

Technical note

Forensic application of the affymetrix human mitochondrial resequencing array[☆]

P.M. Vallone^{a,*}, J.P. Jakupciak^a, M.D. Coble^b

^a National Institute of Standards and Technology (NIST), Biochemical Sciences Division, 100 Bureau Dr., M/S 8311, Gaithersburg, MD 20899-8311, United States

^b Armed Forces DNA Identification Laboratory, 1413 Research Boulevard, Building 101, Rockville, MD 20850, United States

Received 22 January 2007; accepted 27 January 2007

Abstract

In the field of forensic DNA testing, sequencing regions of the mitochondrial genome is performed when insufficient genomic DNA is present for traditional autosomal short tandem repeat (STR) testing. Sequencing coding region polymorphisms in the mitochondrial genome can be useful for resolving individuals who have the identical HV1 and HV2 control region sequence. Various methods and strategies have been established to interrogate coding region polymorphisms. These range from SNP assays probing sites most likely to differentiate individuals based on their HV1/HV2 sequence to the use of mass spectrometry to pyrosequencing. Here we evaluate the potential of the Affymetrix GeneChip Mitochondrial Resequencing Array (version 2.0) for forensic applications.

© 2007 Elsevier Ireland Ltd. All rights reserved.

Keywords: Mitochondrial; Microarray; Sequencing; NIST

1. Introduction

The increasing availability of commercial DNA array platforms may be an attractive option for forensic analyses. One such platform is the Affymetrix GeneChip Mitochondrial Resequencing Array version 2.0. The GeneChip platform allows for sequencing of the entire mitochondrial genome by hybridization. The mitochondrial genome is amplified in three separate reactions, fragmented, and hybridized onto the array. Full genome sequencing results can typically be obtained in about 2 days.

It is of importance to understand the performance of the array. Questions to be asked include: the accuracy of the base

calls, effects of the base calling algorithm, and the reproducibility of the chip experiments. Since dideoxy fluorescent sequencing of the entire mitochondrial genome can be routinely performed, this provides us with a context to evaluate the GeneChip performance [1].

Since the tiling of the array utilized sequence information from the revised Cambridge Reference Sequence [2,3], derived from an individual of Western European ancestry, we tested the array's ability to successfully call a sample having a relatively large number of sequence differences compared to the rCRS. Reproducibility of GeneChip experiments was evaluated by running the sample in triplicate. Results were compared to traditional dideoxy fluorescent full genome sequencing experiments.

2. Materials and methods

2.1. GeneChip experiments

Samples were prepared and amplified as directed in the Affymetrix protocol guide (<http://www.affymetrix.com/support/technical/manuals.affx>). The full mitochondrial genome was amplified in three separate singleplex reactions [4]. Approximately 25 ng of DNA template (genomic quantity) was used in

[☆] Contribution of the U.S. National Institute of Standards and Technology. Not subject to copyright. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice. 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 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.

* Corresponding author. Tel.: +1 301 975 4872; fax: +1 301 975 8505.

E-mail address: peter.vallone@nist.gov (P.M. Vallone).

each amplification reaction. Aliquots from the same PCR pool were run on the array in triplicate to determine reproducibility. Array data was analyzed and base calls were made using the Affymetrix GSEQ software package version 4.0. A composite sequence was imported into Sequencher (GeneCodes, Ann Arbor, MI) software version 4.1.4Fb19 for comparison to the rCRS. Data analysis in GSEQ software was performed at three 'Score' levels. The 'Score' level is a base calling GSEQ parameter that determines the stringency of the base calling algorithm. A Score of 12 results in conservative base calling (more ambiguous N base calls) while a Score of 1 results are the most liberal (less N calls, but potential miscalls). Base calls were made with Scores of 1, 6 and 12 for comparisons.

2.2. African American sample AA01

A challenging sample was selected from our NIST U.S. population samples: African American (designated AA01). This sample contained 63 differences from the rCRS as determined from previous dideoxy fluorescent sequencing experiments (AFDIL, unpublished data). A total of 14 base differences were located in the control region and 49 in the coding region of mitochondrial genome.

3. Results and discussion

3.1. Evaluation of GeneChip array performance on sample AA01

Table 1 is a summary of the missed, incorrect, or inconsistent base calls for the three replicates evaluated at three Score values. As indicated at the bottom of Table 1 as the Score increases from 1 to 12 the number of ambiguous N calls

increased from less than to 200 to over 900 (1–6% of the mitochondrial genome). The majority of the N calls are due to poly C stretches (four or more) located throughout the mitochondrial genome.

