

BLOOD GROUPS IN A LARGE SAMPLE FROM THE CITY OF SÃO PAULO (BRAZIL): ALLELE AND HAPLOTYPE FREQUENCIES FOR MNSs, KELL-CELLANO, Rh AND ABO SYSTEMS

Nativa N. Salaru¹ and Paulo A. Otto²

ABSTRACT

In the present paper we estimate the allele and, for the case of complex loci, haplotype frequencies for the following blood group systems in a large population sample consisting of unrelated individuals: MNSs, Kell-Cellano, Rh (5-6 sera) and ABO. The individuals of our sample, studied at the Departamento de Medicina Legal da Faculdade de Medicina da Universidade de São Paulo, were assembled into four major racial groups: whites, blacks, mongoloids and mulattoes. Besides comparing allele and phenotype frequencies among the racial subsamples, we also calculated the degree of racial admixture in the subsample of individuals classified as mulattoes. We also present a simplified method for estimating the frequencies of A_1 and A_2 alleles among A and AB individuals tested with anti-A and anti-B sera and anti- A_1 lectin.

INTRODUCTION

Salzano (1971) pointed out that the available data in relation to blood group system polymorphisms in Brazilian populations was unsatisfactory both in number and in quality and stated that new investigations were needed even in relation to the ABO system, to ascertain on firmer bases the degree of genetic heterogeneity of the Brazilian populations. This author stressed that in the presentation of the results a clear

¹ Departamento de Medicina Legal, Faculdade de Medicina, USP, Laboratório de Investigação Médica LIM-40, São Paulo, SP, Brasil.

² Departamento de Biologia, Instituto de Biociências, Universidade de São Paulo, Caixa Postal 11.461, 05499 São Paulo, SP, Brasil. Send correspondence to P.A.O.

indication of the sampling method employed and the criteria used for racial identification should be given, since their lack constitutes the major flaw found in many papers on this subject hitherto published.

The observation above and the fact that the situation has not changed much in the 17 years that elapsed since it was made, prompted us to publish the results that follow. In the present paper we estimate the allele and, for the case of complex loci, haplotype frequencies for the MNSs, Kell-Cellano, Rh and ABO blood group systems in a large population sample of unrelated individuals studied at the Departamento de Medicina Legal (Laboratório de Investigação Médica LIM-40) da Faculdade de Medicina da Universidade de São Paulo, in the city of São Paulo, Brazil.

MATERIAL AND METHODS

Sample description

All individuals belonging to our sample are unrelated. With a few exceptions, the data were obtained from unrelated couples belonging to trios of putative father, mothers and children. The sample sizes in relation to each blood group system are indicated in Tables I - XI.

Taking into account not only the skin pigmentation but other anthropological characteristics as well, such as hair type, nose, lip and skull shape, and slanting of the palpebral fissures, the individuals of our sample were assembled into four major racial groups: whites, blacks, mongoloids (Japanese) and mulattoes. This latter group was made up of mixed individuals who presented at least one negroid characteristic or declared themselves to be of mixed ancestry.

We also included a sample of individuals whose racial classification was not certain or for whom information about racial origin was not available in our records.

Determination of blood groups

For the determination of blood groups, mono-specific antisera obtained from Ortho Diagnostics (Johnson and Johnson) and from Biotest S.A. Com. Ind. were used throughout. All blood groups were assessed using the hemagglutination method in test tubes, following carefully the directions of the manufacturers.

For the determination of blood groups of the ABO system, anti-A, anti-B and anti-AB sera were used throughout (total of 4,931 tested individuals). A subsample of 609 individuals belonging to A and AB groups was also tested with an anti-A₁ lectin obtained from *Dolichos biflorus*. All individuals were tested not only by the direct method but by the reverse test as well. For the identification of naturally-occurring agglutinins we used pools of A, B and O red blood cells, collected at most 72 hours before the test was performed.

For the determination of the Rh phenotypes all 4,740 individuals examined for this system were tested with at least 5 anti-sera: anti-C, anti-c, anti-D, anti-E and anti-e. Individuals D^u(+) were classified as Rh(+). Part of the Rh total sample, consisting of 3,444 individuals, was also tested with anti-C^w.

MN blood groups were determined in 4,162 individuals and MNSs in 540, with the use of anti-M, anti-N, anti-S and anti-s sera. For the Kell-Cellano blood system, 1,811 individuals were tested with anti-K and anti-k sera.

All individuals belonging to the samples here described denied receiving any blood transfusion at least six months prior to the determination. The possibility of drug interference in the results was considered in each case, and dubious cases were excluded from the sample.

Statistical methods

For the estimation of allele and haplotype frequencies we used either standard counting methods (Mourant *et al.*, 1976) or numerical maximum-likelihood iterative procedures (Stevens, 1938; Fisher, 1946a,b).

