Differences in Lp[a] concentrations and apo[a] polymorphs between black and white Americans

Santica M. Marcovina,[1,*] John J. Albers,* Ellen Wijsman[†,§] ZuHui Zhang,[2,*] Nicola H. Chapman,[†] and Hal Kennedy*

Department of Medicine,* Northwest Lipid Research Laboratories, 2121 N. 35th Street, Seattle, WA 98103; Department of Biostatistics[†] and Division of Medical Genetics,[§] Department of Medicine, University of Washington, Seattle, WA 98195

Abstract Lp[a] concentrations in nmol/L and apo[a] size isoforms, expressed in terms of the relative number of apo[a] kringle 4 (K4) repeats, were determined in 3959 whites and blacks from four U.S. communities. Plasma Lp[a] analyses were performed by an ELISA method insensitive to apo[a] size heterogeneity and apo[a] size isoforms were determined by high resolution agarose gel electrophoresis. Allele frequencies were estimated by maximum likelihood methods in order to account for the presence of null alleles and coalescence of bands on gels. The apo[a] allele frequencies and phenotype distributions differed significantly between blacks and whites ($P < 0.0001$). Blacks had a higher relative frequency of the intermediate alleles $K4_{22}$ through $K4_{28}$ whereas whites had a higher relative frequency of the small alleles $K4_{17}$ through $K4_{24}$ and large alleles $K4_{29}$ through $K4_{33}$. The estimated frequency of the null allele was low in both blacks (1.0%) and whites (6.7%). The Lp[a] distribution was less skewed and Lp[a] concentrations were higher in blacks than whites (mean 94 nmol/L and 48 nmol/L, median 74 nmol/L and 20 nmol/L for blacks and whites, respectively). The relationship between apo[a] size and Lp[a] concentration also differed significantly between these two racial groups. For the large polymorphs (>31 K4 repeats) both blacks and whites exhibited uniformly low Lp[a] values. For the intermediate isoforms $K4_{20}$ through $K4_{30}$, a considerable range of Lp[a] values was evident in blacks; the median Lp[a] for each isoform increased nearly linearly as the apo[a] size decreased. In contrast, in whites there was little change in median Lp[a] concentrations for isoforms $K4_{20}$ through $K4_{30}$. For the small apo[a] size (<20 K4) both blacks and whites exhibited high median Lp[a] levels and a wide variation of Lp[a] levels. The major difference in Lp[a] levels between the two racial groups occurred in the intermediate size isoform range of $K4_{20}$ through $K4_{25}$. In conclusion, whites and blacks differ significantly in Lp[a] concentrations, allele and phenotype frequencies, and in the relationship between apo[a] size isoform and Lp[a] concentration.—**Marcovina, S.M., J.J. Albers, E. Wijsman, Z-H. Zhang, N.H. Chapman, and H. Kennedy.** Difference in Lp[a] concentrations and apo[a] polymorphs between black and white Americans. *J. Lipid Res.* 1996. **37:** 2569–2585.

Supplementary key words lipoprotein[a] • apolipoprotein[a] • ELISA • apo[a] allele • apo[a] phenotype • gel electrophoresis • immunoblotting • racial differences

Lipoprotein[a] (Lp[a]) is a specific class of lipoproteins composed of an LDL particle and a carbohydrate-rich, highly polymorphic glycoprotein, apolipoprotein[a] (apo[a]) (1), disulfide linked to the apoB-100 component of LDL. Early studies in Caucasians indicate that apo[a] is highly variable in size and that plasma concentrations are genetically determined and are inversely related to the apo[a] size (2). The apo[a] size variation is due to the length polymorphism in the apo[a] gene (3). The apo[a] gene encodes three distinct structural domains, each exhibiting a high degree of structural homology with the related domain of plasminogen. Apo[a] is composed by multiple copies of the kringle 4 domain, one copy of kringle 5, and a protease domain (4). Apo[a] kringle 4 sequences can be divided into 10 distinct types based on variation in their amino acid composition. All of the kringle 4 (K4) types, except K4 type 2, appear to be present in one copy per apo[a] molecule (5, 6). K4 type 2 is present in a variable number of repeats and is therefore responsible for the isoform size heterogeneity of apo[a]. Apo[a] isoform can be characterized at the DNA level by pulsed field gel electrophoresis and genomic blotting of an apo[a] DNA restriction fragment containing a variable number of tandem repeats of the K4 type 2 encoding sequence (3) or at the protein level by agarose gel electrophoresis and immunoblotting of plasma apo[a] (7). Elevated plasma Lp[a] levels have been reported to be associated with coronary atherosclerosis in whites (for a review see

Abbreviations: Lp[a], lipoprotein[a]; apo[a], apolipoprotein[a], LDL, low density lipoprotein; ELISA, enzyme-linked immunosorbent assay; CARDIA, coronary artery risk development in young adults study; CHD, coronary heart disease; EDTA, ethylene diamine tetraethyl amine; EM algorithm, expectation maximization algorithm; K4, kringle 4 repeats; HWE, Hardy-Weinberg equilibrium.

[1]To whom correspondence should be addressed.

[2]Present address: West China University of Medical Sciences, Chengdu, Sichuan, P.R. China.

ref. 8) but not in blacks (9, 10). Both Lp[a] concentrations and isoform size distribution have been shown to differ significantly between blacks and whites, with blacks having mean Lp[a] levels approximately two times higher than whites (11–16).

We previously reported the Lp[a] concentrations in over 4000 whites and blacks from four U.S. communities (15) and apo[a] phenotypes were determined on a subset (690) of those participants. However, Lp[a] protein concentrations were measured by an ELISA assay that is sensitive to apo[a] size heterogeneity and apo[a] phenotypes were determined by SDS-polyacrylamide gel electrophoresis and immunoblotting. This apo[a] phenotype method resolved only 7 apo[a] isoforms on the basis of electrophoretic mobility rather than the 34 or more isoforms that are known to occur in plasma (5, 7). Although the frequency distribution of the apo[a] isoforms differed between blacks and whites, this did not explain the difference in Lp[a] levels observed between these two racial groups. In order to further investigate the relationship between Lp[a] concentrations and apo[a] size, we have determined Lp[a] concentrations and apo[a] isoform sizes in 3959 biracial participants from four U.S. communities, using state-of-the-art methodologies.

METHODS

Study population

CARDIA is a multicenter, longitudinal epidemiological study of cardiovascular disease risk factors in young adults aged 18 to 30 years at the initial examination. For details of the recruitment and study design, see Friedman et al. (17) and Cutter et al. (18). Four clinical centers located in Birmingham, AL, Chicago, IL, Minneapolis, MN, and Oakland, CA recruited a total of 5115 participants for an initial examination conducted between April 1985 and June 1986. The study population was approximately balanced within each center for age (18–24 years, 25–30 years), gender, race (blacks, whites), and level of education (high school or less, beyond high school). Three follow-up examinations have been conducted and the fourth is in progress. This report is limited to data collected during the fourth examination (1992 to 1993) on 3959 subjects. Informed consent was obtained from all participants. Examinations were performed in the morning after a 12-h fast. Venous blood was drawn from seated subjects into Vacutainer tubes containing EDTA, using minimal stasis; the cells were separated from plasma and the plasma was stored at −70°C prior to shipment on dry ice to the Northwest Lipid Research Laboratories in Seattle. Of the 4086 CARDIA participants eligible for the fourth examination, adequate plasma was available for Lp[a] quantification and apo[a] phenotype determination on 3959 participants.

Lp[a] quantification

Lp[a] concentration was measured by a direct binding double monoclonal antibody-based enzyme-linked immunoassay (ELISA) (19). The capture monoclonal antibody (a-6) is directed to an epitope present in apo[a] K4 type 2 and the detection antibody (a-40) is directed to an epitope present in apo[a] K4 type 9. Because K4 type 9 is present in only one copy per apo[a] molecule (5, 6), this immunoassay is insensitive to apo[a] isoform size heterogeneity. Lp[a] concentrations are expressed in nmol/L. Fresh-frozen sera from five individuals representing a broad range of Lp[a] concentrations were used as quality controls for the ELISA. Detailed evaluations of the assay, including the analytical performance have been reported (19).

Apo[a] phenotyping

Apo[a] isoforms were characterized using a high-resolution SDS-agarose gel electrophoresis method followed by immunoblotting, performed as previously reported (7). It should be noted that to enhance the isoform detection, a variable volume of plasma, depending on apo[a] concentration, is applied to the gel. The sensitivity of our immunoblot procedure has been determined to be approximately 5 fmol of apo[a]. We have recently demonstrated a consistent relationship (20) between the number of K4 repeats as determined by pulsed-field gel electrophoresis (3) and the relative migration of apo[a] in agarose gel. Therefore, we have replaced the arbitrary numerical identification of the apo[a] isoforms previously used (7) with a size designation based on the respective number of K4 repeats (20).

Statistical methods

Allele frequencies were estimated by maximum likelihood methods using an EM algorithm (21, 22) in order to account for the presence of null alleles and coalescence of bands on gels. Null alleles are those that cannot be detected by the phenotyping method, and are known to exist because a few individuals produce no bands in the gel electrophoresis. Such alleles may represent either absence of protein product (i.e., true null), or presence of the protein in a concentration that is below the sensitivity limit of the method (i.e., apparent null). Coalescence is defined as the inability to distinguish between two alleles because they migrate too close together in the gel, thus appearing as a single band. Therefore, coalescence of bands causes some heterozygotes to be erroneously classified as homozygotes. The issue of coalescence and its effect on allele frequency

estimation and associated hypothesis testing has been previously discussed in the context of DNA fingerprinting (23). Details regarding implementation of the EM algorithm for estimation of allele frequencies are given in the Appendix. In all cases, null allele frequencies were accounted for. Allele frequencies were estimated using three models so that the validity of specific assumptions in the models could be investigated. Using Model I, coalescence probabilities were assumed to be zero (i.e., the possibility of coalescence was ignored). Using Model II, estimates of both allele frequencies and coalescence probabilities were obtained by using the data from only one racial group at a time. Using Model III, allele frequencies were estimated from the data within each of the two racial groups under study, but coalescence probabilities were estimated from the combined data on both racial groups, thus using somewhat more information to estimate coalescence probabilities.

Hardy-Weinberg equilibrium (HWE) tests were performed in both populations for all three models. A test of the null hypothesis that a given population is in HWE was obtained by constructing a likelihood ratio statistic. The likelihood ratio statistic is approximately distributed as χ^2 with degrees of freedom equal to the difference between number of parameters estimated under the two models being compared. Details are given in the Appendix.

