

Reports

Short-read, high-throughput sequencing technology for STR genotyping

Daniel M. Bornman¹, Mark E. Hester¹, Jared M. Schuetter¹, Manjula D. Kasoji¹, Angela Minard-Smith¹, Curt A. Barden¹, Scott C. Nelson¹, Gene D. Godbold², Christine H. Baker², Boyu Yang², Jacquelyn E. Walther¹, Ivan E. Tornes¹, Pearly S. Yan³, Benjamin Rodriguez³, Ralf Bundschuh⁴, Michael L. Dickens¹, Brian A. Young¹, and Seth A. Faith¹

¹Battelle Memorial Institute, Columbus, OH, USA, ²Battelle Memorial Institute, Charlottesville, VA, USA, ³Human Cancer Genetics Program, The Ohio State University Comprehensive Cancer Center, Columbus, OH, USA, and ⁴Department of Physics and Biochemistry, Center for RNA Biology, The Ohio State University, Columbus, OH, USA

DNA-based methods for human identification principally rely upon genotyping of short tandem repeat (STR) loci. Electrophoretic-based techniques for variable-length classification of STRs are universally utilized, but are limited in that they have relatively low throughput and do not yield nucleotide sequence information. High-throughput sequencing technology may provide a more powerful instrument for human identification, but is not currently validated for forensic casework. Here, we present a systematic method to perform high-throughput genotyping analysis of the Combined DNA Index System (CODIS) STR loci using short-read (150 bp) massively parallel sequencing technology. Open source reference alignment tools were optimized to evaluate PCR-amplified STR loci using a custom designed STR genome reference. Evaluation of this approach demonstrated that the 13 CODIS STR loci and amelogenin (AMEL) locus could be accurately called from individual and mixture samples. Sensitivity analysis showed that as few as 18,500 reads, aligned to an *in silico* referenced genome, were required to genotype an individual (>99% confidence) for the CODIS loci. The power of this technology was further demonstrated by identification of variant alleles containing single nucleotide polymorphisms (SNPs) and the development of quantitative measurements (reads) for resolving mixed samples.

Keywords: STR; forensic; next-generation sequencing; high-throughput sequencing; Illumina; Bridge PCR; SNP; genotyping

Analysis of short tandem repeats (STRs) has become a well-established technology for human forensic casework (1–3). STRs, also known as microsatellites or simple sequence repeats (SSRs), are repetitive regions of DNA that contain unique core repeat units of 2–6 nucleotides in length (4). Several STR database systems have been established, including the Combined DNA Index System (CODIS) utilized in the United States (5). This system currently

uses a standard set of 13 STR loci, which are highly polymorphic, genetically unlinked and reside in noncoding regions (2). These STR alleles are routinely analyzed by multiplexed PCR followed by capillary electrophoresis (CE)-based separation (3,6–9). Although the CE-based technique for STR typing is both time and cost-effective, it does not allow for full sequence determination of STR loci and is only semiquantitative. Information on nucleotide variation

within STR alleles would be informative for discriminating alleles in partial profile situations, resolving mixed samples and in kinship analysis (11).

A current limitation of STR typing with CE is its limited bandwidth, restricting the use of additional investigative genetic markers that are potentially informative for ancestry, phenotype, and other attributes. Indeed, additional genetic markers have been recently added to the forensic

Method summary:

Here, we present a systematic method for high-throughput genotyping of the Combined DNA Index System (CODIS) short tandem repeat (STR) loci for human forensics using Illumina GAIIX short-read technology. Our novel contribution to the field is that we show that short-read-based next-generation sequencing technology can accurately genotype the CODIS STR loci from multiple samples and more importantly from mixed samples using quantitative measurements (reads). We also demonstrate method sensitivity, showing that as few as 18,500 reads, aligned to our *in silico* referenced genome, were required to genotype an individual (>99% confidence) for the entire CODIS panel of loci.

