Int J Legal Med (2005) 119: 137–141
DOI 10.1007/s00414-005-0522-3

ORIGINAL ARTICLE

Yamila Carpio · Maria Victoria Sainz de la Peña · Marta Santiesteban · Fermín Amaro · Raul Ferreira · Ricardo Lleonart

Genetic structure of a Cuban population based on nine short tandem repeat loci

Received: 29 April 2004 / Accepted: 7 January 2005 / Published online: 15 February 2005

Abstract Cuba is a multiethnic and multiracial society. Here we describe the genetic variation of a sample of the Cuban population, which include the three most common racial groups, Caucasians, Negroids and Mestizos, by means of a set of nine microsatellites (HUMTH01, HUMTPOX, HUMCFS1PO, HUMVWA, HUMFESFPS, HUMF13A, HUMF13B, HUMLPL and HUMHPRTB). The analysis presented here indicates that these STR loci are highly informative for forensic purposes. The genetic data on the major racial groups is in good agreement with current demographic tendencies and with historic events that took place during the formation of the Cuban population.

Keywords Cuban population · Short tandem repeats · Microsatellites · Genetic structure

Introduction

Short tandem repeats loci have become an extremely useful tool for many applications including forensics, populations biology and anthropology.

The Cuban population at present is the result of very complex and overlapping processes of migrations and interracial admixture between groups from Asia, Africa and Spain (Rivero de la Calle 1984). The autochthonous aborigines present at the time of the discovery (1492) were exterminated in a relatively short period of time. Their presence in the contemporary Cuban genetic background is believed to be very small. The main racial groups at present are Negroids (with background of slaves introduced from the sub-Saharan West Africa, from Guinea-Bissau to Angola), Caucasians (originating from the many migrations from several regions of Spain), and Mestizos, which represents intermediate phenotypic characteristics between the two other groups. Other groups are also present at much lower frequency (Guanche 1996).

Information on the genetic variability of the Cuban population is scarce and almost exclusively based on red cell and serum biochemical markers (Hidalgo et al. 1999). Here we characterize the genetic variability of a sample of Cuban population based on the study of nine short tandem repeats. This study represents the first report on the genetic structure of the Cuban population based on microsatellite loci. The analysis presented here indicate that the nine STR loci are highly polymorphic and provide valuable potential for identification purposes and further studies regarding the population genetics of the Cuban population.

Materials and methods

Subjects and samples

Unrelated volunteers were randomly selected from each major racial group present in the contemporary Cuban population of La Habana city (102 Caucasians, 110 Mestizos, and 99 Negroids). Racial assignments were done as judged from the phenotypic characteristics described by Rivero de la Calle (1984). Individuals clearly having Mongoloid appearance were not included in the study. Samples were taken using a sterile oral swab or from intravenous blood, and DNA was extracted immediately by the Chelex method (Walsh et al. 1991).

For examination of the genetic relationship between the Cuban main racial groups and some of the historically re-

Y. Carpio · R. Lleonart (✉)
Center for Genetic Engineering and Biotechnology,
PO Box 6162, Havana, 10600, Cuba
e-mail: ricardo.lleonart@cigb.edu.cu
Fax: +53-7-331779

M. V. Sainz de la Peña · M. Santiesteban
Laboratorio Central de Criminalística,
Havana, Cuba

F. Amaro
Instituto de Medicina Legal,
Havana, Cuba

R. Ferreira
Hospital Hermanos Ameijeiras,
Havana, Cuba

Table 1 Allele frequencies for nine STR systems on Caucasian sample, including minimal allele frequencies (MAF)

