J Hum Genet (2002) 47:140–145

ORIGINAL ARTICLE

Xin Ma · John Wright · Shujun Dou · Paul Olsen
Larry Teeter · Gerald Adams · Edward Graviss

Ethnic divergence and linkage disequilibrium of novel SNPs in the human *NLI-IF* gene: evidence of human origin and lack of association with tuberculosis susceptibility

Received: December 13, 2001 / Accepted: January 8, 2002

Abstract Sequence variation in the human genome has been used as a tool in studying human diseases and the evolutionary history of man. A human inherited predisposition to tuberculosis has been suggested and studied; however, genetic mechanisms are still ambiguous. In the present study, we scanned the regulatory and coding region of Nuclear LIM Interactor-Interacting Factor gene (*NLI-IF*), which is physically close to the tuberculosis-associated gene *NRAMP1*. Thirteen biallelic single-nucleotide polymorphisms (SNPs) were identified from four ethnic populations (African–American, Caucasian, Hispanic, and Asian) with population-specific distribution of alleles. The extent of linkage disequilibrium (LD) between 402T>C, and 472−42G>A varied distinctly from complete LD in the non-African–American groups to strong but incomplete LD in African–Americans. Both SNPs were in significant LD with the polymorphism 3′ UTR in *NRAMP1* among these ethnic groups ($P < 0.02$), except 402T>C in African–Americans. In a case-control study with a Caucasian population, three cosmopolitan SNPs (204C>A, 402T>C and 472−42G>A) in *NLI-IF* showed no significant association with human susceptibility to tuberculosis. Our results support the "out-of-Africa" model of human origin, and suggest the time for the common ancestor dispersing from Africa could not have been more than approximately 385,620 years ago.

X. Ma · J. Wright · S. Dou · P. Olsen · L. Teeter · G. Adams · E. Graviss
Department of Pathology, Baylor College of Medicine, Houston, TX, USA

E. Graviss (✉)
Department of Pathology, Baylor College of Medicine, One Baylor Plaza (209E), Houston, TX 77030, USA
Tel. +1-713-798-8097; Fax +1-713-798-8895
e-mail: egraviss@bcm.tmc.edu

The SNPs reported in this study have been deposited in the SNP database of the National Center for Biotechnology Information (NCBI) with NCBI assay ID numbers (ss#) 869708, 3171836, 3171837, 3171838, 3171839, 3171840, 3171841, 3171842, 3171843, 3171844, 3171845, 3171846, 3171847.

Key words *NLI-IF* · *NRAMP1* · Linkage disequilibrium · Single-nucleotide polymorphism · Human origin · Tuberculosis

Introduction

Variations in the human genome, especially in coding regions and regulatory elements of genes, have important implications for all biological traits including disease predisposition. In the last few years, growing evidence has strongly supported the hypothesis that the common variants (polymorphisms) of genes in human populations are associated with susceptibility or resistance to the common diseases. For example, a 32-bp-deletion polymorphism in the coding region of the *CCR-5* gene with an allele frequency of up to 0.14 among Caucasian populations leads to strong resistance against human immunodeficiency virus (HIV)-1 infection and acquired immunodeficiency syndrome (Dean et al. 1996). Polymorphisms in the *HLA-DR13* gene and *VDR* gene have been shown to influence the infection of hepatitis B virus in several racial populations (Bellamy et al. 1999; Ahn et al. 2000). Tuberculosis (TB), acting as an important selective force in shaping the human genome during human evolution, has been extensively investigated. Initially, based on the implication of animal models, human natural resistance-associated macrophage protein 1 gene (*NRAMP1*) was isolated and characterized from chromosome region 2q35 (Cellier et al. 1994; Liu et al. 1995). Later, population-based association studies suggested that variants in *NRAMP1* modified TB susceptibility among several ethnic populations, e.g., West African and East Asian (Bellamy et al. 1998; Gao et al. 2000; Ryu et al. 2000). However, the effects of these polymorphisms are mostly weak (odds ratios < 3.0), and genetic analysis of those polymorphisms cannot fully explain *NRAMP1* association with human clinical TB from a functional viewpoint. It has been suggested that some unknown functional polymorphisms could lie nearby and link with them. A family-based linkage study in a Brazilian population, involving 205

