European Journal of Human Genetics (2002) 10, 553–561

www.nature.com/ejhg

ARTICLE

# High-resolution genetic mapping of the ACE-linked QTL influencing circulating ACE activity

Florent Soubrier*,[1], Sabrina Martin[1], Amalia Alonso[1], Sophie Visvikis[1], Laurence Tiret[1], Fumihiko Matsuda[2], G Mark Lathrop[2] and Martin Farrall[3]

[1]*INSERM U 525, Faculté de médecine Pitié-Salpêtrière, Université Paris VI, 91 Bd de l'hôpital, 75013 Paris, France;* [2]*Centre National de Génotypage, 2 rue Gaston Crémieux, CP 5721, 91057 Evry Cedex, France;* [3]*Department of Cardiovascular Medicine, University of Oxford, The Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK*

**Fine-mapping of trait loci through combined linkage and association analysis is an important component of strategies designed to identify causative gene variants, particularly in situations where the trait may be influenced by one or more of many polymorphisms within the same gene. Angiotensin-1 converting enzyme (ACE) provides one of the best models for developing and testing such methodologies, as a major fraction of the heritable variation in the activity of the angiotensin-1 converting enzyme (ACE) is tightly linked to the ACE gene. Moreover, ACE contains many frequent polymorphisms that are in strong linkage disequilibrium with each other. Although none of these variants induces a significant amino-acid change, one or more, either singly or in combination, are likely to have a strong effect on the quantitative phenotype. Here, we show that measured-haplotype analysis of SNP data from a large European family cohort can be used to localise the major ACE-linked genetic factors influencing the trait to a 16 kb interval within the gene, thus limiting the number of ACE variants that need to be considered in future studies designed to elucidate their biological effects. The approaches developed will be applicable to the fine-mapping of other quantitative trait loci in humans.**
*European Journal of Human Genetics* (2002) **10,** 553–561. doi:10.1038/sj.ejhg.5200847

**Keywords:** haplotype mapping; QTL; angiotensin-1 converting enzyme

## Introduction

The angiotensin-1 converting enzyme (ACE) (or dipeptidyl carboxypeptidase 1) is a component of the renin-angiotensin system (RAS) with a major role in the maintenance of salt and water homeostasis and vascular smooth muscle tone. Circulating ACE activity is a highly heritable quantitative trait,[1] and a major QTL has been shown in linkage and gene-association studies to map within or close to the ACE gene itself.[2–4] Many frequent bi-allelic polymorphisms that are to be associated with high/low values of the quantitative trait have been identified within and flanking the gene. The trait-associated polymorphisms are in strong linkage disequilibrium, and none produce significant amino-acid changes. Thus, the identification of the variant or variants that cause the differences in ACE activity is a challenging problem that requires extensive genetic analysis.

These observations provide an unrivalled opportunity to explore high-resolution genetic mapping studies of a human QTL. To this end, we have developed statistical tools to assign quantitative effects to extended haplotypes as they segregate through families (measured haplotype analysis) and have applied them to the ACE-mapping problem.[5] In a study of British families of Caucasian descent, we investigated the haplotype diversity in the ACE gene. Two frequent haplotypes (denoted A and B) were found with frequencies of 43 and 31% respectively. These haplotypes were contrasting, ie carried different alleles, at all the sites with frequent polymorphisms (minor allele

*Correspondence: F Soubrier, INSERM U 525, Faculté de médecine Pitié-Salpêtrière, Université Paris VI, 91 Bd de l'hôpital, 75013 Paris, France;
Tel: +33 1 40779725; Fax: +33 1 40779728;
E-mail: florent.soubrier@chups.jussieu.fr
Received 21 February 2002; revised 8 May 2002; accepted 14 May 2002

with frequency >34%) within a 28 kb region containing the ACE gene. The A and B haplotypes are also associated with contrasting values of circulating ACE activity (low and high respectively). The third most common haplotype, C, (frequency of 16%) was found to resemble the A haplotype at polymorphic sites in the 5′ region of the gene, and to resemble the B haplotype for polymorphic sites in the 3′ region. Farrall *et al*.[6] proposed a simple cladistic structure pivoted on an ancestral recombination event between haplotypes A and B to explain the origin of haplotype C. Sequencing studies refined the localisation of the recombination to an intron 5–exon 8 interval.[6] Measured haplotype analysis demonstrated that the circulating ACE activity associated with the recombinant clade C was very similar to that associated with clade B (high) both of which were distinct from clade A (low). This suggests that the major variant or variants implicated the control of the trait map downstream of the breakpoint.

Sequencing and SNP detection studies have identified additional polymorphisms that map downstream of the ACE transcribed sequences to supplement those characterised in previous studies. Here, we report an extended haplotype analysis with additional variants characterised in two collections of French families of Caucasian origin. One of the haplotypes identified with the new markers is consistent with a second ancestral recombination breakpoint approximately 16 kb downstream from the intron 5–exon 8 breakpoint. Measured haplotype analysis shows that majority of the quantitative variants that influence ACE activity map to the interval flanked by the two ancestral recombination breakpoints. QTL mapping to delimit further the placement of the major causative variants within this interval is shown to be impractical in population of European ancestry due to insufficient haplotype diversity in the critical region.