Allele	STR								
	TH01	TPOX	CSF	VWA	FES	13A	13B	LPL	HPR
3.2	–	–	–	–	–	0.074	–	–	–
4	–	–	–	–	–	0.054	–	–	–
5	0.010	–	–	–	–	0.240	–	–	–
6	0.201	–	–	–	–	0.250	0.123	–	–
7	0.181	–	–	–	–	0.343	0.020	0.005	–
8	0.132	0.539	–	–	0.010	0.010	0.240	–	0.010
9	0.181	0.088	0.034	–	0.010	–	0.250	0.069	0.005
9.3	0.221	–	–	–	–	–	–	–	–
10	0.074	0.049	0.353	–	0.255	–	0.358	0.353	–
11	–	0.294	0.270	–	0.451	–	0.005	0.225	0.112
12	–	0.029	0.284	–	0.235	0.005	0.005	0.279	0.333
13	–	–	0.049	0.010	0.029	0.015	–	0.069	0.357
14	–	–	0.010	0.098	0.010	–	–	–	0.137
15	–	–	–	0.118	–	0.005	–	–	0.044
16	–	–	–	0.265	–	0.005	–	–	–
17	–	–	–	0.270	–	–	–	–	–
18	–	–	–	0.157	–	–	–	–	–
19	–	–	–	0.074	–	–	–	–	–
20	–	–	–	0.010	–	–	–	–	–
21	–	–	–	–	–	–	–	–	–
MAF	0.023	0.022	0.023	0.024	0.023	0.023	0.022	0.022	0.023

lated populations of the world, the populations considered were North Eastern Spain (Crespillo et al. 1997), Andalusia (Andres et al. 1996), Aragon North (Martinez-Jarreta et al. 1998), Catalonia (Gene et al. 1998), USA Caucasians (Budowle and Moretti 1999), Aragon South (Martinez-Jarreta et al. 1998), Central Eastern Spain (Martinez-Jarreta et al. 1998), Asturias (Bell et al. 2000), USA Hispanics (Budowle and Moretti 1999), Equatorial Guinea (Pagalu, Huguet et al. 1998), Rwanda (Tofanelli et al. 2003), USA Afro-Americans (Budowle and Moretti 1999), Equatorial Guinea (Bubi, Gene et al. 2001), and Guinea-Bissau (Goncalves et al. 2002).

Table 2 Allele frequencies for nine STR systems on Mestizo sample, including minimal allele frequencies (MAF)

Allele	STR								
	TH01	TPOX	CSF	VWA	FES	13A	13B	LPL	HPR
3.2	–	–	–	–	–	0.082	–	–	–
4	–	–	–	–	–	0.095	–	–	–
5	0.014	–	0.005	–	–	0.241	–	–	–
6	0.214	0.023	–	–	–	0.214	0.209	–	–
7	0.241	–	0.018	–	0.018	0.282	0.118	–	–
8	0.168	0.373	0.045	–	0.073	0.027	0.168	–	–
9	0.191	0.136	0.023	–	0.064	–	0.264	0.091	0.031
9.3	0.127	–	–	–	–	–	–	–	–
10	0.045	0.068	0.223	–	0.214	0.005	0.236	0.400	0.009
11	–	0.355	0.282	–	0.414	0.009	0.005	0.223	0.104
12	–	0.041	0.318	–	0.159	0.009	–	0.241	0.295
13	–	0.005	0.073	0.005	0.059	0.014	–	0.041	0.281
14	–	–	0.009	0.050	–	0.009	–	0.005	0.213
15	–	–	0.005	0.177	–	0.009	–	–	0.059
16	–	–	–	0.259	–	0.005	–	–	0.005
17	–	–	–	0.255	–	–	–	–	–
18	–	–	–	0.141	–	–	–	–	–
19	–	–	–	0.091	–	–	–	–	–
20	–	–	–	0.023	–	–	–	–	–
21	–	–	–	–	–	–	–	–	–
MAF	0.022	0.021	0.022	0.022	0.022	0.022	0.020	0.020	0.021

Table 3 Allele frequencies for nine STR systems on Negroid sample, including minimal allele frequencies (MAF)

Allele	STR								
	TH01	TPOX	CSF	VWA	FES	13A	13B	LPL	HPR
3.2	–	–	–	–	–	0.045	–	–	–
4	–	–	–	–	–	0.121	–	–	–
5	0.005	–	–	–	–	0.328	–	–	–
6	0.157	0.061	–	–	–	0.111	0.308	–	–
7	0.379	0.061	0.040	–	0.015	0.227	0.192	0.015	–
8	0.217	0.283	0.066	–	0.121	0.066	0.111	0.005	0.005
9	0.116	0.207	0.035	–	0.040	0.010	0.258	0.121	0.020
9.3	0.061	–	–	–	–	–	–	–	–
10	0.066	0.091	0.298	–	0.202	0.010	0.126	0.394	0.010
11	–	0.283	0.273	–	0.359	0.005	0.005	0.192	0.060
12	–	0.015	0.227	–	0.212	0.010	–	0.207	0.262
13	–	–	0.056	0.030	0.051	0.030	–	0.061	0.277
14	–	–	0.005	0.056	–	0.025	–	0.005	0.242
15	–	–	–	0.242	–	0.005	–	–	0.090
16	–	–	–	0.247	–	0.005	–	–	0.030
17	–	–	–	0.217	–	–	–	–	–
18	–	–	–	0.141	–	–	–	–	–
19	–	–	–	0.066	–	–	–	–	–
20	–	–	–	–	–	–	–	–	–
21	–	–	–	–	–	–	–	–	–
MAF	0.023	0.024	0.024	0.024	0.024	0.026	0.022	0.023	0.025