DNA analysis repertoire. For example, STR genotypes on the Y chromosome are often used in mixture analysis particularly in sexual assault cases (10), and various single nucleotide polymorphism (SNP) markers are now being used as predictors of externally visible characteristics (EVCs). The recent proliferation of genetic markers usable in forensic DNA analysis has resulted in a profusion of analytical platforms required to perform these assays (12). High-throughput sequencing (HTS) offers a single analytical platform for multiple forensic DNA analysis.

Previous reports have evaluated the potential of HTS for analyzing STR loci (13–15). In one of them, a pyrosequencing approach was used to examine 10 common markers for forensic analysis in Swedish individuals (13). Although this approach offers several advantages over CE-based STR analysis, including indentifying sequence variants within or proximal to STR loci and sequencing shorter DNA fragments, it was not successful in distinguishing between STRs with compound repeat motifs. Furthermore, this technology is not compatible with providing a high level of sequence coverage offered by HTS platforms. In a subsequent report, the Genome Sequencer FLX System (GS-FLX) (Roche Diagnostics, Branford, CT, USA) was evaluated for its ability to examine five STR loci from 10 human samples (14). This study demonstrated the benefits of STR genotyping using HTS technology (i.e., deeper resolution of STRs) and of using bioinformatic tools for sorting and evaluating sequence read lengths and frequencies critical for showing reliable and consistent results. Although this report showed the advantages of utilizing the Roche GS-FLX sequencer in STR-typing analyses, it is unknown whether a short-read next generation sequencing (NGS) platform (e.g., Illumina, San Diego, CA, USA or SOLiD, Applied Biosystems, Foster City, CA, USA) could provide the level of sensitivity required to detect an allelic mixture within a sample or accurately and reliably genotype all 13 CODIS STR loci (16).

In this report, we investigated whether a NGS platform based on short-read length technology could be applied to STR-typing analysis. Although a clear advantage of this technology is its ability to generate a large sequence data set exceeding 320 GB per run (17), it is unclear whether this type of platform, due to its short read output, could successfully sequence complete STR loci. To perform genotyping analysis, the Illumina GAIx was used to provide deep-sequencing of all 13 CODIS STR loci in addition to the amelogenin (AMEL)

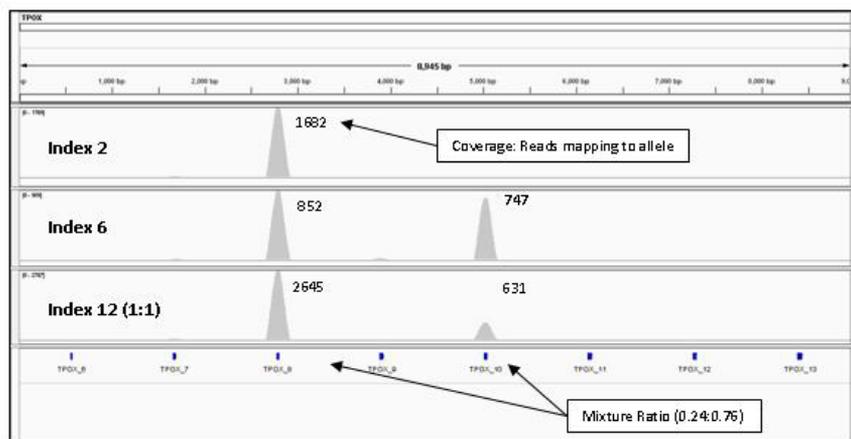


Figure 1. Visualization of mapped reads to specific alleles of the TPOX locus for two individual samples (index 2 and index 6) and a 1:1 mixture (index 12).

locus from five individual human samples and one mixture sample. This large data set was examined with a bioinformatics streamlined approach by aligning reads to an in silico reference comprised of all loci analyzed followed by read-specific filtering and allele association statistics. The Illumina GAIx data, analyzed by reference alignment, provided analogous allele calls for the 13 CODIS loci compared with the standard CE assay protocol based upon the PowerPlex 16 system (18), with the added benefit of detecting variant alleles of the same repeat length. Furthermore, this sequencing platform was able to discriminate STR alleles in a mixed sample preparation providing additional value for its use in forensic analyses.

