

Analysis of Genetic Polymorphisms in Acetylcholinesterase as Reflected in Different Populations

Yehudit Hasin^{1,2,*}, Nili Avidan^{1,*}, Dani Bercovich³, Amos D. Korczyn⁴, Israel Silman⁵, Jacques S. Beckmann^{1,6,**} and Joel L. Sussman^{2,**}

¹Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel, ²Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel, ³Department of Molecular Genetics, MIGAL-Galilee Technology Center, Kiryat Shmona 11016, Israel, ⁴Department of Neurology, Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel, ⁵Department of Neurobiology, Weizmann Institute of Science, Rehovot 76100, Israel, ⁶Service of Medical Genetics, Centre Hospitalier Universitaire Vaudois UNIL, CH-1011 Lausanne, Switzerland

Abstract: Acetylcholinesterase (AChE) plays a crucial physiological role in termination of impulse transmission at cholinergic synapses through rapid hydrolysis of acetylcholine. In addition, it was implicated in amyloid plaque formation, a hallmark of Alzheimer's disease (AD), and most of the drugs used in AD treatment are AChE inhibitors. Thus *ACHE* is an obvious candidate gene for pharmacogenetic study of AD treatment. However, AChE is a highly conserved molecule, and only a few naturally occurring genetic polymorphisms have been reported in the human gene. The goals of this study were to make a systematic effort to identify natural single nucleotide polymorphisms (SNPs) in the human *ACHE* gene, and to reveal their population specific architecture. To this end, the genomic coding sequences for AChE of 96 unrelated control individuals from three distinct ethnic groups, African Americans, Ashkenazi Jews and Israeli Arabs, were analyzed. Thirteen *ACHE* SNPs were identified, ten of which are newly described, and five of which should produce amino-acid substitutions (Arg34Gln, Gly57Arg, Glu344Gly, His353Asn and Pro592Arg). Population frequencies of 11 of the 13 SNPs were established in four different populations, African Americans, Ashkenazi Jews, Sephardic Jews and Israeli Arabs; 17 haplotypes and 5 ethno-specific alleles were identified, and a cladogram of *ACHE* haplotypes was constructed. Among the SNPs resulting in an amino-acid substitution, three are within the mature protein, mapping on its external surface; they are thus unlikely to affect its catalytic properties, yet could have antigenic consequences or affect putative protein-protein interactions. Furthermore, the newly identified SNPs open the door to a study of the possible association of AChE with deleterious phenotypes – such as adverse drug responses to AChE inhibitors employed in treatment of AD patients and hypersensitivity to pesticides.

Keywords: Acetylcholinesterase, *ACHE*, single nucleotide polymorphism, SNP, Alzheimer's disease, adverse drug reaction.

INTRODUCTION

Acetylcholinesterase (AChE) catalyses the rapid hydrolysis of acetylcholine (ACh) to acetate and choline, its principal biological role being termination of impulse transmission at cholinergic synapses. In addition, AChE is believed to play non-classical roles in nerve and muscle development and in haematopoiesis [1-3]. Furthermore, AChE has been implicated in Alzheimer's disease (AD) [4, 5], Gulf War syndrome, and hypersensitivity to pesticides [6].

The human *ACHE* gene (*ACHE* MIM#5334740) is composed of six exons. Three distinct AChE polypeptides, resulting from alternate splicing at the C-terminus, yield a repertoire of isoforms [7]. Although the catalytic properties of these isoforms are similar, they differ in quaternary structure and tissue distribution: The AChE-T transcript is the only one expressed in the brains and muscles of adult mammals

[8]; the AChE-H transcript produces glycoposphatidylinositol (GPI)-anchored dimers which, in higher vertebrates, are expressed mainly in embryonic tissue and on the surface of haematopoietic cells [9]; and the AChE-R transcript is a read-through isoform, which arises due to lack of splicing downstream of the common constitutive exons encoding the catalytic domain, resulting in a soluble monomer which is upregulated in the brain during stress [6].

AChE possesses remarkably high catalytic activity, even though its active site resides at the bottom of a deep and narrow gorge [10]. The substrate molecule, which has to navigate a substantial distance to reach the active site, is assisted by an exceptionally strong electric field [11, 12]. AChE is a highly conserved enzyme; there is 88% sequence identity between the human and mouse enzymes, and the crystal structures of the human, mouse, *Torpedo californica* and *Drosophila* enzymes are fundamentally similar [11, 13-16].

A large repertoire of natural and synthetic AChE inhibitors is used worldwide. Some of these serve as chemical warfare agents and pesticides [17], and others in the treatment of AD [4, 18], and myasthenia gravis [19]. The cholinergic hypothesis of AD suggests that degeneration of cho-

**Address correspondence to these authors at the Department of Structural Biology, Weizmann Institute of Science, Rehovot, 76100 Israel; E-mail: Joel.Sussman@weizmann.ac.il or Jacques S. Beckmann, Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel; E-mail: Jacqui.Beckmann@weizmann.ac.il

* These authors contributed equally to this work.

linergic nerve terminals, resulting in a concomitant decrease in ACh levels in brain regions involved in cognition, produces some symptoms of the disease [4]. Inhibition of AChE was thus predicted to provide symptomatic treatment for AD. Indeed, the first generation of AD medications are all cholinesterase inhibitors, including tacrine (Cognex™), donepezil (Aricept™), rivastigmine (Exelon™) and galanthamine (Reminyl™). Yet a significant proportion of treated patients suffer from severe adverse drug reactions (ADR), thus limiting the maximal dose which can be employed, or even forcing suspension of treatment altogether [20, 21].

In view of the fact that AChE is the biological target of nearly all the currently available medications for treatment of AD, and its possible involvement in amyloid plaque formation, *ACHE* is an obvious candidate gene for a pharmacogenetic study of AD. In such a study one would look for association of any of the *ACHE* genetic polymorphisms (SNPs), or their actual combinations (haplotypes), with the efficacy and/or safety of a given drug. Additional value of such a study would stem from the fact that a cohort of AD patients would be genotyped for *ACHE* SNPs. Their frequency in patients could then be readily compared to that in a normal population. Such an analysis could be of use in assessing the association of natural *ACHE* variants with predisposition to AD.

A prerequisite for such a study is an accurate assessment of genetic polymorphisms (SNPs) in the *ACHE* gene present in the normal population(s) in question. Unfortunately, SNP databases have failed to provide such data for the human *ACHE* gene. Currently, 18 SNPs are listed in the dbSNP database, 13 of which lack any biological validation, and for none of which population frequency data are available. Yet some of these SNPs, if proven to be valid, are predicted to be deleterious, either because they result in a truncated, presumably inactive, protein, or result in substitution of a conserved amino acid.

Until now only two studies focused on identification of naturally occurring genetic polymorphisms in the human *ACHE* gene [22,23]. Neither of these studies can be regarded as exhaustive, due to their limitation to one ethnic group or to the small number of individuals screened, as well as to incomplete coverage of the *ACHE* coding region. Nevertheless, four polymorphisms have been reported in the *ACHE* gene, two of which are of clinical relevance. Polymorphism in the distal promoter of the *ACHE* gene disrupts a putative glucocorticoid-response element. This polymorphism appears to be associated with acute sensitivity to anticholinesterase agents such as pesticides [24], and is suspected of being implicated in Gulf War syndrome [6]. A second polymorphism is the substitution of His322 by Asn, and is responsible for the YT-2 blood group phenotype, an important factor for matching donor and recipient blood types (MIM# 112100). The physiological importance of AChE, combined with the low number of naturally occurring genetic polymorphisms detected, led to the widely accepted hypothesis that “almost every mutation in AChE would be deleterious” [6]. This is supported by the fact that most parts of the AChE protein are important for its function.

In the study presented here, we aimed to assess the validity of the *ACHE* SNPs previously reported, to identify new naturally occurring polymorphisms in the *ACHE* gene, to reveal their distribution and haplotype structure in ethnically distinct populations, and to assess their possible influence on AChE protein structure and function. The long-range objective of our study is to assess possible involvement of naturally occurring *ACHE* polymorphisms in disease mechanisms and in differential drug responses.

