Cladistic analysis of the apolipoprotein AI-CIII-AIV gene cluster using a healthy French Canadian sample. I. Haploid analysis

M. B. HAVILAND,[1] A. M. KESSLING,[2] J. DAVIGNON[3] AND C. F. SING[1]

[1] *Department of Human Genetics, University of Michigan, Ann Arbor, MI 48109-0618, USA*

[2] *Departments of Biochemistry and Molecular Genetics, St. Mary's Hospital Medical School, London, England*

[3] *Hyperlipidemia and Atherosclerosis Research Group, Clinical Research Institute of Montreal, Montreal (QC), Canada*

SUMMARY

A cladistic analysis was carried out to identify haplotypes hypothesized to differ for functional DNA sequence variations within the apolipoprotein (apo) AI-CIII-AIV gene cluster that affect plasma lipid, lipoprotein and apolipoprotein levels. A sample of unrelated healthy French Canadians was studied. First, a cladogram of the observed apo AI-CIII-AIV haplotypes was estimated. Then this cladogram was used to define a statistical analysis of the association between haplotype variation and variation in plasma lipid, lipoprotein and apolipoprotein levels. Three haplotypes were identified which were associated with small (5–12% of the total sum of squares) pleiotropic effects on plasma lipid, lipoprotein and apolipoprotein traits and these effects were context, i.e. gender, dependent.

INTRODUCTION

One of the objectives of human quantitative genetics is to characterize the DNA sequence variations that influence interindividual phenotypic differences in traits that are continuously distributed in the population at large. The first step in accomplishing this is to identify haplotypes which differ for these functional sequence variations. This study is aimed at identifying such haplotypes associated with interindividual variation in quantitative traits known to contribute to the risk of developing coronary artery disease (CAD). We present here a cladistic analysis (Templeton *et al.* 1987, 1988, 1992; Sing *et al.* 1992; Haviland, 1993; Templeton & Sing, 1993) to identify haplotypes hypothesized to have DNA sequence differences, within the apolipoprotein (apo) AI-CIII-AIV gene cluster, which have relatively small effects on plasma lipid, lipoprotein and apolipoprotein levels.

We begin with a justification for the use of a cladistic analysis for identifying haplotypes that differ for functional interindividual DNA sequence variations in a gene hypothesized to be involved in the genetic architecture of a quantitative trait. Genetic architecture is defined by the number of genes involved in the metabolism of the trait, the number of alleles at each gene and their relative frequencies, the arrangement of these alleles into genotypes, the impact of each of the alleles, single locus genotypes and multilocus genotypes on intra- and interindividual trait variation and the impact of each of these alleles and genotypes on the relationships between the trait and other traits (Boerwinkle *et al.* 1986; Sing *et al.* 1992*a*, *b*). There are two

general approaches taken when studying the genetic architecture of continuously distributed risk factor traits: the top-down and the bottom-up approaches (reviewed by Sing *et al.* 1985, 1988; Sing & Moll, 1990).

The top-down approach involves estimating the impact of the effects of unmeasured genetic and environmental factors on trait variation among individuals in the population at large using genealogical relationships (biometrical top-down approach) or characterizing the genotypes associated with specific trait values (molecular top-down approach). The biometrical top-down approach estimates the genetic component of trait variation but cannot be used to address questions about the location of functional sequence variations (reviewed by MacCluer, 1992; Moll, 1993). The molecular top-down approach involves selecting individuals at the tails of the trait distribution, e.g. individuals with very high and very low plasma cholesterol levels, with the ultimate aim of comparing their DNA sequences to determine sequence differences that may be responsible for the extreme phenotypes. The molecular top-down approach has been successful in identifying rare mutations with large effects (e.g. Goldstein & Brown, 1989), but has contributed little to our understanding of the genetic architecture of a quantitative trait in the population at large. The molecular top-down approach, which uses samples of individuals from the tails of a trait distribution, gives no information about the penetrance functions (i.e. distribution of phenotypes in the population at large) associated with each of the many functional interindividual DNA sequence variations that are involved in the genetic architecture of a trait (see for example, Hobbs *et al.* 1989; Roy *et al.* in press). Also, identifying the specific DNA sequence variations with a functional effect is not a trivial problem (e.g. Choudhary & Laurie, 1991; Laurie *et al.* 1991) because there are certain to be many nucleotide differences between sequences which are not functionally involved in the determination of phenotypic differences (Ohno, 1972; Kimura, 1983; Zuckerkandl, 1992). A method to reduce the number of nucleotide differences to those most likely to be responsible for a phenotypic difference is needed.

The bottom-up approach involves measuring genetic variation (e.g. RFLP markers or DNA sequences) and studying its association with phenotypic variation in samples representative of the frequency distribution of the trait in a particular population of inference (reviewed by Cooper & Clayton, 1988; Lusis, 1988; Fisher *et al.* 1989). Haplotypes that associate with significantly different trait levels would then be candidates for sequence comparison studies to find the functional interindividual DNA sequence variation(s). Unfortunately, results from published single locus genotype–phenotype association studies have been largely inconclusive due to the inconsistency of results among the samples studied (Cooper & Clayton, 1988). Inconsistent findings may be a consequence of an unrepresentative sampling criterion or small sample sizes inflating type I and II errors respectively. Furthermore, even if samples are large and representative of the populations of interest, the populations from which the samples were drawn may be genetically different. For example, the functional interindividual DNA sequence variation may be present in some and not other populations, the linkage disequilibria relationships between marker loci and functional interindividual DNA sequence variations may differ among populations, and/or the frequencies of alleles at other genetic loci, involved in gene by gene interactions with the functional interindividual DNA sequence variations, may differ among populations (Bodmer & Thomson, 1977; Cooper & Clayton, 1988; Schull & Hanis, 1990; Haviland *et al.* 1991; Zerba & Sing, 1993). Although single locus genotype–phenotype

association studies have proved useful in confirming the involvement of candidate genes in the determination of CAD risk factor trait levels (Kessling *et al.* 1992), the inferences that can be drawn from these studies about genetic architecture are limited.

Since RFLP sites are common in the genome, many investigators type more than one RFLP in a region and carry out a series of single RFLP–phenotype association studies. RFLP-phenotype association studies that compare results of multiple single RFLP analyses to decide which RFLP is closest to the functional interindividual DNA sequence variation are inappropriate for two reasons. First, this approach inappropriately assumes that the RFLP sites are statistically independent, i.e. the RFLP sites are in linkage equilibria. Secondly, in short regions, there is a lack of correlation between the level of disequilibrium and the distance between loci (e.g. Litt & Jorde, 1986; Borresen *et al.* 1988; Hegele *et al.* 1990; Haviland *et al.* 1991; Zerba *et al.* 1991), so results from multiple single RFLP–phenotype association studies cannot be compared to localize the functional interindividual DNA sequence variations.

Rather than do a series of single RFLP-phenotype association studies, multiple RFLP haplotypes have been constructed for genotype–phenotype association studies (e.g. Aquadro *et al.* 1992). But, as the number of markers increases and the number of individuals in each genotype class decreases, the statistical power to detect significant phenotypic differences among haplotypes decreases. This decreased statistical power becomes critical in the study of quantitative traits because it is expected that the majority of the effects contributing to the genetic architecture of a quantitative trait will be small (Tanksley, 1993; Haviland *et al.* submitted). Also, because each individual has a unique genomic sequence, there is the potential for the number of possible haplotypes to be greater than the number of individuals in the sample or even in the population being studied. Even with modest numbers of RFLP markers, the number of pairwise comparisons among haplotypes needed to identify which haplotypes are associated with significant phenotypic differences may exceed the number of available degrees of freedom (Templeton *et al.* 1987). To solve these problems, Templeton *et al.* developed the theory for a cladistic analysis of the association between haplotype variation and phenotype variation (Templeton *et al.* 1987, 1988, 1992; Sing *et al.* 1992; Haviland, 1993; Templeton & Sing, 1993). This approach uses the available degrees of freedom to test only those contrasts that have the highest probability of calling attention to haplotypes that differ for functional interindividual sequence variation(s).

A cladistic analysis of the apo AI-CIII-AIV gene cluster will be described here. The apolipoproteins, apo AI, apo CIII and apo AIV, are known to be involved in lipoprotein metabolism and many genotype–phenotype association studies have established that variations in the genes coding for these apolipoproteins are associated with interindividual variation in plasma lipid, lipoprotein and apolipoprotein traits (reviewed by Breslow, 1988; Humphries, 1988; Lusis, 1988; Boerwinkle & Hixson, 1990; Chan *et al.* 1990; Kessling *et al.* 1992). But, little progress has been made in the identification of candidate DNA sequences responsible for the impact of these genes on plasma lipid, lipoprotein and apolipoprotein levels and ultimately the progression of disease (Karathanasis *et al.* 1983, 1987; Pagani *et al.* 1989; Smith *et al.* 1992; Tuteja *et al.* 1992; Von Eckardstein *et al.* 1992). In this paper, we report the identification of three apo AI-CIII-AIV haplotypes hypothesized to carry functional DNA sequence variations with small pleiotropic effects on plasma lipid, lipoprotein and apolipoprotein levels that are gender dependent.
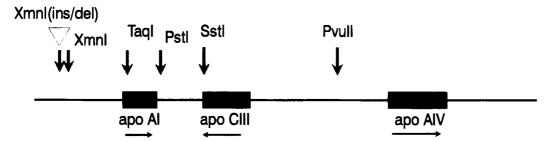
Fig. 1. Diagram of apo AI-CIII-AIV region on chromosome 11 depicting locations of apolipoprotein genes and six RFLP loci. The arrows below the genes denote the direction of transcription.

