APOE *Polymorphism in a Rural Older Population-Based Sample in India*

B.K. THELMA,[1] RAMESH C. JUYAL,[2] HIROKO H. DODGE,[3] RAJESH PANDAV,[3,4] VIJAY CHANDRA,[3,4] AND MARY GANGULI[3,5]

Abstract Allele frequencies are most often reported from small convenience samples of unknown demographics and limited generalizability. We determined the distribution of apolipoprotein E genotype (*APOE*) and allele frequencies for a large, well-defined, representative, rural, population-based sample ($n = 4450$) aged 55–95 years in Ballabgarh, in the northern Indian state of Haryana. The overall *APOE E*2*, *E*3*, and *E*4* allele frequencies were 0.039, 0.887, and 0.073, respectively; frequencies are also reported by age, sex, and religious/caste groups. The *APOE*4* frequency is among the lowest reported anywhere in the world. *APOE* allele frequencies did not vary significantly by age or sex in this study. To our knowledge, this is the largest Indian sample ever genotyped for the *APOE* polymorphism. The representativeness of the sample and its known demographics provide a much-needed normative background for studies of gene-disease associations.

The human apolipoprotein E gene (*APOE*) codes for a 299aa monomeric glycoprotein (apoE) of 34 kD, which acts as a ligand for two specific cell receptors and mediates cellular uptake of apoE-containing lipoproteins. The *APOE* gene is located on chromosome 19 and has three common allelic variants, *E*2, E*3,* and *E*4* (Des et al. 1985). The molecular basis of the *APOE* polymorphism is missense mutations at codons 112 and 158, leading to the *E*2* variant with cysteine at both positions, the *E*3* variant with cysteine at position 112 and arginine at 158, and the *E*4* isoform with arginine at both positions (Assman et al. 1987). Considerable variation in *APOE* allele frequencies has been reported among various ethnic groups (Kamboh 1995;

[1] Department of Genetics, University of Delhi South Campus, New Delhi, India.
[2] National Institute of Immunology, New Delhi, India.
[3] Department of Epidemiology, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA.
[4] Centre for Ageing Research in India, New Delhi, India.
[5] Division of Geriatrics and Neuropsychiatry, Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA.

Human Biology, February 2001, v. 73, no. 1, pp. 135–144.

KEY WORDS: *APOE* ALLELE FREQUENCIES, *APOE* GENOTYPE, INDIAN POPULATION

Hendrie 1998). The majority of allele frequency reports worldwide have been based on studies of small convenience samples of unknown demographics.

Physiologically, *APOE* alleles show variable associations with several diseases. The *APOE*4* allele is associated with higher levels of plasma cholesterol (Utermann 1977), atheroclerosis (Hixon and Group 1991; Davignon et al. 1988), coronary artery disease (Wilson et al. 1996), ischemic cerebrovascular disease (Pedro-Botet et al. 1992), Alzheimer's disease (AD) (Corder et al. 1993; Saunders et al. 1993; Evans et al. 1997), vascular dementias (Slooter et al. 1997; Kalman et al. 1998), and decreased longevity (Cauley et al. 1993; Alberts et al. 1995, Stengard et al. 1996). In the majority of such association studies, the representativeness of neither the diseased nor the nondiseased members appears to have been established.

This study was undertaken as part of an epidemiological study of AD (Chandra et al. 1998); we have previously reported the association between *APOE*4* and AD (Ganguli et al., 2000). The purpose of this article is to provide basic data on the frequencies of *APOE* genotype in a large population-based sample aged 55+ in North India. This report is unique for three reasons: (1) data are derived from a representative and well-defined community sample; (2) genotype and allele frequencies are reported in the overall sample as well as by age, sex, and religion/caste groups; (3) this is the largest sample ever to be described from the Indian subcontinent.

Materials and Methods

This work was carried out as part of the Indo-US Cross-National Dementia Epidemiology Study, a collaboration between the Centre for Ageing Research in India (New Delhi, India) and the University of Pittsburgh (Pittsburgh, Pennsylvania, USA) (Chandra et al. 1998). As part of this project, a large sample was genotyped for the *APOE* polymorphism in the Department of Genetics, University of Delhi South Campus, New Delhi, India, in collaboration with the University of Pittsburgh.