## Methods

### Families

Volunteers for the study were given a systematic medical examination at the Center for Preventive Medicine of Vandoeuvre–Lès-Nancy (France). The two cohorts, RFSK and STAN, were ascertained using the same criteria (both parents and at least two children in apparent good health and willing to participate in the study) but with recruitment separated by a 6-year interval (Table 1). All participants gave informed consent. Further descriptions of the cohorts are provided in Tiret *et al*.[7] and Garner *et al*.[8]

### ACE trait

Plasma ACE concentration was measured on FAPGG substrate as described by Beneteau *et al*.[9] For statistical analysis, the trait data were transformed to z-scores, separately in each cohort, by standardising to zero mean and unit variance separately for each sex and generation (parent and child) combination.

### Polymorphism detection and genotyping

Polymorphic sites were identified by sequencing the whole ACE gene (from 11 kb in the 5′ region, the transcribed region and 3.5 kb in the 3′ region) in a panel of eight unrelated parents from the RFSK cohort that were chosen on the basis of their plasma ACE levels. Four of the selected individuals (four males) had high ACE levels ($1.5 \leqslant z \leqslant 2.3$) while the other four (four males) had low ACE levels ($-2.1 \leqslant z \leqslant -1.2$). The complete cohorts were subsequently genotyped at a subset of the polymorphic sites by direct sequencing (Table 2), or by other techniques as previously described.[4]

The data were verified for Mendelian transmission and absence of apparent recombination events within families. Two microsatellite markers, D17S807 and D17S948 that flank the ACE gene were characterised to aid in verifying the data. The raw sequence data were re-analysed and sequencing was repeated when necessary.

The variant sites and the primers used for sequencing are reported in Table 2a and b respectively. Positions of the polymorphism site relative to the ACE intron/exon structure are depicted in Figure 1. A complete, reference ACE gene sequence, and the list of all polymorphisms found in a series of Caucasians and African subjects are available on the web site http://genecanvas.idf.inserm.fr/ (which is used to assign the current position labels) (Table 2a). For orientation with previous published studies of ACE variation, alternative polymorphism labels are also tabulated.[2,10]

### Statistical analysis

Extensive catalogues of ACE polymorphisms have been compiled following systematic surveys of ACE genetic variation. Strong linkage disequilibrium (LD) is evident between polymorphisms and it is important to model this accurately in high-resolution ACE QTL mapping studies. The selection of markers for haplotype-based analysis needs to balance on one hand the desire to model a representative proportion of ACE genetic variation, and on the other computational limitations due to uncertainties regarding phase assignment and incomplete joint genotypes. For exact likelihood solutions, computational complexity increases exponentially with the number of loci being considered and the best contemporary algorithms can analyse perhaps no more than eight or nine loci simultaneously. For small families, methods such as Genehunter[11] that are built on a hidden Markov model are capable of analysing scores of markers simultaneously; however, this method is inherently unable to model linkage disequilibrium between markers which

**Table 1**

| | *No. of families* | *No. of genotyped individuals* |
|---|---|---|
| RFSK | 102 | 416 |
| STAN | 151 | 620 |

**Table 2a** Numbering and positions of polymorphisms in the ACE gene

| *Actual position* | *Previous number (Villard* et al.*[4])* | *Rieder's position (Rieder* et al.*[10])* | *Primers used for PCR* |
|---|---|---|---|
| C2970T | | | 5 AB-7.9 |
| T7715C | T-3892C | | (see Villard *et al*, 1996) |
| A11377T | A-240T | 2400 | (see Villard *et al*, 1996) |
| T22706A | | 13730 | 43–44 |
| A22906G | | 13930 | 43–44 |
| 23073 | I/D | 14095 | 45–46 |
| G23454C | | 14480 | 45–46 |
| A23462C | | 14488 | 45–46 |
| A23495G | G2350A | 14521 | 45–46 |
| A27888G | | 18912 | 65.3–66.3 |
| C29809T | | 20833 | 73–74 |
| C31226T | | 22251 | 77–78 (77.2)* |
| A31958G | | 22982 | 81–82 |
| (CT)2/3-32915 | $4656(CT)_{2/3}$ | 23945 | 85–86 |
| T33569C | | | 89–90 |
| G33709A | | | 89–90 |
| 33797 insA | | | 89–90 |
| T33889C | | | 89–90 |

*Indicates the primer used for the sequencing reaction.

**Table 2b**

| *Primer sequences* | |
|---|---|
| 5AB | 5′-AGCCTATGTTGCTGTTTTTAC-3′ |
| 7.9 | 5′-GACGGTGATGATGCCAATGAG-3′ |
| 43.0 | 5′-GTGAGTAGATCACTTGAGG-3′ |
| 44.0 | 5′-AATGGGAGAAAGGATGGG-3′ |
| 45.0 | 5′-GGAGAGAGACTCAAGCAC-3′ |
| 46.0 | 5′-GGGTAAAACTGGAGGATGG-3′ |
| 65.3 | 5′-AGGATTCACCTGAGCCAAAA-3′ |
| 66.3 | 5′-TCAGCAGGTTGAGACTGTGC-3′ |
| 73.0 | 5′-ATCGGTTAAGGTGGGAGGCAGG-3′ |
| 74.0 | 5′-CTTGCTGGGCTTCCAGGAAC-3′ |
| 77.0 | 5′-GCAAAGGAGACTTGGCTG-3′ |
| 78.0 | 5′-TCGTGGAACTGGAACTGG-3′ |
| 81.0 | 5′-CATGTTGAGCTACTTCAAGCC-3′ |
| 82.0 | 5′-CCAGTGTTCCCATCCCAG-3′ |
| 85.0 | 5′-CCTGTCTTCCCTCCAGCC-3′ |
| 86.0 | 5′-CCCTCAAAATGACACTCCCTCC-3′ |
| 89.0 | 5′-TCCCAATATCTGTCCCCC-3′ |
| 90.0 | 5′-GCCAACATGATTAAACCCC-3′ |
| 77.2 | 5′-GGCTCAGACAATGCTAAGAG-3′ |

destroys the Markov property. In an attempt to circumvent these problems, we have devised an approximate method to analyse sufficiently long haplotypes to ensure adequate representation of genetic diversity and to incorporate as much phase information and allowance for incomplete genotyping as practically possible.

