ELSEVIER

Forensic Science International
81 (1996) 191–199

Forensic Science International

Northern and southern Croatian population data on seven PCR-based loci

Kathleen M. Keys[a], Bruce Budowle[a,*], Simun Andelinovic[b], Marija Definis-Gojanovic[b], Irena Drmic[b], Marcikic Mladen[c], Dragan Primorac[b,d]

[a]*Forensic Science Research and Training Center, FBI Academy, Quantico, Virginia 22135, USA*
[b]*Zagreb University School of Medicine, Split Branch, Split, Croatia*
[c]*Zagreb University School of Medicine, Osijek Branch, Osijek, Croatia*
[d]*The University of Connecticut Health Center, Farmington, Connecticut, USA*

Received 7 November 1995; revised 14 March 1996; accepted 18 March 1996

Abstract

Northern and southern Croatian sample populations were typed at seven PCR-based loci — LDLR, GYPA, HBGG, D7S8, Gc, HLA-DQA1 and D1S80. The results show that all loci meet Hardy-Weinberg expectations and that there is little evidence for association of alleles between loci. Allelic frequency distributions at all loci, except HLA-DQA1, show no differences between the northern and southern Croatian sample populations. Moreover, the population data for Croatians are similar to U.S. Caucasians; only HLA-DQA1 for southern Croatians was statistically different compared with U.S. Caucasians. A Croatian population database(s) has been created and can be used for forensic analyses to estimate the frequency of a multiple locus DNA profile.

Keywords: Croatia; Population databases; PCR; Hardy-Weinberg Expectations; HLA-DOA1; LDLR; GYPA; HBGG; D7S8; Gc; D1S80

1. Introduction

Little population data exist for forensically important PCR-based loci for

*Corresponding author. Forensic Science Research and Training Center, FBI Academy, Quantico, Virginia 22135, USA.

PII S0379-0738(96)01950-0

Eastern European populations [1–3]. Six loci — low density lipoprotein receptor (LDLR), glycophorin A (GYPA), hemoglobin G gammaglobin (HBGG), D7S8, group-specific component (Gc) (PM loci), and HLA-DQA1 — can be typed using commercially available reverse dot blot kits, and a seventh locus, D1S80, is a variable number of tandem repeats locus. It is desirable to obtain population data on these loci from various subgroups to gain appreciation empirically of the effects of population substructure on forensic DNA profile frequency estimates. For this study, blood samples were obtained from two geographically defined Croatian population samples, i.e. northern and southern Croatian individuals, and typed for LDLR, GYPA, HBGG, D7S8, Gc, HLA-DQA1, and D1S80. Although the heritage of northern and southern Croatians may be different, the results of this study demonstrate that for the seven loci studied here, there is little genetic difference between these two subgroups, as well as between Croatian populations and U.S. Caucasians.

2. Materials and methods

2.1. Sample preparation

Whole blood samples were collected in EDTA Vacutainer tubes from 100 unrelated northern Croatians and 103 unrelated southern Croatians and provided by the DNA Laboratory, Clinical Hospital, Split Branch, Croatia, for study. The samples were air dried on cotton cloth and extracted as described previously [4]. The quantity of DNA in each sample was estimated using the slot-blot procedure described by Waye et al. [5]. Generally 2–5 ng of DNA were amplified by PCR.

2.2. Typing

The Polymarker (PM) loci were typed using the Amplitype PM PCR Amplification and Typing Kit (Perkin-Elmer Corporation, Norwalk, CT). The amplification conditions were those recommended by the manufacturer, except that 16 μg of bovine serum albumin (Sigma, catalog # 3350) were added to the PCR. Amplification was carried out in a Perkin-Elmer 9600 Thermal Cycler.

The population samples were also typed using the Amplitype HLA-DQA1 Forensic DNA Amplification and Typing Kit (Perkin-Elmer Corporation, Norwalk, CT) by following the manufacturer's recommended protocol. The HLA-DQA1 PCR products were derived from the PM multiplex amplifications.

The DNA samples were also typed for D1S80 according to the method of Budowle et al. [6].

