Am. J. Hum. Genet. 50:145–155, 1992

Caucasian Genes in American Blacks: New Data

Ranajit Chakraborty,* Mohammad I. Kamboh,† M. Nwankwo,‡ and Robert E. Ferrell†

*Center for Demographic and Population Genetics, University of Texas Graduate School of Biomedical Sciences, Houston; †Department of Human Genetics, University of Pittsburgh Graduate School of Public Health, Pittsburgh; and ‡University of Benin, Benin City, Nigeria

Summary

Data on 15 polymorphic protein-coding loci are used to estimate the proportion of Caucasian genes in U.S. blacks from the greater-metropolitan area of Pittsburgh. Allele frequencies from U.S. whites of the same region and from a sample of Nigerians are used as representatives of the genetic contributions of the source populations between which admixture has occurred. These materials provide 18 unique variants that occur exclusively in the blacks and 5 variants that are restricted to the Caucasians only. As a result, when all segregating alleles (52) at these 15 loci are considered, the proportion (mean ± SE) of Caucasian genes in U.S. blacks (25.2% ± 2.7%) is estimated with a precision much better than that of all other previous estimates. The estimate based on the frequencies of these 18 unique variants of African origin (24.8% ± 6.2%) is also consistent with the pooled estimate obtained from the 15 loci by the weighted least-square method. The homogeneity of locus-specific estimates of admixture indicates that these loci are appropriate for studying the proportion of black genes in any admixed population involving African admixture. The advantages of employing such loci for genetic-epidemiologic studies in U.S. blacks is discussed in the context of the availability of these large number of unique African variants at these protein loci.

Introduction

The estimation of the proportion of Caucasian genes in the U.S. black populations has been a subject of considerable discussion in the genetic and anthropological literature for almost 5 decades, because of the social, anthropological, genetic, epidemiological, and historical demographic implications of such estimates. Perhaps U.S. blacks are the best-studied human admixed groups, since the origin and evolution of admixture among this group are better documented than the occurrence of admixture in other parts of the world. The availability of shipping lists, records of geographical origin of the imported slaves, and notes on areas of placement of the slaves in the American continent provide direct verification of some of the assumptions of admixture estimation for American blacks (Adams and Ward 1973). In spite of these advantages, several studies noted discrepancies between the estimates of admixture as well as of constituents of the gene pool in U.S. blacks of different regions of the country (Glass and Li 1953; Glass 1955; Roberts 1955; Reed 1969).

More recent compilations (Adams and Ward 1973; Chakraborty 1986) of admixture estimates in U.S. blacks suggest not only that these estimates were different when different sets of loci were employed but also that there are different trends for interlocus variation and for rural/urban classification of the communities. In an incisive study, Adams and Ward (1973) reevaluated these aspects of the extent of Caucasian admixture in U.S. blacks, considering the proportion of slaves from different geographic regions of Africa.

One limitation that is common to all admixture studies thus far conducted on U.S. blacks is that they used allele frequency data on loci that rarely showed unique alleles of African origin, as a result of which the admixture estimates had always a large standard error associated with them. With the exception of the hemoglobin-S allele, peptidase-C_0 allele, and Duffy-null variant, no other unique African allele was de-

Received March 15, 1991; revision received September 3, 1991.

Address for correspondence and reprints: Dr. Ranajit Chakraborty, Center for Demographic and Population Genetics, University of Texas Graduate School of Biomedical Sciences, P.O. Box 20334, Houston, TX 77225.

0002-9297/92/5001-0015$02.00

tected in U.S. blacks, hampering the prospect of reliable single-locus estimation of Caucasian admixture in U.S. blacks. With the discovery of new genetic polymorphisms and the use of isoelectric-focusing/immunoblot methods to improve both resolution and sensitivity in assaying genetic variability, this difficulty can be partially circumvented.

In the present paper, our objective is to use some new data to provide a revised estimate of Caucasian admixture in U.S. blacks sampled from the metropolitan Pittsburgh area. Recent surveys on 15 protein loci revealed 18 alleles that are presumably of African origin, since they were not encountered in the U.S. white population (for references, see table 1). We use the allele frequency data at these 15 loci and show that the estimate of Caucasian admixture in this U.S. black population, an estimate derived from the 18 unique African alleles, agrees fairly well with that obtained from all allele frequencies at these loci. The interlocus estimates of admixture are shown to be statistically homogeneous, suggesting that these loci are appropriate for studying Caucasian admixture in U.S. blacks. While the combined estimate of admixture is generally consistent with the ones suggested before, particularly for U.S. blacks from the northern part of the United States, the present estimate, based on the 15 loci that contain the unique African alleles, is more precise than the ones based on the traditional loci. This implies the advantage of using these discriminatory loci in future admixture studies in U.S. blacks. Finally, we suggest that the availability of such a large number of unique African alleles opens the possibility of estimation of admixture in individual U.S. blacks, which might be of considerable epidemiological interest in tracing the etiology of chronic diseases that occur in U.S. blacks at substantial frequencies (Chakraborty 1986).

Material and Methods

Data

Phenotype data were collected on 15 plasma protein markers (table 1) for black and white blood donors at the Central Blood Bank, Pittsburgh. All participants of the study were residents of the Pittsburgh metropolitan area, and there was no selection beyond their volunteer status. As far as could be determined from the information given by the subjects from whom blood samples were collected, no relatives were included in the analysis. Allele frequencies from the Pittsburgh whites were used as the white parental frequencies in admixture estimation. The African sample consisted of two groups. The first group were first-year medical nursing and midwifery students at the University of Benin in Benin City, Nigeria. The second group consisted of volunteers drawn from civil service workers from Benin City and of teachers from the Oroyo College Secondary School in Benin City. The two groups did not show any significant allele frequency differences at the loci studied, and, hence, the present analysis includes the allele frequency data on a total of 356 unrelated individuals from the pooled sample of these two groups, representing the African parental population frequencies in the admixture computation. Possible pitfalls of using such data to represent the ancestral populations between which admixture might have occurred are addressed in the Discussion section. Descriptions of the sample populations are given by Adams-Campbell et al. (1988) and Bunker et al. (1990). All allele frequencies used in the present analysis are based on the gene-count method of estimation, and the estimates of allele frequencies for the X-linked locus (TBG) are the averages over the males and females included in the sample.

