

ARTICLE

DOI: 10.1038/s41467-017-02576-z

OPEN

# Phylogenomics and antimicrobial resistance of the leprosy bacillus *Mycobacterium leprae*

Andrej Benjak et al. 

Leprosy is a chronic human disease caused by the yet-uncultured pathogen *Mycobacterium leprae*. Although readily curable with multidrug therapy (MDT), over 200,000 new cases are still reported annually. Here, we obtain *M. leprae* genome sequences from DNA extracted directly from patients' skin biopsies using a customized protocol. Comparative and phylogenetic analysis of 154 genomes from 25 countries provides insight into evolution and antimicrobial resistance, uncovering lineages and phylogeographic trends, with the most ancestral strains linked to the Far East. In addition to known MDT-resistance mutations, we detect other mutations associated with antibiotic resistance, and retrace a potential stepwise emergence of extensive drug resistance in the pre-MDT era. Some of the previously undescribed mutations occur in genes that are apparently subject to positive selection, and two of these (*ribD*, *fadD9*) are restricted to drug-resistant strains. Finally, nonsense mutations in the *nth* excision repair gene are associated with greater sequence diversity and drug resistance.

. Correspondence and requests for materials should be addressed to S.T.C. (email: [stewart.cole@epfl.ch](mailto:stewart.cole@epfl.ch)). #A full list of authors and their affiliations appears at the end of the paper.

**M***ycobacterium leprae* is the main causative agent of leprosy, a disease that affects the skin, nerves, and mucosa of the upper respiratory tract in humans<sup>1</sup>. A second, distantly related leprosy bacillus, *Mycobacterium lepromatosis*, was recently discovered in humans and red squirrels (*Sciurus vulgaris*)<sup>2</sup>. Leprosy is curable with multidrug therapy (MDT), but remains a public health problem in South America, Africa, South and Southeast Asia, and Micronesia, where over 200,000 new leprosy cases are reported each year<sup>3</sup>. MDT, comprising rifampicin, dapsone, and clofazimine, has been used intensively since the 1980s and a few second-line drugs, ofloxacin, minocycline, and clarithromycin, are sometimes employed as therapeutic agents<sup>4</sup>. The emergence of drug-resistant (DR) and multidrug-resistant (MDR) *M. leprae* is increasingly reported<sup>5–12</sup>. For dapsone, rifampicin and ofloxacin, the resistance mechanism has been attributed to missense mutations in the drug resistance determining regions (DRDR) of the *folP1*, *rpoB*, and *gyrA* genes, respectively.

*M. leprae* is an obligate intracellular pathogen that has never been cultured axenically but can infect wild or experimental animals. The nine-banded armadillo (*Dasypus novemcinctus*) or the mouse footpad (MFP) can be used to produce bacilli, but both methods are cumbersome and time-consuming<sup>13</sup>. The genome of *M. leprae* is the smallest among mycobacteria (3.3 Mb) with 1614 genes encoding proteins and a remarkable 1300 pseudogenes<sup>14</sup>. Such reductive evolution is a hallmark of bacteria that have changed their lifestyle from free-living to strictly host-associated<sup>15</sup>. Due to its 14-day generation time and the absence of horizontal gene transfer, the genome of *M. leprae* is highly conserved, with <300 single-nucleotide polymorphisms (SNPs) observed between distantly related strains, and only a few SNPs between close relatives<sup>5,16–18</sup>. Four SNP types (branches 1–4) and 16 SNP subtypes (A–P) were defined by surveying 78 informative SNPs and six single-base insertion/deletions (InDels)<sup>16,19</sup>. Genotyping a large panel of *M. leprae* strains revealed strong geographical associations and suggested possible routes of dissemination of leprosy<sup>16</sup> whereas, a recent phylogenetic analysis of 16 whole-genome sequences of modern and ancient *M. leprae* strains, implicated the 3K subtype (branch 0) as the most ancestral<sup>17</sup>.

Leprosy seems to have appeared during the Iron Age (1200–600 BC) and the date of the most recent common ancestor of *M. leprae* was estimated to be from 2543 BC to 36 AD, based on whole-genome sequence analysis<sup>17</sup>. Similarly, the earliest accepted written record of leprosy is from 600 BC<sup>20</sup>, and the earliest osteological evidence dates from around 300 BC<sup>21–25</sup>. The oldest genomic evidence of leprosy is for samples from 80 to 240 AD in Central Asia<sup>26</sup>.

In this study, we develop and apply methods to isolate and purify *M. leprae* DNA that enable whole genome sequences to be obtained directly from human biopsy material, thus removing the necessity for passage through animals. This approach was successfully used to generate 120 new *M. leprae* genome sequences from drug-susceptible and DR strains from around the world, thereby enabling detailed phylogenetic and phylogeographic comparisons to be performed, new mutations associated with antimicrobial resistance to be detected, and the likely origin of leprosy to be proposed.

## Results

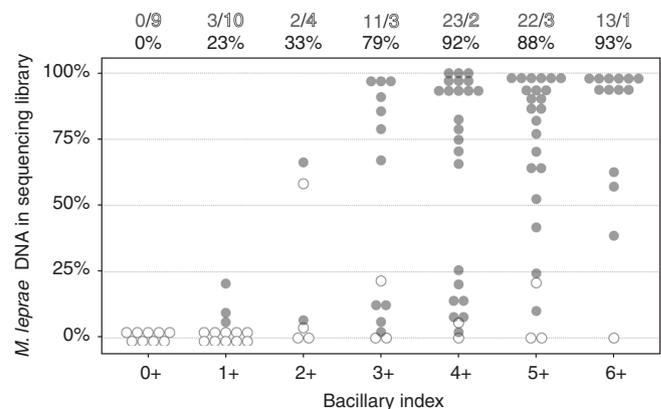
**Isolating *M. leprae* DNA from human skin biopsies.** Genome sequencing has become routine practice in microbiology<sup>27</sup>, especially for micro-organisms that can be readily isolated, which is not the case of the leprosy bacillus. For decades, the sole source of *M. leprae* DNA suitable for genomics was from bacteria

isolated 12 months after infection of armadillos or mice. Recently, we have developed and optimized methods that enable *M. leprae* DNA to be extracted directly from fresh or formalin-fixed skin biopsies from leprosy patients<sup>28</sup>. These methods include enrichment of *M. leprae* DNA by array capture<sup>17</sup> but this is less practical for large population-based investigations.

The DNA extraction method used in this study was applied directly to punch biopsies from clinically well-characterized patients of known bacillary index (BI) and exploits the fact that *M. leprae* resides intracellularly. Host cells are first disrupted and their DNA degraded, leaving the bacilli intact. The bacilli are then lysed and their DNA extracted and used for library preparation. This approach was applied to 106 biopsies whose BI ranged from 0 (no bacilli visible) to 6 (>1000 bacilli per microscopic field) thereby enabling a relationship between BI and sequencing efficiency to be established (Fig. 1). As expected, there was a direct correlation between genome coverage and the BI but, surprisingly, successful coverage could even be achieved with some specimens whose BI was as low as 1+.

**Genome analysis of patient and animal cohort.** We analyzed a total of 154 *M. leprae* genomes from 25 countries (Fig. 2, Supplementary Data 1), of which 120 were newly sequenced and 34 were previously published (Supplementary Data 1). The cohort comprised 147 human samples, 6 from red squirrels and 1 from an armadillo that were all naturally infected. Genome sequences were obtained directly from 109 human samples, 30 from bacilli passed in mice, and 8 from armadillos. Thirty of these strains were from patients who had relapsed or not responded to MDT the remainder (124) were from supposedly drug-susceptible strains (87 were from confirmed primary cases, while disease history was unknown for the others).

A total of 3053 SNPs and 219 InDels (excluding tandem repeats) was found (Supplementary Data 2). The average SNP difference among the 154 genomes was 114. We found a total of 988 non-synonymous alleles (0.62 per protein-coding gene, or 0.61 per kb of protein-coding genes) and 530 synonymous SNPs (0.33 per protein-coding gene or 0.33 per kb of protein-coding genes), and 1763 mutations in intergenic regions and pseudogenes (1.07 mutations per kb of intergenic regions and pseudogenes). The SNP density for each gene is given in Supplementary Data 2. Of the 219 InDels, 58 (27%) were in protein-coding genes.



