Brief Communications

Allele Frequencies of Six Highly Polymorphic DNA Loci in the Croatian Population

I. DRMIĆ,[1] M.S. SCHANFIELD,[2] S. ANDJELINOVIĆ,[1] R. GALAVOTTI,[3] M.D. GOJANOVIĆ,[1] E. TRABETTI,[3] D. MARASOVIĆ,[1] D. PRIMORAC,[1,4] AND P.F. PIGNATTI[3]

Abstract The allele frequency distributions in a series of Croats were analyzed for six unlinked polymorphic DNA loci: *THO1*, *FESFPS*, *VWA01*, *APOB*, *D1S80*, and *D17S5*. The allele frequencies were determined for 100 unrelated genomic DNA samples. The observed heterozygote frequencies of the loci ranged from 0.63 to 0.76; however, the expected heterozygosity ranged from 0.68 to 0.82, with only *D17S5* having a significant excess of homozygous phenotypes ($p < 0.001$). The excess homozygosity seen in the *D17S5* system may be due to allelic drop-out and warrants further technical analysis of that system, given the uniform lack of significant deviation in the other five systems. The forensic usefulness of these systems can be measured using two different statistics: the power of discrimination and the likelihood of a coincidental match. The power of discrimination ranged from 0.85 to 0.94 for the 6 systems with the combined likelihood of a coincidental match based on these 6 systems of 1 in 3.6 million, or slightly less than the population of Croatia. A second, more conservative estimator of the likelihood of a match is based on the most common phenotype for each system. If someone had the most common phenotype for each of the 6 systems, the chance of a coincidental match would be approximately 1 in 64,000. For paternity testing the usefulness of a system is measured by the average power of exclusion or (1 − power of exclusion), the random man not excluded. The average power of exclusion, based on observed heterozygosity, ranged from 0.33 to 0.53, and the average power of exclusion based on the expected heterozygosity ranged from 0.39 to 0.64. The combined average power of exclusion was 99.2% for these 6 systems,

[1] Department of Pathology and Forensic Medicine, Split University Hospital, Spinciceva 1, 21000 Split, Croatia.
[2] Analytical Genetic Testing Center Inc., Denver, CO.
[3] Institute of Biology and Genetics, University of Verona, Verona, Italy.
[4] Department of Pediatrics, University of Connecticut Health Center, Farmington, CT.

Human Biology, October 1998, v. 70, no. 5, pp. 949–957.

KEY WORDS: PCR, INDIVIDUAL IDENTIFICATION, DNA POLYMORPHISM, PATERNITY TESTING, POPULATION GENETICS, CROATIA

using the expected heterozygosity. Based on the results of testing these six systems, there is no significant substructuring within the southern Croatian populations, and these systems provide a useful tool for forensic, paternity, and anthropological applications.

The availability of highly polymorphic DNA loci with large numbers of alleles and high heterozygosity, detected by means of polymerase chain reaction (PCR) testing, is becoming an important tool for forensic, paternity, and anthropological studies. However, for this technology to be useful in a specific population, the collection of a population database is required. To this end, allele and genotype frequencies for six amplified fragment length polymorphisms (AFLPs)—three tetranucleotide short tandem repeat loci [*THO1* (11p15.5) (Puers et al. 1993), *FESFPS* (15q25–qter) (Polymeropoulos et al. 1991), *VWA01* (formerly *VWF*) (intron 40 of the von Willebrand gene, 12p12–pter) (Peake et al. 1990)] and three large tandem repeat loci [the 15*n*-repeat 3′ hypervariable region at *APOB* (2p24–p23) (Boerwinkle et al. 1989), the 16*n*-repeat *D1S80* (1p26–35) (Kasai et al. 1990), and the 70*n*-repeat *D17S5* (17p13.3) (Horn et al. 1989)]—were generated for a southern Croatian population sample.

Materials and Methods

Peripheral blood samples were collected from 100 unrelated healthy individuals whose parents were born in southern Croatia (Split region). Individuals were identified by a questionnaire that provided information on family history for two generations. All potential relatives, such as brothers, sisters, parents, and children, were eliminated from the study population.

DNA was extracted according to the method of Sambrook et al. (1989) and quantitated at 260 nm using a JASCO 7800 Spectrophotometer.

The primer sequences used for PCR are those given in the references cited in the introduction. The PCR reactions consisted of 200–700 ng DNA, 0.15 U Super*Taq* DNA polymerase (HT Biotechnology Ltd., United Kingdom), 12.5 pmol of each primer, and 200 m*M* of each dNTP in 0.01 *M* Tris-HCl (pH 9.0), 1.5 m*M* $MgCl_2$, 0.05 *M* KCl, 0.01% gelatin, and 0.1% Triton X-100.

