

ORIGINAL INVESTIGATION

M. B. Haviland · R. E. Ferrell · C. F. Sing

Association between common alleles of the low-density lipoprotein receptor gene region and interindividual variation in plasma lipid and apolipoprotein levels in a population-based sample from Rochester, Minnesota

Received: 19 March 1996 / Revised: 30 July 1996

Abstract This paper presents an analysis of the relationship between variation in the low-density lipoprotein receptor (LDLR) gene region and interindividual variation in plasma lipid and apolipoprotein levels in a sample representative of the adult population of Rochester, Minn. (217 females and 187 males aged 26 to 63). This relationship was analyzed by estimating the average excesses of alleles of the LDLR gene defined using RFLP markers both singly and simultaneously. We also used a cladistic approach to illustrate the consequences of incorporating evolutionary information into the analysis of genotype-phenotype relationships. Although results from both approaches supported the inference that common variation in the LDLR gene region associates with small effects on plasma lipid and apolipoprotein levels, only the cladistic approach provides direction for further work aimed at identifying the functional DNA sequence variations responsible for the observed associations.

Introduction

One of the aims of human genetics is to identify the DNA sequence variations responsible for quantitative phenotypic variation in a particular trait. To date, most work has concentrated on identifying rare DNA sequence variations with large phenotypic effects in individuals with specific diseases or hyperphenotypes. Searches for such rare variations have been very successful (e.g., Hobbs et al. 1990; Riordan et al. 1989; Wallace et al. 1990). Once the region of the genome containing the gene has been localized, finding the DNA sequence responsible for an extreme

phenotype is straightforward, although in practice it can be hampered by technical problems and be quite labor intensive. The responsible sequence variations are likely to be identifiable because there is a Cartesian or one-to-one relationship between genotype and phenotype. Unfortunately, this strategy has not been effective in the identification of DNA sequence variations responsible for phenotypic variation within the normal range (Sing et al. 1992a, b; Sing and Reilly 1993; Strohmaan 1994).

Identifying the functional DNA sequence variations responsible for variation in traits that are continuously distributed among individuals in the population at large is complicated by the complex genetic architecture underlying the observed trait distribution (Sing et al. 1992c). The genetic architecture of a quantitative trait involves multiple genes with continuously distributed sizes of allelic effects that interact with each other and with variation in the environment to determine trait variation (Sing and Boerwinkle 1985; Sing and Moll 1989; Sing et al. 1992c, 1996; Tanksley 1993; Zerba and Sing 1993). It follows that the causal relationship between a particular genotype and a particular phenotypic value cannot be one-to-one.

We have taken the candidate gene approach to identify genes likely to contain DNA sequence variations that influence quantitative trait variation (Lusis 1988; Sing et al. 1992c). This approach entails the study of genes whose products are involved in the metabolism of the quantitative trait of interest. Once a candidate gene has been selected, the search for functional DNA sequence variations has traditionally begun with single marker association studies aimed at testing the null hypothesis that variation in the gene region does not associate with interindividual trait variation. When available, multiple markers can be used to further resolve the genotype-phenotype relationships (e.g., Aquadro et al. 1992; Kessling et al. 1988). Although such studies have proved useful in confirming the involvement of a candidate gene in influencing variation in a quantitative trait (Kessling et al. 1992), they offer little information to help guide a search for functional DNA sequence variations. The search for functional DNA sequence variations involves first characterizing the varia-

M. B. Haviland (✉) · C. F. Sing
Department of Human Genetics, M4708 Medical Science Bldg. II,
University of Michigan Medical School, Ann Arbor,
MI 48109-0618, USA

R. E. Ferrell
Department of Human Genetics, University of Pittsburgh,
Pittsburgh, PA 15261, USA

tion in a gene region, second identifying those individuals that carry functional variations, and last determining which of the DNA variations are responsible for the functional effects.

Recently, a cladistic approach for characterizing the association between haplotype variation in a candidate gene and phenotype variation has been developed. This analytical strategy organizes complex genic information resulting in increased statistical power to identify individuals carrying haplotypes with functional sequence variations with relatively small effects on trait levels (Haviland 1993; Haviland et al. 1995; Sing et al. 1992b; Templeton et al. 1987, 1988, 1992; Templeton and Sing 1993). A second advantage of the cladistic approach over traditional approaches that do not consider evolutionary information is the identification of those individuals that should be compared in subsequent molecular analyses to exclude background nonfunctional variations and hence maximize the probability of finding functional DNA sequence variations.