The 15 loci tested yield a total of 52 segregating alleles, of which 18 are unique to blacks and of which 5 are unique to whites. The gene products of all 15 loci were screened using isoelectric-focusing and/or immunoblot techniques. The reference sources of allele frequencies in U.S. blacks, African blacks, and U.S. whites, as well as other characteristics of the studied loci, are presented in table 1. Details of experimental protocols for assaying these loci are also given in the references listed in table 1.

Statistical Methods

Since our objective in the present paper is to estimate the proportion of Caucasian admixture in U.S. blacks by utilizing allele frequency data at loci which contain unique African alleles, we used two different approaches. In the first approach, only unique African allele frequencies are considered, while, in the second approach, all allele frequencies are taken into account. For both approaches, several alternative methods of estimation could have been adopted (for reviews of alternative approaches, see Szathmary and Reed 1978; Chakraborty 1986). In the present work, we adopted a weighted least-square method for both approaches, because (*a*) this method yields simple estimators, (*b*) error variance can be computed easily, and (*c*) a heterogeneity test of estimates based on different loci can

Table 1

Characteristics of Marker Loci Used in Analysis

Locus (abbreviation)	Chromosome Location	No. of Alleles Seen across Three Samples	Unique Alleles in		Source of Data
			Blacks	Whites	
Factor XIIIB (FXIIIB).....	1	5	FB*6 and FB*23	FB*3	Kamboh and Ferrell 1986*a* Kamboh and Ferrell 1989
Apolipoprotein D (APO D).....	3	2	APOD*2	. . .	Kamboh et al. 1989*a*, Kamboh and Ferrell 1990
Transferrin (TF).....	3	4	TF*D1	TF*C3	Kamboh and Ferrell 1987*b*
Vitamin D–binding protein (GC)	4	4	GC*1A1	. . .	Kamboh and Ferrell 1986*a* M. I. Kamboh and R. E. Ferrell, unpublished data
Orosomucoid (ORM1)...	9	2	. . .	. . .	Escallon et al. 1987
Orosomucoid (ORM2)...	9	4	ORM2*2, ORM2*3, and ORM2*4	. . .	Escallon et al. 1987
Hemopexin (HPX).....	11	3	HPX*2 and HPX*3	. . .	Escallon et al. 1987 M. I. Kamboh and R. E. Ferrell, unpublished data
Apolipoprotein A-IV (APO A-IV)	11	5	APOAIV*5	APOAIV*2	Kamboh and Ferrell 1987*a*
Complement subcomponent (CIR)	12	3	C1R*5	. . .	Kamboh et al. 1989*b*
α-1-Antitrypsin (PI)	14	5	. . .	PI*S and PI*Z	DeCroo et al., in press
Apolipoprotein H (APO H)	17	4	APOH*4	. . .	Kamboh and Ferrell 1987*a*
Apolipoprotein C-II (APO C-II).....	19	3	APOCII*2 and APOCII*3	. . .	Kamboh and Ferrell 1987*a*
Apolipoprotein E (APO E).....	19	3	. . .	. . .	Kamboh and Ferrell 1987*a*
Thyroxin-binding globulin (TBG).....	X	2	TBG*S	. . .	Kamboh and Ferrell 1986*c* M. I. Kamboh and R. E. Ferrell, unpublished data
Apolipoprotein (LDL)	Unknown	3	LDL*2 and LDL*3	. . .	M. I. Kamboh and R. E. Ferrell, unpublished data

be performed using large-sample properties. Since the estimates of the unique African allele frequencies are subject to comparatively larger sampling error (in terms of coefficient of variation), we suggest a method of admixture estimation from the unique alleles that takes into account errors of estimating such allele frequencies in the U.S. blacks as well as in the African ancestral population. In this section we present, for easy reference, the closed-form expressions of the estimators and their approximate standard errors, since in earlier work they are not readily available.

Estimate Based on Unique African Alleles

Let r denote the total number of unique African alleles. The true frequencies of the ith such allele in the U.S. blacks and in the ancestral African population are represented by h_i and a_i, respectively. Since these alleles are assumed to be absent in the Caucasian population, the proportion μ of Caucasian genes in U.S. blacks is related to h_i and a_i by the structural relationship

$$h_i = (1-\mu)a_i, \qquad (1)$$

so that μ can be estimated by fitting a linear-regression equation (passing through the origin) of h_i on a_i for $i = 1, 2, \ldots, r$. Although a simple least-square estimate μ (Snedecor and Cochran 1967) can be easily obtained from equation (1), the situation is somewhat complicated, since h_i and a_i values are not known without error. We have only estimates of these values, from samples drawn independently from U.S. blacks and from an African population. Assume that estimates of h_i (say, y_i) and a_i (say, x_i) are available from samples of size n_i and m_i individuals randomly drawn from U.S. blacks and an African population. We can, therefore, write

$$x_i = a_i + u_i, \tag{2a}$$

and

$$y_i = h_i + v_i, \tag{2b}$$

for $i = 1, 2, \ldots, r$; where u_i and v_i are the errors of estimation for the unique African alleles observed in the samples. Since in the present work all allele frequencies are estimated by the gene-count method, we can assume that u_i and v_i both have expectation zero and that error variances are given by

