Forensic Science International: Genetics Supplement Series 2 (2009) 529–531

Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

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

Research article

Enhancing resolution and statistical power by utilizing mass spectrometry for detection of SNPs within the short tandem repeats

John V. Planz [a,b,*], Bruce Budowle [a,b], Thomas Hall [c], Arthur J. Eisenberg [a,b], Kristin A. Sannes-Lowery [c], Steven A. Hofstadler [c]

[a] Department of Forensic and Investigative Genetics, UNT Health Science Center at Fort Worth, Fort Worth, TX, USA
[b] Institute of Investigative Genetics, UNT Health Science Center at Fort Worth, Fort Worth, TX, USA
[c] Ibis Biosciences, Inc., Carlsbad, CA, USA

ARTICLE INFO

Article history:
Received 24 August 2009
Accepted 26 August 2009

Keywords:
STR
Mass spectrometry
SNPs
Discrimination power

ABSTRACT

Short tandem repeats (STRs) are used routinely for the analysis of DNA samples from evidentiary items, convicted offenders, relationship testing and other identity testing disciplines. The discriminatory power of the STRs is sufficient in most human identity testing comparisons to render an identification. However, STRs have some limitations in evaluations, such as parentage testing, identification of human remains, or pairwise evaluations of putative relatives by familial searching. A major assumption is that shared alleles in these associations stem from common ancestry, i.e. are Identical by Descent (IBD). However, STR alleles by definition are Identical by State (IBS). Using an electrospray ionization mass spectrometry (ESI-MS) system developed by Ibis Biosciences Inc., population databases were generated for the 13 core CODIS STRs from African Americans, Caucasians and Hispanics capturing both the length of the allele, as well as SNP variation contained within repeat motifs. SNPs were identified in 10 of the loci and some common alleles were subdivided with SNP typing. Inclusion of SNPs increases discrimination power significantly, whereby the seven most polymorphic SNP-containing STR loci have the discriminatory power of 10 traditionally typed loci. A system of nomenclature has been developed that facilitates the databasing, searching and analyses of these combined data forms.

1. Introduction

Short tandem repeat (STR) loci are the standard genetic markers for forensic DNA analysis worldwide. Suites of loci, such as the 13 locus US CODIS panel or the seven consensus STR loci utilized by European Interpol laboratories, often provide sufficient information to attribute an individual as the source of a crime scene sample or provide a strong investigative lead. Although current STR systems offer high levels of differentiation, standard methods have certain limitations when confronting degraded samples, relationship testing cases, or missing persons identification.

Mass spectrometry is a powerful analytical platform for characterizing nucleic acids, such as those generated by the PCR. Because the exact mass of each of the DNA bases is known, accurate mass measurements can be used to derive base compositions for PCR amplicons [1,2]. Since DNA consists of two complementary strands, the compositions of the two strands can be constrained through Watson–Crick base pair complementarity to provide confidence in the proper base composition assignment.

The mass accuracy and resolution obtained with electrospray ionization (ESI) TOF MS, has been shown to simultaneously detect length and nucleotide variability in STRs [3]. Of 21 STR loci tested on an Austrian population ($n = 99$), 11 loci (SE33, D2S1338, vWA, D21S11, D3S1358, D16S539, D8S1179, D7S820, D13S317, D5S818, and D2S441) contained SNPs that were not discernible by capillary electrophoresis.

An automated high-throughput ESI-MS system, the Ibis T5000™ (Ibis Biosciences Inc., Carlsbad, CA) was evaluated for STR typing effectiveness. The T5000™ platform combines automated DNA desalting with direct injection of the PCR product into the mass spectrometer. Coupled with the system is a database which automates data interpretation and a user interface to assist in the evaluation of DNA profiles. This approach has been successfully applied to mtDNA analyses yielding discriminatory power (DP) approaching that attained with sequencing [4].

2. Methods

An STR assay was developed that amplifies the US CODIS STR loci and amelogenin in eight PCRs. A sample is distributed across

* Corresponding author at: Department of Forensic and Investigative Genetics, UNT Health Science Center, 3500 Camp Bowie Blvd., EAD 310B, Fort Worth, TX 76107, USA. Tel.: +1 817 735 2397.
E-mail address: jplanz@hsc.unt.edu (J.V. Planz).

1875-1768/$ – see front matter © 2009 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.fsigss.2009.08.163

Table 1
Descriptive statistics for seven most polymorphic STR loci containing SNPs detected using ESI-MS.

