Forensic Science International: Genetics 28 (2017) 1–9

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Research paper

Performance and concordance of the ForenSeq™ system for autosomal and Y chromosome short tandem repeat sequencing of reference-type specimens

Rebecca S. Just[a,b,*], Lilliana I. Moreno[a], Jill B. Smerick[a], Jodi A. Irwin[a]

[a] DNA Support Unit, Federal Bureau of Investigation Laboratory, 2501 Investigation Parkway, Quantico, VA 22135, USA
[b] Counterterrorism and Forensic Science Research Unit, Visiting Scientist Program, Federal Bureau of Investigation Laboratory, 2501 Investigation Parkway, Quantico, VA 22135, USA

ARTICLE INFO

Article history:
Received 23 September 2016
Received in revised form 30 December 2016
Accepted 1 January 2017
Available online 9 January 2017

Keywords:
Short tandem repeat (STR)
Next generation sequencing (NGS)
Massively parallel sequencing (MPS)
ForenSeq
Concordance

ABSTRACT

Though the utility of next-generation sequencing (NGS) technologies for forensic short tandem repeat (STR) typing has been evident for several years, commercially available assays and software solutions developed specifically to meet forensic needs have only recently become available. One of these, the ForenSeq™ DNA Signature Prep Kit (Illumina, Inc.) sequences 27 autosomal STR (aSTR) and 24 Y chromosome STR (Y-STR) loci (concurrent with additional nuclear markers) per multiplexed sample, with automated secondary and tertiary analyses of the data accomplished via the associated ForenSeq™ Universal Analysis Software (UAS). In this study we investigated the performance of the ForenSeq system for aSTR and Y-STR typing by examination of 151 sample libraries developed from high quality DNAs amplified at the target 1 ng template. Utilizing PCR Primer Mix B, greater than 99.5% of aSTR loci and 97.0% of Y-STR loci were recovered when 42 or fewer sample libraries were pooled for sequencing. A direct comparison of UAS developed fragment length results to capillary electrophoresis (CE) based data identified only two allele call discrepancies when no UAS quality flag was triggered. Review of the ForenSeq data indicated that most samples with total sequence read counts exceeding 40,000 could be interpreted to develop nearly complete aSTR genotypes or Y-STR haplotypes. However, markers D22S1045 and DYS392 produced poor or inconsistent results even when sample read counts were greater than 85,000. Excluding these two loci, analyst-interpreted aSTR and Y-STR ForenSeq profiles were 99.96% and 100% concordant, respectively, with CE data. In addition to demonstrating concordance on par with other CE kit to kit comparisons, the results from this study will assist laboratories seeking to develop workflows for high volume processing and analysis of aSTRs and Y-STRs from reference-type specimens using the ForenSeq system.

1. Introduction

Though highly parallelized sequencing technologies were first introduced more than a decade ago [1] and are widely used in most other genetic disciplines, forensic laboratories have been slower to adopt next generation sequencing (NGS) methods in casework practice. While there are many reasons for this, a primary factor has been the lack of developmentally validated and commercially available start-to-finish kits targeting the marker systems (such as short tandem repeats; STRs) routinely typed for forensic purposes. Several papers in recent years have demonstrated the utility and potential of STR sequencing by NGS methods, however these studies have typically employed custom assays developed in-house ([2–10], for example). An additional barrier to the implementation of STR sequencing had been the absence of software programs specifically intended for STR sequence data assembly, alignment and representation, though this gap has been directly addressed in the past few years by open-source solutions developed with forensic use in mind [11–14]. The recent commercial availability of massively parallel sequencing components and systems in kit format and designed specifically for forensic use has improved the feasibility of routine sequence-based typing of nuclear DNA markers. Results from

* Corresponding author at: DNA Support Unit, Federal Bureau of Investigation Laboratory, 2501 Investigation Parkway, Quantico, VA 22135, USA.
E-mail address: Rebecca.Just@ic.fbi.gov (R.S. Just).

http://dx.doi.org/10.1016/j.fsigen.2017.01.001

testing to date are promising [15–21], and would seem to improve the outlook for nearer term NGS implementation for STR typing in forensic labs.

One of the commercial assays, the ForenSeq™ DNA Signature Prep Kit (compatible with the MiSeq FGx™ instrument; Illumina, Inc., San Diego, CA), simultaneously targets 58 STR loci, along with up to 172 single nucleotide polymorphisms (SNPs) depending on the primer set selected [22]. In addition, the associated ForenSeq™ Universal Analysis Software (UAS) performs all secondary and tertiary data analyses, and presents the resulting STR genotypes in a repeat number format familiar to forensic scientists [23]. We evaluated the potential utility of the ForenSeq™ assay and software system for Y chromosome and autosomal STR (Y-STR and aSTR) typing via examination of 151 sample libraries developed from high quality DNAs amplified at the target template (1 ng). Prior to any interpretation or manual editing of the data, UAS allele calls for 4167 total loci were compared to standard capillary electrophoresis (CE) based fragment length data to investigate the reliability of the software in flagging potential data quality issues (e.g. elevated stutter, allelic dropout) that can result in incorrect STR typing results. Following analyst review and editing of the data, 3060 loci (from 108 complete aSTR genotypes and 38 Y-STR haplotypes) were compared to the CE data to assess concordance of the finalized profiles. Here, we present the results of these examinations, and discuss notable features of the ForenSeq assay and UAS performance that will assist future STR typing efforts using the system.

