

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

Document Title:	Molecular Tools for Low-Template DNA Analysis
Author(s):	John R. Nelson
Document Number:	251935
Date Received:	August 2018
Award Number:	2012-DN-BX-K047

This resource has not been published by the U.S. Department of Justice. This resource is being made publically available through the Office of Justice Programs' National Criminal Justice Reference Service.

Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

RESEARCH PERFORMANCE PROGRESS REPORT

Federal Agency and Organization Element to Which Report is submitted **Department of Justice** Office of Justice Programs National Institute of Justice Federal Grant or Other Identifying Number Assigned by Agency 2012-DN-BX-K047 Project Title Molecular Tools for Low-Template DNA Analysis PD/PI Name, Title and Contact Information John R. Nelson, Principal Scientist, John.Nelson@ge.com, (518) 387-7272 Name of Submitting Official, Title, and Contact if other than PD/PI N/A □Submission Date March 4, 2016 Recipient Organization (Name and Address) General Electric Company **One Research Circle** Niskayuna, NY 12309-1027 Recipient Identifying Number or Account Number, if any Project/Grant Period (Start Date, End Date) 1/1/2013, 3/31/2016 Reporting Period End Date 3/31/2016 Report Term or Frequency (annual, semi-annual, quarterly, other) **Final Summary Overview** Signature of Submitting Official

Final Summary Overview.

Purpose,

Current low-template forensic workflows lack the robustness of standard DNA typing, generating artifactual results from PCR-generated stochastic effects when the input DNA copy number is low. In addition, aside from reducing the amplicon size, there is no consideration for working with DNA that contains template lesions or is fragmented. A successful project will provide new methods for the pre-amplification of trace DNA, environmentally damaged trace DNA, and single cells, avoiding many of the stochastic effects resulting from PCR amplification of limiting DNA. The purpose of this project is to adapt two separate, highly sensitive methods for the treatment of low-template samples that would pre-amplify DNA from a few copies to an amount sufficient for quantification and robust profiling in addition to archiving.

Project design and methods,

The overall goal of this proposal is to adapt two in vitro isothermal amplification methods for use on human forensic Low Copy Number (LCN) samples and evaluate their ability to overcome some of the challenges facing the LCN typing of STR loci including allele and locus dropout and drop-in, allele stutter effects, and allelic imbalance. One method is a Phi29-based MDA method of whole-genome amplification that encompasses several improvements to our currently commercially available WGA kit, GenomiPhi V2. These improvements are directed toward making the kit more suitable for trace-DNA amplification, but have not been tested on forensic samples. The other method, called Ping Pong, is a sequence-specific isothermal amplification method that will be used to target the pre-amplification of desired STR loci. In addition, these pre-amplification methods will be combined with reactions that repair chemically modified and circularize fragmented DNA. Using controlled quantities of undamaged and artificially damaged human DNA, and isolated human cells 1) the proposed enzymatic reactions will be adapted and optimized with the goal of establishing a reliable method for pre-amplification of trace DNA. 2) The optimized pre-amplification methods will then be integrated and validated with current forensic workflows.

whole genome amplification - an improved isothermal whole-genome DNA amplification method that has been previously demonstrated to allow for the highly sensitive amplification of bacterial DNA will be adapted for the pre-amplification of trace human DNA to enable the improved recovery and quality of STR profiles. Our plan is to optimize the reaction methods for four types of samples: 1) Quantitated dilutions of de-identified intact donor genomic DNA; 2) Limiting dilutions of intentionally UVC damaged genomic DNA; 3) Limiting dilutions of intentionally fragmented genomic DNA; and 4) captured intact cells from an established cell line.

