6. L. C. Van Loon, E. A. Van Strien, *Physiol. Mol. Plant Pathol.* **55**, 85 (1999).
7. M. Kinkema, W. Fan, X. Dong, *Plant Cell* **12**, 2339 (2000).
8. A. Vitale, J. Denecke, *Plant Cell* **11**, 615 (1999).
9. E. S. Trombetta, A. J. Parodi, *Annu. Rev. Cell Dev. Biol.* **19**, 649 (2003).
10. J. Fu, G. Kreibich, *J. Biol. Chem.* **275**, 3984 (2000).
11. T. M. Chu, B. Weir, R. Wolfinger, *Math. Biosci.* **176**, 35 (2002).
12. http://genome-www5.stanford.edu
13. H. Walther-Larsen, J. Brandt, D. B. Collinge, H. Thordal-Christensen, *Plant Mol. Biol.* **21**, 1097 (1993).
14. J. Denecke et al., *Plant Cell* **7**, 391 (1995).
15. P. Wick et al., *Plant Physiol.* **132**, 343 (2003).
16. E. P. Jelitto-Van Dooren, S. Vidal, J. Denecke, *Plant Cell* **11**, 1935 (1999).
17. M. Petersen et al., *Cell* **103**, 1111 (2000).
18. J. M. Alonso et al., *Science* **301**, 653 (2003).
19. B. L. Lin et al., *Cell Stress Chaperones* **6**, 201 (2001).
20. S. J. Noh, C. S. Kwon, D. H. Oh, J. S. Moon, W. I. Chung, *Gene* **311**, 81 (2003).
21. D. Wang, N. D. Weaver, M. Kesarwani, X. Dong, data not shown.
22. C. Patil, P. Walter, *Curr. Opin. Cell Biol.* **13**, 349 (2001).
23. I. M. Martinez, M. J. Chrispeels, *Plant Cell* **15**, 561 (2003).
24. T. L. Bailey, C. Elkan, *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2**, 28 (1994).
25. Y. Zhang, M. J. Tessaro, M. Lassner, X. Li, *Plant Cell* **15**, 2647 (2003).
26. E. van Anken et al., *Immunity* **18**, 243 (2003).
27. H. Cao, S. A. Bowling, S. Gordon, X. Dong, *Plant Cell* **6**, 1583 (1994).

On the Origin of Leprosy

Marc Monot,[1]* Nadine Honoré,[1]* Thierry Garnier,[1] Romulo Araoz,[1] Jean-Yves Coppée,[2] Céline Lacroix,[2] Samba Sow,[3] John S. Spencer,[4] Richard W. Truman,[5] Diana L. Williams,[5] Robert Gelber,[6] Marcos Virmond,[7] Béatrice Flageul,[8] Sang-Nae Cho,[9] Baohong Ji,[10] Alberto Paniz-Mondolfi,[11] Jacinto Convit,[11] Saroj Young,[12] Paul E. Fine,[12] Voahangy Rasolofo,[13] Patrick J. Brennan,[4] Stewart T. Cole[1]†

Leprosy, a chronic human disease with potentially debilitating neurological consequences, results from infection with *Mycobacterium leprae*. This unculturable pathogen has undergone extensive reductive evolution, with half of its genome now occupied by pseudogenes. Using comparative genomics, we demonstrated that all extant cases of leprosy are attributable to a single clone whose dissemination worldwide can be retraced from analysis of very rare single-nucleotide polymorphisms. The disease seems to have originated in Eastern Africa or the Near East and spread with successive human migrations. Europeans or North Africans introduced leprosy into West Africa and the Americas within the past 500 years.

