

A new evolutionary scenario for the *Mycobacterium tuberculosis* complex

R. Brosch*, S. V. Gordon†, M. Marmiesse*, P. Brodin*, C. Buchrieser‡, K. Eiglmeier*, T. Garnier*, C. Gutierrez§, G. Hewinson†, K. Kremer¶, L. M. Parsons||, A. S. Pym*, S. Samper**, D. van Soolingen¶, and S. T. Cole**††

*Unité de Génétique Moléculaire Bactérienne, †Laboratoire de Génomique des Microorganismes Pathogènes, and §Centre National de Référence des Mycobactéries, Institut Pasteur, 25-28 Rue du Docteur Roux, 75724 Paris Cedex 15, France; ‡Veterinary Laboratories Agency, Woodham Lane, New Haw, Addlestone, Surrey KT15 3NB, United Kingdom; ¶Mycobacteria Reference Department, Diagnostic Laboratory for Infectious Diseases and Perinatal Screening, National Institute of Public Health and the Environment, 3720 BA Bilthoven, The Netherlands; ||Wadsworth Center, New York State Department of Health and School of Public Health, State University of New York at Albany, David Axelrod Institute, Albany, NY 12201-2002; and **Departamento de Microbiología, Medicina Preventiva y Salud Pública, Universidad de Zaragoza, C/Domingo Miralsn, 50009 Zaragoza, Spain

Edited by John Maynard Smith, University of Sussex, Brighton, United Kingdom, and approved January 9, 2002 (received for review October 15, 2001)

The distribution of 20 variable regions resulting from insertion-deletion events in the genomes of the tubercle bacilli has been evaluated in a total of 100 strains of *Mycobacterium tuberculosis*, *Mycobacterium africanum*, *Mycobacterium canettii*, *Mycobacterium microti*, and *Mycobacterium bovis*. This approach showed that the majority of these polymorphisms did not occur independently in the different strains of the *M. tuberculosis* complex but, rather, resulted from ancient, irreversible genetic events in common progenitor strains. Based on the presence or absence of an *M. tuberculosis* specific deletion (TbD1), *M. tuberculosis* strains can be divided into ancestral and "modern" strains, the latter comprising representatives of major epidemics like the Beijing, Haarlem, and African *M. tuberculosis* clusters. Furthermore, successive loss of DNA, reflected by region of difference 9 and other subsequent deletions, was identified for an evolutionary lineage represented by *M. africanum*, *M. microti*, and *M. bovis* that diverged from the progenitor of the present *M. tuberculosis* strains before TbD1 occurred. These findings contradict the often-presented hypothesis that *M. tuberculosis*, the etiological agent of human tuberculosis evolved from *M. bovis*, the agent of bovine disease. *M. canettii* and ancestral *M. tuberculosis* strains lack none of these deleted regions, and, therefore, seem to be direct descendants of tubercle bacilli that existed before the *M. africanum*→*M. bovis* lineage separated from the *M. tuberculosis* lineage. This observation suggests that the common ancestor of the tubercle bacilli resembled *M. tuberculosis* or *M. canettii* and could well have been a human pathogen already.

evolution | diagnostic | identification

The mycobacteria grouped in the *Mycobacterium tuberculosis* complex are characterized by 99.9% similarity at the nucleotide level and identical 16S rRNA sequences (1, 2) but differ widely in terms of their host tropisms, phenotypes, and pathogenicity. Assuming that they all are derived from a common ancestor, it is intriguing that some are exclusively human (*M. tuberculosis*, *Mycobacterium africanum*, *Mycobacterium canettii*) or rodent pathogens (*Mycobacterium microti*), whereas others have a wide host spectrum (*Mycobacterium bovis*). What was the genetic organization of the last common ancestor of the tubercle bacilli, and in which host did it live? Which genetic events may have contributed to the fact that the host spectrum is so different and often specific? Where and when did *M. tuberculosis* evolve? Answers to these questions are important for a better understanding of the pathogenicity and the global epidemiology of tuberculosis and may help to anticipate future trends in the spread of the disease.

Because of the unusually high degree of conservation in their housekeeping genes, it has been suggested that the members of the *M. tuberculosis* complex underwent an evolutionary bottleneck at the time of speciation, estimated to have occurred

roughly 15,000–20,000 years ago (2). Also, it has been speculated that *M. tuberculosis*, the most widespread etiological agent of human tuberculosis has evolved from *M. bovis*, the agent of bovine tuberculosis, by specific adaptation of an animal pathogen to the human host (3). However, both hypotheses were proposed before the whole genome sequence of *M. tuberculosis* (4) was available and before comparative genomics uncovered several variable genomic regions in the members of the *M. tuberculosis* complex. Differential hybridization arrays identified 14 regions of difference (RD1–14), ranging in size from 2 to 12.7 kb, that were absent from bacillus Calmette–Guérin Pasteur relative to *M. tuberculosis* H37Rv (5, 6). In parallel, six regions, H37Rv related deletions (RvD)1–5, and *M. tuberculosis* specific deletion 1 (TbD1), that were absent from the *M. tuberculosis* H37Rv genome relative to other members of the *M. tuberculosis* complex were revealed by comparative genomics approaches employing pulsed-field gel electrophoresis techniques (5, 7) and *in silico* comparisons of the near complete *M. bovis* AF2122/97 genome sequence and the *M. tuberculosis* H37Rv sequence.

