NCJRS
National Criminal Justice Reference Service

Office of Justice Programs
BJA BJS NIJ OJJDP OVC SMART

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

Document Title:	**Enhanced Sample Preparation and Data Interpretation Strategies for Massively Parallel Sequencing for Human Identification in Missing Persons and DVI Casework**
Author(s):	**Sheree Hughes-Stamm, Ph.D., Bruce Budowle, Ph.D.**
Document Number:	**252576**
Date Received:	**January 2019**
Award Number:	**2015-DN-BX-K066**

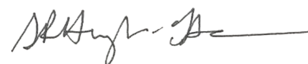
This resource has not been published by the U.S. Department of Justice. This resource is being made publically available through the Office of Justice Programs' National Criminal Justice Reference Service.

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.

Enhanced Sample Preparation and Data Interpretation Strategies for Massively Parallel Sequencing for Human Identification in Missing Persons and DVI Casework

Federal Agency	National Institute of Justice
Award Number	2015-DN-BX-K066
PI Name, Title and Contact Information	Dr. Sheree Hughes-Stamm Asst. Professor of Forensic Science Sam Houston State University sxh039@shsu.edu Tel: 936 294 4359 Dr. Bruce Budowle. (Co.PI and sub-awardee) Prof. and Executive Director, Center for Human Identification, University of North Texas Health Science Center, Fort Worth, Texas Bruce.Budowle@unthsc.edu
Submission Date	11/29/2018
DUNS and EIN Numbers	
Recipient Organization	Sam Houston State University 1003 Bowers Blvd Huntsville, TX 77341
Project/Grant Period	January 1, 2017 – August 31, 2018
Reporting Period End Date	08/31/2018
Report Type:	Final Summary Overview

Signature of Submitting Official



____
Sheree Hughes-Stamm, Ph.D.

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.

PURPOSE

The overall purpose of this project was to investigate the compatibility of common sample preparation methods for highly challenging samples such as those recovered from missing persons' cases with two massively parallel sequencing (MPS) chemistries designed for forensic identification. We also assessed the tolerance of these two chemistries and platforms to high levels of PCR inhibitors commonly associated with skeletal and decomposed human remains.

This project had four specific goals:

1. Assess the baseline tolerance of two forensic MPS systems (HID-Ion AmpliSeq™ Identity panel on the Ion Torrent™ PGM (Life Technologies), and the ForenSeq™ Panel on the MiSeq® (Verogen)) to PCR inhibitors commonly encountered in forensic and missing persons' cases involving human remains.

2. Compare common sample preparation and DNA extraction methods to determine their relative efficiencies to remove common PCR inhibitors prior to MPS and their compatibility with the two MPS chemistries designed for forensic use.

3. Test the performance of both MPS systems when used to sequence a variety of challenging skeletal samples.

4. Developing a data interpretation strategy for the various MPS assays.

STUDY 1 - Assessing the baseline inhibitor tolerance of two forensic MPS systems to PCR inhibitors.

Sample Preparation

Control DNA (N = 3) at 1 ng and 0.1 ng input was spiked with five increasing concentrations of five common PCR inhibitors associated with skeletal remains (Table 1). These samples were first genotyped using two different capillary electrophoresis (CE) based STR kits (GlobalFiler and Investigator 24plex QS) to determine the relative performance of both systems with highly inhibited samples prior to sequencing. The original inhibitor concentrations assessed in CE were then evaluated via massively parallel sequencing

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.

(MPS) using two different platforms and chemistries (Ion Torrent and Illumina systems) and then altered accordingly to seek the limits of tolerance (Table 1).

Inhibitor Preparation

Five inhibitors (humic acid, melanin, hematin, collagen, and calcium) were tested in this study. A range of inhibitor concentrations was used to test the tolerance of CE and MPS to high amounts of PCR inhibition (Table 1). All inhibitor stocks were prepared according to guidelines established in Opel et al (1). All inhibitors were prepared in 10 mL volumes and dilutions were made with deionized water.

Table 1. Concentrations of PCR inhibitors common to human remains spiked into control DNA for CE and MPS analysis.

Melanin	ng/µL	25,35,40,45,50	4,5,7,10,12
Collagen	ng/µL	50,100,112.5,130,160	180,250,300,350,400

CE-STR Sample Preparation

DNA (NIST standard 2372 Component A Male and two male donors) were used for the sensitivity and inhibitor studies. For the sensitivity study, all three DNA sources were amplified in triplicate using template input amounts of 1 ng - 0.0078 ng.