Amplification was performed for 30 cycles in a Perkin Elmer Thermocycler according to the published conditions given for the 6 loci. Fragments were separated by electrophoresis on 20-cm 8% acrylamide gels (*APOB*, *D1S80*), 40-cm 8% polyacrylamide gels (*THO1*, *FESFPS*), 10% polyacrylamide gels (*VWA01*), or 8.5-cm 2% agarose gels (*D17S5*). PCR products were visualized using ethidium bromide fluorescence.

Allele sizes for *APOB* and *D17S5* were calculated by a computer program that interpolated the observed fragment migration value in a molecular weight marker standard curve obtained from appropriate DNA markers co-

electrophoresed in the same gel (NX174-*Hae*III DNA fragments as size marker). Allele assignments for *FESFPS*, *THO1*, *VWA01*, and *D1S80* were determined by comparison to either allele ladders or samples with known repeat numbers.

All individuals were typed using Puers et al.'s (1993) nomenclature for *THO1*; Polymeropoulos et al.'s (1991) nomenclature for *FESFPS* with the *C1–C6* alleles converted to repeat numbers, as is currently recommended; and Peake et al.'s (1990), nomenclature for *VWA01*. For *APOB* a small series of samples was tested using two methods: the interpolation method described in the preceding paragraph and comparison to repeat ladders (Latorra et al. 1994). The second testing was done at the Analytical Genetic Testing Center Inc. (Denver, Colorado). This allowed for the conversion from allele number to the more widely used repeat numbers. *D1S80* nomenclature was determined by repeat ladder (AmpliFLP *D1S80* Allelic Ladder, Perkin Elmer, Applied Biosystems Division, Foster City, California). Horn et al. (1989) provided the nomenclature for *D17S5*.

The results were tested for Hardy-Weinberg equilibrium using the global chi-square test by comparing observed and expected genotype frequencies. All expected values less than 1.0 were pooled to form a residual group to minimize large chi-square values resulting from sampling events. Degrees of freedom were determined by subtracting the number of estimated alleles from the number of tested classes. The overall heterozygosity chi-square was determined using the sum of the homozygous phenotypes versus expected homozygosity (Σp_i^2), where p_i are the individual allele frequencies. The expected heterozygosity is equal to 1 minus the expected homozygosity. A chi-square test compares observed homozygosity and heterozygosity to expected values and has one degree of freedom. Finally, substructuring was tested using the T statistic of Robertson and Hill (1984), as suggested by Crow et al. (1996).

Exclusionary power or power of discrimination was calculated as described by Crow et al. (1996, p. 96) using the formula $2(\Sigma p_i^2)^2 - \Sigma p_i^4$ for multi-allelic systems, whereas the squared genotype array frequencies were used for the Polymarker system (see Discussion). Average power of exclusion was calculated for the highly polymorphic systems using the formula suggested by Brenner and Morris (1990):

$$\bar{A} = h^2(1 - 2hH^2), \tag{1}$$

where h is the heterozygosity and H is the homozygosity. Power of exclusion for the di- and tri-allelic systems found in the Polymarker system were calculated using the method of Dykes (1982).

Allele frequencies were estimated by gene counting. Statistical tests and gene counting were performed using Quattro Pro for Windows spreadsheets. Statistical significance of chi-square and T tests was determined after applying the Bonferroni correction for multiple tests (Westfall and Wolfinger 1997).

Results

Allele frequencies were determined in 100 Croatian individuals for 6 different DNA polymorphisms: *THO1*, *FESFPS*, *VWA01*, *APOB*, *D1S80*, and *D17S5*. Table 1 gives the allele frequencies that were observed in the Croatian population sample for the six DNA polymorphisms and various measures of heterogeneity and genetic information. Because of the large number of phenotypes for each system, individual phenotype data are not presented.

For all systems the common European alleles were detected. With the exception of *VWA01* and *D17S5*, all the systems showed good agreement with Hardy-Weinberg equilibrium and the two substructuring tests. *VWA01* appears to have significant *T* results, primarily because of the occurrence of a single homozygous 18-repeat individual. The pooled substructure test indicates an extremely good fit with expectation. Given that there was only a single exception, it appears that the *VWA01* result may be due to sampling variation. In contrast, all the tests of distortion at the *D17S5* locus are significantly outside expectation, even after applying the Bonferroni corrections.