Materials and methods

Sample collection and isolation

Human DNA (hDNA) was obtained from anonymous adult donors using a protocol approved by the Battelle Memorial Institute Internal Review Board. hDNA was collected and purified from saliva samples using the Oragene-DNA isolation kit (DNA Genotek, Kanata, ON, Canada), according to the manufacturer's recommended protocol. A total of five individual human saliva samples were evaluated. For mixture experiments, two of the five samples were mixed post-PCR amplification at a ratio of 1:1 prior to sequencing.

Illumina GAIx sequencing

The CODIS core 13 loci were individually amplified with Phusion High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA, USA) using custom designed primers covering 1000–2500 bp amplicons per locus (Supplementary Table S1). The length of these amplicons was designed to better simulate the use of genomic

DNA that would be fragmented prior to sequencing, rather than short amplicons formatted for a specific HTS technology. Individual PCR amplifications were pooled and purified with MinElute PCR Purification kit (Qiagen, Valencia, CA, USA). Sequencing libraries were constructed with the TruSeq DNA Library kit (Illumina) using index tags, according to the manufacturer's recommended protocol. Six multiplexed samples (five individual and one mixture) were pooled and sequenced in one lane of an Illumina GAIx at the Nucleic Acid Shared Resource Laboratory, The Ohio State University Medical Center (Columbus, OH, USA), for 150-bp single-end reads. Over 5 million pass-filtered reads were generated for each sample.

CE STR profiling

STR loci were amplified from human DNA samples using the PowerPlex 16 System (Promega, Madison, WI, USA), and the amplified products were detected using an Applied Biosystems 3130 Genetic Analyzer with Data Collection Software Version 3.0. Output data was analyzed using GeneMapper Software Version 4.0 (Applied Biosystems). All process quality values (PQV) were evaluated based on default values to determine the accuracy of genotyping.

STR profiling by NGS

A modified reference alignment method was developed to genotype STR loci. First, each CODIS STR locus sequence was designated by its repeat pattern and a segment of upstream and downstream nonrepetitive sequence. An in silico reference for each STR locus was constructed in FASTA format by generating a linear sequence of concatenated STR alleles as compiled by the National Institute of Standards and Technology (NIST). The Bowtie short read

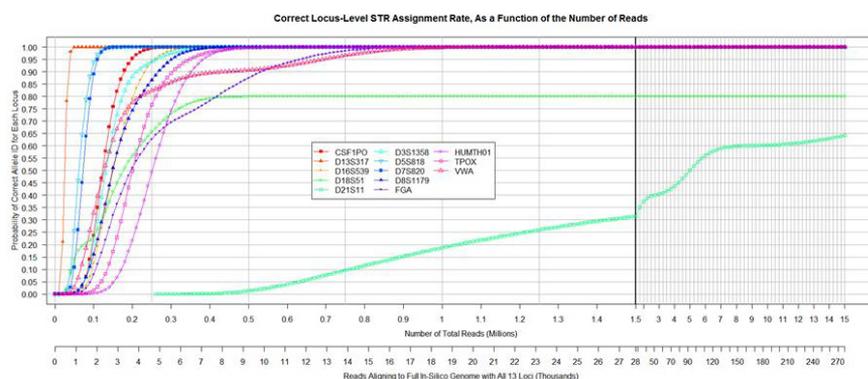


Figure 2. Probability of correct allele assignment for each CODIS STR locus. For each model, 10,000 draws were randomly obtained from a grid of (i) total reads, (ii) reads aligning to the in silico genome, and (iii) the number of reads mapping to the entire locus. The double x-axis represents a cumulative estimate of the (i) number of total raw sequencing reads and (ii) total number of reads aligning to the full STR in silico genome.