The Ballabgarh study site, in the northwestern Indian state of Haryana, consists of 28 contiguous villages, 1 to 15 km apart, with a total population of approximately 60,000, considered representative of rural Haryana. These villages have constituted the Community Medicine Field Practice Site of the All-India Institute of Medical Sciences in New Delhi, India, for the past three decades. A total of 5126 subjects aged 55+ were recruited, representing a 99.84% response rate among all individuals aged 55 years and older living in the 28 study villages. Blood draws for *APOE* genotyping were performed on 86.8% of the original sample or 4450 individuals (385 subjects died before blood could be drawn, 120 subjects refused blood draws, 31 had relocated, 68 were not available for blood draw after multiple visits, and polymerase chain reaction–based genotyping was unsuccessful in 72 specimens). Our

study sample may therefore be considered representative of the age 55+ population of this region of India.

Samples. Finger-stick blood samples were collected from 4450 individuals and air-dried on Guthrie cards. DNA was isolated using QIAmp blood kits from QIAGEN Inc. (Chatworth, CA) using the recommended protocol for blood and body fluids. Genomic DNA was amplified in a 50 µL-reaction volume using a forward primer (5′-GCGGACATGGAGGACGTG-3′) and a reverse primer (5′-GGCCTGGTACACTGCCAG-3′) with a 10 X polymerase chain reaction (PCR) buffer and Taq polymerase (GIBCO-BRL) along with 1 m*M* $MgCl_2$; 7% DMSO and 200 µ*M* dNTPs (Kamboh et al. 1995). PCR was carried out with an initial denaturation at 95°C for 5 min, followed by 35 cycles of 95°C for 30 sec, 51°C for 30 sec, and 70°C for 30 sec. The 177-base-pair (bp) PCR product was digested with *Hha*I (NEB, USA) followed by electrophoresis on an 8% nondenaturing polyacrylamide gel in 1 X TBE buffer (using Protean II xi cell [Biorad] electrophoresis system). Photographic documentation was done on Pharmacia Biotech Image Master VDS. Reliability was established between the laboratories by blind genotyping in the Indian laboratory of specimens provided from the US laboratory (Dr. M.I. Kamboh, Department of Human Genetics, University of Pittsburgh Graduate School of Public Health), and by sending photographs of gels prepared in the Indian laboratory to be read in the US laboratory.

Statistical Analyses. *APOE* allele frequencies were calculated using allele counting. Hardy-Weinberg equilibrium was tested using a χ^2 goodness-of-fit test. Sex, age, and caste/religious differences were assessed for significance using χ^2 tests.

Results

Of the 4450 subjects who were genotyped, 52.0% were men. The mean (SD) age was 67.1 (7.21) years with a median of 65.7 years and a range of 56–96 years at the time of blood draw. Ninety percent of women and eighty-four percent of men were under age 75.

Table 1 shows the overall distribution of *APOE* genotypes by age. As expected, the *E3-3* genotype was the most predominant in all age groups and in both sex groups. The distribution of *APOE* genotypes in this study sample shows a slight deviation from the Hardy-Weinberg expectation (χ^2 with 3 degrees of freedom [*df*] was 9.12 compared to the critical value of 7.81, $p = 0.028$). The deviation comes largely from an excess of *E2*-2s (observed 13 versus expected 6.84) and to a lesser extent from an excess of *E2-4*s (observed 33 versus expected 25.64).