For the determination of the frequencies of A₁ and A₂ alleles among A and AB individuals tested with anti-A and anti-B sera and anti-A₁ lectin, the following method was used: letting p₁, p₂, p = p₁ + p₂, q and r be the frequencies of the alleles A₁, A₂, A = A₁ + A₂, B and O, and n₁, n₂, n₃ and n₄ be the observed numbers of A₁, A₂, A₁B and A₂B individuals out of a panmictic sample of size N of A and AB individuals, then the expected values of A₁, A₂, A₁B and A₂B are respectively:

$$N(p_1^2 + 2p_1p_2 + 2p_1r)/T,$$

$$N(p_2^2 + 2p_2r)/T,$$

$$2Np_1q/T \quad \text{and}$$

$$2Np_2q/T,$$

where:

$$T = 2(p_1 + p_2) - (p_1 + p_2)^2;$$

treating p, q and r as constant values, since they are independently estimated from another sample from the same population, it comes out that the maximum likelihood function has the form:

$$L = \text{constant} + (n_1 + n_3) \ln p_1 + (n_2 + n_4) \ln (p - p_1) + n_1 \ln (2p - p_1 + 2r) + n_2 \ln (p - p_1 + 2r);$$

p_1 is then the solution of the equation:

$$dL/dp_1 = (n_1 + n_3)/p_1 - (n_2 + n_4)/(p - p_1) - n_1/(2p - p_1 + 2r) - n_2/(p - p_1 + 2r) = 0$$

and p_2 is obtained from:

$$p_2 = p - p_1.$$

For contrasting phenotype as well as allele frequencies among subsamples we used standard heterogeneity tests (Stevens, 1938, 1950; Elandt-Johnson, 1971; Beiguelman, 1972, 1977).

For the estimation of the percentage of white ancestry among the individuals of our sample classified as mulattoes we used the following formulae (Cavalli-Sforza and Bodmer, 1971):

$$a) m = (p_m - p_b)/(p_w - p_b),$$

used for calculating the degree of admixture, where p_m , p_b and p_w are the frequencies of the same allele among mulattoes, blacks and whites respectively;

$$b) \text{var}(m) = (\text{var}(p_m) + m^2 \text{var}(p_w) + (1 - m)^2 \text{var}(p_b))/(p_w - p_b)^2,$$

used for calculating the error variance of m ;

$$c) \bar{m} = \sum_i w_i m_i / \sum_i w_i,$$

for combining several estimates of m , where $w_i = 1/\text{var}(m_i)$;

$$d) X_{k-1}^2 = \sum_i (m_i - \bar{m})^2 / \text{var}(m_i),$$

for testing the heterogeneity among m values obtained from the analysis of k different loci.

RESULTS

Phenotype frequencies

Phenotype frequencies are shown in tables I - VI (absolute frequencies). Table

Table I - Individual loci (MN, Ss, Kk, Cc, Dd, and Ee): observed numbers of phenotypes. 1 = whites; 2 = mulattoes; 3 = blacks; 4 = mongoloids; 5 = without racial classification; 6 = total sample.

System	Phenotypes	1	2	3	4	5	6
MN	MM	555	114	36	13	515	1233
	MN	912	273	52	23	814	2074
	NN	367	110	30	11	337	855
	Total	1834	497	118	47	1666	4162
Ss	SS	50	5	2	0	11	68
	Ss	140	40	4	1	56	241
	ss	138	34	6	6	47	231
	Total	328	79	12	7	114	540
Kk	KK	1	0	0	0	4	5
	Kk	22	4	0	0	80	106
	kk	407	82	20	20	1171	1700
	Total	430	86	20	20	1255	1811
Cc	CC	429	65	2	17	284	797
	Cc	1067	283	52	33	817	2252
	cc	727	257	80	6	621	1691
	Total	2223	605	134	56	1722	4740
Dd	D	1989	550	120	56	1533	4248
	dd	234	55	14	0	189	492
	Total	2223	605	134	56	1722	4740
Ee	EE	58	17	3	6	33	117
	Ee	579	142	22	26	452	1221
	ee	1586	446	109	24	1237	3402
	Total	2223	605	134	56	1722	4740

I lists the observed numbers of phenotypes in relation to individual loci (MN, Ss, Kk, Cc, Dd, and Ee); Table II shows the absolute frequencies of ABO phenotypes. Table III shows the subsample of individuals belonging to A and AB groups from this system tested with anti-A₁ lectin. Tables IV - VI list the observed absolute frequencies of MNSs and Rh (classified with 5 or with 6 anti-sera, including anti-C^w) phenotypes.