The usual method of determining the forensic usefulness of a genetic system is the power of discrimination. However, if the worst case situation is of interest, then the most common phenotypes (MCPs) provide useful numbers. The observed versus expected MCPs for the six systems are presented in Table 2. This table also yields an observed combined MCP (CMCP) of 0.004%, or a likelihood of approximately 1 in 23,000 that a Croatian would have the MCP in all 6 systems. This is significantly more frequent than the expected CMCP given the allele frequencies. The expected CMCP is 0.002%, or a likelihood of approximately 1 in 64,000 that a Croatian would have the MCP in all 6 systems. This difference between the two calculations appears to be totally due to the highly significant difference between the observed and expected values for *D17S5* ($\chi^2 = 9.062$, $p < 0.001$) because none of the other differences were significant. No test for linkage disequilibrium was performed.

The average power of discrimination can be obtained by multiplying the individual system (1 − the power of discrimination) values together. The average likelihood that two people will match by chance is 1 in 3.6 million in this case.

Discussion

Allele frequencies for six PCR-based hypervariable DNA polymorphisms in a Croatian population sample have been generated. A detailed comparison of these results to other population studies is ongoing. There is limited data for PCR-based systems in the Croatian population. Keys et al. (1996) recently looked at the distribution of *D1S80*, *HLA-DQA1*, and Polymarker in northern and southern Croatians. Polymarker (Perkin-Elmer, Applied Bio-

Table 1. Allele Frequencies for Six DNA Polymorphisms in the Croatian Population ($N = 100$)

	THO1		*FESFPS*		*VWA01*		*APOB*		*D1S80*		*D17S5*	
Statistic	*Repeats*	*Frequency*	*Repeats*[a]	*Frequency*	*Repeats*[a]	*Frequency*	*Repeats*[a]	*Frequency*	*Repeats*	*Frequency*	*Repeats*	*Frequency*
	6	0.280	7	0.000	6 (1)	0.085	247 (1)	0.000	18	0.260	1	0.040
	7	0.125	8 (C6)	0.015	7 (2)	0.525	267 (2)	0.000	19	0.005	2	0.150
	8	0.150	9 (C5)	0.000	8 (3)	0.015	28 (3)	0.005	20	0.005	3	0.265
	9	0.125	10 (C4)	0.230	9 (4)	0.010	30 (4)	0.075	21	0.010	4	0.265
	9.3	0.240	11 (C3)	0.455	10 (5)	0.100	32 (5)	0.090	22	0.015	5	0.065
	10	0.080	12 (C2)	0.230	11 (6)	0.150	34 (6)	0.260	23	0.030	6	0.025
	11	0.000	13 (C1)	0.070	12 (7)	0.095	36 (7)	0.295	24	0.365	7	0.010
	Total	1.000	14	0.000	13 (8)	0.010	38 (8)	0.115	25	0.080	8	0.030
			Total	1.000	14 (9)	0.010	40 (9)	0.015	26	0.020	9	0.060
					15 (10)	0.000	42 (10)	0.020	27	0.005	10	0.045
					Total	1.000	44 (11)	0.025	28	0.065	11	0.020
							46 (12)	0.005	29	0.050	12	0.025
							48 (13)	0.005	30	0.005	13	0.000
							50 (14)	0.020	31	0.045	14	0.000
							52 (15)	0.060	32	0.005	Total	1.000
							54 (16)	0.005	33	0.005		
							56 (17)	0.000	34	0.005		
							58 (18)	0.005	35	0.000		
							Total	1.000	36	0.000		
									37	0.020		
									38	0.005		
									39	0.000		
									Total	1.000		
Global Hardy-Weinberg equilibrium		22.214		4.573		7.152		20.906		8.534		35.733[b]
d.f.		15		6		9		9		0		15
p		0.105		0.600		0.621		0.014		(c)		0.0022
Observed H_t		0.760		0.640		0.670		0.740		0.750		0.63
Expected H_t		0.804		0.682		0.675		0.813		0.791		0.824
χ^2 (1 d.f.)		1.219		0.815		0.012		3.511		1.056		25.894[b]
p		0.270		0.367		0.913		0.061		0.306		0.000[b]
T		1.703		0.957		3.307[b]		1.633		1.351		6.622[b]
p (one-tailed)		0.044		0.169		0.00049[b]		0.051		0.089		0.000[b]

a. Repeat designation; old allele designation in parentheses.
b. Exceeds Bonferroni corrected significance level 0.0042 for χ^2 and 0.0083 for T.
c. Not calculated.

