Forensic Science International: Genetics 6 (2012) 532–538

Contents lists available at SciVerse ScienceDirect

Forensic Science International: Genetics

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

An investigation of admixture in an Australian Aboriginal Y-chromosome STR database

Duncan Taylor [a], Nano Nagle [b], Kaye N. Ballantyne [c], Roland A.H. van Oorschot [c], Stephen Wilcox [d], Julianne Henry [a], Rust Turakulov [d], R. John Mitchell [b,*]

[a] Forensic Science SA, 21 Divett Place, Adelaide, SA 5000, Australia
[b] Department of Genetics, School of Molecular Sciences, La Trobe University, Victoria 3086, Australia
[c] Victoria Police Forensic Services Department, 31 Forensic Drive, Macleod, Melbourne, Victoria 3085, Australia
[d] Australian Genome Research Facility Ltd, 1G Royal Parade, Parkville, Victoria 3050, Australia

ARTICLE INFO

Article history:
Received 7 March 2011
Received in revised form 3 January 2012
Accepted 4 January 2012

Keywords:
Y chromosome
Haplogroup
Y-STR haplotypes
Admixture
Australian Aboriginals
SNP

ABSTRACT

Y-chromosome specific STR profiling is increasingly used in forensic casework. However, the strong geographic clustering of Y haplogroups can lead to large differences in Y-STR haplotype frequencies between different ethnicities, which may have an impact on database composition in admixed populations.

Aboriginal people have inhabited Australia for over 40,000 years and until ~300 years ago they lived in almost complete isolation. Since the late 18th century Australia has experienced massive immigration, mainly from Europe, although in recent times from more widespread origins. This colonisation resulted in highly asymmetrical admixture between the immigrants and the indigenes.

A State jurisdiction within Australia has created an Aboriginal Y-STR database in which assignment of ethnicity was by self-declaration. This criterion means that some males who identify culturally as members of a particular ethnic group may have a Y haplogroup characteristic of another ethnic group, as a result of admixture in their paternal line. As this may be frequent in Australia, an examination of the extent of genetic admixture within the database was performed. A Y haplogroup predictor program was first used to identify Y haplotypes that could be assigned to a European haplogroup. Of the 757 males (589 unique haplotypes), 445 (58.8%) were identified as European (354 haplotypes). The 312 non-assigned males (235 haplotypes) were then typed, in a hierarchical fashion, with a Y-SNP panel that detected the major Y haplogroups, C–S, as well as the Aboriginal subgroup of C, C4. Among these 96 males were found to have non-Aboriginal haplogroups. In total, ~70% of Y chromosomes in the Aboriginal database could be classed as non-indigenous, with only 169 (129 unique haplotypes) or 22% of the total being associated with haplogroups denoting Aboriginal ancestry, C4 and K* or more correctly K(xL,M,N,O,P,Q,R,S). The relative frequencies of these indigenous haplogroups in South Australia (S.A.) were significantly different to those seen in samples from the Northern Territory and Western Australia. In S.A., K* (~60%) has a much higher frequency than C4 (~40%), and the subgroup of C4, C4(DYS390.1del), comprised only 17%. Clearly admixture in the paternal line is at high levels among males who identify themselves as Australian Aboriginals and this knowledge may have implications for the compilation and use of Y-STR databases in frequency estimates.

1. Introduction

The issue of sub-population effects on DNA profile likelihood estimates is potentially of considerable importance in jurisdictions that comprise cosmopolitan populations [1–3]. This issue is even more pronounced in those jurisdictions that contain indigenous population(s) that have experienced a lifestyle for hundreds of generations that kept overall numbers small and community groupings widely dispersed and small in size. Aboriginal Australians are an indigenous population that fits this description, as, until recent times, they were nomadic foragers and hunters who lived in relatively small groups (~40 persons) within larger tribal clusters and defined geographical areas [4]. Their isolation on the Australian continent for a minimum of 40,000 years, together with the potential for inbreeding coupled with a preferential marriage

* Corresponding author. Tel.: +61 3 94792275; fax: +61 3 94792480.
E-mail addresses: duncan.taylor@sa.gov.au (D. Taylor), n.nagle@latrobe.edu.au (N. Nagle), kaye.ballantyne@police.vic.gov.au (K.N. Ballantyne), roland.vanoorschot@police.vic.gov.au (Roland A.H. van Oorschot), stephen.wilcox@agrf.org.au (S. Wilcox), rust.turakulov@agrf.org.au (R. Turakulov), john.mitchell@latrobe.edu.au (R.J. Mitchell).

1872-4973/$ – see front matter © 2012 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.fsigen.2012.01.001

system and polygamy, has significant ramifications for likelihood estimates of autosomal DNA profile rarity [2,3,5]. The extent of population structure in the Aboriginal forensic autosomal STR databases held by 5 State jurisdictions and also that held by the Northern Territory (NT) was investigated by Walsh et al. [5]. Their analyses of the data in these self-declared Aboriginal databases demonstrated significant differences between the regions in the level of divergence from self-declared Australian European databases. This very large sample of indigenous Australians was heavily biased to New South Wales (NSW) and the NT, with >80% of the total sample coming from those two jurisdictions. The largest and (quite marked) genetic differences were those between Aboriginal tribal regions in the NT and Europeans, whereas the smallest differences were observed between the tribal regions of South Eastern Australia and Europeans. This greater similarity between the Aboriginal and the Europeans databases from these areas was explained by Walsh et al. [5] as a result of considerable admixture over the preceding two centuries.

