Deletion Polymorphism in the Human COL1A2 Gene: Genetic Evidence of a Non-African Population Whose Descendants Spread to All Continents

R.J. MITCHELL,[1] S. HOWLETT,[1] N.G. WHITE,[1] L. FEDERLE,[1] S.S. PAPIHA,[2] I. BRICENO,[2] J. MC COMB,[3] M.S. SCHANFIELD,[4] C. TYLER-SMITH,[5] L. OSIPOVA,[6] G. LIVSHITS,[7] AND M.H. CRAWFORD[3]

Abstract We report the frequencies of a deletion polymorphism at the $\alpha 2$ (1) collagen gene (*COL1A2*) and argue that this distribution has major implications for understanding the evolution of modern humans immediately after their exodus from sub-Saharan Africa as well as their subsequent spread to all continents. The high frequency of the deletion in non-African populations and its complete absence in sub-Saharan African groups suggest that the deletion event occurred just before or shortly after modern humans left Africa. The deletion probably arose shortly after the African exodus in a group whose descendants were among the ancestors of all contemporary populations, except for sub-Saharan Africans. This, of course, does not imply that there was a single migration out of Africa. The *GM* immunoglobulin haplotype *GM*A,X G* displays a similar distribution to that for the *COL1A2* deletion, and these 2 polymorphisms suggest that the exodus from Africa may not have been a rapid dispersion to all other regions of the world. Instead, it may have involved a period of time for the savanna-derived gene pool to adapt to novel selective agents, such as bacteria, viruses, and/or environmental xenobiotics found in both animal and plant foods in their new environment. In this context these polymorphisms are indicators of the evolution that occurred before the diaspora of these populations to the current distribution of modern peoples.

[1] Department of Genetics and Human Variation, La Trobe University, Bundoora, Victoria, Australia 3083.

[2] Department of Human Genetics, University of Newcastle upon Tyne, Newcastle upon Tyne, England.

[3] Department of Anthropology, University of Kansas, Lawrence, KS.

[4] Analytical Genetic Testing Center, Denver, CO.

[5] Department of Biochemistry, University of Oxford, Oxford, England.

[6] Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia.

[7] Department of Anatomy and Anthropology, Sackler School of Medicine, Tel Aviv, Israel.

Human Biology, December 1999, v. 71, no. 6, pp. 901–914.

KEY WORDS: *COL1A2* DELETION POLYMORPHISM, EVOLUTION, ORIGINS OF MODERN HUMANS, OUT OF AFRICA HYPOTHESIS, *GM* HAPLOTYPE, RECURRENT MUTATION

In recent years DNA polymorphisms have been used to reconstruct hominid phylogeny. In particular, studies of mtDNA (Vigilant et al. 1991; Stoneking and Soodyall 1996) and closely linked polymorphic loci at the autosomal *CD4* locus (Tishkoff et al. 1996) have been used to generate a hypothesis of a single origin of modern *Homo sapiens* in sub-Saharan Africa followed by a relatively rapid spread from there (probably through the Near East and Asia) to the rest of the world, replacing earlier hominid forms that may have lived there. Y-chromosome-specific polymorphisms do not refute this hypothesis (Hammer et al. 1997, 1998), and analysis of the worldwide variability in microsatellite polymorphisms also supports this "out of Africa" model (Jorde et al. 1997). The genetic data further suggest that the time required for these events to have occurred has been insufficient to generate discrete human gene pools. It is the greater genetic diversity among sub-Saharan Africans that is the essential ingredient in the concept of sub-Saharan Africa being the epicenter of modern *Homo sapiens* evolution, because 1 suggestion is that greater diversity correlates with a greater age for sub-Saharan Africans under neutral models of evolution. The population size of the earliest human group is unknown, but variation would also have a significant impact on levels of genetic diversity.

After the spread of modern humans from Africa novel polymorphisms evolved in distinct ethnic and even continental groups, some of which are highly restricted geographically and not present in contemporary sub-Saharan Africans. However, a locus that displays polymorphism in a large number of populations that are widely distributed geographically but that shows monomorphism in sub-Saharan Africans would be of considerable anthropological interest. Provided that the locus is not subject to recurrent mutation, the existence of such a polymorphism suggests (1) that there is a common origin of all groups of modern humans, (2) that this particular polymorphism may have been 1 of the first to evolve in the human groups that constituted the exodus out of Africa, and (3) that this migrant population, through its descendants, contributed to all present non-African populations.

The present study reports on an autosomal polymorphism in the human $\alpha 2$ (1) collagen gene (*COL1A2*) that may well fit the description. *COL1A2* is 1 of 2 genes that encode for the polypeptides of type 1 collagen. Type 1 collagen is a hydrogen-bound trimer of 2 $\alpha 1$ chains and 1 $\alpha 2$ chain; it is the major constituent of skin, bone, tendons, ligaments, blood vessels, dentin, and many interstitial tissues (Pepe, Bue et al. 1995). The *COL1A2* polymorphism that is the focus of the present study is an insertion or deletion in an intron within the human *COL1A2* gene that is located on chromosome 7q21.3–q22.1. The polymorphism was first recognized through restriction site (*Rsa*I) analysis (Grobler-Rabie et al. 1985) but was subsequently shown to be an insertion or deletion of a 38 base pair (bp) sequence detectable by a standard polymerase chain reaction (PCR) protocol (Watson and Dalgleish 1990). The

larger allele, containing the 38-bp sequence, is 641 bp in size, and the smaller allele is 603 bp.

