Human Molecular Genetics, 1998, Vol. 7, No. 11 **1745–1751**

Measured haplotype analysis of the angiotensin-I converting enzyme gene

Bernard Keavney[+], Colin A. McKenzie[+], John M. C. Connell[1], Cécile Julier, Peter J. Ratcliffe, Eric Sobel, Mark Lathrop and Martin Farrall*

The Wellcome Trust Centre for Human Genetics, Nuffield Department of Clinical Medicine, University of Oxford, Windmill Road, Oxford OX3 7BN, UK and [1]MRC Blood Pressure Group, Department of Medicine and Therapeutics, University of Glasgow, Glasgow, UK

Received May 9, 1998; Revised and Accepted July 27, 1998

Linkage and segregation analysis have shown that circulating angiotensin-I converting enzyme (ACE) levels are influenced by a major quantitative trait locus that maps within or close to the *ACE* gene. The D variant of a 287 bp insertion/deletion (I/D) polymorphism in intron 16 of the gene is associated with high ACE levels and may also be related to increased risk of cardiovascular disease. Multiple variants that are in linkage disequilibrium with the I/D polymorphism have been described, but it is unknown if any of these are directly implicated, alone or in combination with as yet undiscovered variants, in the determination of ACE levels. An analysis of 10 polymorphisms spanning 26 kb of the *ACE* gene revealed a limited number of haplotypes in Caucasian British families due to strong linkage disequilibrium operating over this small chromosomal region. A haplotype tree (cladogram) was constructed with three main branches (clades A–C) which account for 90% of the observed haplotypes. Clade C is most likely derived from clades A and B following an ancestral recombination event. This evolutionary information was then used to direct a series of nested, measured haplotype analyses that excluded upstream sequences, including the *ACE* promoter, from harbouring the major *ACE*-linked variant that explains 36% of the total trait variability. Residual familial correlations were highly significant, suggesting the influence of additional unlinked genes. Our results demonstrate that a combined cladistic/measured haplotype analysis of polymorphisms within a gene provides a powerful means to localize variants that directly influence a quantitative trait.

INTRODUCTION

The analysis of multiple polymorphisms in a candidate gene or region and the definition of their relationship to the underlying susceptibility or quantitative trait locus (QTL) represents a major theoretical and practical bottleneck in the identification of DNA variants underlying complex inherited diseases and traits. The angiotensin-I converting enzyme (ACE) is strongly influenced by a limited number of QTLs and is an experimentally tractable model trait for the development of strategies for identifying such variants. Physiological, pharmacological and genetic studies have all shown the importance of ACE and its inhibition in the pathogenesis and treatment of a variety of cardiovascular diseases (1–7). Therefore, there is considerable interest in defining the molecular mechanisms determining circulating ACE levels in humans.

Evidence that circulating ACE levels are under substantial genetic control was first detected in a segregation analysis of ACE levels in 87 healthy nuclear families (8). Familial correlations showed a striking pattern that was consistent with a QTL influencing plasma ACE levels. The human *ACE* gene has been cloned and localized to chromosome 17q23 (9) and a 287 bp insertion/deletion (I/D) polymorphism in intron 16 has been identified for use as a genetic marker (10). A population-based analysis of unrelated individuals with the I/D polymorphism showed an association between this polymorphism and plasma ACE levels. The association followed an additive pattern with D/D and I/I genotypes being associated with high and low ACE levels, respectively, and I/D heterozygotes being associated with an intermediate level (10). These associations have been confirmed in family studies in white European (11) and Afro-Caribbean (12) populations. Linkage between the QTL and a neighbouring microsatellite marker has been detected and alternative models incorporating an unlinked genetic component have also been investigated (12).

Following a sequencing study, Villard *et al.* (13) reported a series of polymorphisms within the *ACE* gene and investigated the pattern of linkage disequilibrium both among the polymorphisms and between the polymorphisms and what they refer to as one or two *ACE*-linked QTLs. They proposed that two variants influence circulating ACE levels: one that could be in complete linkage disequilibrium or identical with the I/D polymorphism and a second that mapped upstream of the start of translation of the *ACE* gene.