**Fig. 1** Correlation between bacillary index and successful sequencing. The content of *M. leprae* DNA in sequencing libraries derived from human skin biopsies was determined and found to be proportional to the bacillary index (not available for all samples). Empty circles are samples that were not included in the study due to insufficient genome coverage. Sample count and sequencing success rates are given at the top of each category



**Fig. 2** Geographic distribution of the *M. leprae* samples used in this study. World map shows the number of registered cases of leprosy per 10,000 population (prevalence rates) in 2015 as reported by the World Health Organization ([http://apps.who.int/neglected\\_diseases/ntddata/leprosy/leprosy.html](http://apps.who.int/neglected_diseases/ntddata/leprosy/leprosy.html)). Blue numbers indicate ancient *M. leprae* strains

**Phylogeny of *M. leprae*.** Phylogenetic analysis using both maximum parsimony (MP) and Bayesian inference resulted in consistent tree topologies and revealed distinct lineages and sublineages of *M. leprae* (Fig. 3). Strains belonging to the same SNP subtypes<sup>16</sup> clustered within single branches, with the exception of SNP subtype 3K, which is represented by a newly discovered ancestral lineage, termed here 3K-1, and the ancestral lineage referred to earlier as branch 0<sup>17</sup> and termed here 3K-0 (Fig. 3a). All strains from the two most ancestral lineages, 3K-0 and 3K-1, originated from Japan (8), China (1), Korea (1), the Marshall Islands (1), and New Caledonia (1), in agreement with earlier genotyping studies of hundreds of *M. leprae* strains which confirmed the predominance of the 3K genotype in East Asia, notably in Japan, China, and Korea<sup>16,29–31</sup>.

*M. leprae* in East Africa showed higher diversity with subtypes 2E, 2F, and 2H representing distinct lineages (Fig. 3a). The geographic distribution of those lineages corroborates the findings of earlier studies reporting the presence of SNP type 2 in Medieval Europe, the Middle East and East Africa<sup>16</sup>. Two Ethiopian isolates, belonging to the 2F subtype, clustered closely with medieval European strains dating from the 11th to 12th century (Fig. 3), which supports the hypothesis that the ancient Greek and Roman routes<sup>32</sup> connecting Europe, the Middle East, East Africa, and South Asia<sup>16,33–35</sup> contributed to the dissemination of SNP type 2 *M. leprae*.

West Africa, on the other hand, harbors exclusively SNP type 4, suggesting that overland migration between East and West Africa was limited. SNP subtypes 4N, 4O, and 4P, albeit sharing the same ancestor, do not form a monophyletic clade as previously hypothesized<sup>16</sup>. Rather, the 4O and 4P subtypes cluster together in a branch distinct from 4N (Fig. 3).

Brazil, as expected, contains a great diversity of several *M. leprae* lineages, with the SNP type 4 and SNP subtype 3I being the most prevalent<sup>36</sup>. The 3I genotype was common in medieval Europe<sup>17,21,37,38</sup>, and is still present in red squirrels in the United Kingdom<sup>2</sup>. The modern Brazilian strain Br2016-45 branched between two medieval strains from Europe (Fig. 3), making it the most ancestral contemporary 3I strain in the Americas to date. The broad diversity of 3I genotypes from Brazil probably derives from multiple introductions from Europe. On the other hand, the strains circulating in the Southern USA and associated with zoonosis from the nine-banded armadillo, I-30, NHDP-55 and NHDP-63<sup>18</sup>, originated much more recently (Fig. 3b), in agreement with the rapid expansion and spread of the armadillo population since its introduction to this region about 150 years ago<sup>39</sup>.

Good representation of most *M. leprae* lineages enabled identification of lineage-specific markers. A set of 235 SNPs and 25 InDels were specific to single lineages or groups of related lineages (Supplementary Data 2), of which 73 non-synonymous SNPs and 5 InDels were within protein-coding genes. These new lineage-specific markers can be used for future genotyping schemes.

**Dating analysis.** Dating analysis of *M. leprae* (Fig. 3b and Supplementary Figure 1) was done using BEAST v2.4.4<sup>40</sup> and the results were very similar to those obtained from a ten-fold smaller number of contemporary isolates<sup>17</sup>. The most recent common ancestor (TMRCA) of all *M. leprae* strains was estimated to be 3699 years old (95% Highest Posterior Density (HPD) 2731–4838 ya) and the substitution rate was  $7.8 \times 10^{-9}$  per site per year. Overlapping results were obtained when using different models (Supplementary Note and Supplementary Table 1), indicating that the dataset was robust and sufficiently informative.

A striking observation is the relative youth of the SNP type 1 lineage and its association with South Asia (Fig. 3b). Earlier studies revealed a predominance of SNP subtype 1D in India and Nepal, followed by 1C, 1A<sup>16</sup>, and 2E, 2G, and 2H<sup>33–35</sup>. SNP type 1 predominates in Thailand<sup>41</sup>, Bangladesh, Indonesia, and Philippines<sup>16</sup>. The current phylogeography of *M. leprae* implies that humans brought leprosy to South Asia from other parts of the continent.

**Hypermutated *M. leprae* strains.** Eight *M. leprae* strains (85054, S15, Amami, Zensho-4, Zensho-5, Zensho-9, Br14-3, and Br2016-15), belonging to five different SNP subtypes, had unusually long branches in the MP tree (Fig. 3a) because they contained on average 92 more SNPs than the other strains but approximately the same number of InDels. Comparative analysis revealed one unique feature linking the observed “hypermutated” strains, namely deleterious mutations in the endonuclease III gene *nth* (*ML2301*) due to frameshifts and premature stop codons (Table 1).

**Drug resistance.** DR-associated SNPs were detected in the DRDR in 24 strains for *folP1*, 11 strains for *rpoB*, and 2 strains for *gyrA* (in bold in Table 1). Previously described mutations were identified in *folP1* at codons 53 ( $n=7$ ) and 55 ( $n=17$ ), except in one isolate (Bn8-52), which had mutations at codons 55 and 145. Eleven strains had known mutations that confer rifampicin resistance in their *rpoB*-DRDR, while two strains (Kutatsu-6 and S15) harbor one additional mutation, and one (Br14-3) has two



**Fig. 3** Phylogeny of *M. leprae*. **a** Maximum parsimony tree of 154 genomes of *M. leprae*. The tree is drawn to scale, with branch lengths representing number of substitutions. *M. lepromatosis* was used as outgroup. Bootstrap values (500 replicates) are shown next to the branches. Dots indicate protein-changing mutations in the corresponding gene as given in Table 1. **b** Bayesian phylogenetic tree of 146 genomes of *M. leprae* calculated with BEAST 2.4.4. Hypermutated samples with mutations in the *nth* gene were excluded from the analysis. The tree is drawn to scale, with branch lengths representing years of age. Samples were binned according to geographic origin as given in the legend. Posterior probabilities for each node are shown in gray. Location probabilities of nodes were inferred by the Discrete Phylogeny model