$$V(u_i) \approx x_i(1 - x_i)/2m_i, \tag{3a}$$

and

$$V(v_i) \approx y_i(1 - y_i)/2n_i, \tag{3b}$$

respectively. The average values of these error variances over all r unique African alleles are denoted by σ_u^2 and σ_v^2, respectively, and the ratio

$$\lambda = \sigma_v^2/\sigma_u^2 \tag{4}$$

is assumed to be known without error. Under this formulation, Madansky (1959) and Brown (1982) showed that a robust estimator of $1 - \mu$ ($= \alpha$, say) is given by

$$\hat{\alpha} = [\delta + \delta^2 + 4\lambda s_{xy}^2)^{1/2}]/2s_{xy}, \tag{5}$$

where

$$\bar{x} = \sum_{i=1}^{r} x_i/r,$$

$$\bar{y} = \sum_{i=1}^{r} y_i/r,$$

$$s_{xx} = \sum_{i=1}^{r} (x_i - \bar{x})^2/r,$$

$$s_{yy} = \sum_{i=1}^{r} (y_i - \bar{y})^2/r,$$

$$s_{xy} = \sum_{i=1}^{r} (x_i y_i - \bar{x}\bar{y})/r,$$

and

$$\delta = s_{yy} - \lambda s_{xx}.$$

Madansky (1959, p. 180) also gave an expression for the approximate error variance of $\hat{\alpha}$ by following the suggestion of Creasy (1956), which requires the evaluation of 95% confidence limits of $\Phi = \tan^{-1}(\alpha/\sqrt{\lambda})$. Since $\mu = 1 - \alpha$, the complement of equation (5) gives the weighted least-square estimate of μ, whose sampling error can be evaluated by the method of Creasy (1956) and Madansky (1959). Note that, although this method is comparatively more tedious than the usual least-square estimator (Snedecor and Cochran 1967, p. 166), this is more appropriate for the present analysis, since the allele frequency estimates from both populations involve sampling errors.

Estimate Based on All Allele Frequencies

While the above method uses only the unique African alleles, so that allele frequency data are needed only from the samples from U.S. blacks and the African ancestral population, to utilize data on all alleles at several loci we need samples from the Caucasian ancestral population as well. Using a somewhat different notation, let us denote the frequencies of the ith allele at the ℓth locus in the three populations by $P_{hi}^{(\ell)}$ (for U.S. blacks), $P_{ai}^{(\ell)}$ (for the African ancestral population) and $P_{ci}^{(\ell)}$ (for the Caucasian ancestral population). Suppose that we have allele frequency data on L loci (so that $\ell = 1, 2, \ldots, L$) and that there are r_ℓ segregating alleles at the ℓth locus (i.e., i = 1, 2, . . . , r_ℓ for a fixed ℓ). Long and Smouse (1983) and, more recently, Long (1991) have suggested a weighted least-square estimate of admixture proportion for such a dihybrid admixed population, an estimate whose closed form expression is given by

$$\hat{\mu} = \frac{\sum_{\ell=1}^{L}\sum_{i=1}^{r_\ell}(P_{ci}^{(\ell)} - P_{ai}^{(\ell)})(P_{hi}^{(\ell)} - P_{ai}^{(\ell)})/P_{hi}^{(\ell)}}{\sum_{\ell=1}^{L}\sum_{i=1}^{r_\ell}(P_{ci}^{(\ell)} - P_{ai}^{(\ell)})^2/P_{hi}^{(\ell)}}. \quad (6)$$

The sampling variance of $\hat{\mu}$ is

$$V(\hat{\mu}) = MSE/\sum_{\ell=1}^{L}\sum_{i=1}^{r_\ell}(P_{ci}^{(\ell)} - P_{ai}^{(\ell)})^2/P_{hi}^{(\ell)}, \quad (7)$$

where MSE (mean square error) is given by

$$\frac{\sum_{\ell=1}^{L}\sum_{i=1}^{r_\ell}[(P_{hi}^{(\ell)} - P_{ai}^{(\ell)}) - (\hat{\mu}P_{ci}^{(\ell)} - P_{ai}^{(\ell)})]^2/P_{hi}^{(\ell)}}{r-L}, \quad (8)$$

and $r = \sum_{\ell=1}^{L} r_\ell$ is the total number of alleles at all L loci.

One interesting property of this weighted least-square estimate is that such estimators can be defined for each individual locus as well, simply by dropping the summations over loci in equation (6), to get locus-specific estimates of μ, denoted by $\hat{\mu}_\ell$. The sampling variance of $\hat{\mu}_\ell$ can be computed as

$$V(\hat{\mu}_\ell) = MSE/\sum_{i=1}^{r_\ell}(P_{ci}^{(\ell)} - P_{ai}^{(\ell)})^2/P_{hi}^{(\ell)}, \quad (9)$$

in which MSE remains the same as in equation (8). When the L loci are independently segregating, a test of homogeneity of locus-specific estimators can be conducted by the statistic

$$\chi^2_{(L-1)} = \sum(\hat{\mu}_\ell - \hat{\mu})^2/V(\hat{\mu}_\ell), \quad (10)$$

which, for large samples, should approximately follow a χ^2 distribution with $L-1$ df (Cavalli-Sforza and Bodmer 1971).

Results

Estimate Based on Unique African Alleles

Table 2 presents the allele frequency estimates for each of the 18 unique African alleles that are found at the 15 loci surveyed. The standard errors of these estimates were calculated from the standard binomial theory (eq. [3a] and [3b]).