Sequence-specific DNA amplification- adapt and evaluate a novel locus-specific isothermal reaction also currently under development at GE Global Research. Instead of amplifying the entire human genome, this method would be used to pre-amplify only the desired STR loci from trace DNA samples, providing a substantially increased probability of extracting a more complete profile. This reaction, called "Ping-Pong," involves incubation with nucleotide analog-modified primers, the genetically engineered nicking enzyme

Endonuclease V, and a strand-displacing polymerase. Each primer uses a specifically placed inosine residue (a "G" analog) located 2 nt from the 3' end. After primer extension by polymerase, the endonuclease V nicks the strand 2 nt downstream from the inosine, effectively regenerating the primer for additional repeated extensions. Unlike traditional PCR, hybridization of each primer to the template DNA allows for the synthesis of multiple extension products through a repeated series of nicking and extension reactions, creating many copies of the downstream DNA. The displaced products can then serve as template for additional analog-modified reverse primer hybridization and extensions, further amplifying the desired DNA sequence with geometric kinetics. An advantage of this isothermal method is that the amplification takes place at a lower temperature than with PCR, and so the degree of STR stutter peak formation is expected to decrease because of less DNA strand slippage. In addition, the linear amplification kinetics is potentially advantageous for the maintenance of allelic ratios. The specific regions of the genome that will be amplified will include the 13 core CODIS STR loci in addition to amelogenin and the HTERT gene for quantitation purposes using the Quantifiler[™] kit. Our current objective will be to modify the reaction formulation, temperature, and primer composition in order to optimize the reaction for a multiplexed system. We will also test the single-tube incorporation of the previously developed repair method. Buffer composition, salt, and volumes will be adjusted to optimize reaction compatibility.

Data analysis,

Analysis methods for WGA reactions- Real time amplification kinetics is used by inclusion of SYBR green in reactions and performing reactions in plate reader. Samples quantified with PICO green staining in a plate reader. Gel analysis is used to size and quantify amplification products. Amplified samples sent to Strand Laboratories for STR analysis where peak size and intensity is used to judge amplification.

<u>Findings,</u>

Evaluation of whole-genome amplification formulations

We have been evaluating new formulations of the random primed phi29 DNA polymerase-based amplification reactions to determine if robust amplification of low-template DNA samples can be achieved and to identify the most appropriate formulation for forensics applications. The formulations under consideration include: AT GPhi, which contains random hexamers with locked nucleic acid modifications; sc GPhi, which was previously optimized for single mammalian cell DNA amplification; dN6-hiGC, which can potentially improve amplification of GC-rich regions; and dN6-low dNTP, which has a reduced dNTP concentration that reduces the overall extent of amplification and potential for amplification reactions were performed at 30°C until complete (nucleotides depleted). Either 5 ng of amplified DNA or the indicated quantity of unamplified DNA was analyzed using the Identifiler® Plus kit following the standard 28-cycle PCR protocol (Figs. 1 and 2). Using normal injection conditions, products were separated on the ABI 3130xI Genetic Analyzer and results analyzed. The error bars indicate standard deviations of three replicate reactions.



Figure 1. DNA yield and STR allele recovery after WGA



Figure 2. STR peak heights and STR allele balance after WGA

Our results showed that all of the WGA formulations allowed the detection of many more STR alleles compared with an equivalent initial amount of unamplified DNA from 100 pg input DNA and lower. The improvement was especially dramatic using a DNA input level of 10 pg, in which almost no alleles were identified using unamplified DNA, compared with 50-80% of correct allele identification using WGA reaction products. It should be noted that 10 pg DNA is approaching the DNA quantity in a single cell (6.6 pg), and so it would not be expected that all alleles would be recovered from a random sampling of 10 pg DNA. Although there was a drop in average peak height ratio from 100 pg input DNA and lower, indicating some degree of peak height imbalance, all WGA reactions provided highly improved overall peak heights and reasonably representative amplification, with the AT formulation showing the largest improvement at the lowest input DNA levels.