**Table 1 Mutations in genes associated with drug resistance**

Sample	<i>folP1</i> ML0224	<i>rpoB</i> ML1891c	<i>gyrA</i> ML0006	<i>gyrB</i> ML0005	<i>fadD9</i> ML0484c	<i>ribD</i> ML1340	<i>ethA</i> ML0065	<i>pks4</i> ML1229	<i>nth</i> ML2301c
2188—2007	.	.	.	.	.	G62D <sup>1</sup>	.	.	.
2188—2014	.	.	.	.	.	G62D	.	.	.
85054	<b>P55L</b>	<b>S456L</b> G52E	.	.	W878*	D256N	.	.	R197*
2DDS	<b>P55R</b>	.	.	.	L396P	S58R	.	.	.
LRC-1A	.	.	I851T	V214G	.	.	.	.	.
Airaku-3	<b>T53I</b>	.	.	.	N304fs	.	.	.	.
Amami	<b>P55L</b>	.	.	.	.	R236C	.	.	L145fs
ARLP_08	.	H200Y	.	.	.	.	.	.	.
ARLP_10	.	.	.	.	Y927D	G61C	.	.	.
ARLP_30	.	.	G1115R	.	.	.	.	.	.
ARLP_52	.	.	G1115R	.	.	.	.	.	.
Bn8-46	<b>P55R</b>	.	.	.	.	.	.	.	.
Bn8-52	<b>P55LN145H</b>	.	.	.	W1108fs	.	.	.	.
Br14-1	<b>P55R</b>	<b>S456M</b>	.	.	G148fs	.	.	.	.
Br14-2	<b>P55R</b>	<b>S456M</b>	.	.	G148fs	.	.	.	.
Br14-3	<b>P55L</b>	<b>H451Y</b> G448D T433I	<b>A91V</b>	.	Q107*	G94D	.	.	E173*
Br14-4	<b>P55R</b>	<b>S456L</b> Y171N	.	.	L396P	I56T	.	.	.
Br14-5	<b>P55R</b>	<b>S456L</b>	.	.	A919E	C222W	.	.	.
Br2016-15	<b>P55L</b>	.	V731I	T503I	G796S	A63T	.	M14I	N142fs
Br2016-16	.	.	.	.	L998fs	K267fs	.	.	.
Br2016-18	.	.	S307L	.	.	.	.	.	.
Br2016-21	.	.	.	.	.	.	.	I932fs	.
Kusatsu-6	<b>P55L</b>	T433I <b>D441Y</b>	.	.	.	.	K477fs	.	.
MI10-93	<b>P55R</b>	.	.	.	R73fs	S58R	.	.	.
MI10-98	.	.	.	.	.	A10fs	.	.	.
MI2-10	<b>T53R</b>	.	.	.	.	.	.	.	.
MI6-50	<b>T53R</b>	.	.	.	.	.	.	.	.
MI6-55	<b>T53R</b>	.	.	.	.	.	.	.	.
Ng14-35	<b>P55R</b>	.	S307L	.	A594T	D34del	.	.	.
S15	<b>T53I</b>	G432S <b>H451D</b>	.	.	D466N	Q117*	D63N D323N	T334I	G146fs
S9	<b>T53I</b>	R791Q	.	.	.	S58N	.	.	.
Thai-237	.	.	I851T	.	.	.	.	.	.
US57	.	.	G362E	.	.	.	.	.	.
Zensho-2	<b>P55L</b>	.	.	.	Y562fs	G94D	.	.	.
Zensho-4	<b>T53I</b>	P51S <b>S456L</b>	<b>A91V</b>	D464N	R314C	D77N	A25T	Q1719*	L163fs
Zensho-5	<b>P55L</b>	<b>S456L</b> P583L	.	.	.	G204C	G390A	.	N142fs
Zensho-9	<b>P55L</b>	<b>H451Y</b> G681S	.	.	A973T	P150L	P383L	.	E122*

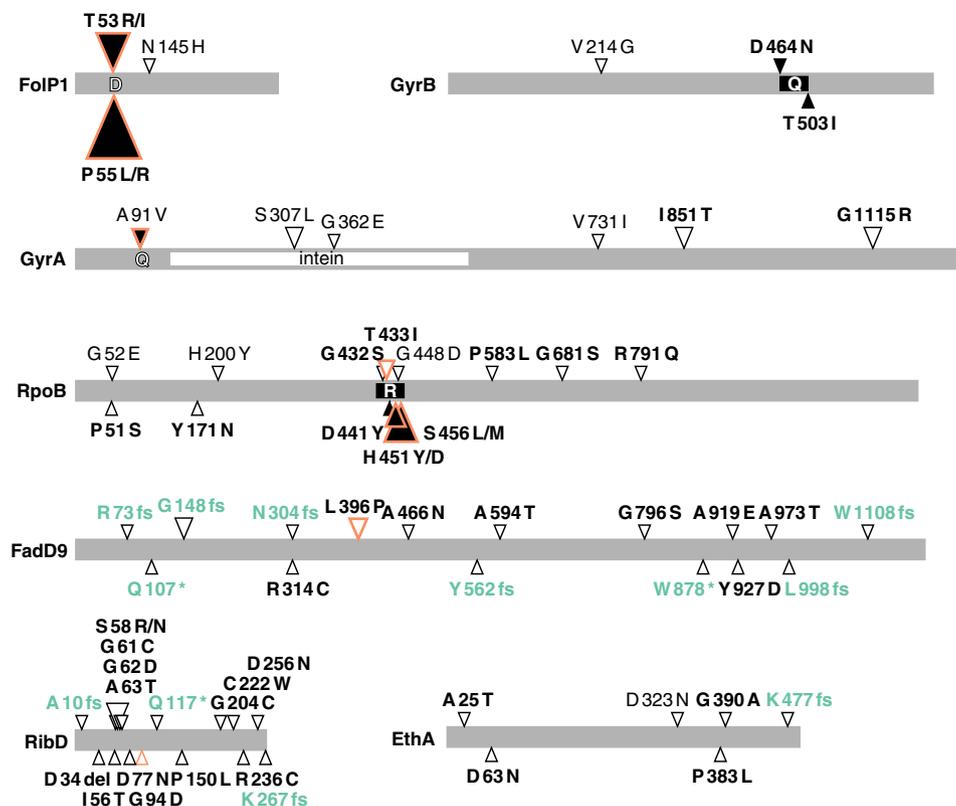
In bold, substitutions or residues known to confer drug resistance in *M. leprae*; \* premature stop codon, fs frameshift, dot (.) wild type. RpoB numbering is based on *M. leprae*, *E. coli* numbering in brackets: 51 (126), 52 (127), 171 (246), 200 (275), 432 (507), 433 (508), 441 (516), 448 (523), 451 (526), 456 (531), 681 (756), 791 (866). <sup>1</sup>Coverage below the threshold, both "2188" isolates come from the same patient but after an interval of 7 years<sup>28</sup>

additional mutations in the DRDR (Table 1). One of these additional mutations (G432S) does not confer rifampicin resistance to recombinant *Mycobacterium smegmatis*<sup>42</sup> whereas no information is available for the remaining three (T433I, G448D, and T508I) except that the G448A substitution does confer rifampicin resistance in *M. tuberculosis*<sup>43</sup>. Also, 5 of the 11 rifampicin-resistant strains had additional missense mutations in *rpoB* (85054, Br14-4, Zensho-4, Zensho-5, and Zensho-9) while 2 strains (ARLP\_08 and S9) presented non-synonymous SNPs outside the DRDR (Fig. 4). Compensatory mutations in *rpoA* and *rpoC*, encoding the alpha and beta-prime subunits of RNA polymerase, can occur in rifampicin-resistant *M. tuberculosis*<sup>44</sup>. We found one non-synonymous SNP in *rpoA*, substitution T187P in the rifampicin-resistant strain Br14-5, and seven non-synonymous SNPs in *rpoC* (Supplementary Data 2), of which

two occurred in the drug-resistant strains S15 (A258T) and Zensho-4 (H1133Y).

Two strains had known quinolone resistance mutations in the DRDR of *gyrA* and six harbored different single mutations elsewhere in the gene. Three isolates had a missense mutation in *gyrB*, including two within the DRDR (Table 1). Five strains harbor deleterious mutations in the *ethA* gene, encoding a mono-oxygenase that activates thioamide prodrugs in *M. tuberculosis*<sup>45,46</sup>. Interestingly, in addition to *ethA* and *nth*, three genes (*fadD9*, *ribD*, *pks4*) were mutated almost exclusively in MDR strains occurring 18, 19, and 4 times, respectively (Table 1).

**Retracing the emergence of drug resistance in leprosy patients.** Prior to the introduction of MDT in the 1980s, patients were treated with dapsone or other antimicrobials as monotherapies of



**Fig. 4** Mutations of *M. leprae* genes associated with antimicrobial resistance. Triangles point to the location of the mutation in the protein. Black triangles indicate known resistance-conferring mutations identified in this study that are situated in the drug resistance determining regions (DRDR): D dapsone, Q quinolone, R rifampicin. Orange border means the mutation was found to be homoplasic. Triangle size reflects the number of isolates from this study harboring the mutation, ranging from 1 to 17. Frameshifts and premature stop codons are in turquoise. Substitutions predicted to have an impact on the biological function of the protein<sup>75</sup> are in bold. Proteins are drawn to scale

varying duration<sup>7,8</sup>. Since genomics uncovered new mutations that are associated with antimicrobial resistance in other bacteria, such as those in *ethA* and *gyrB*, this prompted us to try and retrieve the clinical records of six patients whose strains displayed resistance to three or more drugs (dapsone, rifampicin, quinolones, and thioamides). Four of these extensively drug-resistant (XDR) strains were from multibacillary patients in Japan who had received a succession of monotherapies in the pre-MDT era and our genome analysis enabled the chronology of resistance emergence to be retraced. This is illustrated in Fig. 5, and sadly exemplified by the strain from patient Zensho-4 who was diagnosed in 1963 and first treated with protonamide followed by thiambutosin, both of which show cross-resistance and likely require activation by the EthA mono-oxidase that acquired the A25T missense mutation<sup>47</sup>; then treatment began with dapsone leading to emergence of the T53I mutation in *folP1*, followed by rifapentine that selected the S456L mutation in *rpoB*, and continued with ofloxacin, to which resistance arose from the A91V mutation in *gyrA* and D464N in *gyrB*. Molecular drug susceptibility testing was performed in 1998 and the patient finally cured by a regimen comprising clofazimine, minocycline, chloramphenicol and levofloxacin/sparfloxacin. The fifth XDR strain was from a newly diagnosed Brazilian case (Br2016-15) with no history of treatment for leprosy, confirming the ongoing transmission of primary antimicrobial resistance, while details of the sixth case could not be recovered.