To this end, we analyzed the genomic *ACHE* coding sequences from 96 unrelated control individuals representing 3 distinct ethnic groups. Unexpectedly, we observed 13 single nucleotide polymorphisms (SNPs) in the *ACHE* gene, 10 of which have not been described previously. Moreover, most of the SNPs listed in the database failed to reproduce in our study. Five of the 13 SNPs we observed result in amino-acid substitutions. Allelic frequencies of 11 SNPs were established in four ethnic groups, demonstrating 3 ethno-specific alleles. Inference of haplotypes from genotype data resulted in identification of 17 haplotypes with clear indication of a recombination point in the *ACHE* gene. Cladistic analysis of *ACHE* haplotypes reveals their evolutionary architecture, and shows an additional 2 alleles to be ethno-specific.

MATERIALS AND METHODS

DNA Samples

DNA samples were obtained from the National Laboratory for the Genetics of Israeli Populations at Tel-Aviv University (<http://www.tau.ac.il/medicine/NLGIP/nlgip.htm>) and from Coriell Cell Repositories, Camden, NJ (<http://locus.umdnj.edu/ccr/>).

PCR Amplification

Fragments covering the entire coding region of the *ACHE* gene (NM_015831.1), except for the DNA fragment corresponding to Pro81-Trp117 (Pro50-Trp86), were amplified from genomic DNA samples of 96 individuals of 3 different ethnic origins (32 African Americans, 32 Ashkenazi Jews and 32 Israeli Arabs). Amplification primers were designed using Oligo software (http://bio.weizmann.ac.il/software/mac_software/mol_biol/oligo.html, 2004), and primer sequences and PCR conditions for each amplicon are summarized in Table 1. Amplification was performed in a 60 µl reaction volume, using HotStart Taq polymerase with Q buffer (Qiagen), under standard cycling conditions (Table 1).

SNP Discovery

For SNP discovery, seven fragments, covering most of the coding region of the *ACHE* gene, were analyzed by denaturing high performance liquid chromatography (DHPLC), using a WAVE DNA fragment analysis system (Transgenomics, Omaha, NE) [25-27] in 96 unrelated individuals (192 chromosomes), viz. 32 African Americans, 32 Ashkenazi Jews, and 32 Israeli Arabs. The PCR products were denatured at 95°C for 5 min, and cooled to 65°C at a temperature gradient of 1°C/min. The samples were kept at 4°C until 5 µl were applied to a preheated C18 reversed-phase column based on non-porous poly (styrene-divinylbenzene) particles (DNA-Sep Cartridge, CAT no. 450181; all DHPLC catalog

Table 1. PCR Primers for Amplification of the ACHE Gene Coding Region

Amplicon*	Coding sequence covered	Primer ID	Sequence (5' to 3')	Conc.** (μM)	Annealing temp.***	Product size
C	exon 2	ACHE_C_r	GTGGCAGAAAGCGACGGG	1	65°C	301bp
		ACHE_C_f	CGACCGACCCTTCACCCT	1		
H	exon 2	ACHE_H_r	TGGAGAAGCCCTCATGCCT	0.5	61°C	811bp
		ACHE_H_f	GGAACCCCAACCGTGAGC	0.5		
G	exons 2 & 3	ACHE_G_r	TCGGGATGCAGCCAGTCT	0.5	61°C	824bp
		ACHE_G_f	GTGCCCAATGGACCCT	0.5		
E	exon3	ACHE_E_r	CTATCTGCCCTGTCCC	0.5	61°C	593bp
		ACHE_E_f	TGGTGGGTGTGGTGAAGG	0.5		
J	exons 4 & 5	ACHE_J_r	AATGTCAGGCTCAGTTCC	1	65°C	723bp
		ACHE_J_f	GAGCTGAAGAGTCCGGGA	1		
I	exon 6	ACHE_I_r	CTGGGGCTCGTCTGTGTTA	1	61°C	744bp
		ACHE_I_f	TCCCTCCTCCTCAAACCG	1		
A	exon 6	ACHE_A_r	CCCCTCCTGCATAGACTC	0.5	65°C	342bp
		ACHE_A_f	GTTCGACCACTACAGCAAGC	0.5		

* Amplicon identification matches Fig. 2.

** Concentration refers to the actual primer concentration in the reaction mixture.

*** Annealing temperature refers to the temperature used for PCR (see text).

numbers are from Transgenomic Inc.). DNA was eluted by use of a linear acetonitrile gradient consisting of buffer A (0.1 M triethylammonium acetate (TEAA) – CAT no. SP5890), and buffer B (0.1 M TEAA, 25% acetonitrile – CAT no. 700001). The temperature at which heteroduplex detection occurred was deduced from the Transgenomic software (Wavemaker 4.2) and the Stanford DHPLC melting program (<http://insertion.stanford.edu/meltdoc.html>), which analyzes the melting profile of the specific DNA fragment.

For resequencing, PCR fragments were re-amplified from the corresponding genomic DNA sample and subjected to direct sequencing using dye terminators. The sequencing reaction was performed at 60°C, and sequencing primers were used every 300bp for both the forward and reverse strands. Sequence comparisons and SNP visualization were performed using the STADEN package [28]. Each SNP was identified in at least two independent PCR amplifications, and appeared in at least two sequencing reactions.

SNP Identifiers and Protein Nomenclature

Each new SNP was assigned a unique identifier of the form: “ACHE:c.#### nucleotide substitution”. The assigned SNP numbers correspond to the cDNA sequence (NM_015831.1), with the A of the ATG translation initiation codon being taken as the +1 base. For example, substitution of G to A at position 101 in ACHE cDNA, is identified as ACHE:c.101G>A.

The intronic variation is designated ACHE_1608+21:G>A, meaning that it occurred in the 21st base of the intron

which in cDNA sequence NM_015831.1 is positioned between nucleotides 1608 and 1609.

The protein nomenclature employed is consistent with the convention of numbering the initiation methionine as the +1 amino acid. However, in the cholinesterase literature numbering usually starts from the first amino acid of the mature protein. In the case of human AChE this is the 32nd residue, Glu. In the following, for the sake of clarity, the cholinesterase residue number is given in brackets after the conventional number.

SNP Genotyping

Population-specific frequencies of the detected SNP alleles were established by genotyping 48 individuals (96 chromosomes) from each of the three ethnic groups used in SNP discovery. These 48 individuals include the original 32 individuals used for SNP discovery. In addition, the SNPs were genotyped in 48 DNAs from Sephardic Jews, which had not been used for SNP discovery.

Genotyping was performed using the Sequenom MassARRAY system, which uses MALDI-TOF mass spectrometry to analyze single nucleotide polymorphisms in amplified DNA fragments. The two alleles of a given SNP are represented by primer extensions of different sizes. The software automatically scores each individual sample for the presence of either or both alleles. Two SNPs, ACHE:c.1031A>G and ACHE:c.1903G>A, although clearly visualized by multiple direct sequencing, failed to undergo extension in the Sequenom assay; consequently, the allele

frequencies for these SNPs in the different populations could not be determined.

Haplotype Inference

ACHE haplotypes were inferred from genotype data using PHASE software version 0.21, available at (www.stats.ox.ac.uk/mathgen/software.html), with a minimum probability of 0.88.