The overall frequencies of *APOE E**2, *E**3, and *E**4 alleles in this study sample were 0.039, 0.887, and 0.073, respectively. These results are shown

Table 1. Ballabgarh Sample: *APOE* Genotype by Age Group

Age Group (Years)	*APOE Genotype n (%)*						*Total n (%)*
	E2-2	*E3-3*	*E4-4*	*E2-3*	*E2-4*	*E3-4*	
55–64	6	1553	11	154	17	283	2024
	(0.13)	(34.90)	(0.25)	(3.46)	(0.38)	(6.36)	(45.48)
65–74	6	1516	8	103	13	213	1859
	(0.13)	(34.07)	(0.18)	(2.31)	(0.29)	(4.79)	(41.78)
75–84	1	374	2	27	2	66	472
	(0.02)	(8.40)	(0.04)	(0.61)	(0.04)	(1.48)	(10.61)
85 +	0	72	1	6	1	15	95
	(0.00)	(1.62)	(0.02)	(0.13)	(0.02)	(0.34)	(2.13)
All Ages	13	3515	22	290	33	577	4450
	(0.29)	(78.99)	(0.49)	(6.52)	(0.74)	(12.97)	(100.00)

in Table 2 along with allele frequencies in the different age groups and for both sexes. No statistically significant differences were found among allele frequencies for the various age groups, or between men and women within age groups, by χ^2 tests.

Table 3 shows *APOE* frequencies within the Hindu caste (Brahmin, Kshatriya, Vaish, Shudra) and Muslim groups comprising the study sample. χ^2 tests of independence across the sample showed no significant differences in frequencies of either *APOE**3 ($\chi^2 = 0.49$ with 4 *df*, $p = 0.92$) or *APOE*4* ($\chi^2 = 5.58$ with 4 *df*, $p = 0.23$) alleles. The distribution of the *APOE*2* allele showed significant differences ($\chi^2 = 16.42$ with 4 *df*, $p < 0.01$); contributions to the χ^2 value were largely from Brahmins who showed a greater-than-expected *E*2* frequency, and Kshatriyas who showed a lower-than-expected *E*2* frequency.

Discussion

We have reported the distribution, according to age, sex, and caste/religion, of *APOE* genotypes and allele frequencies in a large, rural population-based sample of older adults from India. No differences were found by age or sex. The homozygous *E3-3* genotype was the most prevalent in all examined age groups. Of the 4450 individuals genotyped, only 22 were homozygous *E4-4*, and the overall *E*4* allele frequency was 0.073 in this study sample.

Other groups have reported low *E*4* frequencies (Table 4), notably 0.065 in a Greek blood bank sample (Sklavounou et al. 1997), 0.074 in a Chinese blood bank sample (Hallmann et al 1991), and 0.126 in a Catalonian

Table 2. *APOE* Allele Frequencies by Age and Sex

	55–64 years			*65–74 years*			*75–84 years*			*85 + years*			*Overall*		
	Men	*Women*	*Both*	*Men*	*Women*	*Both*	*Men*	*Women*	*Both*	*Men*	*Women*	*Both*	*Men*	*Women*	*Both*
	n =	*n* =	*n* =	*n* =	*n* =	*n* =	*n* =	*n* =	*n* =	*n* =	*n* =	*n* =	*n* =	*n* =	*N* =
APOE	935	1089	2024	1016	843	1859	297	175	472	68	27	95	2316	2134	4450
*E*2*	0.044	0.046	0.045	0.033	0.037	0.034	0.034	0.031	0.033	0.044	0.019	0.037	0.038	0.041	0.039
*E*3*	0.880	0.871	0.875	0.904	0.897	0.900	0.891	0.891	0.891	0.853	0.907	0.868	0.891	0.884	0.887
*E*4*	0.076	0.083	0.080	0.064	0.066	0.065	0.076	0.077	0.076	0.103	0.074	0.095	0.072	0.076	0.073