Table 2. Forensic and Paternity Indexes of Usefulness

Locus	*PD*[a]	*Observed MCP*[b]	*Expected MCP*	*RMNE*[c]
Present AFLP data				
APOB	0.058	0.170	0.153	0.377
D1S80	0.073	0.210	0.190	0.416
D7S5	0.052	0.150	0.140	0.356
FESFPS	0.154	0.220	0.209	0.599
THO1	0.065	0.130	0.134	0.394
VWA01	0.134	0.280	0.276	0.609
Combined	0.00000013	0.000004	0.000003	0.00802
1 in match	3,428,201	23,319	31,698	125
Average A[d]		99.20%		
Published data from southern Croatia (Keys et al. 1996)				
LDLR	0.384	0.475	0.482	0.817
GYPA	0.385	0.525	0.480	0.818
HBG2	0.375	0.475	0.500	0.813
D7S8	0.400	0.446	0.456	0.824
GC	0.269	0.356	0.365	0.712
HLA-DQA1	0.095	0.128	0.139	0.482
Combined	0.00057	0.00240	0.00266	0.15348
1 in match	1751	416	375	7
Average A[d]		84.65%		
Combined	1.67×10^{-10}	1.03×10^{-7}	8.4068×10^{-8}	0.0012
1 in match	6,004,371,136	9,709,440	11,895,087	812
Average A[d]		99.88%		

a. Chance that two samples will match by chance; PD, power of discrimination.

b. Most common phenotype for locus.

c. Random man not excluded; likelihood that a nonfather will match by chance.

d. Average power of all the tests to exclude falsely accused men.

systems Division, Foster City, California) is a forensic product that allows for the multiplex PCR detection of five sequence polymorphisms. The polymorphisms are detected using a reverse dot blot technology. In order of appearance from largest to smallest, the fragments are glycophorin A (*GYPA*) (*MN* blood group, two alleles *A*, *B*, same as *M* and *N*); low-density lipoprotein receptor (*LDLR*) (two alleles, *A* and *B*); hemoglobin (*HBGG*) (three alleles, *A*, *B*, and *C*); *D7S8* and anonymous sequence polymorphism (two alleles, *A* and *B*); and group-specific component/vitamin D transport protein (*GC/VDBG*) (three alleles, *A*, *B*, and *C*, which are the same as *GC*2*, **1S*, and **1F*, respectively).

Comparison of the Keys et al. (1996) data for northern and southern Croatia for *D1S80* indicates no significant differences between northern and southern Croatia and no significant differences between the southern Croatians studied by Keys et al. (1996) and those studied here.

The results obtained for *D17S5* indicate a highly significant deficiency of heterozygotes by all the tests used. This could be due to insufficient detection of large alleles that might not amplify well (Deka et al. 1992). It has been suggested that incomplete resolution of alleles can also be a problem (Chakraborty and Jin 1992; Chakraborty et al. 1992); however, given the large size of the repeat versus the fragment size, this is highly unlikely. At the largest fragment size (approximately 1080 bp), a single repeat difference (70 bp) represents a 6.5% difference, which is large enough to be detected by even insensitive technology. Therefore it seems reasonable that additional studies are needed to verify either technical problems with the test or significant substructuring at only this locus.

The data provided by Keys et al. (1996) allow for the addition of *HLA-DQA1* and Polymarker data to the six AFLP systems used here. Although this is not a shared data set and the data cannot be tested for linkage disequilibrium, the data provide insight into the usefulness of a broad-based forensic identification system. Table 2 provides the individual indexes for the six AFLP systems plus those for *HLA-DQA1* and Polymarker taken from Keys et al. (1996). The addition of *HLA-DQA1* and Polymarker to the six AFLP loci increases the total likelihood of a coincidental match to 1 in 6.3 billion, whereas the more conservative CMCP frequency goes to approximately 1 in 23 million.

Unfortunately, because of the low information value of Polymarker for paternity purposes, the combined average power of exclusion only goes from 99.2% to 99.88%, with a random man not excluded probability of 1 in 812.

The PCR-based DNA typing systems used here appear to provide a rapid and reliable typing system for individual identification, parentage testing, forensic analysis, and medical applications. One of the intended forensic purposes is to aid the DNA typing of war victims from mass graves (Primorac et al. 1996). Limitations on the number of genetic markers available have

hampered our ability to identify some samples. The additional markers should allow for better laboratory services in the region.