Previous studies of the distribution of the *COL1A2* polymorphism have indicated that it is a useful marker in anthropological studies. Grobler-Rabie et al. (1985) showed that it could differentiate 3 populations in South Africa. Although the smaller allele (603 bp) was absent in South African blacks, nearly 40% of Afrikaners carried it, and it was also present in 11% of Cape Coloreds. Subsequently, it was shown that the incidence of the 603-bp allele in Caucasians was also high, ranging between 24% and 38%, and similar frequencies have been reported in Asian groups (Bowcock et al. 1987, 1991; Pepe et al. 1990; Borresen et al. 1988; Baker et al. 1991; Pepe, Rickards et al. 1995). The 603-bp allele was absent in West Africans (Pepe, Rickards et al. 1995) and Central African Pygmies (Bowcock et al. 1991). In all populations sampled to date, the 641-bp allele never has a frequency less than 60%, suggesting that the polymorphism is the result of a deletion event.

To further investigate the distribution of the *COL1A2* deletion in worldwide populations, we scored the *COL1A2* marker in 10 widely distributed human groups that represent different degrees of geographic isolation and lifestyle, ranging from small-scale hunter-foragers to large mainstream groups. The status of this polymorphism was also investigated in our nearest relatives, the great apes, to identify the ancestral allele at the locus. Our results confirm the previously observed pattern of *COL1A2* deletion distribution and support the out of Africa hypothesis for modern human origins and dispersal throughout the Old World.

Subjects and Methods

Subjects. In this study we analyzed 219 samples from 10 populations representing each of the major continental areas except Europe, for which there already exists a considerable data set (Figure 1, Table 1). These populations ranged in size from small hunter-forager groups to peasant societies. Four aboriginal populations from central Siberia (Altai, Evenki, Kets, and Sel'kups) and the Chuvash, who reside in the Urals approximately 1000 km east of Moscow, were sampled for this study. The Kizhi Altai, who are pastoralists, were all sampled from the village of Mendor-Sokkon, near the Katun River in the Altai mountains. The samples from the reindeer hunting and herding Evenki were collected in the villages of Poligus and Surinda on the Stony Tunguska River, a tributary of the Yenisey River. Speaking a unique language with no known linguistic affiliation, the Kets were all living in the village of Sulamai, which lies at the juncture of the Yenisey and Stony Tunguska rivers. The Sel'kup samples were obtained from 5 different villages (Farkovo, Rechka, Krasnoselkup, Ratta, and Tolka Pur). The Chuvash were from villages located near the towns of Ceboksary and Zel'onodol'sk.

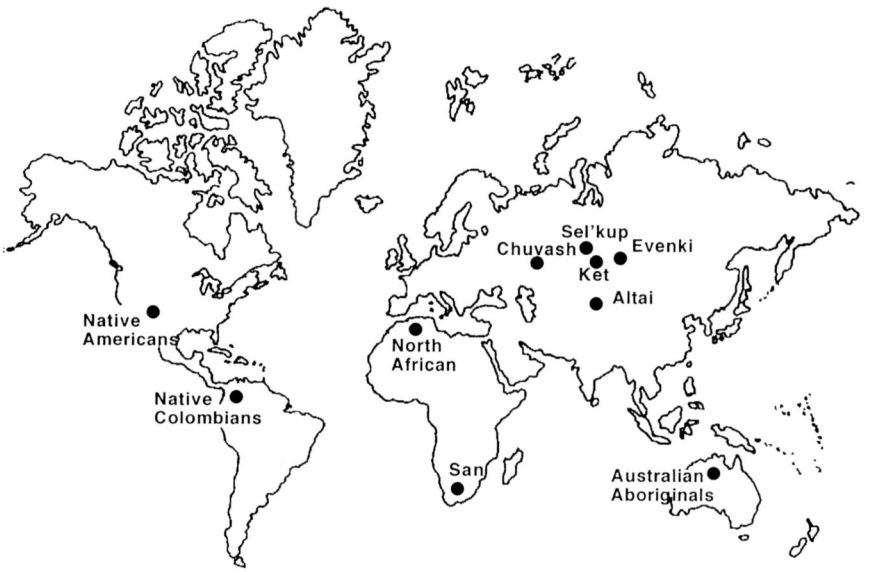

Figure 1. Locations of the 10 population samples scored for the *COL1A2* polymorphism.