nuclear families, suggested a weak linkage for the loci (IL8RB and D2S1471) physically close (<150 kb) to *NRAMP1*, but not for *NRAMP1* itself (Shaw et al. 1997). Another family-based genome-wide linkage study in Gambia and South Africa suggested that no region in the human genome showed evidence of a strong linkage (e.g., a logarithm of differences [LOD] score > 3.0) to TB, and the nearest marker to *NRAMP1* had a weakly positive LOD score of 0.47 (Bellamy et al. 2000). However, recent findings from a linkage study of a large aboriginal Canadian family strongly suggested the effect of a major gene by showing a maximum LOD score of 3.81 for the locus just distal to *NRAMP1* in chromosome region 2q35. Moreover, none of the LOD scores for the *NRAMP1* haplotype fell to < 3.0. As a result, the authors from this linkage study estimated the location of the susceptibility locus to be either at the *NRAMP1* haplotype or between the marker D2S424 and the *NRAMP1* haplotype (Greenwood et al. 2000). In brief, present data from different ethnic populations make the question of genetic predisposition to human TB highly controversial. Nevertheless, each of the previously described cases consistently implied that a potential gene or gene cluster close to *NRAMP1* or *NRAMP1* itself contributed, at least weakly, to TB susceptibility in humans. Therefore, exploration of the *NRAMP1* vicinity has been an absolute necessity.

Most recently, a novel gene, designated the nuclear LIM interactor-interacting factor gene (*NLI-IF*), was identified in the immediate vicinity (~4 kb) of *NRAMP1*, when a 32-kb of genomic DNA overlapping *NRAMP1* was sequenced and annotated using computerized sequence analysis (Marquet et al. 2000). The human *NLI-IF* gene begins approximately 4 kb downstream of the *NRAMP1* stop codon, and consists of seven exons with a 783-bp open reading frame. The predicted NLI-IF protein is highly homologous with the NLI-interacting factor from *Gallus gallus* and a variety of species from yeast to plants, which are probably involved in transcriptional activity of the cell. *NLI-IF* mRNA was consistently detected in human tissues at different expression levels by Northern blot (Marquet et al. 2000). Although present evidence suggests involvement of *NLI-IF* in basic cellular function, the close physical proximity between *NLI-IF* and *NRAMP1* justifies investigating the association between the *NLI-IF* variants and human TB. In this study, we scanned the coding region, promoter, and 5′ and 3′ flanking regions of the human *NLI-IF* gene for sequence variation from four ethnic population samples of African–American, Caucasian, Hispanic, and Asian origin, respectively. Furthermore, to test the role of novel polymorphisms of *NLI-IF* in human TB susceptibility, a case-control study was conducted with a group of TB patients (n = 94) and non-TB controls (n = 145) from a Caucasian population in Houston, Texas.

Over the last decade, genomic variation has become an indispensable tool in studying human natural history. The variability and heritability of the human genome enable it to record the story of human evolution regarding human origin, population migration, and natural selection. For instance, linkage disequilibrium (LD) at the *CD4* locus has been analyzed from 42 geographically dispersed populations, and results suggest a common and recent African origin for the non-African human population (Tishkoff et al. 1996). In the present study, we discuss this issue with data generated from four major ethnic populations, which provides additional evidence for a modern human origin.

Methods

Identification of *NLI-IF* polymorphisms

To identify the polymorphisms in *NLI-IF* and establish their frequencies in the populations, we recruited a total of 200 individuals from four ethnic groups (African–American, Caucasian, Hispanic, and Asian; 50 individuals from each group) from hospitals and clinics in Houston, Texas. Race or ethnicity was determined for each individual by self-identification. This study was approved by the Baylor College of Medicine and Affiliates Institution Review Board.