The measured haplotype analysis method follows three stages; the following description includes a number of modifications and improvements to our initial application of this approach.[5]

Stage 1: genotypes are occasionally incomplete, this problem increases with the number of loci under consideration and omission of incomplete genotypes would be wasteful of information. The UNKNOWN programme from the LINKAGE package[12] is used to recursively assign missing genotypes based on available information for that marker as it segregates through the rest of the family; a check for intragenic recombination is made using the Genehunter programme.[11] Non-recombinant sets are retained for further analysis (stages 2 and 3).

Stage 2: For extended haplotype analyses, a list of alternative non-recombinant phase-assignments is compiled using the SIMWALK programme[13] to obtain a random sample from the total space of phase-assignments for each family. The phase-assignment list was processed using a customised EM algorithm[5,14] to derive maximum likelihood estimates of the haplotype frequencies. SIMWALK was repeatedly run until the sorted list of *common* haplotype frequencies (frequencies $>10^{-8}$) converged. Alternatively, for relatively short haplotype analyses (eg three polymorphisms), the SIMWALK phase-sampling analysis can be bypassed by simply compiling a list of all possible haplotypes.

Stage 3: The output from stages 1 and 2 are processed by a new programme HAPLOTRY which exhaustively compiles alternative phase-assignments using the list of *common* haplotypes for extended haplotype analyses or the complete list for short haplotypes. A final check for any exceptional mendelian inconsistencies was made using the PedCheck programme.[15]

Stage 4: the data are now ready to enter a modified version of the Pedigree Analysis Package (PAP)[16,17] to undertake a measured haplotype analysis in which a co-dominant QTL model is fitted to measure the influence of each haplotype on circulating ACE activity. Likelihoods were maximised numerically using the GEMINI quasi-Newtonian method,[18] which also permits the calculation of standard

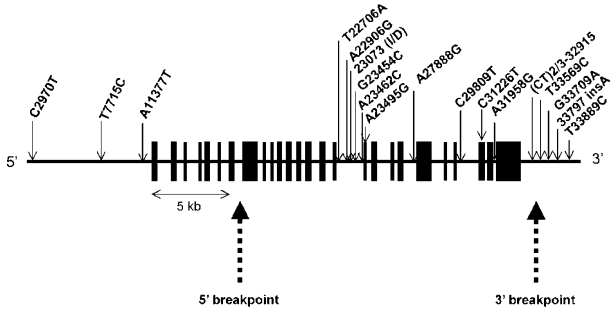


**Figure 1** Schematic representation of 31 kb of ACE outlining the intron–exon structure and location of 18 polymorphic sites identified in the sequencing experiment. Exons are shown by broad solid vertical lines. The sites indicated in bold are those potentially associated with the ACE trait based on the contrast of genotypes in high/low individuals. The locations of two ancestral recombination breakpoints are indicated by dashed arrows.

errors of parameter estimates. Residual intra-familial correlations were accounted for by means of a class D regressive model[19] to include spouse–spouse, mother–offspring, father–offspring and sib–sib correlations. The goodness-of-fit of nested models was assessed by calculating the $-2$ natural logarithm of the likelihood difference and comparing this test statistic with a chi-square distribution with degrees of freedom equal to the difference in the numbers of parameters being evaluated in the general and nested models.

## Results

### Sequence and extended haplotype analysis

A total of 62 polymorphic sites were identified in the gene through comparison of sequence data on four Caucasian individuals with high trait values and four individuals with low trait values (see http://genecanvas.idf.inserm.fr/ for complete data). Initially, we selected seventeen sites to examine in the whole RFSK cohort (Table 3a). Most of these are downstream of the putative haplotype breakpoint identified by Farrall *et al.*,[6] where the principal variant or variants influencing the trait were shown to reside. A smaller number are upstream of the breakpoint; these were included in order to identify recombinant haplotype classes, and to test for possible minor effects of variation in the 5′ region of the gene on the trait. Based on the analysis of the RFSK data, 14 polymorphisms were chosen for characterisation in the STAN cohort, of which 13 were common to both cohorts (Table 3b).

Statistical analysis led to identification of 15 haplotypes with frequency estimates $>1\%$ for the 17 loci characterised in the RFSK cohort, and 12 haplotypes with frequency estimate $>1\%$ for the 14 loci characterised in the STAN cohort (Table 3). The two most frequent haplotypes found in each cohort are completely divergent ie they carry different allelic variants at each of the polymorphic sites. This distinctive pattern was noted in our previous analysis of ACE haplotypes in a British family cohort.[5] Using the classification scheme proposed by these authors, haplotypes 1170 (STAN) and 5249 (RFSK) are assigned to clade A, while haplotypes 15215 (STAN) and 125824 (RFSK) are assigned to clade B. As shown in Table 3, other haplotypes that differ by substitutions at one or two sites from these predominant haplotypes are also categorised into the same clades. The frequency estimates for clades A and B are broadly comparable in the RFSK (41 and 39%), STAN (50 and 43%) and UK families (48 and 34%).