*To whom correspondence should be addressed. Tel: +44 1865 740 012; Fax: +44 1865 742 196; Email: martin.farrall@well.ox.ac.uk
[+]The first two authors contributed equally to this work

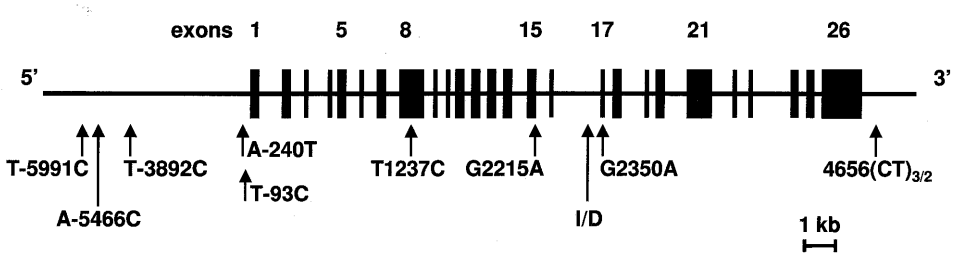

Figure 1. Schematic diagram of the human angiotensin-I converting enzyme (ACE) gene illustrating the location of 10 biallelic polymorphisms. Polymorphisms are numbered in base pairs relative to the start of transcription of the *ACE* gene. Exons 1–26 are indicated with vertical bars and are numbered intermittently for clarity.

Table 1. Ranked ACE haplotype frequencies

Haplotype	T-5991C	A-5466C	T-3892C	A-240T	T-93C	T1237C	G2215A	I/D	G2350A	4656(CT)$_{3/2}$	Frequency
1	T	A	T	A	T	T	G	I	A	3	0.35
2	C	C	C	T	C	C	A	D	G	2	0.28
3	T	A	T	A	T	C	A	D	G	2	0.09
4	T	A	C	A	T	C	A	D	G	2	0.07
5	T	A	T	A	T	C	G	I	A	3	0.05
6	C	C	C	T	C	C	G	D	G	2	0.03
7	T	A	T	A	T	T	A	I	A	3	0.03

Here we report the results of a haplotype analysis with 10 polymorphisms at the *ACE* locus that were genotyped in a series of white British families. An evolutionary tree (cladogram) was proposed which accounted for the majority (90%) of the observed haplotypes. Information from the cladogram was then used to direct a nested, measured haplotype analysis of plasma ACE levels and to investigate both the number of variants influencing the ACE quantitative trait and their likely location.

RESULTS

Cladistic haplotype analysis

Haplotypes of 10 biallelic polymorphisms, distributed throughout the *ACE* gene (Fig. 1), were assigned to 555 members of 83 extended families by following a novel two-stage procedure. Firstly, genotype data were repeatedly run through the SIMWALK program (14) to compile a list of every possible set of haplotypes for each family that included no recombination events. In 24 families a unique set of haplotypes was found, but in other families as many as 48 alternative haplotype sets (with zero recombinants) were found to be consistent with the genotypic data. In the second stage, the haplotype data were subjected to a maximum likelihood analysis in order to estimate the haplotype frequencies in the founders. Haplotype sets with small final probabilities ($<10^{-6}$) were eliminated from further analysis, leaving 239 non-recombinant haplotype sets to enter into the measured haplotype analysis.

The seven most frequent haplotypes, which cumulatively account for 90% of the founders' haplotypes, are listed in Table 1 (a complete listing of the haplotypes is available from the World Wide Web at http://www.well.ox.ac.uk/mfarrall). The polymorphisms are all in strong linkage disequilibrium; sparse contingency tables analyses, testing each polymorphism in turn for association with haplotypes constructed from the remaining polymorphisms, were all highly significant ($P < 0.001$). The 95% confidence interval for the expected number of different haplotypes in 183 unrelated individuals using the observed gene frequencies and assuming linkage equilibrium is 267–295 (versus the 29 found for the 183 founders).

The two most frequent haplotypes, haplotype 1 (T-A-T-A-T-T-G-I-A-3) and haplotype 2 (C-C-C-T-C-C-A-D-G-2), complement each other at all 10 sites. Haplotype 3 (T-A-T-A-T-C-A-D-G-2) is consistent with being generated by an ancestral recombination event between haplotype 2 and haplotype 1, 5 or 7 with a breakpoint in the interval delimited by T-93C and T1237C. Haplotypes 5 (T-A-T-A-T-C-G-I-A-3) and 7 (T-A-T-A-T-T-A-I-A-3) each show one difference from haplotype 1 at the T1237C and G2215A markers, respectively; these differences are presumably due to spontaneous mutations. Haplotype 6 (C-C-C-T-C-C-G-D-G-2) shows one difference from haplotype 2 at the G2215A marker. Haplotype 4 (T-A-C-A-T-C-A-D-G-2) shows one difference from haplotype 3 at the T-3892C marker.

