Available online at www.sciencedirect.com
SCIENCE DIRECT®
Forensic Science International 146S (2004) S147–S149
Forensic Science International
www.elsevier.com/locate/forsciint

Short communication

Toward increased utility of mtDNA in forensic identifications

Rebecca S. Just[a], Jodi A. Irwin[a], Jennifer E. O'Callaghan[a], Jessica L. Saunier[a], Michael D. Coble[a,b], Peter M. Vallone[b], John M. Butler[b], Suzanne M. Barritt[a], Thomas J. Parsons[a,*]

[a]*Armed Forces DNA Identification Laboratory, Armed Forces Institute of Pathology, AFIP Annex, 1413 Research Blvd., Rockville, MD 20850, USA*
[b]*Biotechnology Division, National Institute of Standards and Technology, 100 Bureau Drive, Mail Stop 8311, Gaithersburg, MD 20899-8311, USA*

Available online 18 October 2004

Abstract

The utility of mtDNA in forensic identifications is limited by its low power of discrimination and the absence of high quality mtDNA databases. Single nucleotide polymorphisms (SNPs) in the control region outside of hypervariable regions I and II (HVI/HVII), and in the coding region of the mtDNA genome, can provide additional discrimination in mtDNA testing. We have identified particularly useful SNP sites via high throughput sequencing of the entire mtDNA genome. We report here two cases in which an 11-plex SNP assay (panel "A") targeting the most common HVI/HVII type successfully resolved two cases in which identifications could not be made on the basis of HVI/HVII sequencing. Additionally, we established a database of 286 samples for SNP panel "A" generated with robotic protocols. We have addressed the need for high quality mtDNA control region (CR) databases by developing robotic protocols for lab processing, and a carefully devised electronic data review process. A large-scale databasing effort targeting several populations underrepresented in current mtDNA databases is underway.
Published by Elsevier Ireland Ltd.

Keywords: mtDNA; SNP; Allele-specific primer extension; Control region; Database

1. Introduction

The Armed Forces DNA Identification Laboratory (AFDIL), USA routinely uses mtDNA to assist in the identification of degraded skeletal remains, and mtDNA is being increasingly implemented in many laboratories worldwide. Here, we describe several ongoing projects at AFDIL designed to augment the utility of mtDNA testing in relation to significant, but correctable weaknesses: (1) the low power of discrimination obtained when common HVI/HVII types are encountered in casework, and (2) a lack of high quality population reference databases from many regions of the world.

Though the strengths of mtDNA testing are well characterized [1–3], this method has significant weaknesses which limit its usefulness in forensic casework. One such weakness, the inability to discriminate among individuals who match common HVI/HVII types, we have described previously in some detail [1,2]. To augment current HVI/HVII sequencing we sequenced 241 entire mtGenomes of 18 of the most common HVI/HVII types in West European populations and identified 59 SNPs appropriate for assay development. Taken together, these 59 SNPs greatly increase the discriminatory power of mtDNA among the most

* Corresponding author. Tel.: +1 301 319 0268; fax: +1 301 295 5932.
E-mail address: parsons@afip.osd.mil (T.J. Parsons).

0379-0738/$ – see front matter. Published by Elsevier Ireland Ltd.
doi:10.1016/j.forsciint.2004.09.045

common European mtDNA types. These 59 sites were organized into eight multiplex panels targeting 18 specific common HVI/HVII types [3].

2. SNP typing of degraded remains

We have optimized an allele-specific primer extension assay for one of these multiplex panels (panel "A") that targets the most common HVI/HVII type "H:1" [3,4]. This multiplex allele-specific primer extension assay has many characteristics that are well-suited to the testing of degraded casework extracts. The assay is multiplexed for 11 sites and thus suitable for cases where extract quantities are limited. The assay uses short amplicons and is therefore appropriate for degraded extracts. The assay is sensitive to the 1–2 pg range and has mixture and heteroplasmy detection exceeding what can be reliably detected in mtDNA sequencing [4].

In addition to the non-probative cases previously addressed [4], we have now used this SNP assay to assist in the resolution of two AFDIL cases that required mtDNA testing. In the first case, the HVI and HVII sequences of the case extracts matched the sequences for reference samples from two different families. Subsequent "variable region" (outside of HVI/HVII in the control region) sequencing identified a single site at which the families varied; however, AFDIL policy dictates that two differences are required to report an exclusion. The casework extracts and reference samples were typed for multiplex SNP panel A. The results confirmed the single difference previously observed and identified an additional site at which the families varied. In the second case, the HVI and HVII sequences of the case extracts matched one family, but varied from a second family at only a single site. Variable region sequencing did not identify any additional variable sites. Multiplex SNP panel A typing identified an additional site at which the families varied. In both cases, SNP typing of the skeletal remains extracts gave full results, permitting exclusion of the mismatching families. The HVI/HVII type for the first case was H:1, and the HVI/HVII type for the second case was H:1 plus an additional polymorphism. These results indicate that multiplex SNP panel A may be as useful for near-H:1 types as it is for HVI/HVII type H:1.

3. SNP databasing

To establish SNP databases necessary for casework application we have developed high-throughput robotic protocols to database samples for multiplex SNP panel A. Extractions in 96-well format are performed on a Qiagen 9604, a four-probe instrument with disposable tips. Amplification set-up is performed on a Corbett CAS-1200, a fully-enclosed instrument with a single probe and disposable tips. Post-amplification reactions are pipetted on a Tecan Genesis BioRobot, an eight-probe instrument with integrated plate sealer and thermal cyclers. Detection is performed on an Applied Biosystems 3100, a capillary electrophoresis instrument with 16 probes. All reactions and thermal cycling are performed as described in Vallone et al. [4].