Markers

The STR systems used were the triplexes FFV (HUMF 13A01, HUMFESFPS, HUMVWA), CTT (HUMCSF1PO, HUMTPOX, HUMTH01) and the monoplexes HUMF13B, HUMHPRTB and HUMLPL as described by the manufacturer (Promega, USA). Loci amplifications, denaturing polyacrylamide electrophoresis, silver stainings and allele assignments were done also as recommended by the manufacturer.

Table 4 Forensic and descriptive statistics for the three Cuban racial groups studied

Group	Locus	He[a]	Ho[b]	M.P.[c]	M.P.[d]	D.P.[e]	P.E.[f]	HWE[g]
Caucasian	TH01	0.8261	0.8529	0.069	14.5	0.931	0.701	0.0726
	TPOX	0.6147	0.5882	0.212	4.70	0.788	0.277	0.0278
	CSF	0.7217	0.8235	0.162	6.2	0.838	0.643	0.0590
	VWA	0.8075	0.7549	0.071	14.1	0.929	0.518	0.1800
	FES	0.6784	0.6764	0.169	5.9	0.831	0.393	0.2116
	F13A	0.7570	0.7058	0.112	8.9	0.888	0.437	0.0139
	F13B	0.7399	0.7549	0.132	7.6	0.868	0.518	0.0128
	LPL	0.7407	0.6568	0.117	8.6	0.883	0.365	0.0137
Mestizos	TH01	0.8168	0.8454	0.074	13.6	0.926	0.686	0.1676
	TPOX	0.7131	0.7181	0.146	6.9	0.854	0.457	0.1742
	CSF	0.7648	0.7454	0.105	9.5	0.895	0.502	0.0576
	VWA	0.8091	0.7363	0.064	15.6	0.936	0.487	0.7403
	FES	0.7481	0.7000	0.102	9.8	0.898	0.428	0.0529
	F13A	0.8034	0.8272	0.076	13.2	0.924	0.651	0.5432
	F13B	0.7922	0.7272	0.081	12.4	0.919	0.472	0.1832
	LPL	0.7299	0.6363	0.117	8.6	0.883	0.337	0.5080
Negroids	TH01	0.7672	0.7979	0.119	8.4	0.881	0.595	0.0090
	TPOX	0.7852	0.7070	0.083	12	0.917	0.439	0.0577
	CSF	0.7788	0.7070	0.085	11.8	0.915	0.439	0.2495
	VWA	0.8085	0.7373	0.070	14.2	0.930	0.488	0.1602
	FES	0.7703	0.8484	0.105	9.5	0.895	0.692	0.4510
	F13A	0.8093	0.7575	0.067	14.8	0.933	0.523	0.0687
	F13B	0.7775	0.7979	0.092	10.8	0.908	0.595	0.7130
	LPL	0.7502	0.7171	0.100	10.0	0.900	0.455	0.6011

Exact *P* values estimated by Markov chain Monte Carlo procedure following the parameters: dememorization steps, 1,000; batch number, 1,000; iterations per batch, 1,000. Significance value, 0.05. Bonferroni-corrected value, 0.00625

[a] Expected heterozygosity, calculated as Nei (1987)

[b] Observed heterozygosity

[c] Matching probability

[d] Matching probability, computed as “1 in...”

[e] Discrimination power

[f] Power of exclusion

[g] Hardy–Weinberg equilibrium test

Statistical analysis

Allelic frequencies, matching probability, power of discrimination, and power of exclusion were estimated using Excel spreadsheet (PowerStats, Promega, USA). Minimal allele frequencies were calculated (Budowle et al. 1996).