Following the principle of maximum parsimony, we propose a putative haplotype tree linking haplotypes 1–7, which is illustrated in Figure 2. Three closely related branches (clades) are proposed: A, comprising haplotypes 1, 5 and 7; B, comprising haplotypes 2 and 6; and C, comprising haplotypes 3 and 4. The other rare haplotypes were pooled into a common group (X) for subsequent analyses.

Measured haplotype analysis

The results of a pedigree analysis, to estimate by maximum likelihood the influence of each haplotype on the ACE quantitat-

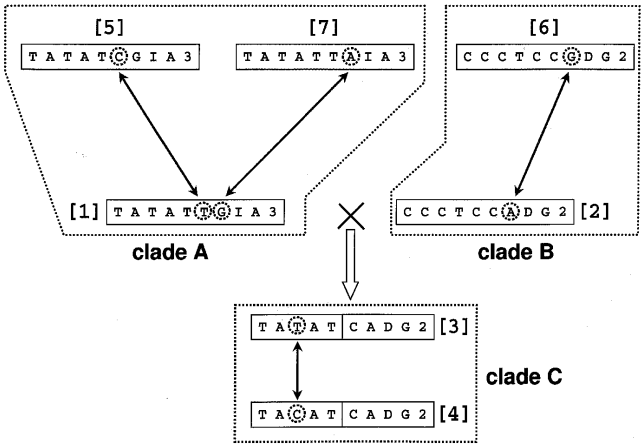

Figure 2. Evolutionary tree (cladogram) for 10 ordered *ACE* polymorphisms. Alternative sites for the polymorphisms are shown in their physical order as follows: T-5991C, A-5466C, T-3892C, A-240T, T-93C, T1237C, G2215A, I/D, G2350A and 4656(CT)$_{3/2}$. Three main branches (clades) are evident, clade A (haplotypes 1, 5 and 7), clade B (haplotypes 2 and 6) and clade C (haplotypes 3 and 4). Haplotype 3 is derived following a recombination event (indicated by a cross) between haplotypes 1 and 2. Single substitutions between clade members are indicated with double-headed arrows.

ive trait, are shown in Table 2. Additive (co-dominant) models were fitted, as supplementary analyses showed no significant support for a dominance variance component (data not shown). The most general model ([1, 2, 3, 4, 5, 6, 7, X]) includes seven haplotype-specific parameters to model the mean effect on the ACE quantitative trait for each haplotype shown in Table 1 and a single parameter to model the common effect for all the other haplotypes (group X). Residual familial correlations that were not accounted for by the ACE-linked effects were modelled by following a class D regressive approach (15).

The goodness-of-fit of the nested model [A, B, C, X] was evaluated; no significant difference (P = 0.7) was found in a comparison with the general model ([1, 2, 3, 4, 5, 6, 7, X]). This result indicates that clade members resemble each other phenotypically to the limit of resolution in these data. The models [A ≡ C, B, X] and [A, B ≡ C, X] were then fitted to test if the phenotypic effects of recombinant clade C resembled more closely those of clades A or B. When these nested models were compared with the general model ([1, 2, 3, 4, 5, 6, 7, X]), model [A ≡ C, B, X] was strongly rejected ($P < 0.001$) but model [A, B ≡ C, X] was not rejected (P = 0.4). These results demonstrate that clades B and C have very similar effects on the ACE trait. Table 3 shows the mean and standard deviation for plasma ACE levels for each genotype (comprising two haplotypes) after fitting model [A, B, C, X]. There is substantial support for residual familial correlations that are not accounted for by the *ACE*-linked major gene, for example the [A, B, C, X] model fitted without residual correlations is rejected ($P < 0.001$) when tested against a similar model including residual correlations. Using the parameters fitted in the [A, B ≡ C, X] model including residual familial correlations, the proportion of the total phenotypic variance accounted for by the *ACE*-linked variant is 36% and the residual mother–offspring, father–offspring and sibling correlations are 0.15, 0.35 and 0.33, respectively.

