B. W. Grunbaum, Ph.D., M.Crim.; Steve Selvin,[1] Ph.D.; B. A. Myhre,[2] M.D., Ph.D.; and Nello Pace,[1] Ph.D.

Distribution of Gene Frequencies and Discrimination Probabilities for 22 Human Blood Genetic Systems in Four Racial Groups

REFERENCE: Grunbaum, B. W., Selvin, S., Myhre, B. A., and Pace, N., **"Distribution of Gene Frequencies and Discrimination Probabilities for 22 Human Blood Genetic Systems in Four Racial Groups,"** *Journal of Forensic Sciences*, JFSCA, Vol. 25, No. 2, April 1980, pp. 428-444.

ABSTRACT: Gene frequencies were computed in four racial categories from 5956 blood donors from California, Hawaii, Mexico City, and Texas. Calculations were based on the phenotypic distribution of 22 blood genetic systems including 7 blood groups and 15 genetically controlled polymorphic proteins and enzymes. Matching probabilities for 20 systems were approximately 1 in 100 000 Asians, 1 in 200 000 blacks, 1 in 330 000 Mexicans, and 1 in 1 000 000 whites. The complementary discrimination probability, which measures the likelihood that two random individuals do not match, was, for practical purposes, unity. The combined new technology for blood grouping and electrophoresis using cellulose acetate membranes provides a powerful individualizing and discriminating tool for forensic science investigation.

KEY WORDS: pathology and biology, genetic typing, blood

Civil and criminal investigations often employ analysis of human blood group data as a tool for identification. When two blood samples compared for identification do not match, the innocence of an accused person is established. When two compared samples match, the accused person is not eliminated from suspicion on the basis of blood analysis. In this case, mathematical probabilities are the best method available to assess the likelihood associated with the failure of genetic evidence to give a conclusive answer. The calculation of these probabilities depends on accurate knowledge of the phenotypic or gene frequencies of the genetic variants being compared. The probability of matching [*1,2*], the probability of discrimination [*2,3*], and the probability of paternity [*2,4*] are all examples

The preparation of these materials was financially assisted through a federal grant from the Law Enforcement Assistance Administration (LEAA) and the California Office of Criminal Justice Planning (OCJP) and under Title I of the Crime Control Act of 1973. The opinions, findings, and conclusions in this publication are those of the authors and not necessarily those of OCJP or LEAA, and OCJP reserves a royalty-free, nonexclusive, and irrevocable license to reproduce, publish, and use these materials, and to authorize others to do so. Received for publication 3 Aug. 1979; revised manuscripts received 5 Oct. and 6 Dec. 1979; accepted for publication 7 Dec. 1979.

[1] Research biochemist, Environmental Physiology Laboratory; associate professor of biostatistics, School of Public Health; and professor of physiology, emeritus, respectively, University of California, Berkeley.

[2] Professor of pathology, University of California Los Angeles School of Medicine, Harbor UCLA Medical Center Campus, Harbor General Hospital, Torrance, Calif.

of mathematical probabilities computed from blood group data employed in medicolegal situations.

A prerequisite for useful probabilities is reliable estimates of the phenotypic and gene frequencies from specific populations. An earlier report from this laboratory [*1*] described the frequency distribution and discrimination probability of twelve genetically controlled systems in four human racial categories. The present study extends the previous work to the frequency distribution of phenotypes in 22 human blood genetic systems based on newly collected data. These data were used to estimate the gene frequencies associated with 20 human blood group systems for four racial categories. Also included is a test of the "goodness of fit" for each of these estimates along with the discriminant probabilities that measure the relative efficacy of each system for identification.

Method

As part of a project entitled "Admissibility of Technical Physical Evidence" funded by the California Office of Criminal Justice Planning, 5956 blood samples were analyzed to determine the phenotypes for up to 22 separate genetic systems for each individual sampled (Table 1). All determinations were made on fresh blood specimens from blood bank donors. Each blood sample received from a blood bank was contained in a segment of donor tubing. A section (two adjoining sealed segments) was snipped off at the blood bank and then shipped with others in a refrigerated container to the University of California laboratory at Berkeley. The blood bank identified each sample by code number, to preserve the anonymity of the donor, and provided the age, sex, and ethnic group of the donor.

At the Berkeley laboratory, the sample segments were separated so that a coded half of each sample could be shipped immediately in a refrigerated container to the Harbor General Hospital laboratory at Torrance, Calif., for blood typing for ABO, Rh, MNSs, Kell, Duffy, Kidd, and P systems.

The blood grouping was done with immunological techniques. Red cell groupings were performed with Microtiter® plates using slight modifications of the technic of Crawford et al [*5*]. The red cell samples and appropriate controls were placed vertically. The resultant mixtures were agitated. One plate was used for room temperature reactions, another plate for 37°C reactions, and a third plate for anti-globulin testing. All reactions took place under identical conditions and were read at the same time. Immediate comparisons between reactions were made. When a permanent record was needed, the plates were photographed. More detailed methods will be described elsewhere.[3]

At the Berkeley laboratory each blood sample was analyzed electrophoretically for up to 15 variants. Haptoglobin typing was done by electrophoresis in a horizontal 1-mm-thick acrylamide gel for about 60 min, using a newly designed apparatus [*6*]. The electrophoresis of the other 14 protein and enzyme systems was done on Sartorius cellulose acetate membranes.

The data were collected so that adequate numbers of individuals were present for four ethnic groups (white, black, Mexican, and Asian). Unselected samples from California blood banks contain very few specimens from minority ethnic groups. To correct this inequality, special effort was made to collect samples from blood banks in areas where there are concentrations of particular ethnic groups. Participating blood banks were requested first to give all samples collected from specific ethnic minorities, then to supplement their quota for a given period with blood from white donors. When it became apparent that an adequate statistical balance could not be secured from California blood

[3]B. W. Grunbaum, *A Manual for the Forensic Analyst for Phenotyping of Human Fresh Blood and Bloodstains,* in preparation.

