

Is *Mycobacterium tuberculosis* 15,000 Years Old?

Colleagues—One-third of the world's population is infected with *Mycobacterium tuberculosis*, the cause of tuberculosis, and 8 million new cases occur each year [1]. Moreover, nearly 3 million people die annually from tuberculosis, making it the leading cause of death due to an infectious agent worldwide [1]. Paleopathologic evidence suggests that tuberculosis afflicted humans as early as 3700 BC in Egypt and 2500–1500 BC in Europe and also in pre-Columbian North and South America [2]. The recent recovery [3] of *M. tuberculosis* DNA from lung lesions in a 1000-year-old Peruvian mummy confirms that the disease existed in the pre-Columbian New World.

Inasmuch as *M. tuberculosis* has long been associated with human populations, there is a very large global pool of infected individuals, and chromosomal heterogeneity based on restriction fragment length polymorphism (RFLP) patterns generated by probing with mobile elements such as IS6110 has been reported [4], it is reasonable to expect that substantial levels of genetic diversity exist among isolates of the species. However, characterization of eight loci (192,875 nucleotides) from unassociated isolates recovered in North America and Europe revealed almost a complete absence of nucleotide variation (table 1). To rule out the possibility that restricted geographic sampling biased the data set, 350-bp fragments of *rpoB*, *hsp65*, *gyrA*, *aroA* (GenBank M62708; positions 714–1063), and *recA* (GenBank X58485; positions 1646–1995) and a 1435-bp region of the *katG* gene were sequenced from 1 randomly selected isolate each from 7 countries (Switzerland, Turkey, Algeria, Somalia, Papua New Guinea, Vietnam, and Tibet) with well-differentiated human populations. We attempted to increase the likelihood of identifying allelic variation by analysis of a region of *hsp65* that encodes an immunodominant T cell epitope [5] and an area of *recA* that codes for a protein intein [6]. A virtual lack of nucleotide variation was found in the 7 isolates. The only variation was a missense mutation in *katG* and 1 nucleotide change in *rpoB*, which are associated with resistance to isoniazid

(unpublished data) and rifampicin [7, 8], respectively. Telenti et al. [7] also reported absence of variation in a 411-bp segment of *rpoB* among 56 rifampin-sensitive isolates cultured from patients in 11 countries worldwide, including many not sampled in our study.

Two hypotheses that can be invoked to explain the unusual lack of nucleotide diversity in *M. tuberculosis* from global sources include low mutation rate as a result of unusual DNA replication fidelity or repair capability and an evolutionarily recent origin and worldwide spread of a clone in an episode of periodic selection. The report [9] of a spontaneous mutation frequency of 10^{-5} – 10^{-8} determined by drug resistance studies effectively eliminates the hypothesis of high replication fidelity or unusual DNA repair capacity. To place the amount of nucleotide variation among *M. tuberculosis* isolates on an evolutionary time scale, we combined the 752 codons of four genes (*hsp65*, *rpoB*, *orf1*, and *inhA*) sequenced in the same 31 strains and estimated the average number of synonymous substitutions per synonymous site (K_s) [10] across all pairwise comparisons of these strains. The value of $K_s = 0.00012$ (± 0.00002) indicates that on average, *M. tuberculosis* isolates have only one synonymous difference per 10,000 synonymous sites. Sharp [11] examined 67 homologous genes from *Escherichia coli* and *Salmonella typhimurium* and found an average K_s of 0.94. With the divergence time between *E. coli* and *S. typhimurium* (as estimated by Ochman and Wilson [12]) of 120–160 million years, we estimate the synonymous substitution rate to be 0.0058–0.0078 per site per million years. Under the assumption that *M. tuberculosis* accumulates synonymous changes at this rate, the amount of synonymous divergence among contemporary isolates of this pathogen is estimated to have occurred in 15,300–20,400 years. This calculation dates speciation and global dissemination of *M. tuberculosis* to roughly the same time as the paleomigration into the New World. The time frame is also consistent with speculation [13] that the agent of human tuberculosis arose from the very closely related cattle pathogen *Mycobacterium bovis* by host specialization occurring since the domestication of this animal some 8000–10,000 years ago.

The findings have several implications for *M. tuberculosis* pathobiology research. For example, the virtual absence of substitutions, except those believed to confer drug resistance and undoubtedly observed due to selective pressure, together with the subclonal chromosomal heterogeneity based on IS6110 RFLPs, suggests that insertion sequence-associated polymor-

Reprints or correspondence: Dr. James M. Musser, Section of Molecular Pathobiology, Dept. of Pathology, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030.