Four control region sites are listed in Table 1. Position 146 is a T–C polymorphism. The change from T to C at this position results in a poly C stretch (four bases) that is not called using a Score of 12. At the lower score values 146 T–C was correctly called twice. The C insertion at 315.1 was not called by the software, nor was a AC deletion at 523–524.

Seven coding region sites are also listed in Table 1. The T to C call at sites 2416 and 10,873 results in a poly C stretch which the array has problems calling (similar to 146 above). Sites 11,719 and 11,914 are incorrectly called at lower Score values due to overlap with the PCR primers used for amplifying the mitochondrial genome. At 13,650 is an N call until the Score is lowered to 1 then the correct call is made.

Sample AA01 contained two instances of point heteroplasmy detected by traditional fluorescent sequencing (at 1709R and 15978Y), but were not called in the GSEQ software. Further analysis of the array cell intensity suggests a 1:1 signal ratio at those sites (data not shown), but each point heteroplasmy would not have been identified without the fluorescent sequencing data.

It should be noted that we have only focused on the 63 differences from the rCRS and not fully investigated the details of the 200–900 ambiguous N calls that were called on the array.

3.2. Summary

The GeneChip Mitochondrial Resequencing Array is a means to perform full genome sequencing on an array-based platform. The amount of DNA needed for the array is

Table 1
Summary of discordant base calls for AA01

position	Score Replicate	1			6			12		
		1	2	3	1	2	3	1	2	3
146	T-C	C	C	N	C	N	C	N	N	N
315.1	C-ins.	missed								
523	A-del.	missed	missed	missed	missed	missed	missed	N	N	N
524	C-del.	missed								
1709	G-R	G	G	G	G	G	G	G	G	G
2416	T-C	N	N	N	N	N	N	N	N	N
10873	T-C	N	C	C	N	N	N	N	N	N
11719	G-A	G	G	G	N	N	N	N	N	N
11914	G-A	G	G	G	G	G	G	N	N	N
13650	C-T	T	T	T	N	N	N	N	N	N
15978	C-Y	C	C	C	N	N	N	N	N	N
	Total N Calls	176	184	176	542	572	557	926	996	983

Shaded cells indicate a missed or incorrect base call.

comparable to that required by fluorescent methods for sequencing the entire mitochondrial genome. However, it should be noted that the array does not provide full sequence coverage (as indicated by the number N calls in Table 1). Because of this relatively high sample requirement the array may have restrictions for running a limited quantity of casework sample. However the platform should have utility in running family reference samples for the elucidation of SNPs in the coding region that will help resolve individuals. These array-determined polymorphisms found in reference sample can then be probed in the limited casework sample extract [5,6]. Comparisons between the GeneChip and traditional sequencing indicated the array platform had difficulty calling instances of point heteroplasmy, insertions, deletions, poly C regions as well as some closely spaced polymorphisms.

Acknowledgement

This work was funded in part by the National Institute of Justice through interagency agreement 2003-IJ-R-029 with the NIST Office of Law Enforcement Standards.

References

- [1] 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.
- [2] S. Anderson, A.T. Bankier, B.G. Barrell, M.H. de Bruijn, A.R. Coulson, J. Drouin, I.C. Eperon, D.P. Nierlich, B.A. Roe, F. Sanger, P.H. Schreier, A.J. Smith, R. Staden, I.G. Young, Sequence and organization of the human mitochondrial genome, *Nature* 290 (1981) 457–465.
- [3] R.M. Andrews, I. Kubacka, P.F. Chinnery, R.N. Lightowlers, D.M. Turnbull, N. Howell, Reanalysis and revision of the Cambridge Reference Sequence for human mitochondrial DNA, *Nat. Genet.* 23 (1999) 147.
- [4] A. Maitra, Y. Cohen, S.E. Gillespie, E. Mambo, N. Fukushima, M.O. Hoque, N. Shah, M. Goggins, J. Califano, D. Sidransky, A. Chakravarti, The Human MitoChip: a high-throughput sequencing microarray for mitochondrial mutation detection, *Genome Res.* 14 (2004) 812–819.
- [5] 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.
- [6] B. Budowle, J.V. Planz, S. Campbell, A.J. Eisenberg, Single nucleotide polymorphisms and microarray technology in forensic genetics—development and application to mitochondrial DNA, *Forensic Sci. Rev.* 16 (2004) 21–36.