Two native American populations were also included in this study. The native Americans from the United States are mainly from Colorado and Arizona and include Na Dene and Amerind speakers. The Colombian native American samples are from 3 tribal groups: the Coreguaje, who inhabit the Amazon region and belong to the Tukano linguistic group, and the Arsario and the Kogui, who inhabit the mountainous region of Sierra Nevada de Santa Marta in northeast Colombia and belong to Chibcha linguistic stock. One Oceanic population was also included. The Australian Aboriginal samples are of Yolngu people from North East Arnhem Land in the Northern Territory. A small North African sample was also included, consisting of individuals from Mauritania, Algeria, Tunisia, and Morocco. In addition, 7 San samples from southern Africa were also included.

The nonhuman primate samples comprised 11 common chimpanzees (*Pan troglodytes*), 3 bonobos (*Pan paniscus*), 5 gorillas (*Gorilla gorilla*), and 9 orangutans (*Pongo pygmaeus*).

Methods. Most samples were received in the form of DNA, and those obtained as blood samples had the DNA extracted using standard procedures (Sambrook et al. 1989).

Table 1. Distribution of Genotypes and Alleles at the *COL1A2* Locus

Population	*Number Tested*	*COL1A2 Genotype Number and Allele Frequency*				
		641/641 bp	*641/603 bp*	*603/603 bp*	*641 bp*	*603 bp*
Chuvash, Russia	34	17	15	2	0.72	0.28
Evenki	16	8	5	3	0.66	0.34
Ket	9	5	2	2	0.67	0.33
Sel'kup	27	16	9	2	0.76	0.24
Altai	53	36	15	2	0.82	0.18
US native Americans	19	12	5	2	0.76	0.24
Native Americans, Colombia	27	17	8	2	0.78	0.22
Australian Aboriginals	17	12	4	1	0.82	0.18
North Africa	10	6	3	1	0.75	0.25
San	7	7	0	0	1.0	0.0
Total	219	136	66	17	0.77	0.23

The PCR protocol used to score the *COL1A2* locus is described by Watson and Dalgleish (1990). The primers are 5′-TCAGTGTATGTTGCTATCAG-3′ and 5′-ATTCCACAGTCAACATCAAC-3′. The 2 primers were used for both the human and the nonhuman primate samples. Amplification was carried out in a 20-μl volume in a Perkin Elmer Gene-Amp PCR System 2400 Thermocycler for 30 cycles of 94°C for 1 min, 50°C for 1 min, and 72°C for 1.5 min. The amplified product was analyzed directly by electrophoresis in a 2% agarose gel run at 80 V and 80 mA. The alleles were detected by staining with ethidium bromide, and the size marker used was ϕx174/*Hae*III. The 3 genotypes detected are shown in Figure 2.

A direct association between phenotype and genotype was assumed, and allele frequencies were determined by direct count. Deviation from Hardy-Weinberg equilibrium of genotypes was examined using the chi-square test. Differences among population samples were measured by means of contingency chi-square analysis. The Kruskal-Wallis test was performed [using Statmost 3.0 (Dataxiom) Software, Los Angeles, California] to determine whether k samples were drawn from the common population or from a population of equal means.

Results

Testing of Nonhuman Primate Samples. The *COL1A2* locus was amplified in gorillas, chimpanzees, and bonobos but could not be amplified in orangutans. Annealing temperatures were lowered by 2°C in an attempt to amplify the sequence in orangutans, but no product was detected. Because the amplified product in African apes was identical in size to the 641-bp

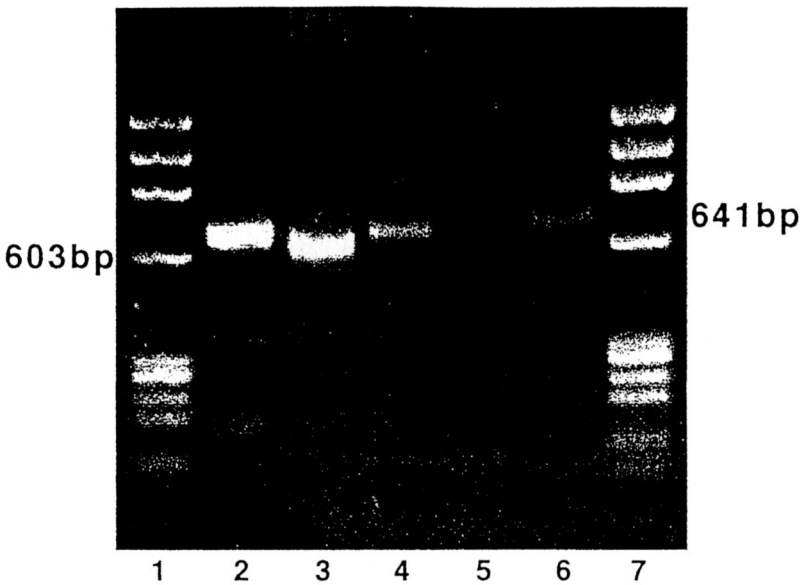

Figure 2. Agarose gel electrophoresis of *COL1A2* phenotypes. Lanes 1 and 7, ϕX174/*Hae*III size marker; lane 2, human, 641/641 bp; lane 3, human, 603/603 bp; lane 4, chimpanzee (*Pan troglodytes*), 641/641 bp; lane 5, orangutan, no amplified product; lane 6, gorilla, 641/641 bp.

fragment in humans (Figure 2), each of these was then restricted with *Rsa*I to confirm that the 38-bp sequence contained an *Rsa*I site in all 3 species. Both human and ape alleles contained an *Rsa*I site, and the fragments generated by the enzyme were the same in each after electrophoresis in a 2% agarose gel (data not shown).