STR Amplification and Capillary Electrophoresis

STR typing was performed using the GlobalFiler® PCR Amplification kit (Thermo Fisher Scientific) and the Investigator® 24plex QS kit (QIAGEN) as per manufacturer's instructions. Bone and tissue samples were amplified with 0.8 ng of DNA (or maximum sample volume (15 µL) if DNA was less than 0.053 ng/µL). Inhibited samples were amplified with 10 µL of each inhibitor at the required concentration (Table 1) and 5 µL of DNA (0.2 ng/µL or 0.02 ng/µL). Inhibitor controls were performed using sterilized deionized water in lieu of the inhibitor. PCR products were separated and detected via capillary electrophoresis according to the respective manufacturer protocols using a 3500™ Genetic Analyzer and GeneMapper® ID-X v 1.4 software.

Ion PGM Sequencing

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.

All sequencing reactions were performed with 1 ng of input DNA with various concentrations of humic acid, melanin, hematin, collagen, or calcium (Table 1). Each sample was amplified in triplicate using the Ion AmpliSeq™ Library Kit 2.0 and ID panel according to the manufacturer's specifications. Pooled libraries were batched according to concentration (10 or 25 pM), added to the Ion Chef™ and loaded onto 316 barcoded semiconductor chips and sequenced on the Ion PGM™ System. Data analysis was conducted using Torrent Suite v4.6, the HID_SNP_Genotyper plugin v4.3.1 and an in-house workbook created at UNTHSC. STR and SNP allele typing success was calculated as the percentage of concordant alleles reported. An arbitrary detection threshold was set at 2X.

MiSeq FGx Sequencing

The same DNA samples were amplified in triplicate using ForenSeq™ DNA Signature Prep Kit (using Primer Mix A) according to the manufacturer's specifications and sequenced on the Illumina FGx™ system (10 µL pooled libraries were used). Three sequencing runs were performed. Data analysis was conducted using STRaitRazor v2s and R software.

Results & Discussion

<u>CE-STR Kit Comparison</u>

Conclusions

In this study, we examined the comparative sensitivity and performance of two commercial STR kits (GlobalFiler® PCR Amplification kit and Investigator® 24plex QS kit) with low template and highly inhibited (humic acid, melanin, hematin, collagen, and calcium) samples. The results of this research suggest that the GlobalFiler® kit may be slightly more sensitive than the Investigator® 24plex QS kit, producing more complete and balanced STR profiles. However, the Investigator® kit was more tolerant than the GlobalFiler® kit to all of the PCR inhibitors tested in this study when both 1 ng and 0.1 ng of DNA was amplified (results published in Elwick et al (2)).

<u>Massively Parallel Sequencing</u>

Noise

The influence of five inhibitors on noise generated during SNP sequencing with the AmpliSeq™ Library Kit and ID panel and SNP and STR sequencing (ForenSeq™ DNA Signature Prep Kit) was

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.

investigated. Overall noise (i.e., PCR/sequence error) did not appear to increase with exposure to inhibitors. However, percent noise did increase with increasing concentrations of inhibitors as a result of a decrease in read depth of the true allele sequence.

Conclusions

The two multiplexes with different chemistries were exposed to DNA samples containing a number of inhibitors over a range of concentrations. As expected, increasing concentrations of inhibitors had an inverse effect on locus read depth and typing success, with a few exceptions. The most noted outcome was that the two kits were not always susceptible to the effects of inhibitors in a similar fashion. For example, the Ion AmpliSeq panel was more susceptible to the presence of hematin and calcium with little or no effect observed for the ForenSeq panel. In contrast, the Forenseq panel was more susceptible to melanin and collagen compared with the Ion AmpliSeq kit. Overall, large amplicon STR loci were less resistant to inhibitors compared with small STR and especially SNP loci. In some of the higher inhibitor concentrations, there were differences within a series in which read depth and/or typing success (although slight) were not always consistent. The overall outcome was that inhibitors, when they do have a negative effect on typing performance, can reduce typing success but do not contribute to sequencing error.

For complete results of Study 1 please see:

- **CE-based STR results:** Kyleen Elwick, Carrie Mayes, Sheree Hughes-Stamm. Comparative Sensitivity and Inhibitor Tolerance of GlobalFiler PCR Amplification and Investigator 24plex QS Kits for Challenging Samples. Legal Medicine. 2018 Feb 17;32:31-36. doi: 10.1016/j.legalmed.2018.01.006.
- **MPS results:** Kyleen Elwick, Xiangpei Zeng, Jonathan King, Bruce Budowle, Sheree Hughes-Stamm. Comparative Tolerance of Two Massively Parallel Sequencing Systems to Common PCR Inhibitors. International Journal of Legal Medicine. 2017. DOI 10.1007/s00414-017-1693-4.