Locus	Population	STR-SNP analysis on IBIS T5000					STR only analysis on IBIS T5000				
		n	Alleles detected	DP	H_e	H_o	*n*	Alleles detected	DP	H_e	H_o
D13S317	Caucasian	181	12	0.9705	0.8735	0.9061	182	7	0.9213	0.7820	0.8297
	African Am.	213	12	0.9528	0.8345	0.7981	214	7	0.8607	0.7026	0.7056
	Hispanic	193	13	0.9751	0.8847	0.9016	193	7	0.9445	0.8241	0.8290
D21S11	Caucasian	181	23	0.9780	0.8925	0.9006	182	14	0.9540	0.8390	0.8681
	African Am.	213	33	0.9708	0.8688	0.8357	214	20	0.9589	0.8459	0.8178
	Hispanic	193	25	0.9752	0.8845	0.8446	193	14	0.9521	0.8335	0.8083
D3S1358	Caucasian	181	18	0.9671	0.8641	0.8895	182	8	0.9226	0.7900	0.8077
	African Am.	213	18	0.9775	0.8907	0.9061	214	8	0.8923	0.7485	0.7850
	Hispanic	193	18	0.9455	0.8101	0.8187	193	8	0.8939	0.7391	0.7409
D5S818	Caucasian	181	15	0.9260	0.7870	0.8177	182	9	0.8432	0.6858	0.6703
	African Am.	213	17	0.9102	0.8396	0.8310	214	9	0.8932	0.7449	0.7009
	Hispanic	193	13	0.9554	0.7471	0.7617	193	9	0.8679	0.6998	0.6891
D7S820	Caucasian	181	15	0.9600	0.8525	0.8066	182	8	0.9349	0.8089	0.7857
	African Am.	213	12	0.9376	0.8117	0.7887	214	8	0.7	0.7957	0.7710
	Hispanic	193	14	0.9482	0.8271	0.7876	193	9	0.7358	0.7895	0.7358
D8S1179	Caucasian	181	14	0.9627	0.8554	0.8785	182	10	0.9324	0.7970	0.8187
	African Am.	213	19	0.9489	0.8215	0.8122	214	10	0.9239	0.7860	0.7523
	Hispanic	193	16	0.9639	0.8581	0.8860	193	9	0.9303	0.7983	0.8394
vWA	Caucasian	181	22	0.9580	0.8471	0.8674	182	10	0.9388	0.8152	0.8022
	African Am.	213	26	0.9766	0.8882	0.8920	214	11	0.9403	0.8166	0.8224
	Hispanic	193	16	0.9305	0.7955	0.7979	193	7	0.9108	0.7692	0.7876

DP, discrimination power; H_e, expected heterozygosity; H_o, observed heterozygosity; *n*, sample size.

eight wells of a pre-fabricated PCR plate such that 12 samples occupy a 96-well plate. PCR primers were placed close to the repeat unit to produce amplicon lengths of 82–266 bp. After thermocycling, samples were analyzed on the T5000™ [2]. Automated data processing software was used to deconvolve raw mass spectra [4] and associate amplicon masses directly to STR alleles without the need for allelic ladders. Mass determinations were sufficiently accurate to determine amplicon base compositions [2], allowing accurate assignment of nucleotide polymorphisms relative to nominal alleles.

Population samples from 587 individuals (African American, Caucasian, Hispanic), previously typed with Powerplex®16 (Promega Corp., Madison, WI), were evaluated using the mass spectrometry STR assay. Allelic distributions at 13 STR loci (CSF1PO, FGA, TH01, TPOX, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11) were determined. Robustness of primer design was assessed and typing concordance was evaluated using Powerplex®16 results. Possible departure from HWE was tested using Fisher's Exact Test [5]. Linkage disequilibrium (LD) between loci was examined using an interclass correlation criterion for two locus associations [6].

3. Results

Mass spectrometry analysis revealed 10 of the 13 loci exhibited SNPs distributed within the STR motifs. The number of observed SNPs was low for CSF1P0, D18S51, and FGA, and did not contribute substantially to increased DP at these loci. The other seven loci contained SNPs which for some loci doubled the number of observed alleles compared with standard typing methods (Table 1). In several instances, multiple SNPs were detected within a single allele. Departure from HWE and LD was not observed. Nominal STR allele frequencies generated by mass spectrometry did not differ from published data [7]. STR typing results were concordant for all 587 individuals.