Hardy–Weinberg equilibrium was checked by means of an exact test as implemented in the Genepop 3.2a software (Raymond and Rousset 1995a). The same software was used to estimate possible linkage disequilibrium between pairs of loci. The differences in allele frequencies across all Cuban groups were tested as in Raymond and Rousset (1995b), while the corresponding analysis for genotype frequencies was done as described by Goudet et al. (1996). The observed and expected proportions of heterozygotes were estimated using the GDA software (d15, Lewis and Zaykin 2001), as well as the values of the *F*-statistics.

As a measure of genetic distance among the three Cuban racial groups and other related populations of the world, pairwise Fst distance was estimated as described by Reynolds et al. (1983) using the program MICROSAT (v1.5d; Minch). For this estimation, published population data on the markers HUMCSF1PO and HUMVWA was used. Non-metric multidimensional scaling was performed on the Fst genetic distance matrix by using the PROXSCAL algorithm as implemented in the SPSS package (v11.5).

Results and discussion

Short tandem repeats loci have been useful before for identification purposes in the Cuban population (Lleonart et al. 2000). However, a detailed genetic characterization of this population for this type of markers was still needed for the routine use of STR in the forensic context. When considering the nine markers, no genotype was observed twice in the whole sample, as expected from the known variability of these systems. One rare, shorter allele 5 was observed in the Mestizo group at the system HUMCSF1PO (Lleonart et al. 2003; informed to the Internet STR database at http://www.cstl.nist.gov/biotech/strbase/var_csf.htm).

The allelic frequencies distributions found for the examined STRs in the Cuban groups resembled the ones reported for several Caucasian and Negroid populations in the world (Pérez-Lezaun et al. 2000). Some of the allelic frequencies reported here differ notably from the ones informed for a South American population (Albeza et al. 2002), especially at the systems HUMVWA, HUMTH01 and HUMF13A01. As expected, the Mestizo group frequently showed intermediate frequencies as compared with the Caucasian and Negroid groups (Tables 1, 2, 3). Private alleles were found in all three Cuban groups, the allele 7 at the HUMTPOX system (frequency, 0.06) and the allele 9 at the HUMF13A01 (frequency, 0.01) being the most frequent, both on the Negroids sample (Tables 1, 2, 3).

Several population genetics parameters where estimated from the genotypic data for each of the three most important Cuban racial groups (Table 4). The values for observed and expected heterozygosity were high for all the systems and in agreement of the values reported for several populations abroad using some of the markers presented here (Pérez-Lezaun et al. 2000). For these calculations, the HUMHPRTB system was excluded given the fact that our sample contained an undetermined proportion of males (Szibor et al. 2003). The overall figures for the discrimination power were very high for the three racial groups. These figures confirm the very high usefulness of this set of STR for individualization and forensic purposes in the Cuban population.

We checked for compliance with the Hardy–Weinberg proportions and also for linkage disequilibrium among pairs of loci. The system HUMHPRTB was excluded from the analysis. After Bonferroni's correction for multiple tests (Weir 1996), there was no indication of major departures from the equilibrium proportions (Table 4). There was also no indication of extensive linkage disequilibrium for the samples tested. After the correction for multiple tests, only the pair HUMHPRTB–HUMF13A showed a highly significant value ($p<0.0013$). This result is probably an artifact, consequence of the mixed composition of our sample.

When comparing allelic and genotypic frequencies among the Cuban groups, highly significant differences were found for almost all the loci (data not shown). The genotypic frequencies were significantly different among the three groups for all the STR except for the HUMHPRTB ($p=$

