Insights from the Genome Sequence of *Mycobacterium lepraemurium*: Massive Gene Decay and Reductive Evolution

Andrei Benjak, Tanvi P. Honap, Charlotte Avanzi, Enrique Becerril-Villanueva, Iris Estrada-García, Oscar Rojas-Espinosa, Anne C. Stone, and Stewart T. Cole

Global Health Institute, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland; School of Life Sciences, Arizona State University, Tempe, Arizona, USA; Departamento de Psicoinmunología, Instituto Nacional de Psiquiatría Ramón de la Fuente, Mexico City, Mexico; Departamento de Inmunología, Escuela Nacional de Ciencias Biológicas, Instituto Politécnico Nacional, Mexico City, Mexico; Center for Evolution and Medicine, Arizona State University, Tempe, Arizona, USA

**ABSTRACT** *Mycobacterium lepraemurium* is the causative agent of murine leprosy, a chronic, granulomatous disease similar to human leprosy. Due to the similar clinical manifestations of human and murine leprosy and the difficulty of growing both bacilli axenically, *Mycobacterium leprae* and *M. lepraemurium* were once thought to be closely related, although it was later suggested that *M. lepraemurium* might be related to *Mycobacterium avium*. In this study, the complete genome of *M. lepraemurium* was sequenced using a combination of PacBio and Illumina sequencing. Phylogenomic analyses confirmed that *M. lepraemurium* is a distinct species within the *M. avium* complex (MAC). The *M. lepraemurium* genome is 4.05 Mb in length, which is considerably smaller than other MAC genomes, and it comprises 2,682 functional genes and 1,139 pseudogenes, which indicates that *M. lepraemurium* has undergone genome reduction. An error-prone repair homologue of the DNA polymerase III α-subunit was found to be nonfunctional in *M. lepraemurium*, which might contribute to pseudogene formation due to the accumulation of mutations in nonessential genes. *M. lepraemurium* has retained the functionality of several genes thought to influence virulence among members of the MAC.

**IMPORTANCE** *Mycobacterium lepraemurium* seems to be evolving toward a minimal set of genes required for an obligatory intracellular lifestyle within its host, a niche seldom adopted by most mycobacteria, as they are free-living. *M. lepraemurium* could be used as a model to elucidate functions of genes shared with other members of the MAC. Its reduced gene set can be exploited for studying the essentiality of genes in related pathogenic species, which might lead to discovery of common virulence factors or clarify host-pathogen interactions. *M. lepraemurium* can be cultivated in vitro only under specific conditions and even then with difficulty. Elucidating the metabolic (in)capabilities of *M. lepraemurium* will help develop suitable axenic media and facilitate genetic studies.

**KEYWORDS** *Mycobacterium lepraemurium*, comparative genomics, genome sequencing, murine leprosy

**M**urine leprosy is a chronic, granulomatous disease caused by *Mycobacterium lepraemurium*. Murine leprosy mainly affects the skin, mucosa of the upper respiratory tract, and eyes; however, unlike human leprosy, the viscera are commonly affected, whereas the peripheral nerves are not (1). Murine leprosy was first reported in the early 20th century in rats in Ukraine (2), after which similar cases were reported from...
other countries (3, 4). M. lepraemurium is one of the causative agents of leprosy in cats and causes granulomatous skin lesions that often involve ulceration (5).

In humans, leprosy is primarily caused by Mycobacterium leprae and Mycobacterium lepromatosis. Numerous similarities exist between human and murine leprosy, including disease transmission through abrasions in the skin and the mucosal respiratory surfaces, a similar spectrum of disease progression, suppression of cell-mediated immunity, and strong humoral immunity (1). This led to the hypothesis that these species were closely related, and hence, it was thought that murine leprosy might serve as a model for human leprosy (6–8). DNA hybridization studies and analysis of the 16S rRNA gene sequences suggested that M. lepraemurium might actually be more related to the Mycobacterium avium complex (MAC) (9, 10). However, to date, no phylogenomic study of M. lepraemurium has been conducted, and the lack of a genome sequence has restricted our understanding of its biology and evolution.