2. Materials and methods

High quality DNA extracts from 103 individuals (collected and typed with informed consent), plus two positive control DNAs (2800M and 9947A), were used for the study. Including 2800M, 36 of the samples were male. Amplification, library preparation, sequencing and data analysis of samples using the ForenSeq™ system (Illumina, Inc.) occurred in five total runs. Runs 1–4 used a pre-release version of the assay, whereas Run 5 used the commercialized version.

Samples were amplified, some in replicate, using the ForenSeq™ DNA Signature Prep Kit with DNA Primer Mix B (Illumina, Inc.) and 1 ng template according to the manufacturer's protocol (see the DNA Signature Prep Kit Reference Guide [22]). Though DNA Primer Mix B targets 231 total nuclear DNA markers, including 27 aSTRs, 24 Y-STRs, 7 X chromosome STRs, 94 identity SNPs, 56 ancestry SNPs, 22 phenotypic SNPs and Amelogenin, only the aSTR and Y-STR markers were considered for this study. Library preparation from the amplified products, including target enrichment and indexing, library purification and library normalization, also followed the manufacturer's protocol. A MiSeq instrument (Illumina, Inc.) was used to sequence the normalized libraries in five pools, and for all but one pool the number of libraries exceeded the manufacturer's recommended limit of 32 libraries for "casework samples"/DNA Primer Mix B [22]. The five sequencing runs (1–5) were comprised of 42, 61, 33, 18 and 36 libraries (including positive and negative controls), respectively. Runs 2 (61 libraries) and 3 (33 libraries) included some evidentiary-type specimens that were not examined for this study. Sequencing Run 3 proceeded from a subset of the normalized libraries pooled for Run 2.

ForenSeq aSTR and Y-STR data were analyzed using UAS version 1.1, with data from each of the five MiSeq instrument runs imported separately for analysis. Data for the D22S1045 locus from Runs 1–4 could not be analyzed via any commercially available version of the UAS, and thus D22S1045 was only examined for Run 5. With the exception of the STR intralocus balance threshold, the default UAS v1.1 analysis settings for STRs (which are detailed in full in the ForenSeq Universal Analysis Software Guide for v1.1 [23], and are identical to those for version 1.2 [24]) were used. The STR intralocus balance threshold was lowered from the default 60% to 50% for all analyses to reduce the total number of loci flagged with UAS quality indicators. This adjustment followed an initial examination of samples typed in replicate that demonstrated the only quality issue at some flagged aSTR loci was heterozygote balance in the 50–60% range (data not shown). Of further note, the UAS default settings for aSTR and Y-STR analyses typically apply an analytical threshold of 1.5% of sequence reads, and an interpretation threshold of 4.5% of sequence reads (higher percentages for each threshold are used for three Y-STR loci). These percentages are applied to the complete read coverage (the sum of all reads) for the STR marker in a given sample, except when marker coverage is below 650 reads. At low read coverage, 650 reads is used by the UAS as a minimum value to which the analytical and interpretation threshold percentages are applied. In effect, then, the minimum analytical and interpretation threshold values are 10 and 30 reads, respectively. In addition, the interpretation threshold is used by the UAS as a minimum marker coverage threshold. When no identical set of sequence reads for a locus exceeds the interpretation threshold value, the UAS will designate the locus as inconclusive.

The UAS-analyzed aSTR and Y-STR data were examined via both the UAS interface and two standard UAS report outputs (the project level Genotype Report and genotype level Sample Details Report) in Microsoft Excel. Per-sample read counts were captured from the Sample Representation histogram in the project Quality Metrics tab within the UAS.

Comparative CE aSTR data for the 103 individuals and two positive control DNAs were produced using the PowerPlex® Fusion System (Promega Corporation, Madison, WI), while comparative Y-STR data for 36 individuals and one positive control DNA were generated with the Yfiler® Plus PCR Amplification Kit (Thermo Fisher Scientific, Waltham, MA), according to established laboratory protocols. GeneMapper® *ID-X* Software version 1.4 (Thermo Fisher Scientific) was used for sizing and allele calling of the CE data. In assessing concordance between ForenSeq and CE data, only loci that overlapped between the assays were compared. PowerPlex Fusion types 22 aSTRs that overlapped with the ForenSeq data analyzed from Runs 1–4, and 23 that overlapped with the ForenSeq data analyzed from Run 5. Yfiler Plus types 19 Y-STRs that are also typed by the ForenSeq system. For a few samples, the GlobalFiler® PCR Amplification Kit (Thermo Fisher Scientific) was also used to investigate aSTR markers that were inconsistent between the ForenSeq and Fusion profiles.

To assess both 1) ForenSeq aSTR and Y-STR recovery rates, and 2) UAS performance with regard to the detection of poor quality or inconclusive data that could lead to erroneous typing results, the first examinations in this study considered the UAS-determined genotyping results *prior* to any analyst review or manual editing of the ForenSeq data. For this portion of the study, CE-based allele calls were compared directly to the Genotype Report exported from the UAS for each sequencing run, and loci for which no alleles were recovered with the ForenSeq assay (i.e., any locus designated inconclusive by the UAS) were ignored.

To finalize ForenSeq aSTR and Y-STR profiles for the concordance examination, loci flagged by the UAS with quality control indicators (including inconclusive loci) from Runs 1, 3, 4 and 5 were reviewed, and allele calls were manually edited as necessary within the UAS without reference to the CE data. Run 2 results were not analyst-reviewed in full nor edited, as the high incidence of loci for which no sequence reads were recovered generally precluded the development of complete STR genotypes. Ultimately, finalized ForenSeq Genotype Reports were exported from the UAS, and all samples for which a complete aSTR or Y-STR profile was developed