Figure 1 shows the plot of this relationship, where

Table 2

Frequencies of Unique African Alleles among West Africans and U.S. Blacks

	Africans		U.S. Blacks	
Allele	N^a	Mean Frequency ± SE	N^a	Mean Frequency ± SE
HPX*2	382	.017 ± .005	193	.018 ± .007
HPX*3	382	.089 ± .010	193	.046 ± .011
LDL*2	191	.298 ± .023	67	.216 ± .036
LDL*3	191	.000	67	.008 ± .008
APOCII*2	194	.049 ± .011	137	.025 ± .009
APPCII*3	194	.003 ± .003	137	.000
APOD*2	364	.022 ± .005	263	.013 ± .005
ORM2*2	187	.005 ± .004	181	.025 ± .008
ORM2*3	187	.005 ± .004	181	.006 ± .004
ORM2*4	187	.003 ± .003	181	.011 ± .006
TBG*S	45	.080 ± .029	101	.120 ± .023
APOH*4	167	.015 ± .007	148	.013 ± .007
APOAIV*5	171	.015 ± .007	127	.008 ± .006
FB*6	370	.036 ± .007	220	.028 ± .008
FB*23	370	.023 ± .006	220	.011 ± .005
C1R*5	247	.016 ± .006	136	.011 ± .006
GC*1A1	750	.025 ± .004	273	.015 ± .005
TF*D1	131	.080 ± .017	194	.030 ± .009
Average		.043 ± .016		.034 ± .012

[a] Number of individuals sampled.

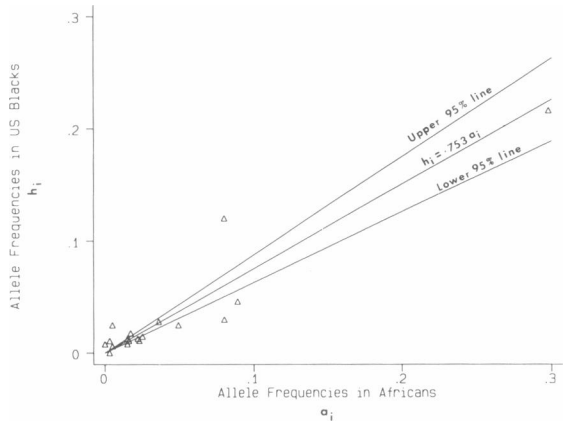

Figure 1 Relationship of frequencies of unique African alleles in U.S. blacks and in west Africans. The middle solid line is the fitted regression line (eq. [5]), and the other two solid lines are the upper and lower 95% confidence limits of the regression equation that pass through the origin. The slope (α) is the complement of μ.

the slope of the regression equation ($\hat{\alpha} = 0.752$) is based on equation (5). This is the complement of the estimated μ (mean $\pm$ SE), which becomes .248 $\pm$ .062. Although some of the data points are outside the 95% confidence limit of the regression equation, most of the deviant points have also large standard errors in both directions. In general the fit of the regression line (of frequencies of unique African alleles in U.S. blacks, on the same in the African population) is statistically adequate, since deviation from linearity passing through the origin (Student's t-statistic with 16 df = 0.51) is not significant ($P > .69$).

Since our compilation of allele frequencies at these 15 loci also revealed five unique Caucasian alleles not found in the Nigerians, we can use the above approach to estimate μ by using these five unique Caucasian alleles as well. Table 3 shows the frequencies (and sample sizes) for these five alleles in the U.S. black and U.S. white samples. Using equation (5) on these data, we get μ directly, giving $\hat{\mu} = .301 \pm .051$ as the estimate of Caucasian genes in U.S. blacks. Although this estimate is somewhat higher than the one obtained from the unique African alleles, their difference is not statistically significant.

Estimate Based All Alleles at the 15 Loci

When all allele frequency data at the 15 loci were used, the weighted least-square estimate of μ (by eq. [6] and [7]) becomes .252 $\pm$ .027. Note that while this point estimate is very close to the one obtained by using the 18 unique African alleles ($\hat{\mu} = .248$), the precision of this estimator is much improved when all data are utilized. Since the least-square estimation of admixture proportion, when all allele frequency data are used, presupposes a linearity of $(P_{hi}^{(l)} - P_{ai}^{(l)})$ with $(P_{ci}^{(l)} - P_{ai}^{(l)})$ for all alleles at all loci, figure 2 plots this relationship for the 52 alleles encountered at the 15 loci surveyed. The regression line (with a slope of 0.252) is obviously in good agreement with the data points, suggesting that the admixture model is adequate for the present data.

Homogeneity of Interlocus Estimates

Table 4 shows the summary of the homogeneity test (eq. [10]) of the locus-specific estimates of μ when the least-square approach is used. It is clear that, although two loci (ORM2 and TBG) gave estimates outside the valid range ($0 < \mu < 1$), their deviations from the pooled estimate ($\hat{\mu} = .252$) are not significant, judged from the individual χ^2 values. Furthermore, note that, even though the pooled estimate is quite precise, the

Table 3

Frequencies of Unique Caucasian Alleles among U.S. Whites and U.S. Blacks

	U.S. Whites		U.S. Blacks	
Allele	N^a	Mean Frequency $\pm$ SE	N^a	Mean Frequency $\pm$ SE
FB*3	171	.202 $\pm$ .022	260	.056 $\pm$ .010
TF*C3	300	.040 $\pm$ .008	194	.010 $\pm$ .005
APOAIV*2	159	.088 $\pm$ .016	127	.035 $\pm$ .012
PI*S	192	.021 $\pm$ .007	96	.015 $\pm$ .009
PI*Z	192	.005 $\pm$ .004	96	.0
Average		.071 $\pm$ .036		.023 $\pm$ .010

[a] Number of individuals sampled.