The third most frequent haplotype in the RFSK or STAN cohorts (2927 – STAN and 11136 – RFSK) carries the same alleles as clade A haplotypes for the three (RFSK) or two (STAN) most 5′ marker sites characterised, and the same alleles as clade B haplotypes for the remaining 14 (RFSK) or 12 (STAN) marker sites. (The differences in the numbers shared in RFSK and STAN is due to the choice of polymorphic sites that were examined in the two cohorts.) These patterns are consistent with the results from British families described in Farrall *et al.*,[6] in which it was proposed that this pattern is representative of a third clade (C), that could have arisen from an ancestral recombination with a breakpoint localised between intron 5 and exon 8. Haplotypes 2927 and 11136 are also members of the recombinant clade C. As with clades A and B, the frequency estimates for clade C haplotypes are similar in the two cohorts studied here and the UK cohort: 18% (RFSK), 13% (STAN) and 18% (UK).

The additional markers from the 3′ region of the gene that have been genotyped in the RFSK and STAN families reveal a fourth haplotype clade of interest. Haplotypes 15202 (STAN) and 125817 (RFSK) carry the same marker alleles as clade B haplotypes at 32915 and other polymorphic sites that are 5′ of this (14 sites in RFSK and 10 sites in STAN), and the same marker alleles as clade A haplotypes at 33709 (RFSK) or 33569 (STAN) and other 3′ sites (three in RFSK and four in STAN). This pattern is consistent with an ancestral recombination breakpoint located in the 655 bp interval, flanked by the markers 32915 and 33569. We will refer to this newly identified recombinant clade as clade D. Table 3a (RFSK analysis) and 3b (STAN analysis) include provisional assignments of the other haplotypes to clades A–C; each haplotype differs by no more than two sites from the most frequent clade member.

### Measured haplotype analysis: extended haplotypes

Computational constraints limit the number of haplotypes that can be expediently incorporated into the likelihood calculations in the measured haplotype analysis, and estimates of the mean trait values associated with rare haplotypes have large standard errors. Consequently, relatively rare haplotypes ($<1\%$) were pooled into a single class leaving 12 and 15 haplotype classes respectively for

**Table 3a** 17-locus haplotype analysis in the RFSK cohort

| *Haplotype* | *2970 1* | *7715 2* | *11377 3* | *22706 4* | *22906 5* | *23073 6* | *23454 7* | *23462 8* | *23495 9* | *27888 10* | *29809 11* | *31226 12* | *31958 13* | *32915 14* | *33709 15* | *33797 16* | *33889 17* | *Frequency* | *Clade* | *Mean* | *s.e.* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 125824 | T | C | T | T | A | D | G | A | G | A | T | T | G | $(CT)_2$ | G | 2A | C | 0.25 | B | 0.45 | 0.05 |
| 5249 | C | T | A | A | G | I | C | C | A | G | C | C | A | $(CT)_3$ | A | A | T | 0.22 | A | −0.50 | 0.06 |
| 11136 | C | T | A | T | A | D | G | A | G | A | T | T | G | $(CT)_2$ | G | 2A | C | 0.09 | C | 0.27 | 0.10 |
| 13441 | C | T | A | T | G | I | C | C | A | G | C | C | A | $(CT)_3$ | A | A | T | 0.04 | A | −0.45 | 0.15 |
| 9345 | C | T | A | T | A | I | C | C | A | G | C | C | A | $(CT)_3$ | A | A | T | 0.04 | A | −0.58 | 0.14 |
| 43904 | C | C | A | T | A | D | G | A | G | A | T | T | G | $(CT)_2$ | G | 2A | C | 0.03 | C | 0.24 | 0.15 |
| 125817 | T | C | T | T | A | D | G | A | G | A | T | T | G | $(CT)_2$ | A | A | T | 0.02 | D | 0.34 | 0.21 |
| 129920 | T | C | T | T | G | D | G | A | G | A | T | T | G | $(CT)_2$ | G | 2A | C | 0.01 | B | 0.50 | 0.28 |
| 11104 | C | T | A | T | A | D | G | A | G | A | T | C | G | $(CT)_2$ | G | 2A | C | 0.01 | C | 0.21 | 0.27 |
| 126848 | T | C | T | T | A | D | C | A | G | A | T | T | G | $(CT)_2$ | G | 2A | C | 0.01 | B | 0.05 | 0.22 |
| 125823 | T | C | T | T | A | D | G | A | G | A | T | T | G | $(CT)_2$ | G | 2A | T | 0.01 | B | 0.69 | 0.28 |
| 5329 | C | T | A | A | G | I | C | C | A | G | T | C | G | $(CT)_3$ | A | A | T | 0.01 | A | 0.19 | 0.31 |
| 125760 | T | C | T | T | A | D | G | A | G | A | C | T | G | $(CT)_2$ | G | 2A | C | 0.01 | B | 0.99 | 0.40 |
| 121728 | T | C | T | A | G | D | G | A | G | A | T | T | G | $(CT)_2$ | G | 2A | C | 0.01 | B | 0.16 | 0.34 |
| 60288 | C | C | T | T | A | D | G | A | G | A | T | T | G | $(CT)_2$ | G | 2A | C | 0.01 | B | 0.71 | 0.39 |