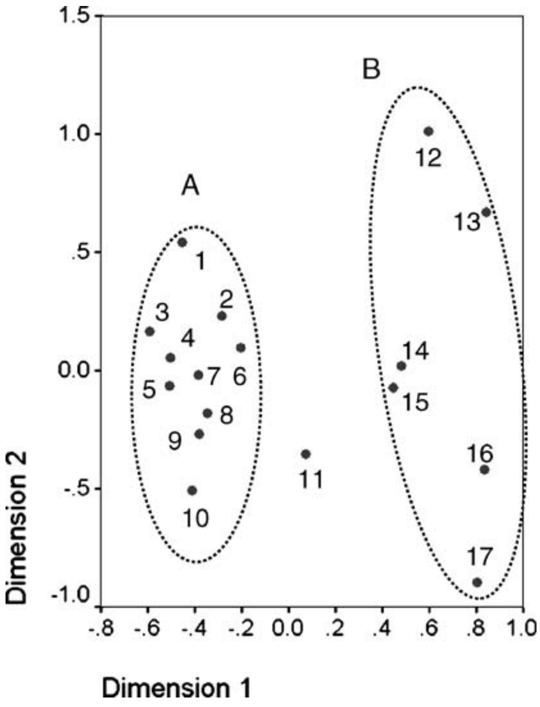

Fig. 1 Two-dimensional solution for a non-metric multidimensional scaling analysis of Reynold's Fst genetic distances among Cuban and related populations (details in Materials and methods). *Dotted circles* show clearly defined Caucasians (**a**) and Negroid populations (**b**). Populations included are: *1* north-eastern Spain, *2* Andalusia, *3* Aragon north, *4* Catalonia, *5* USA Caucasians, *6* Cuban Caucasians, *7* Aragon south, *8* Central Eastern Spain, *9* Asturias, *10* USA Hispanics, *11* Cuban Mestizos, *12* Equatorial Guinea (Pagalu), *13* Rwanda, *14* Cuban Negroids, *15* USA Afro-americans, *16* Equatorial Guinea (Bubi), *17* Guinea Bissau

0.1149) and HUMLPL (p=0.4351), while the allelic frequencies were different for all the systems except for the HUMLPL (p=0.3531).

To characterize further the genetic structure of the Cuban population, the F-statistics were estimated by analysis of variance. The Fst value obtained (1.4%) was well in agreement with values reported for other populations of the world (Martínez-Labarga et al. 1999). The estimated value of F_{is} for the Cuban groups was not statistically different from zero.

In order to compare the three main Cuban groups to populations that contributed to the contemporary Cuban population, Fst genetic distances were calculated. The non-metric multidimensional analysis (Fig. 1) was indicative of a clear distinction among African and Caucasian populations. It is interesting to note that while Cuban Caucasics shows a resemblance to samples from Andalusia and Aragon, Hispano-americans seems to be less related. This finding may have been also related to the fact that this group has been prepared by pooling data from subpopulations that almost certainly contains other backgrounds.

The Afro-Americans, however, looks very close to the Cuban Negroids. The Cuban Mestizos group appears intermediate, confirming findings at the level of allelic frequencies. The microsatellite data presented here confirm other lines of evidence about the very intense process of admixture that is under way among these Cuban groups.

Overall, these findings corroborate the high efficiency of this group of short tandem repeats loci for their use both at the forensic context and for further studies on the population biology of the Cuban and related populations.

References

Albeza MV, Picornell A, Acreche N, Tomas C, Castro JA, Ramon MM (2002) Genetic variability at 14 STR loci in the Puna population of northwestern Argentina. Int J Leg Med 116:126–132

Andres MI, Prieto V, Flores IC, Sanz P (1996) Population genetics of three STRs: TH01, CSF1PO and TPOX in Southern Spain. Adv Forensic Haemog 6:489–491

Bell B, Nievas P, Abecia E, Martínez-Jarreta B, Hinojal R, Martinez-Cordero A (2000) Population genetics of the STR loci HUMCS F1PO, HUMF13A01, HUMFES/FPS and D12S391 in Asturias (Northern Spain). Forensic Sci Int 113:21–23

Budowle B, Moretti TR (1999) Genotype profiles for six population groups at the 13 CODIS short tandem repeat core loci and other PCR-based loci. Forensic Sci Commun 1(2)

Budowle B, Monson KL, Chakraborty R (1996) Estimating minimum allele frequencies for DNA profile frequency estimates for PCR-based loci. Int J Leg Med 108:173–176

Crespillo M, Luque JA, Fernández R, Ramírez E, García P, Valverde JL (1997) Allele frequency distributions of 13 PCR-based systems in a population from North-east Spain. Int J Leg Med 110:223–225

Gene M, Carracedo A, Huguet E, Pérez-Pérez A, Moreno P (1998) Population genetics of the D12S391, CSF1PO and TPOX loci in Catalonia (Northeast Spain). Int J Leg Med 111:52–54