Table 2. Measured haplotype analysis of plasma ACE levels

Model	r^2	No. of parameters	$-2 \times \log_e$ likelihood	χ^2	P-value
1, 2, 3, 4, 5, 6, 7, X	+	20	3048.0	–	–
1, 2, 3, 4, 5, 6, 7, X	–	16	3084.9	37.0	2×10^{-7}
A, B, C, X	+	16	3050.2	2.3	0.7
A, B, C, X	–	12	3088.1	40.1	3×10^{-6}
A, B ≡ C, X	+	15	3053.1	5.1	0.4
A, B ≡ C, X	–	11	3090.9	42.9	2×10^{-6}
A ≡ C, B, X	+	15	3117.3	69.4	1×10^{-13}

r^2, residual familial correlations (class D regressive model) optionally fitted (+) or omitted (–).

Table 3. Genotype means table of plasma ACE levels

Genotype	μMHA	μ	σ	N
A/A	15.94	16.00	3.82	72
A/B	21.72	21.75	4.77	95
A/C	20.90	21.74	5.49	50
A/X	18.05	17.08	3.58	24
B/B	27.51	28.10	5.10	30
B/C	26.69	26.76	5.32	35
B/X	23.83	22.49	4.69	17
C/C	25.87	27.18	5.68	8
C/X	23.01	20.69	4.79	16
X/X	20.16	24.62	3.36	6

μMHA, genotype-specific means (within-genotype SD = 4.8) fitted in a measured haplotype analysis for the model [A, B, C, X] including residual familial correlations (class D regressive model); μ, σ and N, genotype-specific means, SDs and number, respectively, of individuals with each genotype for a simple statistical analysis pooling data for all family members with unambiguous haplotypes.

DISCUSSION

The two most frequent haplotypes at the *ACE* locus in the white British population are complementary to each other for all 10 polymorphisms typed in this study. This is perhaps not an unexpected finding, given the strategy employed by Villard *et al.* (13) in defining the polymorphisms. They sequenced four subjects with contrasting ACE levels, thus maximizing the chances of detecting new polymorphisms that are in strong disequilibrium with the I/D polymorphism [and in all likelihood with the trait variant(s) associated with the I/D polymorphism]. The third most frequent haplotype shares the proximal half of its sites with haplotypes in clade A and the distal half with haplotype 2. We propose that haplotype 3 was most likely created by an ancestral recombination event in an individual with the genotype A/2. A very similar haplotype pattern can be deduced from the data reported by Villard *et al.* (13); however, they subdivided the haplotype at the breakpoint and analysed pairs of markers rather than the complete haplotype.

It is important to consider alternative molecular mechanisms to explain how haplotype 3 arose. A gene conversion or double recombination event involving the transfer of a continuous segment of at least 6 kb of DNA, including the five polymorphisms that map upstream of the *ACE* promoter, from clade A to haplotype 2 cannot be excluded from consideration. Under this scenario, the major variant influencing ACE levels could then be located upstream of the *ACE* gene. However, at least 6 kb of sequence including the *ACE* promoter would then be excluded from harbouring the major variant influencing ACE levels.

In our data, haplotypes 4–7 each showed a single internal difference with haplotypes 1, 2 and 3 and could therefore be assigned to one of the three clades, adhering to the principle of maximum parsimony. Several rare haplotypes not listed in Table 1 could also be assigned to one of the three clades. We have not measured the individual effects of these haplotypes on ACE levels as the numbers of phenotyped subjects carrying these haplotypes is too small for a meaningful analysis, but instead we pooled all rare haplotypes into group X.

The finding of a limited number of clades provides a boundary for the potential model space that should be explored in attempts to structure the underlying variation that directly influences ACE levels. Consider two alternative models, a single variant model and a two variant model with one variant on either side of the ancestral breakpoint. For the single variant model, we may suppose that clade A carries a low trait allele and clade B carries a high trait allele. We then directly test the location of the variant relative to the ancestral breakpoint in a nested statistical analysis. For the two variant model, we suppose that clade A carries two low trait alleles, clade B carries two high trait alleles and clade C one low and one high trait allele. If we assume that one variant shows no dominance over the other (this assumption has been confirmed in supplementary analyses and it is also evident from an inspection of the A/A, A/B and B/B genotype means shown in Table 3), then we would expect that clade C would have an effect on the ACE trait which is intermediate between the effects of clades A and B. The results from the measured haplotype analysis (Table 2 and 3) show that clade C very closely resembles clade B in its influence on ACE levels. We therefore conclude that the major variant influencing the ACE levels is located downstream of the ancestral breakpoint and that there is no significant variant in the promoter or other upstream sequences. Thus, the finding of Villard *et al.* (13) of a major *ACE* variant mapping to the promoter region is not confirmed by our study. We suggest that analysis of their data by the methods presented here might be helpful.