**Table 3b** 14-locus haplotype analysis in the STAN cohort

| *Haplotype* | *2970 1* | *7715 2* | *22706 3* | *22906 4* | *23073 5* | *23454 6* | *23462 7* | *23495 8* | *31958 9* | *32915 10* | *33569 11* | *33709 12* | *33797 13* | *33889 14* | *Frequency* | *Clade* | *Mean* | *s.e.* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1170 | C | T | A | G | I | C | C | A | A | $(CT)_3$ | C | A | A | T | 0.29 | A | −0.49 | 0.05 |
| 15215 | T | C | T | A | D | G | A | G | G | $(CT)_2$ | T | G | 2A | C | 0.23 | B | 0.45 | 0.07 |
| 2927 | C | T | T | A | D | G | A | G | G | $(CT)_2$ | T | G | 2A | C | 0.07 | C | 0.23 | 0.12 |
| 15207 | T | C | T | A | D | G | A | G | G | $(CT)_2$ | C | G | 2A | C | 0.05 | B | 0.75 | 0.15 |
| 7023 | C | C | T | A | D | G | A | G | G | $(CT)_2$ | T | G | 2A | C | 0.04 | B | −0.01 | 0.18 |
| 2919 | C | T | T | A | D | G | A | G | G | $(CT)_2$ | C | G | 2A | C | 0.04 | C | 0.12 | 0.15 |
| 15202 | T | C | T | A | D | G | A | G | G | $(CT)_2$ | C | A | A | T | 0.02 | D | 0.18 | 0.32 |
| 7015 | C | C | T | A | D | G | A | G | G | $(CT)_2$ | C | G | 2A | C | 0.02 | B | −0.01 | 0.30 |
| 1169 | C | T | A | G | I | C | C | A | A | $(CT)_3$ | C | A | A | C | 0.02 | A | −1.07 | 0.41 |
| 1174 | C | T | A | G | I | C | C | A | A | $(CT)_3$ | C | G | A | T | 0.01 | A | −0.41 | 0.31 |
| 15208 | T | C | T | A | D | G | A | G | G | $(CT)_2$ | C | G | 2A | T | 0.01 | B | 0.93 | 0.40 |
| 5266 | C | C | A | G | I | C | C | A | A | $(CT)_3$ | C | A | A | T | 0.01 | A | −0.44 | 0.28 |

Clade denotes the provisional clade assignment. Mean indicates the maximum likelihood estimate of the additive haplotype effect on the trait. s.e. denotes the standard error of this estimate.

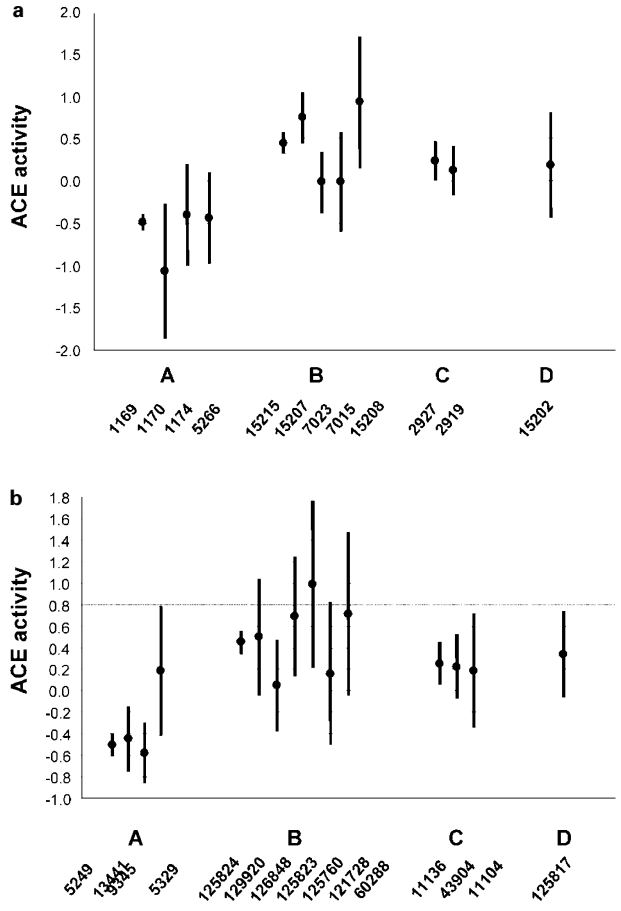


**Figure 2** (**a**) Measured haplotype analysis of 17-markers in the RFSK series of families. Individual haplotypes are identified on the abscissa with oblique integer labels and are grouped into four clades (**A**–**D**). A solid circle marks the mean ACE activity estimated for each haplotype, vertical lines indicate the 95% confidence interval for these estimates. (**b**) Measured haplotype analysis of 14-markers in the STAN series of families.

STAN and RFSK. The pooled haplotype classes represented $<22\%$ of the total in each cohort.

The means and standard deviations of the ACE trait in each of the haplotype classes as estimated in the measured haplotype analysis under a general model of residual familial correlation (see Methods) are given in Table 3a,b. Figure 2 shows the 95% confidence intervals for the estimates of the means under an additive model, assuming identity of the haplotype mean within each clade. The latter hypothesis was not rejected when compared to a model allowing for different means for each haplotype (RFSK, $P$=0.28 and STAN $P$=0.052). In the STAN cohort, residual intra-familial correlations were significant ($P$=0.01), and they were compatible with an additive model ($P$=0.54) with a common parent–offspring and sib–sib correlation of $0.182 \pm 0.054$; the spouse–spouse correlation was non-significant. In the RFSK families, there was no significant evidence for residual familial correlations ($P$=0.29).

