Forensic Science International: Genetics 7 (2013) 1–9

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Comparative analysis of human mitochondrial DNA from World War I bone samples by DNA sequencing and ESI-TOF mass spectrometry

Rebecca Howard [a,*], Vesela Encheva [a], Jim Thomson [a], Katherine Bache [b], Yuen-Ting Chan [b], Simon Cowen [a], Paul Debenham [a], Alan Dixon [b], Jens-Uwe Krause [c], Elaina Krishan [b], Daniel Moore [b], Victoria Moore [b], Michael Ojo [b], Sid Rodrigues [b], Peter Stokes [a], James Walker [b], Wolfgang Zimmermann [c], Rita Barallon [a]

[a] LGC Limited, Queens Road, Teddington, Middlesex TW11 0LY, UK
[b] LGC Forensics, Queens Road, Teddington, Middlesex, TW11 0LY, UK
[c] LGC Genomics GmbH, Ostendstrasse 25, TGS Haus 8, 12459 Berlin, Germany

ARTICLE INFO

Article history:
Received 25 November 2010
Received in revised form 4 May 2011
Accepted 21 May 2011

Keywords:
Mitochondrial DNA
Forensic
Mass Spectrometry
DVI
Skeletal
Teeth

ABSTRACT

Mitochondrial DNA is commonly used in identity testing for the analysis of old or degraded samples or to give evidence of familial links. The Abbott T5000 mass spectrometry platform provides an alternative to the more commonly used Sanger sequencing for the analysis of human mitochondrial DNA. The robustness of the T5000 system has previously been demonstrated using DNA extracted from volunteer buccal swabs but the system has not been tested using more challenging sample types. For mass spectrometry to be considered as a valid alternative to Sanger sequencing it must also be demonstrated to be suitable for use with more limiting sample types such as old teeth, bone fragments, and hair shafts. In 2009 the Commonwealth War Graves Commission launched a project to identify the remains of 250 World War I soldiers discovered in a mass grave in Fromelles, France. This study characterises the performance of both Sanger sequencing and the T5000 platform for the analysis of the mitochondrial DNA extracted from 225 of these remains, both in terms of the ability to amplify and characterise DNA regions of interest and the relative information content and ease-of-use associated with each method.

1. Introduction

The use of mitochondrial DNA in forensic science to identify potential familial relationships is well-established. The mitochondrial genome control region, or D-loop, contains two polymorphic regions known as HyperVariable regions 1 and 2 (HV1 (bases 16024–16365) and HV2 (bases 73–340)). [1,2]. These regions contain a large number of known polymorphisms, inherited as stable haplotypes, that can be used as a basis for familial matching. Mitochondria of individuals with different geographic origins also have patterns of variation so population databases, *e.g.* the European Mitochondrial POPulation database (EMPOP) [3] have been compiled to provide information on the frequencies of haplotypes in various population groups [4].

The Sanger sequencing method [5] is commonly used for mitochondrial DNA analysis since it is able to provide information on the position and base change of every polymorphic position within the analysed region when compared with a reference standard, known as the revised Cambridge Reference Sequence (rCRS) [Genbank NC_012920]. DNA sequencing is also able to provide indications of heteroplasmy in some cases.

Despite being commonly used in the scientific community there are a number of issues with current DNA sequencing methods. The use of automation has substantially reduced the amount of hands-on time required both for sample preparation and subsequent sequence analysis but automated base-calling can be prone to errors, particularly in the detection of unbalanced heteroplasmies and mononucleotide repeats. Manual checking of automated results can therefore be time-consuming and labour-intensive. Nonetheless, precise sequence information is currently required for the interrogation of most available databases, *e.g.* EMPOP, which renders other analysis methods less attractive.

The Abbott T5000 system utilises an Electrospray Ionisation Time-Of-Flight mass spectrometer (ESI TOF MS) to characterise human mitochondrial DNA. The system consists of an automated DNA clean-up station, an Electrospray Ionisation Time-Of-Flight mass spectrometer (ESI TOF MS) and the incorporated IBIS Track software for data analysis. It has been demonstrated that ESI TOF MS has the potential as an analytical tool for the characterisation of nucleic acids such as oligonucleotides and PCR products, *e.g.* [6–9].

* Corresponding author. Tel.: +44 02089437540.
E-mail address: rebecca.howard@lgc.co.uk (R. Howard).

1872-4973/$ – see front matter © 2011 LGC Limited. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.fsigen.2011.05.009

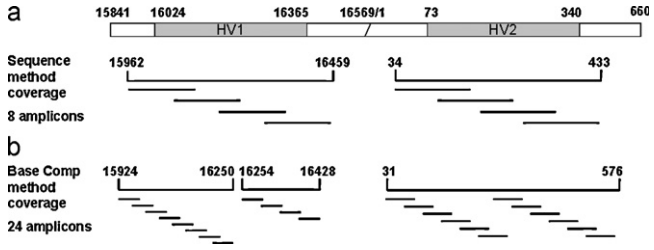

Fig. 1. Coverage map to show regions covered by each analysis method aligned to HV1 and HV2 in the D-loop control region of the mitochondrion. Not to scale.