Here, we describe the complete genome of M. lepraemurium, which was sequenced using single-molecule real-time (SMRT; Pacific Biosciences) and Illumina technologies. The M. lepraemurium genome was found to be circular, with a total GC content of 68.99%, and 4,050,523 bp in length. No plasmids were found. The genome comprises 3,821 "protein-coding" genes, of which 2,682 are functional genes and 1,139 are pseudogenes (Fig. 1A). M. lepraemurium belongs to the MAC and is more closely related to the M. avium clade than to the M. intracellularre clade (Fig. 1B and see Fig. S1 and S2 in the supplemental material).

Genome downsizing and pseudogene formation. At 4.05 Mb, the M. lepraemurium genome is the smallest within the MAC and one of the smallest among mycobacteria (9th out of the 350 sequenced mycobacterial species). More strikingly, within the mycobacteria, only the M. leprae and M. lepromatosis genomes contain fewer functional protein-coding genes than the M. lepraemurium genome. The presence of 1,139 pseudogenes indicates that M. lepraemurium underwent reductive evolution, which is characteristic for strictly host-associated organisms. Analyses of pseudogene...
families within a diverse set of prokaryotes have shown that pseudogenes are most likely to occur in ABC transporter, short-chain dehydrogenase/reductase, sugar transporter, cytochrome P450, and proline-glutamate (PE)/proline-proline-glutamate (PPE) gene families (11). The *M. lepraemurium* genome was found to contain pseudogenes in all these families.

*M. lepraemurium* is the third mycobacterial species known to have undergone reductive evolution. The other species include the common ancestor of *M. leprae* and the closely related *M. lepromatosis*, which underwent genome reduction before the two species diverged, as well as *Mycobacterium ulcerans*, which is in a state of intermediate reductive genome evolution (12–14). Genome size among *M. avium* subspecies varies considerably (around 4.8 to 5.5 Mb), but no extensive pseudogenization was observed in these organisms. Curiously, *M. lepraemurium* seems to be evolving convergently toward a minimal gene set such as the one retained by *M. leprae* (Fig. 2). This may be due to both species having adapted to a similar niche, consequently resulting in similar pathological manifestations.

Loss of the DnaQ-mediated proofreading mechanism of DNA polymerase III \( \alpha \)-subunit has been hypothesized as the cause of pseudogene formation in *M. leprae* (12), and a similar pattern has been observed in other obligate pathogens and symbionts (15, 16). In *M. lepraemurium*, a homologue of the error-prone repair DNA polymerase III \( \alpha \)-subunit (MLM_3495) is nonfunctional, which might contribute to a higher error rate leading to pseudogene formation in this species.

The AT content of the pathogen/symbiont genome seems to correlate with the age of the strict association with the host (16). As expected, the genome of *M. leprae* has the highest AT content of all mycobacteria (42.2%), especially in pseudogenes (43.5%). On the other hand, the genome of *M. lepraemurium* is as AT-poor as that of its closest relative, *M. avium* (31%), and shows only a small difference in the AT content of functional genes and pseudogenes (30.46% and 31.05%, respectively). This suggests

---

**FIG 2** Heatmap of the gene orthology between *M. avium* subsp. *paratuberculosis* K-10, *M. lepraemurium*, *M. leprae*, and *M. ulcerans*. Genes are shown in black, and the absence of genes is shown in white. The raw height corresponds to the number of genes.
that M. lepraemurium adopted the “strictly intracellular” niche relatively recently, as evidenced by its phylogenetic relatedness to the free-living M. avium (Fig. S1). Genomic data for other M. lepraemurium strains would allow us to determine whether the pseudogene content varies between strains, which would indicate ongoing reductive evolution.