A likelihood ratio statistic was also used to determine whether or not the allele frequencies were significantly different between the two populations. Because the results of the tests for HWE showed that it was important to allow for coalescence, the frequencies used in comparing allele frequencies in the two populations were based on a model that allows for coalescence. For consistency, the coalescence probabilities were forced to be the same for each of the three sets of estimates (the frequencies in each of the two populations and the combined data set), and were taken to be equal to the coalescence estimates based on data from both populations (Model III). Details are given in the Appendix.

A second test to compare the phenotype distributions in the two populations was also performed. This test involved a contingency table test of equality of phenotype frequencies. An estimate of the *P*-value for a Fisher exact test was obtained with Monte Carlo Markov chain methods for sparse contingency tables (24). This approach of comparing the phenotype distribution on two populations has lower power than the likelihood ratio test of comparison of allele frequencies so that failure to find a statistically significant result does not imply that the allele frequencies are the same in the two populations. However, this approach is not sensitive to the model used for allele frequency estimation. Phenotype frequency differences could reflect allele frequency differences between the two populations. Unlike the likelihood ratio statistic for comparing allele frequencies, comparison of phenotype frequencies between the two populations did not require developing a model under assumptions of HWE to estimate allele frequencies. This approach could not be used in its current form for comparison of allele frequencies in the two populations because of the ambiguities introduced by null alleles and band coalescence.

Standard statistical procedures including χ^2 test, linear regression analysis, estimates of skewness and kurtosis, were performed using Complete Statistical System Software (CSS, Statistica, Tulsa, OK). For comparison of groups for differences in Lp[a] levels, the Mann-Whitney test was used. Tests of differences in slopes and intercepts were performed as described by Kleinbaum, Kupper, and Muller (25).

TABLE 1. Lp[a] concentrations in blacks and whites

Study Group	n	Percentiles						
		5	10	25	50	75	90	95
		nmol/L						
Blacks								
Men	807	5.6	13.1	32.5	70.0	129.4	193.8	228.2
Women	1,092	8.1	18.1	38.8	76.9	130.0	203.2	248.2
Total	1,899	6.9	15.6	36.3	74.4	130.0	199.4	234.4
Whites								
Men	985	1.3	2.5	7.5	19.4	63.8	140.0	175.1
Women	1,075	1.3	2.5	8.1	21.9	67.5	137.5	180.1
Total	2,060	1.3	2.5	7.5	20.0	66.6	138.2	176.9

RESULTS

Lp[a] concentrations in blacks and whites

Blacks have consistently higher Lp[a] concentrations than whites (Mann-Whitney U test, $P < 0.001$) with the largest relative difference occurring at the lower percentiles, but the largest absolute difference occurring at the higher percentiles **(Table 1).** For example, the 10th percentile value for blacks was more than 6 times the 10th percentile for whites whereas the median value for blacks was 3.7 times the median value for whites. In whites, men and women had very similar Lp[a] percentile values while in blacks, women had higher levels of Lp[a] than men (Mann-Whitney U Test, $P < 0.05$). No significant difference was found in Lp[a] concentrations at each of the four clinical sites **(Table 2).** The plasma Lp[a] concentration varied widely in both racial groups. The distribution of Lp[a] in whites was highly skewed, with most Lp[a] levels in the lower range of values (skewness 2.02, kurtosis 4.53) with a median of

TABLE 2. Plasma Lp[a] concentrations in black and white Americans

Site, Sex	Blacks n	Blacks Median	Blacks Mean	Whites n	Whites Median	Whites Mean
		nmol/L			*nmol/L*	
Birmingham, AL						
Men	203	68.1	83.7	227	20.0	48.4
Women	265	80.0	95.9	192	20.0	51.4
Total	468	76.0	90.6	419	20.0	49.8
Chicago, IL						
Men	160	64.7	85.1	228	20.3	45.6
Women	238	79.1	93.7	232	26.9	52.1
Total	398	75.0	90.3	460	23.4	48.9
Minneapolis, MN						
Men	195	77.5	95.4	321	18.8	46.4
Women	197	68.1	99.7	351	20.0	46.1
Total	392	74.4	97.6	672	18.8	46.2
Oakland, CA						
Men	249	68.8	93.3	209	19.4	47.0
Women	392	76.9	96.3	300	19.1	48.5
Total	641	73.8	95.1	509	19.4	47.9
All sites						
Men	807	70.0	89.7	985	19.4	46.8
Women	1,092	76.9	96.3	1,075	21.9	49.0
Total	1,899	74.4	93.5	2,060	20.0	47.9

20.0 nmol/L and a mean of 47.9 nmol/L **(Fig. 1).** The distribution of Lp[a] in blacks with a median of 74.4 nmol/L and a mean of 93.5 nmol/L, was less skewed toward lower Lp[a] levels than whites (skewness 1.73, kurtosis 4.73). The overall plasma Lp[a] distribution in whites was significantly different from that in blacks (Mann-Whitney U Test, $P < 0.001$). In blacks, the frequency of Lp[a] was nearly uniform up to 100 nmol/L, whereas in whites there was a dramatic decrease in frequency of Lp[a] in the concentration range of 10 to 50 nmol/L (Fig. 1).

The frequency of subjects with two apo[a] isoforms was lower in whites (75.7%) as compared to blacks (85.2%) (**Table 3,** $\chi^2_1 = 52.4$, $P < 0.0001$). One black and 8 white participants had no detectable apo[a] isoform by immunoblotting. All individuals with undetectable apo[a] by immunoblotting had Lp[a] concentration less than 0.5 nmol/L. The distribution of apo[a] phenotypes in blacks and whites is shown in **Table 4** and **Table 5,** respectively. Blacks exhibited 294 phenotypes and whites exhibited 307 phenotypes out of the possible 595, based on 34 different isoforms. Estimates of allele frequencies and coalescence probabilities using the three models are given in **Table 6** for both populations. The likelihood ratio test statistic for comparing allele frequencies between populations is $\chi^2_{33} = 326.801$, $P < 0.0001$. A similar conclusion is obtained from comparison of phenotype distributions in the two populations: the Monte Carlo Markov Chain estimate of the *P*-value for the Fisher exact test indicated that there was a significant difference in the phenotype distributions,

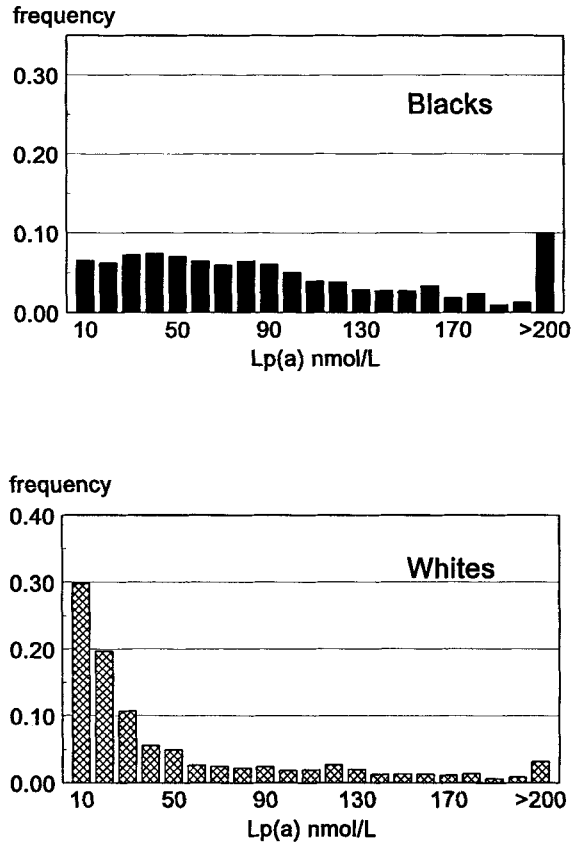


Fig. 1. Frequency distribution of plasma Lp[a] concentrations in blacks and whites. Plasma Lp[a] concentrations were measured by a direct binding ELISA as described in Methods. The width of each bar corresponds to a plasma Lp[a] concentration interval of 10 nmol/L.

and thus presumably also in the allele frequency distributions, in the two populations ($P < 0.0001$). The allele frequency distribution in blacks appeared nearly Gaussian **(Fig. 2),** while it had a bimodal appearance in whites. The allele frequency distributions in both racial groups do not change when the analysis includes only those individuals in which both alleles could be determined (data not shown). Blacks had a higher frequency of the intermediate isoforms ($K4_{22}$ to $K4_{28}$) whereas whites had a higher frequency of the small isoforms ($K4_{17}$ to $K4_{20}$) and large isoforms ($K4_{29}$ to $K4_{33}$) as evident from the frequency difference plot (Fig. 2).

TABLE 3. Frequency of single and double band phenotypes

	Blacks n	Blacks %	Whites n	Whites %
Apo[a] undetectable	1	0.05	8	0.4
Single band	281	14.8	493	23.9
Double band	1,617	85.2	1,559	75.7
Total	1,899		2,060	

TABLE 4. Distribution of apo[a] phenotypes in blacks (n = 1,898)

	Band 1																																	
Band 2	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43
10	1																																	
11																																		
12																																		
13																																		
14				1	1																													
15	1			1	3	2																												
16	3		1		1	2																												
17		1				4	1	1																										
18		2	2	3	1	2	4	1	9																									
19			1		2		2	1	1	5																								
20		1				2	2	2	6	1	11																							
21			1	1	3	6	9	5	11	8	14	11																						
22	1	1		1		5	5	4	11	13	11	4	13																					
23		1	1	2	3	7	9	7	5	8	13	23	3	19																				
24						1	3	5	11	7	18	21	25	4	25																				
25						5	4	6	9	16	21	17	32	19	3	32																			
26	2			1	1	2	4	3	6	9	15	13	21	36	19	3	36																	
27					2	3	6	6	10	9	17	18	20	24	33	27	1	36																
28		1			1	3	4	6	6	5	16	19	19	36	37	28	16	3	27															
29						1	1	5	2	3	8	14	17	18	19	28	26	13	3	22														
30		1				2	1	6	2	3	6	9	11	12	11	17	25	16	9	1	14													
31						3	2	2	5	2	3	7	3	1	9	12	8	12	14	3		4												
32					1	1	3	4		2	1	2	3	1	5	6	8	8	5	6	3		3											
33				1			2	1			3	2	5	1	6	4	3	9	2	5	2			1										
34					1			1	2	3	1	4	1	2	6	2	9	1	6	4	3	2			2									
35	1						1				1	1		1	4	5		1	3	3	1	2	2			1								
36										2	1	2	2	2	1	6	1	2	4	2	2	4		2			3							
37							1					1	1		1	2	1		1	1		1												
38								1		1		1	3	1	1	1	2	1	2	4	1				3		1		1					
39										1										1						1								
40																		1		1					1									
41																																		
42																																		
43																1																		1

Each number reflects the number of individuals with a given apo[a] phenotype. The columns under Band 1 refer to the number of kringle 4 repeats of the smaller apo[a] polymorph.