Gene M, Moreno P, Borrego N, Pique E, Brandt C, Mas J, Luna M, Corbella J, Huguet E (2001) The Bubi population of Equatorial Guinea characterized by HUMTH01, HUMVWA31A, HUMCS F1PO, HUMTPOX, D3S1358, D8S1179, D18S51 and D19S253 STR polymorphisms. Int J Leg Med 114:298–300

Goncalves R, Jesús J, Fernandes AT, Brehm A (2002) Genetic profile of a multi-ethnic population from Guiné-Bissau (west African coast) using the new PowerPlex 16 System kit. Forensic Sci Int 129:78–80

Goudet J, Raymond M, De Meeüs T, Rousset F (1996) Testing differentiation in diploid populations. Genetics 144:1933–1940

Guanche J (1996) Componentes Etnicos de la Nación Cubana. Fundación Fernando Ortiz, Ediciones Unión, Ciudad de La Habana

Hidalgo PC, Cabrera J, Castellanos T, López N, González O, Villar M, Bauer I (1999). Abnormal hemoglobins and racial admixture in Villa Clara province. Cuban J Hum Genet 1(2)

Huguet E, Borrego N, Pinheiro MF, Luna M, Corbella J, Mas J, Gené M, Moreno P (1998) Annobon Island population (Equatorial Guinea) characterised by five VNTRs-PCR polymorphisms. Prog Forensic Genet 7:329–331

Lewis PO, Zaykin D (2001) 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://www.lewis.eeb.uconn.edu/lewishome/software.html

Lleonart R, Riego E, Sainz de la Pena MV, Bacallao K, Amaro F, Santiesteban M, Blanco M, Currenti H, Puentes A, Rolo F, Herrera L, de la Fuente (2000) Forensic identification of skeletal remains from members of Ernesto Che Guevara's guerrillas in Bolivia based on DNA typing. Int J Leg Med 113(2):98–101

Lleonart R, Carpio Y, Sainz de la Peña MV, Santiesteban M, Rodríguez G, Ferreira R, Amaro F, Riego E, Bacallao K, Blanco M (2003) Genetic structure of the contemporary Cuban population for nine STR loci. Int Congr Ser 1239:147–148

Martínez-Jarreta B, Diaz Roche P, Budowle B, Abecia E, Castellano M, Casalod Y (1998) Pyrenean population data on 3 tetrameric short tandem repeat loci—HUMTH01, TPOX and CSF1PO-derived using a STR multiplex system. Prog Forensic Genet 7:312–314

Martínez-Labarga C, Rickards O, Scacchi R, Corbo RM, Biondi G, Peña JA, Varas de Vieira C, Guevara AEB, Mesa Saturino MS, De Stefano GF (1999) Genetic population structure of two African-Ecuadorian communities of Esmeraldas. Am J Phys Anthropol 109:159–174

Minch E, MICROSAT, http://www.hpgl.stanford.edu/projects/microsat/) (Stanford University)

Pérez-Lezaun A, Calafell F, Clarimón J, Bosch E, Mateu E, Gusmao L, Amorim A, Benchemsi N, Bertranpetit J (2000) Allele frequencies of 13 short tandem repeats in population samples from the Iberian Peninsula and northern Africa. Int J Leg Med 113:208–214

Raymond M, Rousset F (1995a) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J Heredity 86:248–249

Raymond M, Rousset F (1995b) An exact test for population differentiation. Evolution 49:1280–1283

Reynolds J, Weir BS, Cockerham CC (1983) Estimation of the coancestry coefficient: basis for a short-term genetic distance. Genetics, 105:767–779

Rivero de la Calle, M (1984) Antropología de la Población Adulta Cubana. Editorial Científico-Técnica, La Habana

Szibor R, Krawczak M, Hering S, Edelmann J, Kuhlisch E, Krause D (2003) Use of X-linked markers for forensic purposes. Int J Leg Med 117:67–74

Tofanelli S, Boschi I, Bertoneri S, Coia V, Taglioli L, Franceschi MG, Destro-Bisol G, Pascali V, Paoli G (2003) Variation at 16 STR loci in Rwandans (Hutu) and implications on profile frequency estimation in Bantu-speakers. Int J Leg Med 17:121–126

Walsh PS, Metzger DA, Higuchi R (1991) Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. BioTechniques 10:506

Weir S (1996) Genetic Data Analysis II. Sinauer Associates, Sunderland, MA, USA