This paper presents an analysis of the relationship between common variation in the low-density lipoprotein receptor (LDLR) gene region and interindividual variation in plasma lipid and apolipoprotein levels. Although results from both the cladistic analysis and an analysis ignoring evolutionary information supported the inference that common variations in the LDLR gene region are associated with small effects on plasma lipid and apolipoprotein levels, only the cladistic approach provides direction for further work aimed at identifying the functional DNA sequence variations responsible for the observed associations.

Materials and methods

Sample

The first phase of the Rochester Family Heart Study includes 2002 individuals in 276 multigenerational pedigrees representative of the Caucasian population of Rochester, Minnesota (Moll et al. 1989; Turner et al. 1989). We typed 426 unrelated individuals from the parental generation of this sample for four RFLP markers in the LDLR gene region; unambiguous haplotypes could be determined for 409 of them (genotyping and haplotyping is described below). Five individuals were removed because their plasma levels for one or more of the nine lipid and apolipoprotein traits (apo AI, AII, B, CII, CIII and E, total cholesterol, HDL-C and triglycerides) were greater than, or less than, 5 standard deviations from the gender-specific mean. None of the individuals in this sample was recorded as having a clinical diagnosis of familial hypercholesterolemia. Thus, the sample analyzed here consisted of 217 females (average age of 41.7 years, range 26.0 to 62.9 years) and 187 males (average age of 43.5 years, range 27.2 to 61.5 years).

Laboratory methods

All blood samples were collected in EDTA by venipuncture. Plasma levels of apo AI, AII, B, CII, CIII and E were measured using radioimmunoassays (Kottke et al. 1991). Total plasma cholesterol and triglyceride levels were measured using standard enzymatic methods, and plasma HDL-C was measured following the precipitation of apo B-containing lipoproteins (Izzo et al. 1981). Lipid and apolipoprotein values are presented here in mg/dl.

Four RFLP markers in the LDLR gene region were typed by DNA amplification using primers flanking the *TaqI* (intron 4), *StuI* (exon 8), *HincII* (exon 12) and *AvaII* (exon 13) variable restriction sites followed by digestion of the amplification product with the appropriate restriction endonuclease and electrophoresis in 1% agarose gel (Leitersdorf and Hobbs 1988; Leitersdorf et al. 1989). Grandparents and children of the parents studied here were also typed for the purposes of error checking and haplotyping. The computer program PATCH (Wijsman 1987) was used to search for segregation inconsistencies in the pedigrees. All members in a pedigree, detected by PATCH as having an inconsistency, were re-typed to resolve typing errors. Two pedigrees were removed from the analysis because pedigree inconsistencies could not be resolved. PATCH was then used to construct haplotypes based on the four RFLP markers.

Analytical methods

All analyses were done using the SAS statistical package (SAS Institute 1989). The means, variances and ranges of the concomitants (age, height and weight) and plasma lipid and apolipoprotein traits (apo AI, AII, B, CII, CIII and E, total cholesterol, HDL-C and triglycerides) in females and males were estimated. Homogeneity of means between females and males was tested using the *t*-test (the use of the nonparametric Wilcoxon sum rank test did not alter the inferences, results not shown). Homogeneity of variances between the genders was tested using the *F*-test. The *F*-test is sensitive to deviations from normality, so the results of the nonparametric Ansari-Bradley test of homogeneity of dispersion with a modification for unequal locations (Randles and Wolfe 1979) are also presented. The levels of the nine plasma lipid and apolipoprotein traits were then adjusted by multiple regression for date of assay, age, age², age³, height, height², height³, weight, weight² and weight³ within each gender.

The association between variation in the LDLR gene region and adjusted phenotypic variation within each gender was then analyzed. Allelic effects, where alleles were defined using each RFLP marker singly and all four RFLP markers simultaneously, were estimated by the average excess statistic (Templeton 1987). Because the theoretical sampling distribution of this statistic is unknown, we used permutation testing (Edgington 1987) to test the null hypothesis that an average excess was equal to zero. To estimate the null distribution of the average excess statistic, the phenotypic values were randomly permuted over the genotypes using a random number generator (Wichmann and Hill 1982) and the average excess calculated using the permuted data. One thousand permutations of the data were made. The relative frequency of the 1000 permuted average excesses that exceeded the magnitude of the average excess computed for the original data set was taken as the estimate of the probability of observing an average excess equal to or greater than the original average excess under the null hypothesis.