aligner was utilized for reference alignment based on several evaluated criteria: open-source accessibility, flexibility in specifying search and output parameters, and an ungapped alignment approach that was most applicable to STR alignment requirements (19). Optimized parameters of the algorithm were defined as seed length equal to 100 nucleotides, three allowed mismatches, and suppress all alignments for a particular read if more than one reportable alignment exists. Output from the aligner was generated in the SAM format and converted to BAM format using Samtools (20). Picard (<http://picard.sourceforge.net/>) was used to remove potential PCR duplicates, and BedTools (21) was used to convert BAM files to a BED file format. A custom script was used in the R statistical programming environment (22) to perform read filtering, retaining only reads that spanned the entire repeat region of any STR allele in the in silico genome and 5 bp of the 5'- and 3'-flanking regions. Finally, to make allele calls, a heuristic decision model based on Fisher's Exact Test was applied to evaluate the magnitude of reads mapping to each allele. A probability score was generated for each allele in the in silico genome based on: (i) the number of reads mapping to the allele, (ii) the total number of reads mapping to locus, (iii) the number of reads aligning to the in silico genome, and (iv) the total number of reads generated for a particular sample. This metric provided objective criteria on which to base STR genotyping calls from the NGS data.

Sensitivity analysis

A simulated analysis was performed to estimate the confidence of STR genotyping calls as related to number of reads. A Monte Carlo probabilistic model was applied to read counts that mapped to the in silico

reference alignment. For each model, 10,000 draws were randomly obtained from a grid of: (i) total reads, (ii) reads aligning to the in silico genome, and (iii) the number of reads mapping to the entire locus. Here, each draw resulted in a simulated table of alignment counts with each draw designated as either pass or fail. A passing draw had true allele counts with locus proportions significantly larger than 0.10, and all other alleles with proportions not significantly larger than 0.10. To test the Monte Carlo estimations, a separate analysis was also performed on subsets (10% and 1%) of the raw sequence data using the reference alignment method described previously.

Results and discussion

Data generated on the Illumina GAIIX system and analyzed with an optimized reference alignment method consistently identified 13 CODIS STR core loci and sex (AMEL) locus from single individuals and a mixed sample as compared with results that were obtained by CE using the PowerPlex 16 assay (Table 1). Since the Illumina GAIIX platform is capable of producing an extremely large number of reads per run, >39 million high-quality sequence reads in one lane (one-eighth of a flow cell) were produced that corresponded to the six multiplexed samples in this study (Supplementary Table S2). This high number of reads allowed for a high level of stringency in accurately calling STR alleles within a Fisher Exact Test, thereby discriminating against stutter or sequencing errors (Supplementary Table S3).

Overall, the NGS method was highly accurate in determining the STR allele types from simple, compound, and complex repeats as benchmarked against the CE-based method (Table 1). The STR allelic genotypes

for TPOX, D3S1358, FGA, CSF1PO, D5S818, D7S820, D8S1179, HUMTH01, VWA, D18S51, and D16S539 were identical between CE- and NGS-based methods. However, some discrete differences were observed for D18S51 and D21S11. First, some inconsistencies were observed with the NGS compared with the CE-based method for the D18S51 STR locus. For index samples 5 and 7, allele 14 could not be identified by reference alignment. This observation was unexpected given that this locus has a simple type of repeat pattern ($[AAGA]_n$), and the STR allele length at this locus is completely spanned by 150-bp sequence reads. Although the exact cause was not determined for the allele drop-out by NGS, it is possible that the custom PCR primers used to make the libraries did not sufficiently amplify this allele. Hence, the allele was not present nor observed in the reference alignment analysis.

One of the observed limitations of the current NGS method is in the length of sequence reads for detecting longer STR alleles. The data used in this study was 150-bp single-end reads, which is currently the maximum high-quality read length for this DNA sequencing platform. Although an analysis of read quality showed that quality diminished toward the 3'-end of the reads, a significantly large number of reads maintained an average score of Q20 and Q30 throughout the entire 150-bp read length, as expected (Supplementary Table S2). Due to the length criteria, the short-read NGS method was unsuccessful in detecting longer alleles (34.2 in index 2, and 32.2 in index 7) from the D21S11 CODIS STR. Additional corroborating evidence for this observation was also seen in the mixture sample, index 12, in which alleles 32.2 and 34.2 were not detected as well. The longest allele typed in this study was 31.2 at the D21S11 locus. Approximately ~17% D21 alleles and <1% FGA alleles are longer than this allele (23). Importantly, at about 6% in most populations, the D21S11 32.2 allele is the most common allele among the CODIS loci that are longer than 32 repeats. Thus, a marginal increase in read lengths on the Illumina GAIIX platform will capture this allele. Significant improvement in read length will be required to capture all of the rare but extremely long alleles reported at the FGA locus (24).