METHODS

Cladistic analysis

The cladistic analysis has been described fully elsewhere (Templeton et al. 1987, 1988, 1992; Sing et al. 1992; Haviland, 1993; Templeton & Sing, 1993) and is only briefly summarized here. Two steps are involved. In the first step an unrooted cladogram, i.e. network, is built by organizing the haplotypes such that the mutational steps, hypothesized to reflect the evolutionary history that produced the current array of haplotypes, are displayed. The method of maximum parsimony was used to estimate the cladogram (Felsenstein, 1983; Sober, 1983; Stewart, 1993). The gene cladogram is then used to define a statistical design with which to analyse the association between haplotype variation and phenotype variation.

In the second step, instead of executing an analysis which considers all possible pairwise comparisons between haplotypes, a nested testing scheme is defined by the cladogram structure. The nesting starts by defining each unique haplotype as a 0-step clade, where step refers to single mutational changes that result in the gain or loss of a restriction cut site or the creation of the insertion/deletion polymorphism. One-step clades are defined by starting at terminal haplotypes (i.e. ones that are connected to only one other haplotype) and moving in one step. Two-step, and higher, clades are defined analogously. The phenotypic values of a trait are then compared among individuals who are grouped according to haplotype relationships defined by the cladogram. Haplotypes that have few RFLP differences, i.e. are near each other in the cladogram, but are associated with statistically significant phenotypic differences, are those hypothesized to differ for functional DNA sequence variations. The functional DNA sequence variations will most likely not be the RFLP markers used to define the haplotypes.

Sample, lipid, lipoprotein and apolipoprotein determination and RFLP typing

The sample consisted of 167 individuals of French Canadian (FC) ancestry drawn from a larger sample of healthy unrelated white-collar workers at Hydro-Quebec, a utility company in Montreal, Canada. This larger sample has been described elsewhere (Haviland et al. 1991; Xhignesse et al. 1991; Zerba et al. 1991; Kessling et al. 1992). The a priori exclusion criteria used to define the healthy sample were: age below 20 or above 59 years, known hypertension, diabetes, glucose intolerance, hyperlipidemia, gross obesity (body mass index > 30 kg/m^2), non-fasting state, or the taking of regular, or recent, prescription or non-prescription medication (except women taking contraceptive or replacement hormones). An a posteriori exclusion procedure was also used, including the following criteria: hyperuricemia, hyperglycemia, abnormal liver function tests, urea or electrolytes, non-FC ancestry, missing results, known

kinship with others in the sample, or lipid, lipoprotein or apolipoprotein levels greater than 4 s.d. from the sample mean. To determine ancestry, the sampled individuals were asked whether they knew of any ancestor of theirs who was not FC and to list the surnames and maiden names of their grandparents. 167 individuals were successfully typed for six RFLP loci in the apo AI-CIII-AIV gene cluster. This sample has five fewer individuals than the one described in Haviland *et al.* (1991) because one individual was removed due to missing plasma lipoprotein data and four individuals were removed during a subsequent verification of the RFLP typings (see below).

Plasma levels of total cholesterol, LDL-C, apo B, VLDL-C, apo AI, HDL-C, triglycerides and lipoprotein (a) (Lp(a)) were determined as described elsewhere (Xhignesse *et al.* 1991; Kessling *et al.* 1992). DNA extraction, digestion, blotting, hybridization, and autoradiography have also been described elsewhere (Kessling *et al.* 1992). Two probes were used. Taken 5′ to 3′, they were a 2·2 kb *Pst*I genomic fragment (Kessling *et al.* 1985), which was used to detect the *Xmn*I (Kessling *et al.* 1985; Coleman *et al.* 1986; Kessling *et al.* 1988*b*), *Taq*I (Cohen *et al.* 1986), *Pst*I (Kessling *et al.* 1985) and *Sst*I RFLP loci (Rees *et al.* 1983), and 1·05 kb (Kessling *et al.* 1988*c*), a genomic probe which was used to detect the *Pvu*II RFLP (Oettgen *et al.* 1986). Fig. 1 depicts the location of the six RFLP sites in relation to the apolipoprotein genes. In this paper, 0 and 1 denote uncut and cut RFLP alleles (insert and no insert) respectively.

The RFLP data used in this paper were collected and checked in a multistage process. At the time that the RFLP loci were originally typed, data exclusion occurred when three investigators independently reading the autoradiographs did not agree on the typing of an individual's genotype. Typings were repeated when DNA was present in sufficient amounts. At a latter stage, RFLP typings were checked again in four individuals who carried a rare haplotype (seen only once in the sample) and 24 individuals for whom haplotypes could not be determined unambiguously because they were heterozygous at two or more RFLP sites. The *Sst*I and *Pst*I RFLP loci were retyped using a PCR based assay (A. Minnich, personal communication), and the other RFLP loci were checked by re-reading the blots from the original analyses. Due to this data verification, four individuals were removed from the analysis because the original typings could not be confirmed.

The genotype frequencies based on the 147 individuals for whom haplotypes could be determined unambiguously (i.e. they were heterozygous at no more than one RFLP site) are given in Table 1. As shown in the Table, the majority of the sample for whom haplotypes could be unambiguously determined carried at least one copy of a specific haplotype designated haplotype 6. The cladistic analysis of the apo AI-CIII-AIV gene cluster presented in this paper includes only those individuals for whom haplotypes could be determined unambiguously and who carried at least one copy of haplotype 6 ($n = 140$). In other words, only one allele for the apo AI-CIII-AIV gene cluster varied among individuals, the other allele was invariant. Thus, the genotypes could be analysed as if they were haploid and a parametric statistical test could be used to test for differences among individuals grouped according to haplotype relationships (Haviland, 1993).

Statistical analysis

All analyses were done using the SAS statistical package (SAS Institute Inc. 1989). To guard against type II errors, which were likely given that the size of the association between haplotype and trait variation was expected to be small, statistical tests that gave p-values

Table 1. *Genotype frequencies*

| | Haplotype II | | | | | | | | | |
Haplotype I	1	2	3	4	5	6	7	8	9	10
1	0	0	0	0	0	20	0	0	0	0
2	0	0	1	0	0	0	0	0	0	0
3	0	0	0	0	0	5	0	0	0	0
4	0	0	0	0	0	0	0	1	0	0
5	0	0	0	0	0	9	0	0	0	0
6	0	0	0	0	0	74	0	0	0	0
7	0	0	0	0	0	12	0	0	0	0
8	0	0	0	0	0	14	0	2	0	0
9	0	0	1	0	0	0	0	0	0	0
10	0	0	0	0	0	6	0	0	0	2
Total	0	0	2	0	0	140	0	3	0	2

equal to or less than 0·10 were considered significant. The results of the statistical tests are reported using two significance levels, marginally significant $0·05 \leqslant p \leqslant 0·10$ and significant $p \leqslant 0·05$. All analyses were done separately in females and males. Those traits, VLDL-C, triglycerides and Lp(a), that were not normally distributed (results not shown) were transformed with the natural logarithm to reduce skewness. The means and variances for the concomitants and plasma lipid, lipoprotein and apolipoprotein traits were estimated. Homogeneity of variance between the genders was tested using the F-test and the means were tested for homogeneity using the t-test (Sokal & Rohlf, 1981). Satterthwaite's t-test correction was used if the variances were significantly heterogeneous between the genders.

The plasma lipid, lipoprotein and apolipoprotein levels were adjusted separately in females and males by multiple polynomial regression for the concomitants: age, height, and weight to the third power, smoking and, in females, hormone use (Neter *et al.* 1985). The residuals were used in a one-way analysis of variance. Then single degree of freedom contrasts, defined by the cladogram, were made to detect statistically significant associations between haplotype variation and variation in adjusted plasma lipid, lipoprotein and apolipoprotein levels. Tests of the assumption of homogeneity of variance between the pairs of haplotypes contrasted in the analysis of variance were not statistically significant (results not shown).

RESULTS

The means and standard deviations for the concomitants and plasma lipid, lipoprotein and apolipoprotein levels for females and males are given in Table 2. The females were significantly younger, shorter and weighed less than the males. The mean plasma lipid, lipoprotein and apolipoprotein levels were significantly different between females and males for all traits except Lp(a). HDL-C and apo AI levels were higher in females than males while the rest of the plasma lipid, lipoprotein and apolipoprotein levels were lower. The variances for weight, apo B and ln VLDL-C were significantly smaller in females than males. The variances of total cholesterol, LDL-C and HDL-C were also heterogeneous at a less stringent level of statistical significance $(p \leqslant 0·10)$. These results confirm that the distributions of the concomitants and plasma lipid, lipoprotein and apolipoprotein traits were different in females and males and therefore the cladistic analysis was carried out separately by gender.