The cladistic approach has been described fully elsewhere (Haviland 1993; Haviland et al. 1995; Sing et al. 1992b; Templeton et al. 1987, 1988, 1992; Templeton and Sing 1993) and is only briefly summarized here. A cladistic analysis involves two steps. In the first step an unrooted cladogram is built by organizing the haplotypes, defined by the four RFLP markers, such that those hypothesized to be close in evolutionary history are connected. The method of maximum parsimony was used here (Felsenstein 1983; Sober 1983; Stewart 1993). Rare haplotypes, haplotypes with a relative frequency less than or equal to 1% in either gender, were not included in the cladogram. Coalescent theory has been used to demonstrate that rare haplotypes are likely to be connected to only one other haplotype (Crandall and Templeton 1993) so the removal of the rare haplotypes should not affect the structure of the rest of the cladogram. A scheme for testing hypotheses about the association between haplotype and phenotype variation is defined by nesting (i.e., grouping) the haplotypes according to the cladogram structure in the second step of the cladistic analysis (Templeton et al. 1987).

The LDLR cladogram defined pairwise contrasts between haplotypes and between groups of haplotypes. Thus, rather than using the nested sum-of-squares statistics derived by Templeton et al. (1988), the test statistic, $a_i - a_j$, where a_i and a_j are the average excesses for the i^{th} and j^{th} haplotypes, or groups of haplotypes, respectively, was calculated. To estimate the distribution of the test statistic under the null hypothesis, permutation testing was used (see above).

When choosing a statistical significance level, both the type-I and type-II error rates must be considered. Because the size of the association between haplotype and trait variation was expected to be small and this is the first such study of this type, a liberal test criteria of 0.10 was used to minimize type-II errors.

Table 1 Haplotype frequencies by gender

	Haplotype Designation ^a				Haplotype Frequency (%)	
	<i>TaqI</i>	<i>StuI</i>	<i>HincII</i>	<i>AvaII</i>	Females ^b	Males
H1	+	+	+	-	41 (9.5)	39 (10.4)
H2	-	+	-	+	195 (44.9)	146 (38.8)
H3	-	+	+	+	7 (1.6)	2 (0.5)
H4	-	+	-	-	1 (0.2)	0 (0.0)
H5	-	+	+	-	176 (40.6)	166 (44.7)
H6	-	-	+	-	14 (3.2)	21 (5.6)

^a + denotes presence of restriction cut site, - denotes absence of restriction cut site

^b RFLP allele and haplotype frequencies were not significantly different between females and males (results not shown)

Results

The means, standard deviations and ranges for age, height, weight and the nine plasma lipid and apolipoprotein traits for females and males separately are given in the Appendix. The significant gender differences in most of the concomitant and plasma lipid trait distributions justified separate treatment of females and males. When the four RFLP markers were considered simultaneously, 6 unambiguous haplotypes, out of a possible 16, were observed. These haplotypes were labeled H1-H6. The haplotype frequencies are given in Table 1. Four haplotypes, H1, H2, H5 and H6, accounted for more than 98% of the haplotypes.

The average excesses for alleles defined by each RFLP marker singly are given in Table 2. The null hypothesis, average excess equal to zero, could not be rejected for alleles defined by any of the RFLP markers for any of the adjusted lipid or apolipoprotein traits in females (see Table 2). In males, the average excesses associated with the *TaqI* alleles differed significantly from zero for adjusted apo B and apo CII, and the average excesses associated with the *StuI* alleles differed significantly from zero for adjusted HDL-C.

The average excesses for alleles defined using all four RFLP markers simultaneously are given in Table 3. The average excesses associated with haplotypes H3 and H4

Table 2 Average excess for the allele of each RFLP with the restriction site present (+) for each adjusted plasma lipid and apolipoprotein by gender

	Females				Males			
	<i>TaqI</i> (fr(+)=0.09)	<i>StuI</i> (fr(+)=0.97)	<i>HincII</i> (fr(+)=0.55)	<i>AvaII</i> (fr(+)=0.47)	<i>TaqI</i> (fr(+)=0.10)	<i>StuI</i> (fr(+)=0.94)	<i>HincII</i> (fr(+)=0.61)	<i>AvaII</i> (fr(+)=0.40)
Apo AI	-1.64	-0.18	-0.43	0.30	-1.19	-0.26	0.12	-0.14
Apo AII	-0.28	-0.01	-0.15	0.16	-0.30	-0.05	0.01	-0.03
Apo B	-0.12	-0.01	-0.15	0.32	4.73**	-0.05	0.84	-1.24
Apo CII	0.02	0.01	0.03	-0.03	0.30**	-0.00	0.02	-0.02
Apo CIII	-0.60	-0.01	-0.15	0.15	1.20	-0.00	0.16	-0.21
Apo E	-0.26	0.01	0.03	-0.04	0.29	0.01	-0.14	0.26
T-Cholesterol	-1.04	-0.11	-1.60	1.42	9.01	-0.07	0.74	-1.19
HDL-C	-0.47	-0.11	-0.51	0.27	-1.87	-0.18*	-0.38	0.53
Triglycerides	-9.46	0.10	1.16	-1.14	20.32	0.96	0.20	1.44