Comparative genomics enables us to establish solid genealogical relationships with greater precision than ever before. Leprosy (*1*) has plagued human populations for thousands of years and puzzled scientists since the identification of its etiological agent, *Mycobacterium leprae*, by Hansen in 1873 (*2*). The main difficulties of working with *M. leprae* are that it cannot be grown in axenic culture and that its doubling time in tissue is slow, nearly 13 days (*3*). It was only when it was discovered that the nine-banded armadillo, *Dasypus novemcinctus*, could be infected (*4*) that sufficient quantities of *M. leprae* were obtained for biological and immunological analysis. Comparison of the genome sequence of the armadillo-passaged strain of *M. leprae* from Tamil Nadu, India (TN strain) with that of the close relative *Mycobacterium tuberculosis* (*5*), led to a major breakthrough (*6*). *M. leprae* was shown to have embarked upon a path of reductive evolution in which the genome underwent downsizing and accumulated more than 1130 pseudogenes. The concomitant loss of catabolic and respiratory functions appears to have resulted in severe metabolic constraints (*6*, *7*).

To establish whether all strains of *M. leprae* had undergone similar events and to determine their level of relatedness, we used technological approaches that have successfully detected polymorphic regions in the *M. tuberculosis* complex (*8–10*). First, genomic DNA, prepared from seven different strains of leprosy bacilli (Table 1), was hybridized to microarrays corresponding to the complete genome of the TN strain, but no evidence for further gene loss was uncovered in these isolates (fig. S1). Second, to establish whether differences existed in the copy number of insertion-sequence–like, dispersed repetitive sequences, quantitative polymerase chain reaction was performed to target the repetitive sequences RLEP, REPLEP, LEPREP, and LEPRPT (*11*). Again, within the limits of sensitivity of this approach, no differences were detected between the TN strain and the other isolates (fig. S2).

A major source of variability in tubercle bacilli is the mycobacterial interspersed repetitive unit (MIRU), which serves as the basis of a robust typing system that exploits differences in the variable number of the tandem repeats

[1]Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, Paris, France. [2]PF2, Génopole, Institut Pasteur, Paris, France. [3]Centre National d'Appui à la lutte Contre la Maladie, Bamako, Mali. [4]Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins, Colorado 80523–1682, USA. [5]National Hansen's Disease Program, DHHS/HRSA/BPHC, Lousiana State University, Baton Rouge, Lousiana 70894, USA. [6]Leonard Wood Memorial Center for Leprosy Research, Cebu, Philippines. [7]Instituto Lauro de Souza Lima, Bauru, São Paulo, Brazil. [8]Hôpital Saint-Louis, Paris, France. [9]Yonsei University College of Medicine, Seoul, Republic of Korea. [10]Bactériologie-Hygiène, Faculté de Médecine Pitié-Salpêtrière, Paris, France. [11]Instituto de Biomedicina, Caracas, Venezuela. [12]London School of Hygiene and Tropical Medicine, London, UK. [13]Institut Pasteur de Madagascar, Antananarivo, Madagascar.

*These authors contributed equally to this work.
†To whom correspondence should be addressed: E-mail: stcole@pasteur.fr

Table 1. Strains of armadillo-derived *M. leprae* and VNTR profile.

Strain	Patient's country of origin	Source	3-Hexa	21-TTC	9-GTA	14-AT	15-AT	17-AT	18-AT
Tamil Nadu*	India	IP	3	21	9	14	15	17	18
Africa	Ethiopia	IP	3	29	8	14	19	13	13
India 2	India	IP	3	15	11	18	14	13	9
Br4923	Brazil	NHDP	3	12	12	20	20	15	18
NHDP98	Mexico	CSU/NHDP	3	10	9	22	14	11	12
Thai-53	Thailand	CSU/NHDP	3	15	9	16	17	10	13
NHDP63	USA	CSU/NHDP	3	10	10	18	18	13	16

*Numbers refer to the repeat copy number for the Tamil Nadu strain (*11*), whereas numbers in the rest of the table are the copy numbers found in the respective isolates. IP, Institut Pasteur; NHDP, National Hansen's Disease Program; CSU, Colorado State University.