In the present study, we have analyzed the distribution of these 20 variable regions situated around the genome (see Table 1, which is published as supporting information on the PNAS web site, www.pnas.org) in a representative and diverse set of 100 strains belonging to the *M. tuberculosis* complex. The strains were isolated from different hosts and from a broad range of geographic origins, and exhibit a wide spectrum of typing characteristics like IS6110 and spoligotype hybridization patterns or variable-number tandem repeats of mycobacterial interspersed repetitive units (MIRU-VNTR; refs. 8 and 9). We have found striking evidence that deletion of certain variable genomic regions did not occur independently in the different strains of the *M. tuberculosis* complex; assuming that there is little or no recombination of chromosomal segments between the various lineages of the complex, this allows us to propose a completely new scenario for the evolution of the *M. tuberculosis* complex and the origin of human tuberculosis.

Material and Methods

Bacterial Strains. The 100 *M. tuberculosis* complex strains were composed of 46 *M. tuberculosis* strains isolated in 30 countries,

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: TbD1, *M. tuberculosis* specific deletion 1; RvD, H37Rv related deletion; RD, region of difference.

Data deposition: The sequences reported in this paper have been deposited in the EMBL database (accession nos. AJ426486, AJ003103, AJ007301, AJ131210, Y18604, and AJ132559).

††To whom reprint requests should be addressed. E-mail: stcole@pasteur.fr.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

14 *M. africanum* strains, 28 *M. bovis* strains originating in 5 countries, 2 *M. bovis* bacillus Calmette–Guérin vaccine strains (Pasteur and Japan), 5 *M. microti* strains, and 5 *M. canettii* strains. The strains were isolated from human and animal sources and were selected to represent a wide diversity, including 60 strains that have been used in a multicenter study (8). The *M. africanum* strains were retrieved from the collection of the Wadsworth Center (New York State Department of Health, Albany, NY), whereas the majority of the *M. bovis* isolates came from the collection of the University of Zaragoza (Zaragoza, Spain). Four *M. canettii* strains are from the culture collection of the Pasteur Institute (Paris, France). The strains have been extensively characterized by reference typing methods, i.e., IS6110 restriction fragment length polymorphism (RFLP) typing and spoligotyping. *M. tuberculosis* H37Rv, *M. tuberculosis* H37Ra, *M. tuberculosis* CDC1551, *M. bovis* AF2122/97, *M. microti* OV254, and *M. canettii* CIPT 140010059 were included as reference strains. DNA was prepared as described (10).

Genome Comparisons and Primer Design. For preliminary genome comparisons between *M. tuberculosis* and *M. bovis*, web sites <http://genolist.pasteur.fr/TubercuList/> and <http://www.sanger.ac.uk/Projects/M.bovis/>, as well as in-house databases, were used. For primer design, sequences inside or flanking RD and RvD regions were obtained from the same web sites. Primers were designed by using the Primer3 web site <http://www-genome.wi.mit.edu/cgi-bin/primer/primer3-www.cgi> that would amplify ≈500 bp fragments in the reference strains (Table 1).

RD-PCR Analysis. Reactions were performed in 96-well plates and contained per reaction 1.25 μ l of 10 \times PCR buffer (600 mM Tris-HCl, pH 8.8/20 mM MgCl₂/170 mM (NH₄)₂SO₄/100 mM β -mercaptoethanol), 1.25 μ l 20 mM nucleotide mix, 50 nM of each primer, 1–10 ng of template DNA, 10% DMSO, 0.2 units *Taq* polymerase (GIBCO-BRL), and sterile distilled water to 12.5 μ l. Thermal cycling was performed on a PTC-100 amplifier (MJ Research, Cambridge, MA) with an initial denaturation step of 90 s at 95°C, followed by 35 cycles of 30 s at 95°C, 1 min at 58°C, and 4 min at 72°C.

Sequencing of *katG*, *gyrA*, *oxyR*, *pncA*, *mmpL6* Genes and RD, TbD1 Junction Regions. PCR products were obtained as described above, using the primers listed in Table 1.

For primer elimination, 6 μ l of PCR product was incubated with 1 unit of shrimp alkaline phosphatase (United States Biochemical), 10 units of exonuclease I (United States Biochemical), and 2 μ l of 5 \times buffer (200 mM Tris-HCl, pH 8.8/5 mM MgCl₂) for 15 min at 37°C and then for 15 min at 80°C. To this reaction mixture, 2 μ l of Big Dye sequencing mix (Applied Biosystems), 2 μ l (2 μ M) of primer, and 3 μ l of 5 \times buffer (5 mM MgCl₂/200 mM Tris-HCl, pH 8.8) were added, and 35 cycles (96°C for 30 sec, 56°C for 15 sec, and 60°C for 4 min) were performed in a thermocycler (MJ Research, Cambridge, MA). DNA was precipitated by using 80 μ l of 76% ethanol, centrifuged, rinsed with 70% ethanol, and dried. Reactions were dissolved in 2 μ l of formamide/EDTA buffer, denatured and loaded onto 48-cm 4% polyacrylamide gels, and electrophoresis was performed on 377 automated DNA sequencers (Applied Biosystems) for 10 to 12 h. Alternatively, reactions were dissolved in 0.3 mM EDTA buffer and subjected to automated sequencing on a Model 3700 DNA sequencer (Applied Biosystems). Reactions generally gave between 500–700 bp of unambiguous sequence.