**Repetitive sequences and mobile elements.** Repeats and mobile elements are the major vehicles leading to genomic rearrangements and large deletions, and hence, they are thought to play an important role in genome downsizing. Yet, there seems to be a negative correlation between the stage of reductive evolution and repeat content (17). This is likely because deletions and pseudogenization impact fitness as more genes become nonfunctional, and the remaining genes become indispensable. Therefore, it has been suggested that repeats might be involved in the early phase of reductive evolution and that a different mechanism is responsible for the gradual deletion of pseudogenes (17). M. lepraemurium has only a few repeats (Table S1), which mostly derive from three families of dysfunctional insertion sequences (IS), each consisting of up to six copies.

Curiously, one IS family consisting of six copies (MLM_1414, MLM_1873, MLM_2691, MLM_2913, MLM_3078, and MLM_3876) shows 84% nucleotide identity with the ISMsm2 mobile element from Mycobacterium smegmatis (NCBI accession number WP_003887303). A BLAST search showed that this IS family resides in only a few other unrelated mycobacteria and not in the MAC, with the exception of a more diverged copy (with 78% identity) present in the plasmid of M. avium subsp. hominisuis strain 88Br (GenBank accession number KR997898.1) and a truncated copy (with 81% identity) in the genome of M. avium subsp. hominisuis strain H87 (GenBank accession number CP018363.1). Strikingly, when we analyzed the genomic synteny between M. lepraemurium, M. intracellulare, and M. avium subsp. hominisuis, all the synteny breaks in M. lepraemurium occurred at the locations of this particular IS (Fig. S3A and B). The synteny plot between M. lepraemurium and M. avium subsp. paratuberculosis looked different from the aforementioned genomes, which is probably due to genomic rearrangements in M. avium subsp. paratuberculosis (Fig. S3C).

The presence of identical copies of dysfunctional transposable elements indicates that M. lepraemurium lost functional transposases relatively recently. As mentioned above, at least one IS family was responsible for large genomic rearrangements in M. lepraemurium, and a deeper analysis of the genome sequence might shed light on the driving force that caused major deletions in this genome.

**M. lepraemurium-specific sequences.** Comparison of M. lepraemurium with other members of the MAC revealed a total of 32.2 kb of genomic sequence that was not found in any publicly available genome sequence. These regions contain 35 genes, of which only 12 are functional (Table S2). One of these genes, MLM_3300, codes for a Fic family protein. Fic proteins have been associated with pathogenicity in bacteria, often acting as toxins that interfere with the host cell in different ways (18).

**Interaction with macrophages.** After entering the macrophage, members of the MAC reside within phagosomes (19). There they inhibit phagosome maturation by preventing acidification to a pH below 6.4, and thus, prevent fusion with the extremely acidic lysosome. In M. avium, mutations in the PPE gene MAV_2928 and the PE gene MAV_1346 inhibit maturation and acidification of phagosomes, resulting in decreased virulence (20, 21). In M. lepraemurium, gene MLM_2357 (homologous to MAV_2928) and MLM_1265 (homologous to MAV_1346) are functional, suggesting that they may help its survival in macrophages. Additionally, the gene MLM_2012 encodes the lipoprotein, LppM, which is an important virulence factor in M. tuberculosis and interferes with phagosomal maturation in macrophages (22).

Reactive oxygen species (ROS) are produced by the macrophage as a bactericidal mechanism in response to infection. The ability of pathogens to produce enzymes such as catalase-peroxidase, epoxide hydrolase, and superoxide dismutase (SOD), which remove ROS, enable their survival within macrophages. While it has been suggested
that *M. lepraemurium* abolishes the production of ROS upon entering the macrophage (29), probably as a survival mechanism, it remains unclear whether small quantities of ROS are produced upon infection and handled by the bacterium. *M. lepraemurium* has retained four functional epoxide hydrolases (*MLM_0642, MLM_0684, MLM_1194, and MLM_1485*) and one catalase-peroxidase (*MLM_2092*), which explains its observed catalase-peroxidase activity (23). Moreover, *M. lepraemurium* is able to produce two superoxide dismutases, SodA and SodC (*MLM_0123 and MLM_2650*).