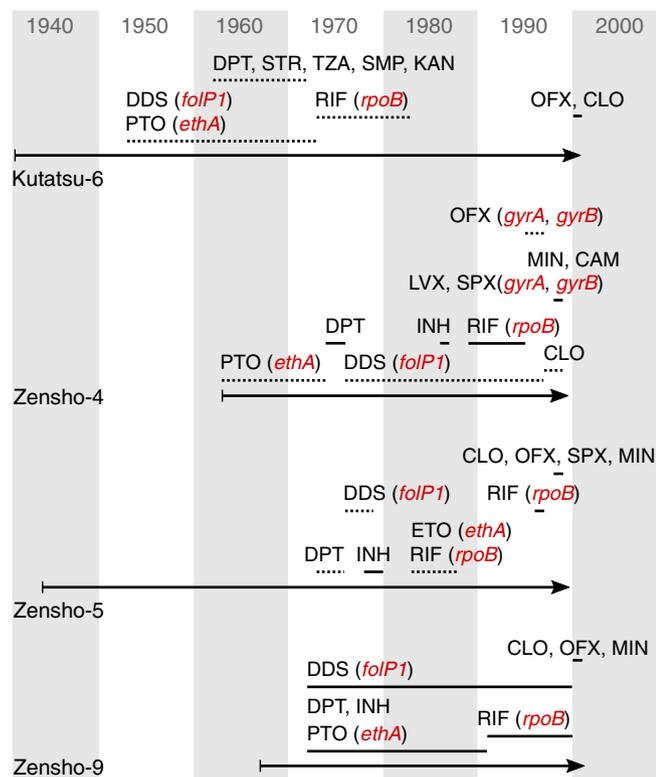
**Genes under positive selection.** We also identified genes containing an unusually high number of polymorphisms, multiple alleles, and homoplasies (Supplementary Figure 2), which could

be indicative of positive selection<sup>48</sup>. Strikingly, the distribution of these polymorphic sites around the genome was not random as they were often clustered, especially proximal to either side of the origin of replication (Supplementary Figure 3). Protein-changing mutations were found in 540 genes, with an average of 1.77 mutations per gene (STD 2.12). Table 2 contains a ranking of genes with at least five non-synonymous mutations or regions with one or more homoplasies (excluding VNTRs). The most polymorphic gene by far was *ML0411*<sup>49</sup> encoding the serine-rich antigen, a member of the immunogenic, surface-exposed PPE protein family. Two other known T-cell antigens whose genes display variability are *Lsr2* and *EsxA* (Table 2). Other than *nth*, three other polymorphic genes (*ML1040c*, *ML1750c*, and *ML1512c*) code for proteins that appear to function in nucleic acid or cyclic nucleotide metabolism (Table 2).

## Discussion

Here we have optimized and applied highly sensitive procedures to extract *M. leprae* DNA directly from human skin biopsies that is suitable for whole-genome sequencing. The resultant genome sequences were analyzed phylogenetically and used to retrace the origin of the leprosy bacillus, and to identify polymorphisms that had been positively selected during evolution. Such polymorphisms might reflect pressure from the human immune system, from MDT or other forces.

It is striking that the ancestral lineages of *M. leprae* predominate in East Asia, although we should keep in mind that Central Asia has been understudied, so it would be interesting to sequence more samples spanning the East–West axis of Asia, including the Middle East, where the 3K genotype is also



**Fig. 5** Timeline of the leprosy treatment and emergence of drug resistance in the XDR strains. Mutated genes conferring resistance to the corresponding drugs are shown in red. Arrows span from the onset of disease to the end of treatment. Horizontal lines show the period when a drug was given. Dotted lines mean irregular treatment. CAM chloramphenicol, CLO clofazimine, DDS dapsone, DPT thiambutosine (diphenylthiourea), ETO ethionamide, INH isoniazid, KAN kanamycin, LVX levofloxacin, MIN minocycline, OFX ofloxacin, PTO protionamide, RIF rifampicin, SMP sulfamethoxy-pyridazine, SPX sparfloxacin, STR streptomycin, TZA thiozamin

present<sup>16</sup>. Nevertheless, given the current data on the distribution of the 3K subtype we can deduce that the ancestor of *M. leprae* originated within Eurasia, probably in the Far East.

Endonuclease III (Nth) and the formamidopyrimidine and endonuclease VIII family (Fpg/Nei) of DNA glycosylases are central to the base excision repair pathway in bacteria<sup>50</sup>. Mycobacterial genomes usually contain a single *nth* and two *fpg/nei* genes but *M. leprae* has lost both *fpg/nei* orthologues and retained the *nth* gene. Nth, Fpg, and Nei may have overlapping functions and, in enteric bacteria, mutator phenotypes were observed when *nth* was inactivated in combination with the *fpg* and *nei* genes<sup>51–53</sup>. In *M. smegmatis*, deletion of *nth* and both the *nei* homologs resulted in elevated spontaneous mutation frequencies and increased sensitivity to oxidative stress<sup>54</sup>. Therefore, in the absence of Nei, inactivation of *nth* in *M. leprae* should lead to increased sequence variability, which is consistent with our results.

Strikingly, all *nth* mutants were also drug-resistant so Nth loss likely favors emergence of drug resistance, and *nth* mutations might serve as a surrogate marker for potential drug resistance and treatment failure. A link between a higher mutation rate and drug resistance was observed in strains of *M. tuberculosis* (which has *nth* and two *fpg/nei* genes) belonging to lineage 2, but the molecular basis for this is unknown<sup>55</sup>. For a pathogen with an extremely reduced genome such as *M. leprae*, a hypermutator phenotype could be detrimental and ultimately lethal.

Drug resistance is alarming for leprosy control. There is growing evidence for primary quinolone resistance in strains of *M. leprae* from patients who have never been treated with quinolones for leprosy but may have received this drug for other infections<sup>56</sup>. Five new GyrA mutations were identified in this study, but their effect on FQ resistance remains to be determined. Since two of them arose independently in the GyrA intein, which is removed by protein splicing, they may not impact quinolone activity (Fig. 4). Three non-synonymous mutations were found in *gyrB* (Table 2) and experimental evidence exists for two of them conferring quinolone resistance in *in vitro* assays or in *M. tuberculosis*<sup>57,58</sup>. To our knowledge, this is the first report of *M. leprae* clinical isolates harboring mutations in *gyrB*. Thus, despite their apparent rarity, mutations in *gyrB* should be systematically assessed in drug resistance screening.

A range of known and new mutations was detected in the DRDR and elsewhere in *rpoB* (Fig. 4; Table 1). Some of these might have a compensatory role in restoring fitness, that is known to be reduced to various degrees in rifampicin-resistant mutants of *M. tuberculosis*<sup>59,60</sup>. Similarly, compensatory mutations in *rpoA* and *rpoC* can occur in rifampicin-resistant *M. tuberculosis*<sup>44</sup>. The *rpoA* substitution T187P in the rifampicin-resistant *M. leprae* strain Br14-5 was shown to be compensatory in *M. tuberculosis*<sup>44</sup>. Rifampicin-resistant isolates of *M. tuberculosis* harbor more mutations in *rpoC* compared to rifampicin-susceptible isolates<sup>44,61</sup>. In our case, we observed no clear correlation between rifampicin resistance and mutations in *rpoC*, which occurred in two resistant and two wild-type strains.