The comparative data indicate that the *COL1A2* polymorphism arose as a result of a deletion of a 38-bp sequence and that it is restricted to *Homo sapiens*. The unsuccessful amplification of *COL1A2* alleles in the orangutan suggests that there may be sequence differences in 1 or both of the primers between orangutans and the African apes and humans. Because collagen is essential to mammalian development, the *COL1A2* gene must be present in the orangutan. Consequently, it should be possible to confirm its presence by hybridizing an appropriate *COL1A2* probe with *Rsa*I digested orangutan genomic DNA. Unfortunately, insufficient orangutan DNA was available for Southern blot analysis in this study, leaving this question to be explored at another time.

Distribution of *COL1A2* Polymorphism. Table 1 shows the distribution of the *COL1A2* deletion in the 10 populations of our study. Where sample size permitted appropriate statistical analysis, no significant deviation from Hardy-Weinberg genotype proportions was observed. The deletion allele (603

bp) was absent in the San but present in all other samples, ranging in frequency from a low of 18% in both the Altai of Siberia and Australian Aborigines to a maximum of 34% in the Evenki. Contingency chi-square analysis showed no significant difference in the distribution of alleles among the 9 non-sub-Saharan African populations in Table 1 ($\chi^2 = 13.9$, 9 d.f., $p > 0.05$).

Pooling all available data on the distribution of the *COL1A2* deletion (Table 2) demonstrates that it is a marker of significance for understanding both the origin and subsequent spread of modern humans. The deletion allele is present, usually at a relatively high frequency (>20%), in all populations with the exception of the inhabitants of sub-Saharan Africa, among whom it is absent. Of 528 sub-Saharan African chromosomes scored for this polymorphism only 4 deletion alleles (<1%) were identified, and all of these occurred among the coastal dwelling Fon people of southern Benin, West Africa, suggesting they may have been acquired through admixture. In contrast, among the non-sub-Saharan African populations included in the present study there were 100 deletions (or 24%) among 424 chromosomes scored. In 31 Solomon Islanders the deletion was absent, but this may be explained by the nonnegligible number of relatives in this sample (Bowcock et al. 1987, 1991). If confirmed, however, its absence may be a result of a founder effect and/or subsequent bottlenecks in this small archipelago. In addition, the presence of the deletion allele in our Oceania sample and also in ethnic groups of Indonesia at a frequency greater than 15% (unpublished observations) supports the notion that the polymorphism is present in the Southeast Asia–Oceania region.

Most of the populations found in Table 2 were tested for homogeneity using the Kruskal-Wallis test. The populations were Europeans ($n = 12$) excluding North Africa but including US whites and South African Afrikaners; Asians ($n = 7$); native Americans ($n = 3$); Oceanic groups ($n = 2$); and sub-Saharan Africans ($n = 8$) excluding South African Coloreds and Afrikaners and including South African blacks. The data set included 32 samples from 5 world regions. The results were highly significant ($H = 21.318$, 4 d.f., $p = 0.0003$). Removal of sub-Saharan Africans yielded a nonsignificant result for the remaining populations ($H = 7.491$, 3 d.f., $p = 0.058$). Comparison of the combined 4 world samples versus sub-Saharan Africans gave a highly significant result ($H = 15.857$, 1 d.f., $p = 0.00007$). No correction was made for multiple tests, given the level of significance. Thus the *COL1A2* deletion polymorphism clearly separates sub-Saharan Africans from all other groups.

Discussion

Our study indicates that the *COL1A2* polymorphism arose as a result of a deletion. Insertions and deletions are 1 of the major sources of DNA

Table 2. Distribution of the *COL1A2* Deletion (603 bp) Allele in Worldwide Populations