The Journal of Infectious Diseases 1994;170:1348–9
© 1994 by The University of Chicago. All rights reserved.
0022-1899/94/7005-0053\$01.00

Table 1. Eight *M. tuberculosis* loci analyzed for nucleotide sequence diversity.

Locus	Size (bp)	No. of strains	Synonymous substitutions	Nucleotides sequenced	GenBank no.
<i>rpoB</i>	350	200	3 (4 isolates)	83–432	LO5910
<i>inhA</i>	810	37	0	22–831	UO2492
<i>hsp65</i>	366	30	0	453–818	M15467
<i>orf1</i>	700	31	1 (1 isolate)	1–700	—
<i>gyrA</i>	318	50	0	2384–2701	L27512
<i>rpsL</i>	350	15	0	10–359	X70995
<i>gyrB</i>	351	15	0	1580–1930	L27512
<i>katG</i>	1435	15	0	2191–3610*	X68081

* Does not total 1435 bp because of errors in initially deposited sequence (unpublished data; Cole S, personal communication).

phism is accumulating more rapidly than simple nucleotide variation. Also, restricted allelic diversity means that it is probable that only nominal amino acid variation will occur in proteins that are of potential immunophylaxis or virulence interest. To gain further insight into the molecular evolutionary genetics of *M. tuberculosis*, it will be necessary to determine the level of sequence variation present in the other closely allied members of the *M. tuberculosis* complex (*M. bovis*, *Mycobacterium africanum*, and *Mycobacterium microti*) and the related slowly growing pathogen *Mycobacterium leprae*, for which there is a report of limited chromosomal diversity based on restriction endonuclease analysis [14].

Vivek Kapur, Thomas S. Whittam, and James M. Musser

Section of Molecular Pathobiology, Department of Pathology, Baylor College of Medicine, Houston, Texas; Institute of Molecular Evolutionary Genetics, Pennsylvania State University, University Park

References

1. Bloom BR, Murray CJL. Tuberculosis: commentary on a reemerging killer. *Science* 1992;257:1055-64.
2. Ayvazian LF. History of tuberculosis. In: Reichman LB, Hershfield ES, eds. *Tuberculosis: a comprehensive international approach*. New York: Marcel Dekker, 1993:1-20.
3. Salo WL, Aufderheide AC, Buikstra J, Holcomb TA. Identification of *Mycobacterium tuberculosis* DNA in a pre-Columbian Peruvian mummy. *Proc Natl Acad Sci USA* 1994;91:2091-4.
4. van Embden JDA, van Soolingen D, Small PM, Hermans PWM. Genetic markers for the epidemiology of tuberculosis. *Res Microbiol* 1992;143:385-91.
5. Young D, Lathigra R, Hendrix R, Sweetser D, Young RA. Stress proteins are immune targets in leprosy and tuberculosis. *Proc Natl Acad Sci USA* 1988;85:4267-70.
6. Davis EO, Jenner PJ, Brooks PC, Colston MJ, Sedgwick SG. Protein splicing in the maturation of *M. tuberculosis* RecA protein: a mechanism for tolerating a novel class of intervening sequence. *Cell* 1992;71:201-10.
7. Telenti A, Imboden P, Marchesi F, et al. Detection of rifampicin-resistance mutations in *Mycobacterium tuberculosis*. *Lancet* 1993;341:647-50.
8. Kapur V, Li LL, Iordanescu S, et al. Characterization by automated DNA sequencing of mutations in the gene (*rpoB*) encoding the RNA polymerase β subunit in rifampin-resistant *Mycobacterium tuberculosis* strains from New York City and Texas. *J Clin Microbiol* 1994;32:1095-8.
9. David HL, Newman CM. Some observations on the genetics of isoniazid resistance in the tubercle bacilli. *Am Rev Respir Dis* 1971;104:508-15.
10. Li WH, Wu CI, Luo CC. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol Biol Evol* 1985;2:150-74.
11. Sharp PM. Determinants of DNA sequence divergence between *Escherichia coli* and *Salmonella typhimurium*: codon usage, map position, and concerted evolution. *J Mol Evol* 1991;33:23-33.
12. Ochman H, Wilson AC. Evolutionary history of enteric bacteria. In: Neidhardt FC, Ingraham JL, Low KB, Magasanik B, Schaechter M, Umberger HE, eds. *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. Washington, DC: American Society for Microbiology, 1987:1649-54.
13. Webb GB. *Tuberculosis*. New York: Hoeber, 1936.
14. Williams DL, Gillis TP, Portaels F. Geographically distinct isolates of *Mycobacterium leprae* exhibit no genotypic diversity by restriction fragment-length polymorphism analysis. *Mol Microbiol* 1990;4:1653-9.