Since forensic samples often contain the genetic material of more than one individual, it becomes even more critical to accurately genotype STR alleles from complex mixture samples (25–27). Therefore, the NGS method was tested using a sample (index 12) based on a mixture

Table 1. STR genotype reported for each loci examined by CE and NGS across five individuals (indices 2, 4, 5, 6, and 7) and a 1:1 mixture of indices 2 and 6 (index 12)

Sample	TPOX		D3S1358		FGA		CSF1PO		D5S818		D7S820		D8S1179	
	CE	NGS	CE	NGS	CE	NGS	CE	NGS	CE	NGS	CE	NGS	CE	NGS
Index 2	8	8	16	16	22/23	22/23	11/12	11/12	11/13	11/13	10/11	10/11	13	13
Index 4	10/12	10/12	14	14	22/25	22/25	12	12	11/12	11/12	10/11	10/11	13/14	13/14
Index 5	8	8	17/19	17/19	21	21	10/12	10/12	13	13	10/11	10/11	13/15	13/15
Index 6	8/10	8/10	16	16	22/26	22/26	10/12	10/12	11/13	11/13	11	11	13/15	13/15
Index 7	8/11	8/11	15/18	15/18	21/25	21/25	10/11	10/11	10/13	10/13	12/13	12/13	13/14	13/14
Index 12 (1:1 mixture)	NA	8/10	NA	16	NA	22/23/ 26	NA	10/11/ 12	NA	11/13	NA	10/11	NA	13/15

Sample	TH01		VWA		D13S317		D16S539		D18S51		D21S11		AMEL	
	CE	NGS	CE	NGS	CE	NGS	CE	NGS	CE	NGS	CE	NGS	CE	NGS
Index 2	6/9.3	6/9.3	16/17	16/17	11/13	11/13	12/13	12/13	13/17	13/17	30/34.2	*30	X	X
Index 4	6/9.3	6/9.3	15/19	15/19	8/12	8/12	11/12	11/12	13/17	13/17	28/30	28/30	XY	XY
Index 5	7/9.3	7/9.3	15/16	15/16	11	11	9/11	9/11	14/15	*15	30	30	XY	XY
Index 6	6/9	6/9	14/18	14/18	11/12	11/12	11/14	11/14	13/15	13/15	28/30.2	28/30.2	XY	XY
Index 7	6/9.3	6/9.3	15	15	11/12	11/12	12/13	12/13	14/17	*17	31.2/32.2	*31.2	XY	XY
Index 12 (1:1 mixture)	NA	6/9/9.3	NA	14/16/ 17/18	NA	11/12/ 13	NA	11/12/ 13/14	NA	13/15/ 17	NA	28/30/ 30.2	NA	XXY

of indexes 2 and 6. Equal amounts of DNA from these individuals were mixed and sequenced on the Illumina GAIIX. Table 2 shows that the quantification of mapped reads from NGS analysis yielded expected ratios for the distribution of alleles from the two individuals when mixed at a 1:1 ratio. For example, index 2 was homozygous for TPOX 8, while index 6 was heterozygous for alleles 8 and 10. The resulting TPOX 10:8 ratio was 0.24:0.76 as expected (Figure 1 and Table 2).

Another benefit from the NGS-based STR typing method was sequence determination of STR alleles and detection of variant alleles that were not initially observed by the CE-based method. For example, using NGS variants of allele 16 at the D3S1358 locus from index 2 and 6 were detected. Alignment to the in silico STR genome, containing a representation of each D3S1358 allele variant, revealed that index 2 presented alleles 16a/16b and index 6 presented alleles 16b/16c (Table 2). Analysis of index 12 successfully identified all three allele variants in expected relative abundances based on the number of mapped reads for each allele within the mixture (0.25:0.48:0.28 or 1:2:1). These results demonstrate the advantage of using NGS compared with CE, as samples remained indistinguishable using CE (size-based), but can be discriminated utilizing the NGS method (SNP- or sequence-based).