TABLE 1—*Genetic systems phenotyped.*

Genetic System	Abbreviation	White	Black	Mexican	Asian	Other	Total	Percentage of Total	Method Reference
		Available Specimens							
		1121	924	2047	1605	259	5956	100	
ABO	...	914	713	1212	1342	38	4219	71	*5*, Footnote 3
Rh	...	914	713	1212	1342	38	4219	71	*5*, Footnote 3
MNSs	...	751	710	1189	1342	38	4030	68	*5*, Footnote 3
Kell	...	911	713	1212	1342	38	4216	71	*5*, Footnote 3
Duffy	...	913	713	1212	1342	38	4218	71	*5*, Footnote 3
Kidd	...	514	697	1143	1340	37	3731	63	*5*, Footnote 3
P	...	914	713	1212	1342	38	4219	71	*5*, Footnote 3
Haptoglobin	Hp	860	463	775	1105	37	3240	54	*6*, Footnote 3
Group specific component	Gc	1050	867	1908	1566	249	5640	95	*7*
Transferrin	Tf	801	502	765	1295	32	3395	57	*8*, Footnote 3
Hemoglobin	Hb	1040	792	1569	1451	144	4996	84	*9*
Adenylate kinase	AK	1021	736	1380	1410	83	4630	78	*10*, Footnote 3
Adenosine deaminase	ADA	1005	726	1329	1391	70	4521	76	*10*, Footnote 3
Erythrocyte acid phosphatase	EAP	1044	845	1797	1542	179	5407	91	*11*
Esterase D	EsD	1025	770	1478	1428	109	4810	81	*12*
Glucose-6-phosphate dehydrogenase	G-6-PD	1054	834	1778	1527	218	5411	91	*13*
Phosphoglucomutase	PGM	1069	802	1621	1500	140	5132	86	*14*
6-Phosphogluconate dehydrogenase	6-PGD	951	828	1806	1541	215	5341	90	*10*, Footnote 3
Phosphoglucose isomerase	PGI	769	547	785	1260	33	3394	57	Footnote 3
Glutamic pyruvate transaminase	GPT	517	558	1021	1076	25	3197	54	*15*, Footnote 3
Glyoxalase I	GLO-I	313	308	1080	884	11	2596	44	*15*, Footnote 3
Peptidase A	PEP-A	301	492	766	108	17	1684	28	*16*, Footnote 3

banks alone, we asked the Blood Bank of Hawaii for Asian samples and the Banco Central de Sangre of Mexico City for Mexican samples. The Hawaiian blood bank has kindly given us a more complete breakdown of the Asian donors, marking them as Japanese, Chinese, Korean, Filipino, and so forth. However, only Japanese and Chinese were considered as Asians for this report. This more precise racial determination has enabled us to make some interesting statistical observations that will be presented in a separate report.

The distribution of the samples by ethnic groups is given in Table 2. The white sample is mostly from southern California (63%), the black sample predominantly from northern California (92%), the Mexican group primarily from Mexico City (73%), and the Asian sample is almost entirely from Hawaii (86%).

The data were not classified by variables other than race for statistical analysis. Categories such as age, residence within California, and sex (except for the glucose-6-phosphate dehydrogenase system) were analyzed and found to have no measurable influence on the distribution of phenotypic and, therefore, genotypic frequencies. The assumption that these data were homogeneous for demographic characteristics other than race is consistent with genetic theory and empiric data [*1*].

Results

In Table 3 the data obtained from the laboratory phenotyping is presented by frequency for each phenotype for the 22 genetic systems in four racial groups. However, the basic measure of the distribution of a genetic variant in a population is the gene frequency. The estimated gene frequencies for 20 of the 22 analyzed systems[4] are given in Table 4 along with their standard errors. These gene frequencies were estimated by the widely used method of maximum likelihood [*17*], which has several optimum statistical properties. For codominant genetic systems with two alleles, the maximum likelihood estimates of the two gene frequencies p_1 and p_2 are

$$\hat{p}_1 = (2x + y)/2N$$

and

$$\hat{p}_2 = 1 - \hat{p}_1$$

where x symbolizes the number of one homozygotic type, y symbolizes the number of the heterozygotic type, and N is the total number of individuals analyzed. The standard error of the estimates $\hat{p}_1$ and $\hat{p}_2$ is $(p_1p_2/2N)^{1/2}$. Estimates and standard errors for more complicated systems were calculated by a computer algorithm. Rare phenotypes (Table 3) (occurrence of less than 1% in the general population) were excluded from the estimation procedure since inclusion complicated the estimation and does not appreciably improve the gene frequency estimates.

TABLE 2—*The distribution of ethnic groups in the data by areas sampled (percentages).*

Area	White	Black	Mexican	Asian
Northern California	12.0	92.4	13.6	13.8
Southern California	63.2	5.6	7.9	...
Dallas, Texas	24.6	1.9	5.4	...
Hawaii	0.2	0.1	0.6	86.2
Mexico City	...	...	72.5	...

[4] Phosphoglucose isomerase and transferrin had only rare phenotypes, and thus the gene frequency estimates are 1.0.

TABLE 3—*Frequency distribution of 22 human blood genetic systems in four racial categories.*

