International Congress Series 1239 (2003) 67–69

Comparison of Y-chromosome haplotypes in three racial groups and the possibility of predicting ethnic origin

D. Syndercombe Court*, D. Ballard, C. Phillips, A. Revoir, C. Robson, C. Thacker

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

Abstract

Currently, most STRs found on the Y-chromosome exhibit much lower levels of polymorphism when compared to autosomal STRs. However, unlike autosomal STRs, they often show marked differences in allele frequency distributions between racial groups. Two hundred unrelated males from each of the three principal ethnic groupings within the UK were typed for 11 loci and used to build a predictive model for those classifications. An additional 50 individuals from each group were used in a blind trial to validate the model and the utility of the assignments assessed by calculating likelihood ratios. Use of a haplotype consisting of only three Y-chromosome STRs correctly identified 81%, 96% and 70% of individuals who defined themselves as white, black or South Asian.

Keywords: Y-chromosome STRs; Race; Ethnic origin

1. Introduction

Scientists have made attempts in the past to predict racial origin from genetic markers found in blood and DNA. Some blood group markers are prevalent in the black population, but can only be recognised amongst individuals who inherit the markers from

* Corresponding author. Tel.: +44-20-7377-7076; fax: +44-20-7377-7629.
E-mail address: y.d.syndercombe-court@qmul.ac.uk (D. Syndercombe Court).

0531-5131/03
PII: S0531-5131(02)00229-7

both parents. Other markers, while being found more often amongst blacks, are not common enough to be of much use. Little is available to help distinguish South Asians from others. Additionally, for forensic utility, any method should be able to make use of minute quantities of material. Most recently, six autosomal STRs have been used to infer ethnic origin [1], correctly predicting 56%, 67% and 43% of the Caucasian, Afro-Caribbean and South Asians they tested. They hypothesised that it would be more difficult to distinguish Caucasians and South Asians than other groupings. Large differences in allele distributions amongst Y-chromosome STRs between the racial groups led us to investigate these as possible predictors.

2. Materials and methods

Six hundred male individuals who described themselves and their parents as being 'white', 'black' or 'from the Indian Sub-continent (South Asian)', 200 in each group, were typed for 11 Y-chromosome STRs (DYS 19, 385, 389-I/II, 390, 391, 392, 393, 437, 438 and 439). A further 159 individuals (around 50 from each group) were used for validation purposes. A sub-group of individuals was also typed at the Gc locus using conventional electrophoretic techniques or Polymarker (Applied Biosytems) to see if the Gc 1F subtype, common amongst black individuals, could add further improvements. A data-mining approach based on the development of classification trees was used with

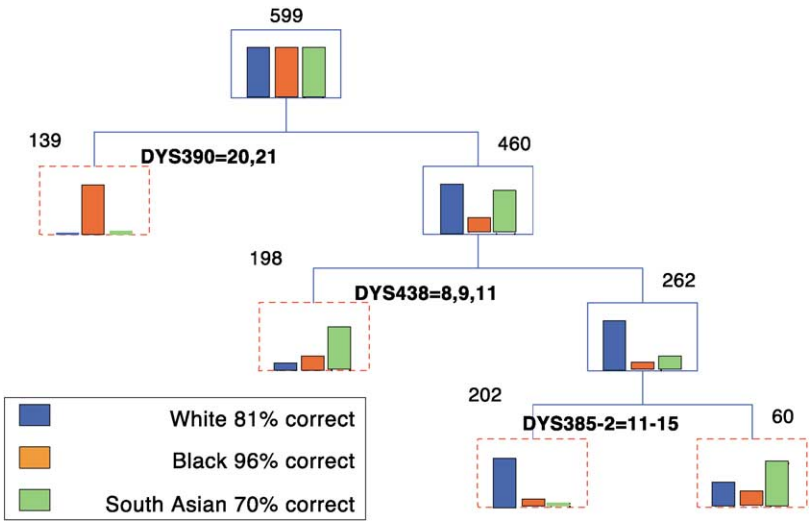

Fig. 1. Classification tree for ethnicity based on allelic markers. The total numbers in each classified group are shown above the box; within the box histograms illustrate the proportion identified in each group. The rule used for classification is shown between the boxes, with those individuals meeting the rule moving to the left. DYS385-2 refers to the larger of the alleles in this biallelic marker (8,9,11 implies those alleles only, whereas 11–15 is meant to imply the range of alleles).

Table 1
Likelihood ratios for competing hypotheses

Model prediction	Likelihood that the individual would have described himself as predicted compared with other groupings		
	White	Black	South Asian
White		10×	4×
Black	56×		34×
South Asian	23×	20×	

selection of a tree based on lowest possible misclassification and simplicity of the tree [2].

3. Results

The selected classification model is illustrated in Fig. 1 and makes use of binary classifications to correctly classify 81% of white individuals, 96% of blacks but only 70% of South Asians. In the model particular use is made of the common DYS390 (21) allele amongst black individuals. Three alleles in the DYS438 locus helped to identify some South Asians and more were identified with the DYS385 locus where South Asians are more represented amongst the larger alleles within the larger of the pair in this complex STR. Addition of Gc types to a subgroup of white and black individuals increased their correct classification of whites and blacks to 85% and 98%, respectively.

Table 1 illustrates the utility of the classification by presenting the competing likelihood ratios based on the best predictive model.

4. Discussion

Use of a small constellation of Y-chromosome STR markers has produced a useful predictive ability for broad ethnic classification, particularly where the prediction is not ‘white’. Lowe et al. predicted from Fst values that it would be more difficult to distinguish Caucasian from Asian, than Afro-Caribbean from Asian, but this model has shown that prediction of someone as ‘white’ has the least utility. The model has the lowest sensitivity (70%) for correctly identifying South Asians, compared with 98% for blacks and we are currently researching further markers to improve the former. Although the prediction has some important utility for intelligence purposes it should be employed with caution. The model presented here has been validated with a UK population and should be further validated with other populations where other markers are more discriminating.

References

[1] A.L. Lowe, A. Urquhart, L.A. Foreman, I.W. Evett, Inferring ethnic origin by means of an STR profile, Forensic Sci. Int. 11 (1) (2001) 17–22.

[2] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, Classification and Regression Trees, Wadsworth & Brooks/Cole Advanced Books & Software, Monterey, CA, 1984.