Statistical Analysis

Statistical analysis was performed using version 8 of the SAS program. Analysis of each SNP was tested by a Pearson goodness of fit χ^2 test. Rare alleles were tested by a Fisher exact test. An overall test statistic was computed for each pair of populations by summing the χ^2 test-statistics across loci. Significance levels were obtained from the χ^2 distribution, whose number of degrees of freedom was equal to the sum of the degrees of freedom for the individual tests. Probability values were computed for each SNP, and multiple comparisons were corrected by the False Discovery Rate (FDR) method [29]. Haplotype analysis for each pair of populations was tested using the χ^2 test. A contingency table was constructed with absolute haplotype frequencies for each pair of populations and for each locus. Significance was assigned by a test analogous to Fisher's exact test

Cladistic Analysis

The cladogram was constructed manually. It was guided by adopting the most parsimonious explanation for haplotype divergence, and allowing recombination only if it involved at least one frequent haplotype, *i.e.* Hap0 or Hap1. In the absence of a reference sequence, we assumed the most frequent haplotype to be the most ancient. A recombination hot spot between *ACHE:c.2126G>A* and *ACHE:c.1775C>G* was allowed, since it is supported by gamete combinations of numerous SNPs. Whenever possible, the more frequent SNP was assumed to be the more ancient. The SNPs were assumed to appear independently, one at a time, and recurrent mutations were not allowed. For example, we can look at the system of haplotypes 0, 1, 4 and 12, shown in Fig. 1.

Hap0 is considered to be the ancestral haplotype, since it is the most frequent haplotype in all 4 populations. Hap1 differs from Hap0 by one SNP (*ACHE:c.1775*), Hap12 also differs from Hap0 by one SNP (*ACHE:c.2126*), and Hap4 carries both these SNPs. Theoretically, Hap1 and Hap12 could both have arisen by one SNP from Hap0, in which case Hap4 would require recombination between Hap1 and Hap12. However, Hap4 was observed 11 times, and Hap12 only once. It was, therefore, assumed that the more frequent haplotype, *viz.* Hap4, occurred prior to Hap12 (Fig. 1), and that Hap12 is the recombinant one.

RESULTS

Using DHPLC and direct sequencing, we identified 13 *ACHE* SNPs (Fig. 2). Only 3 of them match the database records, and 10 have not been reported previously. Population-specific frequencies of 11 out of the 13 *ACHE* SNPs were established in four different populations – African Americans, Ashkenazi Jews, Sephardic Jews and Israeli Ar-

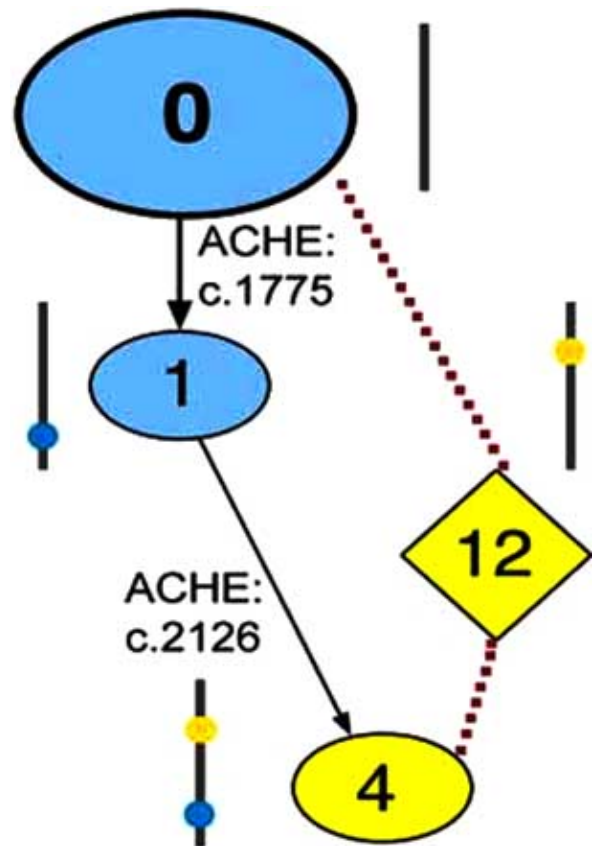


Fig. (1). Principals of cladistic analysis

Construction of the cladogram was guided by adopting the most parsimonious explanation for haplotype divergence, and allowing recombination only if it involved at least one frequent haplotype, *i.e.* Hap0 or Hap1. The most frequent haplotype, Hap0, was assumed to be the most ancient. A recombination spot between *ACHE:c.1775C>G* and *ACHE:c.2126G>A* was assumed. Haplotype number is indicated in the nodes. Appearance of new haplotypes is indicated by directional arrows. The appropriate mutation is indicated on the arrows. Haplotypes that, supposedly, originated by recombination are represented by rectangles, connected to the original haplotypes by dotted lines. Grey lines represent the locus, the blue circle represents *ACHE:c.1775* and the yellow circle represents *ACHE:c.2126*.

abs (Table 2). The other two SNPs, *ACHE:c.1031A>G* and *ACHE:c.1903G>A*, although clearly visualized by multiple direct sequencing, failed to undergo extension in the Sequenom assay, which may be an additional consequence of the characteristics of the *ACHE* sequence, as discussed below. Consequently, the population frequency of the corresponding alleles could not be established.

Statistical analysis of the population frequencies of the 11 *ACHE* SNPs indicates that the African American population is significantly different from the other populations examined in this study ($p < 0.01$, Table 3), while the three Middle East populations are virtually indistinguishable. These results are expected, since the African population diverged much earlier than the three Israeli populations

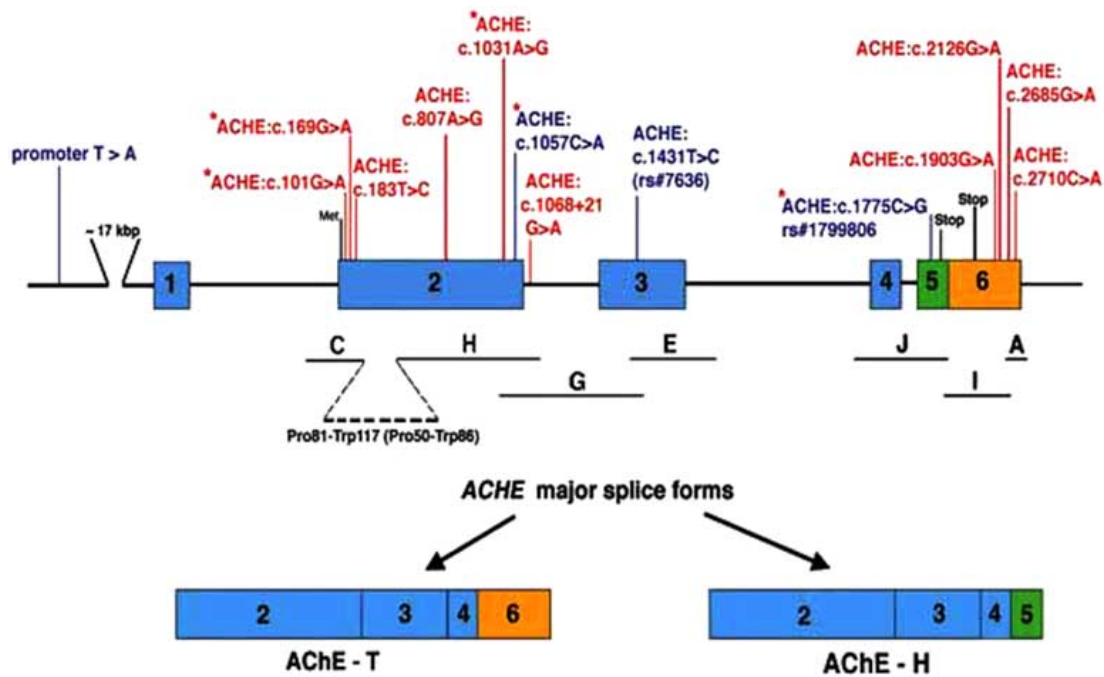


Fig. (2). Polymorphisms and genomic organization of *ACHE*

Human AChE is encoded by a single *ACHE* gene composed of six exons, which generates two major alternatively spliced forms that differ in quaternary structure and tissue distribution. Exons are depicted as boxes, and labeled from 1 to 6. PCR amplicons used for DHPLC analysis (A,C,E,G,H,I, and J) in SNP detection are represented under the gene. Coverage of the coding regions by the amplicons is representative. The DNA segment corresponding to amino-acids Pro81–Trp117 (Pro50–Trp86) was not amplified. Previously reported SNPs are in blue, and novel SNPs in red. Non-synonymous SNPs are marked with *. SNPs are identified based on their position in the cDNA sequence, NM_015831, taking the A nucleotide of the ATG as the +1 base. For detection of SNPs, seven fragments, covering most of the coding region of the *ACHE* gene, were analyzed both by DHPLC and by direct sequencing.

screened, and serves as an additional internal control for our data [23, 30].