Arguably the most intriguing finding of the present investigation was the remarkably high frequency of mutations in the *fadD9* and *ribD* genes and in 19/23 cases these occur in strains that have at least one mutation that is associated with resistance to a leprosy drug (Fig. 4; Table 1). Functional information for *fadD9* is scarce and the mutations found are predicted to either abolish protein production (8/16) or to cause detrimental amino acid changes (Fig. 4). In the case of *ribD*, 14 different missense mutations were found in a group of 17 variant alleles, indicating that this is likely an essential function. From studies with *M. tuberculosis* it is known that *ribD* encodes an alternative dihydrofolate reductase, with relatively low activity compared to that conferred by the bona fide dihydrofolate reductase gene, *dfrA*<sup>62</sup>. In clinical isolates of *M. tuberculosis*, a promoter mutation causes overexpression of *ribD* that is associated with resistance to the old drug, *para*-amino salicylic acid (PAS), and to certain DHFR inhibitors<sup>63</sup>. This suggests that the mutations detected in the *M. leprae ribD* gene may also confer resistance to PAS and support for this is provided by the fact that vadrine (2-pyridyl-(4)-1,3,4-oxydiazolone-(5)-*p*-aminosalicylate) was used as a drug to treat leprosy before dapsone became widely available<sup>64</sup>. It is thus possible that the *ribD* mutations we report here arose nearly 60 years ago following treatment with vadrine or another PAS derivative. Our discovery of these mutations and those in *fadD9* should encourage further experimentation in order to establish their true role and contribution to antimicrobial resistance, especially to clofazimine.

## Methods

**Sample collection.** Samples were taken from leprosy patients as punch biopsies of skin (preserved in 70% ethanol or formalin-fixed and paraffin-embedded (FFPE)), which is standard diagnostic procedure for leprosy, or from mouse foot-pads. Details about the samples used in this study are given in Supplementary Data 1 and below.

**Origin of S15:** Strain S15 corresponds to strain 92041<sup>65</sup>, which was isolated from a lepromatous leprosy patient originally from Martinique. The origin of S15 was erroneously attributed to New Caledonia in Monot et al.<sup>16</sup>, and the error was subsequently propagated in several publications<sup>2,17,66,67</sup>.

**Table 2 Highly polymorphic genes and genomic regions of *M. leprae***

Gene or region	Description	Non-synonymous mutations (multi-allele loci)	Synonymous mutations	Homoplasmy
<i>MLO411</i>	Serine-rich antigen	32 (4)	1	4
<i>ML1040c</i>	Putative ATP-dependent helicase	19 (1)	0	0
<i>fadD9</i>	Probable fatty-acid-CoA ligase	16	1	1
<i>ML1750c</i>	Putative nucleotide cyclase	17	1	0
<i>ML1512c</i>	Putative ribonuclease J	17	0	0
<i>ribD</i>	Bifunctional enzyme riboflavin biosynthesis protein	17	0	1
<i>rpoB</i>	DNA-directed RNA polymerase (beta chain)	13 (1)	1	3
<i>ML1753c</i>	Probable transcriptional regulatory protein	9	0	0
<i>gyrA</i>	DNA gyrase (subunit A)	8	8	1
<i>ML1300</i>	Conserved hypothetical protein	8	0	0
<i>nth</i>	Endonuclease III	7	0	1
<i>ctpC</i>	Metal cation-transporting P-type ATPase C	7	2	0
<i>rpoC</i>	DNA-directed RNA polymerase (beta chain)	7	5	0
<i>ML2687c</i>	Probable conserved transmembrane protein	7	0	0
<i>ML1052c</i>	Conserved hypothetical protein	7	1	0
<i>MLO009</i>	Hypothetical protein	6	1	0
<i>MLO283</i>	Cation-efflux transporter component	6	1	0
<i>ML2700</i>	Conserved transmembrane protein	6	7	0
<i>mfd</i>	Transcription-repair coupling factor	6	1	0
<i>ppsC</i>	Phenolphthiocerol and DIM synthesis	6	1	0
<i>ethA</i>	Activates the pro-drug ethionamide	5	0	0
<i>trpE</i>	Biosynthesis of tryptophan (at the first step)	5	1	0
<i>pknB</i>	Transmembrane serine/threonine-protein kinase	5	3	0
<i>fas</i>	Fatty acid synthase	5	1	0
<i>esxA</i>	Early secretory antigenic target	2	0	1
<i>lsr2</i>	Dominant T-cell antigen and stimulates lymphoproliferation.	4	1	3
<i>folP1</i>	Dihydropteroate synthase	5 (2)	1	2
<i>ML1752c</i>	Conserved hypothetical protein	2	1	1
<i>ppsA</i>	Phenolphthiocerol and DIM synthesis	4	3	1
<i>MLO803</i>	Two-component sensor kinase	2 (1)	0	0
<i>MLO237</i>	Conserved hypothetical protein (pseudogene)	NA	1	1
<i>ML0010c-ppiA</i>	Intergenic region	NA	3	1

Origin of the case Ng14-3: We mention in the Results that West Africa harbors only SNP-type 4 strains. Strain Ng14-35 (subtype 1D) is an exception, but the patient (a native of Nigeria) developed leprosy during a prolonged stay in Libya.

LRC-1A, an unidentified strain from Japan: A sample obtained from the Leprosy Research Center (LRC) in Tokyo, initially labeled as Airaku-2, did not contain the mutations in *folP1* and *rpoB* previously detected in Airaku-2 by PCR sequencing<sup>7</sup>. Therefore, we renamed this sample to LRC-1A (standing for SNP subtype 1A sample from LRC).

**DNA extraction and library preparation.** DNA was extracted from 101 human skin biopsies with known BI using a customized in-house protocol combining host tissue digestion and the QIAmp microbiome kit for host DNA depletion, strong bacterial cell lysis and silica-based purification. Punch biopsies (6 mm) in 70% ethanol were first rehydrated in Hank's balanced solution prior to mincing with scissors. Cells were detached from the tissue by 30 min incubation at 37 °C with a mixture of 0.5 U of collagenase and dispase, followed by incubation at 56 °C with 10 mg/ml of trypsin until complete digestion. Free cells were then suspended in 1 ml of phosphate-buffered saline (PBS) and DNA was extracted using the QIAmp DNA microbiome extraction kit according to the manufacturer's recommendations. Each run of extraction included a batch of five to nine samples and one blank control (500 µl of Hank's balanced solution). The presence of *M. leprae* was assessed by PCR using RLEP primers<sup>2</sup> prior to library preparation. Libraries were prepared from 50 µl of extracted DNA using the Kapa Hyperprep kit as described previously<sup>2,5</sup>. DNA from FFPE samples was extracted using the truXTRACTM FFPE DNA kit (Covaris) as described previously<sup>28</sup>. Libraries prepared from the extracted DNA were used directly for shotgun sequencing. *M. leprae* DNA extraction quality was assessed from the percentage of *M. leprae* DNA present in the library inferred by alignment to the reference genome sequence, with a minimum threshold set at 1%. This threshold was chosen because it yields an average genome coverage of at least 5× per sample in a multiplexed run of 10 samples on one HiSeq 2500 lane (yielding around 20 million reads per sample and 100 bases per read).

**Library enrichment.** Libraries with low *M. leprae* content underwent enrichment using whole-genome tiling arrays as described previously<sup>17</sup>. Briefly, Illumina

libraries were hybridized onto custom Agilent SureSelect Capture Arrays containing ca. one million DNA probes (60 bp) spanning the entire *M. leprae* genome (tiled every 4 bp), followed by elution and PCR amplification.

**Sequencing.** Sequencing was performed on Illumina Hi-Seq 2000, Hi-Seq 2500, or Mi-Seq instruments.

**Sequence processing.** We took precautions in analyzing the data to avoid false-positive SNP calls. All raw reads were adapter- and quality-trimmed with Trimmomatic v0.33<sup>68</sup>. The quality settings were "SLIDINGWINDOW:5:15 MINLEN:40". Paired-end (PE) data were additionally processed with SeqPrep (<https://github.com/jstjohn/SeqPrep>) to merge overlapping pairs. This increases the accuracy of sequence in the overlapping area, avoids problems in estimating coverage and creates longer reads, which facilitates InDel calling. Duplicate reads were omitted from downstream analyses. This is especially important for libraries with insufficient *M. leprae* DNA fragments, which is not uncommon for low BI samples or samples that are difficult to process, like FFPE samples. In these cases, library enrichment with array-capture, or very deep sequencing often produce a high number of duplicate reads (DNA fragments that were sequenced multiple times, seemingly increasing the overall genome coverage), with each read having dozens or even hundreds of copies. Such reads will amplify possible artefacts and sequence errors, resulting in false SNP calls.