Table II - ABO blood-groups: observed numbers of phenotypes. 1 = whites; 2 = mulattoes; 3 = blacks; 4 = mongoloids; 5 = without racial classification; 6 = total sample.

Phenotypes	1	2	3	4	5	6
O	1120	281	66	22	917	2406
A	867	221	38	17	655	1798
B	231	77	24	14	194	540
AB	85	22	3	6	71	187
Total	2303	601	131	59	1837	4931

Table III - Subsample of individuals belonging to A and AB groups from ABO system tested with anti-A₁ lectin: observed numbers of phenotypes. 1 = whites; 2 = mulattoes; 3 = blacks; 4 = mongoloids; 5 = without racial classification; 6 = total sample.

Phenotypes	1	2	3	4	5	6
A ₁	304	56	5	10	114	489
A ₂	49	13	3	0	15	80
A ₁ B	13	2	0	1	10	26
A ₂ B	10	3	1	0	0	14
Total	376	74	9	11	139	609

Table IV - MNSs blood-groups: observed numbers of phenotypes. 1 = whites; 2 = mulattoes; 3 = blacks; 4 = mongoloids; 5 = without racial classification; 6 = total sample.

Phenotypes	1	2	3	4	5	6
MMSS	25	4	1	0	3	33
MMSs	52	13	0	1	21	87
MMss	25	6	1	2	9	43
MNSS	19	1	1	0	6	27
MNSs	72	24	4	0	24	124
MNss	82	18	3	0	23	126
NNSS	6	0	0	0	2	8
NNSs	16	3	0	0	11	30
NNss	31	10	2	4	15	62
Total	328	79	12	7	114	540

Table V - Rh blood-groups: observed numbers of phenotypes with 5 sera (anti-C, anti-c, anti-D, anti-E and anti-e). 1 = whites; 2 = mulattoes; 3 = blacks; 4 = mongoloids; 5 = without racial classification; 6 = total sample.

Phenotypes	1	2	3	4	5	6
ccddee	209	49	8	0	168	434
ccD ee	178	117	56	1	195	547
ccddEe	0	0	1	0	4	5
ccD Ee	288	77	13	0	224	602
ccddEE	0	0	0	0	0	0
ccD EE	52	14	2	5	30	103
Ccddee	24	6	5	0	15	50
CcD ee	758	209	38	7	590	1602
CcddEe	1	0	0	0	2	3
CcD Ee	278	65	8	26	207	584
CcD EE	6	3	1	0	3	13
CCddee	0	0	0	0	0	0
CCD ee	417	65	2	16	269	769
CCD Ee	12	0	0	0	15	27
CCD EE	0	0	0	1	0	1
Total	2223	605	134	56	1722	4740

Allele and haplotype frequency estimates

Table VII lists the estimates of allele frequencies in relation to individual loci (MN, Ss, Kk, Cc, Dd and Ee) together with their respective standard errors.

Table VIII lists the estimates of ABO system allele frequencies together with their respective standard errors; results of goodness of fit tests are also given, since the estimates were obtained under the assumption of panmixia.

Estimates of A_1 and A_2 allele frequencies from the total subsample shown in Table III (using the method described in the section 'statistical methods') were respectively $p_1 = 0.1876$ and $p_2 = 0.0392$, using as constant values the estimates p , q and r obtained from the total sample (Table VIII). The adherence of the expected absolute frequencies to the observed ones using the above estimates was poor however ($\chi^2_3 = 10.65$; $P = 0.02-0.01$) though this could be explained by the fact that p , q and r are not exactly the same in the two samples.

Table IX shows the estimates of MNSs haplotype frequencies obtained in the

Table VI - Rh blood-groups: observed numbers of phenotypes with 6 sera (anti-C, anti-C^W, anti-c, anti-D, anti-E and anti-e). 1 = whites; 2 = mulattoes; 3 = blacks; 4 = mongoloids; 5 = without racial classification; 6 = total sample.

Phenotypes	1	2	3	4	5	6
ccddee	148	37	6	0	105	296
ccD ee	130	85	43	1	124	383
ccddEe	0	0	1	0	3	4
ccD Ee	216	49	10	0	141	416
ccddEE	0	0	0	0	0	0
ccD EE	43	6	1	3	21	74
Ccddee	17	5	5	0	11	38
CcD ee	575	157	24	3	415	1174
CcddEe	1	0	0	0	1	2
CcD Ee	229	43	6	21	131	430
CcD EE	5	3	1	0	1	10
CCddee	0	0	0	0	0	0
CCD ee	325	40	1	13	187	566
CCD Ee	12	0	0	0	11	23
CCD EE	0	0	0	1	0	1
CC ^W Dee	6	1	0	0	4	11
CC ^W DEe	5	3	0	0	5	13
C ^W cDEe	3	0	0	0	0	3
Total	1715	429	98	42	1160	3444

present study. Table X lists the estimates of Rh haplotype frequencies among the samples tested with 5 anti-sera; Table XI shows the estimates obtained with part of the Rh total sample tested with six anti-sera (anti-C, anti-c, anti-C^W, anti-D, anti-E and anti-e).