The results of the likelihood ratio tests for Hardy-Weinberg equilibrium (HWE) are given in **Table 7,** and indicate that there is no evidence for deviation from HWE when coalescence of bands is accounted for (Models II and III). However, when coalescence is not incorporated into the model both the black and the white populations give highly significant evidence that allele frequencies deviate from the expected Hardy-Weinberg frequencies. In addition, for both populations, likelihood ratio tests comparing fit of Model II with coalescence versus Model I without coalescence strongly suggested that the models that included coalescence fit the observed data significantly better than models that did not estimate coalescence probabilities ($\chi^2_4 = 377.6$, $P < 0.0001$ for the black population, and $\chi^2_4 = 238.6$, $P < 0.0001$ for the white population). The coalescence probabilities indicate that there is a high probability (84%) of coalescence of two isoforms on a gel if they are within 2 mm of each other, with successively smaller probabilities of coalescence for larger migration distances (see Table 6).

The relationship between apo[a] size and plasma Lp[a] concentration was first examined in blacks and whites who exhibited a single apo[a] isoform **(Fig. 3).** A considerable range of values was evident within each isoform but the values were clustered towards low Lp[a] values for the large isoforms, particularly in whites, whereas blacks had a broader range of Lp[a] values than whites, from negligible to greater than 200 nmol/L, for the intermediate apo[a] sizes ($K4_{21}$ to $K4_{27}$). Both blacks and whites had a broad range of Lp[a] values for the small isoforms with less than 21 K4 repeats (Fig. 3). Only high Lp[a] values were observed for apo[a] with less than 15 K4 repeats.

The change in median Lp[a] as a function of apo[a] size differed greatly between the two racial groups **(Fig. 4).** Among the individuals exhibiting a single apo[a] isoform, whites demonstrated little change in median Lp[a] for isoforms $K4_{23}$ to $K4_{30}$ while blacks demonstrated a nearly linear decrease in median Lp[a]. (Fig. 4, top). Rank order correlations were calculated between adjacent K4 repeats ($K4_{18}$ to $K4_{22}$) for Lp[a] val-

TABLE 5. Distribution of apo[a] phenotypes in whites (n = 2,052)

	Band 1																																
Band 2	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43
10	2																																	
11																																		
12		2	1																															
13			1																															
14			2	3	3																													
15				1	3	5																												
16					2	4	7																											
17		2			2	2	5	9																										
18	1	1	1	3	4	2	8	9	24																									
19			2	6	1	1	1	10	10	23																								
20	1			1	5	8	2	5	23	5	31																							
21				1	3	2	3	7	13	15	6	36																						
22		1	1	3		2	4	3	8	8	18	1	18																					
23					1	4	4	4	6	7	10	6	1	20																				
24		1	1	1		1	5	8	9	5	14	14	8	3	20																		
25			1	1	2	1	2	8	10	13	12	18	15	5	2	28																	
26						3	3	8	13	9	19	18	7	6	14	1	38																	
27				1		1	3	5	8	11	21	22	10	10	13	11	6	38															
28	1						3	7	6	7	10	14	11	15	13	27	13	3	39															
29				1	1	1	2	6	8	13	11	16	19	19	13	16	25	15	2	46														
30				1	1	1		3	10	8	13	8	7	12	15	16	17	15	16	3	27													
31					1			4	7	6	7	8	5	10	13	11	8	15	18	5	2	21												
32				1		2	1	1	3	4	7	5	3	5	4	7	11	12	9	19	9	2	25											
33						1		1	2	2	2	2	4	5	1	8	4	8	8	3	10	2	3	13										
34									1		3	4	2	2		5	1	4	5	4	9	6	4		9									
35							1		1				1		1	2	2	1	4	1	5	3	2			3								
36								1			1	5	3		2	1		1	4	4	1	1	3	2	3	1	1							
37										1	1				1	1	1	4	1	5				1				1						
38						1			1	1	1	2	1			3		2		5	2	2	4	2	1	1			3					
39												1						1						1										
40												1			1	1											1				1			
41																								1								1		
42																																		
43																																		

Each number reflects the number of individuals with a given apo[a] phenotype. The columns under Band 1 refer to the number of kringle 4 repeats of the smaller apo[a] polymorph.

ues for blacks and whites. Lp[a] was significantly correlated with adjacent K4 repeats between $K4_{18}$ to $K4_{21}$ for whites but not for blacks. Among the individuals heterozygous for apo[a] isoforms, whites demonstrated only a slight change in Lp[a] from $K4_{23}$ to $K4_{33}$ (Fig. 4, bottom) while blacks exhibited a general decrease in Lp[a] for this same size range. An abrupt change in Lp[a] median value is exhibited among whites between $K4_{19}$ and $K4_{20}$. For the small apo[a] size ($K4_{13}$ to $K4_{19}$) there was no consistent change in median Lp[a] values in either racial group with blacks having only slightly higher Lp[a] levels than whites overall. Furthermore, for the large isoforms (K4 > 31) blacks and whites exhibited little difference in Lp[a] levels. Therefore, the major difference in median Lp[a] levels between blacks and whites occurred in the intermediate range ($K4_{20}$ to $K4_{25}$), ranging from 122 nmol/L for $K4_{20}$ to 44 nmol/L for $K4_{25}$ (Fig. 4, bottom).

Because in heterozygous subjects the second apo[a] isoform also contributes to the plasma Lp[a] concentration, we evaluated the relationship of plasma Lp[a] level to the total number of K4 repeats in all heterozygous individuals **(Fig. 5).** As we noted previously, blacks have a higher variance in plasma Lp[a] levels than whites. The relationship between the sum of the kringle 4 repeats and Lp[a] concentration appeared biphasic for whites in that median plasma Lp[a] concentrations were uniformly low for the group of individuals with ≥50 K4 repeats (Fig 5, bottom). For the most part, there was a direct relationship between the increase in Lp[a] concentration and the decrease in the sum of K4 repeats. Spearman's rank correlation between the sum of the kringle 4 repeats and plasma Lp[a] values was −0.487 for blacks and −0.461 for whites, both significantly different from zero ($P < 0.001$). Linear regression analysis between the sum of the kringle 4 repeats and the plasma Lp[a] concentration indicated that the slope of the line for blacks (−5.07) is significantly different from the slope for whites (−3.64) $P < 0.001$. The intercept for the regression line for blacks (348.8 vs.

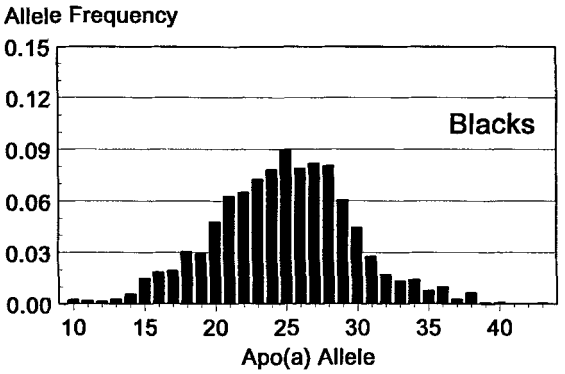
TABLE 6. Estimated allele frequencies and coalescence probabilities

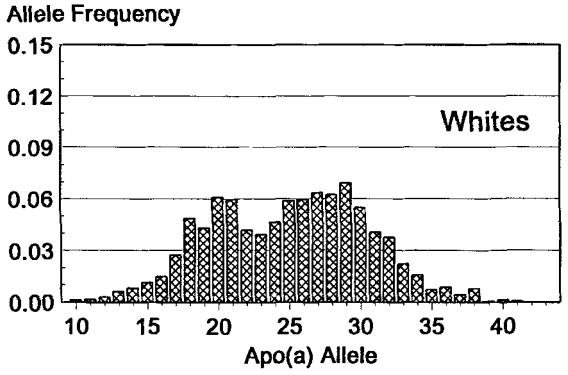
Allele	White Population			Black Population		
	Model I	Model II	Model III	Model I	Model II	Model III
0	0.1007	0.0728	0.0670	0.0443	0.0082	0.0105
10	0.0012	0.0012	0.0012	0.0024	0.0024	0.0024
11	0.0017	0.0017	0.0017	0.0021	0.0021	0.0021
12	0.0029	0.0029	0.0029	0.0016	0.0016	0.0016
13	0.0061	0.0061	0.0061	0.0029	0.0029	0.0029
14	0.0083	0.0083	0.0083	0.0055	0.0056	0.0056
15	0.0112	0.0113	0.0113	0.0148	0.0150	0.0150
16	0.0147	0.0147	0.0148	0.0187	0.0188	0.0187
17	0.0270	0.0273	0.0275	0.0193	0.0197	0.0197
18	0.0477	0.0486	0.0489	0.0298	0.0305	0.0304
19	0.0420	0.0432	0.0435	0.0280	0.0291	0.0290
20	0.0592	0.0610	0.0614	0.0468	0.0479	0.0479
21	0.0580	0.0595	0.0598	0.0609	0.0628	0.0626
22	0.0405	0.0417	0.0420	0.0633	0.0652	0.0650
23	0.0384	0.0392	0.0394	0.0704	0.0727	0.0726
24	0.0453	0.0464	0.0467	0.0754	0.0783	0.0781
25	0.0568	0.0589	0.0595	0.0847	0.0897	0.0892
26	0.0571	0.0594	0.0595	0.0757	0.0788	0.0790
27	0.0606	0.0635	0.0641	0.0779	0.0817	0.0815
28	0.0594	0.0624	0.0629	0.0760	0.0805	0.0802
29	0.0666	0.0692	0.0697	0.0578	0.0605	0.0604
30	0.0526	0.0550	0.0554	0.0428	0.0445	0.0444
31	0.0389	0.0405	0.0407	0.0263	0.0275	0.0274
32	0.0364	0.0375	0.0377	0.0170	0.0173	0.0173
33	0.0212	0.0220	0.0221	0.0129	0.0133	0.0133
34	0.0154	0.0157	0.0158	0.0143	0.0145	0.0145
35	0.0071	0.0072	0.0072	0.0074	0.0077	0.0076
36	0.0085	0.0086	0.0087	0.0098	0.0100	0.0010
37	0.0041	0.0041	0.0041	0.0026	0.0028	0.0027
38	0.0078	0.0078	0.0078	0.0063	0.0064	0.0064
39	0.0007	0.0007	0.0007	0.0008	0.0008	0.0008
40	0.0012	0.0012	0.0012	0.0008	0.0008	0.0008
41	0.0005	0.0005	0.0005	0.0000	0.0000	0.0000
43	0.0000	0.0000	0.0000	0.0005	0.0005	0.0005
Pr(Coal|$d < 2$)	NA	0.7723	0.8426	NA	0.9035	0.8426
Pr(Coal|$d = 2$–3)	NA	0.4825	0.6071	NA	0.6967	0.6071
Pr(Coal|$d = 3$–4)	NA	0.0216	0.1232	NA	0.2101	0.1232
Pr(Coal|$d = 4$–5)	NA	0.0000	0.0000	NA	0.0000	0.0000

NA, not applicable, assumed to be 0.0.