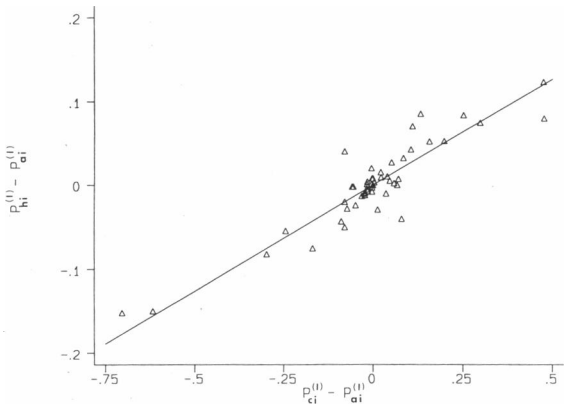

Figure 2 Relationship between $(P_{hi}^{(l)} - P_{ai}^{(l)})$ and $(P_{ci}^{(l)} - P_{ai}^{(l)})$ for 52 alleles at 15 polymorphic loci. This relationship should be linear (shown by the solid line) with a slope of $\hat{\mu} = .252$.

standard errors of the locus-specific estimates are rather high. Nevertheless, the homogeneity of the locus-specific estimates is revealed by the pooled χ^2 (with 14 df), whose observed value 10.65 is not statistically significant ($P > .69$). This homogeneity, along with the good fit of allele frequency differences (fig. 2) suggest that these 15 loci, most of which contain unique African alleles, are appropriate for studying the extent of African genes in admixed black populations.

Table 4

Locus-specific Estimates of μ, and Test of Homogeneity of These Estimates

Locus (no. of alleles)	Mean $\hat{\mu}_{(l)} \pm$ SE	$\chi^2(l)$
HPX (3)	.434 ± .173	1.15
LDL (3)	.269 ± .181	.01
APO-CII (3)	.503 ± .289	.76
APO-D (2)	.409 ± .341	.21
ORM1 (2)	.034 ± .658	.11
ORM2 (4)	−1.202 ± 1.018	2.04
TBG (2)	−.500 ± .435	2.99
APO-H (4)	.117 ± .229	.35
APO-E (3)	.225 ± .250	.01
APO-AIV (5)	.395 ± .193	.55
FXIIIB (5)	.264 ± .046	.07
CIR (3)	.178 ± .421	.03
PI (5)	.245 ± .137	.00
GC (4)	.214 ± .041	.85
TF (4)	.385 ± .107	1.56
Pooled (52)	.252 ± .027	10.65[a]

[a] Pooled χ^2 has 14 df, and hence it is not significant ($P > .69$).

Discussion and Conclusion

In the present work our main purpose has been to show that with the modified approach of electrophoretic detectability of alleles it is now possible to detect a large number of unique alleles ("private" variants in the terminology of Neel [1973]) that are of African origin. We detected 18 such alleles at 15 loci, which greatly enhanced the reliability of estimating African ancestry of admixed black populations such as the U.S. blacks of the metropolitan Pittsburgh area. Our estimate of the μ (.252 ± .027), based on all allele frequency data at the 15 loci examined here, agree well with previous estimates (Glass 1955; Adams and Ward 1973), although the loci employed here are different from the ones utilized by these previous workers. The increased precision of our estimate is due to the presence of both a large number of unique African alleles (18) and five unique Caucasian alleles at these loci. Even though we used the weighted least-square method for the present work, because it provides the additional test of homogeneity of locus-specific estimators, other methods of estimating admixture on the basis of the data also give similar results. For example, when Chakraborty's (1985) gene-identity method is employed, the present data yield the weighted least-square estimate (based on gene identity) of $\hat{\mu} = .236 \pm .005$, which is even more precise.

The congruence between such estimates based on the pooled data and the ones obtained from only the unique alleles (African as well as Caucasian) is reassuring in spite of the fact that the unique-allele frequencies are subject to larger coefficient of variation due to estimation error. The use of unique alleles in estimating admixture proportions that was originally suggested by Szathmary and Reed (1978) is somewhat different from the one presented here. Their method yields an estimate of maximum possible admixture, while the present method gives a point estimate. Byard et al. (1985) and Williams et al. (1986) used unique-allele frequencies in estimating the degree of Caucasian admixture in some Amerindian groups, giving point estimates. Their method does not take into account the errors of estimation of the unique-allele frequencies, while the present method does. As a result of such measuremental errors, the precision of the estimate based on the unique African alleles is somewhat lower than that of the pooled estimate. Nevertheless, the congruence with the pooled estimate is not distorted, irrespective of the methods of estimation.

It is worthwhile to mention here a comment about

the method of incorporating the sampling errors of variables in fitting a linear-regression equation. The estimator suggested by Madansky (1959), given in equation (5), was called, by Brown (1982), the "best" maximum likelihood (BML) estimator, under the assumption that the errors of measurement (u_i and v_i) have a bivariate normal distribution with a zero correlation and that the ratio $\lambda = \sigma_v^2/\sigma_u^2$ is known without error. While the independence of u_i and v_i is assured when the samples from the two populations are independently chosen, λ is not strictly known in our case. We estimated λ from the average error variances of the estimates of the unique alleles. Through a series of simulations Brown (1982) showed that the BML estimator is robust for such imperfect knowledge of λ, particularly when the sample sizes are large. Since for most of the loci the number of individuals sampled is larger than 100, we argue that this approximation is reasonable.