Population	*Number Tested*	*Allele 603 bp*	*Reference*
Africa			
North Africa	10	0.25	Present study
Benin, Bariba	32	0.0	Pepe, Rickards et al. (1995)
Benin, Berba	19	0.0	Pepe, Rickards et al. (1995)
Benin, Dendi	50	0.0	Pepe, Rickards et al. (1995)
Benin, Fon	49	0.04	Pepe, Rickards et al. (1995)
Biaka Pygmy, Central African Republic	45	0.0	Bowcock et al. (1991)
Mbuti Pygmy, Zaire	53	0.0	Bowcock et al. (1991)
San	7	0.0	Present study
South African blacks	9	0.0	Grobler-Rabie et al. (1985)
South African Cape Colored	9	0.11	Grobler-Rabie et al. (1985)
South African Afrikaner	24	0.38	Grobler-Rabie et al. (1985)
Europe			
England	35	0.34	Sykes et al. (1986)
England	28	0.36	Watson and Dalgleish (1990)
England	51	0.28	Baker et al. (1991)
Italy, Calabria	74	0.28	Pepe et al. (1990)
Italy, central	44	0.24	Pepe et al. (1990)
Italy, Sardinia	76	0.32	Pepe, Bue et al. (1995)
Italy	100	0.35	Mottes et al. (1989)
Norway	180	0.34	Borresen et al. (1988)
Chuvash	34	0.28	Present study
Europeans	48	0.27	Bowcock et al. (1991)
Asia			
Sel'kup	27	0.24	Present study
Evenki	16	0.34	Present study
Ket	9	0.33	Present study
Altai	53	0.18	Present study
Chinese	75	0.35	Bowcock et al. (1991)
Nepal, Tharus	55	0.28	Pepe, Rickards et al. (1995)
Indonesia	36	0.22	Pepe, Rickards et al. (1995)
Americas			
US native Americans	19	0.24	Present study
Colombian native Americans	27	0.22	Present study
Cayapa, Ecuador	38	0.34	Pepe et al. (1994)
US whites	97	0.34	Mottes et al. (1989)
Oceania			
Australian Aboriginals	17	0.18	Present study
Solomon Islands	31	0.0	Bowcock et al. (1991)

sequence variation, but the mechanism that causes their formation is not well understood (Gu and Li 1995). Deletions occur more frequently than insertions in nuclear noncoding sequences, with the ratio of deletions to insertions for nuclear sequences varying from 1.89 to 3.50, and short insertions or deletions

(1 or 2 bp) are much more frequent than long insertions or deletions (Gu and Li 1995). These data on the evolution of insertion-deletion polymorphisms strongly suggest that the *COL1A2* polymorphism is the result of a single event in the evolution of the human species.

An alternative explanation for the observed *COL1A2* distribution is that the deletion is a product of recurrent mutation. The mechanism of insertion-deletion evolution is unclear, but it does appear that long insertion-deletion sequences are rare events. For insertions, for example, Batzer et al. (1994, 1996) argued that *Alu* insertions are good markers for reconstructing the human phylogeny and that there is no reason to think that large deletions, although somewhat more frequent, are not equally informative as tracers for human evolution. In addition, if recurrent deletion was the mechanism creating the distribution, it is difficult to explain how such events occurred in 1 or possibly more of the less densely populated continents but not in sub-Saharan Africa. Furthermore, the absence of the polymorphism in sub-Saharan Africans is unlikely to be a product of biased sampling. Although far more worldwide samples still need to be tested, Africans are not unfairly represented in Table 2 (>2000 non-African alleles and >500 African alleles scored). Therefore, if the *COL1A2* deletion is a unique event in human history, then this may be traceable through the use of 1 or more closely linked microsatellites or single nucleotide polymorphisms. However, the data suggest that the *COL1A2* deletion may have occurred up to 100,000 years ago (a minimum of 4000 generations), and therefore recombination (and recurrent mutation of a microsatellite) may make detection extremely difficult. To strengthen the argument that the deletion represents a unique event would require sequencing the PCR products of the deletion allele from a few individuals from around the world and demonstrating that they have the same deletion.

It is possible that the deletion arose in an African population that has not been sampled and that it was their ancestors who contributed the deletion to non-African groups, perhaps during the initial exodus of modern humans from Africa. East Africa is a prime candidate for such a population, and it is unfortunate that there are no samples from this region. Only a more adequate sampling of sub-Saharan populations will resolve this question.

Another explanation to be considered is that the *COL1A2* polymorphism may have evolved in the first *Homo sapiens* inhabiting sub-Saharan Africa but that through a bottleneck effect it has been eliminated from the contemporary population of the region. However, if drift is the explanation for the absence of this deletion in sub-Saharan Africans, then it would be expected that the same evolutionary mechanism would have had an effect on other markers in Africans. There is no evidence, however, that sub-Saharan African populations experienced severe bottlenecks in their evolution. In fact, most studies of mtDNA and autosomal sequence nuclear DNA indicate greater

levels of diversity in sub-Saharan Africans compared with all other human groups, a feature incompatible with major bottleneck effects.

The relatively high frequency of the deletion allele in most non-African populations (>20%) raises the possibility that selection may be acting on either the *COL1A2* locus or a neighboring gene in these groups and that this force is absent in sub-Saharan Africans. However, the deletion itself lies in an intron and is unlikely to have a phenotypic effect. No plausible candidates for selective force or neighboring genes are known, but this possibility could be investigated further by searching for variants in genes that are in linkage disequilibrium with the deletion.