Because of the brevity of this paper, there have been no attempts to look for linkage disequilibrium among the loci, primarily due to a lack of software when this study was undertaken. Furthermore, we decided to limit the length and, given the limited scope of this paper, it was not possible to look at many aspects of forensic validation and microvariation that might have an impact on day-to-day applications of these systems. It is apparent that the systems are all polymorphic and potentially useful for identification and parentage testing purposes, but specific additional studies may be needed to fully implement all these systems.

Acknowledgments We thank Vesna Capkun for helping us with the statistical analysis of the data. This project was partially funded by an Italian CNR Target Project ("Cultural Goods"), awarded to P.F. Pignatti.

Received 7 July 1997; revision received 25 November 1997.

Literature Cited

Boerwinkle, E., W. Xiong, E. Fourest et al. 1989. Rapid typing of tandemly repeated hypervariable loci by the polymerase chain reaction: Application to the apolipoprotein B 3′ hypervariable region. *Proc. Natl. Acad. Sci. USA* 86:212–216.

Brenner, C., and J. Morris. 1990. Paternity index calculations in single-locus hypervariable DNA probes: Validation and other studies. In *The International Symposium on Human Identification, 1989: Data Acquisition and Statistical Analysis for DNA Laboratories.* Madison, WI: Promega Corp., 21–53.

Chakraborty, R., and L. Jin. 1992. Heterozygote deficiency, population substructure, and their implications in DNA fingerprinting. *Hum. Genet.* 88:267–272.

Chakraborty, R., M. De Andrade, S.P. Daiger et al. 1992. Apparent heterozygote deficiencies observed in DNA typing data and their implications in forensic applications. *Ann. Hum. Genet.* 56:45–57.

Crow, J.F., M.A. Berger, D.H. Kaye et al. 1996. *The Evaluation of Forensic DNA Evidence.* Washington, DC: National Research Council, National Academy Press.

Deka, R., S. DeCroo, Yu.L. Mei et al. 1992. Variable number of tandem repeat (VNTR) polymorphism at locus *D17S5* (*YNZ22*) in four ethnically defined human populations. *Hum. Genet.* 90:86–90.

Dykes, D.D. 1982. The use of frequency tables in parentage testing. In *Probability of Inclusion in Paternity Testing: A Technical Workshop,* H. Silver, ed. Arlington, VA: American Association of Blood Banks, 15–44.

Horn, G.T., B. Richards, and K.W. Klinger. 1989. Amplification of a highly polymorphic VNTR segment by the polymerase chain reaction. *Nucleic Acids Res.* 17:2140.

Kasai, K., Y. Nakamura, and R. White. 1990. Amplification of a VNTR locus (pMCT118) by the polymerase chain reaction (PCR) and its application to forensic science. *J. Forensic Sci.* 35:1196–1200.

Keys, K.M., B. Budowle, S. Andelinovic et al. 1996. Northern and southern Croatian population data on seven PCR-based loci. *Forensic Sci. Int.* 81:191–199.

Latorra, D., C.M. Stern, and M.S. Schanfield. 1994. Characterization of human AFLP systems apolipoprotein B, phenylalanine hydroxylase, and *D1S80*. *PCR Meth. Appl.* 3:351–358.

Peake, I.R., D. Bowen, P. Bignell et al. 1990. Family studies and prenatal diagnosis in severe von Willebrand disease by polymerase chain reaction amplification of a variable number of tandem repeat region of the von Willebrand factor gene. *Blood* 76:555–561.

Polymeropoulos, M.H., D.S. Rath, H. Xiao et al. 1991. Tetranucleotide repeat polymorphism at the human c-fes/fps proto-oncogene (FES). *Nucleic Acids Res.* 19:4018.

Primorac, D., S. Andelinovic, M. Definis Gojanovic et al. 1996. Identification of war victims from mass graves in Croatia, Bosnia, and Herzegovina by the use of standard forensic methods and DNA typing. *J. Forensic Sci.* 41:891–894.

Puers, C., H.A. Hammond, C.T. Caskey et al. 1993. Identification of repeat sequence heterogeneity at the polymorphic short tandem repeat locus HUM *THO1* [AATG]*n* and reassignment of alleles in population analysis by using a locus specific allelic ladder. *Am. J. Hum. Genet.* 53:953–958.

Robertson, A., and W. Hill. 1984. Inbreeding coefficients. *Genetics* 107:703–718.

Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. *Molecular Cloning: A Laboratory Manual,* 2d ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 9.17–9.19.

Westfall, P.H., and R.D. Wolfinger. 1997. Multiple tests with discrete distributions. *Am. Stat.* 51:1–8.