233.7) was also higher than for whites. As reflected by the steeper slope and the higher intercept of the regression line, for blacks, the median Lp[a] concentration was inversely related to the kringle 4 repeats number and this relationship was nearly linear. Blacks consistently had higher Lp[a] levels than whites at each K4 repeat sum except at large kringle number sum (>61 kringle repeats) constituted by individuals with two large apo[a] isoforms and at very small kringle number sum (<33 kringle repeats) constituted by individuals with two small apo[a] isoforms. Examination of Lp[a] concentration distributions by the predominantly expressed apo[a] isoform revealed both isoform-specific and race-specific differences in Lp[a] distribution **(Fig. 6).** If the predominant isoform was relatively small in size, such as for $K4_{18}$ or $K4_{19}$, the Lp[a] distribution was unimodal in both whites and blacks with little racial differences. In contrast, for the intermediate sizes, $K4_{20}$ through $K4_{25}$ there was an evident bimodal distribution in whites but not in blacks. A large proportion of whites but very few blacks exhibiting intermediate apo[a] size had a low Lp[a] concentration (Fig. 6). On the other hand, when the predominant isoform was large, such as for $K4_{27}$ and $K4_{28}$, the Lp[a] distribution was unimodal in both whites and blacks but the distribution in whites was shifted to lower Lp[a] concentrations.

Potential differences in the secondary isoform do not appear to contribute to the bimodal Lp[a] distribution in whites with predominant isoform of intermediate size as the distribution of the secondary apo[a] isoform was similar among whites expressing high or low levels of Lp[a]. These isoform- and racial-specific differences in Lp[a] distributions were also observed in blacks and whites exhibiting a single apo[a] isoform (data not shown). Whites and blacks with either small or relatively large apo[a] size isoforms had a unimodal Lp[a] distri-





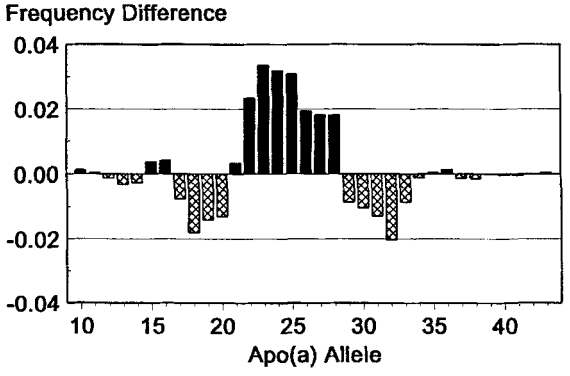


Fig. 2. Apo[a] allele frequencies in blacks and whites. Estimates of allele frequency take into account for the presence of null alleles and coalescence of bands on gels as described under Model II (see Statistical Methods). The difference in apo[a] allele frequency (blacks minus whites) is shown in the bottom figure.

bution whereas in the range of intermediate apo[a] size isoforms, Lp[a] in whites, but not in blacks, had a bimodal distribution. If we eliminate the whites in the low Lp[a] modal group for $K4_{20}$ to $K4_{26}$ then the resultant slope of median values more closely corresponds to blacks, although Lp[a] values are approximately 25% lower. Also, for this select range of apo[a] sizes, Lp[a] values for blacks and whites remain significantly different (Mann-Whitney Test, $P < 0.0001$).

TABLE 7. Likelihood ratio tests of Hardy-Weinberg equilibrium

Population	Model	χ^2	df	P
Black	I	770.556	561	<0.0001
Black	II	392.946	557	1.0
Black	III	434.538	557	1.0
White	I	750.867	561	<0.0001
White	II	512.246	557	0.913
White	III	460.381	557	0.999

The relationship between plasma Lp[a] and apo[a] phenotype in heterozygotes was examined separately in the two racial groups (**Fig. 7**). A preliminary regression model was constructed which predicted Lp[a] levels as a function of the number of repeats in the two K4 polymorphs in the apo[a] phenotype. Because of concerns that one allele may contribute more to Lp[a] levels than the other when one band in a heterozygote exhibits

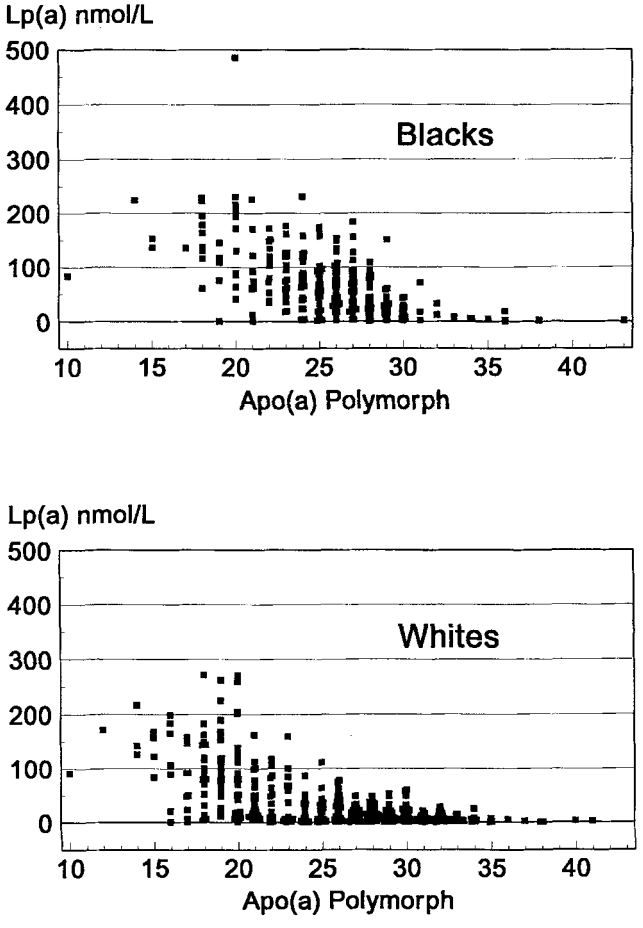


Fig. 3. Relationship between Lp[a] concentration and apo[a] size in blacks and whites who exhibited a single apo[a] polymorph.

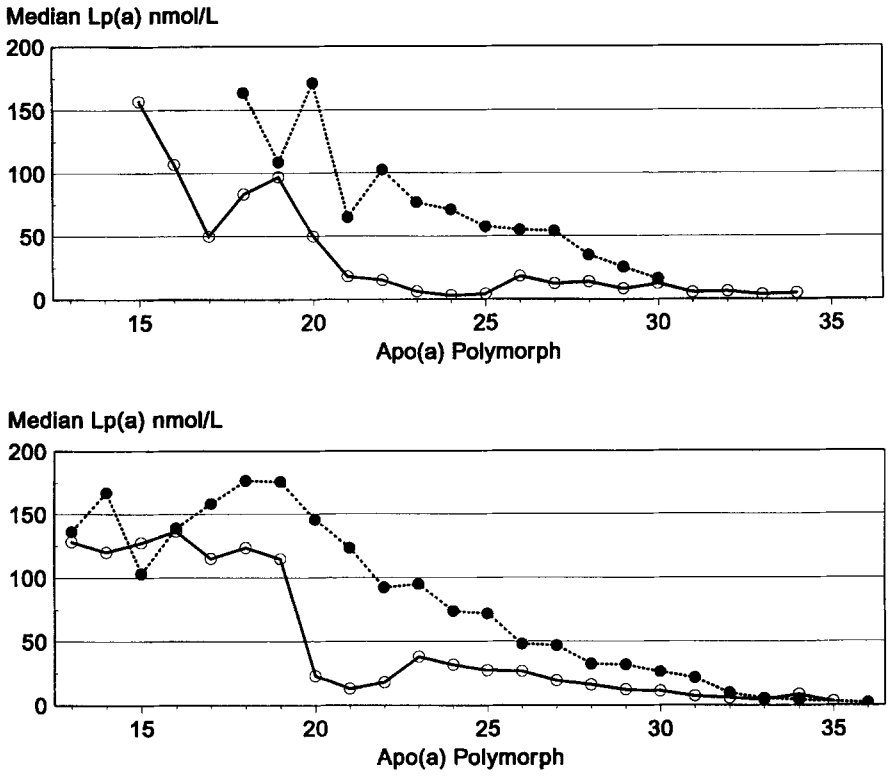


Fig. 4. Relationship between median Lp[a] concentration and apo[a] size in blacks (closed circles) and whites (open circles) who exhibited a single apo[a] polymorph (top figure) or predominantly expressed apo[a] polymorph (bottom figure). Those with a predominantly expressed apo[a] polymorph included 1366 blacks and 1370 whites exhibiting two polymorphs with one polymorph predominant in concentration from visual inspection of the immunoblot. Median Lp[a] concentrations are shown for each polymorph group having at least five individuals.

higher intensity on a gel than does the other, only the 1370 whites and 1366 blacks were included for whom one band could be visually identified as the major band in the genotype. An analysis which could take into account the relative amounts of the gene products derived from each of the two alleles is not yet possible as it awaits quantitation of the relative proportions of the two allele products. The model constructed was:

$$y_i = C + \alpha(k_{1i} + k_{2i}) + \beta k_{1i} + \gamma k_{2i} + \delta k_{1i}k_{2i} + \varepsilon_i$$

where y_i is the Lp[a] for individual i, K_{1i} is the number of K4 repeats for major band for individual i, K_{2i} is the number of K4 repeats for the minor band for individual i, and ε_i is a normally distributed error term with 0. For the whites, all terms in this model were significant at $P < 0.005$. For the blacks, all terms were significant at $P < 0.005$ except the quadratic term, $\gamma(p = 0.389)$. Thus there is evidence that the two populations have different models which best explain the observed Lp[a] levels. It should be noted that in both populations both α and β are significantly different from 0, indicating that both alleles contribute to Lp[a] levels, but that the predominantly expressed allele contributes more to the Lp[a] level than does the secondary allele. The resulting model for the whites was:

$$y_i = 359.19 - 5.71(k_{1i} + k_{2i}) - 6.80k_{1i} + 1.08k_{2i} + 0.18k_{1i}k_{2i} + \varepsilon_i$$