The maximum parsimony cladogram using only those haplotypes observed in haplotype 6

Table 2. *Means and standard deviations (s.d.) for the concomitants and plasma lipid, lipoprotein and apolipoprotein levels in females and males*

	Females mean	(n = 50) s.d.	Males mean	(n = 90) s.d.	Test of means† p-value	Test of variances p-value
Age (years)	36·18	6·65	39·40	7·85	0·0155	0·2032
Height (m)	1·62	0·05	1·73	0·06	0·0001	0·2179
Weight (kg)	57·90	5·84	72·71	8·65	0·0001	0·0033
T-Chol. (mg/dl)	179·68	27·76	192·58	35·54	0·0283	0·0601
LDL-C (mg/dl)	112·86	24·95	126·91	31·12	0·0070	0·0926
Apo B (mg/dl)	102·40	20·76	122·24	33·08	0·0001	0·0005
ln VLDL-C (ln mg/dl)	2·78	0·37	3·12	0·48	0·0001	0·0474
ln Trig. (ln mg/dl)	4·18	0·36	4·49	0·50	0·0001	0·0158
HDL-C (mg/dl)	49·58	11·08	40·52	8·96	0·0001	0·0817
Apo AI (mg/dl)	135·70	20·94	126·39	20·99	0·0130	1·0000
ln Lp(a) (ln mg/dl)	1·81	1·23	2·02	1·41	0·3847	0·2777

† Satterthwaites *t*-test correction was used if the variances were significantly heterogeneous between the genders.
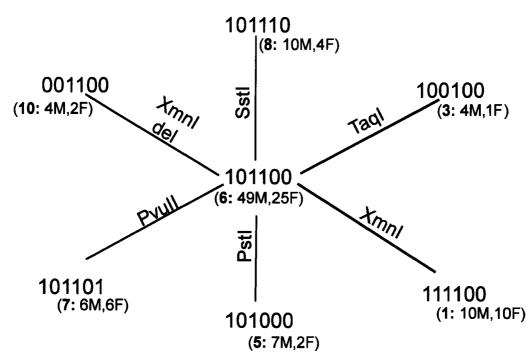

Fig. 2. Maximum parsimony cladogram of the apo AI-CIII-AIV haplotypes. The RFLP sites are listed 5' to 3' (*Xmn* I (*Ins/Del*), *Xmn* I, *Taq* I, *Pst* I, *Sst* I and *Pvu* II) and 1 and 0 denote cut and uncut (no insert and insert) respectively. The designated haplotype number, number of males and number of females are given in parentheses. The RFLP site that changes between the haplotypes is given on each branch.

homo- (n = 74) and heterozygotes (n = 66) is given in Fig. 2. Each connection in the cladogram occurred between haplotypes differing for a single RFLP marker and defined a contrast between individuals heterozygous for haplotype 6 and individuals homozygous for haplotype 6. With seven haplotypes there are 21 possible pairwise contrasts. By using the cladogram, six evolutionarily relevant contrasts were selected representing exactly the number of degrees of
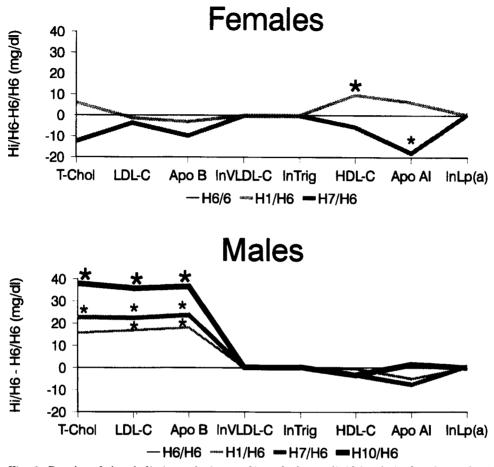
Females

Males

Fig. 3. Results of the cladistic analysis on adjusted plasma lipid levels in females and males. The profiles give the differences in the adjusted plasma lipid, lipoprotein and apolipoprotein levels between H6 heterozygotes and H6 homozygotes for those genotypes that had at least one statistically significant contrast. *Denotes marginally significant differences, $0.05 \leqslant p \leqslant 0.10$, and *denotes significant differences, $p \leqslant 0.05$.

freedom when there are seven haplotypes. Thus, the mean values estimated for each of the six different types of heterozygotes carrying haplotype 6 were compared with the mean of those homozygous for haplotype 6.

Before testing these contrasts, the traits were adjusted separately in females and males for variability in age, height, and weight to the third power, smoking and, in females, hormone use. The means and standard deviations for the concomitants and the adjusted plasma lipid, lipoprotein and apolipoprotein levels by gender are given for each genotype in Appendix I. The one-way analysis of variance and single degree of freedom contrasts are presented in Appendix II. The profile of mean differences for those haplotypes, in genotypes with haplotype 6, that had at least one statistically significant difference in adjusted plasma lipid, lipoprotein or apolipoprotein levels relative to haplotype 6 homozygotes are depicted in Fig. 3.

Using the cladogram to define a priori tests, there were statistically significant contrasts between specific genotypes. Haplotypes 1, 7 and 10, when in genotypes with haplotype 6, were associated with significant differences in adjusted levels of at least one plasma lipid, lipoprotein or apolipoprotein trait in each gender when contrasted with haplotype 6 homozygotes. As depicted in Fig. 3, the 1/6 genotype was significantly different from 6/6 homozygotes in the

level of adjusted HDL-C levels in females; the 1/6 heterozygotes had average HDL-C levels 9·87 mg/dl higher than those observed in 6/6 homozygotes ($p \leqslant 0·02$). Also, the females with the 7/6 genotype had average adjusted apo AI levels 17·96 mg/dl lower than the 6/6 homozygotes and this difference was marginally significant ($p \leqslant 0·07$). In males, haplotypes 1, 7 and 10, when in genotypes with haplotype 6, were associated with significantly higher adjusted apo B, LDL-C and total cholesterol levels relative to haplotype 6 homozygotes and these differences are also depicted in Fig. 3. Males with the 10/6 genotype had average adjusted lipid, lipoprotein and apolipoprotein levels higher than 6/6 homozygotes by 37·93 mg/dl for total cholesterol levels ($p \leqslant 0·03$), 35·77 mg/dl for LDL-C levels ($p \leqslant 0·01$), and 36·72 mg/dl for apo B levels ($p \leqslant 0·02$). The total cholesterol, LDL-C and apo B contrasts between males with the 7/6 and 6/6 genotypes were marginally significant; males with the 7/6 genotype had 22·79 mg/dl higher average adjusted total cholesterol levels ($p \leqslant 0·10$), 22·36 mg/dl higher average adjusted LDL-C levels ($p \leqslant 0·06$) and 23·84 mg/dl higher average adjusted apo B levels ($p \leqslant 0·06$) than 6/6 homozygotes. The average adjusted LDL-C and apo B levels in 1/6 males were also marginally higher than those in 6/6 males, by 16·94 mg/dl ($p \leqslant 0·08$) and 18·07 mg/dl ($p \leqslant 0·07$) respectively. Because certain plasma lipid, lipoprotein and apolipoprotein levels are positively correlated, for example total cholesterol, LDL-C and apo B, it was not unexpected that they were similarly associated with haplotype variation. Thus, using only those individuals for whom haplotypes could be determined unambiguously and who carried at least one copy of haplotype 6, haplotypes with gender specific effects on correlated lipid, lipoprotein and apolipoprotein levels were identified.

DISCUSSION

The analysis reported here is part of an ongoing study of a sample of healthy French Canadians (Haviland *et al.* 1991; Xhignesse *et al.* 1991; Zerba *et al.* 1991; Kessling *et al.* 1992). In a previous study, Kessling *et al.* (1992) established that the apo AI-CIII-AIV gene region should be considered as a contributor to the genetic architecture of lipid, lipoprotein and apolipoprotein traits, rather than just a *candidate* gene region. In the present study, using a subset of the sample, the aim was to identify haplotypes hypothesized to differ for functional DNA sequence variations within the apo AI-CIII-AIV gene cluster that influence variation in plasma lipid, lipoprotein and apolipoprotein traits. The majority of individuals in this sample carried at least one copy of a specific haplotype, designated haplotype 6. This haplotype frequency structure made it possible to carry out a haploid cladistic analysis to address the question: does variation in the other haplotype, carried by haplotype 6 heterozygotes, associate with variation in plasma lipid, lipoprotein and apolipoprotein traits.

The cladogram estimated using the haplotypes present in this sample was consistent with many of the hypotheses from coalescent theory concerning the evolution of a gene region in a species (Crandall & Templeton, 1993). For example, coalescent theory suggests that interior haplotypes (those with more than one connection to other haplotypes) should be more frequent than tip haplotypes. This is supported by the estimated cladogram for the apo AI-CIII-AIV region in which the most frequent haplotype, haplotype 6, was interior and the other haplotypes were all tips. Also, coalescent theory suggests that the most frequent haplotype should have the most connections. Again, the apo AI-CIII-AIV cladogram supports this hypothesis because haplotype 6, the most frequent haplotype, was connected to six other

haplotypes while all the other haplotypes were connected to only one. Thus, haplotype 6 is most likely the oldest haplotype and the other haplotypes descendants of it.