Table 3. *APOE* Allele Frequencies within Religion/Caste Groups

	Hindu Castes					
APOE	*Brahmin n = 793*	*Kshatriya n = 1926*	*Vaish n = 115*	*Shudra n = 1552*	*Muslim n = 64*	*Overall N = 4450*
*E*2*	0.050	0.031	0.030	0.043	0.070	0.039
*E*3*	0.886	0.889	0.896	0.886	0.875	0.887
*E*4*	0.064	0.080	0.074	0.071	0.055	0.073

paternity testing sample (Gene et al. 1997). Mastana et al. (1998) reported an even lower *E*4* frequency (0.033) in a sample of 60 Baiga tribals from Madhya Pradesh, India. Previous studies were based on much smaller samples than ours and did not report allele distributions by age. Since *E*4* alleles are associated with several life-shortening diseases, individuals with one or two *E*4* alleles may be prone to early mortality, thus resulting in a lower *E*4* frequency in elderly or medically underserved populations. Our own data do not at present permit a test of this hypothesis. However, other workers (Cauley et al. 1993) have shown significantly lower *E*4* frequencies in older compared to younger populations, and *E*4* predicts mortality from both cardiovascular (Stengard et al. 1996) and cerebrovascular (Alberts et al. l 995) disease.

APOE allele frequencies for samples of individuals from India, or of Indian origin residing elsewhere, have previously been reported by others from relatively small convenience samples (Table 4). India has an ethnically and linguistically heterogeneous population. Based upon previous genetic studies in present-day Indians, linguistic differences (reflecting underlying ethnic differences) appear to be strongly associated with genetic diversity (Cavalli-Sforza et al. 1996). This study was conducted in a particular region of northern India where the inhabitants speak an Indo-European language (the Haryanvi dialect of Hindi). In a study of two states neighboring our study area, Indians speaking the Punjabi and Hindi languages, respectively, reportedly had Caucasoid genetic features and at least one marker suggesting ancient East Asian background (Passarino et al. 1996). This finding supports the likelihood of mixed ancestry for the Indian population. Given recent interest in genetic polymorphisms among ethnic (caste and tribal) and other putatively endogamous subgroups in Indian populations (Majumder et al. 1999, Mukerjee et al. 1999), we have also reported *APOE* allele frequencies among the four broad Hindu caste (Brahmin, Kshatriya, Vaish, Shudra) groups and the Muslim group, which together constitute our study sample. The statistically significant caste differences in *E*2* frequency (lower than expected in Kshatriyas and higher than expected in Brahmins) do not appear to have immediate explanations or implications; the lack of other significant differences may be due to the disparity in sample sizes among the different

Table 4. *APOE* Allele Frequencies in Different Studies

Author (Year)	*Source of Sample*	*Sample Size*	*E*2*	*E*3*	*E*4*
	*Previous Reports of Low E*4 Frequency Worldwide*				
Gene et al. (1997)	Individuals (mean age 36 years) with unrelated parents, sampled for paternity testing, from Catalonia, Spain	226	0.064	0.810	0.126
Sklavanou et al. (1997)	Healthy blood donors (mean age 35.6 years) from different parts of Greece, visiting National Blood Bank Centers in Greece	216	0.053	0.882	0.065
Hallman et al. (1991)	Unrelated blood bank donors in Singapore (Chinese individuals)	190	0.097	0.829	0.074
	Reports of Indian Samples				
Hallmann et al. (1997)	Unrelated blood bank donors in Singapore (Indian individuals)	142	0.046	0.827	0.127
Gounden et al. (1995)	Female Indian nurses (ages 25–55 years) employed at 3 major hospitals in Durban, South Africa	173	0.012	0.876	0.127
Mastana et al. (1998)	Unrelated random donors:				
	"Indian castes" from Bombay	305	0.067	0.818	0.115
	Baiga tribals from Madhya Pradesh	60	0.042	0.925	0.033
Kamboh (unpublished)	Random donors, Brahmins from Madras	68	0.088	0.809	0.103
Thelma et al. (this study)	Community-based rural sample aged 55 +, mean age 67.1 years, from Ballabgarh, India	4450	0.039	0.887	0.073

groups. We would urge caution in drawing inferences about the degree of homogeneity within these groups, or between them and the same religious/caste categories in different regions of India. Although ours is certainly the largest Indian sample that has been genotyped for *APOE*, it should not be considered representative of the entire Indian population.