The T5000 system takes advantage of the fact that the accurate masses of each of the four bases that constitute DNA are known. Therefore, for each measured amplicon mass, there are a limited number of possible base compositions [10]. This is additionally constrained by taking into account the rules of base complementarity and the resultant correlation between complementary DNA strands. Using unmodified nucleotides erroneous assignments could be made when two products have very similar masses, *e.g.* in cases where a double SNPs G → A and C → T result in only a 1 Da mass difference. To avoid this difficulty "heavy" ^{13}C-dGTP is used which introduces mass differences of 10 Da or more and allows the discrimination of such potentially ambiguous products. The T5000 approach (which will be referred to during comparisons as the 'base composition method') makes use of twenty-four overlapping PCR products of approximately 80–100 bp in length spanning the majority of sequence covering the HV1 and HV2 regions (Fig. 1). Note that a three base pair region of HV1 (positions 16251–16253) is not covered by the assay as it falls between two adjacent amplicons. Following amplification in triplex PCRs, the masses of both strands of each amplicon are determined by mass spectrometry and a base composition is assigned to pairs of complementary strands. Base compositions are then compared *in silico* to those predicted from the rCRS. A minimum number of differences is calculated for each amplicon and for the sample as a whole.

Although base compositions cannot be used to search sequencing databases directly the T5000 system avoids this problem by providing its own database permitting conversion of sequence data into base compositions for profile searching. This provides the facility to convert any known DNA sequence, or group of sequences, spanning the HV1 and HV2 loci into base composition thus permitting the interrogation of existing databases.

The T5000 system has previously been validated by the FBI using DNA extracted from fresh buccal swabs [6,7]. While buccal swabs will no doubt constitute a proportion of samples submitted for mitochondrial DNA analysis a large proportion of mitochondrial analysis in forensic casework is of old or degraded samples. The T5000 system must therefore also be capable of analysis of these sample types if it is considered to be a realistic alternative to the Sanger method.

Direct comparison between two methods is difficult, since one method is invariably favoured over the other. For example, there is a fivefold difference in recommended input DNA volume (5 μl in aT5000 reaction vs. 1 μl in a Sanger sequencing reaction) which may favour successful amplification in the T5000 PCRs. The T5000 conditions also use one additional cycle of PCR and the assay targets shorter amplicons. This is partially offset by the fact that amplification of T5000 products takes place in less efficient triplex reactions, while the Sanger sequencing is carried out in singleplex reactions. What can be compared, however, is the results obtained from each method when used under recommended conditions. It is this comparison that is the subject of the study described here.

The opportunity to test the two systems in parallel on appropriate sample types was presented during a Commonwealth War Graves Commission project to identify the remains of 250 soldiers buried during the First World War. The T5000 was tested under manufacturer-recommended conditions for the purpose of this comparison.

On 19th July 1916 the Australian Army's 5th Division and the British Army's 61st Division engaged German troops near Fromelles in northern France. By the time the battle, now known as the Battle of Fromelles, was over it is believed that 1708 Australian and 503 British soldiers had been killed. Many of these soldiers could not be accounted for in the aftermath of the conflict. In 2007, work carried out by the Glasgow University Archaeology Research Division (GUARD) led to the discovery of 250 sets of remains in a mass grave at Pheasant Wood, an area just outside Fromelles. In 2009, the Commonwealth War Graves Commission announced a project to identify as many sets of these remains as possible and, in 2010, the remains were reburied in a designated military cemetery. The last of these was interred at a commemorative ceremony on 19th July 2010, the 94th anniversary of the battle. Headstones on at least ninety-six of the graves carry the names of Australian soldiers identified as a result of the Commission-sponsored project (Commonwealth War Graves Commission, http://www.cwgc.org/fromelles/).

Mitochondrial DNA analysis of the remains was carried out in the forensic laboratories at LGC. Of the 250 sets of remains discovered by the archaeological team directly comparable DNA data was available from both the Sanger sequencing and base composition methods for 225. These 225 samples were used as an example dataset. This paper will evaluate both methods using several criteria including coverage within regions of interest, quality of information available and the discriminatory power when profiles generated by each method are searched against a standard sequence database of potential relatives.

2. Materials and methods

All reagents were obtained from Abbott Molecular unless otherwise specified.

2.1. Bone decalcification and DNA extraction

Individual teeth or bone fragments (0.9–1.5 g) were ground to a fine powder and resuspended in 16 ml 0.5 M EDTA to decalcify. The suspension was mixed by inversion and then on a Luckham R100 Rotatest Shaker for 24 h. The suspension was then centrifuged and the supernatant was then aspirated from the pellet. This process was repeated 5–6 times, before a final centrifugation step to pellet the decalcified material to be used for DNA extraction.

The ground and decalcified tooth or bone sample was mixed with 15 ml ATL buffer and 750 μl proteinase K (Qiagen, UK) and incubated at 56 °C for a minimum of 12 h. The samples were centrifuged and 15 ml AL buffer added (Qiagen, UK). This was vortexed gently until all powder was suspended and incubated at 70 °C for 1 h. Samples were then washed with absolute ethanol and purified on a Qiagen Maxi Column (Qiagen, UK). The DNA was eluted in 3 ml Qiagen AE buffer prior to downstream analysis.

2.2. DNA yield

Extracted DNA was concentrated using Amicon columns (Millipore, Billerica, MA). Prior to amplification, human genomic DNA was quantified using the Quantifiler Human DNA Quantification kit (Life Technologies, Carlsbad, CA) under manufacturer-recommended conditions.