Letter to the Editor

No Fixed Nucleotide Difference Between Africans and Non-Africans at the Pyruvate Dehydrogenase E1 α-Subunit Locus

Ning Yu and Wen-Hsiung Li

Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637

Manuscript received February 11, 2000
Accepted for publication March 22, 2000

THE genetic diversity within and between human populations has been a topic of intense research for the past three decades because of its relevance to human genetics and evolution (Cavalli-Sforza 1969; Nei and Roychoudhury 1972). Of particular interest is the degree of genetic differentiation between Africans and non-Africans because such data are needed for resolving the controversy over the origin and evolution of modern humans. Major questions in this debate are (1) how geographically spread were the archaic and early modern humans, (2) were they divided into subpopulations, and (3) if so, how much migration occurred between subpopulations? To address these issues, Harris and Hey (1999) have recently obtained very interesting data from the X-linked pyruvate dehydrogenase E1 α-subunit (PDHA1) locus. They sequenced a 4200-bp region of the PDHA1 gene in 35 male individuals (6 French, 7 Chinese, 5 Vietnamese, 1 Mongolian, 4 Southern Bantu speakers, 6 Senegalese, 3 African pygmies, and 3 Khoisan speakers). Strikingly, a fixed nucleotide difference at position 544 distinguished the Africans from the non-Africans sampled: all African sequences studied have nucleotide G, whereas all non-African sequences studied have nucleotide A. (It is a difference between two synonymous alanine codons, GCG and GCA.) The fixed difference was taken as a strong indication of historical population subdivision between Africans and non-Africans. However, since only 35 males were sampled, it is not certain that the difference is really fixed between Africans and non-Africans. In view of the importance of this issue, we decided to sequence more individuals. As mentioned in Harris and Hey (1999), a larger sample could be expected to reveal some shared haplotypes. On the other hand, if the fixed difference holds up for a larger sample, it will be an extremely interesting observation.

We examined a geographically diverse set of 101 individuals (35 Africans, 33 Europeans, and 33 Asians). A PCR primer pair was designed to amplify a 456-bp region covering position 544 and two other polymorphic sites (positions 494 and 595) observed by Harris and Hey (1999). An internal sequencing primer was used to obtain sequences in the forward direction for the individuals who were heterozygous for an insertion (duplication) to be described below. A reliable sequence of 331 bp was obtained from each of 139 X chromosomes (Table 1); this segment includes 149 nucleotides of intron 7, 72 nucleotides of exon 8, and 110 nucleotides of intron 8. Among the sequences obtained, we found that four non-African sequences (three from mainland Italy and one from Sardinia) have G instead of A at position 544 and that one sequence from Africa (Nigeria) has A instead of G at position 544. Therefore, this site shows no fixed difference between Africans and non-Africans.

Within this short region, the Africans showed a higher nucleotide diversity than the non-Africans (Table 2). The observed value of 0.31% in the Africans is three times higher than the average nucleotide diversity at fourfold degenerate sites (0.11%) estimated by Li and Sadler (1991). Of course, the high diversity observed partly reflects our selection of a known polymorphic region. This selection notwithstanding, there is no nucleotide variation among the Asians studied. The nucleotide diversity among the Europeans studied is fairly high (0.17%) because the Italian samples contain four polymorphic nucleotide sites (see Table 3 and discussion below). The low diversity in Asians could be due to selection sweep or random drift.

The fact that the nucleotide difference at position 544 is not fixed has some implications. From their observation, Harris and Hey (1999) stated that the mutation causing the fixed difference probably arose in the presence of some population subdivision and then became fixed in the population in which it arose. Since the two known chimpanzee sequences have G at position 544 (Harris and Hey 1999), it is reasonable to assume that A is the derived nucleotide. Their statement then implies that this mutation occurred in non-Africans. This may still be true. However, we now note that nucleotide

Corresponding author: Wen-Hsiung Li, Department of Ecology and Evolution, University of Chicago, 1101 E. 57th St., Chicago, IL 60637. E-mail: whli@uchicago.edu