2.3. Statistical analysis

The frequency of each allele for each locus was calculated from the numbers of each genotype in the sample set. Unbiased estimates of expected heterozygosity

were computed as described by Edwards et al. [7]. Possible divergence from Hardy-Weinberg expectations (HWE) was determined by calculating the unbiased estimate of the expected homozygote/heterozygote frequencies [8–10], the likelihood ratio test [7,11,12], and the exact test [13]. An interclass correlation criterion was used for detecting disequilibrium between loci pairs [14]. Evidence for lack of independence across the PM markers, HLA-DQA1 and D1S80, was also evaluated by examining whether or not the observed variance of the number of heterozygous loci in the population sample is outside its confidence interval under the assumption of independence [15,16]. A 2 × C contingency table exact test [17,18] was used to generate a G-statistic (1000 shuffling experiments) to test for homogeneity between the Croatian sample populations and among the Croatians and a U.S. Caucasian population sample.

3. Results and discussion

The distribution of observed allele and genotype frequencies for the PM, HLA-DQA1 and D1S80 loci in the Croatian populations are shown in Tables 1, 2, 4–6. The genotype distributions for all seven PCR-based loci do not deviate from Hardy-Weinberg expectations based on the homozygosity test, likelihood ratio, and the exact test (Tables 3, 4 and 6).

Table 1
Observed genotype frequency distributions of PM loci in 98 unrelated Northern Croatians (N.C.) and 101 unrelated Southern Croatians (S.C.)

Genotype	Frequency N.C.	Frequency S.C.
LDLR AA	0.143	0.168
LDLR AB	0.541	0.475
LDLR BB	0.316	0.356
GYPA AA	0.235	0.337
GYPA AB	0.571	0.525
GYPA BB	0.194	0.139
HBGG AA	0.337	0.267
HBGG AB	0.439	0.475
HBGG BB	0.224	0.257
HBGG AC	0.000	0.000
HBGG BC	0.000	0.000
HBGG CC	0.000	0.000
D7S8 AA	0.439	0.426
D7S8 AB	0.439	0.446
D7S8 BB	0.122	0.129
Gc AA	0.071	0.050
Gc AB	0.112	0.089
Gc BB	0.000	0.010
Gc AC	0.337	0.337
Gc BC	0.122	0.158
Gc CC	0.357	0.356

Table 2
Observed allele frequency distributions for PM loci in 98 unrelated Northern Croatians (N.C.) and 101 unrelated Southern Croatians (S.C.)

Allele	Frequency N.C.	Frequency S.C.
LDLR A	0.413	0.406
LDLR B	0.587	0.594
GYPA A	0.520	0.599
GYPA B	0.480	0.401
HBGG A	0.556	0.505
HBGG B	0.444	0.495
HBGG C	0.000	0.000
D7S8 A	0.658	0.649
D7S8 B	0.342	0.351
Gc A	0.296	0.262
Gc B	0.117	0.134
Gc C	0.587	0.604

Table 3
Distribution of observed HLA-DQA1 genotype frequencies in 97 unrelated Northern Croatians (N.C.) and 94 unrelated Southern Croatians (S.C.)

Genotype	Frequency[a] N.C.	Frequency[b] S.C.
1.1–1.1	0.010	0.021
1.1–1.2	0.052	0.085
1.1–1.3	0.010	0.043
1.1–2	0.021	0.000
1.1–3	0.031	0.021
1.1–4	0.082	0.074
1.2–1.2	0.082	0.043
1.2–1.3	0.052	0.043
1.2–2	0.053	0.053
1.2–3	0.093	0.064
1.2–4	0.124	0.202
1.3–1.3	0.000	0.000
1.3–2	0.021	0.000
1.3–3	0.021	0.000
1.3–4	0.010	0.106
2–2	0.021	0.000
2–3	0.041	0.011
2–4	0.062	0.064
3–3	0.010	0.000
3–4	0.072	0.043
4–4	0.134	0.128

[a]Northern Croatians: observed homozygosity = 25.8%; expected homozygosity (unbiased) = 21.2%; HWE — Homozygosity test (P = 0.268), Likelihood ratio test (P = 0.893), Exact test (P = 0.804).
[b]Southern Croatians: observed homozygosity = 19.2%; expected homozygosity (unbiased) = 24.1%; HWE — Homozygosity test (P = 0.261), Likelihood ratio test (P = 0.278), Exact test (P = 0.634).