Another possible explanation for the presence of the *COL1A2* deletion in all contemporary groups other than sub-Saharan Africans is that it arose in an archaic population outside Africa. This hypothesis requires interbreeding between the modern humans coming out of Africa approximately 100,000 years ago and ancient non-African hominids (probably inhabiting the Middle East or West Asia), which led to the inclusion of the *COL1A2* polymorphism in the gene pool and its spread by migration into all modern non-Africans. Descendants of this hybrid group, however, did not move back into Africa.

The literature was searched for other genes that displayed a similar distribution to that seen for *COL1A2,* that is, a polymorphism present in all world populations but absent in sub-Saharan Africans. Two such markers were found. The first was associated with the *GM* immunoglobulin system. There are extensive data for the *GM*A,X G* haplotype of the *GM* immunoglobulin system. This haplotype is a result of a simple point mutation (Schanfield and Loghem 1986) and is therefore unlikely to have evolved more than once in *Homo sapiens. GM*A,X G* is observed in all human groups except sub-Saharan Africans (Table 3), in whom it is absent or, if present, is explained as the result of admixture (Jenkins et al. 1978). Although some population samples in Table 3 represent the same individuals tested for the *COL1A2* polymorphism, most were selected from the literature to complement as closely as possible those scored for *COL1A2.* The distribution of a newly reported single nucleotide polymorphism further supports the observations for *COL1A2* and *GM*A,X G* (Pompei et al. 1998). The nucleotide 107 point mutation, C → T, in the human beta pseudogene of hemoglobin was found to be present at polymorphic frequencies in 9 non-African populations (n = 238, average frequency of 0.11) but absent in 2 African groups (n = 75). The 9 non-African groups were Europeans (Italians, Sardinians, Dutch, Czechs, and Lithuanians), Hindus from India, and 3 Asian populations (Chinese, Indo-Chinese, and Tharus of Nepal).

Given that the out of Africa model is currently the most widely accepted explanation of the origin of modern humans, the *COL1A2* deletion and *GM*A,X G* distributions (and, if confirmed, the nt107 mutation in the hemoglobin beta pseudogene) have to be viewed in the context of that model. It is most probable that these evolutionary events occurred only once in hu-

Table 3. *GM*A,X G* Haplotype Frequencies in World Populations

Population	*Number Tested*	*GM*A,X G Frequency*	*Reference*
Africa			
North Africa	1848	0.028	Benabadji et al. (1965), Ducos et al. (1965), Lefranc et al. (1979), Ropartz et al. (1963), Ruffie et al. (1962)
Biaka Pygmies, Central African Republic	1179	0.001	van Loghem (1986)
San	1017	0.000	Jenkins et al. (1978)
Nigerians	?	0.000	van Loghem et al. (1978)
South Africa, Colored	146	0.000	Jenkins et al. (1978)
South Africa, whites	4483	0.095	Schanfield (1990)
Europe			
British	146	0.093	Stevenson and Schanfield (1981)
Italians	237	0.042	Stevenson and Schanfield (1981)
Russians	48	0.042	Stevenson and Schanfield (1981)
Asia			
Altai	95	0.109	Unpublished data[a]
Evenki	197	0.129	Unpublished data[a]
Ket	33	0.114	Unpublished data[a]
Sel'kup	640	0.131	Sukernik et al. (1992)
Chinese	700	0.090	Schanfield et al. (1972)
Americas			
US native Americans	>2000	0.101	Pooled unpublished data[a]
South American native Americans	165	0.431	Barrantes et al. (1982)
Mixed US groups	4483	0.089	Schanfield (1990)
Oceania			
Australian Aboriginals, mainland	39	0.091	Curtain et al. (1972)
Australian aboriginals, Mornington Island	113	0.068	Curtain et al. (1972)
Solomon Islands	1434	0.036	Steinberg et al. (1972)

a. Data collected by M.S. Schanfield, of Analytical Genetic Testing Center, Denver, Colorado, over several years.

man evolution and, further, only after some members of *Homo sapiens* left sub-Saharan Africa, probably between 60,000 and 120,000 years ago. These 2 mutations may have evolved in the first migrants to leave Africa because it is clear that descendants of the group in whom the mutations occurred have contributed to all present non-African populations. This finding does not mean that there was only 1 migration out of Africa. However, the data suggest that there was no substantial migration back to Africa. This finding is of significance because of the suggestion of Hammer et al. (1998) that there was a recent back-migration to Africa of Asian Y chromosomes. If there was a relatively recent migration of Asian groups back to Africa, it is difficult to

see how these migrants did not also carry the *COL1A2* deletion (as well as haplotype *GM*A,X G* and the hemoglobin beta pseudogene mutation) along with their Y chromosomes.