Phenotype	White, $N = 1121$	Black, $N = 924$	Mexican, $N = 2047$	Asian, $N = 1605$	Other,[a] $N = 259$	Total, $N = 5956$
ABO						
A_1	288	147	239	442	16	1132
A_2	83	37	39	4	0	163
B	85	137	113	287	5	627
OO	412	359	801	475	13	2060
A_1B	36	20	15	125	4	200
A_2B	10	13	5	9	0	37
Total	914	713	1212	1342	38	4219
Rh[b]						
DCCee	150	32	324	560	13	1079
dccee	165	61	32	2	0	260
DCcee	284	154	215	112	6	771
DccEe	113	94	91	46	2	346
DCcEe	122	38	316	466	15	957
Dccee	37	283	19	23	0	362
DccEE	13	14	84	95	1	207
dccEE	2	0	2	0	0	4
DCCEe	4	2	46	14	0	66
dCcee	7	12	2	3	0	24
dccEe	5	0	2	3	0	10
DCcEE	3	0	47	6	0	56
dCCEe	0	0	0	1	0	1
DCCEE	0	0	28	0	0	28
D^uCCee	1	0	0	4	0	5
D^uccee	1	12	0	1	1	15
D^uCcee	6	7	0	2	0	15
D^uccEe	1	4	0	1	0	6
D^uCcEe	0	0	0	3	0	3
dCCEE	0	0	4	0	0	4
Total	914	713	1212	1342	38	4219
MNSs						
MNss	177	198	215	553	18	1161
MNSs	158	75	205	80	4	522
MMSs	111	68	277	79	2	537
NNss	119	145	85	234	6	589
MMss	81	128	183	347	7	746
MMSS	49	16	133	8	0	206
NNSs	22	48	31	30	1	132
MNSS	28	21	52	8	0	109
NNSS	6	11	8	3	0	28
Total	751	710	1189	1342	38	4030
Kell						
kk	824	694	1189	1339	38	4084
Kk	84	19	23	3	0	129
KK	3	0	0	0	0	3
Total	911	713	1212	1342	38	4216
Duffy						
Fy(a+b−)	201	117	493	1093	29	1933
Fy(a−b+)	291	149	182	15	3	640
Fy(a+b+)	408	33	517	232	6	1196
Fy(a−b−)	13	414	20	2	0	449
Total	913	713	1212	1342	38	4218
Kidd						
Jk(a+b−)	144	303	288	262	11	1008
Jk(a−b+)	126	41	206	311	7	691
Jk(a+b+)	244	353	649	767	19	2032
Total	514	697	1143	1340	37	3731
P						
P_1^+	648	622	835	454	16	2575
P_1^-	266	91	377	888	22	1644
Total	914	713	1212	1342	38	4219

TABLE 3—*Continued.*

Phenotype	White, $N = 1121$	Black, $N = 924$	Mexican, $N = 2047$	Asian, $N = 1605$	Other,[a] $N = 259$	Total, $N = 5956$
Hp						
1-1	154	135	218	148	11	666
1-2	418	236	393	444	14	1505
2-2	288	89	163	512	12	1064
Rare[c]	0	3	1	1	0	5
Total	860	463	775	1105	37	3240
Gc						
1-1	537	638	1160	780	148	3263
1-2	429	183	655	590	82	1939
2-2	74	18	80	118	13	303
Rare[c]	10	28	13	78	6	135
Total	1050	867	1908	1566	249	5640
Tf						
CC	786	467	742	1259	31	3285
Rare[c]	15	35	23	36	1	110
Total	801	502	765	1295	32	3395
Hb						
AA	1040	716	1561	1448	144	4909
AS	0	54	3	1	0	58
AC	0	18	3	0	0	21
Rare[c]	0	4	2	2	0	8
Total	1040	792	1569	1451	144	4996
AK						
1-1	944	722	1347	1406	81	4500
1-2	77	13	31	4	2	127
2-2	0	0	2	0	0	2
Rare[c]	0	1	0	0	0	1
Total	1021	736	1380	1410	83	4630
ADA						
1-1	910	701	1284	1317	63	4275
1-2	91	24	44	73	7	239
2-2	3	1	1	1	0	6
Rare[c]	1	0	0	0	0	1
Total	1005	726	1329	1391	70	4521
EAP						
AA	106	31	109	67	9	322
BA	426	305	640	502	64	1937
BB	391	481	986	972	99	2929
CA	45	2	19	0	2	68
CB	69	8	40	1	5	123
Rare[c]	7	18	3	0	0	28
Total	1044	845	1797	1542	179	5407
EsD						
1-1	794	626	1077	561	48	3106
1-2	225	140	370	674	51	1460
2-2	6	4	31	193	10	244
Total	1025	770	1478	1428	109	4810
G-6-PD (males)						
B	699	389	1433	1236	161	3918
A	1	98	13	...	...	112
BA	...	...	...	...	...	...
A−	...	45	7	1	1	54
Rare[c]	3	12	6	13	6	40
Total	703	544	1459	1250	168	4124
G-6-PD (females)						
B	347	182	311	275	49	1164
A	1	28	1	...	...	30
BA	1	67	3	...	1	72
A−	0	2	0	0	0	2
Rare[c]	2	11	4	2	0	19
Total	351	290	319	277	50	1287

TABLE 3—*Continued.*

Phenotype	White, $N = 1121$	Black, $N = 924$	Mexican, $N = 2047$	Asian, $N = 1605$	Other,[a] $N = 259$	Total, $N = 5956$
PGM						
1-1	626	530	961	851	78	3046
1-2	393	238	563	557	56	1807
2-2	48	27	87	77	5	244
Rare[c]	2	7	10	15	1	35
Total	1069	802	1621	1500	140	5132
6-PGD						
AA	918	752	1719	1257	188	4834
AC	31	73	85	276	25	490
Rare[c]	2	3	2	8	2	17
Total	951	828	1806	1541	215	5341
PGI						
1-1	768	546	785	1252	33	3384
Rare[c]	1	1	0	8	0	10
Total	769	547	785	1260	33	3394
GPT						
1-1	198	359	180	400	10	1147
2-2	127	53	401	220	7	808
1-2	189	148	438	454	8	1237
Rare[c]	3	0	2	2	0	7
Total	517	558	1021	1076	25	3197
GLO-I						
1-1	54	39	111	4	0	208
2-2	94	146	525	755	9	1529
1-2	165	125	444	125	2	861
Total	313	308	1080	884	11	2596
PEP-A						
1-1	300	442	761	108	17	1628
2-2	0	5	0	0	0	5
1-2	0	41	3	0	0	44
Rare[c]	1	4	2	0	0	7
Total	301	492	766	108	17	1684

[a] Other = race not available, or American Indian, Filipino, Hawaiian, and other South Sea Islanders.

[b] D = DD or Dd phenotypes.

[c] Rare = phenotypes occurring in less than 1% in the general population.