The traditional analysis, using the one-way analysis of variance, did not detect an association between haplotype variation and adjusted plasma lipid trait variation in this sample. But, by using the estimated cladogram to define an *a priori* hypothesis testing scheme, the average adjusted plasma lipid, lipoprotein and apolipoprotein levels of three heterozygote classes, 1/6, 7/6 and 10/6, were found to differ from those of the 6/6 homozygotes. In females, haplotypes 1 and 7 were associated with variation in the adjusted levels of HDL-C and apo AI, although 1/6 females had on average high HDL-C levels while 7/6 females had on average low apo AI levels, as compared to 6/6 homozygotes. In males, haplotypes 1, 7 and 10, when in genotypes with haplotype 6, associated with elevated average adjusted total cholesterol, LDL-C and apo B levels as compared to 6/6 homozygotes. This study cannot distinguish between the possibilities that female and male inferences differed because different functional interindividual DNA sequence variations were acting in the genders, because the same functional interindividual DNA sequence variations had differing effects in the genders, or because of type I sampling errors. Assuming that the latter is not the case, it is unlikely that the interindividual DNA sequence variations acting in females and males were different because two different functional interindividual DNA sequence variations would have to be associated with the three haplotypes, each acting in only one gender. Rather, we hypothesize that the same functional interindividual DNA sequence variations were acting, but the effects of these variations differed between the genders and hence were context dependent.

It was not unexpected that the plasma lipid, lipoprotein and apolipoprotein traits associated with haplotype variation were different in females and males. The distributions of plasma lipid, lipoprotein and apolipoprotein traits in females and males have been shown to be quite different, suggesting that the underlying biology may also be different (e.g. Table 2, National Center for Health Statistics *et al.* 1986; Reilly *et al.* 1990; Kottke *et al.* 1991; The Bezafibrate Infarction Prevention (BIP) Study Group Israel, 1992). Also, there is a growing body of literature demonstrating that genotype effects on measures of lipid metabolism are gender specific (e.g. Hanis *et al.* 1991; Kaprio *et al.* 1991; Reilly *et al.* 1991, 1992, 1994; Xhignesse *et al.* 1991; Kessling *et al.* 1992; Hegele *et al.* 1994). These gender differences may be due, in part, to the association between variation in sex hormone levels and variation in lipid, lipoprotein and apolipoprotein trait levels (Kim & Kalkhoff, 1979; Gorbach *et al.* 1989) and/or variation in sex hormone levels leading to differential regulation of the genes involved in lipid metabolism (e.g. Archer *et al.* 1986). Thus, it is reasonable to hypothesize that the same functional interindividual DNA sequence variations in the apo AI-CIII-AIV gene cluster were acting in females and males in the study reported here but interactions with different modifying factors associated with gender resulted in different penetrance functions.

Many association studies, using single RFLP loci in the apo AI-CIII-AIV gene cluster as the measure of genetic variation, have found associations with variation in plasma lipid, lipoprotein and apolipoprotein levels (reviewed by Humphries *et al.* 1987; Humphries, 1988; Lusis, 1988; Boerwinkle & Hixson, 1990). Most studies, of males only or males and females analysed separately, have found associations between variation in the apo AI-CIII-AIV gene cluster and variation in HDL-C and apo AI levels (Kessling *et al.* 1988b; Paulweber *et al.* 1988; Wile *et al.* 1989; Jeenah *et al.* 1990; Kessling *et al.* 1992). However, the exact RFLP in association with

the differing plasma lipid levels was not the same in all studies (Haviland *et al.* 1991). It is generally hypothesized that the interindividual DNA sequence variation(s) responsible for the association with plasma apo AI and HDL-C levels is located in the apo AI gene. However, until the functional interindividual DNA sequence variation(s) have been identified, the hypothesis that the responsible sequences are located in or around the apo CIII and AIV genes, or in regulatory sequences within or to either side of the cluster, must also be considered.

The association between haplotype variation in the apo AI-CIII-AIV gene cluster with variation in HDL-C and apo AI levels was only observed in females in this cladistic analysis; in males the association was found with variation in apo B, LDL-C and total cholesterol levels. Only two previous analyses found associations between apo AI-CIII-AIV genetic variability and variation in LDL-C levels in males (Kessling *et al.* 1988*a*; Aalto-Setala *et al.* 1991). The association of variation in the apo AI-CIII-AIV gene cluster with variation in many different lipid, lipoprotein and apolipoprotein traits is expected given that lipid metabolism is a complex coherent system of interacting agents (Sing & Reilly, 1993).

Within males, the effect of the three haplotypes identified by the cladistic analysis were similar, but it is unclear whether the functional interindividual DNA sequence variation(s) detected were the same in all three haplotypes or different. It is tempting to hypothesize that the three haplotypes had a common ancestor carrying 'the' functional interindividual DNA sequence variation(s) and the three haplotypes diverged from it, at other sites, over time. This hypothesis is similar to Weiss's characterization of the evolution of functional alleles as a fractal process (Weiss, 1993). Under this characterization, a new mutation will evolve into a clade of descendent alleles, each clonal for the mutation. The alleles in this clade continue to evolve as other mutations accumulate. In this way, evolution generates a fractal structure of alleles and allele effects. In contrast to this hypothesis of a common ancestral DNA sequence variation, it may be that the functional interindividual DNA sequence variations in the haplotypes arose independently and probably at different sites. Of course, the answer will not be definitive until the functional DNA sequence variations have been identified.

A cladistic analysis can be used to generate hypotheses about which haplotypes should be compared to have the greatest chance of finding candidate DNA sequences responsible for the observed phenotypic effects. In the apo AI-CIII-AIV region, this analysis suggests that haplotypes 1, 7 and 10 should be compared to haplotype 6 because they are hypothesized to differ for one or more functional interindividual DNA sequence variation(s) but are similar for a large part of the gene region. A cladistic analysis can also be used to generate hypotheses about which haplotypes are unlikely to differ for functional DNA sequence variations. From this analysis there is no evidence to suggest that haplotypes 3, 5 and 8 differ from haplotype 6 at a functional sequence variation and thus, it is unlikely that the DNA sequence variations that differentiate them have a functional effect on plasma lipid, lipoprotein or apolipoprotein levels.

An iterative cladistic approach is presently being developed aimed at identifying functional DNA sequence variations. This approach expands on the use of the cladistic approach to generate hypotheses about which haplotypes are likely or unlikely to differ for functional DNA sequence variations, and subsequently which DNA sequence variations are likely or unlikely to be functional, by applying the cladistic approach in an iterative fashion (Haviland *et al.* in preparation).

In summary, three haplotypes in the apo AI-CIII-AIV gene cluster were identified that, when in genotypes with haplotype 6, were associated with significant differences in adjusted plasma lipid, lipoprotein and apolipoprotein levels relative to haplotype 6 homozygotes. In females, the haplotypes associated with variation in adjusted levels of HDL-C and apo AI, while in males they were associated with variation in adjusted levels of total cholesterol, LDL-C and apo B. Thus, the study presented here suggests that interindividual DNA sequence variations in the apo AI-CIII-AIV gene region have pleiotropic effects on several measures of lipid metabolism and the nature of these effects is gender dependent.

REFERENCES

AALTO-SETALA, K., VIIKARI, J., AKERBLOM, H. K., KUUSELA, V. & KONTULA, K. (1991). DNA polymorphisms of the apolipoprotein B and A-I/C-III genes are associated with variations of serum low density lipoprotein cholesterol level in childhood. *J. Lipid Res.* **32**, 1477–1487.

AQUADRO, C. F., JENNINGS, R. M., BLAND, M. M., LAURIE, C. C. & LANGLEY, C. H. (1992). Patterns of naturally occurring restriction map variation, dopa decarboxylase activity variation and linkage disequilibrium in the *DdC* gene region of *Drosophila melanogaster*. *Genetics* **132**, 443–452.

ARCHER, T. K., TAM, S.-P. & DEELEY, R. G. (1986). Kinetics of estrogen-dependent modulation of apolipoprotein A–I synthesis in human hepatoma cells. *J. Biol. Chem.* **261**, 5067–5074.

BODMER, W. & THOMSON, G. (1977). Population genetics and evolution of the HLA system. In *HLA and Disease* (ed. J. Dausset & A. Svejgaard), pp. 280–295. Baltimore: Williams and Wilkins Co.

BOERWINKLE, E., CHAKRABORTY, R. & SING, C. F. (1986). The use of measured genotype information in the analysis of quantitative phenotypes in man. I. Models and analytical methods. *Ann. Hum. Genet.* **50**, 181–194.

BOERWINKLE, E. & HIXSON, J. E. (1990). Genes and normal lipid variation. *Curr. Opin. Lipidol.* **1**, 151–159.

BORRESEN, A. L., MOLLER, P. & BERG, K. (1988). Linkage disequilibrium analyses and restriction mapping of four RFLPs at the proα2 (I) collagen locus: lack of correlation between linkage disequilibrium and physical distance. *Hum Genet.* **78**, 216–221.

BRESLOW, J. L. (1988). Apolipoprotein genetic variation and human disease. *Physiol. Rev.* **68**, 85–132.

CHAN, L., BOERWINKLE, E. & LI, W.-H. (1990). Molecular genetics of the plasma apolipoproteins. In *Molecular Biology of the Cardiovascular System* (ed. S. Chien), pp. 183–219. Philadelphia: Lea and Febiger.