Accession Numbers. The sequence of the TbD1 region from the ancestral *M. tuberculosis* strain no. 74 (8) containing genes *mmpS6* and *mmpL6* was deposited in the EMBL database under accession no. AJ426486. Sequences bordering RD4, RD7, RD8,

RD9, and RD10 in bacillus Calmette–Guérin were deposited in the EMBL database under accession nos. AJ003103, AJ007301, AJ131210, Y18604, and AJ132559, respectively.

Results

Variable Genomic Regions and Their Occurrence in the Members of the *M. tuberculosis* Complex. The PCR screening assay for the 20 variable regions (Table 1) within 46 *M. tuberculosis*, 14 *M. africanum*, 5 *M. canettii*, 5 *M. microti*, 28 *M. bovis*, and 2 bacillus Calmette–Guérin strains used oligonucleotides internal to known RDs and RvDs, as well as oligonucleotides flanking these regions (Table 1). This approach generated a large data set that was robust, highly reliable, and internally controlled, because PCR amplicons obtained with the internal primer pair correlated with the absence of an appropriately sized amplicon with the flanking primer-pair, and *vice versa*.

According to the conservation of junction sequences flanking the variable regions, three types of regions were distinguished, each one having a different importance as an evolutionary marker. The first type included mobile genetic elements, like the prophages ϕ Rv1 (RD3) and ϕ Rv2 (RD11) and insertion sequences IS1532 (RD6) and IS6110 (RD5), whose distribution in the tubercle bacilli was highly divergent (see Table 2, which is published as supporting information on the PNAS web site). The second type of deletion is mediated by homologous recombination between adjacent IS6110 insertion elements resulting in the loss of the intervening DNA segment (RvD2, RvD3, RvD4, and RvD5; ref. 7) and is variable from strain to strain (Table 2).

The third type includes deletions whose bordering genomic regions typically do not contain repetitive sequences. Often, this type of deletion occurred in coding regions resulting in the truncation of genes that are still intact in other strains of the *M. tuberculosis* complex. The exact mechanism leading to this type of deletion remains obscure, but possibly rare strand slippage errors of DNA polymerase may have contributed to this event. As shown in detail below, RD1, RD2, RD4, RD7, RD8, RD9, RD10, RD12, RD13, RD14, and TbD1 are representatives of this third group whose distribution among the 100 strains allows us to propose an evolutionary scenario for the members of the *M. tuberculosis* complex that identifies *M. tuberculosis* and/or *M. canettii* as most closely related to the common ancestor of the tubercle bacilli.

***M. tuberculosis* Strains.** Investigation of the 46 *M. tuberculosis* strains by deletion analysis revealed that most RD regions were present in all *M. tuberculosis* strains tested (Table 2). Only regions RD3 and RD11, corresponding to the two prophages ϕ Rv1 and ϕ Rv2 of *M. tuberculosis* H37Rv (4), RD6, containing the insertion sequence IS1532, and RD5, that is flanked by a copy of IS6110 (5), were absent in some strains. This observation is an important one, as it implies that *M. tuberculosis* strains are highly conserved with respect to RD1, RD2, RD4, RD7, RD8, RD9, RD10, RD12, RD13, and RD14, and that these RDs represent regions that can differentiate *M. tuberculosis* strains independently of their geographical origin and their typing characteristics from certain other members of the *M. tuberculosis* complex. Furthermore, this finding suggests that these regions may be involved in the host specificity of *M. tuberculosis*.

In contrast, the presence or absence of RvD regions in *M. tuberculosis* strains was variable. The region that showed the greatest variability was RvD2, because 18 of 46 tested *M. tuberculosis* strains did not carry the RvD2 region. Strains with a high copy number of IS6110 (>14) missed regions RvD2 to RvD5 more often than strains with only a few copies. As an example, all six tested strains belonging to the Beijing cluster (8) lacked regions RvD2 and RvD3. This finding is in agreement

with the proposed involvement of recombination of two adjacent copies of IS6110 in this deletion event (7).