Standard errors are an assessment of the magnitude of the influence of sampling variation on the estimated gene frequencies. In general, the standard error associated with each gene frequency estimate is less than 0.015, showing that the number of blood samples analyzed was large enough to produce rather stable estimates (that is, fluctuations of ± 3 [0.015] = ± 0.045 are extremely unlikely to be a result of strictly sampling variation). For example, differences within genetic systems among the four racial categories are consistently larger than would be expected from the effects of only sampling variation. The difference between gene frequencies almost always exceeds two or three times the standard error for the 20 systems analyzed (Table 4). This observation reaffirms that biological variation among racial groups is very much an issue when blood group frequencies are applied for purposes of identification.

Goodness of Fit

The estimated gene frequencies were tested against the observed data to measure their goodness of fit. Each genetic system was investigated assuming simple or multi-allelic inheritance with gene frequencies at Hardy-Weinberg equilibrium, and the lack of fit was expressed in terms of a χ^2 statistic. This assumption is typically made when one is dealing with human blood group genetics. A blood group system was considered worthy of special note when the probability of observing the specific χ^2 value by chance fluctuation alone was less than 0.05. For example, consider the phosphoglucomutase system for whites as a typical case. The data and the χ^2 calculation are given in Table 5. The lack of fit of the gene frequency estimates to the observed data was likely a result of random variation. The probability of observing the lack of fit that produced the χ^2 value of 1.935 by change variation is 0.164 (p value). The other systems were similarly analyzed and a summary of the results is given in Table 6.

The race-specific gene frequencies given in Table 4 were used to calculate expected phenotypic frequencies. This computation is the basis for the χ^2 test of goodness of fit used to produce Table 6. For example, the Kell system phenotype Kk has an estimated frequency of $2(0.049)(0.951) = 0.093$, where 0.049 is the frequency of Gene K and 0.951 is the frequency of Gene k among whites, as taken from Table 4. For some systems, the calculation is slightly more complicated but follows the same principle. The A_1 phenotype of the ABO system, for example, is made up of two genotypes A_1A_1 and A_1O. Therefore, the total frequency of A_1 types is the sum of the frequencies of the two relevant genotypes or $[(0.196)^2 + 2(0.196)(0.665)] = 0.299$, using the gene frequencies A_1 and O from the white sample (Table 4).

Fifteen blood group systems showed good to extremely good fit to the theoretical frequencies (calculated under the assumption that the system is at Hardy-Weinberg equilibrium) for each of the four racial categories. Four blood group systems, MNSs, Duffy, Kidd, and glutamic pyruvic transaminase, consistently failed to fit the theoretical expectations. The MNSs system had a significant excess of NNSS genotypes for all races and lacked MNSs genotypes for all races except white. The Duffy system had excess "silent" Fy types among whites and an excess of Fy(a+b+) among blacks, Mexicans, and Asians. The Kidd system also significantly did not fit the hypothesis of Hardy-Weinberg equilibrium of gene frequencies for blacks, Mexicans, and Asians [lack of Jk(a−b+) and excess of Jk(a+b−) and of Jk(a+b+)]. The twentieth system, the P system, could not be tested for goodness of fit because only two phenotypes were identified.

The only enzyme system that did not fit the Hardy-Weinberg model was glutamic pyruvic transaminase; all races lacked 1-2 types and had an excess of 1-1 and 2-2 types. The glutamic pyruvic transaminase system is subject to a classification bias. The 1-1 and 2-2 types can be accurately identified but the 1-2 types may be ambiguous and, consequently, not accurately classified. An estimate of gene frequencies can be made by compensating for the misclassification of heterozygous types [*18*]. Under the assumption that the misclassified 1-2 type is equally likely to be classified as a 1-1 or 2-2 type, the estimated gene frequencies given in Table 4 are unaffected (that is, unbiased) but the standard errors are increased. However, if the ambiguous phenotypes are not misclassified at random, the gene frequencies can be highly influenced. A few other systems failed to fit the theoretical expectations for isolated phenotypes but not in a consistent pattern and are noted in Table 5 (for example, the Rh system).

Computations Using Gene Frequencies

Phenotypic frequencies, either estimated from gene frequencies or derived directly from the data, are necessary for the calculation of the probability that two randomly chosen

TABLE 4—*Gene frequencies and standard errors for 20 genetic systems analyzed for four racial categories.*

System	Gene Frequencies and Standard Errors				Number Analyzed
ABO	A_1	A_2	B	O	
White	0.196(0.010)	0.066(0.077)	0.074(0.006)	0.664(0.012)	914
Black	0.125(0.009)	0.041(0.006)	0.127(0.009)	0.707(0.013)	713
Mexican	0.111(0.007)	0.021(0.003)	0.056(0.005)	0.812(0.008)	1212
Asian	0.239(0.008)	0.006(0.002)	0.171(0.007)	0.584(0.010)	1342
Rh	CDE	CDe	CdE	cDE	
White	0.008(0.003)	0.382(0.012)	0.000(0.001)	0.136(0.009)	905
Black	0.005(0.003)	0.163(0.012)	0.000(0.000)	0.114(0.009)	690
Mexican	0.061(0.008)	0.490(0.001)	0.015(0.006)	0.235(0.010)	1212
Asian	0.000(0.000)	0.631(0.025)	0.009(0.002)	0.262(0.013)	1342
Rh (continued)	cdE	cDe	Cde	cde	
White	0.012(0.004)	0.044(0.007)	0.010(0.003)	0.408(0.013)	905
Black	0.000(0.000)	0.397(0.016)	0.029(0.008)	0.292(0.018)	790
Mexican	0.016(0.006)	0.040(0.009)	0.005(0.004)	0.138(0.011)	1212
Asian	0.003(0.004)	0.020(0.009)	0.013(0.007)	0.062(0.010)	1342
MNSs	MS	Ms	NS	Ns	
White	0.245(0.012)	0.317(0.013)	0.059(0.007)	0.379(0.013)	751
Black	0.116(0.010)	0.389(0.014)	0.086(0.091)	0.409(0.014)	710
Mexican	0.316(0.010)	0.381(0.010)	0.062(0.006)	0.241(0.010)	1189
Asian	0.057(0.005)	0.505(0.010)	0.028(0.004)	0.410(0.098)	1342
Kell	K	k			
White	0.041(0.005)	0.951(0.005)			911
Black	0.013(0.003)	0.987(0.003)			713
Mexican	0.001(0.002)	0.999(0.002)			1212
Asian	0.001(0.006)	0.999(0.006)			1342
Duffy	$Fy^{(a)}$	$Fy^{(b)}$	Fy		
White	0.410(0.013)	0.500(0.013)	0.090(0.013)		913
Black	0.111(0.009)	0.136(0.009)	0.753(0.012)		713
Mexican	0.570(0.012)	0.337(0.010)	0.093(0.011)		1212
Asian	0.876(0.011)	0.096(0.006)	0.028(0.011)		1342
Kidd	$Jk^{(a)}$	$Jk^{(b)}$			
White	0.518(0.016)	0.482(0.016)			514
Black	0.688(0.012)	0.312(0.012)			697
Mexican	0.536(0.010)	0.464(0.010)			1143
Asian	0.482(0.010)	0.518(0.010)			1340