Direct estimates (that is, using the gene counting method) of the frequency of the allele C^W (in the series C, c, C^W) can be obtained straightforwardly from Table VI and were as follows: a) among whites: 0.004 ± 0.008 ; b) among mulattoes: 0.005 ± 0.002 ; c) among blacks and mongoloids: 0; d) among the individuals without racial classification and in the total sample: 0.004 ± 0.001 .

We also calculated the values of linkage disequilibrium values for MNSs and Rh system haplotypes; these values show, in modulus, a variation from 0.035 to 0.109 (MNSs) and from 0.002 to 0.212 (Rh), considering all groups.

Table VII - Estimates of allele frequencies for individual loci (MN, Ss, Kk, Cc, Dd and Ee). n = number of sampled individuals; p_i = allele frequency estimates (in each case the frequencies of the alternative allelomorph can be obtained from $q_i = 1 - p_i$); s.e. (p_i) = s.e. (q_i) = standard error of estimates.

Group	n	Alleles	p_i	s.e. (p_i)
Whites	1834	M	0.5513	0.0082
	328	S	0.3659	0.0188
	430	K	0.0279	0.0056
	2223	C	0.4330	0.0074
	2223	D	0.6756	0.0100
	2223	E	0.1563	0.0054
Mulattoes	497	M	0.5040	0.0159
	79	S	0.3165	0.0370
	86	K	0.0233	0.0115
	605	C	0.3413	0.0136
	605	D	0.6985	0.0194
	605	E	0.1455	0.0101
Blacks	118	M	0.5254	0.0325
	12	S	0.3333	0.0962
	20	K	0.0000	0.0000
	134	C	0.2090	0.0248
	134	D	0.6768	0.0409
	134	E	0.1045	0.0187
Mongoloids	47	M	0.5213	0.0515
	7	S	0.0714	0.0688
	20	K	0.0000	0.0000
	56	C	0.5982	0.0463
	56	D	1.0000	0.0000
	56	E	0.3393	0.0447
Without racial classification	1666	M	0.5534	0.0086
	114	S	0.3421	0.0314
	1255	K	0.0351	0.0037
	1722	C	0.4021	0.0084
	1722	D	0.6687	0.0114
	1722	E	0.1504	0.0061
Total sample	4162	M	0.5454	0.0055
	540	S	0.3491	0.0145
	1811	K	0.0320	0.0029
	4740	C	0.4057	0.0050
	4740	D	0.6778	0.0069
	4740	E	0.1535	0.0037

Table VIII - ABO allele frequencies. n = number of sampled individuals; p_i = allele frequency estimates using the maximum likelihood method or Bernstein's adjusted estimates; s.e. (p_i) = standard error of estimates obtained from the variance-covariance matrix of the iterative numerical method or through Li's explicit formulae; χ_1^2 = goodness of fit test (chi-squared value with 1 degree of freedom); the figures in parentheses were obtained using Stevens' approximation for the test.

Group	n	Alleles	p_i	s.e. (p_i)	χ_1^2
Whites	2303	A	0.2338	0.0067	1.33 (1.32)
		B	0.0710	0.0039	
		O	0.6952	0.0072	
Mulattoes	601	A	0.2284	0.0130	0.17 (0.16)
		B	0.0862	0.0083	
		O	0.6854	0.0143	
Blacks	131	A	0.1720	0.0245	1.04 (1.04)
		B	0.1095	0.0199	
		O	0.7184	0.0292	
Mongoloids	59	A	0.2171	0.0404	0.52 (0.51)
		B	0.1854	0.0377	
		O	0.5974	0.0488	
Without racial classification	1837	A	0.2219	0.0073	2.30 (2.28)
		B	0.0748	0.0044	
		O	0.7033	0.0080	
Total sample	4931	A	0.2268	0.0045	2.04 (2.02)
		B	0.0766	0.0027	
		O	0.6966	0.0049	

Degree of racial admixture among mulattoes from our sample

For the estimation of the percentage of white ancestry among the individuals of our sample classified as mulattoes we selected the haplotypes CDe from Rh and the allele B from ABO, on grounds that their frequencies were quite distinct among whites and blacks and their error variances small in relation to other markers, obtaining estimates of 0.6718 and 0.6052, respectively.