An alternative statistical test for other variants at the *ACE* locus is provided by a comparison of models [A, B, C, X] and [A, B ≡ C, X]. This test shows that there is no support ($P > 0.05$) for an additional variant mapping upstream of the promoter. We also note that the polymorphisms T1237C and G2215A appear in their alternative states in members of clades A and B, which makes them very unlikely to be the downstream major *ACE* variant. The I/D, G2350A and 4656$(CT)_{3/2}$ polymorphisms show no substitutions within the clades, consequently no differential inferences may be made regarding their influence on ACE levels. The presence of residual familial correlations that are not accounted for by the ACE-linked QTL was strongly supported in our analyses. These residual correlations are compatible with the presence of an unlinked genetic factor(s) consistent with our previous findings based on a different population (12).

Sequential examination of individual polymorphisms within a gene or region of interest in an attempt to narrow the region in which a causative variant may be found is fraught with problems of interpretation. For correct inferences to be drawn, disequilibrium models need to be correctly specified in terms of the number of variants, which is impossible to verify until the exact nature of the quantitative variation is known. Also, the dimensionality of disequilibrium models increases exponentially with the number of polymorphisms and variants in the model and strategies to reduce the complexity of the analytical problem are desirable. Haplotype-based analyses summarize linkage disequilibrium relationships and provide a natural approach to mapping quantitative variation onto genetic diversity that is biologically appealing. To this end, Templeton and colleagues (for a recent review see ref. 16) have investigated cladistic methodologies in the study of quantitative variation. They argue that mutations that influence traits have arisen during the evolutionary history of the set of polymorphisms that are associated with the trait. Consequently, it is expected that members of clades of haplotypes will resemble each other phenotypically and the overall haplotype cladogram will be useful in devising powerful statistical tests to distinguish the most important mutations. Using data sets for *Drosophila* traits, Templeton and Sing (17) have shown how useful recombination can be to refine the physical localization of phenotype-modifying mutations. Here we have adopted these methods to the analysis of a human quantitative trait with no prior assumptions regarding the number or location of the underlying variants. Our method takes full advantage of the opportunity offered by a simple haplotype cladogram with one clade derived from an ancestral recombination event.

Sequencing of the seven introns between the T-93C and T1237C polymorphisms will be required to identify further polymorphisms, which will refine the localization of the ancestral breakpoint and will also characterize the variant(s) lying downstream of the T-93C polymorphism that influences ACE levels. Sequencing studies in other populations with different (recent) evolutionary histories may reveal alternative polymorphisms, breakpoints and cladograms that may refine the localization of the major variant(s) influencing ACE levels. These approaches will be necessary to identify a limited number of candidate variants in the *ACE* gene, which will require examination in biological assays in order to establish their functional significance.