**Sequence analysis.** Preprocessed reads were mapped onto the *M. leprae* TN reference genome (GenBank AL450380.1) with Bowtie2 v2.2.5<sup>69</sup>. We filtered out all reads with mapping quality below 8 and omitted repetitive regions in the reference sequence. We also omitted ribosomal RNA (rRNA) genes because alignments in these regions tend to be error prone. This is because rRNA genes are highly conserved in bacteria, so sequences from other species could map to the *M. leprae* reference sequence. This usually happens when the content of *M. leprae* DNA in a sequencing library is scarce and is even more pronounced when libraries with low *M. leprae* content undergo array-capture. However, because lineage-specific mutations were previously observed in the *M. leprae* *rrs* gene<sup>30</sup>, we manually checked the alignments corresponding to the rRNA genes and added the curated results to Supplementary Data 2.

SNP calling was done using VarScan v2.3.9<sup>70</sup>. To avoid false-positive SNP calls the following cutoffs were applied: minimum overall coverage of five non-duplicated reads, minimum of three non-duplicated reads supporting the SNP, mapping quality score >8, base quality score >15, and a SNP frequency above 80%. InDel calling was done using Platypus v0.8.1<sup>71</sup> followed by manual curation. Completed genome sequences of *M. leprae* Br4923 and *Mycobacterium lepromatosis* (GenBank JRPY000000000.1) were aligned against the *M. leprae* TN reference using LAST<sup>72</sup> using the default parameters for the former and the gamma-centroid option for the latter.

**Mixed samples.** A large number of missing values, especially in lineage-specific loci, points to the presence of more than one strain in a sequencing library. Although not thoroughly tested, in our opinion mixed data sets are mostly due to technical problems or contamination because in some cases we were able to identify the problematic strains. The possible presence of multiple *M. leprae* strains in single skin lesions was not tested in this study, but we expect it to be extremely low. Overall, a few mixed data sets were detected and some were removed from this study, except for samples that we deemed important and describe below. Nevertheless, results were not biased because loci with mixed alleles were treated as missing values.

Zensho-4 seems to contain a fraction of another strain (possibly around 40%) that is closely related to it. Only a few loci had mixed alleles, and these include the A91V substitution in *gyrA* (supported by 62% of reads) and the D464N substitution in *gyrB* (supported by 41% of reads). The latter was attributed to Zensho-4 for simplicity. Similarly, Zensho-5 seems to contain around 30% of Zensho-4. This is the main reason why we could not detect SNPs specific only to Zensho-5 (Fig. 3a), since such SNPs would be “diluted” with wild-type alleles from Zensho-4 and could not pass the SNP “purity” threshold. We included these two samples in this study because they are multi-drug-resistant and belong to the SNP-type 3K-0. Furthermore, mutations in genes conferring drug resistance from this study match with those from earlier reports of these samples, confirming their identity<sup>73</sup>.

Thai-311 contains <20% of an unidentified 3K-0 strain that belongs to the Kyoto-1/Zensho-5 cluster of strains (Fig. 3a). SNP calling was not significantly affected. Finally, sample Ye2-3 contained around 25% of an unidentified strain belonging to SNP-type 4. Because we only have few samples from Yemen, we decided to keep Ye2-3 in this study.

**Phylogeny and dating analysis.** Concatenated SNP alignments were used for the analyses. MP trees were constructed in MEGA6<sup>74</sup> using 500 bootstrap replicates. Sites with missing data were partially deleted (80% coverage cutoff), resulting in 3046 variable sites used for the tree calculation. The Subtree-Pruning-Regrafting algorithm was used as the MP search method. Dating analysis and discrete phylogeography were done using BEAST2 v2.4.4<sup>40</sup>. Details are given in the Supplementary Note.

**Data availability.** Sequence data are available from the NCBI Sequence Read Archive (SRA) under accession number SRP072827. Accession numbers for all samples used in this study are given in the Supplementary Data 1. Other relevant data supporting the findings of the study are available in this published article and its Supplementary Information files, or from the corresponding author upon request.

Received: 25 August 2017 Accepted: 12 December 2017

Published online: 24 January 2018

## References

- Britton, W. J. & Lockwood, D. N. J. Leprosy. *Lancet* **363**, 1209–1219 (2004).
- Avanzi, C. et al. Red squirrels in the British Isles are infected with leprosy bacilli. *Science* **354**, 744–747 (2016).
- World Health Organization. Global leprosy update 2015: time for action, accountability and inclusion. *Wkly Epidemiol. Rec.* **91**, 405–420 (2016).
- Scollard, D. M. et al. The continuing challenges of leprosy. *Clin. Microbiol. Rev.* **19**, 338–381 (2006).
- Avanzi, C. et al. Transmission of drug-resistant leprosy in Guinea-Conakry detected using molecular epidemiological approaches. *Clin. Infect. Dis.* **63**, 1482–1484 (2016).
- Cambau, E., Perani, E., Guillemin, I., Jamet, P. & Ji, B. Multidrug-resistance to dapsone, rifampicin, and ofloxacin in *Mycobacterium leprae*. *Lancet* **349**, 103–104 (1997).
- Maeda, S. et al. Multidrug resistant *Mycobacterium leprae* from patients with leprosy. *Antimicrob. Agents Chemother.* **45**, 3635–3639 (2001).
- Matsuoka, M., Kashiwabara, Y. & Namisato, M. A. *Mycobacterium leprae* isolate resistant to dapsone, rifampin, ofloxacin and sparfloxacin. *Int. J. Lepr. Mycobact. Dis.* **68**, 452–455 (2000).
- Matsuoka, M., Kashiwabara, Y., Liangfen, Z., Goto, M. & Kitajima, S. A second case of multidrug-resistant *Mycobacterium leprae* isolated from a Japanese patient with relapsed lepromatous leprosy. *Int. J. Lepr. Mycobact. Dis.* **71**, 240–243 (2003).
- Matsuoka, M. et al. The frequency of drug resistance mutations in *Mycobacterium leprae* isolates in untreated and relapsed leprosy patients from Myanmar, Indonesia and the Philippines. *Lepr. Rev.* **78**, 343–352 (2007).
- da Silva Rocha, A. et al. Drug and multidrug resistance among *Mycobacterium leprae* isolates from Brazilian relapsed leprosy patients. *J. Clin. Microbiol.* **50**, 1912–1917 (2012).
- You, E.-Y., Kang, T. J., Kim, S.-K., Lee, S.-B. & Chae, G.-T. Mutations in genes related to drug resistance in *Mycobacterium leprae* isolates from leprosy patients in Korea. *J. Infect.* **50**, 6–11 (2005).
- Lahiri, R. & Adams, L. B. Cultivation and Viability Determination of *Mycobacterium leprae*. In *International Textbook of Leprosy* (eds. Scollard, D. M. & Gillis, T. P.) <http://www.internationaltextbookofleprosy.org/> (2016).
- Cole, S. T. et al. Massive gene decay in the leprosy bacillus. *Nature* **409**, 1007–1011 (2001).
- Moya, A., Peretó, J., Gil, R. & Latorre, A. Learning how to live together: genomic insights into prokaryote–animal symbioses. *Nat. Rev. Genet.* **9**, 218–229 (2008).
- Monot, M. et al. Comparative genomic and phylogeographic analysis of *Mycobacterium leprae*. *Nat. Genet.* **41**, 1282–1289 (2009).
- Schuenemann, V. J. et al. Genome-wide comparison of medieval and modern *Mycobacterium leprae*. *Science* **341**, 179–183 (2013).
- Truman, R. W. et al. Probable zoonotic leprosy in the southern United States. *N. Engl. J. Med.* **364**, 1626–1633 (2011).
- Monot, M. et al. On the origin of leprosy. *Science* **308**, 1040–1042 (2005).
- Dharmendra Leprosy in ancient Indian medicine. *Int. J. Lepr.* **15**, 424–430 (1947).
- Taylor, G. M. et al. *Mycobacterium leprae* genotype amplified from an archaeological case of lepromatous leprosy in Central Asia. *J. Archaeol. Sci.* **36**, 2408–2414 (2009).
- Donoghue, H. D. et al. Co-infection of *Mycobacterium tuberculosis* and *Mycobacterium leprae* in human archaeological samples: a possible explanation for the historical decline of leprosy. *Proc. Biol. Sci.* **272**, 389–394 (2005).
- Blau, S. & Yagodin, V. Osteoarchaeological evidence for leprosy from western Central Asia. *Am. J. Phys. Anthropol.* **126**, 150–158 (2005).
- Tayles, N. & Buckley, H. R. Leprosy and tuberculosis in Iron Age Southeast Asia? *Am. J. Phys. Anthropol.* **125**, 239–256 (2004).
- Mariotti, V., Dutour, O., Belcastro, M. G., Facchini, F. & Brasili, P. Probable early presence of leprosy in Europe in a Celtic skeleton of the 4th–3rd century BC (Casalecchio di Reno, Bologna, Italy). *Int. J. Osteoarchaeol.* **15**, 311–325 (2005).
- Donoghue, H. D. et al. A migration-driven model for the historical spread of leprosy in medieval Eastern and Central Europe. *Infect. Genet. Evol.* **31**, 250–256 (2015).
- Loman, N. J. & Pallen, M. J. Twenty years of bacterial genome sequencing. *Nat. Rev. Microbiol.* **13**, 787–794 (2015).
- Stefani, M. M. A. et al. Whole genome sequencing distinguishes between relapse and reinfection in recurrent leprosy cases. *PLoS Negl. Trop. Dis.* **11**, e0005598 (2017).
- Weng, X. et al. Molecular, ethno-spatial epidemiology of leprosy in China: Novel insights for tracing leprosy in endemic and non endemic provinces. *Infect. Genet. Evol.* **14**, 361–368 (2013).
- Yuan, Y. et al. Characterization of *Mycobacterium leprae* genotypes in China—identification of a new polymorphism C251T in the 16S rRNA gene. *PLoS ONE* **10**, e0133268 (2015).
- Kim, J. P. SNP genotypes of *Mycobacterium leprae* isolated in Korea. *Korean Lepr. Bull.* **45**, 3–19 (2012).
- McLaughlin, R. *Rome and the Distant East: Trade Routes to the Ancient Lands of Arabia, India and China* (Continuum, London, 2010).
- Lavanaia, M. et al. Single nucleotide polymorphisms typing of *Mycobacterium leprae* reveals focal transmission of leprosy in high endemic regions of India. *Clin. Microbiol. Infect.* **19**, 1058–1062 (2013).
- Lavanaia, M. et al. Genotyping of *Mycobacterium leprae* strains from a region of high endemic leprosy prevalence in India. *Infect. Genet. Evol.* **36**, 256–261 (2015).
- Das, M., Chaitanya, V. S., Kanmani, K., Rajan, L. & Ebenezer, M. Genomic diversity in *Mycobacterium leprae* isolates from leprosy cases in South India. *Infect. Genet. Evol.* **45**, 285–289 (2016).
- Fontes, A. N. B. et al. Genotyping of *Mycobacterium leprae* present on Ziehl-Neelsen-stained microscopic slides and in skin biopsy samples from leprosy patients in different geographic regions of Brazil. *Mem. Inst. Oswaldo Cruz.* **107**, 143–149 (2012).
- Inskip, S. A. et al. Osteological, biomolecular and geochemical examination of an early Anglo-Saxon case of lepromatous leprosy. *PLoS ONE* **10**, e0124282 (2015).
- Mendum, T. A. et al. *Mycobacterium leprae* genomes from a British medieval leprosy hospital: towards understanding an ancient epidemic. *BMC Genom.* **15**, 270 (2014).