Emulsion-based whole-genome amplification

At very low template DNA quantities it is possible for amplification bias to create regions of the genome with variable representation levels that can cause STR peak height ratios to be altered. To address this issue, we have developed a formulation that is compatible with Phi29 polymerase-based amplification inside a 1 nl droplet emulsion. By dividing the amplification reaction into many small compartments, the extent by which any given sequence can become over-represented is limited, thus producing an overall more balanced reaction. Starting with 50 pg of chromosomal DNA, amplification reactions were divided into 20,000 droplets using the QX200[™] droplet generator (Bio-Rad). Following the reaction, the emulsion was broken using a chloroform extraction, and the DNA was purified using SureClean[™] reagent (Bioline). DNA was analyzed using qPCR with a series of primers targeting several CODIS STR loci. Comparing bulk reactions with emulsion-based reactions,





we found that overall DNA yields were similar, with both methods producing micrograms of amplified DNA (Fig. 3). However, whereas bulk reactions always produce DNA in the absence of added template because of primer-primer interactions, we found that performing reactions in emulsions nearly completely suppressed this template-independent amplification. In addition, the emulsion-based reactions had significantly lower levels of amplification bias as measured by qPCR of 11 different CODIS STR loci, producing DNA with representation levels more similar to unamplified control DNA.

Evaluation of reduced volume whole-genome amplification workflows

While standard 20 µl bulk WGA reactions generally produce in the range of 3-6 µg of amplified DNA product, this is far more than is necessary for analysis using standard commercial STR PCR amplification kits that require only approximately 1 ng of DNA. Excess whole-genome amplification can potentially introduce artifacts into the DNA including overand under-representation of sequences from amplification bias. To limit the extent of amplification, we evaluated two methods: a reduced, 500 nl reaction assembled using the Labcyte Echo® device, a tipless liquid handler that uses acoustic energy to accurately dispense small volumes; and an emulsion-based reaction in which standard 20 µl reactions are split into 20,000 individual droplets prior to amplification. By dividing the amplification reaction into many small compartments, the extent by which any given sequence can become over-represented is limited, theoretically producing an overall more balanced reaction.

Using a dilution series of purified 2800M control DNA (Promega), we performed standard AT GenomiPhi WGA reactions (20 µl), low volume reactions (500 nl), and emulsion-based reactions, in which standard 20 µl reactions were divided into 1 nl droplets using the QX200[™] droplet generator (Bio-Rad). Following the reaction, the emulsion was broken and DNA was purified using ethanol precipitation. DNA was analyzed using the Identifiler® Plus kit using the standard 28-cycle PCR, and normal injection conditions on the ABI 3730 DNA Analyzer (Figs. 4 and 5). Error bars indicate standard deviations of three replicate reactions. Our results show that compared with unamplified DNA, we were able to detect many more correct alleles at low input quantities using all of the WGA workflows, albeit with some reduction in average peak-height ratios. Unfortunately, the low volume reactions showed only a slight improvement in allele detection and no detectible improvement to the allelic balance. Two of the challenges faced when performing enzymatic reactions at sub-microliter volumes are the increased surface area contact of the reaction plate with the droplet, and difficulties in preventing evaporation that can alter reaction conditions. It is possible that addressing these issues would provide a benefit for the smallvolume reactions. For some of the emulsion-based amplification reactions, we observed a high frequency of drop-in and off-ladder peaks and variable allele detection that we believe result from the additional manipulations required for this workflow. Though some further optimization would be required to ensure that small reaction volume workflows retain sufficient low-template DNA and function robustly, we continue to observe that the AT GenomiPhi formulation provides highly increased allele detection and overall peak heights with reasonably representative amplification from low input DNA levels.



Figure 4. DNA yield and STR allele identification after WGA



Evaluation of single cell whole-genome amplification workflows

In addition to purified DNA, another form of low-template sample is intact multiple and individual human cells. In cases of sexual assault in which the suspect's cells are mixed with the victim's epithelial cells or in other cases where multiple donors have contributed cells to a surface, it would be essential to separate the individual cell or cell types from the mixture in order to be able to accurately resolve individual profiles. Cases in which only a single cell is available are at the boundary of detection. In order to provide sufficient DNA for robust STR profiling, we evaluated amplification of 7 single cell replicates using three different single cell whole-genome amplification workflows (Fig. 6):