TABLE 1

The distribution of A and G nucleotides at site 544 in different populations

Continent	Populations	No. of females	No. of males	Total no. of sequences	Site 544	
					A	G
Asia	Han Chinese	3	7	13	13	
	Taiwanese aborigine	1	3	5	5	
	Vietnamese	0	5	5	5	
	Cambodian	3	2	8	8	
	Japanese	4	5	13	13	
Europe	Italian (mainland)	4	2	10	7	3
	Italian (Sardinian)	0	10	10	9	1
	French	0	5	5	5	
	Finnish	1	1	3	3	
	Hungarian	3	1	7	7	
	Swedish	3	3	9	9	
Africa	Nigerian	13	14	40	1	39
	Pygmies	3	3	9		9
	!Kung	0	1	1		1
	Kenyan	0	1	1		1
Total		38	63	139	85	54

The four individuals that have no gender information were treated as males.

A appears at position 544 in one of the 51 African sequences we sampled, a woman whose parents are also Africans. Although the possibility of earlier ancestral immigration and the possibility of a new G to A mutation cannot be excluded, one should consider also the alternative explanation that the mutation occurred in Africa. The plausibility of this explanation would increase if future investigation reveals that the frequency of nucleotide A at position 544 is not negligible among Africans; note that the frequency (1/51) in the present study is ~2%, not a very low value. Of special interest is the possibility that the mutation arose before the emergence of modern humans. In this respect, we note Harris and Hey's estimate of 189,000 YBP for the mutation, which is considerably older than the estimated age of modern humans (100,000–130,000 YBP).

The presence of nucleotide G at position 544 in 4 Italian sequences provides further evidence against a fixed difference between Africans and non-Africans. Of course, this presence may represent recent immigration from Africa to nearby Italy. However, our data suggest that the subdivision between Africans and non-Africans may not have been as strong as Harris and Hey (1999) stated, for the following reasons. First, the presence of 4 "G sequences" among the 20 Italian sequences we sampled represents a fairly high frequency (20%), so that it is unlikely to be due to recent migration. Second, the fact that the Sardinian G sequence carries a C to T mutation at position 595 (Table 3) may imply a relatively long history of this sequence in Sardinia. Third, whether the G sequence is absent in the other parts of Europe and in Asia is uncertain because our sample sizes are fairly small.

It is also interesting to examine the haplotypes. As in Harris and Hey (1999), we found the presence of a duplication relative to chimpanzee at positions 364–370 (GGCCAA) in more than half of the sequences we sampled. The haplotypes defined by this polymorphism and the four single nucleotide polymorphisms are listed in Table 3. The chimpanzee haplotype (*i.e.*, haplotype O) is not observed in any of the human sequences we sampled; all of the human sequences that lack the duplication have A instead of C at position 494. As every A nucleotide at position 544 is associated with the duplication GGCCAA, one may assume that the mutation from G to A at position 544 occurred after the duplication. It is interesting that all of the human sequences that do not carry the duplication have haplotype B (*i.e.*, A, G, C, and G at positions 494, 544, 595, and 672, respectively), except for one Sardinian sequence (haplotype C) that has a C→T mutation at position 595. Figure 1 shows the parsimony tree for the haplotype evolution. In one lineage, haplotype B was derived from haplotype O by a C→A mutation at position 494 and subsequently haplotype C was derived from haplotype

TABLE 2

Nucleotide diversity (%) within and between populations

Population	African	Asian	European
African	0.31		
Asian	0.65	0.0	
European	0.62	0.089	0.17

Note: The nucleotide diversity is 0.37% among all sequences, 0.086% among non-African sequences, and 0.64% between Africans and non-Africans.

TABLE 3

Haplotype distribution among different populations

Population	Haplotype	No. of sequences	Nucleotide position				
			365–370[b]	494	544	595	672
Chimpanzee[a]	O	2	—	C	G	C	G
Asians	A	44	GGCCAA	C	A	C	C
Italians (mainland)	A	7	GGCCAA	C	A	C	C
	B	3	—	A	G	C	G
Italians (Sardinia)	A	9	GGCCAA	C	A	C	C
	C	1	—	A	G	T	G
Other Europeans	A	24	GGCCAA	C	A	C	C
Nigerians	B	26	—	A	G	C	G
	D	13	GGCCAA	C	G	C	C
	A	1	GGCCAA	C	A	C	C
Pygmies	D	6	GGCCAA	C	G	C	C
	B	3	—	A	G	C	G
Kenyan	B	1	—	A	G	C	G
!Kung	D	1	GGCCAA	C	G	C	C

[a] The chimpanzee sequences are from Harris and Hey (1999).
[b] A duplication of the nucleotides GGCCAA at positions 359–364.