P	P_1	P_2			
White	0.461(0.014)	0.539(0.014)			914
Black	0.643(0.017)	0.357(0.017)			713
Mexican	0.442(0.012)	0.558(0.012)			1212
Asian	0.187(0.008)	0.813(0.088)			1342
Hp	$Hp^{(1)}$	$Hp^{(2)}$			
White	0.422(0.012)	0.578(0.012)			860
Black	0.550(0.016)	0.450(0.016)			460
Mexican	0.536(0.013)	0.464(0.013)			774
Asian	0.335(0.010)	0.665(0.010)			1104
Gc	$Gc^{(1)}$	$Gc^{(2)}$	$Gc^{(ab)}$	Rare	
White	0.721(0.010)	0.277(0.010)	0.002(0.001)	...	1044
Black	0.855(0.008)	0.128(0.008)	0.017(0.003)	...	867
Mexican	0.784(0.007)	0.215(0.007)	0.001(0.001)	...	1902
Asian	0.706(0.008)	0.272(0.008)	...	0.022(0.003)	1558
Hb	A	S	C		
White	1.000(0.000)	0.000(0.000)	0.000(0.000)		1040
Black	0.953(0.005)	0.035(0.005)	0.012(0.003)		789
Mexican	0.998(0.008)	0.001(0.005)	0.001(0.006)		1567
Asian	1.000(0.003)	0.000(0.003)	0.000(0.000)		1449
AK	$AK^{(1)}$	$AK^{(2)}$			
White	0.962(0.004)	0.038(0.004)			1021
Black	0.991(0.002)	0.009(0.002)			735
Mexican	0.987(0.002)	0.013(0.002)			1380
Asian	0.999(0.007)	0.001(0.007)			1410
ADA	$ADA^{(1)}$	$ADA^{(2)}$			
White	0.952(0.005)	0.048(0.005)			1004
Black	0.982(0.003)	0.018(0.003)			726
Mexican	0.983(0.003)	0.017(0.003)			1329
Asian	0.973(0.003)	0.027(0.003)			1391
EAP	A	B	C		
White	0.327(0.001)	0.612(0.001)	0.061(0.005)		1044
Black	0.223(0.010)	0.771(0.010)	0.006(0.002)		827
Mexican	0.244(0.007)	0.739(0.007)	0.017(0.002)		1795
Asian	0.206(0.007)	0.794(0.007)	0.000(0.000)		1542
EsD	$EsD^{(1)}$	$EsD^{(2)}$			
White	0.884(0.007)	0.116(0.007)			1025
Black	0.904(0.008)	0.096(0.008)			770
Mexican	0.854(0.006)	0.146(0.006)			1478
Asian	0.629(0.009)	0.371(0.009)			1428

TABLE 4—*Continued.*

System	Gene Frequencies and Standard Errors				Number Analyzed
G-6-PD	B	A	A^-	B^-	
White	0.992(0.003)	0.003(0.001)	0.000(0.000)	0.005(0.003)	1052
Black	0.743(0.013)	0.193(0.012)	0.058(0.008)	0.006(0.003)	820
Mexican	0.985(0.003)	0.009(0.002)	0.004(0.001)	0.002(0.002)	1771
Asian	0.995(0.002)	0.001(0.001)	0.001(0.001)	0.003(0.002)	1517
PGM	$PGM^{(1)}$	$PGM^{(2)}$			
White	0.771(0.009)	0.229(0.009)			1067
Black	0.816(0.010)	0.184(0.010)			795
Mexican	0.771(0.008)	0.229(0.007)			1611
Asian	0.761(0.008)	0.239(0.008)			1485
6-PGD	$PGD^{(A)}$	$PGD^{(C)}$			
White	0.983(0.003)	0.017(0.003)			950
Black	0.955(0.005)	0.045(0.005)			826
Mexican	0.976(0.002)	0.024(0.002)			1804
Asian	0.906(0.005)	0.094(0.005)			1539
GLO-I	$GLO^{(1)}$	$GLO^{(2)}$			
White	0.436(0.020)	0.564(0.020)			313
Black	0.327(0.019)	0.673(0.019)			310
Mexican	0.308(0.010)	0.692(0.010)			1080
Asian	0.075(0.060)	0.925(0.060)			884
PEP-A	$PEP\text{-}A^{(1)}$	$PEP\text{-}A^{(2)}$			
White	1.000(0.001)	0.000(0.001)			300
Black	0.948(0.007)	0.052(0.007)			488
Mexican	0.998(0.001)	0.002(0.001)			764
Asian	1.000(0.002)	0.000(0.002)			108
GPT	$GPT^{(1)}$	$GPT^{(2)}$			
White	0.569(0.015)	0.431(0.015)			514
Black	0.773(0.013)	0.227(0.013)			560
Mexican	0.391(0.011)	0.069(0.011)			1019
Asian	0.584(0.011)	0.416(0.011)			1074

TABLE 5—*A typical χ^2 calculation (goodness-of-fit) using the phosphoglucomutase system for whites.*

Phenotype	Observed Number	Expected Frequency[a]	Expected Number	Contribution to χ^2 Statistic
1-1	626	0.594	634.026	0.101
1-2	393	0.353	376.948	0.683
2-2	48	0.052	56.026	1.151
Total	1067	1.000	1067.000	1.935

[a]Based on gene frequencies $PGM^{(1)} = 0.771$ and $PGM^{(2)} = 0.229$.

individuals match by chance for all compared genetic systems. Employing the gene frequencies to estimate the phenotypic frequencies is preferable from the point of view of statistical precision. Taking the phenotypic frequencies directly from the sample data incurs a small loss of statistical efficiency (that is, larger standard errors). The probability of matching is precisely defined by Fisher [*2*] and Jones [*3*]. Table 7 gives the probability of matching for each system within each of the four racial categories. The probability of matching indicates the chance a system will produce two matching phenotypes when two randomly chosen individuals are compared.

