Forensic Science International: Genetics 20 (2016) 20–29

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Research paper

# Evaluation of the Illumina® Beta Version ForenSeq™ DNA Signature Prep Kit for use in genetic profiling

Jennifer D. Churchill[a,*], Sarah E. Schmedes[a], Jonathan L. King[a], Bruce Budowle[a,b]

[a] Institute of Applied Genetics, Department of Molecular and Medical Genetics, University of North Texas Health Science Center, 3500 Camp Bowie Blvd., Fort Worth, TX 76107, USA
[b] Center of Excellence in Genomic Medicine Research (CEGMR), King Abdulaziz University, Jeddah, Saudi Arabia

ARTICLE INFO

*Article history:*
Received 30 June 2015
Received in revised form 1 September 2015
Accepted 17 September 2015
Available online 21 September 2015

*Keywords:*
Massively parallel sequencing (MPS)
ForenSeq
Illumina MiSeq
SNPs
STRs

ABSTRACT

While capillary electrophoresis-based technologies have been the mainstay for human identity typing applications, there are limitations with this methodology's resolution, scalability, and throughput. Massively parallel sequencing (MPS) offers the capability to multiplex multiple types of forensically-relevant markers and multiple samples together in one run all at an overall lower cost per nucleotide than traditional capillary electrophoresis-based methods; thus, addressing some of these limitations. MPS also is poised to expand forensic typing capabilities by providing new strategies for mixture deconvolution with the identification of intra-STR allele sequence variants and the potential to generate new types of investigative leads with an increase in the overall number and types of genetic markers being analyzed. The beta version of the Illumina ForenSeq DNA Signature Prep Kit is a MPS library preparation method with a streamlined workflow that allows for targeted amplification and sequencing of 63 STRs and 95 identity SNPs, with the option to include an additional 56 ancestry SNPs and 22 phenotypic SNPs depending on the primer mix chosen for amplification, on the MiSeq desktop sequencer (Illumina). This study was divided into a series of experiments that evaluated reliability, sensitivity of detection, mixture analysis, concordance, and the ability to analyze challenged samples. Genotype accuracy, depth of coverage, and allele balance were used as informative metrics for the quality of the data produced. The ForenSeq DNA Signature Prep Kit produced reliable, reproducible results and obtained full profiles with DNA input amounts of 1 ng. Data were found to be concordant with current capillary electrophoresis methods, and mixtures at a 1:19 ratio were resolved accurately. Data from the challenged samples showed concordant results with current DNA typing methods with markers in common and minimal allele drop out from the large number of markers typed on these samples. This set of experiments indicates the beta version of the ForenSeq DNA Signature Prep Kit is a valid tool for forensic DNA typing and warrants full validation studies of this MPS technology.



## 1. Introduction

The current capabilities of forensic DNA typing allow forensic analysts to reliably characterize biological evidence to support criminal investigations. DNA typing methods are used to provide identifications or exculpations, build databases that help find associations and link cases, and generate investigative leads. For years, capillary electrophoresis (CE)-based technologies have been the standard method for DNA typing applications. However, with CE, there are limits to the number of markers that can be multiplexed together as similarly sized amplicons must be labeled with different fluorescent markers in order to accurately assign alleles to the appropriate locus. CE methods also tend to focus on only one marker type at a time, due to a limit on the number of markers that can be multiplexed or different analytical streams.

Massively parallel sequencing (MPS), also known as next-generation sequencing, systems are capable of overcoming the limitations mentioned above and expanding forensic analysts' capabilities to characterize forensic biological evidence. MPS technologies have the potential to sequence at a greater capacity and speed with an overall reduced cost per nucleotide. MPS with the ForenSeq DNA Signature Prep Kit (Illumina, San Diego, CA, USA) costs approximately 84 US dollars (USD)/45 USD a sample with 32/96 samples sequenced per run, and the user generates data on over 200 markers. The manufacturer's protocols recommend up to 32

* Corresponding author at: 3500 Camp Bowie Blvd., CBH-250, Fort Worth, TX 76107, USA.
*E-mail address:* Jennifer.Churchill@unthsc.edu (J.D. Churchill).

http://dx.doi.org/10.1016/j.fsigen.2015.09.009
1872-4973/

casework samples or up to 96 database samples per run with the ForenSeq DNA Signature Prep Kit. Meanwhile, for example, CE kits such as GlobalFiler® Express (Thermo Fisher Scientific, Waltham, MA USA) and Yfiler® Plus (Thermo Fisher Scientific) cost 20 USD and 32 USD a sample, respectively, and the user generates data on 24 markers with GlobalFiler® Express and 27 markers with Yfiler® Plus kits. With MPS, a large number of markers and samples can be sequenced simultaneously. Different types of forensically relevant genetic markers can be multiplexed, and sequence variants within shared STR alleles (or allele and stutter product), undetectable by CE, can be identified. Targeted amplicons in the ForenSeq DNA Signature Prep Kit are clonally amplified and sequenced all at once generating a read depth for each target and allowing for a consensus in sequence calls that produces a greater resolution in sequence accuracy.