DNA was isolated from blood samples using the Nucleon DNA extraction and purification kit (Amersham International and Scotlab, Buckinghamshire, England). Six pairs of primers were designed to amplify the 5′ promoter (up to 1077 bp ahead of start codon) and untranslated regions (UTRs), the coding regions of seven exons with their flanking noncoding sequences at least 50 bp up or downstream of the exon/intron or UTR boundaries (Table 1). Polymerase chain reactions (PCRs) were performed using HotStarTaq DNA polymerase kit (QIAGEN, Hilden, Germany). The PCR products were purified by Microcon Centrifugal Filter Devices (Millipore Corporation Bedford, MA, USA). Each amplified fragment was sequenced on an ABI 377 DNA sequencer using the BigDye DNA Sequencing Kit (Applied Biosystems, Foster City, CA, USA). The rare sequence variants were confirmed by amplifying and sequencing the fragments twice from the original DNA. Restriction fragment length polymorphism analysis was used to genotype the individuals in the case-control study by using the restriction enzymes (see Table 2) under the conditions recommended by the provider (New England Biolabs, Beverly, MA, USA). The sequence variants in the *NRAMP1* gene were detected by the methods described previously in other studies (Liu et al. 1995; Bellamy et al. 1998). Haplotype frequencies and linkage disequilibrium among 400 alleles from four ethnic groups were calculated by using Arlequin software (Genetics and Biometry Laboratory, Geneva, Switzerland).

A case-control study

Ninety-four Caucasian patients with clinical TB (mean age ± SD, 51.9 ± 13.9 years; 73.4% male) were enrolled from an ongoing population-based tuberculosis surveillance study, the Houston Tuberculosis Initiative, including 81 patients with pulmonary TB, 11 with extrapulmonary TB, and two with both pulmonary and extrapulmonary TB. All pa-

Table 1. Primers for PCR amplification and sequencing of *NLI-IF* gene fragments

Fragment	Primer name[a]	Sequence	Region	Size (bp)
1	NLP1F	5′-ATCGGCAGCGAGTCCAGTGAC-3′	Promoter	540
	NLP1R	5′-GCTCCACCTCTGCGTTTTCTG-3′		
2	NL1F	5′-ACGCAGCCGCAGAGACTCAG-3′	Exon 1	568
	NL1R	5′-AAGATCCCGGCCCCTTCTCTC-3′		
3	NL2F	5′-TGGGCACACATGCAGATTCT-3′	Exon 2	280
	NL2R	5′-TTCCCTAAAGCCCTGAAGAC-3′		
4	NL3F	5′-AGCCTGCGTGGGAGGAAGGAC-3′	Exon 3	478
	NL3R	5′-GGGCCTCTGACATTCCACATG-3′		
5	NL5F	5′-GGCCCAGTCTCCTTCCTGTGC-3′	Exon 5 and 6	531
	NL5R	5′-GCTCCTGTCCCAGTCCCAGTC-3′		
6	NL7F	5′-AGGGCTGGCCTGGAAATTCTG-3′	Exon 7	254
	NL7R	5′-TGGGAGGAGGTGTGGGTATCA-3′		

PCR, Polymerase chain reaction
[a]The suffixes F and R of the primer names indicate forward and reverse primers, respectively. The primer sequences are based on the Genbank sequences with accession numbers AF229162 and AF229163

Table 2. Novel single-nucleotide polymorphisms (SNPs) in the *NLI-IF* gene

				Allele frequency[b]			
SNP[a]	Location	Consequence	Restriction enzymes	African	Caucasian	Hispanic	Asian
68−27C>T	Intron 1	Noncoding	*Mnl*I	0	0	0.01	0
166G>A	Exon 2	Ala → Thr	*Cac*8I	0.01	0	0	0
174C>T	Exon 2	Silent	*Msp*A1I	0.01	0	0	0
204C>A	Exon 2	Silent	*Hin*P1I	0.10	0.29	0.29	0.22
216+6T>G	Intron 2	Splicing	—	0	0	0	0.01
216+25C>T	Intron 2	Noncoding	*Hsp*92II	0.01	0	0	0
217−24C>A	Intron 2	Noncoding	—	0.06	0	0.01	0
379−68C>A	Intron 4	Noncoding	*Hsp*92II	0.08	0	0.01	0
402T>C	Exon 5	Silent	*Pml*I	0.87	0.47	0.43	0.30
472−85C>T	Intron 5	Noncoding	*Hap*II	0.01	0	0	0
472−54G>A	Intron 5	Noncoding	*Pae*I	0	0.02	0	0
472−42G>A	Intron 5	Noncoding	*Bmr*I	0.60	0.47	0.43	0.30
474C>T	Exon 6	Silent	*Afa*I	0.01	0	0	0