In order to estimate the number of short reads required to accurately assess each STR allele present in each sample, a Monte Carlo probabilistic model was fit to mapped read counts to an in silico reference alignment.

Figure 2 shows the proportion of passing subsample draws at varying values of read assignments for each of the models. The STR locus detected at greater than 95% probability with the fewest number of reads was D13S317. D13S317 also provided the highest number of mapped (aligned) reads relative to the other loci (Table 2 and Supplementary Tables S3 and S4) and is also one of the shortest loci in the panel (28–60 bp within the repeat). However, it was difficult to ascertain from the data (Figure 2) if this method of STR identification was more accurate for shorter alleles (e.g., D13S317), or if PCR amplification bias and/or sequencing bias had a greater influence in the model. Additional samples would be required to fully evaluate this observation.

The average number of total raw sequenced reads required to correctly identify the CODIS STR alleles with greater than 99% confidence in each sample was less than one million. In addition, the minimum number of reads mapped to the entire in silico reference required to make the same calls was as few as 18,500. However, the limitations of determining longer alleles of D21S11 and potential allele drop-out from D18S51 was further illustrated with the Monte Carlo Model, as increasing number of reads failed to provide 100% accurate genotyping for these loci using this sample set (Figure 2). Similar results were obtained from submitting subsets (10% and 1%) of total sequence reads to the reference aligner (Supplementary Table S4). These data showed that significantly fewer reads than were initially generated were

required to make STR genotyping calls by the NGS approach. Consequently, in order to maximize the number of individuals included within a single sequencing run and thereby reduce the cost of analyzing STRs by this method, multiplexing of samples within a large cohort of individuals is recommended.

Here we have demonstrated the potential of using a widely adopted NGS platform for accurately genotyping the CODIS STR loci. Due to the short length of sequence reads attributed to this platform, sequencing STR loci was previously assumed impractical because of their relatively large repeat size (14). The D21S11 locus was the only problematic STR in this regard, as repeat regions extending beyond the 150 bp maximum read length were not resolved. Paired-end (PE) sequencing with overlapping reads could possibly be beneficial for the Illumina GAIIX platform, potentially generating assembled reads longer than 150 bp. However, with simple repeats, the degree of overlap between the paired ends cannot be determined with confidence. Furthermore, STR analysis would not profit from nonoverlapping PE reads with unsequenced inserts, since the reference alignment method requires a contiguous sequence covering the entire STR repeat and some of the unique flanking regions both up- and downstream of the repeat. A final benefit of using the Illumina NGS platform in STR genotyping analyses is the high level of coverage obtained from one sequencing run. Not only did the large sequence data allow for high probability in accurately

Table 2. Mapped read counts associated with each allele variant detected by NGS that was confirmed by CE

STR Allele	NGS Read Counts Associated with each STR Allele			Expected Mixture Ratio	Actual Mixture Ratio
	Index 2	Index 6	Index 12 1:1 mixture		
TPOX_8	1682	852	2645	0.75	0.76
TPOX_10		747	631	0.25	0.24
D3S1358_16a	1573		1798	0.25	0.25
D3S1358_16b	1346	2273	3460	0.50	0.48
D3S1358_16c		2204	2015	0.25	0.28
FGA_22	1355	1232	2244	0.50	0.60
FGA_23	965		927	0.25	0.25
FGA_26		713	563	0.25	0.15
CSF1PO_10		3007	2226	0.25	0.30
CSF1PO_11	1375		1736	0.25	0.23
CSF1PO_12	1420	2514	3445	0.50	0.47
D5S818_11	2989	5224	8119	0.50	0.54
D5S818_13	2695	4459	6791	0.50	0.46
D7S820_10	2550		3568	0.25	0.21
D7S820_11	2210	12364	13083	0.75	0.79
D8S1179_13	2645	1926	4390	0.75	0.78
D8S1179_15		1621	1210	0.25	0.22
HUMTH01_6	744	1183	2031	0.50	0.49
HUMTH01_9		1146	1096	0.25	0.27
HUMTH01_9.3	677		977	0.25	0.24
VWA_14b		1239	998	0.25	0.11
VWA_16	1714		2809	0.25	0.31
VWA_17	1370		2543	0.25	0.28
VWA_18		3009	2555	0.25	0.29
D13S317_11	7062	14699	22407	0.50	0.50
D13S317_12		14192	12492	0.25	0.28
D13S317_13	6553		9511	0.25	0.21
D16S539_11		2758	2474	0.25	0.31
D16S539_12	1377		2007	0.25	0.25
D16S539_13	1130		1766	0.25	0.22
D16S539_14		2324	1803	0.25	0.22
D18S51_13	1197	3594	5570	0.50	0.54
D18S51_15		3135	2881	0.25	0.28
D18S51_17	942		1862	0.25	0.18
D21S11_28		565	482	0.50	0.65
D21S11_30	77		83	0.25	0.11
D21S11_30.2b		244	181	0.25	0.24
AMELX	90464	152915	240154	0.75	0.67
AMELY		147219	118299	0.25	0.33