Of 11 *ACHE* alleles screened, three seem to be population-specific, two in African-Americans (*ACHE:c.101G>A* and *ACHE:c.2685G>A*, $p=0.0079$ and $p=0.0273$, respectively) and one in both Jewish populations (*ACHE:c.2126G>A*, $p = 0.0014$). The minor alleles for three other SNPs, although detected only in one population (*ACHE:c.807G>A*, *ACHE:c.1068+21 G>A*, and *ACHE:c.183T>C*), had a very low frequency and did not, therefore, attain statistical significance. Yet, cladistic analysis of *ACHE* haplotypes strongly implies that both SNPs, *ACHE:c.807 G>A*, and *ACHE:c.1068+21 G>A*, are specific to the African American population (Fig. 3), because they occurred on the background of an African American specific allele, *ACHE:c.101G>A*.

Two of the SNPs, *ACHE:c.1057C>A* (Yt^b antigen) and *ACHE:c.1431T>C*, have been previously reported to be in full linkage disequilibrium [22, 23]. Our results indicate high, but not full, linkage disequilibrium between these two SNPs in the Israeli populations ($D^{\prime}=0.884$), and even less in African Americans, in whom *ACHE:c.1431T>C* was found to be much more prevalent than *ACHE:c.1057C>A*. This suggests that the previously observed linkage disequilibrium may be attributed to the ethnic groups studied and to the limited number of DNA samples used.

Our survey identified 5 non-synonymous polymorphisms in the *ACHE* gene (Fig. 2). Multiple alignment (Fig. 4a) showed that none of them involves a conserved amino acid. Of the 5 coding SNPs, only 3 can be assigned to the AChE crystal structure (*ACHE:c.169G>A*, *ACHE:c.1031A>G*, *ACHE:c.1057C>A*, Fig. 4b) [16]. However, although residue 34(3) (*ACHE:c.101G>A*) is not visible in the crystal structure, it is present in the mature protein and could possibly affect the biological properties of the mature AChE. Residue 592(561) (*ACHE:c.1775C>G*), although absent from the mature protein, could possibly have an influence prior to processing (e.g. on folding) or on processing itself. Fig. 4b clearly shows that the first three variants map on the outer surface of the crystal structure. Based on their locations (Fig. 4b) and on multiple alignment (Fig. 4a), these non-synonymous changes are unlikely to influence the catalytic properties of AChE [31]. However, it was earlier shown that the His353Asn mutation (*ACHE:c.1057C>A*) influences the antigenic properties of AChE, and is responsible for the YT blood group phenotype [22]. It is thus possible, based on their peripheral localization, that the other two non-synonymous changes found in this study (*ACHE:c.169G>A* *ACHE:c.1031A>G*) may also affect the antigenicity of AChE.

Human *ACHE* haplotypes were inferred from genotype data using PHASE software. We were thus able to show that 11 SNPs are arranged in 15 different haplotypes (Table 4). These data are consistent with a survey performed on 313

Table 2. Minor allele Frequency of ACHE SNPs in Four Ethnically Distinct Populations

SNP identifier	Public data-base accession # *	Position in ACHE gene	Substitution Type**	Deduced Amino acid change	Minor allele frequency in			
					African Americans	Ashkenazi Jews	Sephardic Jews	Israeli Arabs
<i>ACHE:c.101G>A</i>	Not reported	exon2	Coding non-syn.	Arg34Gln (Arg3Gln)	5.2 %	NP	NP	NP
<i>ACHE:c.169G>A</i>	Not reported	exon2	Coding non-syn.	Gly57Arg (Gly26Arg)	1%	2.1%	NP	2.1%
<i>ACHE:c.183 T>C</i>	Not reported	exon2	Coding, synonym.	Ser61Ser (Ser30Ser)	NP	NP	NP	1%
<i>ACHE:c.807G>A</i>	Not reported	exon2	Coding, synonym.	Thr269Thr (Thr238Thr)	2.1%	NP	NP	NP
<i>ACHE:c.1031A>G</i>	Not reported	exon2	Coding, non-syn	Glu344Gly (Glu313Gly)	ND	ND	ND	ND
<i>ACHE:c.1057C>A</i>	MIM #112100 Yt blood group antigen	exon2	Coding non-syn	His353Asn (His322Asn)	1%	5.2%	9.4%	8.3%
<i>ACHE:c.1068+21 G>A</i>	Not reported	intron2	Intronic		2.1%	NP	NP	NP
<i>ACHE:c.1431T>C</i>	rs#7636	exon3	Coding, synonym.	Pro477Pro (Pro446Pro)	29.1%	4.2%	7.3%	8.3%
<i>ACHE:c.1775C>G</i>	rs#1799806	exon5	Coding non-syn	Pro592Arg (Pro561Arg)	11.5%	39.5%	33.3%	33.3%
<i>ACHE:c.1903G>A</i>	Not reported	3'UTR	untranslated		ND	ND	ND	ND
<i>ACHE:c.2126G>A</i>	Not reported	3'UTR	untranslated		NP	4.2%	8.3%	NP
<i>ACHE:c.2685G>A</i>	Not reported	3'UTR	untranslated		4.2%	NP	NP	NP
<i>ACHE:c.2710C>A</i>	Not reported	3'UTR	untranslated		28%	5.2%	9.4%	7.3%

Thirteen *ACHE* SNPs were identified using denaturing HPLC and direct sequencing. Each SNP was sequenced at least twice in at least one individual. Allele frequencies were determined by genotyping 48 unrelated individuals from each of four populations for 11 of the 13 SNPs, using the Sequenom MassARRAY system. Two SNPs, *ACHE:c.1031A>G* and *ACHE:c.1903G>A*, although clearly visualized by multiple direct sequencing in at least two independent PCR amplifications, failed to undergo extension in the Sequenom assay. SNP identifier corresponds to the position in the *ACHE* cDNA sequence (NM_015831.1) as presented in (Fig. 2), and the amino acid numbering corresponds to the position in the NP_056646 protein.

* rs# represents the dbSNP accession # and MIM# represents the OMIM database accession #.

** Type of SNP indicates whether the SNP resides in the coding, intronic or untranslated region (UTR). For coding SNPs, it is also indicated whether the nucleotide change results in an amino-acid substitution (non-synonymous) or not (synonymous).

Table 3. Genetic Differences Among Four Populations, as Reflected by ACHE SNPs

	African Americans	Ashkenazi Jews	Sephardic Jews	Israeli Arabs
African Americans		4	5	5
Ashkenazi Jews	2 = 75.03 DF=10		0	0
Sephardic Jews	2 = 64.78 DF=10	2 = 7.8 DF=6		1
Israeli Arabs	2 = 53.73 DF=10	2 = 8.7 DF=7	2 = 12.22 DF=7	

Upper triangle indicates the number of SNPs significantly ($p < 0.05$) different between the two populations compared.

Lower triangle summarizes the statistical difference between populations, as reflected by all SNPs.

Cells colored in grey are highly significant ($p < < 0.01$); uncolored cells stand for non-significant differences ($p > 0.05$).