(VNTR) that make up this repetitive element (12). Unlike *M. tuberculosis,* none of the 20 MIRU loci in the TN strain contains tandem repeats of the element (11) and, on examination of the additional strains, no copy number differences were detected. Furthermore, the 20 MIRUs were of identical sequence in all seven strains studied (Table 1). Seven other VNTR, with two to six base-pair (bp) repeats, were also targeted, because some of them have proved useful for tracking strains over short epidemiological distances (13–16). No variation was seen in a hexanucleotide repeat situated within the coding sequence of the *sigA* (*rpoT*) gene (17), whereas on examination of two trinucleotide repeats and four dinucleotide repeats, located in pseudogenes or noncoding regions of the genome, extensive differences were seen in copy number (Table 1). However, as expected for such sequences, which are highly prone to slipped-strand mispairing during replication (18), the level of variability was too great to allow patterns to be detected.

Although these results rule out the existence of the most likely insertion and deletion events, they are less informative about genome topology and global organization. These features were surveyed by fingerprinting and end-sequencing 1466 cosmids from a library of a second Indian strain of *M. leprae*, leading to an integrated genome map that showed perfect cocircularity with that of the TN strain (19, 20). To increase the likelihood of detecting single-nucleotide polymorphisms (SNPs), selected genes, noncoding regions, and pseudogenes were sequenced from a Brazilian strain, Br4923 (table S1). This strain was chosen for two reasons: the relative geographic remoteness of the country and the severity of the disease burden in Brazil, which is second highest worldwide after India (1). By this

means, five SNPs were revealed in 142 kb of sequenced DNA, one in an apparently noncoding region and four in pseudogenes. When all seven strains were analyzed, only three of the SNPs were found in two or more of the strains tested (Fig. 1A), whereas the remaining two were restricted to the TN strain of *M. leprae*. Overall, the SNP frequency observed in *M. leprae* of ~1 per 28 kb was significantly less than that seen in other human pathogens, such as the tubercle bacilli (8, 9, 21, 22), *Salmonella typhi* (23), and *Helicobacter pylori* (24) (Table 2). Taken together, these findings indicate that the *M. leprae* genome is exceptionally well conserved and that the leprosy bacillus is highly clonal (25).

To gain insight into the worldwide distribution of the *M. leprae* SNPs, we sought the three informative SNPs in a total of 175 clinical and laboratory specimens from 21 countries and all five continents. We discovered that of a possible 64 permutations only 4 occurred (Table 3 and table S2), referred to as SNP types 1 to 4. When the VNTR panel was probed, extensive variability was found for six VNTRs (table S2), but no particular VNTR pattern was associated with a given SNP type. In contrast, a correlation exists between the geographical origin of the leprosy patient and the SNP profile, because type 1 occurs predominantly in Asia, the Pacific region, and East

Africa, type 4 in West Africa and the Caribbean region, and type 3 in Europe, North Africa, and the Americas. SNP type 2 is the rarest and has only been detected in Ethiopia, Malawi, Nepal/North India, and New Caledonia.

Ancient texts describe the existence of leprosy in China, India, and Egypt in about 600 BC, and skeletal remains bearing hallmarks of the disease have been found in Egypt (26). Leprosy is believed to have originated in the Indian subcontinent and to have been introduced into Europe by Greek soldiers returning from the Indian campaign of Alexan-

Table 2. Comparison of SNP frequency in other bacterial pathogens.

Pathogen	SNP frequency/bp	Reference
M. leprae	1 in 28,400	This work
M. tuberculosis complex	1 in ~3,000	(21, 22)
S. typhi	1 in 1,112	(23)
H. pylori	1 in 3.2	(24)

Table 3. SNP analysis of *M. leprae* from different countries.