Table IX - Estimates of MNSs haplotype frequencies. n = number of sampled individuals; p_i = haplotype frequency estimates using the maximum likelihood method (figures in parentheses are the adjusted frequency estimates obtained with Mourant's counting method); s.e. (p_i) = standard error of estimates obtained from the variance-covariance matrix of the iterative numerical method (Newton-Raphson generalized procedure); χ^2_5 = goodness of fit test (chi-squared value with 5 degrees of freedom).

Group	n	Haplotypes	p_i	s.e. (p_i)	χ^2_5
Whites	328	MS	0.2649 (0.2630)	0.0188	9.79
		Ms	0.3098 (0.3117)	0.0196	
		NS	0.1009 (0.1028)	0.0140	
		Ns	0.3244 (0.3225)	0.0198	
Mulattoes	79	MS	0.2770 (0.2646)	0.0376	2.99
		Ms	0.2863 (0.2987)	0.0380	
		NS	0.0395 (0.0518)	0.0197	
		Ns	0.3972 (0.3849)	0.0408	
Blacks	12	MS	0.2757 (0.3333)	0.0966	—
		Ms	0.2243 (0.1667)	0.0909	
		NS	0.0576 (0.0000)	0.0572	
		Ns	0.4424 (0.5000)	0.1060	
Mongoloids	7	MS	0.0714 (0.0714)	0.0688	—
		Ms	0.3571 (0.3571)	0.1281	
		NS	0.0000 (0.0000)	0.0000	
		Ns	0.5714 (0.5714)	0.1323	
Without racial classification	114	MS	0.2134 (0.2139)	0.0304	4.05
		Ms	0.3085 (0.3081)	0.0352	
		NS	0.1287 (0.1283)	0.0261	
		Ns	0.3494 (0.3498)	0.0344	
Total sample	540	MS	0.2523 (0.2516)	0.0145	5.46
		Ms	0.3060 (0.3068)	0.0153	
		NS	0.0967 (0.0975)	0.0108	
		Ns	0.3449 (0.3441)	0.0157	

Table X - Estimates of Rh haplotype frequencies. n = number of sampled individuals; p_i = haplotype frequency estimates using the maximum likelihood method (figures in parentheses are the adjusted frequency estimates obtained with Mourant's counting method); s.e. (p_i) = standard error of estimates obtained from the variance-covariance matrix of the iterative numerical method (Newton-Raphson generalized procedure); χ^2_{df} = goodness of fit test (chi-squared value with df degrees of freedom).

Group	n	Haplotypes	p_i	s.e. (p_i)	χ^2_{df}	df
Whites	2223	CDE	0.0067 (0.0061)	0.0016	2.36	3
		CDe	0.4088 (0.4095)	0.0079		
		Cde	0.0175 (0.0173)	0.0034		
		cDE	0.1490 (0.1502)	0.0055		
		cDe	0.1107 (0.1105)	0.0075		
		cdE	0.0007 (0.0000)	0.0007		
		cde	0.3066 (0.3063)	0.0090		
Mulattoes	605	CDE	0.0000 (0.0000)	0.0000	4.83	2
		CDe	0.3231 (0.3232)	0.0145		
		Cde	0.0166 (0.0181)	0.0066		
		cDE	0.1433 (0.1455)	0.0101		
		cDe	0.2358 (0.2344)	0.0180		
		cdE	0.0000 (0.0000)	0.0000		
		cde	0.2812 (0.2788)	0.0185		
Blacks	134	CDE	0.0000 (0.0000)	0.0000	5.75	2
		CDe	0.1477 (0.0982)	0.0308		
		Cde	0.0583 (0.1107)	0.0262		
		cDE	0.0882 (0.0915)	0.0203		
		cDe	0.4413 (0.4438)	0.0434		
		cdE	0.0129 (0.0130)	0.0126		
		cde	0.2516 (0.2427)	0.0408		
Mongoloids	56	CDE	0.0199 (0.0000)	0.0140	-	
		CDe	0.5783 (0.5982)	0.0469		
		Cde	0.0000 (0.0000)	0.0000		
		cDE	0.3194 (0.3393)	0.0443		
		cDe	0.0824 (0.0625)	0.0264		
		cdE	0.0000 (0.0000)	0.0000		
		cde	0.0000 (0.0000)	0.0000		

Continued

Table X - Continued

Group	n	Haplotypes	p_i	s.e. (p_i)	χ^2_{df}	df
Without racial classification	1722	CDE	0.0097 (0.0109)	0.0023	1.92	4
		CDe	0.3773 (0.3775)	0.0089		
		Cde	0.0152 (0.0138)	0.0036		
		cDE	0.1356 (0.1356)	0.0062		
		cDe	0.1468 (0.1465)	0.0092		
		cdE	0.0051 (0.0039)	0.0021		
		cde	0.3104 (0.3118)	0.0106		
Total sample	4740	CDE	0.0081 (0.0070)	0.0012	3.26	4
		CDe	0.3797 (0.3815)	0.0054		
		Cde	0.0179 (0.0172)	0.0024		
		cDE	0.1427 (0.1448)	0.0038		
		cDe	0.1505 (0.1500)	0.0056		
		cdE	0.0027 (0.0017)	0.0009		
		cde	0.2984 (0.2979)	0.0063		

Combining the above estimates by the usual method of weighing them by the inverse of their variances, we obtain the final figure of 0.668 (percentage of white ancestry among the mulattoes of our sample). A chi-squared value of 0.045 (1 d.f.) indicates that the two estimates used for obtaining this figure are homogeneous and that the estimate is adequate.