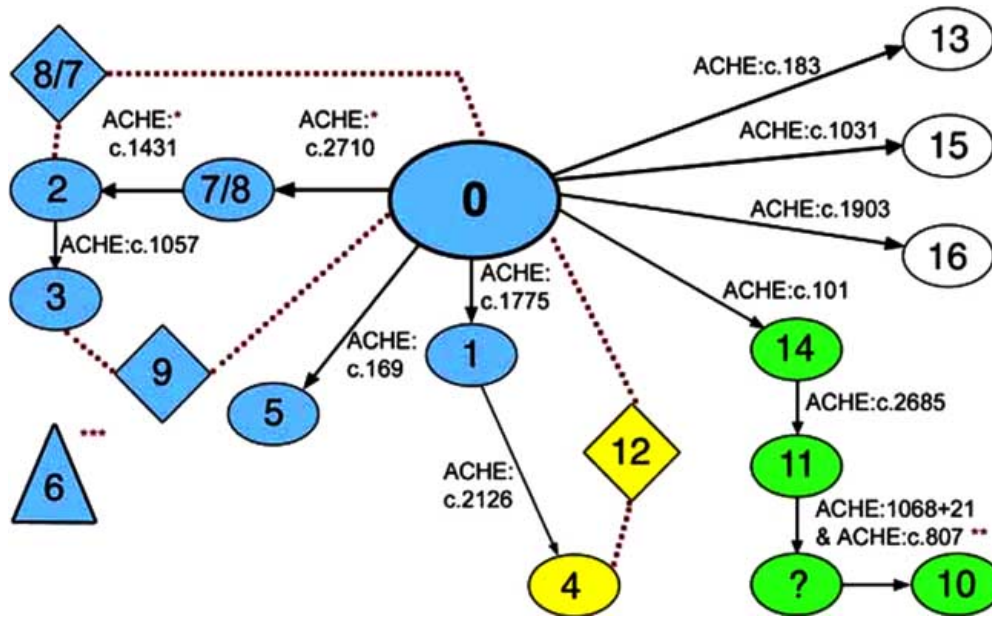


Fig. (3). Cladogram of ACHE haplotypes

Each node represents a haplotype with corresponding Hap ID# taken from Table 4. Appearance of new haplotypes is indicated by directional arrows. The corresponding mutations are indicated on the arrows. Haplotypes that, supposedly, originated by recombination are represented by rectangles, connected to the original haplotypes by red dotted lines. Haplotypes present in more than one population are colored blue, other colors representing population-specific haplotypes. (Green=African Americans, Yellow=Jews). Uncolored nodes represent haplotypes with inconclusive data regarding ethnic specificity. Hap0 is assumed to be the ancestral haplotype. Recombination between ACHE:c.1775C>G and ACHE:c.2126G>A is allowed, if it involves at least one of the common haplotypes (i.e. Hap0 or Hap1). Hap7 and Hap8 are equally probable at the positions indicated.

* The sequence of occurrence of these SNPs cannot be uniquely deduced from our data.

** The exact sequence of occurrence of these two SNPs could not be determined from our data.

***Haplotype 6 could not be explained by this model, as discussed.

4a)

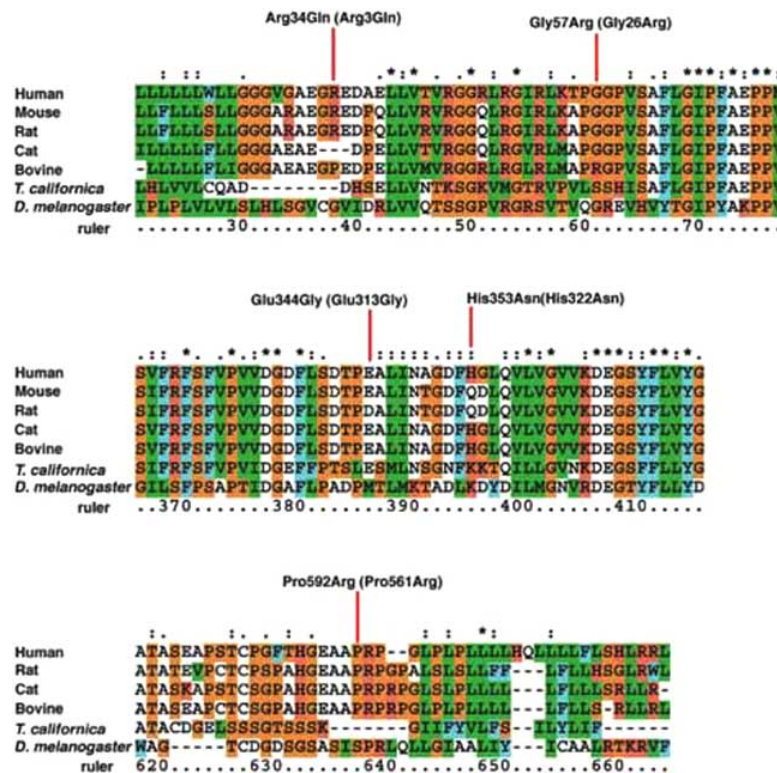


Fig. (4). contd....

4b)

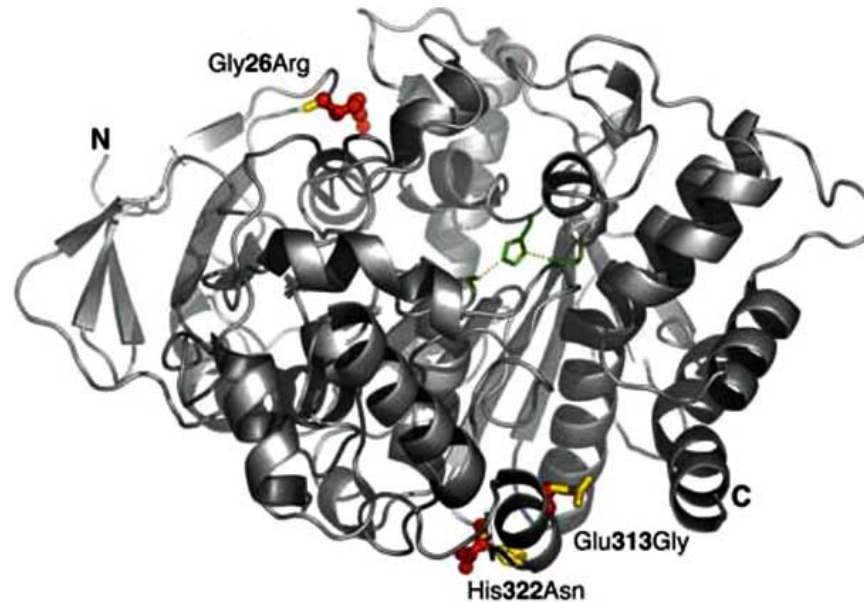


Fig. (4). Multiple alignment and structural representation of AChE

a) A multiple alignment of segments of human AChE (top line) with the corresponding segments of some of its homologs. The sequences were selected using a BLASTP search of hAChE against the non-redundant protein database. Multiple alignment of the entire protein was performed with the ClustalW program, using the default parameters. Acidic amino acids are in **orange**, basic amino acids in **blue**, and others in **green**. Amino-acid substitutions corresponding to the non-synonymous SNPs are labeled. The human AChE sequence presented in the upper two panels corresponds to both major human AChE splice forms. Pro592Arg (Pro561Arg) is expressed only in the AChE-H isoform; therefore, in the bottom panel only AChE-H isoforms of the different species are aligned, with the mouse AChE-H isoform; sequence not being available. Accession numbers of the aligned proteins are: human AChE-H (GPI-linked form precursor): NP_056646.1; *Mus musculus* AChE: NP_033729.1; *Rattus norvegicus* P37136-2 (Swiss-Prot ID¹); *Bos taurus* P23795-2 (Swiss-Prot ID¹); *Felis silvestris catus* O62763-2 (Swiss-Prot ID¹); *Drosophila melanogaster* AChE: NP_476953.1; *Torpedo californica* AChE: P04058. The ruler under the alignment represents the position in a consensus sequence generated by the alignment of all sequences. In the line above the alignment '*' - indicates positions which have a single, fully conserved residue; ':' indicates 'strong' conservation; '.' indicates 'weaker' conservation. ¹Sequences with Swiss-Prot IDs can be retrieved at <http://us.expasy.org/sprot/>.

