International Congress Series 1261 (2004) 30–32

www.ics-elsevier.com

Population studies using single nucleotide polymorphisms—how important is detailed sample origin information?

C. Thacker[a,*], C. Harrison[a], M. Brion[b], C. Phillips[b], P. Sanchez-Diz[b], D. Ballard[a], A. Carracedo[b], D. Syndercombe Court[a]

[a] *Haematology Department, Barts and The London, Queen Mary's School of Medicine and Dentistry, Turner Street, London E1 2AD, UK*

[b] *Institute of Legal Medicine, Faculty of Medicine, University of Santiago de Compostela, Galicia, Spain*

Abstract. Within the UK there is an increase in the request for information about an individual's ancestory. UK population databases are not uncommon but the detailed origin of samples has not traditionally been collected. Typically samples are classified using very general identifiers such as 'black British'. We have conducted a study to establish whether this lack of detailed information proves to be a major problem when considering population studies relying on SNP data. A panel of Y-chromosome and autosomal SNPs have been used to genotype individuals from a 'black British' population and from six Mozambique population groups. The resulting population data sets were contrasted and differences between a well-defined and a broadly characterized population set were used to illustrate the importance of good sample information. A comparison of the different SNP typing methodologies used throughout this study was conducted and all methods were found to work well.

Keywords: Single nucleotide polymorphism; SNP typing methodologies; Sample information

1. Introduction

An early study used the UK national DNA database to infer the ethnic origin of individuals [1]. Since then, marked population differences in allelic distributions of Y chromosome STRs have been exploited to predict ethnic origin, although the relatively high mutation rates necessitate proceeding with care. With mutation rates of only 2.5×10^{-8} per base per generation [2], SNPs look a more promising marker for this purpose. The study of SNPs is increasing rapidly and consequently there is a large amount of population data available. The majority of studies have concentrated on well-defined populations and as the

* Corresponding author. Tel.: +44-2073-777076; fax: +44-2073-777629.
E-mail address: c.r.thacker@qmul.ac.uk (C. Thacker).

0531-5131/
doi:10.1016/j.ics.2003.12.090

significance of forensic analysis often depends upon the population frequency of a finding [3], the question arises as to how valuable this resource is to the UK population.

The possible uses of SNPs are well published but the selection of an appropriate detection platform is something that needs careful consideration.

2. Materials and methods

2.1. DNA samples and DNA extraction

Two databases composed of unrelated, consenting individuals were used for this study. The first database was broadly defined as 'black British'. It was composed of individuals resident within the UK who were identified as belonging to this racial group either by photograph or personal description. The second database was a well-characterized population and was composed of samples donated by indigenous Bantu speakers of Mozambique. The geographical origin of each individual's four grandparents was known. DNA was extracted using Chelex® 100 chelating resin [4].

2.2. SNP selection and primer design

SNPs were not selected for the populations studied. Ten Y-chromosome SNPs and three autosomal SNPs were used. Primers were designed using Primer Express™ 1.5 software (Applied Biosystems). Design constraints were applied to enable the same primers to be used for all the SNP typing methodologies utilised in this study. All primers were HPLC-purified.

2.3. DNA amplification, minisequencing and silver staining

The extracted DNA (10 μl) was amplified in a total PCR reaction volume of 50 μl. For each reaction, 8 μl of PCR product was subjected to a restriction digest. Digested products were run on 9% polyacrylamide gels and visualised following silver staining [5]. The remaining PCR products were pooled and cleaned-up using either ExoSAP-IT® (USB) or MinElute PCR purification columns (Qiagen). Minisequencing was performed using SNaPshot™ (Applied Biosystems) and this was followed by SAP purification (Roche Diagnostics). The purified minisequencing products were analysed using the ABI Prism™ 310 Genetic Analyser (Applied Biosystems).

2.4. TaqMan® detection

TaqMan® detection was performed using the ABI Prism™ 7700 Sequence Detector (Applied Biosystems). As many of the sequences were AT- or GC-rich, MGB probes (Applied Biosystems) were designed for the real-time analysis. Chelex® extract (7 μl) was used in a 25 μl total reaction volume.

3. Results

All technologies worked well however some discrepancies were noted between the manual and SNaPshot™ results. There was no difference in the quality of purification obtained using either ExoSAP-IT® or the MinElute PCR purification columns.

There were noticeable differences between the two populations studied. The well-defined Mozambique samples showed very little intra-group variation. In contrast, the UK black population exhibited a substantial amount of variation between samples. M40 can help define the largely African haplogroup E [6]. The G to A transition characterizes this SNP with all the Mozambique samples exhibiting an A allele. Almost 40% of the UK population tested possessed the G allele. P25 can be used to help define haplogroup R, a European haplogroup [6]. No polymorphism was seen in the Mozambique population (all exhibited the ancestral state, a C allele) whereas 15.4% of the UK black population showed P25 in the derived state (an A allele).

4. Discussion

4.1. Typing techniques

Smaller amplicons were sometimes difficult to accurately size using manual techniques. The discrepancies seen between this method and SNaPshot™ were attributed to this interpretation difficulty. Financial restraints, access to instrumentation and availability of consumables and reagents means that some laboratories have a very limited choice when determining their method of SNP detection. This study has shown that this need not prohibit these laboratories making a significant contribution to the progression of SNP research.

4.2. The black British population

The black British population shows a high level of admixture and as a consequence requires a great deal of caution when assigning a sample to an ethnic group. The UK has strong links with the African Commonwealth countries and its black population will have been influenced in part by the slave trades of the Caribbean and, to a lesser extent, the Atlantic. Although it has been shown that haplotypes can be used to provide useful information on ancestry, it is not possible to use them to reliably predict how an individual from the UK black population would describe themselves. This study has shown that detailed sample information is essential if compiling a population database to be used as a reference source. In the case of populations exhibiting a high level of diversity, published data is not always applicable.

References

[1] A.L. Lowe, et al., Inferring ethnic origin by means of an STR profile, Forensic Science International 119 (2001) 17–22.
[2] R. Thomson, et al., Recent common ancestry of human Y-chromosomes: evidence from DNA sequence data, Proceedings of the National Academy of Sciences of the United States of America 97 (13) (2000) 7360–7365.
[3] M.A. Jobling, C. Tyler-Smith, The human Y-chromosome: an evolutionary marker comes of age, Nature Reviews. Genetics 4 (2003) 598–612.
[4] P.S. Walsh, et al., Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material, BioTechniques 10 (4) (1991) 506–513.
[5] C.P. Phillips, et al., Manual electrophoretic methods for genotyping amplified STR loci, in: P.J. Lincoln, J. Thomson (Eds.), Methods in Molecular Biology, Forensic DNA Profiling Protocols, vol. 98, Humana Press, Totowa, NY, 1998, pp. 181–192.
[6] Y Chromosome Consortium, A nomenclature system for the tree of human Y-chromosomal binary haplogroups, Genome Research 12 (2002) 339–348.