The population of India represents approximately one-sixth that of the world. This report is the first from the Indian subcontinent, and one of few worldwide, to report *APOE* frequencies for a representative community sample. Knowledge about allele frequencies in the base population is essential to determine the extent to which case-control samples (on which most association studies are based) are biased or representative of the population. One example of such volunteer bias was demonstrated in a sample of AD cases in a research center, where patients had higher *APOE*4* frequency than did AD cases identified in a community sample from the same area (Tsuang et al. 1996). Similarly, nondiseased "controls" who volunteer for research may not be representative of the unaffected base population. Thus, our population-based allele frequency data provide useful background information for any future studies of the association between *APOE* polymorphism and diseases in elderly Indian samples.

Acknowledgments The authors gratefully acknowledge the contributions of Prachi Semwal and Poonam Juyal for assistance with DNA isolation and PCR and Anjali Prashar for data entry, at the Department of Genetics, University of Delhi South Campus; M.I. Kamboh of the Department of Human Genetics, University of Pittsburgh Graduate School of Public Health, for his valuable expertise in establishing genotyping reliability and useful comments on the manuscript; Janet M. Johnston for earlier assistance with statistical analysis; at New Delhi, Arun Mehta of Indata Systems for software development, technical support, and quality control; at the Centre for Ageing Research in India, G. Binu and B.S. Nair for project management and data entry; R.K. Kaushik for field supervision; Suresh Kapoor, Guresh Kumar, and other staff of the CRHSP of the Centre for Community Medicine of the All-India Institute of Medical Sciences for their cooperation and for providing access to their facilities in Ballabgarh. Finally, the authors thank the senior citizens of Ballabgarh for their participation in the study.

The work reported here was supported in part by the National Institute of Aging of the National Institutes of Health (US Department of Health and Human Services) through grants AG07562 and AG09292.

Received 7 December 1999; revision received 1 June 2000.

Literature Cited

Alberts, M.J., C. Graffagnino, C. McClenny et al. 1995. ApoE genotype and survival from intracerebral hemorrhage. (Letter.) *Lancet* 346:575.

Assman, G., H. Funke, and H.U. Jabs. 1987. Analytical procedure for the detection and characterisation of apolipoprotein E mutants. *Am. Heart J.* l 113:598–602.

Cauley, J.A., J.E. Eichner, M.I. Kamboh et al. 1993. APOE E allele frequencies in younger (age 4250) vs older (age 65–90) women. *Genetic Epidemiol.* 10:27–34.

Cavalli-Sforza, L.L., P. Menozzi, and A.Piazza. 1996. Asia. In: *The History and Geography of Human Genes.* Princeton, NJ: Princeton University Press, 195–254.

Chandra, V., M. Ganguli, R. Pandav et al. 1998. Prevalence of Alzheimer's and other dementias in rural India: The Indo-US study. *Neurology* 51:1000–1008.

Corder, E.H., A.M. Saunders, W.J. Strittmater et al. 1993. Gene dose of apolipoprotein E allele E4 and the risk of Alzheimer's disease in late-onset families. *Science* 261:921–923.

Davignon, J., R.E. Gregg, and C.F. Sing. 1988. Apolipoprotein E polymorphism and atherosclerosis. *Arteriosclerosis* 8:1–21.

Des, H.K., J. Mcpherson, G.A. Bruns et al. 1985. Isolation, characterization and mapping to chromosome 19 of the human apolipoprotein E gene. *J. Biologic. Chem*. 260:6240–6247.

Evans, DA., L.A. Beckett, T.S. Field et al. 1997. Apolipoprotein E e4 and incidence of Alzheimer disease in a community population of older persons. *J.A.M.A.* 277:822–824.

Ganguli, M., V. Chandra, M.I. Kamboh et al. 2000. APOE polymorphism and Alzheimer's disease: the Indo-US Cross-National Dementia Study. *Archives of Neurology*. 57:824–830.

Gene, M., P. Moreno, M. Ezquerra et al. 1997. Low apolipoprotein E epsilon 4 allele frequency in the population of Catalonia (Spain) determined by PCR-RFLP and Laser fluorescent sequencer. *Eur. J. Epidemiol.* 13:841–843.