b) Ribbon diagram of the 3D structure of human AChE showing the positions of the three nonsynonymous SNPs: Gly57Arg (Gly26Arg, *ACHE:c.169G>A*); Glu344Gly (Glu313Gly, *ACHE:c.1031A>G*); His353Asn (His322Asn, *ACHE:c.1057C>A*). The hAChE structure (16) (PDB ID1b41) is shown in grey, and the catalytic triad is shown in **green**, with the hydrogen bonds indicated as dashed-lines between the residues. For each SNP, the ancestral amino acid is shown in **yellow** (stick representation), and the corresponding polymorphism in **red** (ball-and-stick representation). His353Asn (His322Asn) corresponds to the YT blood type. The labeled N-terminus of the structure is marked, and corresponds to Asp36 (Asp5). Thr574 (Thr543) is both the last residue in the 3D structure and in the last common constitutive exon. Each SNP is marked adjacent to its position in the structure.

randomly chosen genes [32], which reported average values of one SNP per 185bp, and 12.5 biallelic SNPs arranged in 14 haplotypes per gene. All *ACHE* haplotypes are summarized in (Table 4), together with their respective frequencies in the four ethnic groups studied. We have found the ancestral haplotype, Hap0, to be ~ 50% prevalent in all four ethnic groups, while the second, third, fourth and fifth most common haplotypes (Hap1, Hap2, Hap3 and Hap4), three of which carry amino acid substitutions, are highly prevalent in at least one of the ethnic groups. Consequently, every 3 out of 4 individuals, on average, carry at least one polymorphic variant of the *ACHE* gene. Thus, if these polymorphisms have any effect on AChE expression, structure or function, their characterization and identification may potentially have important benefits. We have found that the frequency of

Hap3 in African Americans differs significantly from its frequency in the Jewish and Israeli Arab populations ($p < 0.01$), while Hap4 appears to be specific to Jews. This finding might reflect either the population bottleneck that accompanied migration out of Africa, or a difference in the selective pressures in the new environment.

Cladistic analysis of haplotypes may add to our insight into the population specificity of the SNPs and into the divergence of the populations studied. Based on haplotype frequencies, population specificity and possible recombination we propose a cladogram for the *ACHE* haplotypes (Fig. 3). Indeed, cladistic analysis of *ACHE* haplotypes strongly implies that both SNPs *ACHE:c.807G>A* and *ACHE:c.1068+21G>A* are specific to the African American population (Fig.

Table 4. Haplotype Structure of the ACHE Gene

Hap ID#	Haplotype structure											Haplotype frequency (%) in				
	ACHE:c.101	ACHE:c.169	ACHE:c.183	ACHE:c.807	ACHE:c.1057 (Y ^b)	ACHE:c.1068+21	ACHE:c.1431	ACHE:c.1775	ACHE:c.2126	ACHE:c.2685	ACHE:c.2710	African Americans	Ashkenazi Jews	Sephardic Jews	Israeli Arabs	All 4 populations
0	0	0	0	0	0	0	0	0	0	0	0	52	53	55.2	55.2	54
1	0	0	0	0	0	0	0	1	0	0	0	11.5	35.5	26	33.3	26.5
2	0	0	0	0	0	0	1	0	0	0	1	25			1	6.5
3	0	0	0	0	1	0	1	0	0	0	1	1	4.2	6.25	7.3	4.7
4	0	0	0	0	0	0	0	1	1	0	0		4.2	7.3		2.9
5	0	1	0	0	0	0	0	0	0	0	0	1	2		2	1.3
6	0	0	0	0	1	0	0	0	0	0	1		1	2		0.8
7	0	0	0	0	0	0	1	0	0	0	0	2				0.5
8	0	0	0	0	0	0	0	0	0	0	1	1		1		0.5
9	0	0	0	0	1	0	1	0	0	0	0			1	1	0.5
10	1	0	0	1	0	1	0	0	0	1	0	2				0.5
11	1	0	0	0	0	0	0	0	0	1	0	2				0.5
12	0	0	0	0	0	0	0	0	1	0	0			1		0.26
13	0	0	1	0	0	0	0	0	0	0	0				1	0.26
14	1	0	0	0	0	0	0	0	0	0	0	1				0.26

0 represents a common allele, 1 represents a rare allele for all SNPs

SNPs identifiers correspond to Table 2.

AA=African Americans, Ashk=Ashkenazi Jews, Seph=Sephardic Jews, IA=Israeli Arabs

3), because they occurred on the background of two African American specific alleles, *ACHE:c.101G>A* and *ACHE:c.2685G>A*.

One of the haplotypes, Hap6, could not be resolved by use of the basic rules applied, as discussed below. Yet, we prefer to present the cladogram that is most consistent with our results, leaving Hap6 unexplained. Moreover, due to low frequencies of Hap7 and Hap8, it is not possible to conclude which of these haplotypes is the original haplotype, and which is the recombinant one. Consequently, both are assigned to the same nodes (Fig. 3).

DISCUSSION

The physiological importance of AChE, combined with the low number of naturally occurring genetic polymorphisms previously detected, led to the conclusion that AChE is highly conserved, suggesting that its function cannot readily tolerate non-synonymous DNA sequence variants. Thus,

we did not expect to find many polymorphisms in the human *ACHE* gene. However, we ended up identifying 13 polymorphisms in *ACHE*, arranged in 17 different haplotypes.

Although the GC content of the *ACHE* gene is not especially high (63%), it is unevenly distributed, reaching over 90% in some coding regions. Moreover, the presence of multiple repeat sequences in the untranslated regions of the gene makes primer design and amplification of the *ACHE* coding region a challenging task. We believe that the low number of previously identified polymorphisms is attributable to unusual sequence properties of the *ACHE* gene, and does not reflect the true heterozygosity of this locus.

As already mentioned, we have identified 13 SNPs, arranged in 17 haplotypes, and these SNPs can be further subdivided into 5 coding non-synonymous, 3 coding synonymous and 5 non-coding SNPs. All five coding non-synonymous SNPs identified in this study are either absent from the mature protein or reside far from the active site.

Thus, they are unlikely to have a direct impact on the catalytic properties of AChE, and should not, presumably, have dramatic phenotypic consequences. By extension, this suggests that changes having an impact on the active site would lead to a severe phenotype, and would, therefore, be selected against. This stands in contrast both to findings regarding the paralogous enzyme, butyrylcholinesterase (BChE), and the mouse AChE knockout model [33, 34]. AChE $-/-$ mice do not display an abnormal phenotype, and the human BChE gene has been shown to have more than 30 alleles, several of which code for an enzyme displaying significantly altered levels of catalytic activity [35]. However, in contrast to AChE, BChE has no recognized biological role, and the mouse AChE knockout phenotype may stem from species-specific characteristics.

It should be emphasized that our focus on coding regions, and the sample sizes employed, may have precluded identification both of rare mutations (<1-2% prevalence) and of potentially functional polymorphisms in the intronic, untranslated and regulatory regions of the gene. Moreover, an 118bp fragment corresponding to amino acids Pro81-Trp117 (Pro50-Trp86) has not yet been amplified, and has not, accordingly, been screened for polymorphisms.

Interestingly, 13 of the 18 *ACHE* SNPs listed in the dbSNP database (http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?locusId=43) lack biological validation status. None of these SNPs was observed in our study, including SNP 398, that could be of potential interest since it would create a stop codon at Gln102(Gln71), which should result in an inactive gene product. There may be a number of reasons for these discrepancies: some SNPs may not be covered in our amplified fragments, or their frequency may lie below the detection level of our study. However, others may well be "electronic SNPs", representing either sequencing errors or electronic mistakes, rather than authentic DNA sequence variants. Currently, major efforts are being made to generate systematically an exhaustive and accurate catalogue of all human SNPs, and to identify their underlying chromosomal or haplotype structures. Our findings suggest that many SNPs throughout the proteome may have escaped identification due to inherent local features of the target genomic DNA sequences. Thus, the high-throughput, but crude, approach to SNP discovery may be inadequate, and numerous loci will require particular and time-consuming efforts in order to accurately estimate their genetic variation.