However, the most surprising finding concerning the RvD regions was that TbD1 was absent from 40 of the tested *M. tuberculosis* strains (87%), including representative strains from major epidemics such as the Haarlem, Beijing, and Africa clusters (8). To accentuate this result, we named this region “*M. tuberculosis* specific deletion 1” (TbD1). *In silico* sequence comparison of *M. tuberculosis* H37Rv with the corresponding section in *M. bovis* AF2122/97 revealed that, in *M. bovis*, this locus comprises two genes encoding membrane proteins belonging to a large family, whereas in *M. tuberculosis* H37Rv, one of these genes (*mmpS6*) was absent and the second was truncated (*mmpL6*). Unlike the RvD2-RvD5 deletions, the TbD1 region is not flanked by a copy of IS6110 in *M. tuberculosis* H37Rv, suggesting that insertion elements were not involved in the deletion of the 2,153-bp fragment. To investigate further whether the 40 *M. tuberculosis* strains lacking the TbD1 region had the same genomic organization of this locus as *M. tuberculosis* H37Rv, we amplified the TbD1-junction regions of the various strains by PCR using primers flanking the deleted region (Table 1). This approach showed that the size of the amplicons obtained from multiple strains was uniform (Fig. 1A), and subsequent sequence analysis of the PCR products revealed that in all tested TbD1-deleted strains the sequence of the junction regions was identical to that of *M. tuberculosis* H37Rv (Fig. 1B). The perfect conservation of the junction sequences in TbD1-deleted strains of wide geographical diversity suggests that the genetic event which resulted in the deletion occurred in a common progenitor. However, six *M. tuberculosis* strains, all characterized by very few or no copies of IS6110, and spoligo-types that resembled each other (Fig. 1C), still had the TbD1 region present. Interestingly, these six strains also were clustered together by MIRU-VNTR analysis (9).

Analysis of partial gene sequences of *oxyR*, *pncA*, *katG*, and *gyrA* which have been described as variable between different tubercle bacilli (2, 11–13) revealed that all tested *M. tuberculosis* strains showed *oxyR* and *pncA* partial sequences typical for *M. tuberculosis* [*oxyR*, nucleotide 285 (*oxyR*²⁸⁵):G; *pncA*, codon 57 (*pncA*⁵⁷):CAC]. Based on the *katG* codon 463 (*katG*⁴⁶³) and *gyrA* codon 95 (*gyrA*⁹⁵) sequence polymorphism, Sreevatsan *et al.* (2) defined three groups among the tubercle bacilli: group 1 showed *katG*⁴⁶³ CTG (Leu), *gyrA*⁹⁵ ACC (Thr); group 2 exhibited *katG*⁴⁶³ CGG (Arg), *gyrA*⁹⁵ ACC (Thr); and group 3 showed *katG*⁴⁶³ CGG (Arg), *gyrA*⁹⁵ AGC (Ser). According to this scheme, 16 of the 46 tested *M. tuberculosis* strains in our study belonged to group 1, whereas 27 strains belonged to group 2, and only 3 isolates belonged to group 3. From the 40 strains that were deleted for region TbD1, 9 showed characteristics of group 1, including the strains belonging to the Beijing cluster, 28 showed characteristics of group 2, including the strains from the Haarlem and Africa clusters, and 3 showed characteristics of group 3, including H37Rv and H37Ra. Most interestingly, all six *M. tuberculosis* strains where the TbD1 region was not deleted, contained a leucine (CTG) at *katG*⁴⁶³, which was described as characteristic for ancestral *M. tuberculosis* strains (group 1; ref. 2). As shown in Fig. 2, this finding suggests that during the evolution of *M. tuberculosis*, the *katG* mutation at codon 463 CTG (Leu) → CGG (Arg) occurred in a progenitor strain that had region TbD1 deleted. This proposal is supported by the finding that strains belonging to group 1 may or may not have deleted region TbD1, whereas all 30 strains belonging to groups 2 and 3 lacked TbD1 (Fig. 2). Furthermore, all strains of groups 2 and 3 characteristically lacked spacer sequences 33–36 in the direct repeat (DR) region (Fig. 1C). It appears that such spacers may be lost but not gained (14). Therefore, TbD1-deleted strains will be referred to hereafter as “modern” *M. tuberculosis* strains.