**Glycopeptidolipid synthesis.** Glycopeptidolipids (GPLs) are synthesized by several nontuberculosis mycobacteria, including members of the MAC. GPLs are present on the outermost layer of the cell wall, and therefore, they play an important role in sliding motility, biofilm formation, and pathogenesis (24). This is even more relevant within the MAC, whose members can produce a number of serovar-specific variations of GPLs. The *M. lepraemurium* cell wall contains glycolipids and amino acids (25, 26); however, production of GPLs in *M. lepraemurium* has not been studied. Most of the genes known to be involved in the biosynthesis of non-serovar-specific GPLs are intact in *M. lepraemurium*, albeit with a few exceptions (Table S3). Although it is difficult at this moment to predict the effects of these mutations, it appears that *M. lepraemurium* should be able to produce GPLs, as all of the core genes are present.

**Virulence.** While the ESX-1 system is the main determinant of virulence in *M. tuberculosis* and *M. leprae* (27), it is absent in the MAC. However, ESX-5, which is also associated with virulence in pathogenic mycobacteria, is mostly intact in *M. lepraemurium*, except for the cytochrome P450 hydroxylase (*MLM_2361*), which is also nonfunctional in *M. leprae*. The duplicated four-gene region, ESX-5a, encoding EssI, EssJ, a PPE, and a PE protein may serve as an accessory system for transport of a subset of ESX-5 proteins (28) and is still functional in *M. lepraemurium*. In *M. lepraemurium*, the PPE and PE genes are merged into a single open reading frame, as in *M. avium* subsp. *paratuberculosis* but not in other members of the MAC. An interesting observation is that in MAC, essI and essJ are 100% identical to essN and essM from the main ESX-5 locus (unlike in *M. tuberculosis*), which indicates a novel and recent duplication event and suggests that a crucial function might lie behind redundancy of the ESX-5 components.

The *pks12* gene is involved in the biosynthesis of mannosyl-β-1-phosphomycoketides (MPM) and is found only in the slow-growing mycobacteria. In different pathogenic mycobacteria, including *M. avium* (20), *pks12* was shown to be necessary for the virulence, and this gene is functional in *M. lepraemurium* (*MLM_2156*).

**Experimental procedures.** *M. lepraemurium* strain Hawaii was grown in BALB/c mice. The DNA was sequenced using Illumina and PacBio technologies, followed by sequence assembly and annotation. More details are given in Text S1 in the supplemental material.

**Accession number(s).** The annotated genome was submitted to GenBank under GenBank accession number CP021238.

**SUPPLEMENTAL MATERIAL**

Supplemental material for this article may be found at https://doi.org/10.1128/mBio.01283-17.

**TEXT S1,** DOCX file, 0.03 MB.
**FIG S1,** EPS file, 0.2 MB.
**FIG S2,** EPS file, 0.2 MB.
**FIG S3,** TIF file, 2.8 MB.
**TABLE S1,** DOCX file, 0.01 MB.
**TABLE S2,** DOCX file, 0.01 MB.
**TABLE S3,** DOCX file, 0.02 MB.

**ACKNOWLEDGMENTS**

We thank the Genomic Technologies Facility at the University of Lausanne, Switzerland, for Illumina and PacBio sequencing. We thank Philippe Busso (Ecole Polytech-
REFERENCES

1. Rojas-Espinosa O. 2009. Murine leprosy revisited, p 97–140. In Tomioka H (ed). Current topics on the profiles of host immunological response to mycobacterial infections, 1st ed. Research Signpost, Trivandrum, Kerala, India.