CHOUDHARY, M. & LAURIE, C. C. (1991). Use of *in vitro* mutagenesis to analyze the molecular basis of the difference in Adh expression associated with the allozyme polymorphism in *Drosophila melanogaster*. *Genetics* **129**, 481–488.

COHEN, T., KARATHANASIS, S. K., KAZAZIAN, H. H. & ANTONARAKIS, S. E. (1986). DNA polymorphic sites in the human apo AI-CIII-AIV cluster: Taq I and Ava I. *Nucl. Acids Res.* **14**, 1924.

COLEMAN, R. T., GONZALEZ, P. A., FUNKE, H., ASSMANN, G., LEVY-WILSON, B. & FROSSARD, P. M. (1986). Polymorphisms in the apolipoprotein AI-CIII gene complex. *Mol. Biol. Med.* **3**, 213–228.

COOPER, D. N. & CLAYTON, J. F. (1988). DNA polymorphism and the study of disease associations. *Hum. Genet.* **78**, 299–312.

CRANDALL, K. A. & TEMPLETON, A. R. (1993). Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics* **134**, 959–969.

FELSENSTEIN, J. (1983). Parsimony in systematics: biological and statistical issues. *Ann. Rev. Ecol. Syst.* **14**, 313–333.

FISHER, E. A., COATES, P. M. & CORTNER, J. A. (1989). Gene polymorphisms and variability of human apolipoproteins. *Annu. Rev. Nutr.* **9**, 139–160.

GOLDSTEIN, J. L. & BROWN, M. S. (1989). Familial hypercholesterolemia. In *The Metabolic Basis of Inherited Disease* (ed. C. R. Scriver, A. L. Beaudet, W. S. Sly & D. Valle), pp. 1215–1250. New York: McGraw-Hill.

GORBACH, S. L., SCHAEFER, E. J., WOODS, M., LONGCOPE, C. DWYER, J. T., GOLDIN, B. R., MORRILL-LABRODE, A. & DALLAL, G. (1989). Plasma lipoprotein cholesterol and endogenous sex hormones in healthy young women. *Metabolism* **38**, 1077–1081.

HANIS, C. L., HEWETT-EMMETT, D., DOUGLAS, T. C., BERTIN, T. K. & SCHULL, W. J. (1991). Effects of the apolipoprotein E polymorphism on levels of lipids, lipoproteins and apolipoproteins among Mexican-Americans in Starr County, Texas. *Arterioscler. Thromb.* **11**, 362–370.

HAVILAND, M. B. (1993). *A cladistics approach to the genetic analysis of human quantitative data.* Ph.D. thesis, University of Michigan.

HAVILAND, M. B., FERRELL, R. E. & SING, C. F. A cladistic analysis of the relationship between variation in the low density lipoprotein receptor gene region and interindividual variation in plasma lipid, lipoprotein and apolipoprotein levels. (submitted).

HAVILAND, M. B., KESSLING, A. M., DAVIGNON, J. & SING, C. F. (1991). Estimation of Hardy–Weinberg and pairwise disequilibrium in the apolipoprotein AI-CIII-AIV gene cluster. *Am. J. Hum. Genet.* **49**, 350–365.

HAVILAND, M. B., TEMPLETON, A. R. & SING, C. F. An iterative cladistic approach to identifying functional DNA sequence variations in candidate genes. (In preparation).

HEGELE, R. A., EVANS, A. J., TU, L., IP, G., BRUNT, J. H. & CONNELLY, P. W. (1994). A gene-gender interaction affecting plasma lipoproteins in a genetic isolate. *Arterioscler. Thromb.* **14**, 671–678.

HEGELE, R. A., PLAETKE, R. & LALOUEL, J. M. (1990). Linkage disequilibrium between DNA markers at the low-density lipoprotein receptor gene. *Genet. Epid.* **7**, 69–81.

HOBBS, H. H., LEITERSDORF, E., LEFFERT, C. C., CRYER, D. R., BROWN, M. S. & GOLDSTEIN, J. L. (1989). Evidence for a dominant gene that suppresses hypercholesterolemia in a family with defective low density lipoprotein receptors. *J. Clin. Invest.* **84**, 656–664.

HUMPHRIES, S. E. (1988). DNA polymorphisms of the apolipoprotein genes – their use in the investigation of the genetic component of hyperlipidaemia and atherosclerosis. *Atherosclerosis* **72**, 89–108.

HUMPHRIES, S. E., TALMUD, P. J. & KESSLING, A. M. (1987). Use of DNA polymorphisms of the apolipoprotein genes to study the role of genetic variation in the determination of serum lipid levels. In *Molecular Approaches to Human Polygenic Disease* (ed. G. Bock & G. M. Collins), pp. 128–149. Chichester: Wiley.

JEENAH, M., KESSLING, A., MILLER, N. & HUMPHRIES, S. (1990). G to A substitution in the promoter region of the apolipoprotein AI gene is associated with elevated serum apolipoprotein AI and high density lipoprotein cholesterol concentrations. *Mol. Biol. Med.* **7**, 233–241.

KAPRIO, J., FERRELL, R. E., KOTTKE, B. A., KAMBOH, M. I. & SING, C. F. (1991). Effects of polymorphisms in apolipoproteins E, A-IV, and H on quantitative traits related to risk for cardiovascular disease. *Arterioscler. Thromb.* **11**, 1330–1348.

KARATHANASIS, S. K., FERRIS, E. & HADDAD, I. A. (1987). DNA inversion within the apolipoproteins AI/CIII/AIV-encoding gene cluster of certain patients with premature atherosclerosis. *Proc. Natl. Acad. Sci. USA* **84**, 7198–7202.

KARATHANASIS, S. K., ZANNIS, V. I. & BRESLOW, J. L. (1983). A DNA insertion in the apolipoprotein A-I gene of patients with premature atherosclerosis. *Nature, Lond.* **305**, 823–825.

KESSLING, A., OUELLETTE, S., BOUFFARD, O., CHAMBERLAND, A., BETARD, C., SELINGER, E., XHIGNESSE, M., LUSSIER-CACAN, S. & DAVIGNON, J. (1992). Patterns of association between genetic variability in apolipoprotein (apo) B, apo AI-CIII-AIV, and cholesterol ester transfer protein gene regions and quantitative variation in lipid and lipoprotein traits: influence of gender and exogenous hormones. *Am. J. Hum. Genet.* **50**, 92–106.

KESSLING, A. M., HORSTHEMKE, B. & HUMPHRIES, S. E. (1985). A study of DNA polymorphisms around the human apolipoprotein AI gene in hyperlipidaemic and normal individuals. *Clin. Genet.* **28**, 296–306.

KESSLING, A. M., NANJEE, M. N., MILLER, N. E. & HUMPHRIES, S. E. (1988a). Variations in the apolipoprotein AI-CIII-AIV gene region and in lecithin:cholesterol acyltransferase concentration are determinants of plasma cholesterol concentrations. *Atherosclerosis* **70**, 13–19.

KESSLING, A. M., RAJPUT-WILLIAMS, J., BAINTON, D., SCOTT, J., MILLER, N. E., BAKER, I. & HUMPHRIES, S. E. (1988b). DNA polymorphisms of the apolipoprotein AII and AI-CIII-AIV genes: a study in men selected for differences in high-density-lipoprotein cholesterol concentration. *Am. J. Hum. Genet.* **42**, 458–467.

KESSLING, A. M., TAYLOR, R., TEMPLE, A., HUTSON, J., HIDALGO, A. & HUMPHRIES, S. E. (1988c). A PvuII polymorphism in the 5′ flanking region of the apolipoprotein AIV gene: its use to study genetic variation determining serum lipid and apolipoprotein concentration. *Hum. Genet.* **78**, 237–239.

KIM, H.-J. & KALKHOFF, R. K. (1979). Changes in lipoprotein composition during the menstrual cycle. *Metabolism* **28**, 663–668.

KIMURA, M. (1983). *The Neutral Theory of Molecular Evolution.* Cambridge: Cambridge University Press.

KOTTKE, B. A., MOLL, P. P., MICHELS, V. V. & WEIDMAN, W. H. (1991). Levels of lipids, lipoproteins, and apolipoproteins in a defined population. *Mayo Clin. Proc.* **66**, 1198–1208.

LAURIE, C. C., BRIDGHAM, J. T. & CHOUDHARY, M. (1991). Associations between DNA sequence variation and variation in expression of the *Adh* gene in natural populations of *Drosophila melanogaster.* *Genetics* **129**, 489–499.

LITT, M. & JORDE, L. B. (1986). Linkage disequilibria between pairs of loci within a highly polymorphic region of chromosome 2q. *Am. J. Hum. Genet.* **39**, 166–178.

LUSIS, A. J. (1988). Genetic factors affecting blood lipoproteins: the candidate gene approach. *J. Lipid Res.* **29**, 397–429.

MACCLUER, J. W. (1992). Biometrical studies to detect new genes with major effects on quantitative risk factors for atherosclerosis. *Curr. Opin. Lipidol.* **3**, 114–121.

MOLL, P. P. (1993). Biometrical searches for new genes with effects on quantitative risk factors for atherosclerosis: problems and prospects. *Curr. Opin. Lipidol.* **4**, 144–151.