The sampling variance of the estimator given in equation (5) is perhaps an underestimate of the true variance, because σ_u^2 and σ_v^2 used in estimating this quantity are associated with allele frequencies that are rare. Alternative formulations of estimating error variance that are based on the observed numbers of heterozygotes and/or homozygotes in the samples may provide somewhat improved estimators. However, since the genotype distributions at these loci in all three samples did not significantly differ from their Hardy-Weinberg expectations (data not shown), we surmise that the underestimation of error variance is not a serious problem in the context of the present data. We might also note that, in view of the possible unreliability associated with the small numbers of heterozygotes or homozygotes which contributed to the rare-unique-allele frequencies, the incorporation of the sampling errors of x and y in estimating the admixture proportion, as suggested here, is important even though a direct application of a weighted regression (which ignores both sampling variances) is easier because of its availability in statistical packages. (A copy of the FORTRAN-77 algorithm of the method proposed here may be obtained from R.C. on request.) The homogeneity of locus-specific estimates (table 4) shows that the loci containing the unique alleles should be particularly useful for estimating the African ancestry for any admixed black population, as long as the other parental populations contributing such admixture do not possess these unique alleles.

The present study suffers from a limitation common to all studies of admixture. Seldom are allele frequencies in the "parental" populations known precisely, and the definition of "parental" population(s) contributing to any admixed group is often uncertain. This is certainly true of the present study. African slaves originated from a variety of geographic locales, and their disposition in the New World was not random. Local or regional variation in estimates of allele frequency for the loci used in the present work cannot be assessed from the limited population genetic data available on undisturbed African populations. Thus, a single African population sample is, at best, a poor representative of the true parental population. The only approach may be to employ a large number of loci, in the hope that local fluctuations in allele frequencies will average out over a larger set of markers. Even where data on African geographical origins and disposition of slaves in the New World are available from historical records, subsequent large-scale migrations of American blacks have further homogenized the population. In this regard, Reed (1969) estimated that 20% of U.S. blacks originated in central Nigeria and that 46% were from countries west of Nigeria. Nigeria is in the west-African linguistic group, which is more homogeneous than groups to the east and south and, hence, perhaps best approximates the true unknown situation.

The same problems plague, perhaps to a lesser extent, the choice of a "parental" white population representative of the population contributing to the admixture. Older industrial American cities such as Pittsburgh attracted European immigrants because they offered economic opportunities. These immigrants maintained their cultural separateness, still reflected in the urban neighborhoods of Pittsburgh. American blacks were attracted to the region for the same, primarily economic, reasons as were European immigrants, and the extent and nature of non-African admixture in the present-day Pittsburgh black population represents admixture events spanning the whole history of blacks in America. The contemporary white population of Pittsburgh is probably not representative of the "parental" population contributing to the past admixture.

Given these limitations, there seems to be only two alternatives—(1) to abandon efforts to use admixture as a tool in understanding ethnic variation in disease or (2) to accept the weaknesses in the approach and cautiously use genetic data to ask, in a general sense, whether admixture plays a role in determining the distribution of disease in a mixed population. In the latter case, consistency of data across studies may increase

our confidence in the conclusions. The Nigerian population used in the present study is a mixed urban population representing a variety of ethnic subgroups of the Nigerian population and to some extent represents the potential variation in allele frequencies in a broader geographical area. The degree to which it is representative of the founders of the American black population is unknown. We argue that the use of a large number of loci characterized by unique alleles from either of the parental populations will yield more accurate admixture estimates for the study of the genetic etiology of common diseases.

Keeping in mind such caveats, we note two other important implications of data on unique alleles. First, there are suggestions that many chronic diseases occur in high frequencies in the admixed black populations, in comparison with other populations of the same environment. Hypertension in the continental United States is a good example of this phenomenon (MacLean et al. 1974; Harburg et al. 1978; Dischinger et al. 1981). Some previous reports have argued that this may be due to the fact that the arrival of the Africans also brought hypertension-susceptibility genes to this continent. This argument is contested by others, who show (*a*) that not all African populations exhibit an increased level of hypertension in the African continent (Kaminer and Lutz 1960) and (*b*) not all individuals of African origin experience the marked increase in blood pressure with advanced age that is evident among the black Americans (Abrahams et al. 1960; Oyediran et al. 1976; Oviasu 1978). The caveat of these studies is that, since blood pressure is probably under polygenic control with a considerable effect of environment, a comparison of prevalence of hypertension across populations is not quite appropriate. This is so because the variation of contributing environmental factors is difficult to measure, nor can it be appropriately controlled. On the other hand, if the African-ness of admixed black individuals from the same environment can be measured by estimating the proportion of African genes at the individual level, then the question of whether a higher susceptibility of hypertension is of African origin can be answered more definitively. The large number of unique African alleles demonstrated in the present work leads to this possibility. For example, for a sample of U.S. blacks who live in a relatively homogeneous environment and on whom blood-pressure measurements are available, we might ask whether the individuals belonging to different deciles of the blood-pressure distribution have equivalent numbers of unique African alleles. Combined with the present 18 such unique African alleles, the total number of recognized unique African alleles is now approaching two dozen, making such a study feasible with enough statistical power. This would be a study parallel to the trend of an increase in the susceptibility to non-insulin-dependent diabetes mellitus in admixed Amerindian communities in the New World, a trend observed with an increasing level of Amerindian admixture (Weiss et al. 1984; Chakraborty et al. 1986; Chakraborty and Weiss 1986). Such studies on hypertension and African admixture should yield new data pertinent to the possible resolution of the above controversy and can provide insight pertinent to the etiology of hypertension.