In general, only ~5% of the disease-causing non-synonymous mutations identified so far have a direct effect on catalytic or ligand-binding properties of the proteins studied [31, 36]. Over 80% of disease-causing missense mutations have been shown to have a minor destabilizing effect on protein structure, but no direct effect on binding, catalysis or regulation. Five coding, non-synonymous polymorphisms that were identified in our study are not likely to have a direct effect on the catalytic properties of AChE, but may affect protein stability and/or antigenicity. This possibility finds support in the fact that one of the SNPs, *ACHE:c169G>A*, which causes substitution of Gly, a small non-charged amino acid, by Arg, a large amino-acid with a basic side-chain, occurs at lower frequencies than expected. This presumably ancient SNP, that is present in all tested

populations, would be expected to occur at a similar frequency as other ancient SNPs, such as *ACHE:c.1057C>A*, *ACHE:c.1431T>C*, *ACHE:c.2710C>A*, and to be more frequent than population-specific SNPs, such as *ACHE:c.2126G>A* and *ACHE:c.101G>A*. Its low frequency may reflect a negative selection for this change. Functional consequences of this amino acid substitution should, therefore, be investigated further.

Another newly identified non-synonymous polymorphism, *ACHE:c.101G>A*, is at lower frequency than might be expected from the fact that two other polymorphisms (*ACHE:c.807G>A*, and *ACHE:c.1068+21G>A*) occur on an *ACHE:c.101G>A* background. Combination of these two events would be highly unlikely ($p < 0.0025$ chance) if at the time that *ACHE:c.807G>A*, and *ACHE:c.1068+21G>A* arose, the prevalence of *ACHE:c.101G>A* was similar to the currently observed ~5%. It is, therefore, plausible that the *ACHE:c.101G>A* SNP was once much more prevalent in the African American population. The decline in frequency of this SNP may stem from one of several causes, including random drift, population bottleneck, change in selective pressure or any combination of them.

Two previously identified polymorphisms in the *ACHE* gene, *ACHE:c.1057C>A* and a promoter region polymorphism, are responsible, respectively, for the YT blood group phenotype and for hypersensitivity to pesticides. Two novel SNPs, *i.e.* *ACHE:c.169G>A* and *ACHE:c.1031A>G*, are of type and position similar to that of *ACHE:c.1057C>A*, and might thus have a similar impact on the antigenic properties of the enzyme. Moreover, functional consequences of any of the SNPs discovered in our study may find expression only under a particular set of conditions, such as stress or response to medication.

Haplotype analysis of the *ACHE* SNPs reveals a recombination spot in the region between *ACHE:c.1775C>G* and *ACHE:c.2126G>A*. Cladistic analysis of haplotypes provides an explanation for different haplotype frequencies in different populations. Such an analysis may add to our insight into the population specificity of the SNPs, and guide the optimal SNP selection for an association study. The proposed cladogram provides a plausible explanation for all the haplotypes observed, except for Hap6. In our sample, a specific combination of three SNPs, *ACHE:c.1057C>A*, *ACHE:c.1431T>C*, and *ACHE:c.2710C>A*, creates 7 out of the 8 theoretically possible haplotypes (2, 3, 6, 7 and 8). Considering the rare recombination rate expected [37, 38], and the physical proximity of the SNPs (0.4 kb apart), these results are surprising. Each of these haplotypes was observed more than once in our sample, and are thus unlikely to represent genotyping errors, although this possibility cannot be ruled out completely. Furthermore, several improbable events, such as additional recombination between *ACHE:c.1057C>A* and *ACHE:c.1431T>C*, or recurrent mutation, may explain our observations. Resolution of this issue demands more comprehensive research, employing larger sample sizes and more ethnic populations. We choose, therefore, to present the cladogram most consistent with our results, leaving Hap6 unexplained.

Considering the small sample sizes used (64 chromosomes from each of the 3 populations), this study did not

have the power to detect rare (1-2% frequency) polymorphisms. Thus, studying possible association of the detected SNPs with diverse phenotypes seems justified. It has been shown that non-coding polymorphisms may be associated with diseases, including autoimmunity [39], asthma [40], expression of specific molecular markers by certain cancers [41], and AD [42]. Such associations may be indirect, due to tight linkage of the non-coding SNP to the deleterious mutation, or the non-coding SNP itself may cause the phenotype by creating or abolishing splice sites [43], modifying transcription [44], or altering mRNA stability [39]. We have identified 5 non-coding and 3 synonymous polymorphisms in the *ACHE* gene. Though they do not change the primary sequence of the protein, these SNPs may also be associated with alterations in protein expression or distribution. Thus, their possible (direct or indirect) influence on the transcription and stability of *ACHE* mRNA should be further investigated.

To summarize, our survey is the first comprehensive study of naturally occurring changes in the *ACHE* gene and their population-specific structure. Our unexpected findings challenge the accepted view, and suggest that *ACHE* is no less polymorphic than the average gene. Furthermore our study provides the necessary foundation for functional and pharmacogenetic characterization of naturally occurring variants of the *ACHE* protein.

ACKNOWLEDGEMENTS

This study was supported by a grant from the Institute for the Study of Aging, an Israel Ministry of Science and Technology (MOST) grant for Interdisciplinary Studies and a MOST grant to the Israel Structural Proteomics Center; by the European Commission (EC) Vth Framework 'Structural Proteomics in Europe (SPINE)' Project grant: QL2-CT-2002-00988, by the Kalman and Ida Wolens Foundation, by the Jean and Julia Goldwurm Memorial Foundation, by the Divadol Foundation, and by the Benozio Center for Neuroscience. Jacques S. Beckmann is the Hermann Mayer Professor of Molecular Genetics, Joel L. Sussman is the Morton and Gladys Pickman Professor of Structural Biology and Amos D. Korczyn is the Heinrich Sieratzki Professor of Neurobiology.