Analyses revealed that SNPs subdivide the more common nominal alleles, reapportioning the original frequency over two or more alleles. SNP-bearing alleles are not evenly distributed among population groups and further investigation into patterns of SNP distribution within populations is underway. The seven most informative SNP-containing loci (Table 1) have a combined DP of greater than 99.999999, comparable to any ten traditionally typed STRs. DP increased 3.5–5 percent per locus compared to nominal typing, with the greatest increase observed within the African American population. Pedigree evaluations demonstrated that SNP-containing alleles could be traced though several generations in some pedigrees. The presence of distinct alleles can increase the resultant Kinship Index (KI) when a putative missing person shares SNP-containing alleles with members of the pedigree thus reducing ambiguity resulting from individuals sharing alleles that are IBS.

4. Conclusion

By capitalizing on the presence of SNPs within the STRs, the information content and discrimination power (DP) can be increased for a number of forensically relevant STR loci. The most notable effect was an increase in heterozygosity, as some individuals who are typed as homozygous with conventional methods are actually same length heterozygotes. An assessment of the effects of population substructure is warranted due to the reapportionment of alleles though SNP-STR typing.

Due to the limitation of amplicon size effectively typed using ESI-MS, an expansion into "non-CODIS" miniSTR loci is worth pursuing. The European forensic community has discussed the inclusion of miniSTR loci into an expanded core STR panel [8], specifically D10S1248, D2S441, and D22S1045. Additional candidate loci have been described [9] that are being evaluated for potential SNP distributions and efficacy in forensic casework.

Role of funding

The UNT Health Science Center received NIJ grant (Nos. 2007-DN-BX-K200 and 2008-DN-BX-K157) and Ibis Biosciences Inc. received NIJ grant (Nos. 2006-DN-BX-K011 and 2008-DN-BX-K304).

The NIJ had no direct role in experimental design, data interpretation or publication.

Conflict of interest

Salaried employees of Ibis Biosciences Inc., a Subsidiary of Abbott Molecular Inc. with financial interest in the development and marketing of products and services related to human forensics.

Acknowledgements

The following individuals contributed to this project: Suzanne D. Gonzalez, Helen Costa, Patricia Gibson, Jinhua Yang, Dixie Peters, Barbara Frankovich, Sarah Schultheis, Rhonda K. Roby, Preston Burnley, Amy Smuts, Jennifer Thomas.

References

[1] D.C. Muddiman, G.A. Anderson, S.A. Hofstadler, R.D. Smith, Length and base composition of PCR-amplified nucleic acids using mass measurements from electrospray ionization mass spectrometry, Anal. Chem. 69 (1997) 1543–1549.

[2] S.A. Hofstadler, R. Sampath, L.B. Blyn, M.W. Eshoo, T.A. Hall, Y. Jiang, J.J. Drader, J.C. Hannis, K.A. Sannes-Lowery, L.L. Cummins, B. Libby, D.J. Walcott, A. Schink, C. Massire, R. Ranken, N. White, V. Samant, J.A. McNeil, D. Knize, D. Robbins, K. Rudnik, A. Desai, E. Moradi, D.J. Ecker, TIGER: the universal biosensor, Int. J. Mass Spectrom. 242 (2005) 23–41.

[3] H. Oberacher, F. Pitterl, G. Huber, H. Niederstatter, M. Steinlechner, W. Parson, Increased forensic efficiency of DNA fingerprints through simultaneous resolution of length and nucleotide variability by high-performance mass spectrometry, Hum. Mutat. 29 (2008) 427–432.

[4] T.A. Hall, K.A. Sannes-Lowery, L.D. McCurdy, C. Fisher, T. Anderson, A. Henthorne, B. Budowle, S.A. Hofstadler, Base composition profiling of human mitochondrial DNA using PCR and direct automated electrospray ionization mass spectrometry, Anal. Chem. 81 (2009) 7515–7526.

[5] S. Guo, E. Thompson, Performing the exact test of Hardy-Weinberg proportion for multiple alleles, Biometrics 48 (1992) 361–372.

[6] R.C. Lewontin, J. Felsenstein, The robustness of homogeneity tests in $2 \times N$ tables, Biometrics 21 (1965) 19–33.

[7] B. Budowle, T.R. Moretti, A.L. Baumstark, D.A. Defenbaugh, K.M. Keys, Population data on the thirteen CODIS core short tandem repeat loci in African Americans, U.S. Caucasians, Hispanics, Bahamians, Jamaicans, and Trinidadians, J. Forensic Sci. 44 (1999) 1277–1286.

[8] P. Gill, L. Fereday, N. Morling, P.M. Schneider, The evolution of DNA databases—recommendations for new European STR loci, Forensic Sci. Int. 156 (2006) 242–244.

[9] 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.