These genetic data raise questions about how the migration from sub-Saharan Africa occurred. They argue against a number of small bands separating into pockets of widely dispersed, noninteracting groups in the new lands of the Middle East or western Asia, some of whose descendants subsequently migrated to even more distant lands. Instead, the widely shared heritage of these polymorphisms suggests a single group keeping close together in the new region for a period of time. In other words, the present distributions of these genetic polymorphisms suggest that there was a distinct pause after the initial exodus from Africa and before the later radiation to all regions of the planet. How long this period was is speculative, as are estimates of the size of the migrant group. If the first migrants were restricted in their new environment for some considerable time, what factors might have been responsible? One such pressure could have been the need for the African savanna adapted gene pool to evolve in response to new infectious agents, such as parasites, bacteria, viruses, and/or novel environmental xenobiotics found in both animal and plant foods. In this context the *COL1A2* and *GM*A,X G* variants are indicators of the genetic differentiation that must have occurred during this time and that became the property of all non-Africans.

Acknowledgments DNA aliquots of nonhuman primates were kindly made available by W. Schemp, Institut für Humangenetik und Anthropologie der Universitat, Freiburg, Germany, by M. Jackson, Department of Human Genetics, University of Newcastle upon Tyne, Newcastle upon Tyne, England, and by the Institute of Zoology, London, England.

Received 28 September 1998; revision received 18 March 1999.

Literature Cited

Baker, R., J. Lynch, L. Ferguson et al. 1991. PCR detection of five restriction site dimorphisms at the type 1 collagen loci *COL1A1* and *COL1A2*. *Nucleic Acids Res.* 19:4315.

Barrantes, R., P.E. Smouse, J.V. Neel et al. 1982. Migration and infrastructure of the Central American Guayami and their affinities with other tribal groups. *Am. J. Phys. Anthropol.* 58:201–214.

Batzer, M., S. Arcot, J. Phinney et al. 1996. Genetic variation of recent *Alu* insertions in human populations. *J. Molec. Evol.* 42:22–29.

Batzer, M., M. Stoneking, M. Alegria-Hartman et al. 1994. African origin of human specific polymorphic *Alu* insertions. *Proc. Natl. Acad. Sci. USA* 91:12,228–12,292.

Benabadji, M., J. Ruffie, G. Larrouy et al. 1965. Etude hemotypologique des populations du massif du Hoggar et du Plateau de l'Air. II. Les groupes serique. *Bull. Soc. Anthropol.* 72:181–184.

Borresen, A.L., P. Moller, and K. Berg. 1988. Linkage disequilibrium analyses and restriction mapping of four RFLPs at the pro α2(1) collagen locus: Lack of correlation between linkage disequilibrium and physical distance. *Hum. Genet.* 78:216–221.

Bowcock, A.M., C. Bucci, J.M. Hebert et al. 1987. Study of 47 DNA markers in five populations from four continents. *Gene Geogr.* 1:47–64.

Bowcock, A., J. Herbert, J. Mountain et al. 1991. Study of an additional 58 DNA markers in five human populations from four continents. *Gene Geogr.* 5:151–173.

Curtain, C.C., E. van Loghem, H.H. Fudenberg et al. 1972. Distribution of the immunoglobulin markers at the IgG1, IgG2, IgG3, IgGA2, and γ-chain loci in Australian Aborigines: Comparison with New Guinea populations. *Am. J. Hum. Genet.* 24:145–155.

Dalgleish, R., G. Williams, and J.R. Hawkins. 1986. Length polymorphism in the pro α (I) collagen gene: An alternative explanation in a case of Marfan syndrome. *Hum. Genet.* 73:91–92.

Ducos, J., P. Fernet, and H. Vignes. 1965. Etude hemotypologique des populations du Tidikelt: Les groupes seriques. *Bull. Mem. Soc. Anthropol. Paris* 7:175–177.

Grobler-Rabie, A.F., G. Wallis, D.K. Brebner et al. 1985. Detection of a high frequency *Rsa*I polymorphism in the human pro α2(I) collagen gene which is linked to an autosomal dominant form of osteogenesis imperfecta. *EMBO J.* 4:1745–1748.

Gu, X., and W.-H. Li. 1995. The size distribution of insertions and deletions in human and rodent pseudogenes suggests the logarithmic gap penalty for sequence alignment. *J. Molec. Evol.* 40:464–473.

Hammer, M.F., T. Karafet, A. Rasanayagam et al. 1998. Out of Africa and back again: Nested cladistic analysis of human Y-chromosome variation. *Molec. Biol. Evol.* 15:427–441.

Hammer, M.F., A.B. Spurdle, T. Karafet et al. 1997. The geographic distribution of human Y-chromosome variation. *Genetics* 145:787–805.

Jenkins, T., H. Harpending, and G.T. Nurse. 1978. Genetic distances among certain Southern African populations. In *Evolutionary Models and Studies in Human Diversity,* R.J. Meier, C.M. Otten, and F. Abdel-Hameed, eds. The Hague, The Netherlands: Mouton, 227–243.

Jorde, L.B., A.R. Rogers, M. Bamshad et al. 1997. Microsatellite diversity and the demographic history of modern humans. *Proc. Natl. Acad. Sci. USA* 94:3100–3103.

Lefranc, G., G. de Lange, L. Rivat et al. 1979. *Gm, Am,* and *Km* immunoglobulin allotypes of two populations in Tunisia. *Hum. Genet.* 50:199–211.