* Null hypothesis of average excess equal to zero rejected at a significance level of $0.05 < P < 0.10$

** Null hypothesis of average excess equal to zero rejected at a significance level of $P < 0.05$

Table 3 Average excess for each adjusted lipid and apolipoprotein by haplotype and gender

	Females				Males			
	H1	H2	H5	H6	H1	H2	H5	H6
Apo AI	-1.64	0.49	-0.45	5.44	-1.19	-0.18	-0.15	4.39
Apo AII	-0.28	0.17	-0.16	0.37	-0.30	-0.01	-0.01	0.84
Apo B	-0.12	0.21	-0.32	0.23	4.73**	-1.31	-0.11	0.79
Apo CII	0.02	-0.03	0.05	-0.15	0.30**	-0.03	-0.05	0.01
Apo CIII	-0.60	0.18	-0.05	0.21	1.20	-0.24	-0.10	0.02
Apo E	-0.26	-0.02	0.15	-0.40	0.29	0.22	-0.28*	-0.14
T-Cholesterol	-1.04	1.99	-1.61	3.26	9.01	-1.15	-1.21	1.21
HDL-C	-0.47	0.58	-0.50	3.21	-1.87	0.60	-0.42	3.07*
Triglycerides	-9.46	-1.31	3.88	-3.12	20.32	-0.31	-4.01	-16.22

* Null hypothesis of average excess equal to zero rejected at a significance level of $0.05 < P < 0.10$

** Null hypothesis of average excess equal to zero rejected at a significance level of $P < 0.05$

were not estimated because of small sample size. Defining the alleles using all four RFLP markers simultaneously revealed information about functional DNA sequence variations that was similar to that found using each marker singly. In fact, haplotype H1 is equivalent to the *TaqI* allele with the restriction site present and haplotype H6 is equivalent to the *StuI* allele with the restriction site absent, so the average excesses for these haplotypes were exactly what was estimated above. As seen before, none of the average excesses associated with any of the haplotypes differed significantly from zero for any of the traits in females. In males, as with alleles defined using the *TaqI* RFLP, haplotype H1 was associated with significant average excesses for adjusted apo B and apo CII. As with alleles defined using the *StuI* RFLP, haplotype H6 was associated with a significant average excess for adjusted HDL-C. One additional association was revealed using the RFLP loci simultaneously that was not revealed by the single RFLP association study: haplotype H5 associated with a significant average excess for adjusted apo E.

The LDLR cladogram estimated in the first step of the cladistic analysis is depicted in Fig. 1. The cladogram, nested into one-step clades C1 (composed of H1, H5, and H6) and C2 (composed of H2), was then used to design the analysis of phenotypic variation in the second step of the cladistic analysis. We first contrasted the average excesses associated with the one-step clades C1 and C2. None of the average excesses associated with C1 and C2 were significantly different in females or males for any of the plasma lipid or apolipoprotein traits (see Table 4). The

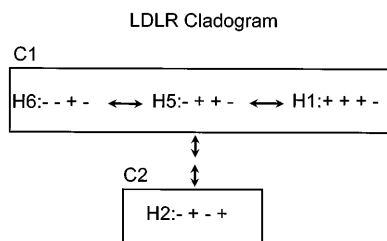


Fig. 1 Maximum parsimony cladogram of the LDLR haplotypes. The RFLP markers are listed 5' to 3' and + and - denote cut and uncut, respectively. The *arrows* represent the number of mutational steps between haplotypes and haplotype groups or clades. The one-step clades, C1 and C2, are designated by *boxes*

Table 4 LDLR cladistic analysis of adjusted plasma lipid and apolipoprotein levels by gender (Relative frequency of 1000 permuted test statistics that exceeded the magnitude of the test statistic computed for the original data set.)

	Females			Males		
	C1 vs C2	H5 vs H1	H5 vs H6	C1 vs C2	H5 vs H1	H5 vs H6
Apo AI	0.70	0.67	0.20	0.95	0.71	0.29
Apo AII	0.43	0.72	0.65	0.97	0.71	0.47
Apo B	0.71	0.89	0.87	0.17	0.05	0.75
Apo CII	0.40	0.91	0.36	0.51	0.02	0.66
Apo CIII	0.32	0.49	0.75	0.45	0.12	0.90
Apo E	0.70	0.31	0.28	0.32	0.26	0.77
T-Cholesterol	0.23	0.94	0.58	0.56	0.12	0.67
HDL-C	0.51	0.97	0.26	0.29	0.32	0.07
Triglycerides	0.75	0.21	0.69	0.93	0.12	0.54

haplotypes within the one-step clade C1 were compared next and the results are given in Table 4. As with the analyses described above ignoring evolutionary information, there were no statistically significant associations in females. In males, haplotype H1 was associated with significantly elevated apo B and apo CII as compared to H5. Haplotype H6, when compared to H5, was associated with an elevation in adjusted HDL-C. The size of all three of these effects was approximately one-third of a standard deviation of the mean.