STUDY 2 - Assessing sample preparation compatibility with two MPS chemistries

Sample and Inhibitor Preparation

Blood and hair samples were obtained from the same live donor in accordance with Sam Houston State University Institutional Review Board Guidelines #2015-12-26123. A high concentration of each PCR inhibitor was added to the appropriate biological sample prior to DNA extraction: hematin - 27.5 mM, melanin - 750 ng, calcium - 22.5 mM, and humic acid - 3750 ng. Three spiked replicates and one control

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.

(no inhibitor) of each of the blood and hair were subjected to four different extraction methods: an organic extraction method, DNA IQ™, PrepFiler® BTA, and QIAamp® DNA Investigator all following recommended protocols. In addition to the commercial kits, bone samples also underwent two previously reported total demineralization protocols, TD1 (4) and TD2 (5).

Ion S5 Library Preparation and Sequencing

Extracted DNA (1 ng) from blood, hair, and bone samples was amplified using the Precision ID DL8 Kit and a custom STR and iiSNP panel on the Ion Chef™ System. This panel includes 32 STR markers, 1 Y-indel, 2 amelogenin sex markers, 41 iiSNPs, and 34 Y-SNPs. Four sequencing runs were performed using the Ion S5™ System and the Ion S5™ Precision ID Chef and Sequencing Kit. Data analysis was performed using Converge™ 2.0 and in-house workbooks. Reportable alleles, mean read depth, and average heterozygote balance were calculated for each system.

MiSeq FGx Library Preparation and Sequencing

Extracted DNA (1 ng) from blood, hair, and bone samples was amplified using the ForenSeq™ DNA Signature Prep Kit (using Primer Mix A) according to the manufacturer's specifications. Sequencing was performed using the MiSeq FGx system (10 µL pooled libraries were analyzed). Three sequencing runs were performed. Data analysis was conducted using STRait Razor v2s.

Conclusions

The overall results of this study demonstrate that all extraction methods tested were effective in removing high amounts of inhibitors from blood, hair, and bone tissues. These extraction methods all produced clean extracts that were equally compatible with both the Precision ID chemistry on the Ion S5™ System the ForenSeq™ DNA Signature Prep Kit on the MiSeq FGx™ system. All extraction methods produced quantifiable DNA with little or no PCR inhibition detected, demonstrating that all extraction methods were efficient and are suitable for preparing samples for MPS. Furthermore, the MiSeq platform appears to be slightly more tolerant to PCR inhibitors. However, due to the small sample size, no definitive statement can be made. Both platforms produced near complete profiles throughout this study, and neither platform was more tolerant to a specific inhibitor than the other. Forensic laboratories can be confident that

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.

when DNA is extracted from skeletal and hair samples using one of the common methods tested in this study, extracts are equally compatible with both MPS workflows.

For complete results of Study 2 please see: Kyleen Elwick, Xiangpei Zeng, Carrie Mayes, Jonathan L. King, Bruce Budowle, Sheree Hughes-Stamm. Assessment of impact of extraction methods on analysis of human remain samples on massively parallel sequencing success. 2018. International Journal of Legal Medicine. doi.org/10.1007/s00414-018-1955-9

STUDY 3 - Evaluation of two MPS systems and chemistries with extremely challenging skeletal samples

Sample Preparation

Bone and teeth samples (N=24) were harvested from 14 bodies willed to the Applied Anatomical Research Center (AARC) at Sam Houston State University in Huntsville, Texas. Bodies were either cremated embalmed, highly decomposed, or burned. Three replicates of each bone or tooth sample were extracted (300 mg powder) using a total demineralization protocol (4).

CE-based STR Analysis

PCR amplification of STRs was performed using the GlobalFiler™ PCR Amplification Kit in accordance with the manufacturer's protocol. DNA target input was 0.8 ng when possible. Reportable alleles, average peak height (APH), average peak height ratios (APHR), and allelic dropout were calculated.

Ion S5 Library Preparation and Sequencing

All low template samples (i.e., <0.16 ng/µL) were amplified and prepared using the DL8 kit and DNA samples greater than or equal to 0.16 ng were prepared manually using the Precision ID Library Kit. A custom AmpliSeq™ STR and iiSNP primer panel was used to amplify the extracted DNA. All samples (N = 81) were sequenced in four runs. Templating and chip loading were performed using the Ion Chef™ System on a 530™ semiconductor chip, and sequencing was performed using the Ion S5™ Precision ID Chef and Sequencing Kit with the Ion S5™ System. Data analyses were performed using Converge™ 2.0 and in-house excel workbooks. For STRs and SNPs, mean read depth, heterozygote balance, and reportable alleles were assessed. A minimum arbitrary detection threshold of 5X was used for both systems.