We did not use the frequencies of any of the MNSs haplotypes for estimating the degree of black and white ancestry among mulattoes because the haplotype frequencies in one of the parental samples (blacks) were obtained from a very small sample, therefore with a large probability of sampling error that could explain the fact that the frequencies for Ms and MS obtained in the sample of blacks are outside the usual range reported for typical African populations (see, for example, Cavalli-Sforza and Bodmer, 1971).

The estimates shown above were obtained using as parental populations our own samples of blacks and whites, and it is well known that even these categories present a significant degree of interracial admixture in Brazil. Using for the parental black population the frequency of 0.091 for the CDe haplotype and for the parental white population the frequency of 0.413 (frequencies for Bantu negroes of S. Africa and for mediterranean Portuguese - Beiguelman, 1983) we obtain the following estimates of white ancestry for Brazilian negroes, mulattoes and whites:

Table XI - Estimates of Rh haplotype frequencies (part of the Rh total sample tested with anti-C^W); p_i = haplotype frequency estimates using the maximum likelihood method; s.e. (p_i) = standard error of estimates obtained from the variance-covariance matrix of the iterative numerical method (Newton-Raphson generalized procedure); χ^2_{df} = goodness of fit test (chi-squared value with df degrees of freedom).

Group	n	Haplotypes	p_i	s.e. (p_i)	χ^2_{df}	df
Whites	1715	CDE	0.0104	0.0022	1.91	4
		CDe	0.4135	0.0091		
		Cde	0.0169	0.0039		
		cDE	0.1526	0.0063		
		cDe	0.1088	0.0086		
		cdE	0.0009	0.0009		
		cde	0.2928	0.0103		
		C ^W De	0.0041	0.0011		
Mulattoes	429	CDE	0.0000	0.0000	5.69	2
		CDe	0.3177	0.0172		
		Cde	0.0180	0.0079		
		cDE	0.1256	0.0114		
		cDe	0.2396	0.0216		
		cdE	0.0000	0.0000		
		cde	0.2944	0.0222		
		C ^W De	0.0047	0.0023		
Blacks	98	CDE	0.0000	0.0000	3.42	1
		CDe	0.1121	0.0369		
		Cde	0.0776	0.0349		
		cDE	0.0874	0.0236		
		cDe	0.4494	0.0501		
		cdE	0.0151	0.0150		
		cde	0.2583	0.0471		
		C ^W De	0.0000	0.0000		
Mongoloids	42	CDE	0.0259	0.0181		
		CDe	0.5931	0.0539		
		Cde	0.0000	0.0000		

Continued

Table XI - Continued

Group	n	Haplotypes	p_i	s.e. (p_i)	χ^2_{df}	df
Mongoloids		cDE	0.3193	0.0512	—	
		cDe	0.0616	0.0267		
		cdE	0.0000	0.0000		
		cde	0.0000	0.0000		
		C ^W De	0.0000	0.0000		
Without racial classification	1160	CDE	0.0129	0.0031	1.57	2
		CDe	0.3869	0.0109		
		Cde	0.0157	0.0044		
		cDE	0.1269	0.0075		
		cDe	0.1452	0.0114		
		cdE	0.0051	0.0025		
		cde	0.3034	0.0130		
		C ^W De	0.0039	0.0013		
Total sample	3444	CDE	0.0124	0.0018	5.18	4
		CDe	0.3838	0.0064		
		Cde	0.0187	0.0029		
		cDE	0.1388	0.0044		
		cDe	0.1494	0.0067		
		cdE	0.0028	0.0012		
		cde	0.2901	0.0075		
		C ^W De	0.0039	0.0008		

$(0.1477-0.091)/(0.413-0.091) = 17.6\%$, $(0.3231-0.091)/(0.413-0.091) = 78.1\%$ and $(0.4088-0.091)/(0.413-0.091) = 98.7\%$, respectively.

DISCUSSION

The phenotype, haplotype and allele frequencies reported here are in the range of frequencies described for the same blood group systems in other samples of Brazilian white, black and mongoloid individuals.

