Forensic Science International: Genetics Supplement Series 3 (2011) e357–e358

Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

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

Short tandem repeat sequencing on the 454 platform

Melissa Scheible, Odile Loreille, Rebecca Just, Jodi Irwin *

Armed Forces DNA Identification Laboratory, Armed Forces Medical Examiner System, Rockville, Maryland, USA

ARTICLE INFO

Article history:
Received 25 August 2011
Accepted 15 September 2011

Keywords:
454 sequencing
Short tandem repeat
Next generation sequencing

ABSTRACT

To investigate the feasibility of next generation sequencing (NGS) technology for the multiplex detection and sequence production of short tandem repeats (STRs), thirteen STR markers (CSF1PO, D2S441, D3S1358, D5S818, D7S820, D8S1179, D10S1248, D13S317, D16S539, D21S11, D22S1045, TPOX, and vWA) were amplified using an optimized multiplex reaction with primer sequences designed for reduced size amplicons. Each sample multiplex was barcoded with a different sample-specific multiplex identifier (MID) for subsequent parallel tagged sequencing on the GS Junior System (454 Life Sciences, Branford, CT).

Published by Elsevier Ireland Ltd.

1. Introduction

The development of reduced size amplicons for STR markers has increased the success of typing samples with degraded DNA [1–4]. However, when separating these alleles using capillary electrophoresis, the number of loci that can be multiplexed together is limited by the number of available dyes. Those loci with overlapping size ranges must either be labeled with different dyes or amplified in separate multiplexes [4]. One potential solution to this limitation of standard chemistries and detection platforms is next generation sequencing technology based on clonal amplification by emulsion polymerase chain reaction (emPCR) followed by pyrosequencing [5]. With this strategy, many loci with overlapping amplicon size ranges can be multiplexed in one reaction, which allows conservation of valuable sample extract.

2. Materials and methods

A multiplex of thirteen STR markers was amplified as described in [6] with a total amplicon size range of 69–211 base pairs. Individual samples were tagged with different multiplex identifiers (MIDs) and processed according to the GS Junior Titanium Series user manuals [7–9] with a few exceptions. First, no fragmentation by nebulization was necessary, since amplified targets were already smaller than the length that requires fragmentation. Second, for the removal of small fragments such as adaptors and adaptor dimers, the Agencourt® AMPure® XP (Beckman Coulter, Inc., Danvers, Massachusetts) protocol was used according to manufacturer recommendations [10] rather than the bead preparation recommended by the 454 protocol that intentionally selects for fragments larger than 300 base pairs [7]. Third, one molecule of sample library per capture bead was targeted for emPCR, as previous experiments showed that two molecules per bead produced an overabundance of beads containing multiple templates. This, in turn, leads to a drastic reduction in successful reads since those reads representing multiple templates are discarded during the data filtering process. Lastly, the quantity of amplification primer for the emPCR reaction was reduced as suggested for amplicon libraries of very short fragments. This decreases crosstalk and is intended to eliminate incomplete extension during the sequencing reaction [11].

Post-processing data analysis was performed with the CLC Genomics Workbench software (CLC bio, Aarhus, Denmark). Once high throughput sequencing data were imported, reads representing individual samples were separated based on the unique MID sequence tag. Sequences were then aligned to references designed for this project. Reference sequences were created for each allele of every locus and spanned both amplification primers in length. Default mapping and alignment parameters were adjusted so that only the sequences matching a reference by 90% of its length and 90% of its base similarity were captured. Non-specific matches, reads that could be aligned equally well to multiple reference sequences [12], were ignored (not aligned). These stringent parameters were used to both ensure that STR markers with similar repeats would not align to the incorrect references and eliminate reads ending within the repeat region.

3. Results and discussion

The preliminary 454 data were extremely promising in terms of total data recovery, despite an unequal distribution of reads among

* Corresponding author at: AFDIL, 1413 Research Boulevard, Rockville, MD 20850, United States. Tel.: +1 301 319 0244; fax: +1 301 295 5932.
E-mail address: jodi.a.irwin@us.army.mil (J. Irwin).

1875-1768/$ – see front matter. Published by Elsevier Ireland Ltd.
doi:10.1016/j.fsigss.2011.09.041

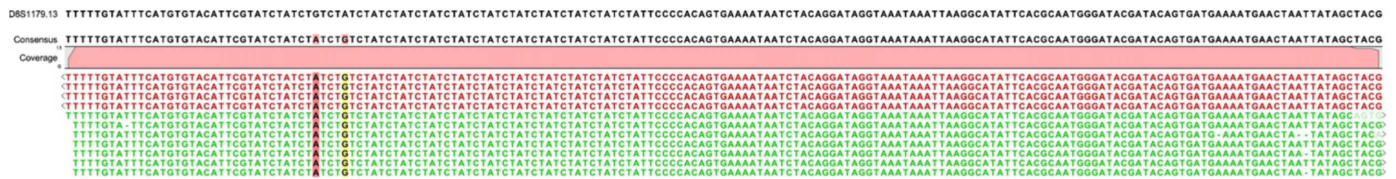

Fig. 1. Sequences aligned to a reference showing single nucleotide polymorphisms within the repeat region.