Table 7 shows the well-known fact that the discrimination probability increases for each additional system used for identification. More importantly, Table 7 shows the exact amount of gain in discrimination efficacy expected from each system. For example, the probability that two randomly chosen white individuals will match with respect to any phenotype within the Kell system is 0.825. The haptoglobin system has a matching probability of 0.381, indicating that this system is more than twice as effective as the Kell system for discrimination among white individuals. The probability of matching for several systems is the product of the specific matching probabilities for the systems being compared. The discrimination probability is the complementary probability [P(discrimination) $= 1.0 - P$(matching)] and measures the likelihood that two randomly chosen individuals do not match for at least one of the systems being compared. The matching probability computed for all 20 systems among males, for example, is extremely small for each racial category. For whites the matching probability is 9.65×10^{-7} (approximately 1 in 1 000 000); for blacks it is 5.20×10^{-6} (about 1 in 200 000); for Mexicans it is 3.23×10^{-6} (about 1 in 330 000); and for Asians it is 9.61×10^{-6} (about 1 in 100 000). Therefore, the discrimination probability is correspondingly close to 1.0. The females have a similar matching probability for all 20 systems since they differ only for the G-6-PD system.

Discussion

The lack of fit of the observed frequencies to the theoretical phenotypic frequencies (estimated under the hypothesis that the genetic system is a Hardy-Weinberg equilibrium) results, at least in part, from one or more known sources of bias. The data used to estimate the gene frequencies are the result of a carefully conducted large-scale screening project. Nevertheless, it is expected that errors will occur when close to 6000 individual blood samples are analyzed for a large number of genetic systems. In particular, blood group systems like the Duffy and Kidd, where the agglutination tests involve a degree of subjective judgment, are prone to laboratory misclassification. Also, the quality of the antisera can be a source of error with several of the agglutination tests. The GPT enzyme system, found in an extremely low concentration in human blood, is subject to ambiguity of heterozygous phenotypes. The lack of fit observed for this system came about because

TABLE 6—*Results of the goodness of fit tests in terms of χ^2 values and significance probabilities (p value for four racial categories).*[a]

System	χ^2	*p* Value	Comments
ABO			
White	5.00	0.082	...
Black	5.31	0.070	...
Mexican	1.83	0.400	...
Asian	21.00	0.000	excess A_2B, lack A_2
Rh			
White	11.363	0.327	...
Black	12.636	0.245	...
Mexican	102.057	0.000	lack DCCEe, excess DCCEE
Asian	12.861	0.169	...
MNSs			
White	13.014	0.023	excess NNSs, excess NNSS
Black	41.278	0.000	lack MNSs, excess NNss, MNss, NNSS
Mexican	14.544	0.013	lack MNSs, excess NNss, NNSS
Asian	14.760	0.011	lack MNSs, excess NNss
Kell			
White	0.301	0.584	...
Black	0.123	0.719	...
Mexican	0.111	0.739	...
Asian	0.001	0.966	...
Duffy			
White	10.263	0.001	excess Fy(a−b−)
Black	7.942	0.005	excess Fy(a+b+)
Mexican	20.529	0.000	excess Fy(a+b+), Fy(a−b−)
Asian	2.103	0.147	...
Kidd			
White	1.255	0.263	...
Black	22.478	0.000	lack Jk(a−b+), excess Jk(a+b+)
Mexican	22.880	0.000	lack Jk(a−b+), excess Jk(a+b+), Jk(a+b−)
Asian	28.684	0.000	lack Jk(a−b+), excess Jk(a+b+), Jk(a+b−)
Hp			
White	0.012	0.913	...
Black	0.611	0.434	...
Mexican	0.330	0.565	...
Asian	10.504	0.001	lack 1-2, excess 1-1
Gc			
White	1.852	0.604	...
Black	1.648	0.649	...
Mexican	1.300	0.729	...
Asian	0.288	0.962	...
Hb			
White	...	...	no polymorphism
Black	1.29	0.731	...
Mexican	...	...	six variants other than AA
Asian	...	...	no polymorphism
AK			
White	1.567	0.211	...
Black	0.058	0.809	...
Mexican	0.096	0.757	...
Asian	...	...	four variants other than 1-1
ADA			
White	0.203	0.652	...
Black	0.025	0.876	...
Mexican	0.942	0.332	...
Asian	0.000	0.991	...
EAP			
White	4.202	0.240	...
Black	4.280	0.233	...
Mexican	2.312	0.510	...
Asian	0.296	0.961	...

TABLE 6—*Continued.*

System		*p* Value	Comments
EsD			
White	0.338	0.561	...
Black	1.666	0.197	...
Mexican	0.014	0.906	...
Asian	0.177	0.674	...
G-6-PD			
White	1.500	0.682	...
Black	45.273	0.000	lack A, AB , excess AA
Mexican	6.991	0.072	...
Asian	6.740	0.081	...
PGM			
White	1.935	0.164	...
Black	0.002	0.965	...
Mexican	0.146	0.702	...
Asian	1.334	0.248	...
6-PGD			
White	1.838	0.175	...
Black	0.318	0.573	...
Mexican	1.049	0.306	...
Asian	5.045	0.025	lack CC
GLO-I			
White	1.614	0.204	...
Black	2.212	0.137	...
Mexican	1.411	0.235	...
Asian	0.235	0.628	...
PEP-A	...	...	lack of common polymorphism
GPT			
White	32.200	0.000	lack 1-2, excess 1-1, 2-2
Black	34.005	0.000	lack 1-2, excess 1-1, 2-2
Mexican	9.767	0.002	lack 1-2, excess 1-1, 2-2
Asian	18.187	0.000	lack 1-2, excess 1-1, 2-2

[a] No test was possible for the P system because only two phenotypes were identified.

when the 1-2 GPT phenotypes were poorly resolved, they were not classified as heterozygotes.