Mottes, M., L. Cugola, and P.F. Pignatti. 1989. Haplotype frequencies of the collagen type-I genes in the Italian population. *Hum. Genet.* 83:369–372.

Pepe, G., M. Muglia, and C. Brancati. 1990. Studies of four restriction fragment length polymorphisms of the type 1 collagen genes in two Italian populations. *Hum. Hered.* 40:368–380.

Pepe, G., C. Bue, P. Orlandi et al. 1995. Are the SSTR alleles stable enough to be considered monophyletic and hence reliable anthropogenetic markers? Linkage disequilibrium study on the $(ACT)_n$ *COL1A2* SSTR. *Hum. Biol.* 67:703–715.

Pepe, G., O. Rickards, C. Bue et al. 1994. *Eco*RI, *Rsa*I, and *Msp*I RFLPs of the *COL1A2* gene (type 1 collagen) in the Cayapa, a native American population of Ecuador. *Hum. Biol.* 66:979–989.

Pepe, G., O. Rickards, C. Jodice et al. 1995. Allele and haplotype frequency distribution of the *Eco*RI, *Rsa*I, and *Msp*I *COL1A2* RFLPs among various human populations. *Hum. Biol.* 67:905–920.

Pompei, F., B.M. Ciminelli, and G. Modiano. 1998. Two ethnic-specific polymorphisms in the human beta pseudogene of hemoglobin. *Hum. Biol.* 70:659–666.

Ropartz, C., P.Y. Rousseau, and L. Rivat. 1963. Interet des groupes de g-globulin *Gm* et *Inv* dans l'appreciation de metissage des populations: Etude de ces groupes seriques dans l'ouest African et l'extreme-orient. *Rev. Fr. Etud. Clin. Biol.* 8:465–472.

Ruffie, J., R. Cabannes, and G. Larrouy. 1962. Etude hemotypologique des populations Berbres de M. Sirda-Fouaga (Nord Ouest Orenais). *Bull. Mem. Soc. Anthropol. Paris* 3:294–314.

Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Schanfield, M.S. 1990. Forensic applications of immunoglobulin allotypes (*GM, AM,* and *KM*). In *DNA and Other Polymorphisms in Forensic Science,* H.C. Lee and R.E. Gaensslen, eds. Chicago, IL: Yearbook Medical Publishers, 160–213.

Schanfield, M.S., and E. van Loghem. 1986. Human immunoglobulin allotypes. In *Handbook of Experimental Immunology,* v. 3, *Genetics and Immunology,* D.M. Weir, L.A. Herzenberg, C. Blackwell et al., eds. Oxford, England: Blackwell, ch. 94, 1–18.

Schanfield, M.S., H. Gershowitz, K. Okhura et al. 1972. Studies on the immunoglobulin allotypes of Asiatic populations. II. *Gm* and *Inv* allotypes in Chinese. *Hum. Hered.* 22:138–143.

Steinberg, A.G., A. Damon, and J. Bloom. 1972. Gammaglobulin allotypes of Melanesians from Malaita and Bougainville, Solomon Islands. *Am. J. Phys. Anthropol.* 36:77–84.

Stevenson, J.C., and M.S. Schanfield. 1981. Immunoglobulin allotypes in European populations. III. *Gm, Am,* and *Km* allotypes in people of European ancestry in the United States. *Hum. Biol.* 53:521–542.

Stoneking, M., and H. Soodyall. 1996. Human evolution and the mitochondrial genome. *Curr. Op. Genet. Dev.* 6:731–736.

Sukernik, R.I., L.P. Osipova, and M.S. Schanfield. 1992. Distribution of *GM* allotypes and deleted *IGHG1* haplotype in the Sel'kups in comparison to the Forest Nentsi of northwestern Siberia. *Exp. Clin. Immunogenet.* 9:13–23.

Sykes, B., D. Ogilvie, P. Wordsworth et al. 1986. Osteogenesis imperfecta is linked to both type I collagen structural genes. *Lancet* 2:69–71.

Tishkoff, S., A.E. Dietzsch, W. Speed et al. 1996. Global patterns of linkage disequilibrium at the *CD4* locus and modern human origins. *Science* 271:1380–1387.

van Loghem, E. 1986. Immunoglobulin allotypes. In *African Pygmies,* L.L. Cavalli-Sforza, ed. Orlando, FL: Academic Press; 201–218.

van Loghem, E., L. Salimonu, A.I.O. Williams et al. 1978. Immunoglobulin allotypes in African populations. I. *GM-AM* haplotypes in a Nigerian population. *J. Immunogenet.* 5:143–147.

Vigilant, L., M. Stoneking, H. Harpending et al. 1991. African populations and the evolution of human mitochondrial DNA. *Science* 253:1503–1507.

Watson, C.J., and R. Dalgleish. 1990. PCR detection of a 38-bp length variant in the *COL1A2* gene. *Nucleic Acids Res.* 18:5925.