After eliminating the nonsignificant quadratic term, the resulting model for blacks was:

$$y_i = 390.56 - 4.10(k_{1i} + k_{2i}) - 6.31k_{1i} + 2.21k_{2i} + \varepsilon_i$$

Blacks and whites had significantly different patterns, reflecting differences in apo[a] phenotype frequencies and differences in Lp[a] levels for a given phenotype. For example, whites had a greater frequency of phenotypes in which both isoforms were large and the median

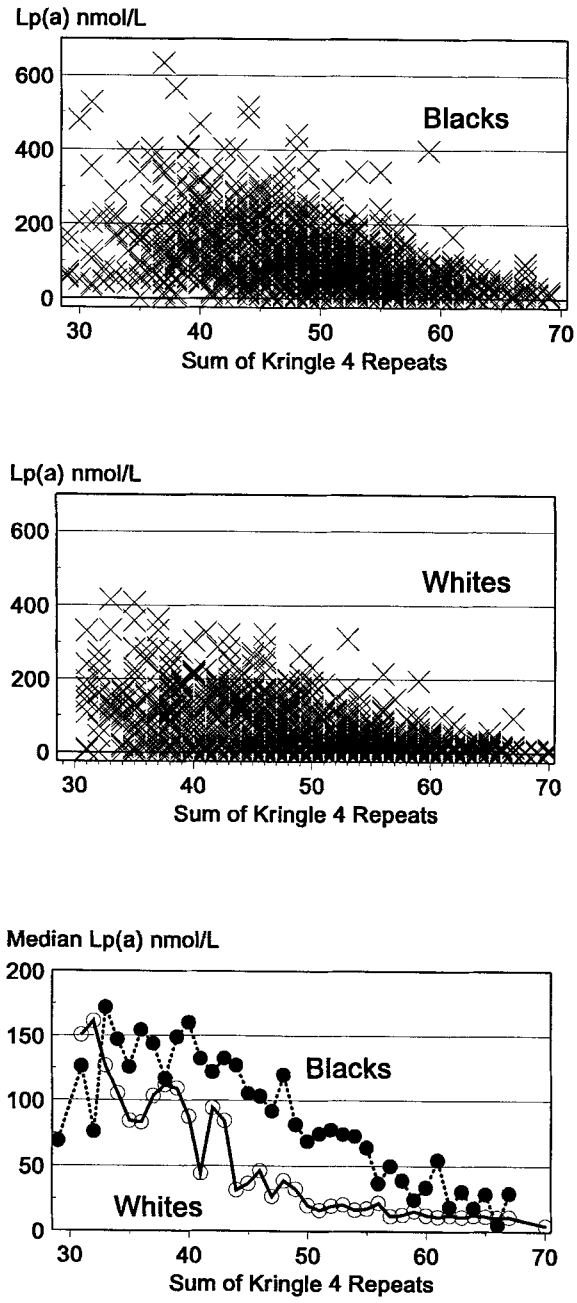


Fig. 5. Relationship between the sum of the kringle 4 repeats in the apo[a] polymorphs and plasma Lp[a] concentrations in blacks (n = 1617) and whites (n = 1559) exhibiting two apo[a] polymorphs. The sum of the kringle 4 repeats of the two apo[a] polymorphs is plotted against the plasma Lp[a] concentrations. Median Lp[a] concentrations are shown for each kringle 4 repeat sum group having at least five individuals.

Lp[a] value was consistently low for these phenotypes. For blacks, there was clearly an inverse relationship between apo[a] size and the median Lp[a] concentration throughout most of apo[a] size range. In contrast, whites with one or both isoforms with K4 > 20, had low median Lp[a] concentrations while whites with one or both isoforms with K4 < 20 had high median Lp[a] levels.

DISCUSSION

We previously reported Lp[a] protein concentrations in the biracial sample from the CARDIA study (15). However, the ELISA method used in this earlier study differed significantly from the present study because the detection antibody recognized the kringle 4 type 2 repeats and therefore the assay was sensitive to apo[a] size heterogeneity which, as previously reported (19), would result in overestimation of Lp[a] values in samples with large apo[a] sizes and underestimation of Lp[a] values in samples with small apo[a] sizes. In contrast, the ELISA method in the present study uses a detecting monoclonal antibody (a40) which does not interact with the variably expressed kringle 4 type 2 repeats and therefore this method is not sensitive to apo[a] size heterogeneity. Consequently, we reported the Lp[a] values in nmol/L. The overall Lp[a] values distribution in Whites, exhibiting a high skewness towards low Lp[a] values, is characteristic of our previous observations and those of others. The frequency distribution in Blacks also exhibited a relatively high skewness but one which was quite different from Whites in that the frequency of Lp[a] in Blacks was nearly constant up to 100 nmol/L. This distribution profile differs somewhat from the Lp[a] frequency distribution we previously reported for Blacks (15). Because the cohort in the two studies is essentially the same, the observed differences likely reflect the difference in Lp[a] measurement method. This observation emphasizes the importance of selecting an accurate analytical method insensitive to apo[a] size heterogeneity to draw valid conclusions from the measurement of Lp[a] concentration in epidemiological or clinical studies.

Because apo B protein mass is constant and the available evidence indicates that Lp[a] particles contain a single molecule of apo B-100 (25), we can readily convert our Lp[a] values expressed in nmol/L to mg/L of apo B in Lp[a] by the formula: nmol/L of Lp[a] × 0.5128 = mg/L of Lp[a]-apo B. Thus, whites had a mean and median Lp[a]-apo B value of 24.6 and 10.3 mg/L, respectively, while blacks had a mean and median Lp[a]-apo B value of 47.9 and 38.2 mg/L, respec-

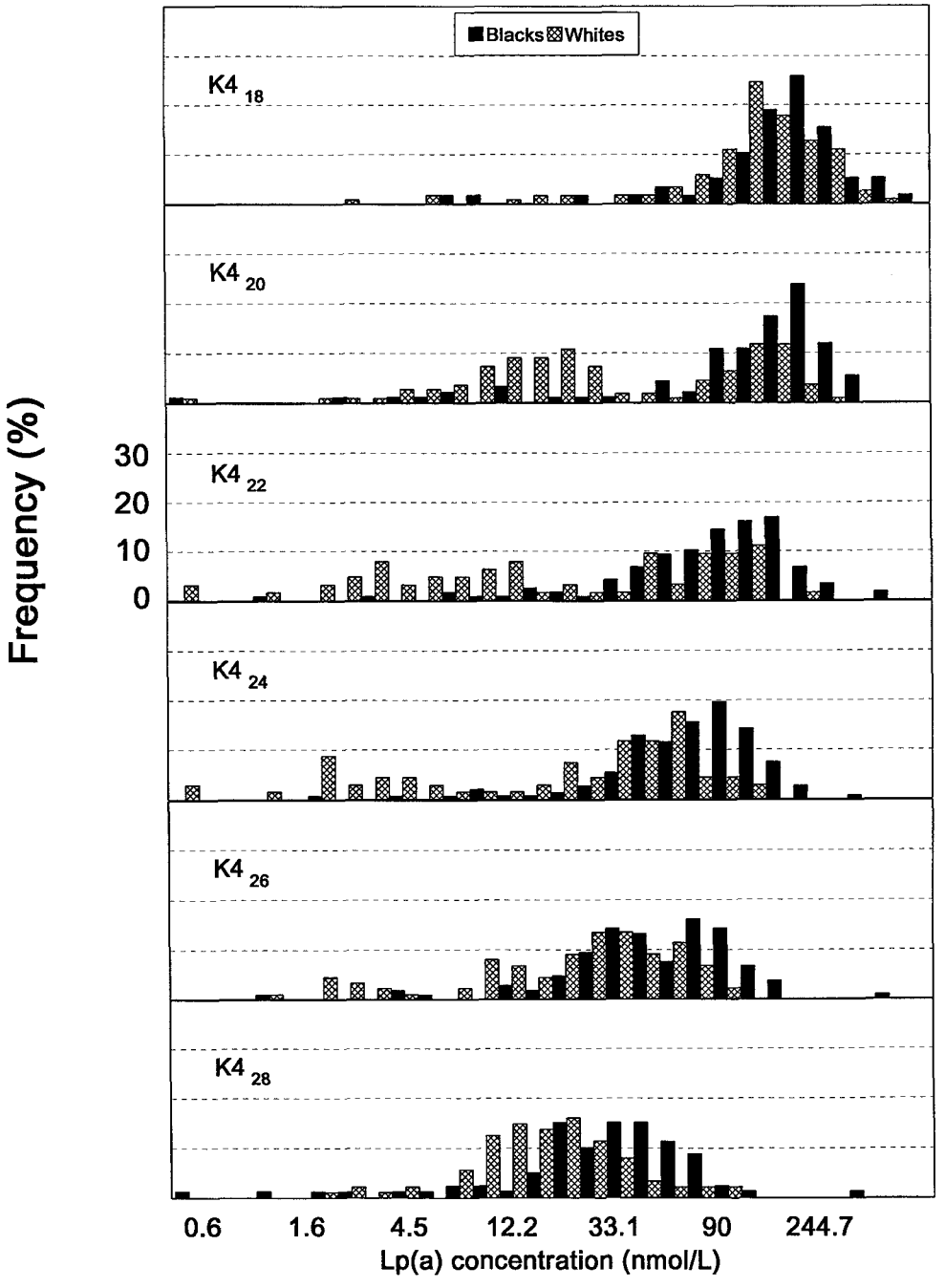


Fig. 6. Kringle specific Lp[a] concentration distribution in blacks and whites exhibiting a predominantly expressed apo[a] polymorph. In order to normalize the Lp[a] frequency distribution, Lp[a] concentrations were grouped into 28 categories of 0.25 natural logarithm intervals. The antilogarithm of the Lp[a] concentration intervals is indicated on the x-axis. Each kringle specific category examined contains at least 50 individuals in each of the two racial groups.