Racial definition is another source of bias that will cause the expected phenotypic frequencies to deviate from the observed. The racial category for each sample was determined at the blood bank and certainly involves the possibility of error. No account was taken, for example, of individuals with racially mixed parents. Racial misclassification will have the most influences on the goodness of fit when the race-specific gene frequencies widely differ within a specific system (for example, the Rh and Duffy systems). A small number of racially misclassified individuals can be the primary source of lack of fit for the Duffy and Kidd systems. For example, the Duffy sample showed 20 Mexican individuals with phenotype Fy(a−b−) where only 10.6 were expected. Deviations of this sort are likely caused by racial admixture (for example, the A_2 frequency in the ABO system among Asians).

Another source of lack of fit is the heterogeneity in gene frequencies observed among the different populations included in this study. For example, the use of Mexican specimens from both California and Mexico adds a small contribution to the χ^2 statistic because of the differences in gene frequencies between the two areas. This source of bias was observed to be small for the groups included in the data in Table 1.

The gene frequencies given in Table 4 were compared to gene frequencies from a variety of other sources. As expected, there is close agreement in some cases and lack of agreement in others. These similarities and differences are undoubtedly a function of genetic

TABLE 7—*Matching probabilities for 20 genetic systems analyzed for four racial categories.*

System	White	Black	Mexican	Asian
1. ABO	0.320	0.334	0.484	0.290
2. Rh	0.205	0.250	0.182	0.308
3. MNSs	0.152	0.191	0.168	0.275
4. Kell	0.825	0.948	0.963	0.995
5. Duffy	0.342	0.404	0.364	0.692
6. Kidd	0.375	0.417	0.376	0.375
7. P	0.587	0.777	0.571	0.552
8. Haptoglobin	0.381	0.378	0.376	0.406
9. Group specific component	0.435	0.585	0.492	0.401
10. Hemoglobin	1.000	0.829	0.992	0.999
11. Adenylate kinase	0.862	0.965	0.951	0.994
12. Adenosine deaminase	0.828	0.931	0.933	0.899
13. Erythrocyte acid phosphatase	0.318	0.474	0.432	0.505
14. Esterase D	0.654	0.697	0.594	0.393
15. Glucose-6-phosphate dehydrogenase (male)	0.994	0.666	0.983	1.000
Glucose-6-phosphate dehydrogenase (female)	0.989	0.499	0.966	1.000
16. Phosphoglucomutase	0.480	0.535	0.481	0.471
17. 6-Phosphogluconate dehydrogenase	0.933	0.838	0.911	0.704
18. Glyoxalase I	0.379	0.410	0.420	0.751
19. Peptidase A	1.000	0.817	0.992	1.000
20. Glutamic pyruvic transaminase	0.380	0.483	0.388	0.382
Total for all 20 systems	9.65×10^{-7}	5.20×10^{-6}	3.23×10^{-6}	9.61×10^{-6}

variation that is observed throughout the world. The data employed to calculate the gene frequencies for Mexicans and Asians are not samples primarily from California (72% Mexico City [Mexicans] and 86% Hawaii [Asian]). The extent to which these populations are unrepresentative of a California population cannot be precisely assessed. However, the overall calculation of the discrimination probability is not highly affected by variation in gene frequencies. The degree of area-heterogeneity among the white and black samples in these data was investigated and found to be small or nonexistent.

Several other sources of lack of fit certainly exist but probably at low levels, particularly in human populations. Phenomena such as selection, mutation, and nonrandom sampling (with respect to blood types) are generally not strong biases in blood group data.

Other relevant sets of data must be found for calculations concerning individuals who come from populations differing in genetic composition.

The MNSs system yielded observed values that deviated from the expected Hardy-Weinberg equilibrium to an extent that cannot be entirely explained by misclassification, antisera problems, or racial admixture. This lack of fit associated with the MNSs has been observed by other investigators [*19*]. Race and Sanger [*19*] concur with Wiener that the antisera used for MNSs testing could produce some spurious results [*20*], but this has not been elaborated further in their latest edition. Among blacks the existence of a silent allele, called $S^{(u)}$, is a complicating factor and accounts for some of the lack of fit observed for the MNSs system. Silent alleles could also be a source of bias in other systems (for example, C^w in the Rh system).

As expected, 911 of the donors (15.3%) in this survey had one or more rare phenotypes.[5] A rare phenotype, as mentioned before, is defined as a phenotype that occurs in the population of 5956 donors with a frequency of less than 1%. Appearance of certain of these phenotypes was more frequent in particular racial subgroups, as for instance, Gc

[5]Combined data on rare variants found in this study and the earlier study [*1*] of blood samples from 11 678 donors will be published separately.

phenotypes in Asians or G-6-PD in blacks (Table 3). While rare variants in all the blood group systems are listed according to their phenotypes in Table 3, the rare variants for each protein and enzyme system are grouped together. For example, six rare phenotypes (3.2%) for the Tf system are grouped together and ten PGI variants are grouped together. Some of these were not identifiable because of a lack of a standard of comparison. However, in medicolegal analysis, the identification of a rare variant, whether known or unknown, significantly increases the individualization of a blood specimen.

Acknowledgments

Without the help of the directors and staff of cooperating blood banks this study would not have been possible. We gratefully acknowledge the invaluable assistance given to us by Dr. Vincent Caggiano, scientific director, The Sacramento Medical Foundation Blood Bank, Sacramento, Calif.; Dr. Julia Ann Frohlich, executive director, Blood Bank of Hawaii, Honolulu, Hawaii; Mrs. Beverly Hayhurst, manager, Blood Bank of San Bernardino-Riverside Counties, San Bernardino, Calif.; Dr. Thomas Hunt, director, Blood Bank of the Alameda-Contra Costa Medical Association, Oakland, Calif.; Dr. Hector Rodriguez Moyado, director, Del Banco Central De Sangre Del Centro Medico Nacional, Mexico City, Mex.; and Dr. Edwin Steane, University of Texas Health Sciences Center, Dallas, Tex.