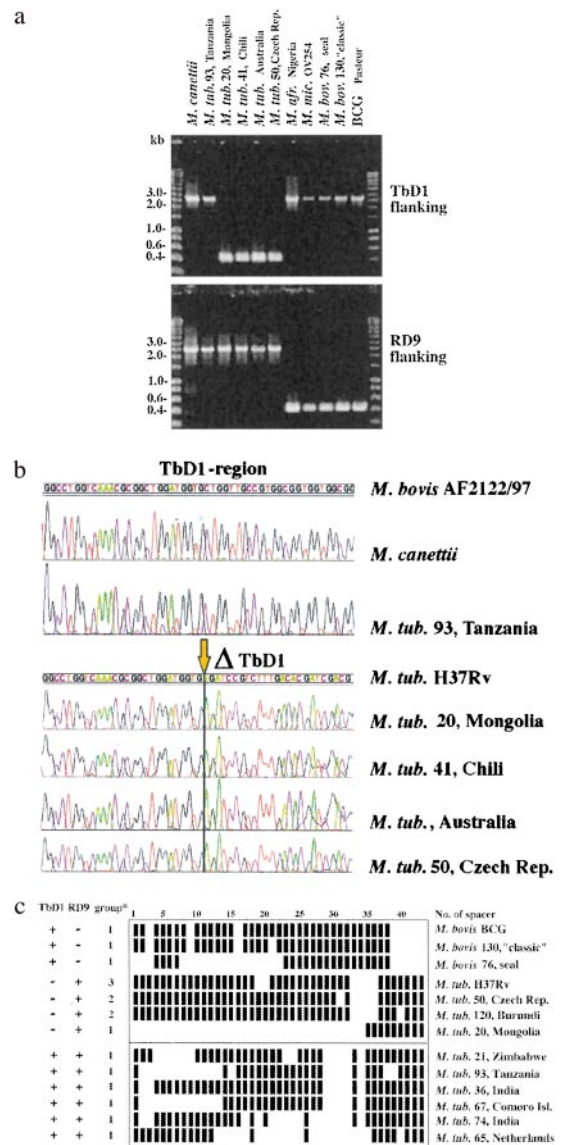


Fig. 1. (A) Amplicons obtained from strains that have the indicated genomic region present or deleted. Sizes of amplicons in each group are uniform. Numbers correspond to strain designation used in refs. 8 and 9. (B) Sequences in the TbD1 region obtained from strains of various geographic regions. Numbers correspond to strain designation used in refs. 8 and 9. (C) Spoligo-types of selected *M. tuberculosis* and *M. bovis* strains. Numbers correspond to strain designation used in refs. 8 and 9. *, groups based on *katG*⁴⁶³/*gyrA*⁹⁵ sequence polymorphism defined by Sreevatsan *et al.* (2).

M. canettii. *M. canettii* is a very rare, smooth variant of *M. tuberculosis* and is usually isolated from patients from, or with connection to, Africa. Although it shares identical 16S rRNA sequences with the other members of the *M. tuberculosis* complex, *M. canettii* strains differ in many respects, including polymorphisms in certain house-keeping genes, *IS1081* copy number, colony morphology, and the lipid content of the cell wall (15, 16). Therefore, we were surprised to find that in *M. canettii*, all of the RD, RvD, and TbD1 regions except the prophages (*phiRv1*, *phiRv2*) were present. In contrast, we identified a region (RD^{can}) being specifically absent from all five *M. canettii* strains that partially overlapped RD12 (Fig. 2).

The conservation of the RD, RvD, and TbD1 regions in the genome of *M. canettii* in conjunction with the many described and observed differences suggest that *M. canettii* diverged from

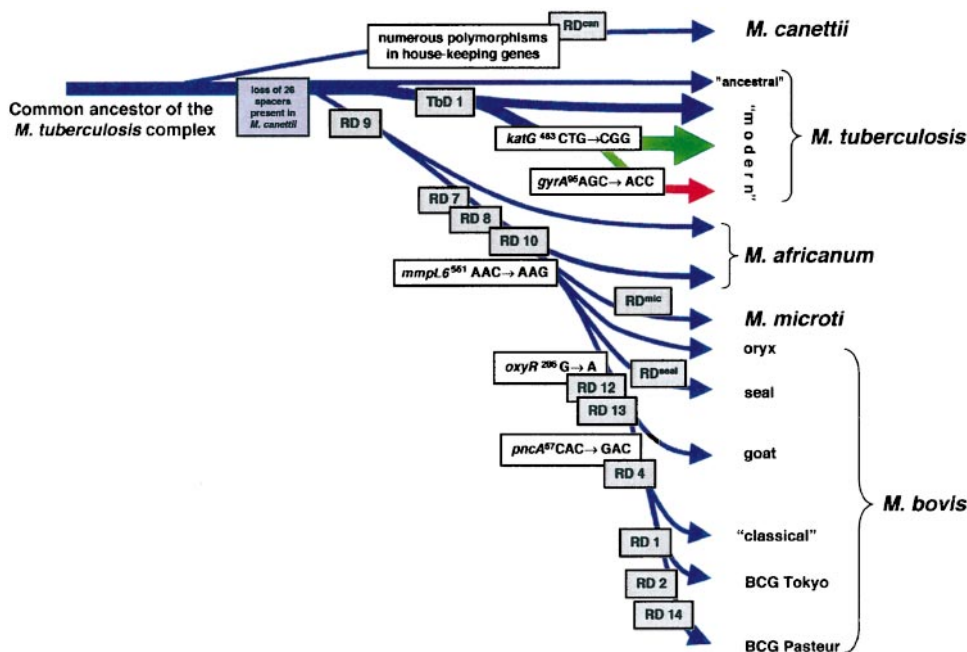


Fig. 2. Scheme of the proposed evolutionary pathway of the tubercle bacilli illustrating successive loss of DNA in certain lineages (gray boxes). The scheme is based on the presence or absence of conserved deleted regions and on sequence polymorphisms in five selected genes. Note that the distances between certain branches may not correspond to actual phylogenetic differences calculated by other methods. Blue arrows indicate that strains are characterized by *katG*⁴⁶³ CTG (Leu), *gyrA*⁹⁵ ACC (Thr), typical for group 1 organisms. Green arrows indicate that strains belong to group 2 characterized by *katG*⁴⁶³ CCG (Arg), *gyrA*⁹⁵ ACC (Thr). The red arrow indicates that strains belong to group 3, characterized by *katG*⁴⁶³ CCG (Arg), *gyrA*⁹⁵ AGC (Ser), as defined by Sreevatsan *et al.* (2).

the common ancestor of the *M. tuberculosis* complex before RD, RvD, and TbD1 occurred in the lineages of tubercle bacilli (Fig. 2). This hypothesis is supported by the finding that *M. canettii* was shown to carry 26 unique spacer sequences in the DR region (14) that are no longer present in any other member of the *M. tuberculosis* complex. Therefore, *M. canettii* represents a fascinating tubercle bacillus, whose detailed genomic analysis may reveal further insights into the evolution of the *M. tuberculosis* complex.