Discussion

Comparison of the cladistic analysis to analyses that did not incorporate evolutionary information

The first step in the search for functional DNA sequence variations is to establish that a statistical relationship exists between variation in a candidate gene region and phenotypic variation. In this paper, both the cladistic analysis and an analysis that did not incorporate evolutionary information established that variations in the LDLR gene region significantly associated with variation in plasma lipid and apolipoprotein traits, supporting the hypothesis that the LDLR gene region contains functional variations that influence phenotypic variation in the normal range. Of course, this hypothesis would be even stronger if our results were replicated in an independent sample. We fully expect that others will replicate the finding that variation in the LDLR gene region significantly associates with normal variation in plasma lipid and apolipoprotein traits. But, given that we expect most effects to be context-dependent (Merry 1995; Sing et al. 1996), and that it will be difficult if not impossible to duplicate all of the genetic and environmental contexts that may be influencing these effects in our sample, replication of the average allele effects found in our study may be an unrealistic expectation.

Our cladistic analysis suggests that the LDLR gene region contains at least two common functional variations with small effects (less than one-half of 1 standard deviation of the mean) that are carried on two separate haplotypes. In general, an analysis that ignores evolutionary information will not allow one to make inferences about how many functional variations are present in a gene region. Without the evolutionary information, it is unclear

whether the *TaqI* and *StuI* RFLP markers are in linkage disequilibria with the same or different functional variations. In contrast, a cladistic analysis, using evolutionary information to construct a testing design that is independent of phenotypic information, efficiently uses the available degrees of freedom to simultaneously test for multiple functional variations. If different haplotypes in discrete regions of the cladogram are found to associate with phenotypic variation, it is likely that the haplotypes are carrying different functional variations. It is unlikely that the same mutation occurred more than once in the evolutionary history of the gene region. So, because haplotypes H1 and H6 were in discrete regions of the LDLR cladogram and were associated with effects on different plasma lipid or apolipoprotein traits, we hypothesize that there are different functional sequence variation(s) in each haplotype.

The second step in identifying variations with a high probability of being functional involves identifying individuals hypothesized to differ for the functional variations. Results from the single RFLP and multiple RFLP association studies suggest that there are functional DNA sequence variations present in the LDLR gene region, but they give little information for directing studies aimed at identifying individuals who differ for the functional variations (Kessling et al. 1992). There are certain to be many nucleotide differences between sequences from any two individuals that are not functionally involved in the determination of phenotypic differences (Kimura 1983; Ohno 1972; Zuckerkandl 1992). A comparison of individuals carrying haplotypes that are hypothesized to be close in evolutionary history, but that associate with significantly different trait levels, will minimize the differences in background 'silent' variations in favor of candidate functional DNA sequence variations. The cladistic analysis suggests that individuals carrying haplotypes H1 and H6 should be compared to individuals carrying haplotype H5, and not to individuals carrying haplotype H2, to maximize the chance of finding candidate DNA sequences responsible for the observed phenotypic effects.

The last step in identifying functional DNA sequence variations involves evaluating all the identified variations to determine those with a high probability of being functional. The traditional methods for statistically testing for functionality, which ignore evolutionary information, will not unequivocally sort out which sites are functional because the variations detected will not be independent, that is they will be in linkage disequilibrium. Thus, any statistically significant association between a single DNA sequence variation and phenotypic variation may be due to other DNA sequence variations in linkage disequilibrium with the one under study. In contrast, as long as the sample size is sufficient, the cladistic analysis can be used iteratively, incorporating new candidate functional variations as they are identified. DNA sequence variations that are repeatedly found to differentiate clades that are significantly different phenotypically are those for which the probability of functionality increases, while those variations that differentiate clades that are not significantly dif-

ferent are those for which the probability of functionality decreases.