REFERENCES

- [1] Sternfeld M, Ming G, Song H, Sela K, Timberg R, Poo M, *et al.* Acetylcholinesterase enhances neurite growth and synapse development through alternative contributions of its hydrolytic capacity, core protein, and variable C termini. *J Neurosci* 18(4):1240-1249 (1998).
- [2] Lev-Lehman E, Deutsch V, Eldor A, Soreq H. Immature human megakaryocytes produce nuclear-associated acetylcholinesterase. *Blood* 89(10):3644-3653 (1997).
- [3] Kawashima K, Fujii T. Extraneuronal cholinergic system in lymphocytes. *Pharmacol. Therapeut.* 86(1):29-48 (2000).
- [4] Giacobini E. Cholinesterases Inhibitors: from the Calabar bean to Alzheimer therapy. In: Giacobini E, editor. *Cholinesterases and Cholinesterase Inhibitors*. London: Martin Dunitz; p. 181-226 (2000).
- [5] Talesa VN. Acetylcholinesterase in Alzheimer's disease. *Mech Ageing Dev* 122(16):1961-1969 (2001).
- [6] Soreq H, Seidman S. Acetylcholinesterase - new roles for an old actor. *Nat Rev Neurosci* 2(4):294-302 (2001).
- [7] Li Y, Camp S, Rachinsky TL, Getman D, Taylor P. Gene structure of mammalian acetylcholinesterase. Alternative exons dictate tissue-specific expression. *J Biol Chem* 266(34):23083-23090 (1991).
- [8] Seidman S, Sternfeld M, Ben Aziz-Aloya R, Timberg R, Kaufman-Nachum D, Soreq H. Synaptic and epidermal accumulations of human acetylcholinesterase are encoded by alternative 3'-terminal exons. *Mol Cell Biol* 15(6):2993-3002 (1995).
- [9] Massoulié J, Anselmet A, Bon S, Krejci E, Legay C, Morel N, *et al.* The polymorphism of acetylcholinesterase: post-translational processing, quaternary associations and localization. *Chem Biol Interact* 119-120:29-42 (1999).
- [10] Sussman JL, Harel M, Frolow F, Oefner C, Tokar L, Silman I. Structural studies on acetylcholinesterase from *Torpedo californica*. In: Massoulié J, Bacou F, Barnard E, Chatonnet A, Doctor BP, Quinn DM, editors. *Cholinesterases: Structure, Function, Mechanism, Genetics and Cell Biology*. Washington, DC: American Chemical Society; p. 7-11 (1991).
- [11] Ripoll DR, Faerman CH, Axelsen PH, Silman I, Sussman JL. An electrostatic mechanism for substrate guidance down the aromatic gorge of acetylcholinesterase. *Proc Natl Acad Sci USA* 90(11):5128-5132 (1993).
- [12] Felder CE, Botti SA, Lifson S, Silman I, Sussman JL. External and internal electrostatic potentials of cholinesterase models. *J Mol Graph Model* 15(5):318-327 (1997).
- [13] Bourne Y, Taylor P, Bougis PE, Marchot P. Crystal structure of mouse acetylcholinesterase. A peripheral site-occluding loop in a tetrameric assembly. *J Biol Chem* 274(5):2963-2970 (1999).
- [14] Gilson MK, Straatsma TP, McCammon JA, Ripoll DR, Faerman CH, Axelsen P, *et al.* Open "back door" in a molecular dynamics simulation of acetylcholinesterase. *Science* 263(5151):1276-1278 (1994).
- [15] Bourne Y, Taylor P, Marchot P. Acetylcholinesterase inhibition by fasciculin: crystal structure of the complex. *Cell* 83(3):503-512 (1995).
- [16] Kryger G, Harel M, Giles K, Tokar L, Velan B, Lazar A, *et al.* Structures of recombinant native and E202Q mutant human acetylcholinesterase complexed with the snake-venom toxin fasciculin-II. *Acta Crystallogr. D Biol Crystallogr* 56(11):1385-1394 (2000).
- [17] Silman I, Sussman JL. Structural studies on acetylcholinesterase. In: Giacobini E, editor. *Cholinesterases and Cholinesterase Inhibitors*. London: Martin Dunitz; p. 9-25 (2000).
- [18] Greenblatt HM, Dvir H, Silman I, Sussman JL. Acetylcholinesterase: a multifaceted target for structure-based drug design of anticholinesterase agents for the treatment of Alzheimer's disease. *J Mol Neurosci* 20(3):369-384 (2003).
- [19] Brenner T, Hamra-Amitay Y, Evron T, Boneva N, Seidman S, Soreq H. The role of readthrough acetylcholinesterase in the pathophysiology of myasthenia gravis. *FASEB J* 17(2):214-222 (2003).
- [20] Michaelis ML. Drugs targeting Alzheimer's disease: some things old and some things new. *J Pharmacol Exp Ther* 304(3):897-904 (2003).
- [21] Scarpini E, Scheltens P, Feldman H. Treatment of Alzheimer's disease: current status and new perspectives. *Lancet Neurol* 2(9):539-547 (2003).
- [22] Bartels CF, Zelinski T, Lockridge O. Mutation at codon 322 in the human acetylcholinesterase (ACHE) gene accounts for YT blood group polymorphism. *Am J Hum Genet* 52(5):928-936 (1993).
- [23] Ehrlich G, Ginzberg D, Loewenstein Y, Glick D, Kerem B, Ben-Ari S, *et al.* Population diversity and distinct haplotype frequencies associated with ACHE and BCHE genes of Israeli Jews from trans-Caucasian Georgia and from Europe. *Genomics* 22(9):288-295 (1994).
- [24] Shapira M, Tur-Kaspa I, Bosgraaf L, Livni N, Grant AD, Grisaru D, *et al.* A transcription-activating polymorphism in the ACHE promoter associated with acute sensitivity to anti-acetylcholinesterases. *Hum Mol Genet* 9(9):1273-1281 (2000).
- [25] Escary JL, Bottius E, Prince N, Reyes C, Fiawoumo Y, Caloustian C, *et al.* A first high-density map of 981 biallelic markers on human chromosome 14. *Genomics* 70(2):153-164 (2000).
- [26] Bercovich D, Beaudet AL. Denaturing high-performance liquid chromatography for the detection of mutations and polymorphisms in UBE3A. *Genet Test* 7(3):189-194 (2003).
- [27] Yaron Y, Ben Zeev B, Shomrat R, Bercovich D, Naiman T, Orr-Urtreger A. MECP2 mutations in Israel: implications for molecular analysis, genetic counseling, and prenatal diagnosis in Rett syndrome. *Hum Mutat* 20(4):323-324 (2002).

- [28] Bonfield JK, Smith K, Staden R. A new DNA sequence assembly program. *Nucleic Acids Res* 23(24):4992-4999 (1995).
- [29] Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. *Behav Brain Res* 125(1-2):279-284 (2001).
- [30] Fullerton SM, Clark AG, Weiss KM, Nickerson DA, Taylor SL, Stengard JH, *et al.* Apolipoprotein E variation at the sequence haplotype level: implications for the origin and maintenance of a major human polymorphism. *Am J Hum Genet* 67(4):881-900 (2000).
- [31] Vitkup D, Sander C, Church GM. The amino-acid mutational spectrum of human genetic disease. *Genome Biol* 4(11):R72 (2003).
- [32] Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68(4):978-989 (2001).
- [33] Xie W, Stribley JA, Chatonnet A, Wilder PJ, Rizzino A, McComb RD, *et al.* Postnatal developmental delay and supersensitivity to organophosphate in gene-targeted mice lacking acetylcholinesterase. *J Pharmacol Exp Therap* 293(3):896-902 (2000).
- [34] Xie W, Wilder PJ, Stribley J, Chatonnet A, Rizzino A, Taylor P, *et al.* Knockout of one acetylcholinesterase allele in the mouse. *Chem Biol Interact* 119-120:289-299 (1999).
- [35] Darvesh S, Hopkins DA, Geula C. Neurobiology of butyrylcholinesterase. *Nat Rev Neurosci* 4(2):131-138 (2003).
- [36] Wang Z, Moulton J. SNPs, protein structure, and disease. *Hum Mut* 17(4):263-270 (2001).
- [37] Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, *et al.* The structure of haplotype blocks in the human genome. *Science* 296(5576):2225-2229 (2002).
- [38] Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. High-resolution haplotype structure in the human genome. *Nat Genet* 29(2):229-232 (2001).
- [39] Ueda H, Howson JM, Esposito L, Heward J, Snook H, Chamberlain G, *et al.* Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature* 423(6939):506-511 (2003).
- [40] Duetsch G, Illig T, Loesgen S, Rohde K, Klopp N, Herbon N, *et al.* STAT6 as an asthma candidate gene: polymorphism-screening, association and haplotype analysis in a Caucasian sib-pair study. *Hum Mol Genet* 11(6):613-621 (2002).
- [41] Miyoshi Y, Ando A, Hasegawa S, Ishitobi M, Yamamura J, Irahara N, *et al.* Association of genetic polymorphisms in CYP19 and CYP1A1 with the oestrogen receptor-positive breast cancer risk. *Eur J Cancer* 39(17):2531-2537 (2003).
- [42] Lambert JC, Goumidi L, Vrieze FW, Frigard B, Harris JM, Cummings A, *et al.* The transcriptional factor LBP-1c/CP2/LSF gene on chromosome 12 is a genetic determinant of Alzheimer's disease. *Hum Mol Genet* 9(15):2275-2280 (2000).
- [43] Richard I, Beckmann JS. How neutral are synonymous codon mutations? *Nat Genet* 10(3):259 (1995).
- [44] Knowles JW, Erickson LM, Guy VK, Sigel CS, Wilder JC, Maeda N. Common variations in noncoding regions of the human natriuretic peptide receptor A gene have quantitative effects. *Hum Genet* 112(1):62-70 (2003).