MiSeq FGx Library Preparation and Sequencing

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.

Libraries were prepared using the ForenSeq™ DNA Signature Prep Kit with Primer Mix A following the manufacturer's protocol. Sequencing was performed on a MiSeq FGx™ instrument using the MiSeq FGx™ Reagent Kit and the manufacturer's protocol. Data analyses were performed using the ForenSeq Universal Analysis Software, STRait Razor v2s, and in-house excel workbooks. The same metrics were calculated on data generated on both the Ion S5 and the MiSeq.

Results & Discussion

Capillary Electrophoresis (CE)

Full female profiles produced 44 alleles and full male profiles produced 46 alleles. All bone and teeth samples amplified with the GlobalFiler™ PCR Amplification Kit produced a STR profile to varying degrees of profile completeness. Average reportable alleles ranged from 10 ± 3 to complete profiles across the various sample types. The thermally degraded teeth, embalmed, and cremated samples produced complete or near complete profiles. Decomposed skeletal remains produced the most degraded DNA profiles, ranging from 10 ± 3 to 32 ± 5.

APH across all samples ranged from 5154 ± 1952 relative fluorescence units (RFUs) to 210578 ± 8846 RFUs. Overall, the trend observed across the sample types when considering the APH was consistent with STR profile completeness. APHRs showed a similar trend to both profile completeness and APH, with highly decomposed samples being the least successful.

No allelic dropout was observed with the thermally degraded (five samples) and embalmed samples (one sample), the cremated sample (one sample) produced one dropout event at the DYS391 locus. The burned samples (eleven samples) produced 153 instances of allele dropout, and the decomposed samples (six samples) resulted in the highest amount of allele dropout with 396 occurrences.

Massively Parallel Sequencing (MPS)

In this study, STR and SNP typing success was assessed via the number of reportable alleles, read depth, and heterozygote balance. Between the two MPS platforms, all comparable results were concordant.

Reportable Alleles

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.

The number of STR and SNP reportable alleles was calculated in the same manner as CE-based STRs. All Ion S5 samples sequenced produced reportable alleles ranging from one allele to full profiles. However, using the MiSeq, two samples (one decomposed and one thermally degraded) produced no DNA profile. The decomposed sample that produced no profile with the MiSeq™ produced one allele with the Ion S5. In contrast, the thermally degraded sample that produced no profile with the MiSeq™ resulted in 98% of alleles using the Ion S5.

Overall, both sequencing platforms produced quality data for the types of challenged remains analyzed. On the Ion S5 the SNPs demonstrated higher profile completeness than that of STRs, producing ~10% more alleles than STRs overall (~93% ± 29% vs ~84% ± 16%). The severely decomposed remains were especially difficult to analyze. In general, profile completeness between STRs and SNPs was comparable for the MiSeq samples.

CE-based STRs produced a greater percentage of total alleles than the Ion S5 for 2 out of 24 samples and for 8 out of 24 samples when using the MiSeq. For the less compromised remains all methods (CE and MPS) were comparable based on the loci common across the three systems. However, for the more severely degraded samples, the systems demonstrated variable results. Although the common loci of CE-based STRs were comparable to the MPS results, MPS panels in general provided more information because they contain more markers (35 STR markers for the Ion S5 and 58 STRs for the MiSeq vs 24 STRs in GlobalFiler™, and many SNPs). However, two decomposed samples produced almost no results via MPS, and CE-based typing was able to recover >30% of alleles.

The success of typing the 20 core CODIS loci was compared across the three platforms (CE, Ion S5, and MiSeq). CE-generated STRs produced alleles ranging from 8 ± 3 to 40 (full profiles), with all but 3 samples yielding >50% of alleles. Samples sequenced using the Ion S5 system generated profiles ranging from 0 to 40 alleles, with all but three samples producing >50% of alleles. CE-generated data produced more complete than those from the Ion S5 for the CODIS loci in 5/24 samples. Using the MiSeq system, profiles ranged from 0 to 40 alleles, with 16/24 samples producing >50% of alleles, with only two samples failing to produce any alleles (one decomposed and one thermally degraded). However, 11/24 CE-generated

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.

profiles showed more alleles for the CODIS loci. CE results demonstrated 14/24 samples (Ion S5) and 19/24 samples (MiSeq) with equal or greater profile completeness than the MPS systems.