Table 4. Primer sequences for PCR and digestion of ACE polymorphisms

Polymorphism	Primer sequence	T_A (°C)[a], $MgCl_2$ (mM)	Restriction endonuclease	Fragment size (bp)
T-5491C	TAC AAC CAT CAC TAC TAA TGT C	55, 2.0	*Bst*NI	110/88
	TAT AAT ATA TGT GAC ATG GCC TG			
A-5466C	GCC ATG TCA CAT ATA TTA TAG GA	50[b], 2.5	*Eco*RV	133/109
	CGT CTT TGG AAA CTT GTC TGC			
T-3892C	ATA GTG TAT ATA GGG CTT GGT AC	55, 2.0	*Pst*I	114/90
	AGA AGA TAT TTG CAA AGT ATG TAC TG			
A-240T	TCG GGC TGG GAA GAT CGA GC	55, 2.0	*Xba*I	137/114
	GAG AAA GGG CCT CCT CTC TCT			
T-93C[c]	TGT CAC TCC GGA GGC GGG AGG CT	60[b], 2.0	*Hin*fI	389/291/100
	CGG CTC TGC CCC TTC TCC TGC GC			
T1237C	AGT GCA CAC GGG TCA CGA TG	58[b], 2.0	*Msp*I	108/88
	CCC CCC GAC GCA GGG AGC C			
G2215A	CAC ACC CTG AAG TAC GGC AC	60, 1.0	*Hae*II	131/109
	TCC TCC AGC TCC TGG GCA G			
G2350A	CTG ACG AAT GTG ATG GCC GC	58, 2.0	*Bst*UI	122/103
	TTG ATG AGT TCC ACG TAT TTC G			
$4656(CT)_{3/2}$	TTG GCT CCT GCT GTA CCA G	60, 2.0	–	112/114
	CTG CTA CAC TCC AGC GTC TG			

Mismatched nucleotides are underlined.
[a]Annealing temperature for PCR.
[b]Touchdown technique.
[c]Villard *et al*. (13).

MATERIALS AND METHODS

Families and ACE trait

Eighty-three extended families with between four and 18 members were collected through an established project (HTO) to investigate the genetic basis of blood pressure variability and cardiovascular risk factor status in the British white population (18). The Central Oxford Research Ethics Committee approved the protocol for collection of the families. Plasma ACE activity was measured by HPLC with the use of a synthetic substrate (19). All samples from one family were assayed in a single run; each sample was analysed in duplicate and re-analysed if the two analyses differed by >15%. A preliminary correlational analysis found no evidence of an association between plasma ACE levels and systolic blood pressure ($\rho = 0.0013$, $P = 0.98$) or diastolic blood pressure ($\rho = -0.045$, $P = 0.45$); consequently no ascertainment correction was made in analyses of plasma ACE levels. Plasma ACE levels were standardized separately for males and females to account for a small but significant ($P < 0.05$) difference. ACE levels showed no significant relationship with age and/or body mass index. Ages of family members ranged from 19 to 90 years.

Genotyping

DNA was extracted using standard methods (20). The I/D polymorphism at the *ACE* locus was typed using a PCR-based method (21). A further nine polymorphisms, T-5991C, A-5466C, T-3892C, A-240T, T-93C, T1237C, G2215A, G2350A and $4656(CT)_{3/2}$, that were originally described by Villard *et al*. (13), were genotyped following PCR amplification (Table 4). Further PCR experimental details may be found on a World Wide Web page (http://www.well.ox.ac.uk/mfarrall/ACE_polymorphisms.html). The genomic location of each polymorphism is shown with respect to *ACE* exonic sequences in Figure 1.

Haplotype assignment

The SIMWALK program (14) *ab initio* finds a set of (possibly recombinant) haplotypes that are consistent with the family structure and the multilocus genotypic data. The program then performs simulated annealing, a combinatorial optimization procedure, to search for more likely states with fewer recombinant haplotypes and terminates immediately on finding a non-recombinant haplotype state. By initiating SIMWALK with different random seeds in a pedigree, one may obtain replicates that are a random sample from the space of all possible underlying haplotype configurations. In SIMWALK, equal allele frequencies at each site were assumed, so that the frequency with which different haplotype configurations (haplotype sets) emerge in the simulation would depend only on the pattern of genotype data and relationships in that family. After many replicate haplotype sets were produced for each family, maximum likelihood estimates of the haplotype frequencies were computed based on a gene counting or expectation maximization (EM) algorithm (22,23), similar to that used for random samples from diploid populations (24). Briefly, equal or other arbitrary values were assumed as initial haplotype frequency estimates. New estimates were

obtained by weighting each haplotype set from the simulation by the frequency estimates applied to the founders. The weighted haplotype frequencies in founding pedigree members were calculated and used as new estimates in the EM procedure. Iterations were continued until the estimates converged. Haplotype sets for each family were then tabulated with their final probabilities. Unfavourable sets associated with very small probabilities ($<10^{-6}$) were discarded, with no effect on subsequent analyses.

Exact significance tests for sparse contingency table analyses were performed using the computer program CoCo 1.3.R2 obtained by anonymous ftp from the server ftp.math.auc.dk (25).