39. Taulman, J. F. & Robbins, L. W. Recent range expansion and distributional limits of the nine-banded armadillo (*Dasypus novemcinctus*) in the United States. *J. Biogeogr.* **23**, 635–648 (1996).
40. Bouckaert, R. et al. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **10**, e1003537 (2014).
41. Phetsuksiri, B. et al. SNP genotypes of *Mycobacterium leprae* isolates in Thailand and their combination with rpoT and TTC genotyping for analysis of leprosy distribution and transmission. *Jpn J. Infect. Dis.* **65**, 52–56 (2012).
42. Nakata, N., Kai, M. & Makino, M. Mutation analysis of mycobacterial rpoB genes and rifampin resistance using recombinant *Mycobacterium smegmatis*. *Antimicrob. Agents Chemother.* **56**, 2008–2013 (2012).
43. Bahrmand, A. R., Titov, L. P., Tasbiti, A. H., Yari, S. & Graviss, E. A. High-level rifampin resistance correlates with multiple mutations in the rpoB gene of pulmonary tuberculosis isolates from the Afghanistan border of Iran. *J. Clin. Microbiol.* **47**, 2744–2750 (2009).
44. Comas, I. et al. Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. *Nat. Genet.* **44**, 106–110 (2012).
45. DeBarber, A. E., Mdluli, K., Bosman, M., Bekker, L. G. & Barry, C. E. Ethionamide activation and sensitivity in multidrug-resistant *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **97**, 9677–9682 (2000).
46. Baulard, A. R. et al. Activation of the pro-drug ethionamide is regulated in mycobacteria. *J. Biol. Chem.* **275**, 28326–28331 (2000).
47. Pattyn, S. R. & Colston, M. J. Cross-resistance amongst thiambutosine, thiacetazone, ethionamide and prothionamide with *Mycobacterium leprae*. *Lepr. Rev.* **49**, 324–326 (1978).
48. Hedge, J. & Wilson, D. J. Practical approaches for detecting selection in microbial genomes. *PLoS Comput. Biol.* **12**, e1004739 (2016).
49. Kai, M. et al. Characteristic mutations found in the *ML0411* gene of *Mycobacterium leprae* isolated in Northeast Asian countries. *Infect. Genet. Evol.* **19**, 200–204 (2013).
50. Zharkov, D. O. Base excision DNA repair. *Cell. Mol. Life Sci.* **65**, 1544–1565 (2008).
51. Jiang, D., Hatahet, Z., Blaisdell, J. O., Melamed, R. J. & Wallace, S. S. *Escherichia coli* endonuclease VIII: cloning, sequencing, and overexpression of the nei structural gene and characterization of nei and nei nth mutants. *J. Bacteriol.* **179**, 3773–3782 (1997).
52. Saito, Y. et al. Characterization of endonuclease III (*nth*) and endonuclease VIII (*nei*) mutants of *Escherichia coli* K-12. *J. Bacteriol.* **179**, 3783–3785 (1997).
53. Suvarnapunya, A. E. & Stein, M. A. DNA base excision repair potentiates the protective effect of Salmonella Pathogenicity Island 2 within macrophages. *Microbiol. Read. Engl.* **151**, 557–567 (2005).
54. Moolla, N., Goosens, V. J., Kana, B. D. & Gordhan, B. G. The contribution of Nth and Nei DNA glycosylases to mutagenesis in *Mycobacterium smegmatis*. *DNA Repair* **13**, 32–41 (2014).
55. Ford, C. B. et al. *Mycobacterium tuberculosis* mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. *Nat. Genet.* **45**, 784–790 (2013).
56. Linder, J. A., Huang, E. S., Steinman, M. A., Gonzales, R. & Stafford, R. S. Fluoroquinolone prescribing in the United States: 1995 to 2002. *Am. J. Med.* **118**, 259–268 (2005).
57. Malik, S., Willby, M., Sikes, D., Tsodikov, O. V. & Posey, J. E. New insights into fluoroquinolone resistance in *Mycobacterium tuberculosis*: functional genetic analysis of gyrA and gyrB mutations. *PLoS ONE* **7**, e39754 (2012).
58. Yokoyama, K. et al. Impact of amino acid substitutions in B subunit of DNA gyrase in *Mycobacterium leprae* on fluoroquinolone resistance. *PLoS Negl. Trop. Dis.* **6**, e1838 (2012).
59. Gagneux, S. et al. The competitive cost of antibiotic resistance in *Mycobacterium tuberculosis*. *Science* **312**, 1944–1946 (2006).
60. Mariam, D. H., Mengistu, Y., Hoffner, S. E. & Andersson, D. I. Effect of rpoB mutations conferring rifampin resistance on fitness of *Mycobacterium tuberculosis*. *Antimicrob. Agents Chemother.* **48**, 1289–1294 (2004).
61. de Vos, M. et al. Putative compensatory mutations in the rpoC gene of rifampin-resistant *Mycobacterium tuberculosis* are associated with ongoing transmission. *Antimicrob. Agents Chemother.* **57**, 827–832 (2013).
62. Cheng, Y.-S. & Sacchetti, J. C. Structural insights into *Mycobacterium tuberculosis* Rv2671 protein as a dihydrofolate reductase functional analogue contributing to para-aminosalicylic acid resistance. *Biochemistry* **55**, 1107–1119 (2016).
63. Zheng, J. et al. Para-aminosalicylic acid is a prodrug targeting dihydrofolate reductase in *Mycobacterium tuberculosis*. *J. Biol. Chem.* **288**, 23447–23456 (2013).
64. Jopling, W. H. & Ridley, D. S. Vadrine (S. 131) in the treatment of lepromatous leprosy: a preliminary report. *Lepr. Rev.* **29**, 143–147 (1958).
65. Honoré, N., Perrani, E., Telenti, A., Grosset, J. & Cole, S. T. A simple and rapid technique for the detection of rifampin resistance in *Mycobacterium leprae*. *Int. J. Lepr. Mycobact. Dis.* **61**, 600–604 (1993).
66. Singh, P. et al. Insight into the evolution and origin of leprosy bacilli from the genome sequence of *Mycobacterium lepromatosis*. *Proc. Natl Acad. Sci. USA* **112**, 4459–4464 (2015).
67. Singh, P. et al. Genome-wide re-sequencing of multidrug-resistant *Mycobacterium leprae* Airaku-3. *Clin. Microbiol. Infect.* **20**, O619–O622 (2014).
68. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
69. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
70. Koboldt, D. C. et al. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22**, 568–576 (2012).
71. Rimmer, A. et al. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat. Genet.* **46**, 912–918 (2014).
72. Kielbasa, S. M., Wan, R., Sato, K., Horton, P. & Frith, M. C. Adaptive seeds tame genomic sequence comparison. *Genome Res.* **21**, 487–493 (2011).
73. Matsuoka, M. The history and characteristics of isolates maintained at the leprosy research center. *Nihon Hansen. Gakkai Zasshi* **79**, 247–256 (2010).
74. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).
75. Choi, Y. & Chan, A. P. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **31**, 2745–2747 (2015).