2. Stefansky WK. 1902. Zabolevanija u krys, vyzvannyj kislotoupornoj palosjokj. Rus Vratj 47:1726–1727.

3. Marchoux E, Sorel F. 1912. Recherches sur la lèpre: la lèpre des rats (Lepra murium). Ann Inst Pasteur 56:778–801.

4. Dean G. 1903. A disease of the rat caused by an acid-fast bacillus. Cent F Bakteriol 34:222–244.

5. O’Brien CR, Malik R, Gliban M, Reppas G, McCowan C, Fyfe JA. 2017. Mycobacterium avium-PPE gene is associated with the ability of the bacterium to grow in macrophages and virulence in mice. Cell Microbiol 19:737–746. https://doi.org/10.1111/j.1462-5822.2017.01776.x

6. Walker EL, Sweeney MA. 1929. The identity of human leprosy and rat leprosy. J Prev Med 3:325–333.

7. Schmitt LS. 1911. On the relation between rat and human leprosy. Pathologie 2:29–37.

8. Dean G. 1905. Further observations on a leprosy-like disease of the rat. J Hyg 5:99–112. https://doi.org/10.1017/S0022172400002370.

9. Athwal RS, Deo SS, Imaeda T. 1984. Deoxyribonucleic acid relatedness among Mycobacterium leprae, Mycobacterium lepraemurium, and selected bacteria by dot blot and spectrophotometric deoxyribonucleic acid hybridization assays. Int J Syst Bacteriol 34:371–375. https://doi.org/10.1099/00221724-34-4-371.

10. Hughes MS, Ball NW, Beck LA, De Lisle GW, Skuce RA, Neill SD. 1997. Determination of the etiology of presumptive feline leprosy by 16S rRNA gene analysis. J Clin Microbiol 35:2464–2471.

11. Liu Y, Harrison PM, Kunin V, Gerstein M. 2004. Comprehensive analysis of pseudogenes in prokaryotes: widespread gene decay and failure of putative horizontally transferred genes. Genome Biol 5:R64. https://doi.org/10.1186/gb-2004-5-9-r64.

12. Cole ST, Eiglemeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, Honoré N, Garnier T, Churcher C, Harris D, Mungall K, Basham D, Brown D, Chillingworth T, Connor R, Davies RM, Devlin K, Duthey S, Feltwell T, Fraser A, Hamlin N, Holroyd S, Hornsby T, Jagels K, Lacroix C, Maclean J, Moule S, Murphy L, Oliver K, Quail MA, Rajandream M-A, Rutherford KM, Rutter S, Seeger K, Simon S, Simmonds M, Skelton J, Squires R, Stevens S, Stevens K. 2001. Massive gene decay in the leprosy bacillus. Nature 409:1007–1011. https://doi.org/10.1038/35095006.

13. Singh P, Benjak A, Schuennemann VJ, Herbig A, Avanzi C, Busso P, Nieselt K, Krause J, Vera-Cabrera L, Cole ST. 2015. Insight into the evolution and origin of leprosy bacilli from the genome sequence of Mycobacterium lepromatosis. Proc Natl Acad Sci U S A 112:4459–4464. https://doi.org/10.1073/pnas.1421504112.

14. Stinear TP, Seemann T, Pitot S, Frigui W, Reyssset G, Garnier T, Meurice G, Simon D, Boucher C, Ma L, Tichit M, Porter JL, Ryan J, Johnson PDR, Davies JK, Jenkin GA, Small PLC, Jones LM, Tekala F, Laval F, Daffe M, Parkhill J, Cole ST. 2007. Reductive evolution and niche adaptation inferred from the genome of Mycobacterium ulcerans, the causative agent of Buruli ulcer. Genome Res 17:192–200. https://doi.org/10.1101/gr.5942807.

15. Dale C, Wang B, Moran N, Ochman H. 2003. Loss of DNA recombinational repair enzymes in the initial stages of DNA degeneration. Mol Biol Evol 20:1188–1194. https://doi.org/10.1093/molbev/msg138.

16. Moya A, Peretó J, Gil R, Latorre A. 2008. Learning how to live together: genomic insights into prokaryote-animal symbioses. Nat Rev Genet 9:218–229. https://doi.org/10.1038/nrg2319.