NATIONAL CENTER FOR HEALTH STATISTICS (FULWOOD, R., KALSBEEK, W., RIFKIND, B., RUSSELL-BRIEFEL, R., MUESING, R., LAROSA, J. & LIPPEL, K.) (1986). Total serum cholesterol levels of adults 20–74 years of age: United States, 1976–1980. Washington: US Government Printing Office, 1986. Vital and health statistics. Series II, No. 236. DHHS Pub. No. (PHS) 86-1686. Public Health Service.

NETER, J., WASSERMAN, W. & KUTNER, M. H. (1985). *Applied Linear Statistical Models.* Homewood, Illinois: Irwin.

OETTGEN, P., ANTONARAKIS, S. E. & KARANTHANASIS, S. K. (1986). PvuII polymorphic site upstream to the human ApoCIII gene. *Nucl. Acids Res.* **14**, 5571.

OHNO, S. (1972). So much "junk" DNA in our genome. In *Evolution of Genetic Systems* (ed. H. H. Smith), pp. 366–70. New York: Gordon-Breach.

PAGANI, F., SIDOLI, A., GIUDICI, G. A., VERGANI, C. & BARALLE, F. E. (1989). Association of a polymorphism in the apo AI gene promoter with hyperalphalipoproteinemia. In *High Density Lipoproteins and Atherosclerosis II* (ed. N. E. Miller), pp. 121–128. Netherlands: Elsevier Science Publishers.

PAULWEBER, B., FRIEDL, W., KREMPLER, F., HUMPHRIES, S. E. & SANDHOFER, F. (1988). Genetic variation in the apolipoprotein AI-CIII-AIV gene cluster and coronary heart disease. *Atherosclerosis* **73**, 125–133.

REES, A., STOCKS, J., SHOULDERS, C. C., GALTON, D. J. & BARALLE, F. E. (1983). DNA polymorphism adjacent to human apoprotein A-1 gene: relation to hypertriglyceridaemia. *Lancet* 444–446.

REILLY, S. L., FERRELL, R. E., KOTTKE, B. A., KAMBOH, M. I. & SING, C. F. (1991). The gender-specific apolipoprotein E genotype influence on the distribution of lipids and apolipoproteins in the population of Rochester, MN. I. Pleiotropic effects on means and variances. *Am. J. Hum. Genet.* **49**, 1155–1166.

REILLY, S. L., FERRELL, R. E., KOTTKE, B. A. & SING, C. F. (1992). The gender specific apolipoprotein E genotype influence on the distribution of plasma lipids and apolipoproteins in the population of Rochester, Minnesota. II. Regression relationships with concomitants. *Am. J. Hum. Genet.* **51**, 1311–1324.

REILLY, S. L., FERRELL, R. E. & SING, C. F. (1994). The gender-specific apolipoprotein E genotype influence on the distribution of plasma lipids and apolipoproteins in the population of Rochester, MN. III. Correlations and covariances. *Am. J. Hum. Genet.* **55**, 1001–1018.

REILLY, S. L., KOTTKE, B. A. & SING, C. F. (1990). The effects of generation and gender on the joint distributions of lipid and apolipoprotein phenotypes in the population at large. *J. Clin. Epidemiol.* **43**, 921–940.

ROY, M., SING, C. F., BETARD, C. & DAVIGNON, J. Impact of a common mutation of the LDL receptor gene, in French Canadian patients diagnosed as FH, on means, variances and covariances among traits of lipid metabolism. *Clin. Genet.* (In press).

SAS INSTITUTE INC. (1989). *SAS/STAT@ User's Guide, Version 6.* Cary, NC: SAS Institute Inc.

SCHULL, W. J. & HANIS, C. L. (1990). Genetics and public health in the 1990s. *Ann. Rev. Public Health* **11**, 105–125.

SING, C. F., BOERWINKLE, E. & MOLL, P. P. (1985). Strategies for elucidating the phenotypic and genetic heterogeneity of a chronic disease with a complex etiology. In *Diseases of Complex Etiology in Small Populations: Ethnic Differences and Research Approaches* (ed. R. Chakraborty & E. J. E. Szathmary), pp. 39–66. New York: Alan R. Liss.

SING, C. F., BOERWINKLE, E., MOLL, P. P. & TEMPLETON, A. R. (1988). Characterization of genes affecting quantitative traits in humans. In *Proceedings of the Second International Conference on Quantitative Genetics* (ed. B. S. Weir, E. J. Eisen, M. M. Goodman & G. Namkoong), pp. 250–269. Sunderland, Mass.: Sinauer Associates.

SING, C. F., HAVILAND, M. B., ZERBA, K. E. & TEMPLETON, A. R. (1992a). Application of cladistics to the analysis of genotype-phenotype relationships. *Eur. J. Epidemiol.* **8**, 3–9.

SING, C. F. & MOLL, P. P. (1990). Genetics of atherosclerosis. *Annu. Rev. Genet.* **24**, 171–187.

SING, C. F. & REILLY, S. L. (1993). Genetics of common diseases that aggregate, but do not segregate, in families. In *Genetics of Cellular, Individual, Family, and Population Variability* (ed. C. F. Sing & C. L. Hanis), pp. 140–161. New York: Oxford University Press.

SING, C. F., ZERBA, K. E. & HAVILAND, M. B. (1992b). Genetic architecture of interindividual variation in plasma cholesterol. In *Genetics of Coronary Heart Disease* (ed. A. G. Bearn), pp. 143–162. Oslo: University of Oslo.

SMITH, J. D., BRINTON, E. A. & BRESLOW, J. L. (1992). Polymorphism in the human apolipoprotein A-I gene promoter region. Association of the minor allele with decreased production rate *in vivo* and promoter activity *in vitro. J. Clin. Invest.* **89**, 1796–1800.

SOBER, E. (1983). Parsimony in systematics: philosophical issues. *Ann. Rev. Ecol. Syst.* **14**, 335–357.

SOKAL, R. R. & ROHLF, F. J. (1981). *Biometry.* New York: W. H. Freeman and Company.

STEWART, C.-B. (1993). The powers and pitfalls of parsimony. *Nature, Lond.* **361**, 603–607.

TANKSLEY, S. D. (1993). Mapping polygenes. *Annu. Rev. Genet.* **27**, 205–233.

TEMPLETON, A. R., BOERWINKLE, E. & SING, C. F. (1987). A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and an analysis of alcohol dehydrogenase activity in Drosophila. *Genetics* **117**, 343–351.

TEMPLETON, A. R., CRANDALL, K. A. & SING, C. F. (1992). A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* **132**, 619–633.

TEMPLETON, A. R. & SING, C. F. (1993). A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analyses with cladogram uncertainty and recombination. *Genetics* **134**, 659–669.

TEMPLETON, A. R., SING, C. F., KESSLING, A. & HUMPHRIES, S. (1988). A cladistic analysis of phenotype associations with haplotypes inferred from restriction endonuclease mapping. II. The analysis of natural populations. *Genetics* **120**, 1145–1154.

THE BEZAFIBRATE INFARCTION PREVENTION (BIP) STUDY GROUP ISRAEL. (1992). Lipids and lipoproteins in symptomatic coronary heart disease. Distribution, intercorrelations, and significance for risk classification in 6,700 men and 1,500 women. *Circulation* **86**, 839–848.

TUTEJA, R., TUTEJA, N., MELO, C., CASARI, G. & BARALLE, F. E. (1992). Transcription efficiency of human apolipoprotein A-I promoter varies with naturally occurring A to G transition. *FEBS Letters* **304**, 98–101.

VON ECKARDSTEIN, A., FUNKE, H., SCHULTE, M., ERREN, M., SCHULTE, H. & ASSMANN, G. (1992). Nonsynonymous polymorphic sties in the apolipoprotein (apo) A-IV gene are associated with changes in the concentration of apo B- and apo A-I-containing lipoproteins in a normal population. *Am. J. Hum. Genet.* **50**, 1115–1128.

WEISS, K. M. (1993). Genetic variation and human disease. *Principles and Evolutionary Approaches.* Cambridge: Cambridge University Press.

WILE, D. B., BARBIR, M., GALLAGHER, J., MYANT, N. B., RITCHIE, C. D., THOMPSON, G. R. & HUMPHRIES, S. E. (1989). Apolipoprotein A-I gene polymorphisms: frequency in patients with coronary artery disease and healthy controls and association with serum apo A-I and HDL-cholesterol concentration. *Atherosclerosis* **78**, 9–18.

XHIGNESSE, M., LUSSIER-CACAN, S., SING, C. F., KESSLING, A. M. & DAVIGNON, J. (1991). Influences of common variants of apolipoprotein E on measures of lipid metabolism in a sample selected for health. *Arterioscler. Thromb.* **11**, 1100–1110.

ZERBA, K. E., KESSLING, A. M., DAVIGNON, J. & SING, C. F. (1991). Genetic structure and the search for genotype–phenotype relationships: An example from disequilibrium in the apo B gene region. *Genetics* **129**, 525–533.

ZERBA, K. E. & SING, C. F. (1993). The role of genome type-environment interaction and time in understanding the impact of genetic polymorphisms on lipid metabolism. *Curr. Opin. Lipidol.* **4**, 152–162.

ZUCKERKANDL, E. (1992). Revisiting junk DNA. *J. Mol. Evol.* **34**, 259–271.