B by a C→T mutation at position 595. In the other lineage, haplotype D was derived from haplotype O by a duplication of GGCCAA at positions 359–364 (denoted by − → + in Figure 1) and a G→C mutation at position 672, and then haplotype A was derived from haplotype D by a G→A mutation at position 544.

Table 3 shows that all Asian and the majority of European sequences studied are of haplotype A. This haplotype is present in one of the African sequences (a Nigerian). Haplotype B is the major haplotype among Africans (30/51 = 59%) but is also observed in three Italian sequences. Thus, both major haplotypes (*i.e.*, A and B) are shared by Africans and non-Africans. In addition, haplotype C, which is found in a Sardinian, can also be counted as shared because it differs from haplotype B by only one mutation (C→T at position 595). The only haplotype that is present at a fairly high frequency and is not shared by Africans and non-Africans is haplotype D, which occurs at a frequency of 39% (20/51) among the Africans sampled.

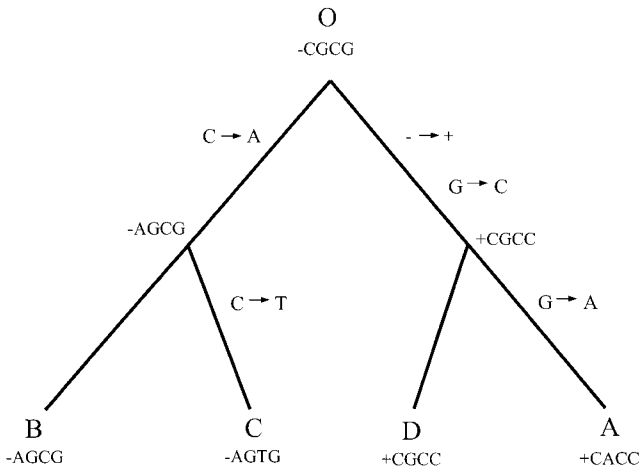

Figure 1.—Parsimony tree for the evolution of haplotypes. Haplotype O is the ancestral (chimpanzee) haplotype and haplotypes A, B, C, and D are the four human haplotypes. −, absence of the duplicated sequence GGCCAA at positions 365–370; +, presence of the duplication; →, the direction of mutation.

In conclusion, there is no fixed nucleotide difference between Africans and non-Africans in the region studied and, as foreseen by Harris and Hey (1999), a larger sample has revealed shared haplotypes. The differences between Africans and non-Africans in nucleotide frequencies at position 544 and in the distribution of haplotypes do indicate substantial population subdivision between the two groups, but the subdivision has likely been weaker than that suggested by Harris and Hey's data.

We thank Drs. Anna Di Rienzo, Lynn Jorde, Marie Lin, Pekka Pamilo, Laszlo Patthy, and MaryEllen Ruvolo for DNA samples and Jody Hey, N. Pearson, Z. Zhao and the reviewers for suggestions. This study was supported by National Institutes of Health grant GM55795.

LITERATURE CITED

Cavalli-Sforza, L. L., 1969 Human diversity. Proceedings of the 12th International Congress on Genetics, Tokyo, Vol. 3, pp. 405–416.

Harris, E. E., and J. Hey, 1999 X chromosome evidence for ancient human histories. Proc. Natl. Acad. Sci. USA **6:** 3320–3324.

Li, W.-H., and L. A. Sadler, 1991 Low nucleotide diversity in man. Genetics **129:** 513–523.

Nei, M., and A. K. Roychoudhury, 1972 Gene differences between Caucasian, Negro, and Japanese populations. Science **177:** 434–436.

Communicating editor: Y.-X. Fu