The 4-clade model as described above was used to examine difference between the trait means for clades C and D, hypothesised to have been formed by intragenic recombination, and the trait means for clades A and B, the putative ancestral clades. In both STAN and RFSK, C and D are associated with a significantly higher ACE activity than A (RFSK: $P=6\times10^{-12}$ and $P$=0.0002 respectively; STAN: $P=2\times10^{-9}$ and $P$=0.036). Although C also has a slightly lower mean ACE activity than B in both cohorts, the difference is not statistically significant in the RFSK cohort ($P$=0.08) and only marginally significant in STAN ($P$=0.024). The mean ACE activity of D closely resembles that of B in both cohorts (STAN: $P$=0.60; RFSK: $P$=0.64).

**Measured haplotype analysis: reduced haplotypes**

As there appeared to be no significant variation in the effects on the ACE trait for different haplotypes that were grouped together within the major clades, we selected three polymorphisms to summarise the cladistic structure of the haplotype variation (STAN: 7715, 23495 and 33797; RFSK: 11377, 23495 and 33709). Because of the limited number of haplotypes in the 3-locus analysis (eight haplotypes), it was simple to enumerate all possible haplotype combinations in each family, and the incorporation of a MCMC procedure was unnecessary (see Methods). Data from RFSK and STAN could be combined based on clade assignment even though different markers were analysed in each cohort.

The conclusions of the 3-locus analysis undertaken separately in each cohort were generally similar to those from the extended haplotype analysis as summarised in Table 4a,b. However, in the 3-locus analyses, clade C was associated with a significantly lower mean ACE activity than that of clade B in both cohorts (RFSK: $P$=0.004; STAN: $P$=0.008), whereas the mean associated with clade D remained statistically indistinguishable from that of clade B (RFSK: $P$=0.58; STAN: $P$=1.00). The residual familial correlations in the 3-locus analysis were also marginally significant in the RFSK cohort ($P$=0.05) as well as in STAN ($P$=0.001).

Finally, a joint analysis of the STAN and RFSK data showed that there was no evidence for heterogeneity between the samples ($P$=0.34). Figure 3 shows the estimated clade means and standard deviations obtained from the joint analysis. Clade C is associated with a significantly lower effect on ACE activity than clade B ($P$=0.0001) whereas clade D is indistinguishable from clade B ($P$=0.74). Residual intra-familial correlations were significant ($P=3\times10^{-5}$) and compatible with an additive pattern (parent–offspring and sib–sib correlation estimated as $0.164 \pm 0.035$). The spouse–spouse trait correlation was not significantly different from 0.

**Table 4a** 3-locus haplotype analysis of markers 11377, 23495 and 33709 in the RFSK cohort

| *Haplotype* | *Clade* | *Frequency* | *Mean* | *s.e.* |
|---|---|---|---|---|
| A-A-A | A | 0.39 | −0.46 | 0.04 |
| A-A-G | – | 0.03 | −0.35 | 0.20 |
| A-G-A | – | 0.02 | −0.43 | 0.24 |
| A-G-G | C | 0.18 | 0.19 | 0.07 |
| T-A-A | – | 0.01 | −0.43 | 0.32 |
| T-A-G | – | 0.00 | – | – |
| T-G-A | D | 0.03 | 0.35 | 0.18 |
| T-G-G | B | 0.35 | 0.45 | 0.05 |

**Table 4b** 3-locus haplotype analysis of markers 7715, 23495 and 33797 in the STAN cohort

| *Haplotype* | *Clade* | *Frequency* | *Mean* | *s.e.* |
|---|---|---|---|---|
| T-A-A | A | 0.36 | −0.50 | 0.05 |
| T-A-2A | – | 0.02 | 0.09 | 0.21 |
| T-G-A | – | 0.01 | −0.33 | 0.26 |
| T-G-2A | C | 0.12 | 0.14 | 0.09 |
| C-A-A | – | 0.03 | 0.12 | 0.17 |
| C-A-2A | – | 0.01 | −0.41 | 0.31 |
| G-G-A | D | 0.04 | 0.42 | 0.17 |
| C-G-2A | B | 0.40 | 0.42 | 0.05 |

Clade denotes the provisional clade assignment. Mean indicates the maximum likelihood estimate of the additive haplotype effect on the trait. s.e. denotes the standard error of this estimate.

## Discussion

Systematic, exhaustive sequencing of the ACE gene in multiple individuals covering the transcribed region, 11 kb in the 5′ region and 3.5 kb in the 3′ region has identified all the common variants of ACE gene. At least 33 of these polymorphisms are strongly associated with circulating ACE activity in European populations (present study;[4,5,6]). It is likely that one, or a combination of these is responsible for up to 40% of the trait variance that maps to the ACE locus. Because of the systematic nature of the sequencing, all the variants within the gene that could account for this genetic effect have been identified (although, at least theoretically, additional causative polymorphisms in strong linkage disequilbrium with the described polymorphisms might be discovered outside of the sequenced segment). Interestingly, none of these polymorphisms has a predicted biological effect based on the type of substitution or other features of its location. Despite the strong linkage disequilibrium, high-resolution genetic analysis can provide important information contributing to the identification of the specific region of the gene and a subset of variants that includes those with the principal effects on the phenotype. Such genetic analysis is an important prerequisite for undertaking focused investigations of the potential functional consequences of these variants, which may be both time-consuming and difficult to interpret.