Read Depth

In general, both platforms performed well and produced high sample read depth. Overall, Ion S5 SNPs produced higher read depth than STRs (74050X vs 17344X) for every sample. Both STRs and SNPs demonstrated proportional read depth across all sample types. The MiSeq STRs produced higher mean read depths than SNPs (70568X vs 31184X) for most samples. Only two samples produced higher mean read depth for SNPs than STRs (both decomposed remains). The MiSeq showed a large increase in mean read depth for burned samples for both STRs and SNPs. All other samples produced very low read depth compared to the burned samples, which may be due to run variability and/or a lower ability to type such samples.

Heterozygote Balance

In general, for the Ion S5, the majority of samples showed average heterozygote balances of >70%. Heterozygote balance averaged ~15% higher for SNPs than STRs. SNPs also demonstrated fewer samples with a heterozygote balance <70% compared with STRs (2 vs 8 samples). All samples except decomposed remains resulted in good heterozygote balance for SNPs. Overall, when sequenced on the MiSeq, just under half of the STR and SNP profiles demonstrated heterozygote balances <70%. For both STRs and SNPS, the burned samples demonstrated good heterozygous balance, while decomposed samples consistently demonstrated poor balance (<10%), and variability in all other samples.

Conclusions

Overall, MPS generated reliable sequencing data from environmentally challenged human skeletal samples, and provided more genetic data in 22 of the 24 samples compared with CE-based fragment analysis. Although CE produced a usable DNA profile for identification purposes, the greater number of loci included in the MPS multiplexes allowed for more genetic information to be obtained from most samples barring the highly decomposed remains. Results suggest that MPS may recover more probative information from most samples, but CE-based STR typing is still a robust method for identifying skeletal

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.

samples when DNA quantity and quality is sufficient. CE chemistry has been continually improved over the past 25 years, while MPS kits for forensic applications have been under development and refinement for less than five years. However, improvement in MPS panel design and chemistries will likely further enhance library preparation and sequencing performance.

For complete results of all studies please see: Elwick, Kyleen Elizabeth, *Enhanced sample preparation and data interpretation strategies using massively parallel sequencing for human identification in missing persons' and DVI casework*. Doctor of Philosophy (Forensic Science), December, 2018, Sam Houston State University, Huntsville, Texas. *(in press.)*

CRIMINAL JUSTICE IMPLICATIONS

The value of new MPS-based forensic panels for human identification has yet to be fully explored. However, the successful completion of this research will inform the forensic DNA community of the utility of both systems (AmpliSeq HID panel, Precision ID Degradation panel, and ForenSeq Panel) for the identification of extremely challenged biological samples.

This study will also provide DNA laboratories with important information on the compatibility of common sample preparation and DNA extraction methods with downstream MPS success, and report on the most effective methods for removing inhibitors, whilst also maximizing DNA yield from decomposed tissues, hair and skeletal material. In addition, this study will assist with developing effective data interpretation strategies for the vast amounts of MPS data obtained, in order to streamline analysis and deliver improved results for the identification of missing persons and DVI cases. The evaluation and development of sample preparation methods that result in successful genotyping using various MPS chemistries may provide more genetic information from which to make a reliable association, and therefore will enable closure to more grieving families more quickly.

References

1. Opel KL, Chung D, McCord BR. A study of PCR inhibition mechanisms using real time PCR. Journal of forensic sciences. 2010;55(1):25-33.
2. Elwick K, Mayes C, Hughes-Stamm S. Comparative Sensitivity and Inhibitor Tolerance of GlobalFiler® PCR Amplification and Investigator® 24plex QS Kits for Challenging Samples. Legal Medicine. 2018.
3. Elwick K, Zeng X, King J, Budowle B, Hughes-Stamm S. Comparative tolerance of two massively parallel sequencing systems to common PCR inhibitors. International journal of legal medicine. 2017:1-13.

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.

4. Loreille OM, Parr RL, McGregor KA, Fitzpatrick CM, Lyon C, Yang DY, et al. Integrated DNA and Fingerprint Analyses in the Identification of 60-Year-Old Mummified Human Remains Discovered in an Alaskan Glacier. Journal of forensic sciences. 2010;55(3):813-8.
5. Identification UCfH. Procedure Manual - Forensic Laboratory. Demineralization Extraction of Skeletal Remains. Rev 2 ed. Fort Worth, TX; 2008;1-4.
6. Hu Q, Liu Y, Yi S, Huang D. A comparison of four methods for PCR inhibitor removal. Forensic Science International: Genetics. 2015;16:94-7.
7. Faber KL, Person EC, Hudlow WR. PCR inhibitor removal using the NucleoSpin® DNA Clean-Up XS kit. Forensic Science International: Genetics. 2013;7(1):209-13.

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.