[a]The designation SNP follows the recommendations of mutation nomenclature in Dunnen and Antonarakis 2000
[b]Allele frequency refers to frequency of the minor allele of the biallelic SNP among the populations studied except 402T>C and 472−42G>A among African–Americans

tients were identified by either *Mycobacterium tuberculosis* culture positivity ($n = 91$) or clinical improvement in response to antimycobacterial treatment ($n = 3$). A total of 145 Caucasian individuals without a history of TB were recruited from hospitals and clinics as controls for this study (mean age ± SD, 55.4 ± 14.8 years; 45.8% male). All patients and controls were HIV seronegative. The overall difference of genotype frequencies between the case group and the control group was tested by using a chi-square test for a two-by-three contingency table with two degrees of freedom. Genotypes were compared by using a chi-square test for two-by-two contingency tables with SAS software (version 8.0, SAS Institute; Cary, NC, USA).

Results

Novel polymorphisms in the *NLI-IF* region

A total of 200 individuals (400 chromosomes) from four ethnic populations (Caucasian, African–American, Hispanic, and Asian; $n = 50$ for each ethnic group) were enrolled for screening the sequence variants in the *NLI-IF* gene region including 783 bp of the open reading frame within seven exons and 1868 bp of their adjacent noncoding regions (promoter, UTRs, and introns). The sample size (50 individuals per ethnic group) provided us with a greater than 99% power to detect variant alleles present at a frequency of 0.05, and 70% power to detect minor alleles present at a frequency of 0.01 for each ethnic group. Using PCR direct sequencing, we detected 13 biallelic single-nucleotide polymorphisms (SNPs) with five variants in exons (38%) and eight in introns (62%) (Table 2). Only one SNP at codon 56 (exon 2) detected from one chromosome of an African–American individual resulted in an amino acid change (alanine to threonine). The remaining four coding SNPs (cSNPs) scattering in exons 2, 5, and 6, respectively, were synonymous variants. Eight intronic SNPs were found within introns 1, 2, 4, and 5, within a range of 6 to 85 bp away from exon/intron boundaries. In general, SNPs were found in this region at a frequency of 1 per 204 bp (2651 bp/13 SNPs); one SNP was found per 156 bp for the coding region and 1 SNP per 233 bp for the noncoding

region. No sequence variant was found in the 5′ promotor region and in the 3′ immediate vicinity (within 58 bp of the stop codon). Of the 13 SNPs, nine (69%) were transitions (6 C↔T, 3 G↔A), and four (31%) were transversions (3 C↔A, 1 T↔G). Among these biallelic SNPs, the G or the C allele was observed as the more common allele much more frequently than the T or the A allele (11 vs. 2).

Population-specific pattern of allele distribution and linkage disequilibrium of *NLI-IF* SNPs

Allele distribution of SNPs in the *NLI-IF* region was observed as a remarkable population-favored pattern; i.e., the occurrence of these variants was different among the four ethnic populations studied. African–Americans appeared distinct from other racial groups. For example, of the 13 SNPs, only 3 (23%) (204C>A, 402T>C, and 472−42G>A) were found in all ethnic groups (also called cosmopolitan SNPs), presenting allele frequencies of more than 0.10 for the minor alleles in each group. The A allele of 204C>A was seen at a lower allele frequency in African–Americans than in non-African–American groups with statistical significance ($P < 0.05$) (Table 2). A striking difference in allele frequency between African–Americans and other ethnic groups was detected at 402T>C and 472−42G>A, where the C allele of 402T>C and the A allele of 472−42G>A were common alleles among African–Americans (allele frequencies 0.87 and 0.60, respectively). In contrast, both alleles were the minor alleles (allele frequencies 0.30 to 0.47) among non-African–American groups. Nevertheless, all population genotype frequencies conformed to Hardy–Weinberg equilibrium. Ten rare SNPs with allele frequency of less than 0.10 for the minor allele in these populations exhibited a population-specific distribution pattern. For instance, the minor alleles of 166G>A, 174C>T, 216+25C>T, 472−85C>T, and 474C>T were observed only once as a single heterozygote exclusively among African–Americans; likewise, 472−54G>A was seen only among Caucasians, 216+6T>G only among Asians, and 68−27C>T only among Hispanics. Therefore, African–Americans had the greatest number of population-specific rare alleles. Two rare variants, 217−24C>T and 379−68C>A, were shared only by Hispanics and African–Americans at different allele frequencies (African–American 0.06 and 0.08, respectively, vs. Hispanic 0.01 and 0.01), which implied the complexity of the ethnic ancestry of Hispanics.

