



Allele frequencies for CODIS loci in a Sicilian population sample

C. Vecchio^{a,*}, L. Garofano^a, L. Saravo^a, S. Spitaleri^a, G. Iacovacci^a, M. Santacroce^a, V. Manzari^b, B. Budowle^c

ARaggruppamento Carabinieri Investigazioni Scientifiche, Viale di Tor di Quinto n.151, 00191 Rome, Italy
 Dipartimento Medicina Università "Tor Vergata"-00100 Rome, Italy
 FBI Laboratory, 2501 Investigation Pkwy., Quantico, VA 22135, USA

Abstract. Data from an island population can be useful for understanding effects of forensic evaluations of DNA evidence. In this study, allele frequencies for the thirteen CODIS loci were generated from 220 unrelated volunteer donors from a Sicilian (Italy) population sample, as well as from casework using both the AmpFISTR ™ Profiler and Co-Filer ™ Amplification Kits (PE Applied Biosystems, Foster City, CA). All loci met Hardy−Weinberg expectations. The data were compared with population studies from individuals residing in other Italian geographic locations and found not be substantially different. The combined power of discrimination is high and comparable to other European population groups. © 2003 Elsevier B.V. All rights reserved.

Keywords: DNA; Short tandem repeat (STR); CODIS; Population data; Forensic

1. Introduction

Short tandem repeat (STR) polymorphisms are the genetic markers of choice for human identification and paternity testing. Moreover, STR loci are the genetic markers of national and international DNA databanks, such as CODIS (FBI, USA), based on a core of 13 STRs. In this preliminary study we investigated the 13 STR CODIS loci in a Sicilian sample (220 unrelated Sicilian donors). The results obtained were compared with previous studies on Italian populations and were found not to be substantially different [1,2].

2. Statistical analysis

Allele designations were determined by comparison of the sample fragments with those of the allelic ladders supplied with each kit. At each locus, the frequency of each allele was calculated from the numbers of each genotype in the sample set (i.e., the gene count method). Unbiased estimates of expected heterozygosity were computed as described by

^{*} Corresponding author. Tel.: +39-333-3458105; fax: +39-06-233212993. E-mail address: evecchio@carabinieri.it (C. Vecchio).

Table 1

All	D8		D21		D18S		D3S		FGA		D5S		D13S		D7S		D16S		vWA		TH01		TPOX		CFS	
	Nr.	Freq	Nr.	Freq	Nr.	Freq	Nr.	Freq	N	Freq	Nr.	Freq.	Nr.	Freq.	Nr.	Freq.	Nr.	Freq.	Nr.	Freq.	Nr.	Freq.	Nr.	Freq.	Nr.	Freq.
6																					117	0.266				
7											1	0.002			6	0.014					89	0.202	2	0.005	1	0.002
8	9	0.020									9	0.020	45	0.102	75	0.170	14	0.032			49	0.111	239	0.543	2	0.005
9	3	0.007									20	0.045	27	0.061	41	0.093	50	0.114			82	0.186	57	0.130	17	0.039
9.3																					92	0.209				
10	37	0.084			3	0.007					32	0.073	29	0.066	137	0.311	32	0.073			11	0.025	28	0.064	132	0.300
11	32	0.073			10	0.023					145	0.330	134	0.305	116	0.264	128	0.291					104	0.236	118	0.268
12	48	0.109			77	0.175	1	0.002			139	0.316	146	0.332	52	0.118	100	0.227	2	0.005			9	0.020	143	0.325
13	127	0.289			66	0.150	1	0.002			86	0.195	44	0.100	10	0.023	52	0.118	2	0.005			1	0.002	27	0.061
14	98	0.223			58	0.132	28	0.064			7	0.016	15	0.034	3	0.007	12	0.027	49	0.111						
15	62	0.141			63	0.143	97	0.220									8	0.018	47	0.107						
16	20	0.045			61	0.139	112	0.255			1	0.002					11	0.025	112	0.255						
17	4	0.009			40	0.091	104	0.236	1	0.002							16	0.036	115	0.261						
18					25	0.057	93	0.211	8	0.018							11	0.025	79	0.180						
19					17	0.039	4	0.009	32	0.073							5	0.011	29	0.066						
19.2									2	0.005																
20					6	0.014			52	0.118							1	0.002	5	0.011						
20.2									3	0.007																
21					7	0.016			61	0.139																
21.2									6	0.014																
22					1	0.002			86	0.195																
22.2									2	0.005																
23									74	0.168																
23.2									2	0.005																
24									46	0.105																
24.2									12	0.027																
25									29	0.066																
25.2									1	0.002																
26			1	0.002					18	0.041																
26.2									1	0.002																
27			12	0.027					2	0.005																
28			63	0.143	2	0.005			1	0.002																
29			96	0.218	1	0.002																				
30			112	0.255	2	0.005																				
30.2			18	0.041																						
31			24	0.055	1	0.002																				
31.2			49	0.111					1	0.002																
32			5	0.011																						
32.2			43	0.098																						
33.2			16	0.036																						
35			1	0.002																						

Edwards et al. [3]. Possible divergence from Hardy—Weinberg expectations (HWE) was tested by calculating the unbiased estimate of the expected homozygote/heterozygote frequencies and also by performing an exact test [4]. The probability of discrimination (PD) and probability of exclusion (PE) was calculated according to Fisher [5].

3. Results and discussion

The distribution of observed alleles for all 13 STR loci is shown in Table 1. All loci are highly polymorphic with the loci D18S51 (86.8%), D21S11 (85.5%) and FGA (84.5%) having the highest observed heterozygosities, and the locus TPOX showing the lowest observed heterozygosity (65.5%). The most polymorphic loci are the most discriminating loci: FGA (97%), D18S51 (96.5%) and D21SII (95.1%). The TPOX locus is the least discriminating (80.3%). For all 13 STR loci the combined power of discrimination is greater than 0.99999999 and the combined probabilities of exclusion are 0.99998. There was no evidence for departures from HWE for any loci based on the exact test. The Sicilian population data for these 13 loci do not differ substantially from other Caucasian data for the loci (data not shown). In conclusion, a database for the 13 STR loci has been established for the Sicilian population. These data can be used for both forensic casework and paternity testing in Sicilian population. These data together with previous studies on Italian populations could be a useful tool to appreciating the diversity of some island populations. Future projects will attempt to increase the number of samples in the Sicilian dataset so that greater confidence can be obtained for statistical estimates of the rarity of a DNA profile.

References

- L. Garofano, et al., Italian population data on thirteen short tandem repeat loci: HUMTH01, D21S11,
 D18S51, HUMVWFA31, HUMFIBRA, D8S1179, HUMTPOX, HUMCSF1PO, D16S539, D7S820.
 D13S317, D5S818, D3S1358, Forensic Science International 97 (1998) 53-60.
- [2] R. Biondo, et al., Regional Italian Allele frequencies at nine short tandem repeat loci, Forensic Science International 115 (1-2) (2001 1 January) 95-98.
- [3] R. Edwards, et al., Genetic variation at five trimeric and tetrameric repeat loci in four human population groups, Genomics 12 (1992) 241–253.
- [4] S.W. Guo, Ea. Thompson, Performing the exact test of Hardy-Weinberg proportion for multiple alleles, Biometrics 48 (1992) 361-372.
- [5] R. Fisher, Standard calculation for evaluating a blood group system, Heredity 5 (1951) 95-102.