- Using a washed culture of K562 cells, we utilized the Fluidigm C1[™] Auto Prep system for whole-genome amplification of single cells. This system utilizes integrated microfluidic circuits to isolate cells, perform lysis, neutralization, and whole-genome amplification using proprietary reagents in a final reaction chamber volume of 300 nl. Our typical DNA yields were approximately 130 ng, more than sufficient for STR analysis.
- We next utilized our standard, commercially available Illustra Single Cell GenomiPhi kit (GE Healthcare) for DNA amplification. K562 cells were isolated by transferring 5 nl droplets of a cell dilution to a reaction plate using the Labcyte Echo® device. Wells containing single cells were identified and DNA was amplified in 20 µl volumes. Typical yields were approximately 6 µg.
- Third, we utilized our AT GenomiPhi formulation for DNA amplification. The workflow is very similar to the Single Cell GenomiPhi kit, except this version contains optimized random hexamers with locked nucleic acid modifications.



Unfortunately, we found that despite producing the expected quantities of amplified DNA using the C1[™] Auto Prep system, no STR data was produced upon analysis with the Identifiler® Plus kit for any of the samples. It's possible that either the device amplified contaminating (non-human) DNA or that it represents background primer-dimer derived material. In contrast, both GE single-cell workflows provided robust STR data with an average of ~70% allele identification per profile (Fig. 7). Though we observed that a similar set of alleles tended to have low peak heights or drop out between the replicates, by creating an average consensus profile from three replicates we were able to assemble a complete profile from DNA amplified using the AT kit. We continue to observe a modest reduction in average peak height ratios with amplified DNA compared with purified, unamplified DNA, but it may be possible to improve this by reducing the extent of whole-genome amplification by reducing the reaction time. Alternatively, because there is only one copy of the genome present, any nick or lesion near an STR locus would cause reduced representation in that area, implementing our enzymatic repair reaction may further improve overall allele detection and balance.



Standard deviations in parenthesis

Inclusion of DNA damage repair with whole genome amplification-

Q-SUN irradiated DNA (simulating sunlight damage to DNA) was prepared. Damage to the DNA was confirmed by qPCR and whole genome amplification kinetics (see figure 8).



Figure 8

NEB's PreCR vs GE repair method vs NEBNext FFPE DNA repair were tested on 100 pg of sun-damaged DNA, and repaired sample amplified by AT GenomiPhi. Results indicate 10 ul PreCR repair is compatible with a workflow where entire sample is added to a 20 ul AT GenomiPhi reaction, and a marked increase in reaction kinetics is observed (see figure 9). The GRC repair formulation and the NEBNext FFPE repair had no effect. We do not yet understand why the GE repair method, which has worked well on UV- treated samples in the past, did not function well, but as this is made from individually purchased enzymes, there could be an issue with one of these components. Nevertheless, we have determined that the NEB PreCR kit can be utilized prior to whole genome amplification, and the repair reaction added in entirety to the whole genome amplification reaction (no loss of sample due to processing).



Figure 9

The enzyme Deoxyribodipyrimidine photo-lyase (phrB)was also tested. PhrB catalyzes the light-dependent monomerization (300-600 nm) of cyclobutyl pyrimidine dimers (in cis-syn configuration), which are formed between closest bases on the same DNA strand upon exposure to ultraviolet radiation. There was no effect on whole genome amplification kinetics (see figure 10).



Figure 10

Evaluation of iSDA for STR pre-amplification-

In other projects we have been comparing Ping Pong with iSDA, both isothermal loci-specific DNA amplification reactions which utilize nicking enzymes and strand displacing DNA polymerase. Because of superior kinetics demonstrated in these other projects, we performed an evaluation of iSDA pre-amplification as an alternative to Ping Pong preamplification. In order to amplify target STR loci with linear reaction kinetics, we utilized a modified form of iSDA, an isothermal DNA amplification method that has been used successfully for the exponential amplification of target bacterial and viral sequences. To be successful for this application however, the method needed to be modified in two ways: to adjust the number of primers so that amplification occurs with linear kinetics in order to best preserve allele balance; and to create multiplexed reactions such that all relevant STR loci are pre-amplified simultaneously. Unfortunately, despite significant work in this area, we had only very limited successful amplification product when an excess of very short (79mer) input template was utilized. All attempts to move to more complex human genomic DNA template resulted in exclusive amplification of off-target products, even for the previously successful exponential amplification system. We conclude that the iSDA reaction method is unlikely to be robust enough for target pre-amplification of multiple STR loci from human genomic DNA.