The second implication of the discovery of such a large number of unique African alleles relates to the issue that, in the context of genetic, epidemiologic, and forensic applications of genetic polymorphism, it is often necessary to establish the genetic homogeneity of populations, particularly when the populations are admixed. Individuals of a homogeneous admixed population may differ in their level of admixture, but the distribution of individual admixture proportions should have a predictable form under genetic homogeneity, whose expectation can be computed from the group-admixture proportion in the entire population. With the help of the unique alleles in one of the ancestral populations, this expected distribution can be evaluated. For example, under the assumption that U.S. blacks of metropolitan Pittsburgh are a homogeneous admixed population with approximately $\alpha = 75\%$ African ancestry, the probability that an individual will have i unique alleles in his or her genome is given by the Poisson distribution with mean $2\alpha X$, where X is the summed allele frequency of all recognized unique alleles in the African population (for this assumption, whose validity is demonstrated by Chakraborty et al. [1991], see Szathmary and Reed [1978]).

$$Q_i = e^{-2\alpha X}(2\alpha X)^i / i! , \qquad (11)$$

for $i = 0, 1, 2$, etc. Unfortunately, in the present work the same individuals were not assayed for all of the 15 loci on which such a distribution can be fitted. But our estimate of $\alpha = .75$ (approximately) leads to an expected distribution for the 18 unique African alleles (listed in table 2) shown in figure 3. The shape of this distribution of individual admixture is in sharp contrast with the U-shaped distribution of individual admixture proportions published in the literature (Chakraborty et al. 1986; Hanis et al. 1986). This is

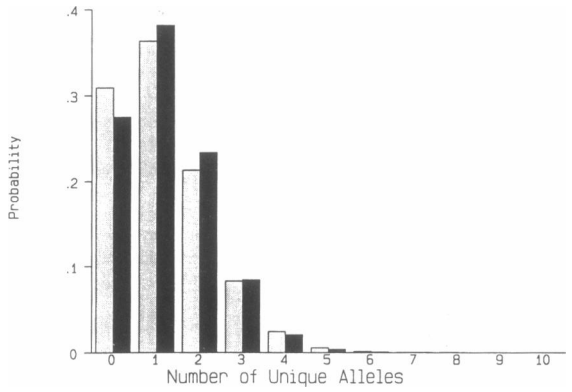

Figure 3 Expected distribution of number of unique African alleles among individuals in population with 75% African genes. The dark bars show the exact expected distribution as derived by Chakraborty et al. (1991), and the lightly shaded bars represent the Poisson approximation (eq. [11]).

so because in these previous works the loci utilized did not contain any unique variant, and hence the genotypes of individuals in an admixed group often were not informative about the individual admixture. A direct assay of these unique African alleles in future works on admixed black populations should help in establishing the genetic homogeneity in view of this predictable distribution shown in figure 3.

In implementing either of these suggestions, one must recognize that the reliability of individual admixture estimates is mitigated both by the unknown ancestry of the parental populations between which admixture may have occurred and by the sampling errors associated with the estimation of rare frequencies of the unique alleles. Empirical studies demonstrate that rare alleles show considerably more geographic variation in their frequencies and that hence, unless a large number of them are available, the conclusions drawn from studies such as the ones mentioned above may be somewhat misleading unless the parental origins of the admixed population are precisely known and unless rare allele frequencies are derived from appropriate reference population.

Finally, we conclude with the cautionary remark that the technique of electrophoresis may not be the best for detecting unique alleles, since this technique does not detect all genetic variation at the DNA level. In spite of the fact that the improvement of electrophoresis through isoelectric focusing revealed at least 18 unique African alleles not found in Caucasians, we have at least one example where a variant that migrated equally in the African and Caucasian samples is perhaps different. This was the case with the FB*6 variant, which we included as an unique African allele, even though in a previous work (Kamboh and Ferrell 1989) this variant was assigned a nonzero frequency (<.01) among the Caucasian populations. With the prospect of discovering more unique alleles by DNA sequencing, the potential of using the unique variants in tracing admixture between populations and in examining genetic homogeneity of populations by studying variation of their occurrences among individuals should be greatly improved.

Acknowledgments

This study was supported by U.S. Public Health Service Research grants NIH GM-41399 and NIJ IJ-90-CX-0038 to R.C. We thank Drs. C. L. Hanis and E. Boerwinkle for their comments on an earlier draft of the manuscript.

References

Abrahams D, Alele CA, Bernard BG (1960) The systemic blood pressure in a rural West African community. W Afr Med J 9:45–58

Adams J, Ward RH (1973) Admixture studies and detection of selection. Science 180:1137–1143

Adams-Campbell LL, Nwankwo M, Omene J, Young FA, Ukoli FA, Haile GT, Kuller LH (1988) An assessment of cardiovascular risk factors in Nigerian medical students. Arteriosclerosis 8:793–796

Brown ML (1982) Robust line estimation with errors in both variables. J Am Stat Assoc 77:71–79

Bunker CH, Nwankwo MU, Ukoli FA, Omene JA, Currier GW, Holifield LR, Freeman DT, et al (1990) Factors related to blood pressure among urban Nigerian workers. J Hum Hypertens 4:82–84

Byard PJ, Schanfield MS, Crawford MH (1985) Admixture and heterozygosity in west Alaskan populations. J Biosoc Sci 15:207–216

Cavalli-Sforza LL, Bodmer WF (1971) The genetics of human populations. WF Freeman, San Francisco

Chakraborty R (1985) Gene identity in racial hybrids and estimation of admixture rates. In: Neel JV, Ahuja Y (eds) Genetic microdifferentiation in man and other animals. Indian Anthropological Association. Delhi University, Delhi, pp 171–180

——— (1986) Gene admixture in human populations: models and predictions. Yearbook Phys Anthropol 29:1–43