The data generated is reviewed through the use of a Genotyper (Applied Biosystems) macro designed for SNP multiplex panel A. The macro makes automatic allele calls for each multiplex site, and is formatted to recognize heteroplasmic and mixed samples as well as null alleles. Data review is performed by two independent analysts who subsequently export data tables directly from Genotyper. A third analyst reviews the data tables and resolves any discrepancies. The results of databasing for 286 unrelated Caucasian samples will be published at a later date.

4. Identifying SNPs for U.S. Hispanic and African American populations

Common HVI and HIVII types are problematic in U.S. Hispanic and African American populations as well, although to a lesser extent than in the Western European Caucasian population. To identify and obtain samples matching the common types that require greater resolution in these populations we have sequenced the entire control region for almost 1000 U.S. Hispanics and African Americans. We have developed robotic protocols for high-throughput entire mtGenome sequencing, and are conducting whole mtGenome sequencing of African American and U.S. Hispanic samples of common HVI and HVII types. The research plan is to identify discriminatory SNPs, design additional multiplex panels, and optimize primer extension assays for common HVI and HVII types for these populations.

5. Control region databasing

A second significant weakness of mtDNA testing is the current lack of high-quality databases for many world populations. Additional problems with the existing forensic CR databases include inconsistencies in nomenclature and, most significantly, errors in the databases. Several studies have identified widespread errors in published databases, most often caused by mistakes in the data analysis and errors in the transcription or tabulation of data [5–9].

To increase the size and range of current mtDNA databases, as well as decrease the potential sources of error in creating databases, the AFDIL Research Section has developed a high-throughput robotic system for population databasing. Robotic processing at every step from pre-extraction through sequence detection (using the same robotic instrumentation described above for SNP databasing) eliminates manual pipetting of samples and reagents, thus greatly reducing the potential for sample switching. Additionally, multiple scientists are used for redundant checking at key

laboratory steps when greater user intervention is required (manual re-amplification, for example), further decreasing the likelihood that errors will be made.

To maintain high standards for data quality and avoid transcriptional errors, a five-step data review process with entirely electronic data transfer was implemented. (1) Primary analysis: Raw sequence data is aligned and a list of polymorphisms generated. A minimum of two strands of data at every position is required for confirmation. (2) Secondary analysis: The alignment is reviewed, double-stranded coverage confirmed and an independent list of polymorphisms generated. The lists of polymorphisms are compared and any discrepancies resolved. (3) Initial review: A summary of polymorphisms is exported directly from the sequence alignment, and the data transferred electronically into a database. The imported list of polymorphisms is compared to the consensus list created in step 2. (4) Final review: The imported list of polymorphisms is independently compared to the consensus list generated in step 2. (5) The raw sequence data is electronically transferred to the EDNAP mtDNA population database (EMPOP). An EMPOP analyst independently aligns the raw sequence data, performs a pairwise comparison with the AFDIL data, and resolves any discrepancies.

This large-scale databasing project aims to sequence on the order of 5000 entire control region sequences per year, with a focus on under-represented populations from around the world. In the last 6 months, more than 2000 CR sequences have been completed. The data will be made publicly available at a later date through publications and submission to GenBank and EMPOP.

Official disclaimer

Certain commercial equipment, instruments, and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the United States Department of Defense, the United States Department of the Army, or the National Institute of Standards and Technology nor does it imply that any of the materials, instruments or equipment identified are necessarily the best available for the purpose. 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 United States Department of Defense, the United States Department of the Army, or the U.S. Department of Commerce.

Acknowledgement

This work was supported by a National Institute of Justice grant 2000-1J-CX-K010 to T.J.P.

References

[1] M.M. Holland, T.J. Parsons, Mitochondrial DNA sequence analysis—validation and use for forensic casework, Forensic Sci. Rev. 11 (1999) 22–50.
[2] T.J. Parsons, M.D. Coble, Increasing the forensic discrimination of mitochondrial DNA testing through analysis of the entire mitochondrial DNA genome, Croat. Med. J. 42 (2001) 304–309.
[3] M.D. Coble, R.S. Just, J.E. O'Callaghan, I.H. Letmanyi, C.T. Peterson, J.A. Irwin, T.J. Parsons, Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in Caucasians, Int. J. Legal Med. 118 (2004) 137–146.
[4] P.M. Vallone, R.S. Just, M.D. Coble, J.M. Butler, T.J. Parsons, A multiplex allele-specific primer extension assay for forensically informative SNPs distributed throughout the mitochondrial genome, Int. J. Legal Med. 118 (2004) 147–157.
[5] H.-J. Bandelt, P. Lahermo, M. Richards, V. Macaulay, Detecting errors in mtDNA data by phylogenetic analysis, Int. J. Legal Med. 115 (2001) 64–69.
[6] H.-J. Bandelt, L. Quintana-Murci, A. Salas, V. Macaulay, The fingerprint of phantom mutations in mtDNA data, Am. J. Hum. Genet. 71 (2002) 1150–1160.
[7] P. Forster, To err is human, Ann. Hum. Genet. 67 (2003) 2–4.
[8] Y.-G. Yao, C.M. Bravi, H.-J. Bandelt, A call for mtDNA data quality control in forensic science, Forensic Sci. Int. 141 (2004) 1–6.
[9] H.-J. Bandelt, A Salas, S. Lutz-Bonengel, Artificial recombination in forensic mtDNA population databases, Int. J. Legal Med. (2004) [Epub ahead of print].