HAL
archives-ouvertes.fr

On the origin of leprosy.

Marc Monot, Nadine Honoré, Thierry Garnier, Romulo Araoz, Jean-Yves
Coppée, Céline Lacroix, Samba Sow, John S. Spencer, Richard W. Truman,
Diana L. Williams, et al.

▶ To cite this version:

HAL Id: pasteur-00204117

https://hal-pasteur.archives-ouvertes.fr/pasteur-00204117

Submitted on 12 Jan 2008

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], and Stewart T. Cole[1+]

[1] *Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, Paris, France*

[2] *PF2, Génopole, Institut Pasteur, Paris, France*

[3] *Centre National d'Appuis 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 whom correspondence should be addressed: E-mail: stcole@pasteur.fr

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 (SNP). 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 difficulty of working with *M. leprae* is that it cannot be grown in axenic culture and 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 over 1,130 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 (Figure S1). Second, to establish whether differences existed in the copy number of Insertion Sequence-like, dispersed repetitive sequences, quantitative PCR 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 (Figure 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 (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 2 − 6 bp repeats, were also targeted, as 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- and four dinucleotide-repeats, located in pseudogenes or non-coding regions of the genome, extensive differences were seen in copy number (Table 1). However, as expected for such sequences, that are highly prone to slipped strand mispairing during replication (*18*), the level of variability was too great to allow patterns to be detected.

While 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 1,466 cosmids from a library of a second Indian strain of *M. leprae* leading to an integrated genome map that showed perfect co-circularity with that of the TN strain (*19, 20*). To increase the likelihood of detecting

single nucleotide polymorphisms (SNP), selected genes, non-coding regions and pseudogenes were sequenced from a Brazilian strain, Br4923 (Table S1), that 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 non-coding 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 (Figure 1A) while 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 like 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 four occurred (Tables 3 and S2), referred to as SNP-types 1 – 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, as 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 about 600BC 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 Alexander 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 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, based on our interpretation of the SNP data, offer two alternative conclusions for the global spread of leprosy that differ from classical explanations. Two equally plausible evolutionary scenarios are possible (Figures 1B, 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, as SNP-type 4 is much closer to type 3, than to type 1 (Figure 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, as 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 (Figure 1B) indicating 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 mid-western states of the USA were settled by Scandinavian immigrants, many cases of leprosy were reported and, at that time, a major epidemic was underway in Norway (*26*). Further support for this hypothesis is provided by the finding that wild armadillos from Louisiana, USA, which are naturally-infected with *M. leprae*, harbor the European/North African SNP-type 3 strain indicating that they were contaminated from human sources. While 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 *Mycobacterium 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 diverse 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 (Figure 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 global spread of leprosy and that it has not accelerated significantly since.

References and Notes

1.   W. J. Britton, D. N. J. Lockwood, *Lancet* **363**, 1209-1219 (2004).

2.   G. H. A. Hansen, *Norsk Magazin for Laegervidenskaben (supplement)* **4**, 1-88 (1874).

3.   C. C. Shepard, D. H. McRae, *J. Bacteriol.* **89**, 365-372 (1965).

4.   W. K. Kirchheimer, E. E. Storrs, *Int. J. Lep.* **39**, 693-702 (1971).

5.   S. T. Cole *et al.*, *Nature* **393**, 537-544 (1998).

6.   S. T. Cole *et al.*, *Nature* **409**, 1007-1011 (2001).

7.   K. Eiglmeier *et al.*, *Lepr. Rev.* **72**, 387-398 (2001).

8.   R. Brosch, A. S. Pym, S. V. Gordon, S. T. Cole, *Trends Microbiol.* **9**, 452-458 (2001).

9.   R. Brosch *et al.*, *Proc. Natl. Acad. Sci. USA* **99**, 3684-9. (2002).

10.   P. Supply *et al.*, *Molec. Microbiol.* **36**, 762-71, (2000).

11.   S. T. Cole, P. Supply, N. Honoré, *Lepr. Rev.* **72**, 449-461 (2001).

12.   E. Mazars *et al.*, *Proc. Natl. Acad. Sci. USA.* **13**, 1901-6. (2001).

13.   Y. C. Shin, H. Lee, G. P. Walsh, J. D. Kim, S. N. Cho, *J. Clin. Microbiol.* **38**, 4535-8. (2000).

14.   R. W. Truman, A. B. Fontes, A. B. de Miranda, P. Suffys, T. Gillis, *J. Clin. Microbiol.* **42**, 2558-2565 (2004).