tively. Because the average total plasma apo B value is approximately 1100 mg/L, the contribution of Lp[a]-apo B to the total plasma apoB is usually quite small. However, for subjects with significant elevations of Lp[a], e.g., ≥90th percentile, Lp[a]-apoB can represent a highly significant or a major portion of the total plasma apoB concentration. Because the mass of apo[a] and Lp[a] is variable it is necessary to know the Lp[a] size to convert Lp[a] values from nmol/L to mg/dl. For an Lp[a] particle with a mass of 3.4×10^6 daltons,

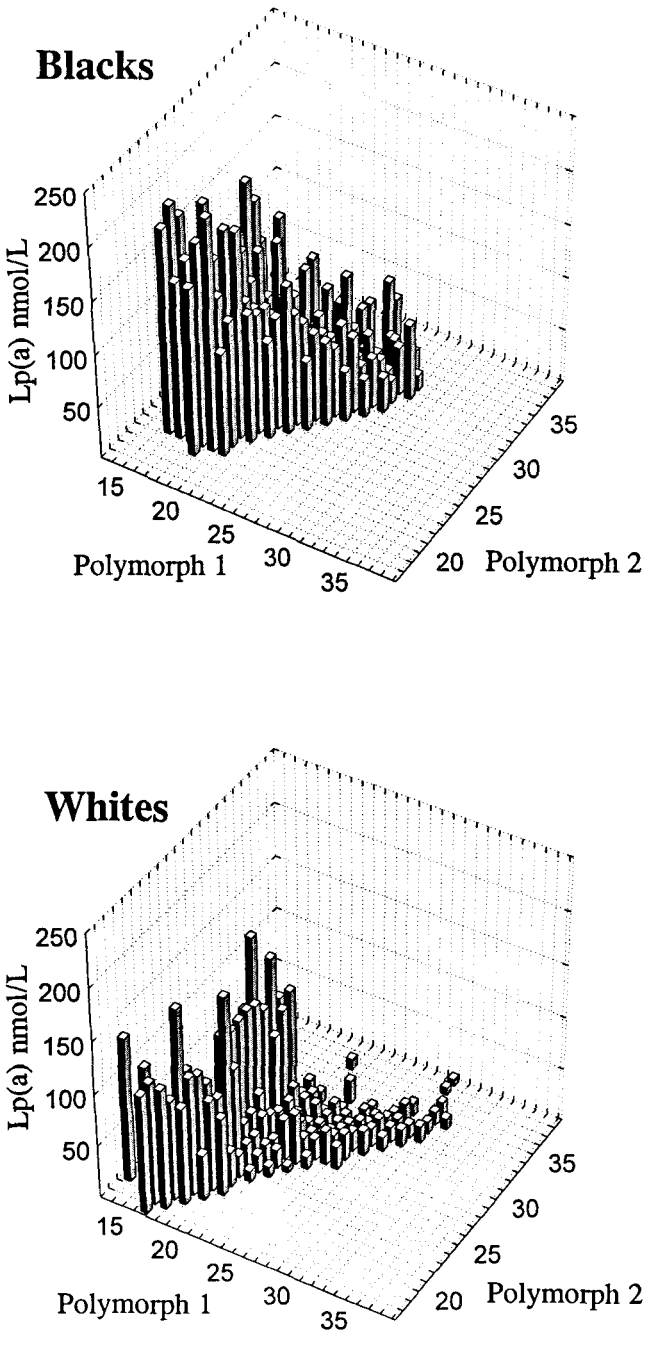


Fig. 7. Relationship between apo[a] phenotype and plasma Lp[a] concentration in blacks and whites. In each three-dimensional plot, polymorph 1 shown on the x-axis on the left corresponds to the smaller apo[a] polymorph and polymorph 2, shown on the y-axis on the right corresponds to the larger apo[a] polymorph. The vertical bars represent the median plasma Lp[a] concentration of each phenotype having at least 5 individuals.

20 nmol/L of Lp[a] would be equivalent to 6.8 mg/dL of Lp[a].

The frequency of individuals who had no detectable apo[a] protein by our sensitive immunoblotting method was 0.39% in whites and 0.05% in blacks. Based on the frequency of the null allele in blacks and whites in the present study using Model III of the likelihood analysis, we would estimate that 6.7% of alleles in whites and 1.0% of the alleles in blacks have undetectable protein products and therefore, operationally have null alleles. It should be emphasized that the frequency of alleles defined as non-expressed or null depends on both the sensitivity of the detection system and the maximum amount of sample applied to the gel.

We previously reported that the frequencies of the apo[a] size isoforms differed between American blacks and whites (7) based on examination of a portion of the CARDIA cohort. In the present study, in which we examined essentially the entire CARDIA cohort available at the fourth examination, we confirmed the previous observation. These findings are consistent with the recent report of Kraft et al. (14) who found that the apo[a] allele frequencies of Tyroleans differ from those of African blacks.

The current study strongly suggests that the immunoblot procedure tends to overestimate the number of homozygotes. The most likely explanation is lack of resolution of apo[a] isoforms differing by only one kringle repeat, particularly among the larger apo[a] sizes. Among allele frequencies estimated with a procedure that allowed for null alleles but not for coalescence between bands, the predicted number of single band phenotypes in blacks and whites is 265 and 459, respectively, as compared to the observed 281 and 493, illustrating the observed excess of single band phenotypes over the number predicted with this model. This model also predicts 21 null phenotypes in whites and 4 in blacks, as compared to the observed 8 and 1, respectively. This high predicted frequency of null alleles is a consequence of the large number of single band phenotypes. The allele frequency estimates that best fit the model are those which assume that 86% of the whites and 61% of the blacks classified as homozygous are in reality heterozygous for the null allele. Therefore, based on the observed excess of homozygotes, deficiency of the null phenotypes and the significant deviation from the Hardy-Weinberg equilibrium, it is apparent that a procedure that allows for null alleles but not for coalescence between closely migrating isoforms in gel electrophoresis, does not adequately explain the observed data. However, when a model that incorporated coalescence of bands was used to estimate allele frequencies, there was no longer evidence for deviation from Hardy-Weinberg equilibrium, and the fit between predicted

and observed numbers of null and single phenotypes was much improved. For example, under model III, which estimates coalescence probabilities simultaneously from both populations, nine whites and no blacks were expected to have a null phenotype, in close agreement with the observed finding of eight whites and one black. Similarly, 426 single phenotypes were expected in whites, and 323 in blacks. Under this model, in whites, 22% of the single phenotypes were predicted to be heterozygotes for the null allele, 60% to be true homozygotes, and 18% to be heterozygous for two similarly sized alleles. In blacks, 35% were predicted to be heterozygotes for the null allele, 12% to be true homozygotes, and 53% to be heterozygotes for similarly sized alleles.

To accurately determine the number of true homozygotes, efforts are currently underway to perform genotype analysis on a subset of the CARDIA population. However, genotype analysis does not permit one to establish whether or not an allele is expressed and thus would not allow one to distinguish between heterozygosity for a null allele or co-migration of two polymorphs in an individual with only one apparent polymorph. For those individuals with two alleles differing significantly in size by genotype analysis it is likely that the product of one allele is not expressed or expressed below the limit of sensitivity. However, for those individuals with two alleles similar in size the appearance of only one polymorph could be due to coalescence of two polymorphs or heterozygosity for a null allele. Family studies would be required to distinguish between these two possibilities.

Although there was an overall inverse relationship between apo[a] size and Lp[a] concentration in both blacks and whites, the nature of this relationship differed significantly. Evaluation of the Lp[a] concentrations as a function of apo[a] isoform indicated considerable variation in Lp[a] concentration at a given apo[a] size. However, the variation was heteroscedastic and differed between the racial groups. The variability was uniformly low for the large apo[a] sizes (K4 $>$ 29) and increased as the apo[a] size decreased. For the intermediate size ($K4_{20}$ to $K4_{29}$) blacks consistently had greater variance in Lp[a] concentrations than whites. Both blacks and whites had the greatest variability associated with the small apo[a] size (K4 $<$ 20).

Despite the strong evidence in case-control studies relating high Lp[a] values to increased CHD, conflicting results have been obtained in prospective studies where six (27–32) out of the nine so far published (33–35) have reported an association between Lp[a] and coronary atherosclerosis. The basis for this apparent ambiguous role of Lp[a] in CHD is unclear. Numerous hypotheses have been formulated to explain the conflicting results; among them are inappropriate statistical analyses, inadequate sample handling and storage, and inaccuracy of the Lp[a] measurement methods due to apo[a] size heterogeneity. Another hypothesis that should be considered is that Lp[a] concentration is not the only factor contributing to the pathological role of Lp[a] but the association of Lp[a] values to apo[a] size isoforms should be taken into consideration. Unfortunately, apo[a] size isoforms have been evaluated in only one of the prospective studies (32). In this study, male cases had a higher frequency of small apo[a] sizes than male controls while no significant difference was found in females. Considering the high number of apo[a] isoforms and the low frequency of the small size polymorphs in the general population, a large number of subjects would need to be studied to achieve unambiguous conclusions.

The intriguing aspect of Lp[a] that strongly points to the possibility that Lp[a] values alone are not responsible for CHD risk comes from the finding that median Lp[a] value of blacks is almost four times that of whites and, despite that, there is no evidence that these increased values result in an increased risk of CHD. Recently, Moliterno et al. (10) reported a lack of association between Lp[a] values and the presence or absence of coronary atherosclerosis in African Americans. One of the major findings of our study, the validity of which is supported by the large number of subjects evaluated, is that whites have a higher frequency of small size apo[a] isoforms and within the small size range, Lp[a] values are high but not significantly different between blacks and whites. The major determinant of the difference in Lp[a] values between these two racial groups is the fact that in the range of intermediate apo[a] sizes blacks have four to five times the Lp[a] values of whites. These findings are in keeping with our hypothesis of an association between the small apo[a] size polymorphs with high Lp[a] level and CHD risk. This association could explain why high Lp[a] levels are less predictive for CHD in blacks than in whites because blacks have fewer individuals within the population with the combination of elevated Lp[a] concentration and small size apo[a] polymorph.

APPENDIX

EM algorithm

The EM algorithm iteratively obtains estimates of allele frequencies and coalescence probabilities by maximizing the likelihood of the observed data, L_{obs}, with respect to the genotype frequencies, p_{ij}:

$$L_{obs} = \frac{n!}{\prod_{ij} n_{ij}!} \cdot \prod_{ij} \hat{p}_{ij}^{n_{ij}}$$

where the products run over all possible genotypes. Given the total number of individuals, n, and an initial estimate of the allele frequencies, p_i, and the number of each genotype, n_{ij}, the iterative steps in the process are: *1*) estimate the coalescence probabilities given the genotype frequencies and number of each genotype; *2*) estimate the number of each genotype given the genotype frequencies and the coalescence probabilities; and *3*) estimate the genotype frequencies, p_{ij}, given the number of each genotype and the coalescence probabilities. These steps cycle until the estimates converge. Details are given below. The algorithm starts with an initial estimate of allele frequencies, and an initial estimate of the number of individuals with each genotype. The precision of these initial estimates is unimportant, so initial estimates of allele frequencies were made by assuming that half of each of the apparent homozygotes for allele i were i/i homozygotes, and half were $i/0$ heterozygotes, where 0 represents the null allele.