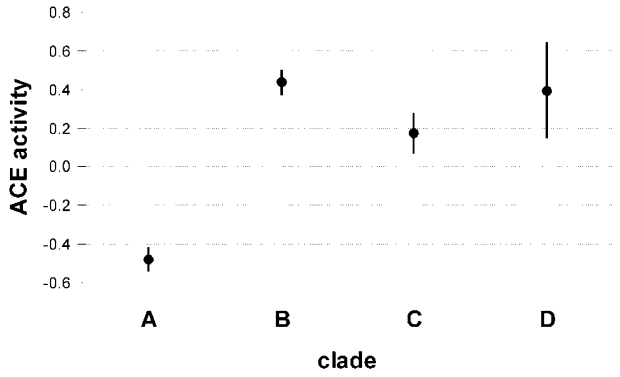


**Figure 3** Measured haplotype analysis of the combined family cohorts using markers that distinguish the four major haplotype clades. A solid circle marks the mean ACE activity estimated for each clade, vertical lines indicate the 95% confidence interval for these estimates.

DNA sequencing studies have previously localised an ACE ancestral recombination breakpoint in Europeans to the intron 5 – exon 8 interval.[6] We have extended our analysis of ACE haplotypes to include 3′ non-transcribed sequences and have identified several new polymorphisms that are in strong linkage disequilibrium with upstream polymorphisms. This analysis has revealed a second ancestral recombination event that localises to a 655 bp interval flanked by the markers 32915 and 33569.

In order to map further the association of ACE variation with circulating ACE activity, we have developed a statistical method to estimate the influences of individual ACE haplotypes segregating in families. We have undertaken two sets of analyses of the RFSK and STAN family cohorts: first, an analysis of extended haplotype involving 14 or 17 polymorphisms distributed across 31 kb of genomic sequence, and second, a reduced haplotype analysis of three polymorphisms that were selected to summarise the cladistic structure of the ACE haplotypes. In both analyses, the estimate of the mean ACE activity associated with the recombinant clade C most closely resembled that of clade B, and it was substantially higher than that estimated for clade A. Interestingly, clade C is associated with slightly lower mean ACE activities than clade B; this trend reaches statistical significance in the combined RFSK and STAN data. Clade D was also associated with substantially higher ACE activity than clade A, and it closely resembled clade B.

We interpret these results to indicate that majority of the genetic variation in ACE that is associated with ACE activity maps to the interval flanked by the two ancestral recombination breakpoints that define clades C and D (Figure 1). This interval includes the '22982' (31958) polymorphism that was identified as most strongly associated with ACE trait variation in a population-based study of Afro-Carribeans.[20] However, because of the strong linkage disequilibrium and consequent lack of haplotype diversity,

it appears that we cannot distinguish further which variant or combination of variants from the interval affects the phenotype in European populations. Based on the lower mean of C compared to B, and the lack of difference between D and B, it also appears that a significant even if (relatively) minor portion of the genetic factors affected the trait variance map upstream of the clade C breakpoint. This is consistent with our previous measured haplotype analysis in British families[5] where the same trend was observed for a lower, but statistically non-significant, mean trait value of C compared to B clades. It also appears to be consistent with the arguments of Villard *et al*.[4] based on analysis of ACE genotypes in the RFSK cohort, which partially overlap with the markers studied here and the analysis of Zhu *et al*.[21] in Africans.

The support for residual intra-familial correlations affected circulating ACE activity, but unlinked to the ACE locus, was obtained in both French cohorts. In the analysis of the combined data, this accounted for 10.5% of the total variance and 22% of the genetic variance of the trait locus. A previous analysis of white British families estimated that residual familial correlations accounted for 10% of the total variance and 22% of the genetic variance in UK families.[5] Together, these results support the notion that an unlinked gene or genes may contribute to the heritability of the ACE trait, which was proposed in a two-QTL linkage analysis of Afro-Caribbean families.[3]

Previous studies have shown that the ACE polymorphism is associated with changes in the levels of circulating immunoreactive enzyme molecules, favouring a modulation of gene expression and not structural differences modifying enzyme kinetics.[22] Common ACE genetic variation affects the plasma enzyme, which corresponds to the secreted form of the enzyme, and the intracellular or membrane-anchored enzyme.[23] In different rat strains, it has been demonstrated that the levels of lung ACE gene transcription is altered according to the ACE genotype, suggesting that variation of plasma ACE levels is the consequence of differing levels of transcription in endothelial cells which represent the main source of plasma ACE.[24] It is likely that a transcriptional mechanism is also at play in the human common variation of plasma ACE, although the transcriptional impact of the region of the ACE gene targeted by our genetic study (between intron 5 and downstream exon 26) needs to be documented. The number of candidate polymorphisms in the region is too high to envision an experimental strategy for testing each of them. Also, testing the internal part of the gene for its influence on the gene expression raises some methodological problems regarding the difference between *in vitro* transfected plasmids and *in vivo* chromatin conformation, which can be responsible for missing transcriptional effects.