Inferences about genetic architecture

The genetic architecture of a quantitative trait is defined as the number of genes involved in the metabolism of the trait, the number of functional alleles at each gene and their relative frequencies, the arrangement of these alleles into genotypes, the impact of each of the alleles, single gene genotypes and multigene 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. 1992b,c). Most research to date suggests that the LDLR alleles that contribute to the genetic architecture of plasma lipid traits are rare and have large effects (greater than 3 standard deviations of the mean) on lipid metabolism (Goldstein and Brown 1989). However, this inference is biased because only those individuals with hyperphenotypes or coronary artery disease (CAD) were studied and thus those alleles in individuals in the normal range of variation have been missed. We expect there to be many LDLR alleles (Weiss 1996) and that the effects of most of these alleles will be less than 1 standard deviation of the mean. It has long been thought that the effects of such alleles would be difficult if not impossible to characterize (Morton 1992). Some evidence from the literature suggests that there may be common allelic variations in the LDLR gene that have small, less than 1 standard deviation from the mean, effects on interindividual lipid variation (Humphries et al. 1991; Pedersen and Berg 1989; Schuster et al. 1990). This study of modest size supports this hypothesis, and in fact detected allele effects of less than half of a standard deviation from the mean. But, if the allele effects are context-dependent, then our inferences are appropriate only for Rochester, Minnesota. Future studies of other populations will be necessary to establish the invariant and context-dependent contributions of common genetic variations in the LDLR gene to plasma lipid and apolipoprotein levels (Sing et al. 1996).

Another way of characterizing the effects of genetic variation is to ask how much phenotypic variation is explained by it. After removing individuals carrying the rare haplotypes H3 and H4 from this sample representative of the Rochester, Minnesota population, the nine LDLR genotypes observed in the males explained 4.6% of the total sum of squares in adjusted apo B levels (4.5% of the unadjusted apo B levels) and 6.3% of the total sum of squares for adjusted CII (5.2% of the unadjusted apo CII levels) using a one-way analysis of variance (results not shown). The results from the cladistic analysis suggested that the H2 and H5 haplotypes were functionally equivalent, so the one-way analysis of variance was repeated after combining the H1/H2 and H1/H5, H6/H2 and H6/H5, and H2/H2, H2/H5, and H5/H5 genotype classes. The resulting five genotype classes explained 83% of the sum of squares explained by the nine genotype classes for ad-

justed apo B levels and 78% of the sum of squares explained by the nine genotype classes for adjusted apo CII levels in this sample of males (results not shown). These results support the inference from the cladistic analysis that haplotypes H2 and H5 can be combined into a single functional allele class. In summary, the LDLR gene contains common functional variations with small average effects, equal to or smaller than those attributable to the apo E polymorphism (Kaprio et al. 1991; Xhignesse et al. 1991), as well as rare functional variations with large effects.

The suggestion from our study that the contribution of the LDLR gene to the genetic architecture of interindividual variation in plasma lipid and apolipoprotein levels can be described as many-to-many, i.e., many haplotypes affecting many traits, illustrates a major advantage of the cladistic approach. Although the cladistic approach does

not presently incorporate the covariance between traits such that a truly multivariate analysis can be done, it does provide a strategy for establishing the mapping between multiple haplotypes and multiple traits. Such a strategy recognizes the complexity of the relationship between genetic variation and phenotypic variation that is not taken into consideration in traditional statistical methods, which seek to establish the relationship between a single mutational change and variation in a quantitative trait.

Acknowledgements We would like to thank Dr. Stephen T. Turner for his supervision of the collection of the Rochester Family Heart Study pedigrees and Yvette Perry and Nancy Petro for their technical expertise in the genotyping of the LDLR RFLP markers. Measurement of plasma lipids and apolipoproteins was carried out in the Mayo Atherosclerosis Research laboratory under the direction of Dr. B.A. Kottke. This research was supported by NIH grant HL39107.

Appendix

Concomitant and plasma lipid and apolipoprotein trait means, standard deviations (SD), and ranges (minimum-maximum)

	Females (n = 217)			Males (n = 187)			P-value	
	Mean	SD	Min-Max	Mean	SD	Min-Max	Mean ^a	Variance ^b
Age (years)	41.7	6.3	26.0–62.9	43.5	6.8	27.2–61.5	0.0064	0.2373
Height (cm)	164.6	5.7	148.5–189.8	178.0	6.6	159.6–195.2	0.0001	0.0342
Weight (kg)	68.9	14.4	43.5–135.7	84.8	13.0	50.6–124.3	0.0001	0.1519
Apo AI (mg/dl)	141.5	18.2	105.0–197.0	132.1	16.8	94.0–196.0	0.0001	0.2345 ^d
Apo AII (mg/dl)	34.4	4.4	22.2–50.1	34.5	5.0	18.9–53.5	0.8231	0.0788
Apo B (mg/dl)	75.3	14.1	45.0–123.0	81.1	14.9	49.0–132.0	0.0001	0.4333
Apo CII (mg/dl)	2.1	0.8	0.2–5.7	2.6	1.0	1.1–6.9	0.0001	0.0440
Apo CIII (mg/dl)	13.5	4.1	5.9–28.4	15.3	5.2	5.0–36.7	0.0002	0.0004 ^c
Apo E (mg/dl)	4.8	1.9	1.4–13.5	5.2	3.1	1.4–29.6	0.0657	0.0001 ^c
T-Cholesterol (mg/dl)	181.7	33.3	101.0–333.0	194.5	39.1	99.0–387.0	0.0005	0.0245 ^c
HDL-C (mg/dl)	50.8	12.9	20.0–96.0	41.6	10.5	21.0–85.0	0.0001	0.0033
Triglycerides (mg/dl)	107.3	59.5	38.0–452.0	148.0	104.6	33.0–942.0	0.0001	0.0001