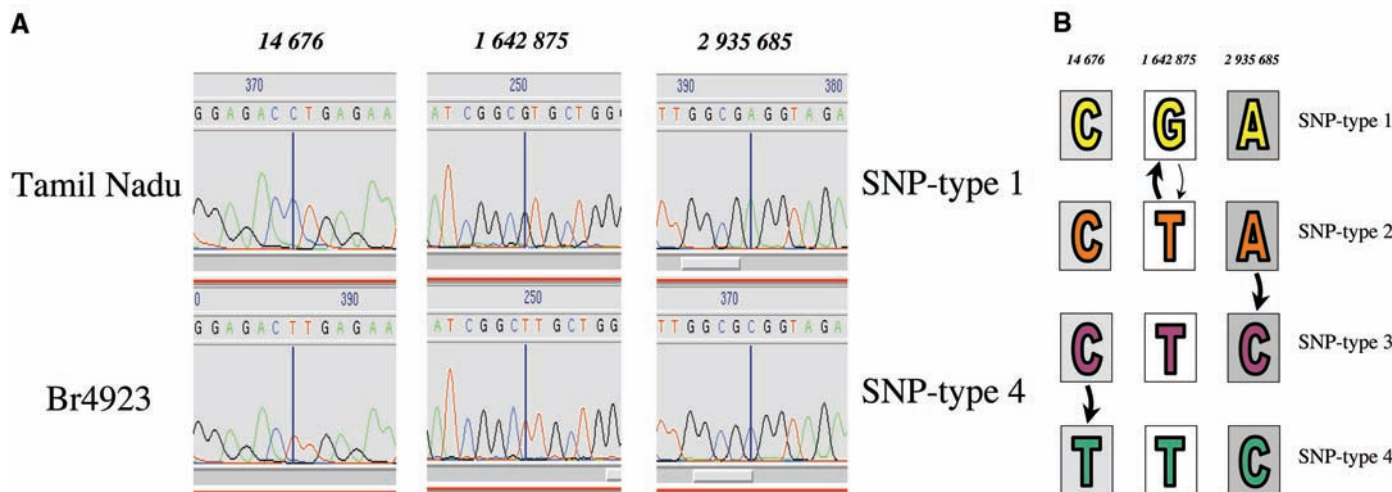| Country	SNP type 1	SNP type 2	SNP type 3	SNP type 4
New Caledonia	3	1	3	
Philippines	19		2	
Korea	3		2	
Thailand	1			
Nepal/North India	23	5		
South India	4			
Madagascar	6			
Ethiopia		2		
Malawi	4	6		
Mali				31
Ivory Coast				6
Guinea				1
Senegal				2
Morocco			2	
France			2	
Brazil			12	2
French West Indies	4		2	14
Venezuela			5	
Mexico			1	
United States			3	
WARM*			4	
Total	67	14	38	56

*Wild armadillos from Louisiana, USA.

Fig. 1. SNP analysis of isolates of different geographical origin and parsimony. (A) Comparison of polymorphic sites in the genomes of the TN and Br4923 strains by automated DNA sequencing. Coordinates are the position in the genome of the TN strain, and the vertical bar indicates the polymorphic base. (B) The most parsimonious route to account for the four SNP types. Bold arrows indicate the most likely direction, based on historical and geographic considerations; the faint arrow denotes an alternative route.

der the Great (*26*). From Greece, the disease is thought to have spread around the Mediterranean basin, with the Romans introducing leprosy into the Western part of Europe. Little is known about its presence in sub-Saharan Africa except that the disease was present prior to the colonial era. From India, leprosy is thought to have spread to China and then to Japan, reaching Pacific Islands like New Caledonia as recently as the 19th century.

Our results provide evidence for a general evolutionary scheme for *M. leprae* and, on the basis of our interpretation of the SNP data, offer two alternative conclusions for the global spread of leprosy that differ from classic explanations. Two equally plausible evolutionary scenarios are possible (Figs. 1B and 2). In the first, SNP type 2, from East Africa/Central Asia, preceded type 1, which migrated eastward, and type 3, which disseminated westward in human populations, before giving rise to type 4. In the second scenario, type 1 was the progenitor of type 2, with SNP types 3 and 4 following in that order.

