

The extent of substructure in the indigenous Australian population and its impact on DNA evidence interpretation

Simon J. Walsh ^{a,*}, R. John Mitchell ^b,
James M. Curran ^c, John S. Buckleton ^d

^a *Centre for Forensic Science, University of Technology Sydney, Australia*

^b *Department of Genetics, LaTrobe University, Melbourne, Australia*

^c *Department of Statistics, University of Auckland, New Zealand*

^d *ESR, Auckland, New Zealand*

Abstract. In order to estimate DNA match statistics it is necessary to understand population genetic features of the major populations of the jurisdiction. Here we summarise some findings of a large-scale research project recently undertaken on autosomal microsatellite data from indigenous Australian donors. This research follows from court decisions that scrutinised the level of understanding in this area. © 2005 Published by Elsevier B.V.

Keywords: DNA profiling; Autosomal STR; Australian Aborigines; Fst; Sub-population diversity

1. Introduction

Indigenous Australians have a unique evolutionary history resulting in a complex system of inter- and intra-tribal relationships. Although European colonisation has disrupted to a varying extent these and other features of Aboriginal life, forensic DNA evidence has recently been questioned with respect to the impact that this evolutionary past may have had on their population structure and thereby, the estimation of DNA match statistics [*R v Bropho* [2004] WADC 182 (unreported)].

The extent of substructure within the indigenous Australian sub-population raises two main questions relevant to the interpretation process: (1) what is the appropriate value of

* Corresponding author. PO Box 123, Broadway, NSW, 2007, Australia. Tel.: +61 2 9514 4064; fax: +61 2 9514 4026.

E-mail address: simon.walsh@uts.edu.au (S.J. Walsh).

the coancestry coefficient, θ or F_{st} ?, and (2) what is the effect of substructure on the performance of the sub-population model? This paper summarises research that focuses on these issues. Research examining classical markers as well as DNA SNPs has shown considerable heterogeneity exists within the indigenous Australian population. The significant question is, to what extent do autosomal microsatellites used in contemporary forensic testing show similar structuring effects?

2. Population data

Autosomal STR profiles from 6312 indigenous Australian and 28,821 non-indigenous donors were collated. The non-indigenous groups included Australian and New Zealand Caucasians, East Timorese, New Zealand Maori and Pacific Islanders. The samples had been analysed using either the nine-locus AMPFISTR® Profiler Plus™, the 15-locus AMPFISTR® Identifiler™, the 10-locus SGM Plus™ (Applied Biosystems, Foster City, CA) or the GenePrint CTT (Promega, Madison, WI) multiplex PCR systems. Ethnicity was by self-declaration.

With respect to the indigenous Australian dataset, many data were associated with a geographic placeholder assigned by either the location of the offence or the place of residence indicated by the donor. Following reference to the language map of Aboriginal Australia developed by Horton [1], contemporary placeholders were assigned to a putative region (regional populations=14) and tribal territory (tribal populations=65). In this way the data were segregated along traditional rather than contemporary lines.

3. F_{st} estimation for aboriginal Australian sub-population

Neighbour-joining (NJ) trees were developed using the Genetic Data Analysis (GDA) software package [2]. Trees were constructed from θ estimates for the full set of populations ($N_{pops}=83$; $N=35,133$) and all the remote Aboriginal groups minus the non-Aboriginal groups and indigenous groups from the urban centres ($N_{pops}=65$; $N=6312$).

Estimates for the parameters f (the inbreeding parameter within populations), F (the total inbreeding coefficient), and θ , were obtained using the method of Weir and Cockerham [3]. In contrast to many previous studies on indigenous Australian populations involving classical or non-autosomal markers, but in agreement with Ayres et al. [4], we find that the autosomal STR θ values are low. Our analysis supports the use of a θ value of 0.03 in the match probability formulae of Balding and Nichols [5] as being sufficiently conservative for estimates calculated from indigenous Australian data.

4. Reliability of the sub-population model

As a component of the recent conjecture, the performance of the sub-population model has been questioned along the lines of: (1) what would occur if the estimate used for θ in the match probability formulae differed from the true parameter in the population? (2) what would happen if the sub-population itself were subdivided?

We have assessed these questions directly by simulation, extending upon the methods of Curran et al. [6]. The full results are given in Buckleton et al. [7]. The match probability was estimated probability and compared to the true match probability. The outcome of

Table 1
Results of simulations to test the performance of the sub-population model

Simulation 1					
θ_{Breed}	θ_{Calc}	Conservative estimates (%)		Non-conservative estimates (%)	
0.03	0.03	99.3		0.7	
0.03	0.05	91.4		8.6	
0.03	0.07	75.8		24.2	
0.03	0.09	58.1		41.9	
Simulation 2					
θ_X	θ_Z	θ_{Calc}	Conservative estimates (%)		Non-conservative estimates (%)
0.03	0.01	0.03	99.3		0.7
0.03	0.03	0.03	99.3		0.7
0.05	0.07	0.03	99.1		0.9
0.03	0.09	0.03	99.1		0.9

these comparisons was categorised as either conservative, i.e., favouring the defendant, or non-conservative, i.e., favouring the prosecution. In simulation 1 (Table 1) the value θ used in the calculation (θ_{Calc}) was altered relative to the “true” or known level of substructure (θ_{Breed}). The results show that if the correct or nearly correct value is used for θ the model appears to be significantly conservative most of the time. In simulation 2 (Table 1) varying levels of sub-structure (θ_Z) were introduced into the sub-population, where the population itself already had a known level of substructure (θ_X). The results show that the induced departures from HWE and LE in the sub-population appear to have very little effect on the conservative performance of the model (Table 1).

Acknowledgements

We wish to acknowledge the support of the contributing laboratories. In addition, the research assistance of Heidi Drummond, Karah Bujtor, Jennifer Morris and Julianne Stevens in partitioning the datasets is gratefully acknowledged. This project was supported by the National Institute of Forensic Sciences, the University of Technology Sydney, and ESR, New Zealand. We acknowledge the coordination role of Dr. Tony Raymond, NIFS and Wayne Chisnall, ESR.

References

- [1] D.R. Horton (creator) Aboriginal Australia Map. AIATSIS and Auslig, Canberra, ACT, 1996.
- [2] P.O. Lewis, D. Zaykin, Genetic Data Analysis: Computer program for the analysis of allelic data. Version 1.0 (d16c). Free program distributed by the authors over the internet from <http://lewis.eeb.uconn.edu/lewishome/software.html>. (2001).
- [3] B.S. Weir, C.C. Cockerham, *Evolution* 38 (1984) 1358–1370.
- [4] K.L. Ayres, J. Chaseling, D.J. Balding, *Forensic Science International* 129 (2002) 90–98.
- [5] D.J. Balding, R.A. Nichols, *Forensic Science International* 64 (1994) 125–140.
- [6] J.M. Curran, J.S. Buckleton, C.M. Triggs, *Forensic Science International* 135 (2003) 1–8.
- [7] J.S. Buckleton, J.M. Curran, S.J. Walsh, How reliable is the sub-population model in DNA testimony? *Forensic Science International*. In press (accepted April 2005).