Among Europeans the maximum value for the frequency of the cDe haplotype does not exceed 3%; the value of 11% found for our white sample is typical of a hybrid white population like the Brazilian. Even in our mongoloid (Japanese)

sample the frequency of this negroid marker exceeds the values usually reported for 'pure' mongoloid populations.

Using the data from Table I (individual loci), it is easy to verify that the phenotype frequencies shown in column 5 (sample without racial classification) are not different from the ones obtained by adding columns 1, 2, 3 and 4 (samples of white, mulatto, black and mongoloid individuals): the respective chi-squared values obtained from the analysis of 3 x 2 contingency tables for MN, Ss, Kk, Cc and Ee loci were 2.21, 1.73, 2.31, 0.28 and 3.59 (each test performed with 2 degrees of freedom); the chi-squared value (with 1 degree of freedom) resulting from the analysis of the 2 x 2 contingency table for the Dd locus was 1.03. These results indicate that the composition of the samples without racial classification was in all instances the one observed in the samples obtained by adding white, mulatto, black and mongoloid individuals; therefore, there is no detectable bias in the sample without racial classification, that is, the reason for not having the racial identification in some of our records is independent of the racial extraction of the sampled individuals.

In order to avoid the unnecessary complications in the comparison of haplotype frequencies among samples, we compared the allele frequencies obtained from the analysis of individual diallelic loci without dominance (Table I) using the usual gene-counting chi-square homogeneity test devised by Stevens (1938); there were no differences among whites, blacks and mongoloids in relation to MN, Ss and Kk loci: the respective chi-squared values (with 2 degrees of freedom) were 0.90, 5.23 and 2.29. In relation to Cc and Ee loci from Rh, however, the chi-squared values (also with 2 d.f.) were highly significant (66.11 and 33.58 respectively); the partition of the contingency tables gave statistically significant differences in all comparisons (whites x blacks, whites x mongoloids, blacks x mongoloids), for both loci.

Allele frequencies from the ABO system were compared using the method of probability equivalent circles (Stevens, 1950), that uses the A and B allele frequencies in a cartesian system with the square root transformation applied to coordinate values. The results are depicted in Figure 1, that shows the 95% confidence intervals for allele frequencies among whites (1), blacks (3) and mongoloids (4). Also shown are the 95% confidence intervals for the mulatto (2) and the total (unidentified) samples (5). At the 95% interval confidence level, the frequencies among whites are different from the ones found among mongoloids; at the 80% confidence level all three samples (whites, blacks and mongoloids) can be considered different as to their allele frequencies.

As for the allele and haplotype frequency estimates shown in Tables VIII to XI, the adherence tests gave fairly good results, even taking into account the usual degree of heterogeneity among subsamples within samples in American countries; the chi-squared values shown in these tables were all non-significant, thus indicating that deviations from panmictic proportions can be considered negligible.

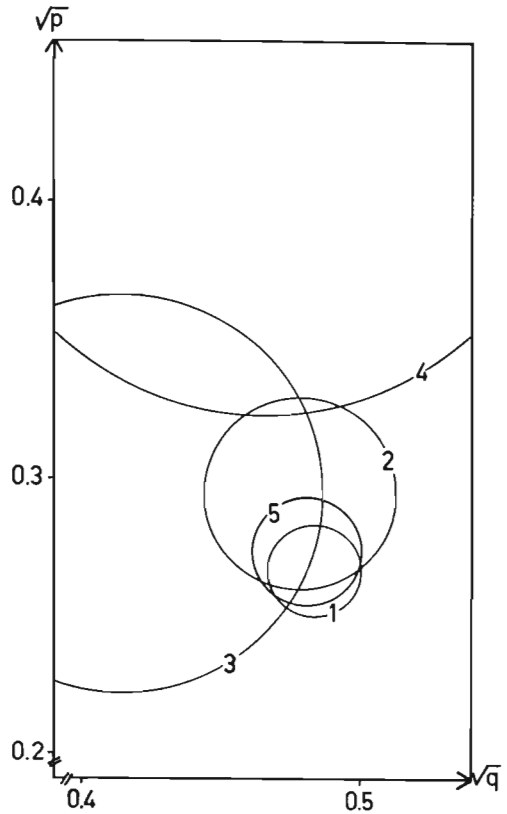


Figure 1 - 95% probability equivalent circles of the white (1), mulatto (2), black (3), mongoloid (4) and total (5) samples. p = frequency of A allele, q = frequency of B allele from ABO system (see text for details).

The adherence was tested also, in the case of haplotype frequencies, using the following method, devised by Fisher (1946a,b): if the haplotype frequency estimates are good, then the sum of the frequencies of the haplotypes involving a given allele, for instance the C allele from Rh, should agree with the proportion of this gene observed by direct gene counting. Using the data for the white sample of Table X, we obtain:

$$\begin{aligned}
 P(C) &= P(CDE) + P(CDe) + P(Cde) = \\
 &= 0.0067 + 0.4088 + 0.0175 = 0.4330,
 \end{aligned}$$

exactly the same value obtained by direct counting of C genes in the analysis of individual loci shown in Table VII.