M. africanum. The isolates designated as *M. africanum* studied here originate from West and East African sources. Eleven strains were isolated in Sierra Leone, Nigeria, and Guinea, and 2 strains were isolated in Uganda. One strain comes from the Netherlands.

For the 11 West African isolates, RD analysis indicated that these strains all lack the RD9 region containing *cobL*. Sequence analysis of the RD9 junction region showed that the genetic organization of this locus in West African strains was identical to that of *M. bovis* and *M. microti* in that the 5' part of *cobL* as well as the genes Rv2073c and Rv2074c were absent. In addition, six strains (two from Sierra Leone and four from Guinea) also lacked RD7, RD8, and RD10 (Table 2). The junction sequences bordering RD7, RD8, and RD10, like those for RD9, were identical to those of *M. bovis* and *M. microti* strains. As regards the two prophages phiRv1 and phiRv2, the West African strains all contained phiRv2, whereas phiRv1 was absent. No variability was seen for the RvD regions. RvD1-RvD5 and TbD1 were present in all tested West African strains. This result shows that *M. africanum* prevalent in West Africa can be differentiated from modern *M. tuberculosis* by at least two variable genetic markers, namely the absence of region RD9 and the presence of region TbD1.

In contrast, for East African *M. africanum* and for the isolate from the Netherlands, no genetic marker was found which could differentiate them from *M. tuberculosis* strains. With the excep-

tion of prophage phiRv1 (RD3), the three strains from Uganda and the Netherlands did not exhibit any of the RD deletions, but lacked the TbD1 region, as do modern *M. tuberculosis* strains. The absence of the TbD1 region also was confirmed by sequence analysis of the TbD1 junction region, which was found to be identical to that of TbD1-deleted *M. tuberculosis* strains. These results indicate a very close genetic relationship of these strains to *M. tuberculosis* and suggest that they should be regarded as *M. tuberculosis* rather than *M. africanum* strains.

M. microti. *M. microti* strains were isolated in the 1930s from voles (17) and, more recently, from immuno-suppressed patients (18). These strains are characterized by an identical, characteristic spoligotype, but differ in their IS6110 profiles. Both the vole and the human isolates lacked regions RD7, RD8, RD9, and RD10, as well as a region that is specifically deleted from *M. microti* (RD^{mic}). RD^{mic} was revealed by a detailed comparative genomics study of *M. microti* isolates (P. Brodin, personal communication) using clones from an *M. microti* bacterial artificial chromosome library. RD^{mic} partially overlaps RD1 from bacillus Calmette–Guérin (data not shown). Furthermore, vole isolates missed part of the RD5 region, whereas this region was present in the human isolate. As the junction region of RD5 in *M. microti* was different from that in bacillus Calmette–Guérin (data not shown), RD5 was not used as an evolutionary marker.

***M. bovis* and *M. bovis* Bacillus Calmette–Guérin.** *M. bovis* has a very large host spectrum infecting many mammalian species, including man. The collection of *M. bovis* strains that were screened for the RD and RvD regions consisted of 2 bacillus Calmette–Guérin strains and 18 “classical” *M. bovis* strains that were generally characterized by only one or two copies of IS6110 from bovine, llama, and human sources, in addition to three goat isolates, three seal isolates, two oryx isolates, and two *M. bovis* strains from humans that presented a higher number of IS6110 copies.

Excluding prophages, the distribution of RDs allowed us to differentiate five main groups among the tested *M. bovis* strains. The first group was formed by strains that lack RD7, RD8, RD9, and RD10. Representatives of this group are three seal isolates and two human isolates containing between three and five copies of IS6110 (data not shown). Two oryx isolates harboring between 17 and 20 copies of IS6110 formed the second group that lacked parts of RD5 in addition to RD7–RD10 and very closely resembled the *M. microti* isolates. However, they did not show RD^{mic}, the deletion characteristic of *M. microti* strains (data not shown). Analysis of partial *oxyR* and *pncA* sequences from strains belonging to groups one and two showed sequence polymorphisms characteristic of *M. tuberculosis* strains (*oxyR*²⁸⁵:G, *pncA*⁵⁷:CAC; refs. 12 and 13).

Group three consists of goat isolates that lack regions RD5, RD7, RD8, RD9, RD10, RD12, and RD13. As described by Aranaz *et al.* (20), these strains exhibited an adenosine at position 285 of the *oxyR* pseudogene that is specific for classical *M. bovis* strains, whereas the sequence of the *pncA*⁵⁷ polymorphism was identical to that in *M. tuberculosis*. This finding is in good agreement with our results from sequence analysis (Table 2) and the finding that, except for RD4, the goat isolates displayed the same deletions as classical *M. bovis* strains.