Table 4
HLA-DQA1 observed allele frequencies in 97 unrelated Northern Croatians (N.C.) and 94 unrelated Southern Croatians (S.C.)

Allele	Frequency N.C.	Frequency S.C.
1.1	0.108	0.133
1.2	0.268	0.266
1.3	0.057	0.096
2	0.119	0.064
3	0.139	0.070
4	0.309	0.372

Table 5
Tests for independence on PM loci

	N.C.	S.C.
LDLR		
Obs. homozygosity	45.9%	52.5%
Exp. homozygosity[a]	51.3%	51.5%
Homozygosity test[b]	0.290	0.849
Likelihood ratio test[b]	0.296	1.000
Exact test[b]	0.296	1.000
GYPA		
Obs. homozygosity	42.9%	47.5%
Exp. homozygosity[a]	49.8%	51.7%
Homozygosity test[b]	0.168	0.399
Likelihood ratio test[b]	0.154	0.423
Exact test[b]	0.212	0.423
HBGG		
Obs. homozygosity	56.1%	52.5%
Exp. homozygosity[a]	50.4%	49.8%
Homozygosity test[b]	0.255	0.585
Likelihood ratio test[b]	0.291	0.688
Exact test[b]	0.291	0.688
D7S8		
Obs. homozygosity	56.1%	55.5%
Exp. homozygosity[a]	54.8%	54.2%
Homozygosity test[b]	0.788	0.799
Likelihood ratio test[b]	0.824	0.832
Exact test[b]	0.824	0.832
Gc		
Obs. homozygosity	42.9%	41.6%
Exp. homozygosity[a]	44.3%	44.9%
Homozygosity test[b]	0.777	0.506
Likelihood ratio test[b]	0.150	0.657
Exact test[a]	0.284	0.757

[a]Expected homozygosity is an unbiased estimate.
[b]These values are probability values.

Table 6
D1S80 observed allele frequencies in 98 unrelated Northern Croatians (N.C.) and 102 unrelated Southern Croatians (S.C.)

Allele	Frequency[a] N.C.	Frequency[b] S.C.
16	0.000	0.000
17	0.010	0.000
18	0.198	0.216
19	0.010	0.025
20	0.015	0.015
21	0.020	0.005
22	0.041	0.044
23	0.010	0.029
24	0.352	0.387
25	0.071	0.049
26	0.020	0.010
27	0.010	0.015
28	0.051	0.059
29	0.056	0.039
30	0.015	0.020
31	0.087	0.049
32	0.000	0.005
33	0.005	0.015
34	0.005	0.000
35	0.000	0.000
36	0.005	0.010
37	0.010	0.005
)41	0.005	0.005

[a]Northern Croatians: observed homozygosity = 14.3%; expected homozygosity (unbiased) = 18.1%; HWE — Homozygosity test (P = 0.323), Likelihood ratio test (P = 0.915), Exact test (P = 0.978).
[b]Southern Croatians: observed homozygosity = 23.5%; expected homozygosity (unbiased) = 20.7%; HWE — Homozygosity test (P = 0.483), Likelihood ratio test (P = 0.439), Exact test (P = 0.666).

An interclass correlation analysis was performed to determine if there were any detectable departures from expectations of independence between any pairs of loci. A departure from independence was observed between HBGG and D7S8 in the northern Croatian population database (P = 0.001) (Table 7). This single departure from independence was the only observation out of a total of 42 pair-wise comparisons. This number of departures is no more than would be expected.