LD was observed among all ethnic groups between 402T>C and 472−42G>A, which are 131 bp apart from each other. Interestingly, the degree of LD presented distinct differences between African–Americans and non-African–Americans. The T allele at 402T>C and the G allele at 472−42G>A were in complete LD among non-African–Americans (the standardized LD coefficient $D' = 1.0000$, $P < 10^{-6}$, which were calculated by using Arlequin software available at http://lgb.unige.ch/arlequin/), whereas they were in strong but not complete LD in the African–American group ($D' = 0.7436$, $P < 10^{-5}$). No LD was observed between either 402T>C or 472−42G>A and 204C>A ($P > 0.05$).

We tested the associations between the cosmopolitan SNPs 204C>A, 402T>C, and 472−42G>A in *NLI-IF* and two sequence variants, 5′$(GT)_n$ and 3′UTR, in *NRAMP1*, which were previously found to be associated with TB (Bellamy et al. 1998; Gao et al. 2000; Ryu et al. 2000). The allele frequencies of the 3′UTR varied remarkably among the four ethnic groups: 0.25 in African–Americans, 0.01 in Caucasians, 0.11 in Hispanics, and 0.12 in Asians. This result, which is consistent with previous reports, may possibly explain why Caucasians are less susceptible to TB than are other ethnic populations. Significant LD was observed between the 3′UTR and 472−42G>A, as well as 402T>C ($P < 0.02$) among these ethnic groups, except for 3′UTR and 402T>C in African–Americans ($P = 0.515$). Thus, the 3′UTR could have originally occurred on the background of a 402C and 472−42A haplotype. No significant LD was shown between 5′$(GT)_n$ and these three SNPs in *NLI-IF*.

Lack of association between three common SNPs of *NLI-IF* and human susceptibility to TB

To test the role of three relatively common SNPs (204C>A, 402T>C, and 472−42G>A of *NLI-IF* in human TB susceptibility), we conducted a case-control study with a Caucasian population sample. There was no statistically significant difference in either allele frequencies or genotype percentages between the case and control groups ($P > 0.05$) (Table 3).

Discussion

It has been suggested by several independent studies that the vicinity of *NRAMP1* probably harbors a gene related to human susceptibility to TB (Shaw et al. 1997; Bellamy et al. 2000; Greenwood et al. 2000). The novel gene, *NLI-IF*, is located only 4-kb downstream of *NRAMP1*, and could therefore be a candidate for a TB-associated gene or possibly functionally linked with *NRAMP1*. In this study, we scanned the coding region and the flanking regions of *NLI-IF* for polymorphisms among several major ethnic groups including African–American, Caucasian, Asian, and Hispanic. Our results, based on the genotyping of 200 individuals, showed that the coding region of the *NLI-IF* gene is evolutionarily conserved, because only a single sequence variant resulting in an amino acid substitution was detected from one chromosome. Two cSNPs (204C>A and 402T>C), and one intronic SNP (472−42G>A) were relatively common with allele frequencies of more than 0.10 for the minor alleles among the populations studied, but all were synonymous or silent variants. The lack of nonsynonymous polymorphisms may be due to the intensive selection pressure during human evolution. Combined with the evidence from the *NLI-IF* expression profile that

Table 3. Analysis of association between the variants in the *NLI-IF* gene and susceptibility to tuberculosis among Houston Caucasian patients