We also acknowledge with thanks the technical assistance by M. Lavins, MPH, Ms. Priscilla Kalish, B.S., Edward J. Gorman, B.S., and Ms. Charmaine Pickett, MT (ASTP).

Addendum

It has been pointed out to us that in our previous paper (see Ref *1*, Table 7) the CB phenotype in the EAP system does not fit the Hardy-Weinberg equilibrium expected values. This lack of fit is not extreme (about 1% of the CB phenotypes were presumably misclassified) and is an inherent problem with EAP phenotyping (see *Journal of Forensic Sciences*, Vol. 23, No. 3, July 1978, pp. 615-618). However, the gene frequencies given in Table 7 of Ref *1* are only slightly affected. In the present paper, no evidence of misclassification exists for any of the EAP phenotypes.

References

[*1*] Grunbaum, B. W., Selvin, S., Pace, N., and Black, D. M., "Frequency Distribution and Discrimination Probability of Twelve Protein Genetic Variants in Human Blood as Functions of Race, Sex, and Age," *Journal of Forensic Sciences*, Vol. 23, No. 3, July 1978, pp. 577-587.

[*2*] Fisher, R. A., "Standard Calculation for Evaluating a Blood-Group System," *Heredity* (London), Vol. 5, 1951, pp. 51-102.

[*3*] Jones, D. A., "Blood Samples: Probability of Discrimination," *Journal of the Forensic Science Society*, Vol. 12, 1972, pp. 355-359.

[*4*] Sussman, L. N., *Paternity Testing by Blood Grouping*, 2nd ed., Charles C Thomas, Publisher, Springfield, Ill., 1976.

[*5*] Crawford, M. N., Gottman, F. E., and Gottman, C. A., "Microplate System for Routine Use in Blood Bank Laboratories," *Transfusion*, 1970, pp. 258-263.

[*6*] Grunbaum, B. W., *The Grunbaum System for Electrophoresis*, NASA Reference Publication 1040, National Aeronautics and Space Administration, Ames Research Center, Technology Utilization Office, Moffett Field, Calif.

[*7*] Grunbaum, B. W. and Zajac, P. L., "Rapid Phenotyping of the Group Specific Component by Immunofixation on Cellulose Acetate," *Journal of Forensic Sciences*, Vol. 22, No. 3, July 1977, pp. 586-589.

[*8*] Kalish, P. I. and Grunbaum, B. W., "Phenotyping of Transferrin in Fresh Blood and Bloodstains by Immunofixation on Sartorius Cellulose Acetate Membranes," presented at the 52nd Semiannual Seminar, California Association of Criminalists, Oct. 1978.

[9] Barnard, P. A. and Grunbaum, B. W., "Alteration of Electrophoretic Mobility of Hemoglobin in Bloodstains," *Journal of Forensic Sciences*, Vol. 24, No. 2, April 1979, pp. 384-388.
[10] Grunbaum, B. W., "A Microanalytical Technique for the Determination of Polymorphic Blood Proteins for Medical and Forensic Applications," *Mikrochimica Acta II*, 1977, pp. 339-352.
[11] Grunbaum, B. W. and Zajac, P. L., "Phenotyping of Erythrocyte Acid Phosphatase in Fresh Blood and in Bloodstains on Cellulose Acetate," *Journal of Forensic Sciences*, Vol. 23, No. 1, Jan. 1978, pp. 84-88.
[12] Grunbaum, B. W., Harmor, G. C., Del Re, B., and Zajac, P., "Electrophoresis of Esterase D in Fresh Blood and Bloodstains on Cellulose Acetate," *Journal of Forensic Sciences*, Vol. 23, No. 1, Jan. 1978, pp. 89-93.
[13] Grunbaum, B. W. and Zajac, P. L., "Differentiation of the Genetic Variants of Glucose-6-Phosphate Dehydrogenase in Bloodstains by Electrophoresis on Cellulose Acetate," *Journal of the Forensic Science Society*, Vol. 16, 1976, pp. 319-324.
[14] Grunbaum, B. W., "A Micro Procedure for Fast Typing of the Genetic Variants of Phosphoglucomutase," *Journal of the Forensic Science Society*, Vol. 14, 1974, pp. 151-157.
[15] Grunbaum, B. W., Kalish, P. I., and Barnard, P. A., "Phenotyping of Glutamic Pyruvic Transaminase and Glyoxalase-I in Fresh Blood and Bloodstains by Electrophoresis on Sartorius Cellulose Acetate Membranes," presented at the 52nd Semiannual Seminar, California Association of Criminalists, Oct. 1978.
[16] Grunbaum, B. W., Noppinger, K., and Morrison, R., "Phenotyping of Peptidase-A in Fresh Blood, Bloodstains and Seminal Fluid Stains by Electrophoresis on Sartorius Cellulose Acetate Membranes," presented at the 52nd Semiannual Seminar, California Association of Criminalists, Oct. 1978.
[17] Kendall, M. G. and Stuart, A., *Advanced Statistical Theory*, C. Griffin & Co., Ltd., London, 1967.
[18] Selvin, S., "A Note on Gene Frequency Estimations with Misclassification Present," *Human Heredity*, Vol. 24, 1980, in press.
[19] Race, R. R. and Sanger, R., *Blood Groups in Man*, 6th ed., Blackwell Scientific Publications, Oxford, London, 1975.
[20] Race, R. R. and Sanger, R., *Blood Groups in Man*, 5th ed., Blackwell Scientific Publications, Oxford, London, 1968.

Address requests for reprints or additional information to
Benjamin W. Grunbaum, Ph.D.
Environmental Physiology Laboratory Bldg. T-2251
University of California
Berkeley, Calif. 94720