the thirteen loci. Further optimization of the multiplex amplification is currently underway to correct this imbalance. Despite this issue, the recovered data and resulting aligned reads were more than sufficient to reveal sequence variation among samples. Single nucleotide polymorphisms (SNPs) were observed in varying frequencies from sample to sample, introducing an opportunity for discrimination not available with commercially available STR kits (Fig. 1). However, this additional information also introduces some complications when viewed from the familiar standpoint of repeat-based allele nomenclature. Conversion of the sequence-based information to repeat-based data will undoubtedly be necessary to facilitate comparisons to existing fragment-based STR profiles, but the benefits of recovering the sequence information would be lost if allele calls are based solely on the number of repeats. Thus, some type of alternative nomenclature system that retains the repeat-based information, yet also captures the SNP variation, should be considered moving forward.

While sequence data for STRs can also be generated using the standard Sanger method, the difficulties associated with amplifying each locus independently, separating heterozygous alleles on a gel, and Sanger sequencing small fragments make this option unfeasible for routine practice. Instead, the simplicity of simultaneously typing multiple samples for numerous mini-amplicon STR markers each highlights the benefits of NGS for STR sequence generation.

4. Conclusion

The relative ease of obtaining large quantities of sequence data via emPCR and pyrosequencing, the additional discriminatory power provided by SNPs otherwise indiscernible with traditional STR typing methods, and the potential to multiplex many markers with overlapping size ranges highlight the benefits of STR sequencing on the 454 platform. Additional multiplexes for a given sample could be amplified separately and then easily combined for the emPCR and sequencing steps to further increase the genetic information recovered. The additional data, and greater discriminatory power resulting from those data, would be valuable to missing persons identification efforts, especially those focused on highly degraded skeletal elements containing fragmented DNA [13,14].

Conflict of interest

None.

Acknowledgements

The authors would like to thank James Canik, Lanelle Chisolm, LTC Louis Finelli, James Ross, Shairose Lalani, Marjorie Bland and the American Registry of Pathology for logistical and administrative support; Suzanne Barritt-Ross for allowing the use of non-probative casework samples; the AFDIL Research Department for continuous support and valuable feedback; and John Butler for permission to use NIST samples for these preliminary experiments.

The opinions and assertions contained herein are solely those of the authors and are not to be construed as official or as views of the US Department of Defense or the US Department of the Army.

References

[1] J.M. Butler, Y. Shen, B.R. McCord, The development of reduced size STR amplicons as tools for analysis of degraded DNA, J. Forensic Sci. 48 (2003) 1054–1064.
[2] M.D. Coble, J.M. Butler, Characterization of new miniSTR loci to aid analysis of degraded DNA, J. Forensic Sci. 50 (2005) 43–53.
[3] P. Grubwieser, R. Muhlmann, B. Berger, H. Niederstatter, M. Pavlic, W. Parson, A new miniSTR-multiplex displaying reduced amplicon lengths for the analysis of degraded DNA, Int. J. Legal Med. 120 (2006) 115–120.
[4] C.R. Hill, M.C. Kline, M.D. Coble, J.M. Butler, Characterization of 26 miniSTR loci for improved analysis of degraded DNA samples, J. Forensic Sci. 53 (2008) 73–80.
[5] M. Margulies, M. Egholm, W.E. Altman, S. Attiya, J.S. Bader, L.A. Bemben, et al., Genome sequencing in microfabricated high-density picolitre reactors, Nature 437 (2005) 376–380.
[6] F. Pitterl, H. Niederstatter, G. Huber, B. Zimmermann, H. Oberacher, W. Parson, The next generation of DNA profiling—STR typing by multiplexed PCR–ion-pair RP LC-ESI time-of-flight MS, Electrophoresis 29 (2008) 4739–4750.
[7] 454 Life Sciences Corp, Rapid Library Preparation Method Manual: GS Junior Titanium Series, Branford, CT, 2010.
[8] 454 Life Sciences Corp, emPCR Amplification Method Manual–Lib-L: GS Junior Titanium Series, Branford, CT, 2010.
[9] 454 Life Sciences Corp, Sequencing Method Manual: GS Junior Titanium Series, Branford, CT, 2010.
[10] Agencourt Bioscience Corporation, A Beckman Coulter Company, Agencourt® AMPure® XP PCR Purification, Beverly, MA, 2009.
[11] 454 Life Sciences Corp, Application Brief: Amplicon (PCR Product) Sequencing Tips for GS FLX Titanium Reagents, Branford, CT, 2010.
[12] CLC bio, CLC Genomics Workbench User Manual, Aarhus, Denmark, 2011.
[13] J. Irwin, M.D. Leney, O. Loreille, et al., Application of low copy number STR typing to the identification of aged, degraded skeletal remains, J. Forensic Sci. 52 (2007) 1322–1327.
[14] T.J. Parsons, R. Huel, J. Davoren, et al., Application of novel mini-amplicon STR multiplexes to high volume casework on degraded skeletal remains, Forensic Sci. Int. Genet. 1 (2007) 175–179.