	Controls (n = 145)	TB cases (n = 94)	Odds ratio	Chi-square	Degree of freedom	P value
Genotypes	n (%)	n (%)				
204 C>A						
CC	72 (50)	44 (47)	1.0			
AA	11 (8)	6 (6)	0.89 (0.27–2.86)	0.42	2	0.81
CA	62 (42)	44 (47)	1.16 (0.65–2.06)			
402 T>C[a]						
TT	44 (30)	30 (32)	1.0			
CC	22 (15)	16 (17)	1.07 (0.45–2.54)	0.29	2	0.86
TC	79 (55)	48 (51)	0.89 (0.48–1.67)			

TB, Tuberculosis

[a]The result of intronic SNP 472−42G>A is the same as 402T>C and not shown in this table, because it is in complete disequilibrium with 402T>C. The overall difference of genotype frequencies between the case and control group was tested by two-by-three chi-square analysis with 2 degrees of freedom. Genotypes were compared by two-by-two chi-square test

mRNA of *NLI-IF* was detected ubiquitously in human tissues (Marquet et al. 2000), the product of *NLI-IF* may be of importance for the fundamental function of cells. Nevertheless, the noncoding SNPs in the regulatory region could result in a subtle phenotypic change by influencing the gene expression quantitatively. For example, the microsatellite polymorphism $(GT)_n$ in the promoter of the *NRAMP1* gene associated with TB risk in several populations has been observed to drive different expression levels of the luciferase gene in vitro (Bellamy et al. 1998; Searle and Blackwell 1999). In other ways, these polymorphisms may be linked with other functional sequence variants. In addition, studies on codon usage have demonstrated that the prevalent codons can result in a substantial increase in expression efficiency (Kim et al. 1997). The cSNPs in *NLI-IF*, though silent, result in a switch-over between prevalent and less used codons; e.g., in codon 68 (204C>A), the change is from GGC (50% usage) to GGA (14% usage), whereas in codon 134 (402T>C), a change from CAT (21%) to CAC (79%) was seen. Their effects on *NLI-IF* expression efficiency and association with some unknown phenotypes remain to be investigated. To test the association between three common SNPs (204C>A, 402T>C, and 472−42G>A) in *NLI-IF* and TB susceptibility, we conducted a case-control study in a Caucasian population. Our results did not show a statistically significant difference between the cases and controls ($P > 0.05$) (Table 3). However, we cannot rule out the potential importance of these polymorphisms when a larger sample size and other ethnic populations are studied, or in regard to other trait(s).

Analyses of human genomic variants, such as SNP, microsatellite, and Alu family, have been used as a critical tool in understanding the origin of mankind, the evolutionary reconstitution of a specific population, and the effect of natural selection. Accumulating evidence from human mitochondrial DNA (mtDNA) (Cann et al. 1987), Y chromosome DNA (Underhill et al. 2000), and nuclear autosome DNA (Tishkoff et al. 1996) studies has strongly argued for the recent African origin model or the "out-of-Africa" model, and against the multiregional evolution model (Relethford 2001). The earlier evidence from mtDNA suggested that all human populations living today could be traced back to one (or a few) closely related humans who emigrated from African 200,000 years ago (Cann et al. 1987). The updated estimates of the most recent common ancestor out of Africa varied considerably among the studies of different genomic loci, including 50,000 years for Y chromosome DNA (Thomson et al. 2000), 143,000 years for mtDNA (Horai et al. 1995), 535,000 years for a noncoding region at Xq13.3 (Kaessmann et al. 1999), 800,000 years for β-globin (Harding et al. 1997), and 1,860,000 years for *PDHA1* (Harris and Hey 1999). In this study, two intriguing phenomena were observed. First, African–Americans showed the greatest sequence variability of the *NLI-IF* region (10/13 SNPs) compared with other racial groups (4–6/13 SNPs) (Table 2). This finding is consistent with the recent data from 313 human genes (Stephens et al. 2001), which showed that 662 rare variants were detected in African–Americans as compared with 294, 223, and 273 in Asian, Caucasian, and Hispanic samples, respectively. Theoretically, the oldest populations accumulated the largest number of mutations; hence, at this point, our results suggest that the African is the oldest ethnic population. Of course, population size and mutation rate can also contribute to genetic diversity.