Moreover, the mechanisms of genetic elevation of ACE are multiple. Firstly, the region of the ACE gene located upstream of the 5′ recombination breakpoint of the ACE haplotypes is responsible for a portion of ACE variation, and this region contains exons, introns, and the promoter. Secondly, a more rare form of familial elevation of plasma ACE has been shown to be due to a mutation changing a proline to a leucine in the C terminal part of ACE.[25] This mutation is responsible for an increased shedding of ACE due to an increased rate of proteolytic cleavage of the membrane bound enzyme by the still unidentified ACE secretase.

The study of complex and quantitative traits is becoming a focus for contemporary human genetic research and is being assisted by comprehensive surveys of polymorphism. Patterns of linkage disequilibrium in human populations, which have been shaped by demographic forces, will need to be modelled when the relationship between genetic and phenotypic variation is characterised. Haplotype-based approaches such as measured haplotype analysis may be useful in this stratagem.

**Acknowledgements**

*The Centre National de Génotypage is supported by the French Ministry of Research. We thank Pr Gérard Siest from the 'Centre de Médecine Préventive de Vandoeuvre-Lès-Nancy' for his continuous support, Bruno Baudin and Monique Vincent-Viry for ACE measurements.*

**References**

1 Cambien F, Alhenc-Gelas F, Herbeth B *et al*: Familial resemblance of plasma angiotensin-converting enzyme level: The Nancy Study. *Am J Hum Gene,* 1988; **43**: 774–780.
2 Tiret L, Rigat B, Visvikis S *et al*: 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* 1992; **51**: 197–205.
3 McKenzie CA, Julier C, Forrester T *et al*: Segregation and linkage analysis of serum angiotensin I-converting enzyme levels: Evidence for two quantitative-trait loci. *Am J Hum Genet* 1995; **57**: 1426–1435.
4 Villard E, Tiret L, Visvikis S, Rakotovao R, Cambien F, Soubrier F: 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* 1996; **58**: 1268–1278.
5 Keavney B, McKenzie CA, Connell JM *et al*: Measured haplotype analysis of the angiotensin-I converting enzyme gene. *Hum Mol Genet* 1998; **7**: 1745–1751.
6 Farrall M, Keavney B, McKenzie C, Delepine M, Matsuda F, Lathrop GM: Fine-mapping of an ancestral recombination breakpoint in DCP1. *Nat Genet* 1999; **23**: 270–271.
7 Tiret L, Rigat B, Visvikis S *et al*: 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* 1992; **51**: 197–210.
8 Garner C, Lecomte E, Visvikis S *et al*: Genetic and environmental influences on left ventricular mass. A family study. *Hypertension* 2000; **36**: 740–746.
9 Beneteau B, Baudin B, Morgant G, Giboudeau J, Baumann FC: Automated kinetic assay of angiotensin-converting enzyme in serum. *Clin Chem* 1986; **32**: 884–886.
10 Rieder MJ, Taylor SL, Clark AG, Nickerson DA: Sequence variation in the human angiotensin converting enzyme. *Nat Genet* 1999; **22**: 59–62.

11 Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES: Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 1996; **58**: 1347–1363.
12 Lathrop GM, Lalouel JM, Julier C, Ott J: Multilocus linkage analysis in humans: detection of linkage and estimation of recombination. *Am J Hum Genet* 1985; **37**: 482–498.
13 Sobel E, Lange K: Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. *Am J Hum Genet* 1996; **58**: 1323–1337.
14 Dempster AP, Laird NM, Rubin DR: Maximum likelihood from incomplete data via the EM algorithm. *J Roy Stat Soc* 1977; **B39**: 1–38.
15 O'Connell JR, Weeks DE: PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 1998; **63**: 259–266.
16 Hasstedt SJ: Variance components/major locus likelihood approximation for quantitative polychotomous multivariate data. *Genet Epidemiol* 1993; **10**: 145–158.
17 Hasstedt SJ: Pedigree Analysis Package revision 4.0. 1994 Department of Human Genetics, University of Utah, Salt Lake City.
18 Lalouel JM: GEMINI – a computer program for optimization of a nonlinear function. Dept of Medical Biophysics and Computing, University of Utah. *Technical Report* 1979, No. 14.
19 Bonney GE: Compound regressive models for family data. *Hum Hered* 1992; **42**: 28–41.
20 Zhu X, McKenzie CA, Forrester T *et al*: Localization of a small genomic region associated with elevated ACE. *Am J Hum Genet* 2000; **67**: 1144–1153.
21 Zhu X, Bouzekri N, Southam L *et al*: Linkage and association analysis of angiotensin I-converting enzyme (ACE)-gene polymorphisms with ACE concentration and blood pressure. *Am J Hum Genet* 2001; **68**: 1139–1148.
22 Danilov S, Savoie F, Lenoir B *et al*: Development of enzyme-linked immunoassays for human angiotensin I converting enzyme suitable for large-scale studies. *J Hypertens* 1996; **14**: 719–727.
23 Danser AH, Schalekamp MA, Bax WA *et al*: Angiotensin-converting enzyme in the human heart. Effect of the deletion/insertion polymorphism. *Circulation* 1995; **92**: 1387–1388.
24 Challah M, Villard E, Philippe M *et al*: Angiotensin I-converting enzyme genotype influences arterial response to injury in normotensive rats. *Arterioscler Throm Vasc Biol* 1998; **18**: 235–243.
25 Eyries M, Michaud A, Deinum J *et al*: Increased shedding of angiotensin-converting enzyme by a mutation identified in the stalk region. *J Biol Chem* 2001; **276**: 5525–5532.