APPENDIX I: MEANS AND STANDARD DEVIATIONS FOR CONCOMITANTS AND ADJUSTED PLASMA LIPID, LIPOPROTEIN AND APOLIPOPROTEIN TRAITS BY GENDER AND APO AI-CIII-AIV GENOTYPE

	Genotype 1/6			
	Females (n = 10)		Males (n = 10)	
	Mean	S.D.	Mean	S.D.
Age (years)	37·10	9·28	38·10	4·63
Height (m)	1·60	0·04	1·77	0·05
Weight (kg)	55·95	5·71	73·60	7·93
AT-Chol. (mg/dl)	186·17	21·48	202·24	38·14
ALDL-C (mg/dl)	112·67	20·12	137·70	31·73
AApoB (mg/dl)	101·64	14·20	133·51	25·54
AlnVLDL-C (ln mg/dl)	2·68	0·33	3·06	0·50
AlnTrig. (ln mg/dl)	4·17	0·22	4·51	0·44
AHDL-C (mg/dl)	58·04	10·27	40·66	5·95
AApoAI (mg/dl)	143·75	18·07	123·23	20·25
AlnLp(a) (ln mg/dl)	1·92	0·99	2·28	1·14

APPENDIX I. (*Cont.*)

Genotype 3/6

	Females (*n* = 1)		Males (*n* = 4)	
	Mean	S.D.	Mean	S.D.
Age (years)	32·00	—	39·25	7·68
Height (m)	1·68	—	1·74	0·06
Weight (kg)	61·70	—	79·68	7·62
AT-Chol. (mg/dl)	156·88	—	192·98	36·98
ALDL-C (mg/dl)	96·80	—	117·51	20·96
AApoB (mg/dl)	95·69	—	121·66	36·94
AlnVLDL-C (ln mg/dl)	2·64	—	3·40	0·42
AlnTrig. (ln mg/dl)	3·70	—	4·54	0·66
AHDL-C (mg/dl)	45·90	—	40·52	6·36
AApoAI (mg/dl)	129·88	—	142·05	30·54
AlnLp(a) (ln mg/dl)	1·12	—	0·67	0·64

Genotype 5/6

	Females (*n* = 2)		Males (*n* = 7)	
	Mean	S.D.	Mean	S.D.
Age (years)	31·50	6·36	34·86	6·01
Height (m)	1·68	0·08	1·71	0·06
Weight (kg)	61·85	3·89	69·04	6·18
AT-Chol. (mg/dl)	183·46	6·65	196·27	35·30
ALDL-C (mg/dl)	113·79	1·37	132·84	29·63
AApoB (mg/dl)	100·03	3·54	124·96	29·06
AlnVLDL-C (ln mg/dl)	3·06	0·29	3·04	0·30
AlnTrig. (ln mg/dl)	4·28	0·08	4·40	0·44
AHDL-C (mg/dl)	46·16	1·88	41·37	4·58
AApoAI (mg/dl)	112·25	24·09	119·43	10·59
AlnLp(a) (ln mg/dl)	1·91	1·13	1·92	1·51

Genotype 6/6

	Females (*n* = 25)		Males (*n* = 49)	
	Mean	S.D.	Mean	S.D.
Age (years)	35·76	6·42	39·51	8·12
Height (m)	1·62	0·05	1·72	0·06
Weight (kg)	58·83	6·25	72·59	8·72
AT-Chol. (mg/dl)	180·13	25·19	186·40	27·78
ALDL-C (mg/dl)	113·98	20·62	120·76	24·23
AApoB (mg/dl)	104·74	16·56	115·44	25·76
AlnVLDL-C (ln mg/dl)	2·82	0·38	3·10	0·50
AlnTrig. (ln mg/dl)	4·21	0·30	4·46	0·47
AHDL-C (mg/dl)	48·17	8·98	41·16	9·04
AApoAI (mg/dl)	137·48	17·80	128·17	19·37
AlnLp(a) (ln mg/dl)	1·78	1·12	2·06	1·29

APPENDIX I. (*Cont.*)

Genotype 7/6

	Females (n = 6)		Males (n = 6)	
	Mean	S.D.	Mean	S.D.
Age (years)	37·17	5·12	35·83	9·52
Height (m)	1·62	0·04	1·71	0·04
Weight (kg)	56·53	6·22	70·88	8·56
AT-Chol. (mg/dl)	167·74	19·77	209·19	33·60
ALDL-C (mg/dl)	110·40	23·58	143·12	32·71
AApoB (mg/dl)	94·82	19·08	139·28	34·14
AlnVLDL-C (ln mg/dl)	2·67	0·24	3·23	0·57
AlnTrig. (ln mg/dl)	4·03	0·21	4·55	0·61
AHDL-C (mg/dl)	42·61	9·39	37·99	9·77
AApoAI (mg/dl)	119·52	14·02	120·59	19·03
AlnLp(a) (ln mg/dl)	2·30	1·25	2·90	1·16

Genotype 8/6

	Females (n = 4)		Males (n = 10)	
	Mean	S.D.	Mean	S.D.
Age (years)	38·50	6·56	43·50	7·29
Height (m)	1·59	0·05	1·74	0·06
Weight (kg)	57·73	4·02	72·82	11·06
AT-Chol. (mg/dl)	179·68	24·46	187·78	12·64
ALDL-C (mg/dl)	113·67	28·35	124·27	17·90
AApoB (mg/dl)	100·98	24·30	120·41	22·01
AlnVLDL-C (ln mg/dl)	2·96	0·32	3·12	0·26
AlnTrig. (ln mg/dl)	4·33	0·26	4·55	0·31
AHDL-C (mg/dl)	46·32	9·83	39·28	9·04
AApoAI (mg/dl)	133·00	25·81	121·57	14·46
AlnLp(a) (ln mg/dl)	1·88	1·40	1·56	1·44

Genotype 10/6

	Females (n = 2)		Males (n = 4)	
	Mean	S.D.	Mean	S.D.
Age (years)	36·00	0·00	44·50	9·95
Height (m)	1·58	0·02	1·75	0·07
Weight (kg)	54·75	7·00	73·75	9·12
AT-Chol. (mg/dl)	184·97	24·73	224·34	44·21
ALDL-C (mg/dl)	112·66	36·29	156·53	33·72
AApoB (mg/dl)	108·24	28·29	152·16	24·13
AlnVLDL-C (ln mg/dl)	2·64	0·65	3·34	0·11
AlnTrig. (ln mg/dl)	4·11	0·05	4·75	0·33
AHDL-C (mg/dl)	57·64	2·15	37·85	7·56
AApoAI (mg/dl)	153·50	10·53	129·68	12·65
AlnLp(a) (ln mg/dl)	0·38	0·91	2·21	1·39

M. B. Haviland and others

APPENDIX II: SUMMARY OF CLADISTIC ANALYSIS OF APO AI-CIII-AIV GENE REGION

T-Chol. in females

Source	D.F.	SS	(%) SS	F	p-value
Total	49	37749	100·0	—	—
Concomitants	11	12078	32·0	1·48	0·19
Genotypes	6	1886	5·0	0·42	0·86
6/6 v. 1/6	1	260	0·7	0·35	0·56
6/6 v. 3/6	1	520	1·4	0·70	0·41
6/6 v. 5/6	1	20	0·1	0·03	0·86
6/6 v. 7/6	1	743	2·0	1·00	0·32
6/6 v. 8/6	1	1	0·0	0·00	0·97
6/6 v. 10/6	1	43	0·1	0·06	0·81
Unexplained	32	23785	63·0	—	—

LDL-C in females

Source	D.F.	SS	(%) SS	F	p-value
Total	49	30512	100·0	—	—
Concomitants	11	9828	32·2	1·40	0·22
Genotypes	6	330	1·1	0·09	0·99
6/6 v. 1/6	1	12	0·0	0·02	0·89
6/6 v. 3/6	1	284	0·9	0·45	0·51
6/6 v. 5/6	1	0	0·0	0·00	0·97
6/6 v. 7/6	1	62	0·2	0·10	0·75
6/6 v. 8/6	1	0	0·0	0·00	0·99
6/6 v. 10/6	1	3	0·0	0·01	0·94
Unexplained	32	20354	66·7	—	—

Apo B in females

Source	D.F.	SS	(%) SS	F	p-value
Total	49	21125	100·0	—	—
Concomitants	11	7703	36·5	1·75	0·11
Genotypes	6	620	2·9	0·26	0·95
6/6 v. 1/6	1	69	0·3	0·17	0·68
6/6 v. 3/6	1	79	0·4	0·20	0·66
6/6 v. 5/6	1	41	0·2	0·10	0·75
6/6 v. 7/6	1	476	2·2	1·19	0·28
6/6 v. 8/6	1	49	0·2	0·12	0·73
6/6 v. 10/6	1	23	0·1	0·06	0·81
Unexplained	32	12802	60·6	—	—

ln VLDL-C in females

Source	D.F.	SS	(%) SS	F	p-value
Total	49	6·83	100·0	—	—
Concomitants	11	0·84	12·3	0·45	0·92
Genotypes	6	0·55	8·1	0·54	0·77
6/6 v. 1/6	1	0·13	1·8	0·76	0·39
6/6 v. 3/6	1	0·03	0·4	0·18	0·67
6/6 v. 5/6	1	0·11	1·6	0·65	0·43
6/6 v. 7/6	1	0·11	1·6	0·65	0·43
6/6 v. 8/6	1	0·07	1·0	0·41	0·53
6/6 v. 10/6	1	0·06	0·9	0·35	0·56
Unexplained	32	5·44	79·6	—	—