Estimation of the coalescence probabilities is done as follows. For a given migration distance category between two bands, d, the probability of coalescence is estimated by:

$$\Pr(coalescence | a \le d < b) = 1 - \frac{obs_{ab}}{\exp_{ab}}$$

where obs_{ab} denotes the observed number of heterozygotes whose alleles were between a mm and b mm apart, and $\exp_{ab}$ denotes the expected number of such heterozygotes, based on the current allele frequency estimates. These quantities can be written as follows:

$$obs_{ab} = \sum_{i<j} n_{ij} \cdot I[a \le |d_i - d_j| < b]$$

$$\exp_{ab} = n \sum_{i<j} 2\hat{p}_i\hat{p}_j \cdot I[a \le |d_i - d_j| < b]$$

where d_i is the migration distance of allele i, I is an indicator function, $\hat{p}_i$ is the current allele frequency for allele i, and n is the number of individuals, either as the total, or when subscripted, the number of heterozygotes for genotype i/j. In both cases the sum is over all heterozygote classes. Note that the expectation assumes HWE. The three coalescent classes are defined as follows: C_1 is the probability that alleles i and $i-1$ will coalesce to allele i, given the distance between the two alleles; C_2 is the probability that alleles i and $i+1$ will coalesce to allele i; C_3 is the probability that alleles $i-1$ and $i+1$ will coalesce; and the C_j are determined by the appropriate coalescence probabilities estimated as described above.

Once the coalescence probabilities have been accounted for, the allele and genotype frequencies are estimated more accurately. The probabilities of interest are thus $\Pr(genotype = x/y | phenotype = i)$ where x and y range over the genotypes of interest. These probabilities can be estimated using Bayes' rule:

$$\Pr(geno = x/y | pheno = i) = \frac{\Pr(pheno = i | geno = x/y) \cdot \Pr(geno = x/y)}{\sum_{x,y} \Pr(pheno = i | geno = x/y) \cdot \Pr(geno = x/y)}.$$

The sum in the denominator ranges over the five possible genotypes which accommodate the possibility of coalescence. The component probabilities in this equation can be easily calculated:

$$\Pr(genotype = x/y) = \begin{cases} 2p_x p_y, & x \ne y \\ p_x^2, & x = y \end{cases}$$

$$\Pr(phenotype = i | genotype = 0/i) = 1$$

$$\Pr(phenotype = i | genotype = i/i) = 1$$

$$\Pr(phenotype = i | genotype = i-1/i) = 0.5 \cdot \Pr(coalescence | d)$$

$$\Pr(phenotype = i | genotype = i/i+1) = 0.5 \cdot \Pr(coalescence | d)$$

$$\Pr(phenotype = i | genotype = i-1/i+1) = \Pr(coalescence | d).$$

New estimates for the number of individuals in each genotypic class can now be obtained as follows. For each allele i, using equation [1] we can obtain the following weights:

$$w_{1i} = \Pr(genotype\ i/i | phenotype\ i)$$

$$w_{2i} = \Pr(genotype\ i/0 | phenotype\ i)$$

$$w_{3i} = \Pr(genotype\ i/i-1 | phenotype\ i)$$

$$w_{4i} = \Pr(genotype\ i/i+1 | phenotype\ i)$$

$$w_{5i} = \Pr(genotype\ i-1/i+1 | phenotype\ i),$$

from which we obtain updated estimates of the sample sizes for the corresponding genotypes:

$$n_{i-1,i+1}^{\exp} = n_{i-1,i+1}^{obs} + w_{5i} \cdot n_{ii}^{obs}$$

$$n_{i-1,j}^{\exp} = n_{i-1,i}^{obs} + w_{3i} \cdot n_{ii}^{obs}$$

$$n_{i,i+1}^{\exp} = n_{i,i+1}^{obs} + w_{4i} \cdot n_{ii}^{obs}$$

$$n_{i,0}^{\exp} = n_{i,0}^{obs} + w_{2i} \cdot n_{ii}^{obs}$$

$$n_{i,i}^{\exp} = n_{ii}^{obs} \cdot w_{1i}.$$

The expected sample sizes, $n_{i,k}^{\exp}$, for genotypes in which $|i-k| > 2$ are just the observed sample sizes, $n_{i,k}^{obs}$.

Finally, given the expected sample sizes of each geno-

type the allele frequencies for each allele are estimated in the usual way:

$$\hat{p}_i = \frac{2n_{ii}^{\exp} + \sum_j n_{ij}^{\exp}}{2n}.$$

LIKELIHOOD RATIO TESTS

Hardy-Weinberg equilibrium within populations

A test of the null hypothesis that a given population is in HWE was performed by constructing the likelihood ratio statistic

$$2 \cdot \ln\left[\frac{L_{obs}}{L_{\mathrm{HWE}}}\right] = 2 \cdot \{\ln(L_{obs}) - \ln(L_{\mathrm{HWE}})\}.$$

Under the null hypothesis of HWE, this statistic will have a χ^2 distribution, where the degrees of freedom are determined by the difference in the number of parameters estimated under each model. In the above expression,

$$L_{obs} = \frac{n!}{\prod_{ij} n_{ij}!} \cdot \prod_{ij} \hat{p}_{ij}^{n_{ij}}$$

where the products run over all possible genotypes, and $\hat{p}_{ij} = n_{ij}/n$. In these expressions, n_{ij} denotes the number of people with genotype i/j, and n denotes the total number in the population. Thus

$$L_{obs} = \frac{n!}{\prod_{ij} n_{ij}!} \cdot \prod_{ij} \left(\frac{n_{ij}}{n}\right)^{n_{ij}} = \frac{n!}{\prod_{ij} n_{ij}!} \cdot \prod_{ij} n_{ij}^{n_{ij}} \cdot \frac{1}{n^n}.$$

Similarly,

$$L_{\mathrm{HWE}} = \frac{n!}{\prod_{ij} n_{ij}!} \cdot \prod_{ij} \hat{p}_{ij}^{n_{ij}}.$$

The values of $\hat{p}_{ij}$ in L_{HWE} depend on which model was used to estimate the allele frequencies. If nulls have been accounted for, but coalescence ignored,

$$\hat{p}_{ij} = 2 \cdot \hat{p}_i \cdot \hat{p}_j, \quad i \neq j$$

and

$$\hat{p}_{ii} = \hat{p}_i^2 + 2 \cdot \hat{p}_i \cdot \hat{p}_0$$

where $\hat{p}_0$ denotes the estimated frequency of the null allele.

On the other hand, if coalescence has been modeled in the estimation of the allele frequencies, then

$$\hat{p}_{ij} = 2 \cdot \hat{p}_i \cdot \hat{p}_j, \quad i \neq j$$

and

$$\hat{p}_{ii} = \hat{p}_i^2 + 2 \cdot \hat{p}_i \cdot \hat{p}_0 + \hat{p}_{i-1} \cdot \hat{p}_i \cdot C_1 + \hat{p}_{i+1} \cdot \hat{p}_i \cdot C_{2p} + 2 \cdot \hat{p}_{i-1} \cdot \hat{p}_{i+1} \cdot C_3$$

where C_1, C_2 and C_3 are defined as in the previous section. Thus,

$$2 \cdot \ln\left[\frac{L_{obs}}{L_{\mathrm{HWE}}}\right] = 2 \cdot \left\{\sum_{ij} n_{ij} \cdot \ln(n_{ij}) - \sum n_{ij} \cdot \ln(\hat{p}_{ij}) - n \cdot \ln(n)\right\}.$$

The number of parameters estimated in the calculation of L_{obs} is simply equal to the number of genotypes, minus one, or 594 for 34 alleles. The number of parameters estimated in the calculation of L_{HWE} is 33 (the number of alleles − 1) if coalescence is ignored, and 37 if coalescence is accounted for. Thus the degrees of freedom for the χ^2 test is 561 if coalescence is ignored, and 557 if coalescence is accounted for.

Comparison of allele frequencies between populations

Another question of interest is whether or not the allele frequencies are the same in the two populations. The null hypothesis that the two sets are the same was tested using the following likelihood ratio statistic:

$$2 \cdot \ln\left[\frac{L_A}{L_O}\right] = 2 \cdot \{\ln(L_A) - \ln(L_O)\}$$

where L_O denotes the likelihood under the null hypothesis that the two sets of allele frequencies are the same, and L_A denotes the likelihood under the alternative hypothesis that the two sets of allele frequencies are different. Thus

$$L_A = \frac{n^w!}{\prod_{ij} n_{ij}^w} \cdot \prod_{ij} (\hat{p}_{ij}^w)^{n_{ij}^w} \cdot \frac{n^b!}{\prod_{ij} n_{ij}^b} \cdot \prod_{ij} (\hat{p}_{ij}^b)^{n_{ij}^b},$$

where the symbols have the same meaning as before, but the superscripts indicate to which population the term applies. There are two sets of allele frequency estimates involved in this likelihood. L_O, on the other hand, is given by

$$L_O = \frac{n^w!}{\prod_{ij} n_{ij}^w} \cdot \prod_{ij} (\hat{p}_{ij})^{n_{ij}^w} \cdot \frac{n^b!}{\prod_{ij} n_{ij}^b} \cdot \prod_{ij} (\hat{p}_{ij})^{n_{ij}^b},$$

where there is only one set of allele frequencies. It is easy to see that

$$2 \cdot \ln\left[\frac{L_A}{L_O}\right] = 2 \cdot \left\{ \sum_{ij} n_{ij}^{w} \cdot \ln(\hat{p}_{ij}^{w}) + \sum_{ij} n_{ij}^{u} \cdot \ln(\hat{p}_{ij}^{b}) - \sum_{ij} (n_{ij}^{w} + n_{ij}^{b}) \cdot \ln(\hat{p}_{ij}) \right\}.$$

In the calculation of L_A, 70 parameter estimates are required, whereas in the calculation of L_O, 37 parameter estimates are required, so the total degrees of freedom for the test is 33.