Adding the elements of the corresponding 3×3 portion of the variance-covariance matrix used for the numerical evaluation of the haplotype frequencies, we obtain:

$$\begin{aligned}
 \text{Var}(C) &= \text{Var}(CDE) + 2\text{Cov}(CDE, CDe) + \text{Var}(CDe) + 2\text{Cov}(CDE, Cde) + \\
 &\quad + 2\text{Cov}(CDe, Cde) + \text{Var}(Cde) = \\
 &= 0.000002432 - 2 \times 0.000001519 + 0.000062857 - \\
 &\quad - 2 \times 0.000000064 - 2 \times 0.000009197 + 0.000011491 = \\
 &= 0.0000552;
 \end{aligned}$$

the standard error is the square root of this value: 0.007431. This is exactly the standard error obtained from the square root of the binomial variance of the allele C frequency estimate shown in Table VII: $\text{Var}(C) = 0.4330(1-0.4330)/4446$, $\text{s.e.}(C) = 0.007431$. This was performed in relation to each Rh and MNSs haplotype, and in all instances the values matched very well (results not shown).

ACKNOWLEDGMENTS

We thank all the members of the staff from the Laboratório de Investigação Médica (LIM-40) do Departamento de Medicina Legal da Faculdade de Medicina da Universidade de São Paulo for helping us to collect from the record books the data presented here and for arranging them conveniently for analysis. We thank also Dr. Sérgio R. Matioli (Departamento de Biologia, USP) for helping us in the preparation of the computer programs and Dr. Oswaldo Frota-Pessoa, from the same institution, for carefully reading some parts of the original manuscript. The careful revision and the useful criticism and suggestions of two anonymous referees from the Editorial Board of the Revista Brasileira de Genética are also deeply acknowledged. Many of these commentaries have been included in the final (revised) version of the manuscript. This work was partially supported by grants from CNPq/PIG, FINEP and FAPESP.

RESUMO

No presente trabalho apresentamos estimativas de frequências alélicas e haplotípicas dos seguintes sistemas de grupos sanguíneos, obtidas da análise de amostras de indivíduos não aparentados: MNSs, Kell-Cellano, Rh (5-6 soros) e ABO. Nossa amostra, estudada no Departamento de Medicina Legal da Faculdade de Medicina da USP, compreende indivíduos pertencentes aos grupos raciais caucasóide, negróide (negros e mulatos) e mongolóides (japoneses). Além de compararmos as frequências alélicas e fenotípicas entre as subamostras, calculamos também o grau de mistura racial na subamostra de indivíduos classificados como mulatos. Apresentamos também um método simplificado para se estimar as frequências dos alelos A_1 e A_2 numa amostra de indivíduos A e AB tipados com anti-soros A e B e lectina anti- A_1 .

REFERENCES

- Beiguelman, B. (1972). *Dinâmica dos genes nas populações e nas famílias*, Edart, São Paulo, 2nd Edition.
- Beiguelman, B. (1977). *Dinâmica dos genes nas famílias e nas populações*, Edart/EDUSP, São Paulo.
- Beiguelman, B. (1983). *Farmacogenética e sistemas sangüíneos eritrocitários*, Guanabara Koogan, Rio de Janeiro.
- Cavalli-Sforza, L.L. and Bodmer, W.F. (1971). *The genetics of human populations*, W.H. Freeman, San Francisco.
- Elandt-Johnson, R.C. (1971). *Probability models and statistical methods in genetics*, John Wiley & Sons, New York.
- Fisher, R.A. (1946a). The fitting of gene frequencies to data on Rhesus reactions. *Ann. Eug.* 13: 150-155.
- Fisher, R.A. (1946b). Note on the calculation of the frequencies of Rh allelomorphs. *Ann. Eug.* 13: 223-224.
- Mourant, A.E., Kopec, A.C. and Domaniewska-Sobczak, K. (1976). *The distribution of human blood groups and other polymorphisms*, Oxford University Press, London, 2nd Edition.
- Salzano, F.M. (1971). Genetic polymorphisms in Brazilian populations. In: *The ongoing evolution of Latin American populations* (Salzano, F.M. ed.), Charles C Thomas, Springfield.
- Stevens, W.L. (1938). Estimation of blood-groups gene frequencies. *Ann. Eug.* 8: 362-375.
- Stevens, W.L. (1950). Statistical analysis of the A-B-O blood groups. *Hum. Biol.* 22: 191-217.

(Received November 8, 1988)