Three common SNPs, 204C>A, 402T>C, and 472−42G>A, shared by all studied ethnic groups, indicated their ancientry. According to the neutral theory of Kimura (Vogel and Motulasky 1997), the great majority of evolutionary changes at the molecular level are caused by selectively neutral or nearly selectively neutral mutants or random genetic drift, rather than by natural selection. Therefore, these three SNPs could be neutral variants with little or no selective advantage or disadvantage. With respect to TB susceptibility, our results from the case-control study support this inference. However, although of little interest for clinical genetics, neutral variants are crucial for the study of human evolution. For instance, the second finding in this study, that the extent of LD between 402T>C and 472−42G>A varied distinctly from complete LD in non-African–Americans to strong but incomplete LD in African–Americans, strongly implies that 1. these ancient

alleles originated from Africa; 2. all non-African populations descended from a single ancestral gene pool because they share the same haplotypes and similar haplotype frequencies; 3. the origin of the non-African populations is comparatively recent because of insufficient time to diminish the complete LD by recombination events; and 4. the population size that emigrated "out of Africa" could be small because the genetic diversity of non-African populations is consistently lower than that of Africans, probably because of a bottleneck effect during emigration. By the principle of Jennings (Vogel and Motulasky 1997), we estimate the required time period for the disappearance of the complete LD in a random mating population to be approximately 19,281 generations, i.e., 385,620 years (20 years per generation). Therefore, based on these findings, our results support the recent African origin model, and suggest that the time for the common ancestor of the non-Africa population leaving Africa should not be more than, and is possibly much less than, 385,620 years ago.

Acknowledgments This project has been funded in whole or in part by Federal Funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, under contract No. N01-AO-02738 and AI 41168. We thank Erwin Schurr for reviewing and making helpful comments on a draft of this manuscript.

References

Ahn SH, Han KH, Park JY, Lee CK, Kang SW, Chon CY, Kim YS, Park K, Kim DK, Moon YM (2000) Association between hepatitis B virus infection and HLA-DR type in Korea. Hepatology 31:1371–1373

Bellamy R, Ruwende C, Corrah T, McAdam KP, Whittle HC, Hill AV (1998) Variations in the *NRAMP1* gene and susceptibility to tuberculosis in West Africans. N Engl J Med 338:640–644

Bellamy R, Ruwende C, Corrah T, McAdam KP, Thursz M, Whittle HC, Hill AV (1999) Tuberculosis and chronic hepatitis B virus infection in Africans and variation in the vitamin D receptor gene. J Infect Dis 179:721–724

Bellamy R, Beyers N, McAdam KP, Ruwende C, Gie R, Samaai P, Bester D, Meyer M, Corrah T, Collin M, Camidge DR, Wilkinson D, Hoal-Van Helden E, Whittle HC, Amos W, van Helden P, Hill AV (2000) Genetic susceptibility to tuberculosis in Africans: a genome-wide scan. Proc Natl Acad Sci USA 97:8005–8009

Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. Nature 325:31–36

Cellier M, Govoni G, Vidal S, Kwan T, Groulx N, Liu J, Sanchez F, Skamene E, Schurr E, Gros P (1994) Human natural resistance-associated macrophage protein: cDNA cloning, chromosomal mapping, genomic organization, and tissue-specific expression. J Exp Med 180:1741–1752

Dean M, Carrington M, Winkler C, Huttley GA, Smith MW, Allikmets R, Goedert JJ, Buchbinder SP, Vittinghoff E, Gomperts E, Donfield S, Vlahov D, Kaslow R, Saah A, Rinaldo C, Detels R, O'Brien SJ (1996) Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the *CKR5* structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. Science 273:1856–1862

Dunnen JT, Antonarakis SE (2000) Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. Hum Mutat 15:7–12