The ability to sequence a larger number of markers and multiple types of markers at once leads to many benefits for the DNA typing field. While short tandem repeats (STRs) have been the primary marker used for DNA typing applications due to their highly polymorphic and informative nature, new markers offer additional types of information. Single nucleotide polymorphisms (SNPs) comprise approximately 85% of human genetic variation and can provide identity, ancestry, phenotypic, and pharmacogenetic information [1–10]. This information can help produce investigative leads that were not possible previously or in cases that had reached dead ends. As SNPs reflect only a single base change, they can be analyzed with smaller amplicons and, therefore, are considered more attractive markers for degraded samples [4,5,9,10].

Because a large number of markers can be sequenced simultaneously, more robust associations can be generated from database searches and more thorough kinship analyses can be completed, such as facilitating familial searching and missing person identifications. With MPS, the traditional repeat length of STRs can be determined along with elucidation of any sequence variants present within the alleles [11–16]. Maintaining the ability to determine the traditional repeat length of STRs promotes backward compatibility with current DNA databases. Identification of intra-allelic sequence variants provides greater discrimination power and opportunities for better mixture deconvolution [11–16]. For example, the presence of an intra-allelic sequence variant could help an analyst discriminate between a stutter peak and a minor contributor in a mixture or even detect a minor contributor peak masked by the major contributor.

The beta version of the ForenSeq DNA Signature Prep Kit in conjunction with the MiSeq desktop sequencer (Illumina) and the beta build of ForenSeq Universal Analysis Software (UAS; Illumina) were evaluated for high throughput genotyping of reference and challenged samples. This study was divided into five sets of experiments: reliability, sensitivity of detection, mixture analysis, concordance, and the ability to evaluate challenged samples. Additionally, informative metrics, including depth of coverage (DoC), allele coverage ratios (ACRs), and sequence coverage ratios (SCRs), were used to evaluate the quality and reliability of the data produced.

## 2. Materials and methods

### 2.1. Samples

Policies and procedures approved by the Institutional Review Board for the University of North Texas Health Science Center (UNTHSC) in Fort Worth, TX were followed for the collection and use of samples in this study. Samples for each experiment were prepared and extracted as outlined below. All DNA was quantified using the Quantifiler® Human DNA Quantification Kit (Thermo Fisher Scientific) and a 7500 Real Time PCR System (Thermo Fisher Scientific), following the manufacturer's recommended protocol [17].

**Experiment 1.** Reliability—Three replicates of the 2800M control DNA, supplied with the kit and each at one ng of total input DNA, were sequenced.

**Experiment 2.** Sensitivity—Dilution series of the 2800M and 9947A control DNAs were prepared at 1000 pg, 500 pg, 250 pg, 100 pg, and 50 pg of total input DNA and sequenced.

**Experiment 3.** Mixture detection—Seven mixtures were prepared, each consisting of two individuals at a total of one ng of input DNA. One mixture was prepared at a 1:1 male/female ratio. Four mixtures were prepared at a 1:9 ratio as follows: 1:9 male/female; 1:9 female/male; 1:9 male/male; and 1:9 female/female. Two mixtures were prepared at a 1:19 ratio as follows: 1:19 male/female; 1:19 female/male. The reference samples used in the mixtures were obtained from anonymous donors at UNTHSC and were extracted with the QIAamp DNA Blood Mini Kit (Qiagen, Valencia, CA, USA) following the manufacturer's protocols [18].

**Experiment 4.** Concordance—Ten reference samples, each of which were unrelated and included one Central Caribbean, two Asian, and seven Caucasian individuals, were sequenced and typed for autosomal STRs and Y-STRs using both MPS and CE methods. These reference samples were sequenced previously on the Ion Torrent Personal Genome Machine® (Ion PGM™) (Thermo Fisher Scientific) and typed for identity informative SNPs on both MPS platforms. The samples used to evaluate concordance were obtained from anonymous donors at UNTHSC and were extracted with a Qiagen DNA Investigator kit or a QIAamp DNA Blood Mini Kit (Qiagen) [18,19].

**Experiment 5.** Challenged samples—Five aged buccal swabs that were collected over six years prior to DNA extraction and five human bone samples, with previously generated STR data and available from the University of North Texas Center for Human Identification missing person identification laboratory, were sequenced. Traditional STR typing using CE was completed as well for comparison purposes. The aged buccal swabs were extracted with the QIAamp DNA Blood Mini Kit (Qiagen) following the manufacturer's protocols [18]. The human bone samples were extracted according to the protocols described in Marshall et al. [20].

### 2.2. Capillary electrophoresis concordance data

Conventional STR typing was completed using the AmpFlSTR® Identifiler® Plus PCR Amplification Kit (Thermo Fisher Scientific) and the AmpFlSTR® Yfiler® PCR Amplification Kit (Thermo Fisher Scientific) with one ng of DNA for each reaction per the recommended manufacturer's protocols [21,22]. The GeneAmp® PCR System 9700 thermal cycler (Thermo Fisher Scientific) was used for PCR amplification, and electrophoresis was carried out on an ABI Prism® 3130xl Genetic Analyzer (Thermo Fisher Scientific). Raw data were analyzed with GeneMapper® ID software v3.2.1 (Thermo Fisher Scientific).

#### 2.2.1. Ion PGM™ concordance data

SNP typing was completed on the Ion PGM™ using the HID-Ion AmpliSeq™ Identiy Panel (Thermo Fisher Scientific) as described in Churchill et al. [23].