To confirm whether or not there is detectable deviation from independence, a second test for association that considers all seven loci was used. This test determines whether or not the observed variance (S^2_k) of the number of heterozygous loci in the population sample is within its confidence interval under the assumption of independence [15,16]. The Croatian sample populations did not show evidence of association for the seven PCR-based loci using this criterion ($S^2_{k\ \text{NORTHERN CROATIAN}}$ = 1.212; 95% confidence interval is 1.132–1.954; $S^2_{k\ \text{SOUTHERN CROATIAN}}$ = 1.759; 95% confidence interval is 1.148–1.994). The results suggest that there is little evidence for departure from independence for

Table 7
Two locus interclass correlation test for HLA-DQA1, PM, and D1S80 loci for unrelated Croatians

	N. Croatians	S. Croatians
LDLR/GYPA	0.737	0.518
LDLR/HBGG	0.194	1.000
LDLR/D7S8	0.230	1.000
LDLR/Gc	1.000	0.579
LDLR/DQA1	0.773	0.903
GYPA/HBGG	0.351	0.788
GYPA/D7S8	0.522	0.830
GYPA/Gc	0.680	0.815
GYPA/DQA1	0.943	0.760
HBGG/D7S8	0.001[a]	0.393
HBGG/Gc	0.646	0.414
HBGG/DQA1	0.834	0.800
D7S8/Gc	0.806	0.892
D7S8/DQA1	0.739	0.602
Gc/DQA1	0.127	0.158
D1S80/LDLR	0.795	0.282
D1S80/GYPA	0.524	0.857
D1S80/HBGG	0.968	0.947
D1S80/D7S8	0.655	0.782
D1S80/Gc	0.259	0.259
D1S80/DOA1	0.520	0.242

[a]Deviation at $P = 0.05$ level.

the seven PCR-based loci in the northern and southern Croatian population databases.

When the allele frequency distributions for each locus were compared between northern and southern Croatians, only the HLA-DQA1 locus was statistically different ($P = 0.037$) (Table 8). Furthermore, the Croatian populations were compared with U.S. Caucasians; again, there was statistical difference only at the HLA-DQA1 locus ($P < 10^{-3}$). In a further comparison, the northern Croatians were compared with the US Caucasians at the HLA-DQA1 locus [19] and no difference was observed ($P = 0.098$). However, the southern Croatians were

Table 8
G-statistic test (*P* values) for homogeneity on PM, HLA-DQA1, D1S80 allele distributions between Northern and Southern Croatians and between a U.S. Caucasian sample population[a]

Locus	Croatians	Croatians/U.S. Caucasians
LDLR	0.907 ± 0.009	0.508 ± 0.016
GYPA	0.126 ± 0.011	0.222 ± 0.013
HBGG	0.314 ± 0.015	0.162 ± 0.012
D7S8	0.923 ± 0.008	0.573 ± 0.016
Gc	0.709 ± 0.014	0.448 ± 0.016
HLA-DQA1	0.037 ± 0.006	10^{-3}
D1S80	0.793 ± 0.013	0.207 ± 0.013

[a]Population data from this study and Budowle et al. [19].

different statistically when compared with U.S. Caucasians at the HLA-DQA1 locus ($P = 0.001$). Overall, there is little difference between the Croatian databases and U.S. Caucasians for the loci studied.

One could argue that the difference between HLA-DQA1 in southern Croatians might suggest that the use of general population databases for deriving allele frequency estimates is inappropriate. However, the data suggest the opposite. First, this difference may be due to sampling and probably is not real. Second, the confidence intervals for the HLA-DQA1 allele frequencies, as determined according to Goodman [20], overlap each other in northern Croatians, southern Croatians, and U.S. Caucasians (data not shown). Thus, the allele frequencies are not substantially different. Third, the final estimate of a multiple locus profile frequency would not be substantially different among the three databases.

In conclusion, Croatian population databases for seven PCR-based loci have been established. This study has shown that the distribution of the genotype frequencies for all the loci meet HWE and there is little evidence for departures from independence of alleles across the loci. These data, by employing the product rule under the assumption of independence, can be used to derive valid estimates of multiple locus profile frequencies for forensic applications in Croatia. Moreover, these data support the notion that there is little anticipated difference in DNA profile frequency estimates among Caucasian population groups for the loci LDLR, GYPA, HBGG, D7S8, Gc, HLA-DQA1, and D1S80.

Acknowledgments

This is publication number 96-04 of the Laboratory Division of the Federal Bureau of Investigation. Names of commercial manufacturers are provided for identification only, and inclusion does not imply endorsement by the Federal Bureau of Investigation.

References

[1] B. Budowle, J. Woller, B. Koons, S. Furedi, J.D. Errera and Z. Padar, Hungarian population data on seven PCR-based loci. *J. Forensic Sci.* (in press).