## Acknowledgements

We thank the Genomic Technologies Facility at the University of Lausanne for Illumina sequencing and technical support and all the patients and clinical staff who participated in the study. This work was supported by the Fondation Raoul Follereau, the Swiss National Science Foundation grant IZRJZ3\_164174, the Swiss Cooperation and Development Center (CODEV), the Heiser Program of the New York Community Trust for Research in Leprosy (grant numbers P15-000827 and P16-000976), and grants CNPq 428964/2016-8 and CAPES PROAMAZONIA 3288/2013.

## Author contributions

S.T.C., P.S., C.A., and A.B. designed the study. C.A., P.S., S.G., A.N.B.F., and P.B. processed the samples, extracted DNA, and prepared sequencing libraries. A.B. and C.L. processed the data. A.B. analyzed the data and prepared figures and tables. A.B., C.A., and S.T.C. interpreted the results and wrote the manuscript with input from other authors. S.G., Y.M., M.N., K.B., C.G.S., M.B.S., R.C.B., M.A.C.F., F.B.F., J.G.B., J.A.C.N., S. B.-S., A.L., A.R.A.-S., Y.A.-Q., A.S.A., G.B., L.V.C., F.S., C.R.J., M.Ko., A.F., S.O.S., M.G., O.K., M.M.A.S., G.O.P., P.N.S., E.N.S., M.O.M., P.S.R., I.M.F.D.B., J.S.S., A.A., M.M., and M.Ka. participated in identification of leprosy cases, patient management, sample collection and preparation, and microscopy.

## Additional information

**Supplementary Information** accompanies this paper at <https://doi.org/10.1038/s41467-017-02576-z>.

**Competing interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018

Andrej Benjak<sup>1</sup>, Charlotte Avanzi<sup>1</sup>, Pushpendra Singh<sup>1,2</sup>, Chloé Loiseau<sup>1,3</sup>, Selfu Girma<sup>4</sup>, Philippe Busso<sup>1</sup>, Amanda N. Brum Fontes<sup>5</sup>, Yuji Miyamoto<sup>6</sup>, Masako Namisato<sup>7</sup>, Kidist Bobosha<sup>4</sup>, Claudio G. Salgado<sup>8</sup>, Moisés B. da Silva<sup>8</sup>, Raquel C. Bouth<sup>8</sup>, Marco A.C. Frade<sup>9</sup>, Fred Bernardes Filho<sup>9</sup>, Josafá G. Barreto<sup>10</sup>, José A.C. Nery<sup>11</sup>, Samira Bühler-Sékula<sup>12</sup>, Andréanne Lupien<sup>1</sup>, Abdul R. Al-Samie<sup>13</sup>, Yasin Al-Qubati<sup>13</sup>, Abdul S. Alkubati<sup>13</sup>, Gisela Bretzel<sup>14</sup>, Lucio Vera-Cabrera<sup>15</sup>, Fatoumata Sakho<sup>16</sup>, Christian R. Johnson<sup>17</sup>, Mamoudou Kodio<sup>18</sup>, Abdoulaye Fomba<sup>18</sup>, Samba O. Sow<sup>18</sup>, Moussa Gado<sup>19</sup>, Ousmane Konaté<sup>19</sup>, Mariane M.A. Stefani<sup>12</sup>, Gerson O. Penna<sup>20</sup>, Philip N. Suffys<sup>5,21</sup>, Euzenir Nunes Sarno<sup>11</sup>, Milton O. Moraes<sup>11</sup>, Patricia S. Rosa<sup>22</sup>, Ida M.F.Dias Baptista<sup>22</sup>, John S. Spencer<sup>23</sup>, Abraham Aseffa<sup>4</sup>, Masanori Matsuoka<sup>6</sup>, Masanori Kai<sup>6</sup> & Stewart T. Cole<sup>1</sup>

<sup>1</sup>Global Health Institute, Ecole Polytechnique Fédérale de Lausanne, Lausanne 1015, Switzerland. <sup>2</sup>Department of Microbiology and Biotechnology Centre, Maharaja Sayajirao University of Baroda, Vadodra 390002, India. <sup>3</sup>Swiss Tropical and Public Health Institute, 4051 Basel, Switzerland. <sup>4</sup>Armauer Hansen Research Institute, PO Box 1005, Addis Ababa 1000, Ethiopia. <sup>5</sup>Laboratory of Molecular Biology Applied to Mycobacteria, Oswaldo Cruz Institute, Fiocruz, Rio de Janeiro 21040-360, Brazil. <sup>6</sup>Leprosy Research Center, National Institute of Infectious Diseases, Higashimurayama, Tokyo 189-0002, Japan. <sup>7</sup>AUEN Polyclinic, Nakatomi, Tokorozawa 359-0002 Saitama Prefecture, Japan. <sup>8</sup>Laboratório de Dermato-Imunologia, Instituto de Ciências Biológicas, Universidade Federal do Pará, Marituba 67200-000 Pará, Brazil. <sup>9</sup>Dermatology Division, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto 14049-900 São Paulo, Brazil. <sup>10</sup>Spatial Epidemiology Laboratory, Federal University of Pará, Castanhal 68746-360 Pará, Brazil. <sup>11</sup>Leprosy Laboratory, Oswaldo Cruz Institute, Fiocruz, Rio de Janeiro 21040-900, Brazil. <sup>12</sup>Tropical Pathology and Public Health Institute, Federal University of Goiás, Goiânia 74605-050, Brazil. <sup>13</sup>Ministry of Health and Population, c/o National Leprosy Elimination Programme, Taiz, Yemen. <sup>14</sup>Division of Infectious Diseases and Tropical Medicine, University Hospital, Ludwig-Maximilians-University, 80802 Munich, Germany. <sup>15</sup>Laboratorio Interdisciplinario de Investigación Dermatológica, Servicio de Dermatología, Hospital Universitario, Universidad Autónoma de Nuevo León, 64460 Monterrey, Mexico. <sup>16</sup>Programme National Lèpre de Guinée, Conakry, Guinea. <sup>17</sup>Centre Interfacultaire de Formation et de Recherche en Environnement pour le Développement Durable, University of Abomey-Calavi, 03 BP 1463 Jericho, Cotonou, Benin. <sup>18</sup>Centre National d'Appui à la Lutte Contre la Maladie, Bamako, Mali. <sup>19</sup>Programme National de Lutte contre la Lèpre, Ministry of Public Health, Niamey, Niger. <sup>20</sup>Tropical Medicine Centre, University of Brasília, Fiocruz, Brasília 70910-900, Brazil. <sup>21</sup>Department of Biomedical Sciences, Mycobacteriology Unit, Tropical Institute of Medicine, 2000 Antwerp, Belgium. <sup>22</sup>Lauro Souza Lima Institute, Bauru 17034-971 São Paulo, Brazil. <sup>23</sup>Department of Microbiology, Immunology, and Pathology, Colorado State University, Fort Collins, CO 80523-1682, USA