Measured haplotype analysis

Likelihood calculations were performed using a modified version of the Pedigree Analysis Package (PAP) v.4.02 (26,27). The PAP FORTRAN subroutine qmlprmv, which models major quantitative traits, was modified to include a parameter to model the mean effect on the ACE trait for each haplotype; these effects were assumed to act additively (i.e. co-dominant inheritance). A single parameter was specified to model the residual genotype-specific variation. Residual familial correlations were accounted for by a class D regressive model that includes spouse, mother–offspring, father–offspring and sibling correlations (15). The most general model [1,2,3,4,5,6,7,X] therefore includes a total of 20 parameters, seven haplotype frequencies, eight mean effects, one residual standard deviation and four to model residual familial correlations. The main likelihood computing subroutine in PAP, paplk, was modified to calculate the overall likelihood of a family as a sum of likelihood terms computed for each alternative haplotype set. The goodness-of-fit of nested models was evaluated using likelihood ratio tests, where the difference in $-2 \log_e$ likelihood between models is assumed to be asymptotically distributed as a χ^2 statistic with degrees of freedom equal to the difference in number of parameters fitted in the two models. The choice of which model comparisons will be the most informative is difficult and should be postponed until a preliminary analysis of the haplotypes, without reference to the quantitative trait, has been completed. For example, it may be possible to reconstruct a plausible cladogram that summarizes the evolutionary relationships between haplotypes. The impact on the goodness-of-fit resulting from pooling the effects of closely related haplotypes (clade members) provides a natural series of tests. The superimposition of an ancestral recombination breakpoint on the haplotype cladogram allows tests between different clades in order to map the quantitative variation with respect to the breakpoint.

ACKNOWLEDGEMENTS

We are grateful to the families who contributed to this project. We thank Drs F. Soubrier and E. Villard for providing information on *ACE* polymorphisms prior to publication, Dr J. Conway and the Oxfordshire general practitioners who contributed index cases to the family collection and nurses Sarah Pollitt, Anna Zawadzka and Jane Cadd and Ms Rosemary Farmer for expert technical help. This work was supported by the Wellcome Trust and the British Heart Foundation (project grant no. 930006 to J.M.C.C.). B.K. is an MRC Training Fellow. C.A.M. is a Rhodes Scholar.