Leprosy was most likely introduced into West Africa by infected explorers, traders, or colonialists of European or North African descent, rather than by migrants from East Africa, because SNP type 4 is much closer to type 3 than to type 1 (Fig. 1B). West and southern Africa are thought to have been settled >50,000 years ago by migrants from East Africa before the arrival of humans in Eurasia (*27, 28*). It seems unlikely that early humans brought leprosy into West Africa with them unless that particular bacterial clone has since been replaced. From West Africa, leprosy was then introduced by the slave trade in the 18th century to the Caribbean islands, Brazil, and

probably other parts of South America, because isolates of *M. leprae* with the same SNP type, 4, are found there as in West Africa.

The strain of *M. leprae* responsible for disease in most of the Americas is closest to the European/North African variety (Fig. 1B), which indicates that colonialism and emigration from the old world most probably contributed to the introduction of leprosy into the new world. For instance, in the 18th and 19th centuries, when the midwestern states of the United States were settled by Scandinavian immigrants, many cases of leprosy were reported and, at that time, a major epidemic was under way in Norway (*26*). Further support for this hypothesis is provided by the finding that wild armadillos from Louisiana, which are naturally infected with *M. leprae*, harbor the European/North African SNP type 3 strain, indicating that they were contaminated by human sources. Although most mycobacteria occur in the soil, there is no convincing evidence for an environmental reservoir of *M. leprae* and, apart from armadillos, which have limited geographical distribution and only very recently became infected, there is no known animal source of the pathogen. Although an ancient zoonotic origin cannot be excluded, insect bites may also have been a possible route of early human infection, particularly as recent studies show that *M. ulcerans*, a related pathogen with many pseudogenes, appears to be transmitted by aquatic insects (*29*).

In conclusion, *M. leprae,* with its exceptionally stable genome, is a helpful marker for tracking the migration of peoples and retracing the steps that led to modern human populations. In this respect, it complements *H. pylori,* which is considerably more di-

verse and thus allows finer understanding of the ethnic origin of humans (*30*). It is noteworthy that the greatest variety of SNP types in the leprosy bacillus is found in islands such as the French West Indies and New Caledonia (Fig. 2), reflecting the passage of, and settlement by, different human populations. Finally, the remarkable clonality seen in isolates of *M. leprae* indicates that genome decay occurred prior to the global spread of leprosy and that it has not accelerated substantially since.

