International Congress Series 1261 (2004) 59–61

www.ics-elsevier.com

Assigning individuals to ethnic groups based on 13 STR loci

Xenia Fosella[a], Fabio Marroni[a], Samuele Manzoni[b], Andrea Verzeletti[b], Francesco De Ferrari[b], Nicoletta Cerri[b], Silvano Presciuttini[a,*]

[a] *Center of Statistical Genetics, University of Pisa, Pisa, Italy*
[b] *Institute of Legal Medicine, University of Brescia, Brescia, Italy*

Abstract. Inferring the ethnic origin of individuals by means of short tandem repeat (STR) profiles has received considerable attention recently. Gene frequency variation among human populations has been extensively documented, and it has been suggested that the differences in allele proportions between ethnic groups could form the basis of an inferential system. We report the use of DNA profiles from 13 STR loci for inferring the ethnic origin of samples of unknown provenance using five populations of immigrants. This preliminary work shows that a population assignment test can be already used in real casework studies. © 2003 Published by Elsevier B.V.

Keywords: Ethnicity; Assignment test; STR; Population data

1. Introduction

The town of Brescia (northern Italy) ranks among the highest in Italy for the proportion of censused immigrants from non-EU countries, amounting to about 10% of the local resident population. Most blood crimes in this area happen within ethnically defined groups, and a test that could assign a biological stain to a subject of a particular population would be highly welcomed. We selected several groups of ethnically defined immigrants in this area and assessed the power of 13 STR loci of tracing back their ethnic origin.

2. Materials and methods

The loci included in the AmpFlSTR Profiler Plus™ and SGM Plus™ commercial kits were typed in the following population samples (not all individuals were typed for all loci): (1) Italians, born in Brescia: $N=120$; Indians: $N=38$; Maghrebians: $N=69$; Mongolians: $N=21$; Blacks, from sub-Saharan countries: $N=122$; Slavonians, including Albanians: $N=66$. We first applied the assignment test of the package Arlequin

* Corresponding author. Tel.: +39-50-2213797; fax: +39-50-2213524.
E-mail address: sprex@biomed.unipi.it (S. Presciuttini).

0531-5131/ © 2003 Published by Elsevier B.V.
doi:10.1016/S0531-5131(03)01699-6

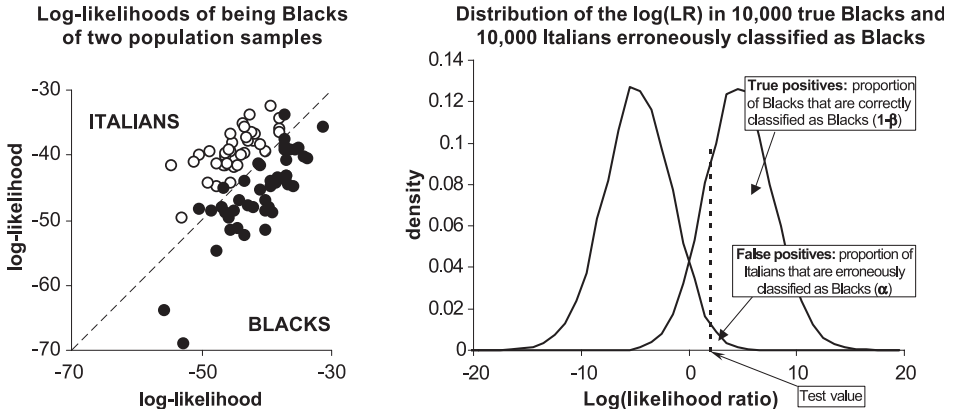

Fig. 1.

http://lgb.unige.ch/arlequin/ to all our data; however, since the assumption made by this program about the frequency of the alleles missing from a sample is not clear, we re-analyzed in more detail the comparison Blacks/Italians using an explicit and conservative assumption.

3. Results

In Arlequin's analysis, the log-likelihoods of each individual from four samples of immigrants (the Mongolians were excluded due to small sample size) were plotted against the log-likelihood of the Italians being falsely attributed to each of these samples. A fair level of discrimination power was evident in all population comparisons (data not shown). Not surprisingly, the sample of Blacks showed the highest level of differentiation; the likelihood of being Italian was higher than being Black in one Black subject only, whereas no Italian was attributed to the Black sample.

The analysis was repeated with our algorithm for the Black sample only, using the allele frequencies estimated from the samples (about 100 subjects typed for each locus in both samples), and assigning a value of 1/200 (= 0.005) to the frequency of the alleles missing from either sample (Fig. 1A). In order to estimate the statistical power of this assignment test with higher accuracy, we simulated 10 000 individuals for both the Italian and the

Table 1

Test value of ln(LR)	α (ratio of false positives)	$1-\beta$ (ratio of true positives)
0.0	6.6%	93.7%
1.0	3.1%	88.4%
2.0	1.5%	80.3%
3.0	0.6%	70.3%
4.0	0.3%	57.9%
5.0	0.1%	45.3%
6.0	0.01%	32.9%

Black samples using the same allele frequencies, and computed the log-likelihood distributions of the true Blacks assigned to the Black sample and the Italians falsely attributed to the Black sample (Fig. 1B).

From this analysis, we obtained the values of α and $1 - \beta$ shown in Table 1.

4. Conclusions

The 13 STR loci included in two commercial kits provided a limited but significant power to infer the ethnicity of non-EU individuals immigrating to an Italian town. Contrasting Blacks with resident Whites achieved the highest level of discrimination. When two alternative hypotheses about the ethnic origin of a sample can be formulated with sufficient confidence, a population assignment test can already be applied to real cases. The present work suggests that a sequential test can be highly efficient in assigning individuals to their ethnic group; given a fixed predefined significance level, investigators can start typing a small number of markers, and typing additional marker sets only for the subjects that failed to be classified at the previous step.