^a Heterogeneity of means between females and males was tested using the *t*-test. Satterthwaites *t*-test correction was used if the variances were significantly heterogeneous between the genders

^b Heterogeneity of variance between females and males was tested using the *F*-test

^c The null hypothesis of homogeneity of dispersion was not rejected using the Ansari-Bradley test ($P > 0.10$)

^d The null hypothesis of homogeneity of dispersion was rejected using the Ansari-Bradley test ($P = 0.0336$)

References

- Aquadro CF, Jennings RM, Bland MM, Laurie CC, Langley CH (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
- Boerwinkle E, Chakraborty R, Sing CF (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
- Crandall KA, Templeton AR (1993) Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics* 134:959–969
- Edgington ES (1987) *Randomization tests*, 2nd edn. Dekker, New York
- Felsenstein J (1983) Parsimony in systematics: biological and statistical issues. *Annu Rev Ecol Syst* 14:313–333
- Goldstein JL, Brown MS (1989) Familial hypercholesterolemia. In: Scriver CR, Beaudet AL, Sly WS, Valle D (eds) *The metabolic basis of inherited disease*, 6th edn. McGraw-Hill, New York, pp 1215–1250
- Haviland MB (1993) A cladistics approach to the genetic analysis of human quantitative data. Ph. D. thesis, University of Michigan
- Haviland MB, Kessler AM, Davignon J, Sing CF (1995) Cladistic analysis of the apolipoprotein AI-CIII-AIV gene cluster using a healthy French Canadian sample. I. Haploid analysis. *Ann Hum Genet* 59:211–231
- Hobbs HH, Russell DW, Brown MS, Goldstein JL (1990) The LDL receptor locus in familial hypercholesterolemia: mutational analysis of a membrane protein. *Annu Rev Genet* 24:133–170
- Humphries S, Coviello DA, Masturzo P, Balestreri R, Orecchini G, Bertolini S (1991) Variation in the low density lipoprotein receptor gene is associated with differences in plasma low density lipoprotein cholesterol levels in young and old normal individuals from Italy. *Arterioscler Thromb* 11:509–516