Ping Pong targeted pre-amplification of STR loci-

Using the Identifiler® Plus kit for analysis, we initially established two important parameters about the workflow: First, that carryover chemical components from the Ping Pong reaction mix had no detectible impact on the resulting STR profiles, meaning that the amplified reactions could be used directly without purification; and second, that carryover modified primers did cause a dropout in signal at the relevant locus, meaning that exonuclease digestion of the primers would be essential.

Targeted Pre-amplification

In order to eliminate the pre-amplification primers from the mixture, we have tried 2 methods; degradation with a single stranded exonuclease, and alternatively, replacement of internal thymine in the primers with uracil. While there were issues with the exonuclease method, we have determined that the uracil replacement is compatible with the Ping Pong enzyme. i.e. the endonuclease required for isothermal targeted pre-amplification does not degrade primers containing uracil (see figure 11).



Figure 11

Also, the addition of uracil DNA glycosylase to the uracil-containing primers completely degrades them rapidly (see figure 12).



Figure 12

The final PP pre-amp protocol may be broken down as follows:

- 1. The template was denatured by heat in the presence of the Two or Four Primer mixes (5 pmol/µl each primer) and then placed on the lab bench to cool,
- 2. An enzyme mix was prepared containing *Bst* DNA polymerase and mutant Endo V diluted in buffer, and
- 3. When the denatured template had cooled the enzymes were added and mixed. The complete PP pre-amp reaction was placed in a pre-heated thermal cycler at 50°C for one hour and then the enzymes were killed by heating at 80°C for 20 minutes.
- 4. Primers were removed by incubation with either UDG (Two Primer) or UDG and Exo I (Four Primer).
- 5. Reactions were precipitated with sodium iodide and resuspended in an equal volume of water.

Denaturation	
Component	Volume (µl)
10X CHP-AAS Buffer	1.0
SSB (5 mg/ml)	0.2
2 or 4 Primer Oligo Mix (5 pmol/µl each)	1.0
Template	1.0
Water	6.8
Total Volume	9.0

Enzyme Mix

Component	Volume		
	(µI)		
Bst DNA Polymerase (120 U/µl)	0.953		
Mutant Endo V (0.623 mg/ml)	1.03		
Enzyme Dilution Buffer*	23.0		
Total Volume	~25		

Complete PP Pre-Amp Reaction

component	volume		
	(µI)		
Denaturation	9.0		
Enzyme Mix	1.0		
Final Volume	10		

*Enzyme Dilution Buffer was 10 mM HEPES, pH 8, 1 mM DTT, 0.5 mM EDTA, 0.01 % Tween 20 and 50 % (v/v) Glycerol

Reactions with either 2 inosine-containing amplification primers (2-primer Ping Pong) or 2 inosine-containing primers with 2 nested regular primers designed to convert the product into a double stranded form (4-primer Ping Pong), were evaluated.



STR pre-amp reactions were designed to have 1 ng of hgDNA in a final reaction volume of 10 µl. Completed 2-Primer pre-amp reactions each had 5 units of UDG added. Completed 4-Primer pre-amp reactions each had 5 units of UDG and 20 units of exonuclease I added. Both sets of reactions were incubated for one hour at 37°C, then 20 minutes at 80°C and finally 10 minutes at 95°C. Each reaction was precipitated using sodium iodide/ethanol (Nal used in place of ammonium acetate or sodium acetate) and suspended in an equal volume of water. Completed reactions were stored at -20°C. One microliter and 10 µl aliquots of each completed pre-amp reaction were analyzed by Strand Analytical Labs (Indianapolis, IN) for their STR profile using PowerPlex® 16 System from Promega.