15.   N. A. Groathouse *et al.*, *J. Clin. Microbiol.* **42**, 1666-1672 (2004).

16.   S. K. Young *et al.*, *J. Clin. Microbiol.* **42**, 4931-6. (2004).

17.   M. Matsuoka *et al.*, *Int. J. Lepr. Other Mycobact. Dis.* **68**, 121-8 (2000).

18.   S. T. Lovett, *Molec. Microbiol.* **52**, 1243-1253 (2004).

19.   K. Eiglmeier, S. Simon, T. Garnier, S. T. Cole, *Lepr. Rev.* **72**, 462-469 (2001).

20.   K. Eiglmeier, N. Honoré, S. A. Woods, B. Caudron, S. T. Cole, *Molec. Microbiol.* **7**, 197-206 (1993).

21.   R. D. Fleischmann *et al.*, *J. Bacteriol.* **184**, 5479-90. (2002).

22.   T. Garnier *et al.*, *Proc. Natl. Acad. Sci. USA* **100**, 7877-82 (2003).

23.   C. Kidgell *et al.*, *Infect. Genet. Evol.* **2**, 39-45 (2002).

24.   D. Falush *et al.*, *Science* **299**, 1582-5 (2003).

25.   J. Maynard-Smith, N. H. Smith, M. O'Rourke, B. G. Spratt, *Proc. Natl. Acad. Sci. USA.* **90**, 4384–8 (1993).

26.   S. G. Browne, in *Leprosy* R. C. Hastings, Ed. (Churchill Livingstone, Edinburgh, 1985).

27.   P. A. Underhill *et al.*, *Nat. Genet.* **26**, 358–361 (2000).

28.   L. L. Cavalli-Sforza, M. W. Feldman, *Nat. Genet. Supp.* **33**, 266–275 (2003).

29.   L. Marsollier *et al.*, *Appl. Environ. Microbiol.* **68**, 4623-8. (2002).

30.   T. Wirth *et al.*, *Proc. Natl. Acad. Sci. USA.* **101**, 4746-51 (2004).

Supporting Online Material

www.sciencemag.org

Materials and Methods

Figs. S1, S2

Tables S1, S2, S3

Legends to figures.

Figure 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, while the faint arrow denotes an alternative route.

Figure 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 Figure 1B. The colored arrows indicate the direction of human migrations predicted by, or inferred from, our SNP analysis; hatched black 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*).

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 in the Tamil Nadu strain (*11*) while those in the Table are the copy numbers found in the respective isolates. IP, Institut Pasteur; NHDP, National Hansen's Disease Program; CSU, Colorado State University.

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	
USA			3	
WARM*			4	
Total	67	14	38	56

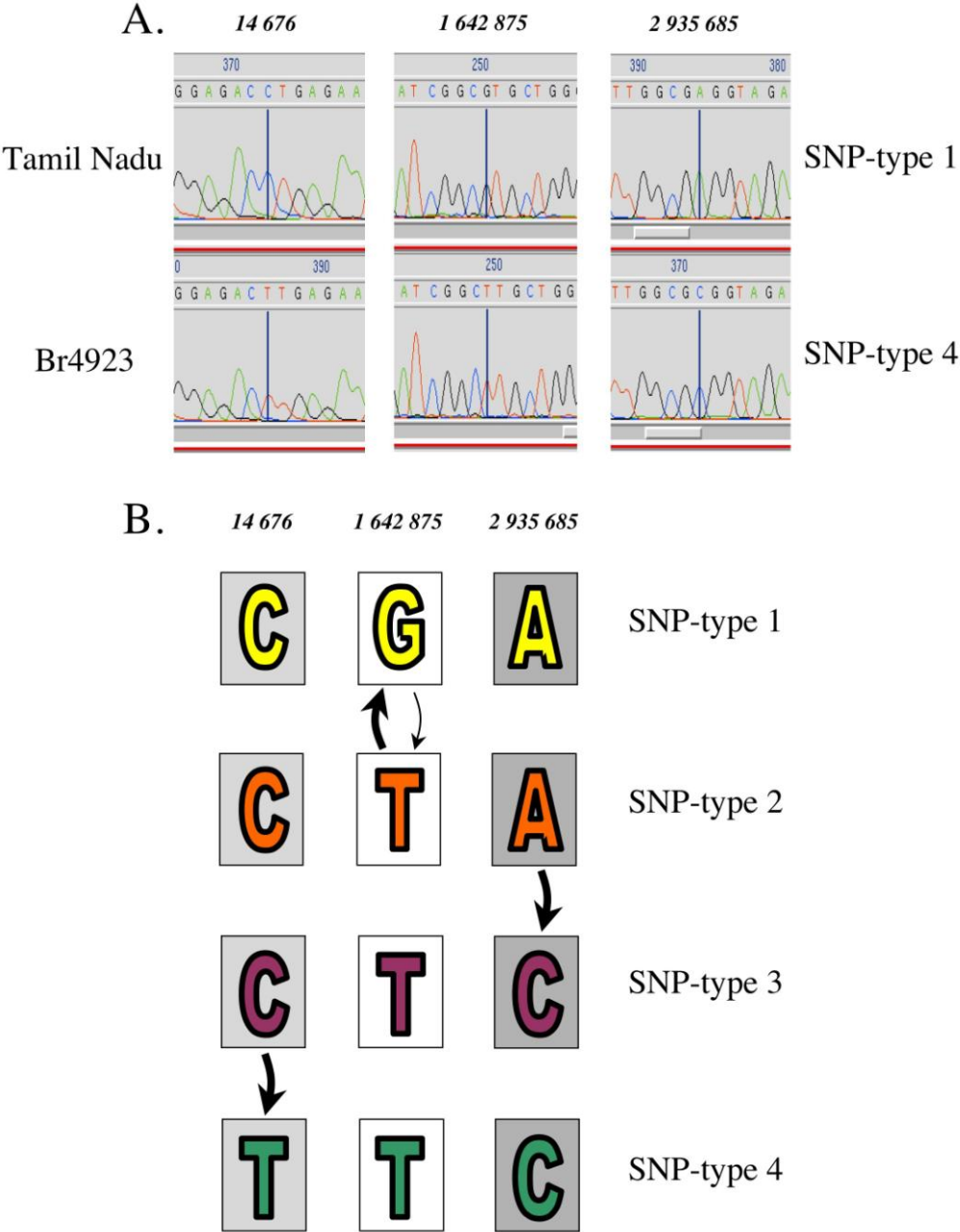*, Wild armadillos from Louisiana, USA

Figure 1.

Figure 2.