Gao PS, Fujishima S, Mao XQ, Remus N, Kanda M, Enomoto T, Dake Y, Bottini N, Tabuchi M, Hasegawa N, Yamaguchi K, Tiemessen C, Hopkin JM, Shirakawa T, Kishi F (2000) Genetic variants of *NRAMP1* and active tuberculosis in Japanese populations. International Tuberculosis Genetics Team. Clin Genet 58:74–76

Greenwood CM, Fujiwara TM, Boothroyd LJ, Miller MA, Frappier D, Fanning EA, Schurr E, Morgan K (2000) Linkage of tuberculosis to chromosome 2q35 loci, including *NRAMP1*, in a large aboriginal Canadian family. Am J Hum Genet 67:405–416

Harding RM, Fullerton SM, Griffiths RC, Clegg JB (1997) A gene tree for beta-globin sequences from Melanesia. J Mol Evol 44:S133–S138

Harris EE, Hey J (1999) X chromosome evidence for ancient human histories. Proc Natl Acad Sci USA 96:3320–3324

Horai S, Hayasaka K, Kondo R, Tsugane K, Takahata N (1995) Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. Proc Natl Acad Sci USA 92:532–536

Kaessmann H, Heissig F, von Haeseler A, Paabo S (1999) DNA sequence variation in a non-coding region of low recombination on the human X chromosome. Nat Genet 22:78–81

Kim CH, Oh Y, Lee TH (1997) Codon optimization for high-level expression of human erythropoietin (EPO) in mammalian cells. Gene 199:293–301

Liu J, Fujiwara TM, Buu NT, Sanchez FO, Cellier M, Paradis AJ, Frappier D, Skamene E, Gros P, Morgan K, Schurr E (1995) Identification of polymorphisms and sequence variants in the human homologue of the mouse natural resistance-associated macrophage protein gene. Am J Hum Genet 56:845–853

Marquet S, Lepage P, Hudson TJ, Musser JM, Schurr E (2000) Complete nucleotide sequence and genomic structure of the human *NRAMP1* gene region on chromosome region 2q35. Mamm Genome 11:755–762

Relethford JH (2001) Ancient DNA and the origin of modern humans. Proc Natl Acad Sci USA 98:390–391

Ryu S, Park YK, Bai GH, Kim SJ, Park SN, Kang S (2000) 3′UTR polymorphisms in the *NRAMP1* gene are associated with susceptibility to tuberculosis in Koreans. Int J Tuberc Lung Dis 4:577–580

Searle S, Blackwell JM (1999) Evidence for a functional repeat polymorphism in the promoter of the human *NRAMP1* gene that correlates with autoimmune versus infectious disease susceptibility. J Med Genet 36:295–299

Shaw MA, Collins A, Peacock CS, Miller EN, Black GF, Sibthorpe D, Lins-Lainson Z, Shaw JJ, Ramos F, Silveira F, Blackwell JM (1997) Evidence that genetic susceptibility to Mycobacterium tuberculosis in a Brazilian population is under oligogenic control: linkage study of the candidate genes *NRAMP1* and *TNFA*. Tuber Lung Dis 78:35–45

Stephens JC, Schneider JA, Tanguay DA, Choi J, Acharya T, Stanley SE, Jiang R, Messer CJ, Chew A, Han JH, Duan J, Carr JL, Lee MS, Koshy B, Kumar AM, Zhang G, Newell WR, Windemuth A, Xu C, Kalbfleisch TS, Shaner SL, Arnold K, Schulz V, Drysdale CM, Nandabalan K, Judson RS, Ruano G, Vovis GF (2001) Haplotype variation and linkage disequilibrium in 313 human genes. Science 293:489–493

Thomson R, Pritchard JK, Shen P, Oefner PJ, Feldman MW (2000) Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. Proc Natl Acad Sci USA 97:7360–7365

Tishkoff SA, Dietzsch E, Speed W, Pakstis AJ, Kidd JR, Cheung K, Bonne-Tamir B, Santachiara-Benerecetti AS, Moral P, Krings M (1996) Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. Science 271:1380–1387

Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonne-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner PJ (2000) Y chromosome sequence variation and the history of human populations. Nat Genet 26:358–361

Vogel F, Motulasky AG (1997) Human genetics–problems and approaches, 3rd edn. Springer, Berlin Heidelberg New York