	Fold Sigr STR loc average	Fold Signal Intensity Increase of STR loci (0.1 ng input DNA) vs average signal from all other STR peaks in trace			
Locus	2 Primer	4 Primer			
D3S1358	38.0x	_			
TH01	1.7x	-			
D21S11	3.4x	* (82/56)			
D18S51	Ох	-			
D5S818	12.6x	* (193)			
D13S317	1.9x	-			
D7S820	7.4x	4.5x (567/450)			

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

D16S539	2.6x	* (262/233)
CSF1PO	2.0x	30.4x (927/6408)
∨WA	1.4x	0
D8S1179	1.3x	* (104)
ΤΡΟΧ	8.9x	_
FGA	2.1x	_

*The only signals observed were from the allele which was pre-amplified

The value(s) listed in parenthesis is(are) the actual signal intensity(ies) of the peak(s).

- Indicates complete absence of signal in the STR PCR, no peaks at all.

These results indicate that the 2-primer scheme is providing pre-amplification of the desired loci, albeit with future work required to vary the amount of amplification of each so that they provide even amplification fold. Upon examination of the 4-primer scheme results, we found that in all reactions the STR peaks that were NOT associated with the desired loci had reduced signal strength, about 5-7 fold overall reduction in signal. This was not anticipated, as control reactions performed on mock samples prior to this test had shown no reduction in signal strength. This made it impossible to provide a "fold" amplification in these reactions as in many cases only the amplified loci were observable. In 6 of the cases, significant signal amplification was observed. Since time was running out on the grant we decided to attempt 4-plex loci pre-amplifications to demonstrate the multiplex potential of this method.

Multiplex Pre-Amplification

Different 4-loci primer mixes were prepared as outlined in Table 1. All 4-loci primer mixes were prepared as 4-Primer Mixes (4 loci x 4 primers each loci), with each oligo included at a concentration of 5 pmol/ μ l. Testing of each 4-loci 4-Primer Mix pre-amplification was essentially unsuccessful, with only three loci (D5, D7 and TPOX) demonstrating an increase in signal intensity.

ID	1	2	3	4	5	6	7	8	9	10
D3S1358										
D5S818										
D7S820										
D8S1179										
D13S317										
D16S539										
D18S51										
D21S11										
CSF1PO										
FGA										
TH01										
ТРОХ										
Кеу										
	Denotes primers included in the multiplex oligo mix									

 Table 1 Multiplex Primer Mixes.
 Four Primer oligo mixes were prepared for each loci as indicated.

These three loci, D5S818, D7S820 and TPOX, were then combined together in a single primer mix and tested for their pre-amplification ability on 0.1 ng of DNA. The results of this multiplex pre-amplification may be found in Table 2.

Locus	Signal Strength
D5S818	215
D7S820	54/60
TPOX	273
All Others	Not observed

 Table 2. 3-loci pre-amplification. These three loci were the only signals obtained during analysis.

While the Ping Pong method of pre-amplification continues to show promise, more work is required to develop any forensic workflow utilizing this method for pre-amplification of all STR loci in trace samples.

Implications for criminal justice policy and practice in the United States.

While more work is required by qualified forensic laboratories to validate whole genome amplification from trace samples, we believe this work has demonstrated that this method of sample pre-amplification could be used to increase robustness and peak heights obtained from trace or low copy number samples. Also, that if 3 or more intact single cells are present, complete STR profiles can be obtained using our protocol using the AT GenomiPhi method.

Outcome/Deliverables

• The team has optimized protocols for pre-amplification of human DNA using wholegenome amplification. Results from this work are being transitioned to the Human Identification group at GE Healthcare for consideration.

• While there are patents and patent applications already existing covering methods discussed within this report, a new patent application on the use of Ping Pong sequence-specific amplification technologies for targeted pre-amplification of samples is being prepared based on this work.