References and Notes
1. W. J. Britton, D. N. J. Lockwood, *Lancet* **363**, 1209 (2004).
2. G. H. A. Hansen, *Norsk Mag. Laegervidenskaben* **4**, 1 (1874) (supplement).
3. C. C. Shepard, D. H. McRae, *J. Bacteriol.* **89**, 365 (1965).
4. W. K. Kirchheimer, E. E. Storrs, *Int. J. Lepr.* **39**, 693 (1971).
5. S. T. Cole *et al.*, *Nature* **393**, 537 (1998).
6. S. T. Cole *et al.*, *Nature* **409**, 1007 (2001).
7. K. Eiglmeier *et al.*, *Lepr. Rev.* **72**, 387 (2001).
8. R. Brosch, A. S. Pym, S. V. Gordon, S. T. Cole, *Trends Microbiol.* **9**, 452 (2001).
9. R. Brosch *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 3684 (2002).
10. P. Supply *et al.*, *Mol. Microbiol.* **36**, 762 (2000).
11. S. T. Cole, P. Supply, N. Honoré, *Lepr. Rev.* **72**, 449 (2001).
12. E. Mazars *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **13**, 1901 (2001).
13. Y. C. Shin, H. Lee, G. P. Walsh, J. D. Kim, S. N. Cho, *J. Clin. Microbiol.* **38**, 4535 (2000).
14. R. Truman, A. B. Fontes, A. B. de Miranda, P. Suffys, T. Gillis, *J. Clin. Microbiol.* **42**, 2558 (2004).
15. N. A. Groathouse *et al.*, *J. Clin. Microbiol.* **42**, 1666 (2004).
16. S. K. Young *et al.*, *J. Clin. Microbiol.* **42**, 4931 (2004).
17. M. Matsuoka *et al.*, *Int. J. Lepr. Mycobact. Dis.* **68**, 121 (2000).
18. S. T. Lovett, *Mol. Microbiol.* **52**, 1243 (2004).
19. K. Eiglmeier, S. Simon, T. Garnier, S. T. Cole, *Lepr. Rev.* **72**, 462 (2001).
20. K. Eiglmeier, N. Honoré, S. A. Woods, B. Caudron, S. T. Cole, *Mol. Microbiol.* **7**, 197 (1993).
21. R. D. Fleischmann *et al.*, *J. Bacteriol.* **184**, 5479 (2002).
22. T. Garnier *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 7877 (2003).
23. C. Kidgell *et al.*, *Infect. Genet. Evol.* **2**, 39 (2002).
24. D. Falush *et al.*, *Science* **299**, 1582 (2003).
25. J. M. Smith, N. H. Smith, M. O'Rourke, B. G. Spratt, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 4384 (1993).
26. S. G. Browne, in *Leprosy*, R. C. Hastings, Ed. (Churchill Livingstone, Edinburgh, 1985).
27. P. A. Underhill *et al.*, *Nat. Genet.* **26**, 358 (2000).
28. L. L. Cavalli-Sforza, M. W. Feldman, *Nat. Genet.* **33** (suppl.), 266 (2003).
29. L. Marsollier *et al.*, *Appl. Environ. Microbiol.* **68**, 4623 (2002).
30. T. Wirth *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 4746 (2004).
31. We thank S. Brisse, R. Brosch, S. V. Gordon, R. Davies, and C. Morey for their help, advice, and encouragement, and the patients and staff of the various leprosy clinics for their invaluable participation. The DNA sequences of SNP-containing regions have been deposited at GenBank with accession numbers AY960580 to AY960582. This work received the financial support of the Institut Pasteur, the Association Française Raoul Follereau, Lepra, the Consortium National de Recherche en Génomique (RNG program), and the NIH National Institute of Allergy and Infectious Diseases (grant RO1-AI47197-01A1 and contract NO1-AI25469).
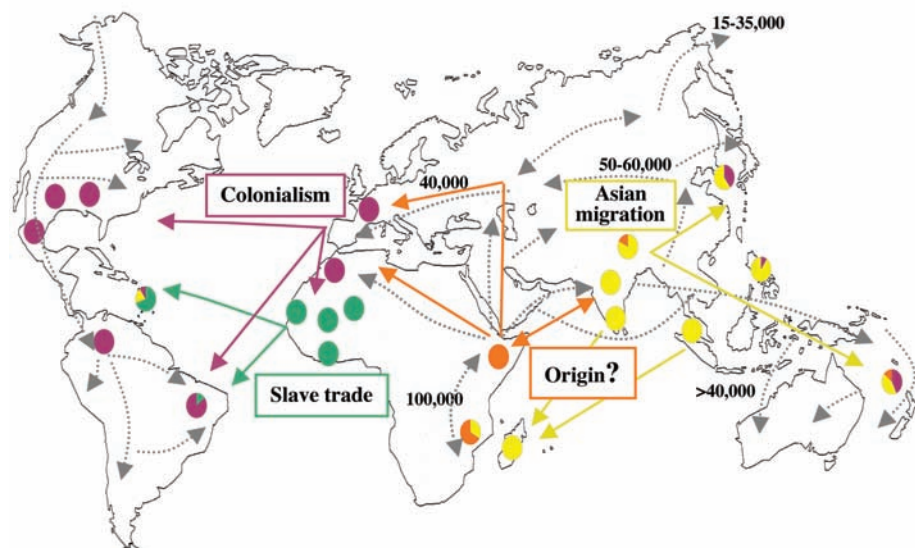
Fig. 2. Dissemination of leprosy in the world. The circles indicate the country of origin of the samples examined and their distribution into the four SNP types, which are color coded as in Fig. 1B. The colored arrows indicate the direction of human migrations predicted by, or inferred from, our SNP analysis; gray arrows correspond to the migration routes of humans derived from genetic, archaeological, and anthropological studies, with the estimated time of migration in years (*27, 28*).