calling alleles at STR loci, it is also amenable for genotyping a large number of individuals. Based on the sensitivity analysis for determining minimum read lengths necessary for accurately calling alleles for a given STR locus, we estimate that >300 individuals could theoretically be STR-genotyped using eight lanes (one flow cell) of an Illumina GAIIx, thereby providing a significant cost and time savings advantage. Alternatively, the massive sequencing bandwidth could be used to evaluate other forensically relevant genetic features, such as ancestry informative markers or forensic DNA phenotypic markers.

Acknowledgments

This research was funded by an internal research and development program at Battelle Memorial Institute. PSY and BR were supported by NCI Comprehensive Cancer Center Support Grant P30 CA016058.

Competing interests

The authors declare no competing interests. The authors and their institutions do not specifically endorse any of the third-party products or technologies described in this report.

References

- Budowle, B., A. Masibay, S.J. Anderson, C. Barna, L. Biega, S. Brenneke, B.L. Brown, J. Cramer, et al. 2001. STR primer concordance study. *Forensic Sci. Int.* 124:47-54.
- Jobling, M.A. and P. Gill. 2004. Encoded evidence: DNA in forensic analysis. *Nat. Rev. Genet.* 5:739-751.
- Butler, J.M. 2007. Short tandem repeat typing technologies used in human identity testing. *BioTechniques* 43:ii-v.
- Ellegren, H. 2004. Microsatellites: simple sequences with complex evolution. *Nat. Rev. Genet.* 5:435-445.
- Budowle, B., T.R. Moretti, S.J. Niezgodna, and B.L. Brown. 1998. CODIS and PCR-based short tandem repeat loci: law enforcement tools, p. 73-88. In *Proceedings of the Second European Symposium on Human Identification* (June 1998, Innsbruck, Austria). Promega Corporation, Madison, WI.
- Butler, J.M., E. Buel, F. Crivellente, and B.R. McCord. 2004. Forensic DNA typing by capillary electrophoresis using the ABI Prism 310 and 3100 genetic analyzers for STR analysis. *Electrophoresis* 25:1397-1412.
- Fregeau, C.J. and R.M. Fourney. 1993. DNA typing with fluorescently tagged short tandem repeats: a sensitive and accurate approach to human identification. *BioTechniques* 15:100-119.
- Kimpton, C.P., P. Gill, A. Walton, A. Urquhart, E.S. Millican, and M. Adams. 1993. Automated DNA profiling employing multiplex amplification of short tandem repeat loci. *PCR Methods Appl.* 3:13-22.