- Izzo C, Grillo F, Murador E (1981) Improved method for determination of high density lipoprotein cholesterol: isolation of high density lipoproteins by use of polyethylene glycol-6000. *Clin Chem* 27:371–374
- Kaprio J, Ferrell RE, Kottke BA, Kambouh MI, Sing CF (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
- Kessling AM, Rajput-Williams J, Bainton D, Scott J, Miller NE, Baker I, Humphries SE (1988) 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, Ouellette S, Bouffard O, Chamberland A, Bétard 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
- Kimura M (1983) *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge
- Kottke BA, Moll PP, Michels VV, Weidman WH (1991) Levels of lipids, lipoproteins, and apolipoproteins in a defined population. *Mayo Clin Proc* 66:1198–1208
- Leitersdorf E, Hobbs HH (1988) Human LDL receptor gene: *HincII* polymorphism detected by gene amplification. *Nucleic Acids Res* 16:7215
- Leitersdorf E, Chakravarti A, Hobbs HH (1989) Polymorphic DNA haplotypes at the LDL receptor locus. *Am J Hum Genet* 44:409–421
- Lusis AJ (1988) Genetic factors affecting blood lipoproteins: the candidate gene approach. *J Lipid Res* 29:397–429
- Merry U (1995) *Coping with uncertainty: insights from the new sciences of chaos, self-organization, and complexity*. Praeger, Westport, Connecticut
- Moll PP, Michels VV, Weidman WH, Kottke BA (1989) Genetic determination of plasma apolipoprotein AI in a population-based sample. *Am J Hum Genet* 44:124–139
- Morton NE (1992) The future of genetic epidemiology. *Ann Med* 24:557–562
- Ohno S (1972) So much “junk” DNA in our genome. In: Smith HH (ed) *Evolution of genetic systems*. Gordon-Breach, New York, pp 366–370
- Pedersen JC, Berg K (1989) Interaction between low density lipoprotein receptor (LDLR) and apolipoprotein E (apoE) alleles contributes to normal variation in lipid level. *Clin Genet* 35:331–337
- Randles RH, Wolfe DA (1979) *Introduction to the theory of non-parametric statistics*. John Wiley, New York
- Riordan JR, Rommens JM, Kerem B-S, Alon N, Rozmahel R, Grzelczak Z, Zielenski J, Lok S, Plavsic N, Chou J-L, Drumm ML, Iannuzzi MC, Collins FS, Tsui L-C (1989) Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* 245:1066–1073
- SAS Institute (1989) *SAS/STAT@ user’s guide, Version 6*. SAS Institute, Cary, North Carolina
- Schuster H, Humphries S, Rauh G, Held C, Keller CH, Wolfram G, Zöllner N (1990) Association of DNA-haplotypes in the human LDL-receptor gene with normal serum cholesterol levels. *Clin Genet* 38:401–409
- Sing CF, Boerwinkle E (1985) The genetics of blood pressure variability: an overview. In: Filer LJ Jr, Lauer RM (eds) *Children’s blood pressure (Report of the 88th Ross Conference on Pediatric Research)* Ross Laboratories, Columbus, Ohio, pp 35–43
- Sing CF, Moll PP (1989) Genetics of variability of CHD risk. *Int J Epidemiol* 18:S183–S195
- Sing CF, Reilly SL (1993) Genetics of common diseases that aggregate, but do not segregate, in families. In: Sing CF, Hanis CL (eds) *Genetics of cellular, individual, family, and population variability*. Oxford University Press, New York, pp 140–161
- Sing CF, Haviland MB, Templeton AR, Zerba KE, Reilly SL (1992a) Biological complexity and strategies for finding DNA variations responsible for inter-individual variation in risk of a common chronic disease, coronary artery disease. *Ann Med* 24:539–547
- Sing CF, Haviland MB, Zerba KE, Templeton AR (1992b) Application of cladistics to the analysis of genotype-phenotype relationships. *Eur J Epidemiol* 8:3–9
- Sing CF, Zerba KE, Haviland MB (1992c) Genetic architecture of interindividual variation in plasma cholesterol. In: Bearn AG (ed) *Genetics of coronary heart disease*. University of Oslo, Oslo, pp 143–162
- Sing CF, Haviland MB, Reilly SL (1996) *Genetic architecture of common multifactorial diseases: variation in the human genome (Ciba Foundation Symposium 197)*. Wiley, Chichester, pp 211–232
- Sober E (1983) Parsimony in systematics: philosophical issues. *Annu Rev Ecol Syst* 14:335–357
- Stewart C-B (1993) The powers and pitfalls of parsimony. *Nature* 361:603–607
- Strohman R (1994) Epigenesis: the missing beat in biotechnology? *Biotechnology* 12:156–164
- Tanksley SD (1993) Mapping polygenes. *Annu Rev Genet* 27:205–233
- Templeton AR (1987) The general relationship between average effect and average excess. *Genet Res* 49:69–70
- Templeton AR, Sing CF (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 AR, Boerwinkle E, Sing CF (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 AR, Sing CF, 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
- Templeton AR, Crandall KA, Sing CF (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
- Turner ST, Weidman WH, Michels VV, Reed TJ, Ormson CL, Fuller T, Sing CF (1989) Distribution of sodium-lithium countertransport and blood pressure in Caucasians five to eighty-nine years of age. *Hypertension* 13:378–391
- Wallace MR, Marchuk DA, Andersen LB, Letcher R, Odeh HM, Saulino AM, Fountain JW, Brereton A, Nicholson J, Mitchell AL, Brownstein BH, Collins FS (1990) Type 1 neurofibromatosis gene: identification of a large transcript disrupted in three NF1 patients. *Science* 249:181–186
- Weiss KM (1996) Is there a paradigm shift in genetics? Lessons from the study of human diseases. *Mol Phylogenet Evol* 5:259–265
- Wichmann BA, Hill ID (1982) Algorithm AS 183: an efficient and portable pseudo-random number generator. *Appl Stat* 31:188–190
- Wijsman EM (1987) A deductive method of haplotype analysis in pedigrees. *Am J Hum Genet* 41:356–373
- Xhignesse M, Lussier-Cacan S, Sing CF, Kessling AM, 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 KE, Sing CF (1993) The role of genome type-environment interaction and time in understanding the impact of genetic polymorphisms on lipid metabolism. *Curr Opin Lipid* 4:152–162
- Zuckerkindl E (1992) Revisiting junk DNA. *J Mol Evol* 34:259–271