard DNA typing, library generation, sequencing, data analysis, and manuscript preparation.

Name: Jonathan King
Nearest person month worked: 2 years
Contribution to Project: Mr. King is the laboratory manager responsible for the overall management of the wet-lab portion of the work.

Name: Magdalena Bus
Nearest person month worked: 0.5 years
Contribution to Project: Dr. Bus, a post doctoral fellow, assisted on RC-PCR design and data analysis.

What other organizations have been involved as partners?
none

Have other collaborators or contacts been involved?
Name: Joop Thelen and Walter van der Vielt
Company: NimaGen B.V. (Nijmegen, Netherlands)

Impact
What is the impact of the project on the criminal justice system?

Nothing to report as of yet. The current impact of this project on the criminal justice system cannot be assessed, as it takes time and resources to translate a technology from research to operation. The potential impact of the RC-PCR results, however, is likely to be substantial. The technology shows a good sensitivity of detection, can be applied to degraded samples, and works in the presence of inhibitors. The advancement leverages the throughput power of MPS. The results from this study will likely expand the number of samples (and hence cases) where forensic biological evidence can provide valuable investigative leads and thus assist in solving more cases and providing identifications with human remains.

How has it contributed to crime laboratories?

Currently this project has not been transferred to crime laboratories. However, UNTCHI has found the initial data from the RC-PCR studies to be sufficiently promising to pursue further with an intent to implement in the near future.

What is the impact on technology transfer?

Nothing to report.

Changes/Problems

The circularization strategies of the project were not successful. Use of CircLigase to circularize DNA has been successful (to a degree) but the process is highly inefficient, requiring substantial amount of input DNA. The template amounts are far greater than encountered in casework. Multiple attempts to overcome this limitation have failed. Testing of the HID MIP panel also was unsuccessful. However, during the course of the project a novel one step PCR technology (RC-PCR) was developed. This novel method targeting 27 human identity SNPs can amplify targets of 50 bps in length with as little as 60 pg of template DNA. Research and validation will continue with this methodology.

9

Proprietary Information

There was no proprietary information related to this work.