Genetic structure of the contemporary Cuban population for nine STR loci

R. Lleonart [a,*], Y. Carpio [a], M.V. Sainz de la Peña [b],
M. Santiesteban [b], G. Rodríguez [c], R. Ferreira [c], F. Amaro [d],
E. Riego [a], K. Bacallao [d], M. Blanco [e]

[a]Center for Genetic Engineering and Biotechnology, P.O. Box 6162, Havana, CP 10600, Cuba
[b]Laboratorio Central de Criminalística, Havana, Cuba
[c]Hermanos Ameijeiras Hospital, Havana, Cuba
[d]Legal Medicine Institute, Havana, Cuba
[e]AIDS Research Laboratory, Havana, Cuba

Short tandem repeats (STR) are, at the moment, the most convenient and powerful way to infer personal identification out of genome data obtained from scarce and highly degraded DNA. In the forensic context, personal identification from bones and other human remains is one of the most challenging tasks to solve in cases where traditional anthropology is insufficient to achieve that goal. In order to assess the usefulness of a set of microsatellites containing the loci HUMTH01, HUMTPOX, HUMCSF1PO, HUMvWA, HUMFESFPS, HUMF13A01, HUMF13B, HUMHPRTB and HUMLPL, reference databases were constructed for the three main racial groups of the contemporary Cuban population (Caucasoids, Mestizos and Negroids). Minimal allele frequencies were estimated following the expression described by Budowle [1] based on the predictions of the Infinite Alleles Model to avoid negative bias when rare alleles are considered. The distributions of allele frequencies were found to be similar to the ones reported by other authors at closely related population groups. A rare allele 5 was found, sequenced and characterized in Mestizo group at the HUMCSF1PO locus. Private alleles were found for the three groups. The three Cubans databases were validated for the presence of Hardy–Weinberg equilibrium and linkage disequilibrium using Markov Chain Monte Carlo methods. A few significant *P*-values were eliminated when Bonferoni-type corrections were done for the control of type I error, indicating both the validity of the sampling scheme done and the possibility of using the standard formulae to estimate the frequency of a multilocus profile in the Cuban population. The way genetic variation is distributed in

---

* Corresponding author. Tel.: +53-7-218008; fax: +53-7-218070.
E-mail addresses: rlleonart@cigb.edu.cu, rlleonart@yahoo.com (R. Lleonart).

the contemporary Cuban population was tested by estimating F- and ρ-statistics and testing their significance by permutation analysis. Genetic distance for microsatellites ($\delta\mu^2$) was also calculated between each racial group as well as the rate of migration (N_m) to further characterize the dynamics of genotypes in the population. Paired comparisons indicate that there is a very high level of genetic information exchange between groups as judged from the estimated values of N_m. Inbreeding coefficients (f) were estimated using the method of moments for each racial group being in close agreement with the values reported by other authors.

Results are also shown which confirm that among all three groups there is a significant genetic differentiation for all the loci except for HUMLPL, and a significant genotypic differentiation for all loci except HUMHPRTB and HUMLPL. The forensic usefulness of this group of nine STR loci was demonstrated by the very high values obtained for combined discrimination power which ranged from $1:1.5\times10^8$ (Caucasic group) to $1:2.6\times10^9$ (Black group).

Taken all together, these results confirm the validity of the use of both the Cuban databases and this STR set for personal identification purposes in this population.

Reference

[1] M. Budowle, Int. J. Leg. Med. 108 (1996) 173–176.