9. Edwards, A., A. Civitello, H.A. Hammond, and C.T. Caskey. 1991. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Am. J. Hum. Genet.* 49:746-756.
10. Ge, J., B. Budowle, and R. Chakraborty. 2010. Interpreting Y chromosome STR haplotype mixture. *Leg Med (Tokyo)* 12:137-143.
11. Pitterl, F., K. Schmidt, G. Huber, B. Zimmermann, R. Delpont, S. Amory, B. Ludes, H. Oberacher, and W. Parson. 2010. Increasing the discrimination power of forensic STR testing by employing high-performance mass spectrometry, as illustrated in indigenous South African and Central Asian populations. *Int. J. Legal Med.* 124:551-558.
12. Zietkiewicz, E., M. Witt, P. Daca, J. Zebracka-Gala, M. Goniewicz, and B. Jarzab. 2012. Current genetic methodologies in the identification of disaster victims and in forensic analysis. *J. Appl. Genet.* 53:41-60.
13. Divne, A.M., H. Edlund, and M. Allen. 2010. Forensic analysis of autosomal STR markers using Pyrosequencing. *Forensic Sci. Int. Genet.* 4:122-129.
14. Fordyce, S.L., M.C. Avila-Arcos, E. Rockenbauer, C. Borsting, R. Frank-Hansen, F.T. Petersen, E. Willerslev, A.J. Hansen, et al. 2011. High-throughput sequencing of core STR loci for forensic genetic investigations using the Roche Genome Sequencer FLX platform. *BioTechniques* 51:127-133.
15. McGuigan, J., N. Shouyong, S. Liu, E. Bouton, M. Manion, C.S.J. Liu, and M. Holland. 2011. Human Identity Analysis with NextGENe Software Using Pyrosequencing Reads. *SoftGenetics*, State College, PA.
16. Hert, D.G., C.P. Fredlake, and A.E. Barron. 2008. Advantages and limitations of next-generation sequencing technologies: a comparison of electrophoresis and non-electrophoresis methods. *Electrophoresis* 29:4618-4626.
17. http://www.illumina.com/documents/products/brochures/brochure_sequencing_systems_portfolio.pdf. Accessed 27 February 2012.
18. Greenspoon, S.A., J.D. Ban, L. Pablo, C.A. Crouse, F.G. Kist, C.S. Tomsey, A.L. Glessner, L.R. Mihalacki, et al. 2004. Validation and implementation of the PowerPlex 16 BIO System STR multiplex for forensic casework. *J. Forensic Sci.* 49:71-80.
19. Langmead, B., C. Trapnell, M. Pop, S.L. Salzberg. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25.
20. Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin and 1000 Genome Project Data Processing Subgroup. 2009. The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics*, 25, 2078-9.
21. Quinlan, A.R. and I.M. Hall. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841-842.
22. Ihaka, R. and R. Gentleman. 1996. R: A language for data analysis and graphics. *Journal of computational and graphical statistics: a joint publication of American Statistical Association, Institute of Mathematical Statistics, Interface Foundation of North America* 5:299-314.
23. Budowle, B., B. Shea, S. Niezgoda, and R. Chakraborty. 2001. CODIS STR loci data from 41 sample populations. *J. Forensic Sci.* 46:453-489.
24. Griffiths, R.A., M.D. Barber, P.E. Johnson, S.M. Gillbard, M.D. Haywood, C.D. Smith, J. Arnold, T. Burke, et al. 1998. New reference allelic ladders to improve allelic designation in a multiplex STR system. *Int. J. Legal Med.* 111:267-272.
25. Weir, B.S., C.M. Triggs, L. Starling, L.I. Stowell, K.A. Walsh, and J. Buckleton. 1997. Interpreting DNA mixtures. *J. Forensic Sci.* 42:213-222.
26. Evett, I.W., C. Buffery, G. Willott, and D. Stoney. 1991. A guide to interpreting single locus profiles of DNA mixtures in forensic cases. *J. Forensic Sci. Soc.* 31:41-47.
27. Budowle, B., A.J. Onorato, T.F. Callaghan, A. Della Manna, A.M. Gross, R.A. Guerrieri, J.C. Luttman, and D.L. McClure. 2009. Mixture interpretation: defining the relevant features for guidelines for the assessment of mixed DNA profiles in forensic casework. *J. Forensic Sci.* 54:810-821.

Received 7 February 2012; accepted 8 March 2012.

Address correspondence to Seth A. Faith, Battelle Memorial Institute, 505 King Avenue, Columbus, OH, USA. e-mail: faiths@battelle.org

Supplementary material for this article is available at www.BioTechniques.com/article/113857.

To purchase reprints of this article, contact: biotechniques@fosterprinting.com