17. Frank AC, Amiri H, Andersson SGE. 2002. Genome deterioration: loss of repeated sequences and accumulation of junk DNA. Genetica 115:1–12. https://doi.org/10.1023/A:1016064511533.

18. Roy CR, Cherflis J. 2015. Structure and function of Fic proteins. Nat Rev Microbiol 13:631–640. https://doi.org/10.1038/nrmicro3520.

19. Ignatov D, Kondratieva E, Azhikina T, Apt A. 2012. Mycobacterium avium-triggered diseases: pathogenomics. Cell Microbiol 14:808–818. https://doi.org/10.1111/j.1462-5822.2012.01776.x.

20. Li Y, Danelishvili L, Wagner D, Petrofsky M, Bermudez LE. 2010. Identification of virulence determinants of Mycobacterium avium that impact on the ability to resist host killing mechanisms. J Med Microbiol 59:8–16. https://doi.org/10.1099/jmm.0.012864-0.

21. Li Y, Miltner E, Wu M, Petrofsky M, Bermudez LE. 2005. A Mycobacterium avium PPE gene is associated with the ability of the bacterium to grow in macrophages and virulence in mice. Cell Microbiol 7:539–548. https://doi.org/10.1111/j.1462-5822.2004.00484.x.

22. Deboosère N, Iantomasi R, Queval CJ, Song OR, Deloison G, Jouny S, Tomioka In Japan, 1989. Proc Natl Acad Sci U S A 86:5964–5968. https://doi.org/10.1073/pnas.86.13.5964.

23. Deboosère N, Iantomasi R, Queval CJ, Song OR, Deloison G, Jouny S, Tomioka In Japan, 1989. Proc Natl Acad Sci U S A 86:5964–5968. https://doi.org/10.1073/pnas.86.13.5964.

24. Pang L, Tian X, Pan W, Xie J. 2013. Catalase-peroxidases and superoxide dismutases in Mycobacterium leprae and other mycobacteria studied by crossed immunoelectrophoresis and polyacrylamide gel electrophoresis. Infect Immun 54:666–672.

25. Frank AC, Amiri H, Andersson SGE. 2002. Genome deterioration: loss of repeated sequences and accumulation of junk DNA. Genetica 115:1–12. https://doi.org/10.1023/A:1016064511533.

26. Luna-Herrera J, Rojas-Espinosa O, Estrada-Parra S. 1994. Recognition of Mycobacterium lepraemurium by sera of mice infected with Mycobacterium lepraemurium. Int J Lepr Other Mycobact Dis 66:365–373.

27. Gröschel MI, Sayes F, Simeone R, Majlessi L, Brosch R. 2016. ESX secretion systems: mycobacterial evolution to counter host immunity. Nat Rev Microbiol 14:677–691. https://doi.org/10.1038/nrmicro3520.

28. Gröschel MI, Sayes F, Simeone R, Majlessi L, Brosch R. 2016. ESX secretion systems: mycobacterial evolution to counter host immunity. Nat Rev Microbiol 14:677–691. https://doi.org/10.1038/nrmicro3520.

29. Rojas-Espinosa O, Estrada-Parra S. 1996. Recognition of lipid antigens by sera of mice infected with Mycobacterium lepraemurium. Int J Lepr Other Mycobact Dis 64:299–305.

30. Gröschel MI, Sayes F, Simeone R, Majlessi L, Brosch R. 2016. ESX secretion systems: mycobacterial evolution to counter host immunity. Nat Rev Microbiol 14:677–691. https://doi.org/10.1038/nrmicro3520.

31. Rojas-Espinosa O, Camarena-Servin V, Estrada-Garcia I, Arez-Bedido P, Wek-Rodriguez K. 1998. Mycobacterium lepraeum, a well-adapted parasite of macrophages: I. Oxygen metabolites. Int J Lepr Other Mycobact Dis 66:365–373.