APPENDIX II (*Cont.*)

ln Trig. in females

Source	D.F.	SS	(%) SS	F	*p*-value
Total	49	6·36	100·0	—	—
Concomitants	11	2·79	43·9	2·65	0·02
Genotypes	6	0·51	8·0	0·89	0·51
6/6 *v.* 1/6	1	0·01	0·2	0·10	0·75
6/6 *v.* 3/6	1	0·25	3·9	2·61	0·12
6/6 *v.* 5/6	1	0·01	0·2	0·10	0·75
6/6 *v.* 7/6	1	0·16	2·5	1·67	0·21
6/6 *v.* 8/6	1	0·05	0·8	0·52	0·48
6/6 *v.* 10/6	1	0·02	0·3	0·21	0·65
Unexplained	32	3·06	48·1	—	—

HDL-C in females

Source	D.F.	SS	(%) SS	F	*p*-value
Total	49	6018	100·0	—	—
Concomitants	11	1128	18·8	0·91	0·54
Genotypes	6	1266	21·0	1·86	0·12
6/6 *v.* 1/6	1	696	11·6	6·14	0·02
6/6 *v.* 3/6	1	5	0·1	0·04	0·84
6/6 *v.* 5/6	1	7	0·1	0·06	0·81
6/6 *v.* 7/6	1	149	2·5	1·32	0·26
6/6 *v.* 8/6	1	12	0·2	0·11	0·74
6/6 *v.* 10/6	1	166	2·8	1·47	0·23
Unexplained	32	3625	60·2	—	—

Apo AI in females

Source	D.F.	SS	(%) SS	F	*p*-value
Total	49	21487	100·0	—	—
Concomitants	11	3174	14·8	0·65	0·77
Genotypes	6	4096	19·1	1·54	0·20
6/6 *v.* 1/6	1	281	1·3	0·63	0·43
6/6 *v.* 3/6	1	55	0·2	0·12	0·73
6/6 *v.* 5/6	1	1179	5·5	2·65	0·11
6/6 *v.* 7/6	1	1561	7·3	3·51	0·07
6/6 *v.* 8/6	1	69	0·3	0·16	0·69
6/6 *v.* 10/6	1	476	2·2	1·07	0·31
Unexplained	32	14217	66·1	—	—

ln Lp(a) in females

Source	D.F.	SS	(%) SS	F	*p*-value
Total	49	73·62	100·0	—	—
Concomitants	11	12·36	16·8	0·65	0·77
Genotypes	6	6·19	8·4	0·60	0·73
6/6 *v.* 1/6	1	0·15	0·2	0·09	0·77
6/6 *v.* 3/6	1	0·41	1·0	0·24	0·63
6/6 *v.* 5/6	1	0·03	0·0	0·02	0·89
6/6 *v.* 7/6	1	1·35	1·8	0·78	0·38
6/6 *v.* 8/6	1	0·04	0·0	0·02	0·89
6/6 *v.* 10/6	1	3·59	4·9	2·09	0·16
Unexplained	32	55·07	74·8	—	—

APPENDIX II (*Cont.*)

Total Chol. in males

Source	D.F.	SS	(%) SS	F	*p*-value
Total	89	112418	100·0	—	—
Concomitants	10	28941	25·7	2·83	0·005
Genotypes	6	8822	7·9	1·44	0·21
6/6 *v.* 1/6	1	2083	1·9	2·04	0·16
6/6 *v.* 3/6	1	160	0·1	0·16	0·69
6/6 *v.* 5/6	1	596	0·5	0·58	0·45
6/6 *v.* 7/6	1	2778	2·5	2·72	0·10
6/6 *v.* 8/6	1	16	0·0	0·02	0·89
6/6 *v.* 10/6	1	5323	4·7	5·20	0·03
Unexplained	73	74655	66·4	—	—

LDL-C in males

Source	D.F.	SS	(%) SS	F	*p*-value
Total	89	86179	100·0	—	—
Concomitants	10	21947	25·5	2·89	0·004
Genotypes	6	8771	10·2	1·92	0·09
6/6 *v.* 1/6	1	2383	2·8	3·14	0·08
6/6 *v.* 3/6	1	39	0·0	0·05	0·82
6/6 *v.* 5/6	1	893	1·0	1·18	0·28
6/6 *v.* 7/6	1	2671	3·1	3·52	0·06
6/6 *v.* 8/6	1	102	0·1	0·13	0·72
6/6 *v.* 10/6	1	4730	5·5	6·23	0·01
Unexplained	73	55461	64·3	—	—

Apo B in males

Source	D.F.	SS	(%) SS	F	*p*-value
Total	89	97407	100·0	—	—
Concomitants	10	29635	30·4	3·68	0·0005
Genotypes	6	8944	9·2	1·85	0·10
6/6 *v.* 1/6	1	2711	2·8	3·36	0·07
6/6 *v.* 3/6	1	143	0·1	0·18	0·67
6/6 *v.* 5/6	1	555	0·6	0·69	0·41
6/6 *v.* 7/6	1	3036	3·1	3·77	0·06
6/6 *v.* 8/6	1	205	0·2	0·25	0·62
6/6 *v.* 10/6	1	4987	5·1	6·19	0·02
Unexplained	73	58828	60·4	—	—

ln VLDL-C in males

Source	D.F.	SS	(%) SS	F	*p*-value
Total	89	20·91	100·0	—	—
Concomitants	10	2·86	13·7	1·20	0·31
Genotypes	6	0·68	3·2	0·48	0·82
6/6 *v.* 1/6	1	0·01	0·0	0·04	0·84
6/6 *v.* 3/6	1	0·34	1·6	1·43	0·24
6/6 *v.* 5/6	1	0·02	0·1	0·08	0·78
6/6 *v.* 7/6	1	0·10	0·5	0·42	0·52
6/6 *v.* 8/6	1	0·00	0·0	0·00	0·99
6/6 *v.* 10/6	1	0·22	1·0	0·92	0·34
Unexplained	73	17·37	83·1	—	—

APPENDIX II (*Cont.*)

ln Trig. in males

Source	D.F.	SS	(%) SS	F	*p*-value
Total	89	21·88	100·0	—	—
Concomitants	10	3·40	15·5	1·38	0·21
Genotypes	6	0·44	2·0	0·30	0·93
6/6 *v.* 1/6	1	0·02	0·1	0·08	0·78
6/6 *v.* 3/6	1	0·02	0·1	0·08	0·78
6/6 *v.* 5/6	1	0·02	0·1	0·08	0·78
6/6 *v.* 7/6	1	0·04	0·2	0·16	0·69
6/6 *v.* 8/6	1	0·07	0·3	0·28	0·60
6/6 *v.* 10/6	1	0·31	1·4	1·25	0·27
Unexplained	73	18·04	82·5	—	—

HDL-C in males

Source	D.F.	SS	(%) SS	F	*p*-value
Total	89	7138	100·0	—	—
Concomitants	10	1158	16·2	1·44	0·18
Genotypes	6	107	1·5	0·22	0·97
6/6 *v.* 1/6	1	2	0·0	0·02	0·89
6/6 *v.* 3/6	1	2	0·0	0·02	0·89
6/6 *v.* 5/6	1	0	0·0	0·00	0·99
6/6 *v.* 7/6	1	54	0·7	0·67	0·42
6/6 *v.* 8/6	1	29	0·4	0·36	0·55
6/6 *v.* 10/6	1	40	0·6	0·50	0·48
Unexplained	73	5873	82·3	—	—

Apo AI in males

Source	D.F.	SS	(%) SS	F	*p*-value
Total	89	39213	100·0	—	—
Concomitants	10	7806	19·9	1·94	0·05
Genotypes	6	2053	5·2	0·85	0·54
6/6 *v.* 1/6	1	203	0·5	0·50	0·48
6/6 *v.* 3/6	1	711	1·8	1·77	0·19
6/6 *v.* 5/6	1	468	1·2	1·16	0·29
6/6 *v.* 7/6	1	308	0·8	0·77	0·38
6/6 *v.* 8/6	1	363	0·9	0·90	0·35
6/6 *v.* 10/6	1	8	0·0	0·02	0·89
Unexplained	73	29354	74·9	—	—

ln Lp(a) in males

Source	D.F.	SS	(%) SS	F	*p*-value
Total	89	177·73	100·0	—	—
Concomitants	10	24·97	14·1	1·32	0·24
Genotypes	6	15·00	8·4	1·32	0·26
6/6 *v.* 1/6	1	0·39	0·2	0·21	0·65
6/6 *v.* 3/6	1	7·14	4·0	3·78	0·06
6/6 *v.* 5/6	1	0·12	0·1	0·06	0·81
6/6 *v.* 7/6	1	3·74	2·1	1·98	0·16
6/6 *v.* 8/6	1	2·12	1·2	1·12	0·29
6/6 *v.* 10/6	1	0·08	0·0	0·04	0·84
Unexplained	73	137·76	77·5	—	—