Recently the South Australian (S.A.) jurisdiction established Y-chromosome specific STR databases comprising males drawn from the autosomal South Australian Criminal Reference and Evidence DNA Database (SACREDD). Given the observations regarding admixture in Aboriginal autosomal databases, the question arises regarding its impact on an Aboriginal Y-STR database. The lack of recombination of the haploid section of the Y chromosome, and the strict paternal inheritance results in reduced variation within paternal lineages compared to autosomes. These properties of the Y chromosome mean any occurrence of admixture of Aboriginal women with non-Aboriginal (commonly European) males, a frequent occurrence historically, will be directly represented in the Y-STR haplotype of his male descendants. Typically such offspring would be raised as Aboriginal rather than European. It is important, therefore, an understanding is gained of the extent of admixture through the paternal line in Y-STR 'Aboriginal' databases before their use in calculation of Y haplotype frequency estimates.

In Australian jurisdictions, the likelihood estimates of an autosomal DNA profile are calculated from databases compiled for the three major ethnic groups – Aboriginal, European and Asian. Often, the likelihood will be calculated in multiple databases, including the database representative of the defendant (if he/she declares their ethnicity), on the basis that if all 18 alleles in the rare profile have been observed once in a member of that group it is more likely that a second occurrence will be seen in that group than any other. This approach replaced that of using a general database of the State's population because the number of Aboriginals in such a database, if any, would be very small and insufficient to represent the frequency of alleles and profiles in that group.

Y-STRs are relatively poor indicators of ethnicity because of their relatively rapid mutation rate [6,7] and, additionally, very little is known of Y-STR variation in indigenous Australians [8–11]. In contrast, single nucleotide polymorphisms (Y-SNPs) can be very powerful markers of geographical origin, with many of them being ethnic specific [12,13]. Based on very limited data, Australian Aboriginals exhibit a set of distinctive Y chromosome haplogroups (a haplogroup is defined after typing for a series of Y-SNPs)—C4 and K* or more accurately K(xLT,M,N,O,P,Q,R,S) [10,14]. Of these haplogroups, C4 is the most frequent, at least in the NT and WA samples published to date [8,10,14]. Haplogroup C4, defined by the SNP M347, is unique to Australian Aboriginals and has at least two sub-lineages, dependent on the presence/absence of a deletion in the STR DYS390 [10,14,15]. This DYS390.1 deletion of 16 bp results in 'short' alleles, between 18 and 20 repeats in size. The major source of admixture with Australian Aborigines over the last 200 years is Europe [5], a region which also has a distinct set of Y haplogroups (e.g. haplogroups R, I and J), none of which are thought to be indigenous to Australia [9,13,14].

This research examines a Y-STR database comprising males of self-declared Aboriginal ancestry in order to detect the extent of admixture and haplogroup/haplotype distribution throughout S.A. To achieve these aims all Y-STR haplotypes were first subjected to analysis using a haplogroup predictor software package [16] in order to specifically identify those associated with haplogroups of European ancestry. Second, haplotypes not assigned to a haplogroup of European ancestry were subjected to hierarchical Y-SNP analysis to determine their haplogroup.

2. Materials and methods

2.1. The Y-STR database

At the time of this research there were 757 male samples in the S.A. Aboriginal Y-STR database. The ethnicity of subjects was obtained chiefly by self-declaration or occasionally by information from the investigating officers, but they were not asked specifically about their paternal ancestry. It is important to clarify that Aboriginality is a culturally based affiliation and not one defined by a person's genetic make-up. Each sample came with a place name that was either the location of the offence or the residence of the donor. This 'placeholder' nomenclature ascribes contemporary geographical locations, such as a city, or town, or a small remote community centre. Reference to Horton's map of Aboriginal languages and traditional regions [17] allowed a contemporary 'placeholder' to be converted to a location within a traditional region of S.A.; Riverine, Desert and Spencer. The Riverine region lies in the extreme south east, the Desert region is in the north and Spencer covers much of the rest of the state. Additionally, those males located in the urban centre of Adelaide were assigned to a separate region, Urban. It is unlikely that all the Aboriginal samples come from the regional or tribal populations to which they were assigned, but given the lack of further information we consider it a reasonable treatment of the data.

2.2. Molecular analysis

All DNA samples were buccal swabs transferred to FTA® paper (Whatman BioScience) and stored at room temperature. DNA was extracted from the FTA® card following a Chelex based method adapted from Walsh et al. [18]. DNA was amplified using the Applied Biosystems Amp*Fl*STR® Yfiler™ PCR Amplification Kit as per the manufacturer's specifications.

The hierarchical Y-SNP typing protocol was as follows: All samples not assigned to an associated haplogroup were typed for a group of 6 'core' SNPs: M168 (defining haplogroup CDE), M130 (C), M89 (F) M9 (K) and M45 (P) and M207 (R). Those samples that were derived for M168, but ancestral for all other core SNPs, were typed for M174 (D) and M96 (E). Those samples derived for M168 and M130 but ancestral for the other SNPs were allocated to haplogroup C and further typed for M347 which defines C4, the uniquely Aboriginal haplogroup. Those samples exhibiting the derived allele for M168 and M89 but were ancestral for M130, M9, M45 and M207 were typed for the SNPs M201, M69, M170, and M304 that define haplogroups G, H, I, and J, respectively. Those samples exhibiting the derived state for M168, M89 and M9 but the ancestral allele for M130, M45 and M207 were typed for M20, M186, LLY22g, M175 and M230 that define haplogroups L, M, N, O and S, respectively. Fig. 1 illustrates a basic Y-chromosome phylogeny based on these SNPs.

2.3. Y-SNP detection

One to 5 ng of extracted genomic DNA was used for each Taqman™ (Applied Biosystems) assay with the reaction mix as