Taken together, this suggests that the *oxyR*²⁸⁵ mutation (G → A) occurred in *M. bovis* strains before RD4 was lost. Interestingly, the most common *M. bovis* strains—classical *M. bovis* (21), isolated from cattle from Argentina, the Netherlands, the UK and Spain, as well as from humans (e.g., multidrug resistant *M. bovis* from Spain; ref. 22)—showed the greatest number of RD deletions and seem to have undergone the greatest loss of DNA relative to other members of the *M. tuberculosis* complex. These lacked regions RD4, RD5, RD6, RD7, RD8, RD9, RD10, RD12, and RD13, confirming results obtained with reference strains (5, 6). These strains, together with the two bacillus Calmette–Guérin strains, were the only ones that showed the *pncA*⁵⁷ polymorphism GAC (Asp) in addition to the *oxyR*²⁸⁵ mutation (G → A) characteristic of *M. bovis*. Analysis of bacillus Calmette–Guérin strains indicate that bacillus Calmette–Guérin lacked the same RD regions as classical *M. bovis* strains, in addition to RD1, RD2, and RD14, which apparently occurred during and after the attenuation process (Fig. 2; refs. 6 and 23).

In contrast to RDs, the RvD regions were highly conserved in the *M. bovis* strains. With the exception of the two IS6110-rich oryx isolates that lacked RvD2, RvD3, and RvD4, all other strains had the five RvD regions present. It is particularly noteworthy that TbD1 was present in all *M. bovis* strains.

However, except for the two human isolates containing between three and five copies of IS6110 from group 1, strains designated as *M. bovis* showed a single nucleotide polymorphism in the TbD1 region at codon 551 (AAG) of the *mmpL6* gene relative to *M. canettii*, *M. africanum*, and ancestral *M. tuberculosis* strains, which are characterized by codon AAC. Even the strains isolated from seals and from oryx with *oxyR* or *pncA* loci like those of *M. tuberculosis*, and with fewer deleted regions than the classical *M. bovis* strains, showed the *mmpL6*⁵⁵¹AAG polymorphism typical for *M. bovis* and *M. microti* (Table 2, Fig. 2). As such, this polymorphism could serve as a very useful genetic marker for the differentiation of strains that lack RD7, RD8, RD9, and RD10 and have been classified as *M. bovis* or *M. africanum* but may differ from other strains of the same taxon.

Discussion

Origin of Human Tuberculosis. For many years, it was thought that human tuberculosis evolved from the bovine disease by adaptation of an animal pathogen to the human host (3). This hypothesis is based on the property of *M. tuberculosis* to be almost exclusively a human pathogen, whereas *M. bovis* has a much broader host range. However, the results from this study

unambiguously show that *M. bovis* has undergone numerous deletions relative to *M. tuberculosis*. This finding is confirmed by the preliminary analysis of the near complete genome sequence of *M. bovis* AF2122/97, a classical *M. bovis* strain isolated from cattle, which revealed no new gene clusters that were confined specifically to *M. bovis*. This result indicates that the genome of *M. bovis* is smaller than that of *M. tuberculosis* (24). It seems plausible that *M. bovis* is the final member of a separate lineage represented by *M. africanum* (RD9), *M. microti* (RD7, RD8, RD9, RD10) and *M. bovis* (RD4, RD5, RD7, RD8, RD9, RD10, RD12, RD13; ref. 25) that branched from the progenitor of *M. tuberculosis* isolates. Successive loss of DNA may have contributed to clonal expansion and the appearance of more successful pathogens in certain new hosts.

Whether the progenitor of extant *M. tuberculosis* strains was already a human pathogen when the *M. africanum* → *M. bovis* lineage separated from the *M. tuberculosis* lineage is a subject for speculation. However, we have two reasons to believe that this was the case. First, the six ancestral *M. tuberculosis* strains (TbD1⁺, RD9⁺; Fig. 1C) that resemble the last common ancestor before the separation of *M. tuberculosis* and *M. africanum* are all human pathogens. Second, *M. canettii*, which probably diverged from the common ancestor of today's *M. tuberculosis* strains before any other known member of the *M. tuberculosis* complex, is also a human pathogen. Taken together, this means that those tubercle bacilli, which are thought to most closely resemble the progenitor of *M. tuberculosis*, are human and not animal pathogens. It is also intriguing that most of these strains were of African or Indian origin (Fig. 1C). It is likely that these ancestral strains predominantly originated from endemic foci (15, 26), whereas modern *M. tuberculosis* strains that have lost TbD1 may represent epidemic *M. tuberculosis* strains that were introduced into the same geographical regions more recently as a consequence of the worldwide spread of the tuberculosis epidemic.

The Evolutionary Time Scale of the *M. tuberculosis* Complex. Because of the high sequence conservation in housekeeping genes, Sreevatsan *et al.* (2) hypothesized that the tubercle bacilli encountered a major bottleneck 15,000–20,000 years ago. As the conservation of the TbD1 junction sequence in all tested TbD1-deleted strains suggests a descent from a single clone, the TbD1 deletion is a perfect indicator that modern *M. tuberculosis* strains that account for the vast majority of today's tuberculosis cases definitely underwent such a bottleneck and then spread around the world.