[2] J. Woller, Z. Padar and S. Furedi, Hungarian population data on the loci HLA-DQA1, LDLR, GYPA, HBGG, D7S8, and Gc. *Int. J. Leg. Med.* (submitted for publication).

[3] J. Woller, S. Furedi and Z. Padar, AMPFLP-analysis of the VNTR loci D1S80 and ApoB in Hungary. *Int. J. Leg. Med.*, 107 (1995) 273–274.

[4] C.T. Comey, B.W. Koons, K.W. Presley, J.B. Smerick, C.A. Sobieralski, D.M. Stanley and F.S. Bacchtel, DNA extraction strategies for amplified fragment length polymorphism analysis. *J. Forensic Sci.*, 39 (1994) 1254–1269.

[5] J.S. Waye, L. Presley, B. Budowle, G.G. Shutler and R.M. Fourney, A simple method for quantifying human genomic DNA in forensic specimen extracts. *Biotechniques*, 7 (1989) 852–855.

[6] B. Budowle, F.S. Baechtel, J.B. Smerick, K.W. Presley, A.M. Giusti, G. Parsons, N. Alevy and R. Chakraborty, D1S80 population data in African Americans, Caucasians, Southeastern Hispanics, Southwestern Hispanics, and Orientals. *J. Forensic Sci.*, 40 (1995) 38–44.

[7] A. Edwards, H. Hammond, L. Jin, C.T. Caskey and R. Chakraborty, Genetic variation at five trimeric and tetrameric repeat loci in four human population groups. *Genomics*, 12 (1992) 241–253.

[8] R. Chakraborty, P.E. Smouse and J.V. Neel, Population amalgamation and genetic variation: observations on artificially agglomerated tribal populations of Central and South America. *Am. J. Hum. Genet.*, 43 (1988) 709–725.

[9] M. Nei and A.K. Roychoudhury, Sampling variances of heterozygosity and genetic distance. *Genetics*, 76 (1974) 379–390.

[10] M. Nei, Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, 89 (1978) 583–590.

[11] R. Chakraborty, M. Fornage, R. Guegue and E. Boerwinkle, Population genetics of hypervariable loci: analysis of PCR-based VNTR polymorphism within a population. In T. Burke, G. Dolf, A.J. Jeffreys and R. Wolff (eds.), *DNA Fingerprinting: Approaches and Applications*, Birkhauser Verlag, Berlin, 1991, pp. 127–143.

[12] B.S. Weir, Independence of VNTR alleles defined by fixed bins. *Genetics*, 130 (1992) 873–887.

[13] S.W. Guo and E.A. Thompson, Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*, 48 (1992) 361–372.

[14] S. Karlin, E.C. Cameron and P.T. Williams, Sibling and parent-offspring correlation estimation with variable family size. *Proc. Natl. Acad. Sci. USA*, 78 (1981) 2664–2668.

[15] A.H.D. Brown, M.W. Feldman and E. Nevo, Multilocus structure of natural populations of *Hordeum spontaneum*. *Genetics*, 96 (1980) 523–536.

[16] R. Chakraborty, Detection of nonrandom association of alleles from the distribution of the number of heterozygous loci in a sample. *Genetics*, 108 (1984) 719–731.

[17] D.A. Roff and P. Bentzen, The statistical analysis of mitochondrial DNA polymorphisms: χ^2 and the problem of small samples. *Mol. Biol. Evol.*, 6 (1989) 539–545.

[18] R.R. Sokal and F.J. Rohlf, *Biometry*, 2nd Edition, W.H. Freeman and Company, San Francisco, 1981.

[19] B. Budowle, J.A. Lindsey, J.A. DeCou, B.W. Koons, A.M. Giusti and C.T. Comey, Validation and population studies of the loci LDLR, GYPA, HBGG, D7S8, and Gc (PM loci), and HLA-DQα using a multiplex amplification and typing procedure. *J. Forensic Sci.*, 40 (1995) 45–54.

[20] L. Goodman, On simultaneous confidence intervals for multinomial proportions. *Technometrics*, 7 (1965) 247–254.