Chakraborty R, Ferrell RE, Stern MP, Haffner SM, Hazuda HP, Rosenthal M (1986) Relationship of prevalence of non–insulin dependent diabetes mellitus with Amerindian admixture in the Mexican Americans of San Antonio, Texas. Genet Epidemiol 3:435–454

Chakraborty R, Kamboh MI, Ferrell RE (1991) Unique alleles in admixed populations: a strategy for determining "hereditary" population differences of disease frequencies. Ethnicity Dis 1:245–256

Chakraborty R, Weiss KM (1986) The frequencies of complex diseases in hybrid populations. Am J Phys Anthropol 70:489–503

Creasy MA (1956) Confidence limits for the gradient in the linear functional relationship. J R Stat Soc B 18:65–69

DeCroo S, Kamboh MI, Ferrell RE. Population genetics of alpha-1-antitrypsin polymorphism in U.S. Whites, U.S. Blacks and African Blacks. Hum Hered (in press)

Dischinger PC, Aposolides AY, Entsisle G, Hebel JR (1981) Hypertension incidence in an inner-city Black population. J Chronic Dis 34:405–413

Escallon MH, Ferrell RE, Kamboh MI (1987) Genetic studies of low-abundance human plasma proteins. V. Evidence for a second orosomucoid structural locus (ORM2) expressed in plasma. Am J Hum Genet 41:418–427

Glass B (1955) On the unlikelihood of significant admixture of genes from the North American Indians in the present composition of the Negroes of the United States. Am J Hum Genet 7:368–385

Glass B, Li CC (1953) The dynamics of racial admixture—an analysis based on the American Negro. Am J Hum Genet 5:1–19

Hanis CL, Chakraborty R, Ferrell RE, Schull WJ (1986) Individual admixture estimates: disease associations and risk of diabetes and gallbladder disease among Mexican Americans of Starr County, Texas. Am J Phys Anthropol 70:433–441

Harburg E, Gleibermann L, Roeper P, Schork MA, Schull WJ (1978) Skin color, ethnicity, and blood pressure. I. Detroit Blacks. Am J Public Health 68:1177–1183

Kamboh MI, Albers JJ, Majumder PP, Ferrell RE (1989*a*) Genetic studies of human lipoproteins. IX. Apolipoprotein D polymorphism and its relation to serum lipoprotein lipid levels. Am J Hum Genet 45:147–154

Kamboh MI, Ferrell RE (1986*a*) Ethnic variation in vitamin D-binding protein (GC): a review of isoelectric focusing studies in human populations. Hum Genet 72:281–293

——— (1986*b*) Genetic studies of low abundance human plasma proteins. II. Population of coagulation factor XIIIB. Am J Hum Genet 39:817–825

——— (1986*c*) A sensitive immunoblotting technique to identify thyroxin-binding globulin protein heterogeneity after isoelectric focusing. Biochem Genet 24:273–280

——— (1987*a*) Genetic studies of low-abundance human plasma proteins. VI. Polymorphism of hemopexin. Am J Hum Genet 41:645–653

——— (1987*b*) Human transferrin polymorphism. Hum Hered 37:65–81

——— (1989) Genetic studies of low abundance human plasma proteins. X. Coagulation factor XIIIB variants in Blacks. Electrophoresis 10:53–57

——— (1990) Genetic studies of low abundance human plasma proteins. XV. An overview of IEF-immunoblotting methods to screen apolipoprotein polymorphisms. Hum Hered 40:193–207

Kamboh MI, Lyons LA, Ferrell RE (1989*b*) Genetic studies of low abundance human plasma proteins. XIII. Population genetics of C1R complement subcomponent and description of new variants. Am J Hum Genet 44:148–153

Kaminer B, Lutz WPW (1960) Blood pressure in Bushman of the Kalahari desert. Circulation 22:289–295

Long JC (1991) The genetic structure of admixed populations. Genetics 127:417–428

Long JC, Smouse PE (1983) Intertribal gene flow between the Ye'cuana and Yanomama: genetic analysis of an admixed village. Am J Phys Anthropol 61:411–422

MacLean CJ, Adams MS, Leyshon WC, Workman PL, Teed TE, Gershowitz H, Weitkamp LR (1974) Genetic studies on hybrid populations. III. Blood pressure in an American Black community. Am J Hum Genet 26:614–626

Madansky A (1959) The fitting of straight lines when both variables are subject to error. J Am Stat Assoc 54:173–205

Neel JV (1973) "Private" genetic variants and the frequency of mutation among South American Indians. Proc Natl Acad Sci USA 70:3311–3315

Oviasu VO (1978) Arterial pressures and hypertension in a rural Nigerian community. Afr J Med Sci 7:137–143

Oyediran ABOO, Osuntokun BO, Akinkugbe OO, Carlisle RO, Bademosi O, Olatunde IA (1976) Community control of hypertensions—results in Epe of blood pressure and urinary survey and preliminary response to control measures. Nigerian Med J 6:248–253

Reed TE (1969) Caucasian genes in American Negroes. Science 165:762–768

Roberts DF (1955) The dynamics of racial admixture in the American Negro: some anthropological considerations. Am J Hum Genet 7:361–367

Snedecor GW, Cochran WG (1967) Statistical methods, 6th ed. Iowa University Press, Ames

Szathmary EJE, Reed TE (1978) Calculation of the maximum amount of gene admixture in a hybrid population. Am J Phys Anthropol 48:29–34

Weiss KM, Ferrell RE, Hanis CL (1984) A New World syndrome of metabolic diseases with a genetic and evolutionary basis. Yearbook Phys Anthropol 27:153–178

Williams RC, Steinberg AG, Knowler WC, Pettitt DJ (1986) *Gm3:5,13,14* and stated-admixture: independent estimates of admixture in American Indians. Am J Hum Genet 39:409–413