Gounden, N., J. Naidoo, R.J. Pegoraro et al. 1995. Apolipoprotein E allele frequencies in a South African Indian female population. *Clin. Genetics* 48:243–245.

Hallmann, D.M., E. Boerwinkle, N. Saha et al. 1991. The apolipoprotein E polymorphism: A comparison of allele frequencies and effects in nine populations. *Am. J. Hum. Genetics* 49:338–349.

Hendrie, H.C. 1998. Epidemiology of dementia and Alzheimer's disease. *Am. J. Geriatric Psychiatry* 6:S3–S18.

Hixon, J., and P.R. Group. 1991. Apolipoprotein E polymorphisms affect atherosclerosis in young males. *Arteriosclerosis and Thrombosis* 11:1237–1244.

Kalman, J., A. Juhasz, A. Csaszar et al. 1998. Increased apolipoprotein E4 allele frequency is associated with vascular dementia in the Hungarian population. *Acta Neurologica Scandinavica* 98:166–1688.

Kamboh, M. I. 1995. Apolipoprotein E polymorphism and susceptibility to Alzheimer's disease. *Hum. Biol.* 67(2):195–215.

Kamboh, M.I., D.K. Sanghera, R.E. Ferrell et al. 1995. APOE*4-associated risk is modified by α-antichymotrypsin polymorphism. *Nature Genetics* 10:486–488.

Majumder, P.P., B. Roy, S. Banerjee et al. 1999. Human-specific insertion/deletion polymorphisms in Indian populations and their possible evolutionary implications. *Eur. J. Hum. Genetics* 7:435–446.

Mastana, S.S., R. Calderon, J. Pena et al. 1998. Anthroplogy of the apolipoprotein E (apoe E) gene: Low frequency of apo E4 allele in Basques and in tribal (Baiga) populations of India. *Ann. Hum. Biol.* 25:137–143.

Mukerjee, N., P.P. Majumder, B. Roy et al. 1999. Variation at 4 short tandem repeat loci in 8 population groups of India. *Hum. Biol.* 71:439–46.

Passarino, G., O. Semino, L.F. Berninin et al. 1996. Pre-Caucasoid and Caucasoid genetic features of the Indian population, revealed by mtDNA polymorphisms. *Am. J. Hum. Genetics* 59:927–934.

Pedro-Botet, J., M. Senti, X. Nogues et al. 1992. Lipoprotein and apolipoprotein profile in men with ischemic stroke, role of lipoprotein(a), triglyceride-rich lipoproteins, and apolipoprotein E polymorphism. *Stroke* 23:1566–1562.

Saunders, A.M., W.J. Strittmater, D. Schmechel et al. 1993. Association of apolipoprotein E allele E4 with late-onset familial and sporadic Alzheimer's disease. *Neurology* 43:1467–1472.

Sklavounou, E., E. Economou-Petersen, G. Karadima et al. 1997. Apolipoprotein E polymorphism in the Greek population. *Clin. Genetics* 52:216–218.

Slooter, A.J.C., M-X. Tang, C.M. van Duijn et al. 1997. Apolipoprotein E e4 and the risk of dementia with stroke: A population-based investigation. *J.A.M.A.* 27:818–821.

Stengard, J.H., J. Pekkanen, C. Ehnholm et al. 1996. Genotypes with the apolipoprotein epsilon 4 allele are predictors of coronary heart disease mortality in a longitudinal study of elderly Finnish men. *Hum. Genetic*s 97:677–684.

Tsuang, D., W. Kukull, L. Sheppard, et al. 1996. Impact of sample selection on APOE e4 allele frequency: A comparison of two Alzheimer's disease samples. *J. Am. Geriatrics Society* 44:704–707.

Utermann, G. 1977. Apolipoprotein E polymorphism in health and disease. *Am. Heart J.* 113:433–439.

Wilson. P., E. Schaefer, M. Larson et al. 1996. Apolipoprotein E alleles and risk of coronary artery disease: A meta-analysis. *Arteriosclerosis, Thrombosis, and Vascular Biology* 16:1250–1255.