REFERENCES

1. Williams, G.H. (1988) Converting-enzyme inhibitors in the treatment of hypertension. *N. Engl. J. Med*., **319**, 1517–1525.
2. The SOLVD Investigators (1992) Effect of enalapril on mortality and the development of heart failure in asymptomatic patients with reduced left ventricular ejection fractions. *N. Engl. J. Med*., **327**, 685–691.
3. Pfeffer, M., Braunwald, E., Moyé, L., Basta, L., Brown, E., Cuddy, T., Davis, B. *et al*. (1992) Effect of captopril on mortality and morbidity in patients with left ventricular dysfunction after myocardial infarction. Results of the Survival and Ventricular Enlargement trial. *N. Engl. J. Med*., **327**, 669–677.
4. Lewis, E.J., Hunsicker, L.G., Bain, R.P. and Rohde, R.D. (1993) The effect of angiotensin-converting-enzyme inhibition on diabetic nephropathy. The Collaborative Study Group. *N. Engl. J. Med*., **329**, 1456–1462.
5. Waeber, B., Nussberger, J. and Brunner, H. (1995) Angiotensin-converting enzyme inhibitors in hypertension. In Laragh, J.H. and Brenner, B.M. (eds), *Hypertension: Pathophysiology, Diagnosis and Management.* Raven Press, New York, NY, pp. 2861–2875.
6. ISIS-4 (Fourth International Study of Infarct Survival) Collaborative Group (1995) ISIS-4: a randomised factorial trial assessing early oral captopril, oral mononitrate and intravenous magnesium sulphate in 58 050 patients with suspected acute myocardial infarction. *Lancet*, **345**, 669–685.
7. Maschio, G., Alberti, D., Janin, G., Locatelli, F., Mann, J., Motolese, M., Ponticelli, C. *et al.* (1996) Effect of the angiotensin-converting-enzyme inhibitor benazepril on the progression of chronic renal insufficiency. *N. Engl. J. Med*., **334**, 939–945.
8. Cambien, F., Alhenc-Gelas, F., Herbeth, B., Andre, J., Rakotovao, R., Gonzales, M., Allegrini, J. *et al*. (1988) Familial resemblance of plasma angiotensin-converting enzyme level: the Nancy study. *Am. J. Hum. Genet*., **43**, 774–780.
9. Mattei, M.-G., Hubert, C., Alhenc-Gelas, F., Roekel, N., Corvol, P. and Soubrier, F. (1989) Angiotensin I-converting enzyme gene is on chromosome 17. *Cytogenet. Cell. Genet*., **51**, 1041.
10. Rigat, B., Hubert, C., Alhenc-Gelas, F., Cambien, F., Corvol, P. and Soubrier, F. (1990) An insertion/deletion polymorphism in the angiotensin I-converting enzyme gene accounting for half the variance of serum enzyme levels. *J. Clin. Invest*., **86**, 1343–1346.
11. Tiret, L., Rigat, B., Visvikis, S., Breda, C., Corvol, P., Cambien, F. and Soubrier, F. (1992) Evidence, from combined segregation and linkage analysis, that a variant of the angiotensin I-converting enzyme (ACE) gene controls plasma ACE levels. *Am. J. Hum. Genet*., **51**, 197–205.
12. McKenzie, C.A., Julier, C., Forrester, T., McFarlane-Anderson, N., Keavney, B., Lathrop, G.M., Ratcliffe, P.J. *et al.* (1995) Segregation and linkage analysis of serum angiotensin I-converting enzyme levels: evidence for two quantitative-trait loci. *Am. J. Hum. Genet*., **57**, 1426–1435.
13. Villard, E., Tiret, L., Visvikis, S., Rakotovao, R., Cambien, F. and Soubrier, F. (1996) Identification of new polymorphisms of the angiotensin I-converting enzyme (ACE) gene and study of their relationship to plasma ACE levels by two-QTL segregation-linkage analysis. *Am. J. Hum. Genet*., **58**, 1268–1278.
14. Sobel, E. and Lange, K. (1996) Descent graphs in pedigree analysis: applications to haplotyping, location scores and marker sharing statistics. *Am. J. Hum. Genet*., **58**, 1323–1337.
15. Bonney, G.E. (1992) Compound regressive models for family data. *Hum. Hered*., **42**, 28–41.
16. Templeton, A.R. (1996) Cladistic approaches to identifying determinants of variability in multifactorial phenotypes and the evolutionary significance of variation in the human genome. *Ciba Foundn Symp*., **197**, 259–283.
17. Templeton, A.R. and 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.
18. Julier, C., Delepine, M., Keavney, B., Terwilliger, J., Davis, S., Weeks, D.E., Bui, T., Jeunemaître, X., Velho, G., Froguel, P., Ratcliffe, P., Corvol, P., Soubrier, F. and Lathrop, G.M. (1997) Genetic susceptibility for human familial essential hypertension in a region of homology with blood pressure linkage on rat chromosome 10. *Hum. Mol. Genet*., **6**, 2077–2085.
19. Chiknas, S.G. (1979) A liquid chromatography-assisted assay for angiotensin-converting enzyme (peptidyl dipeptidase) in serum. *Clin. Chem*., **25**, 1259–1262.
20. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

21. Rigat, B., Hubert, C., Corvol, P. and Soubrier, F. (1992) PCR detection of the insertion/deletion of the human angiotensin converting enzyme gene (DCP1) (dipeptidyl carboxypeptidase 1). *Nucleic Acids Res*., **20**, 1433.
22. Smith, C.A.B. (1957) Counting methods in genetical statistics. *Ann. Hum. Genet*., **21**, 254–276.
23. Dempster, A.P., Laird, N.M. and Rubin, D.R. (1977) Maximum likelihood from incomplete data via the EM algorithm. *J. R. Statist. Soc. B*, **39**, 1–38.
24. Excoffier, L. and Statkin, M. (1995) Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol. Biol. Evol*., **12**, 921–927.
25. Badsberg, J.H. (1995) An environment for graphical models. PhD thesis, Aalborg University, Denmark.
26. Hasstedt, S.J. (1993) Variance components/major locus likelihood approximation for quantitative, polychotomous and multivariate data. *Genet. Epidemiol*., **10**, 145–158.
27. Hasstedt, S.J. (1994) Pedigree Analysis Package revision 4.0. Department of Human Genetics, University of Utah, Salt Lake City, UT.