This research was supported by grant HL30086 from the National Heart, Lung, and Blood Institute. The authors wish to express their appreciation to David R. Jacobs, Jr. and Marlys L. Koschinsky for their careful review of the manuscript and helpful suggestions.

Manuscript received 30 July 1996 and in revised form 27 September 1996.

REFERENCES

1. Marcovina, S. M., and J. D. Morrisett. 1995. Structure and metabolism of lipoprotein[a]. *Curr. Opin. Lipidol.* **6:** 236–145.
2. Utermann, G., H. J. Menzel, H. G. Kraft, H. C. Duba, H. G. Kemmler, and C. Seitz. 1987. Lp[a] glycoprotein phenotypes: inheritance and relation to Lp[a]-lipoprotein concentrations in plasma. *J. Clin. Invest.* **80:** 458–465.
3. Lackner, C., E. Boerwinkle, C. C. Leffert, T. Rahmig, and H. H. Hobbs. 1991. Molecular basis of apolipoprotein[a] isoform size heterogeneity as revealed by pulsed-field gel electrophoresis. *J. Clin. Invest.* **87:** 2153–2161.
4. McLean, J. W., J. E. Tomlinson, W. J. Kuang, D. L. Eaton, E. Y. Chen, G. M. Fless, A. M. Scanu, and R. M. Lawn. 1987. cDNA sequence of apolipoprotein[a] is homologous to plasminogen. *Nature.* **330:** 132–137.
5. Lackner, C., J. C. Cohen, and H. H. Hobbs. 1993. Molecular definition of the extreme size polymorphism in apolipoprotein[a]. *Hum. Mol. Genet.* **2:** 933–940.
6. van der Hoek, Y. Y., M. E. Wittekoek, U. Beisiegel, J. J. Kastelein, and M. L. Koschinsky. 1993. The apolipoprotein[a] kringle IV repeats which differ from the major repeat kringle are present in variably sized isoforms. *Hum. Mol. Genet.* **2:** 361–366.
7. Marcovina, S. M., Z-H. Zhang, V. P. Gaur, and J. J. Albers. 1993. Identification of 34 apolipoprotein[a] isoforms: differential expression of apolipoprotein[a] alleles between American blacks and whites. *Biochem. Biophys. Res. Commun.* **191:** 1192–1196.
8. Marcovina, S. M., D. M. Levine, and G. Lippi. 1994. Lipoprotein[a]: structure, measurement, and clinical significance. *In* Laboratory Measurement of Lipids, Lipoproteins, and Apolipoproteins. N. Rifai and R. Warnick editors. AACC Press, Washington, DC. 235–263.
9. Sorrentino, M. J., C. Vielhauer, J. D. Eisenbart, G. M. Fless, A. M. Scanu, and T. Feldman. 1992. Plasma lipoprotein[a] concentration and coronary artery disease in black patients compared with white patients. *Am. J. Med.* **93:** 658–662.
10. Moliterno, D. J., E. V. Jokinen, A. R. Miserez, R. A. Lange, J. E. Willard, E. Boerwinkle, L. D. Hillis, and H. H. Hobbs. 1995. No association between plasma lipoprotein[a] concentrations and the presence or absence of coronary atherosclerosis in African-Americans. *Arterioscler. Thromb. Vasc. Biol.* **15:** 850–855.
11. Parra, H. J., I. Luyeye, C. Bouramoue, C. Demarquilly, and J. C. Fruchart. 1987. Black–White differences in serum Lp[a] lipoprotein levels. *Clin. Chim. Acta.* **168:** 27–31.
12. Helmhold, M., J. Bigge, R. Muche, J. Mainoo, J. Thiery, D. Seidel, and V. W. Armstrong. 1991. Contribution of the apo[a] phenotype to plasma Lp[a] concentrations shows considerable ethnic variation. *J. Lipid Res.* **32:** 1919–1928.
13. Sandholzer, C., D. M. Hallman, N. Saha, G. Sigurdsson, C. Lackner, A. Csaszar, E. Boerwinkle and G. Utermann. 1991. Effects of the apolipoprotein[a] size polymorphism on the lipoprotein[a] concentration in 7 ethnic groups. *Hum. Genet.* **86:** 607–614.
14. Kraft, H. G., A. Lingenhel, R. W. C. Pang, R. Delport, M. Trommsdorff, H. Vermaak, E. D. Janus, and G. Utermann. 1996. Frequency distributions of apolipoprotein[a] kringle IV repeat alleles and their effects on lipoprotein[a] levels in Caucasian, Asian, and African populations: the distribution of null alleles is non-random. *Eur. J. Hum. Genet.* **4:** 74–87.
15. Marcovina, S. M., J. J. Albers, D. R. Jacobs, Jr., L. L. Perkins, C. E. Lewis, B. V. Howard, and P. Savage. 1993. Lipoprotein[a] concentrations and apolipoprotein[a] phenotypes in Caucasians and African Americans. The CARDIA Study. *Arterioscler. Thromb.* **13:** 1037–1045.
16. Gaw, A., E. Boerwinkle, J. C. Cohen, and H. H. Hobbs. 1994. Comparative analysis of the apo[a] gene, apo[a] glycoprotein, and plasma concentrations of Lp[a] in three ethnic groups: evidence for no common "null" allele at the apo[a] locus. *J. Clin. Invest.* **93:** 2526–2534.
17. Friedman, G. D., G. R. Cutter, R. P. Donahue, G. H. Hughes, S. B. Hulley, D. R. Jacobs, Jr., K. Liu, and P. J. Savage. 1988. CARDIA: study design, recruitment, and some characteristics of the examined subjects. *J. Clin. Epidemiol.* **41:** 1105–1116.
18. Cutter, G. R., G. L. Burke, A. R. Dyer, G. D. Friedman, J. E. Hilner, G. H. Hughes, S. B. Hulley, D. R. Jacobs, Jr., K. Liu, T. A. Manolio, A. Oberman, L. L. Perkins, P. J. Savage, J. R. Serwitz, S. Sidney, and L. E. Wagenknecht. 1991. Cardiovascular risk factors in young adults. The CARDIA baseline monograph. *Controlled Clin. Trials.* **12:** 1S–77S.
19. Marcovina, S. M., J. J. Albers, B. Gabel, M. L. Koschinsky, and V. P. Gaur. 1995. Effect of the number of apolipoprotein[a] kringle 4 domains on immunochemical measurements of lipoprotein[a]. *Clin. Chem.* **41:** 246–255.
20. Marcovina, S. M., H. H. Hobbs, and J. J. Albers 1996. Relation between number of apolipoprotein[a] kringle 4 repeats and mobility of isoforms in agarose gel: basis for a standardized isoform nomenclature. *Clin. Chem.* **42:** 436–439.
21. Ceppellini, R., M. Siniscalco, and C. A. B. Smith. 1955. The estimation of gene frequencies in a random-mating population. *Ann. Hum. Genet.* **20:** 97–115.
22. Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. R. Stat. Soc. B.* **39:** 1–38.
23. Devlin, B., N. Risch, and K. Roeder. 1990. No excess of homozygosity at loci used for DNA fingerprinting. *Science.* **249:** 1416–1420.
24. Guo, S. W., and E. A. Thompson. 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics.* **48:** 361–372.

25. Kleinbaum, D. G., L. L. Kupper, and K. E. Muller. 1988. Applied Regression Analysis and Other Multivariable Methods. 2nd ed. PWS-Kent Publishing Co., Boston, MA. 266–268.
26. Albers, J. J., H. K. Kennedy, and S. M. Marcovina. 1996. Evidence that Lp[a] contains one molecule of apo[a] and one molecule of apoB: evaluation of amino acid analysis data. *J. Lipid Res.* **37:** 192–196.
27. Rosengren, A., L. Wilhelmsen, E. Eriksson, B. Risberg, and H. Wedel. 1990. Lipoprotein[a] and coronary heart disease: a prospective case-control study in a general population sample of middle aged men. *Br. Med. J.* **301:** 1248–1251.
28. Sigurdsson, G., A. Baldursdottir, H. Sigvaldason, U. Agnarsson, G. Thorgeirsson, and N. Sigfusson. 1992. Predictive value of apolipoproteins in a prospective survey of coronary artery disease in men. *Am. J. Cardiol.* **69:** 1251–1254.
29. Schaefer, E. J., S. Lamon-Fava, J. L. Jenner, J. R. McNamara, J. M. Ordovas, C. E. Davis, J. M. Abolafia, K. Lippel, and R. I. Levy. 1994. Lipoprotein[a] levels and risk of coronary heart disease in men. The Lipid Research Clinics Coronary Primary Prevention Trial. *J. Am. Med. Assoc.* **271:** 999–1003.
30. Cremer, P., D. Nagel, B. Labrot, H. Mann, R. Muche, H. Elster, and D. Seidel. 1994. Lipoprotein Lp[a] as predictor of myocardial infarction in comparison to fibrinogen, LDL cholesterol and other risk factors: results from the prospective Gottingen Risk Incidence and Prevalence Study (GRIPS). *Eur. J. Clin. Invest.* **24:** 444–453.
31. Boston, A. G., L. A. Cupples, J. L. Jenner, J. M. Ordovas, L. J. Seman, P. W. F. Wilson, E. J. Schaefer, and W. P. Castelli. 1996. Elevated plasma lipoprotein[a] and coronary heart disease in men aged 55 years and younger. *J. Am. Med. Assoc.* **276:** 544–548.
32. Wild, S. H., S. P. Fortmann, and S. M. Marcovina. 1996. A prospective case control study of lipoprotein[a] levels and apo[a] size and risk of coronary heart disease in Stanford five city project participants. *Arterioscler. Thromb. Vasc. Biol.* In press.
33. Jauhiainen, M., P. Koskinen, C. Ehnholm, M. H. Frick, M. Manttari, V. Manninen, and J. K. Huttunen. 1991. Lipoprotein[a] and coronary heart disease risk: a nested case-control study of the Helsinki Heart Study participants. *Atherosclerosis.* **89:** 59–67.
34. Ridker, P. M., C. H. Hennekens, and M. J. Stampfer. 1993. Prospective study of lipoprotein[a] and the risk of myocardial infarction. *J. Am. Med. Assoc.* **270:** 2195–2199.
35. Alfthan, G., J. Pekkanen, M. Jauhiainen, J. Pikaniemi, M. Karvonen, J. Tuomilehto, J. T. Salonen, and C. Ehnholm. 1994. Relation of serum homocysteine and lipoprotein[a] concentrations to atherosclerotic disease in a prospective Finnish population based study. *Atherosclerosis.* **106:** 9–19.