As described in detail in *Results*, our analysis showed that the *katG*(463) CTG → CGG and the subsequent *gyrA*(95) ACC → AGC mutations that were used by Sreevatsan *et al.* to designate groups 2 and 3 of their proposed evolutionary pathway of the tubercle bacilli (2) occurred in a lineage of *M. tuberculosis* strains that had already lost TbD1 (Fig. 2). Although deletions are more stable markers than point mutations, which may be subject to reversion, a perfect correlation of deletion and point mutation data were found for the tested strains.

This information, together with results from a recent study by Fletcher *et al.* (27), who have shown that *M. tuberculosis* DNAs amplified from naturally mummified Hungarian villagers from the 18th and 19th century belonged to *katG*⁴⁶³/*gyrA*⁹⁵ groups 2 and 3, suggests that the TbD1 deletion occurred in the lineage of *M. tuberculosis* before the 18th century. This could mean that the dramatic increase of tuberculosis cases later in the 18th century in Europe mainly involved modern *M. tuberculosis* strains. In addition, it shows that tuberculosis was caused by *M. tuberculosis* and not by *M. bovis*, a fact which is also described for cases in rural medieval England (28).

There is good evidence that mycobacterial infections occurred in man several thousand years ago. We know that tuberculosis

occurred in Egypt during the reign of the pharaohs because spinal and rib lesions pathognomonic of tuberculosis have been identified in mummies from that period (29). Identification of acid-fast bacilli, as well as PCR amplification of IS6110 from Peruvian mummies (30), also suggest that tuberculosis existed in preColumbian societies of Central and South America. To estimate when the TbD1 bottleneck occurred, it would be helpful to know whether the Egyptian and South American mummies carried *M. tuberculosis* DNA that did or did not have TbD1 deleted.

The other major bottleneck, which seems to have occurred for members of the *M. africanum* → *M. microti* → *M. bovis* lineage, is reflected by RD9 and the subsequent RD7, RD8, and RD10 deletions (Fig. 2). These deletions seem to have occurred in the progenitor of tubercle bacilli that—today—show natural host spectra as diverse as humans in Africa, voles on the Orkney Isles (UK), seals in Argentina, goats in Spain, and badgers in the UK. For this reason, it is difficult to imagine that the spread and adaptation of RD9-deleted bacteria to their specific hosts could have appeared within the postulated 15,000–20,000 years of speciation of the *M. tuberculosis* complex.

However, more insight into this matter could be gained by RD analysis of ancient DNA samples, e.g., mycobacterial DNA isolated from a 17,000-year-old bison skeleton (19). The *Mycobacterium* whose DNA was amplified showed a spoligotype that was most closely related to patterns of *M. africanum* and could have been an early representative of the lineage *M. africanum* → *M. bovis*. With the TbD1 and RD9 junction sequences that we supply here, PCR analyses of ancient DNAs should enable very focused studies to be undertaken to learn more about the time scale within which the members of the *M. tuberculosis* complex have evolved.

Concluding Comments. Our study provides an overview of the diversity and conservation of variable regions in a broad range of tubercle bacilli. Deletion analysis of 100 strains from various hosts and countries has identified some evolutionarily “old” *M. canettii*, *M. tuberculosis*, and *M. africanum* strains, most of them of African origin, as well as modern *M. tuberculosis* strains, the latter including representatives from major epidemic clusters like Beijing, Haarlem, and Africa. The use of deletion analysis in conjunction with molecular typing and analysis of specific mutations was shown to represent a very powerful approach for the study of the evolution of the tubercle bacilli and for the identification of evolutionary markers. In a more practical perspective, these regions, primarily RD9 and TbD1 but also RD1, RD2, RD4, RD7, RD8, RD10, RD12, and RD13 represent very interesting candidates for the development of powerful diagnostic tools for the rapid and unambiguous identification of members of the *M. tuberculosis* complex. This genetic approach for differentiation can now be used to replace the often confusing traditional division of the *M. tuberculosis* complex into rigidly defined subspecies.

Moreover, functional analyses will show whether the TbD1 deletion confers some selective advantage to modern *M. tuberculosis*, or whether other circumstances contributed to the pandemic of the TbD1-deleted *M. tuberculosis* strains.

We thank Philip Supply for his encouragement and help to initiate this study. We also thank Carlos Martin, Tim Stinear, and Veronique Vincent for fruitful discussions. This work was supported by the Institut Pasteur (PTR 2000/35), the Programme de Recherche Fondamentale en Microbiologie et Maladies Infectieuses, the European Union (QLK2-CT-1999-01093, QLRT-CT-2000-00630), the Department of Environment, Food and Rural Affairs (GB), the Association Française Raoul Follereau, and The Wellcome Trust.

- Boddinghaus, B., Rogall, T., Flohr, T., Blocker, H. & Bottger, E. C. (1990) *J. Clin. Microbiol.* **28**, 1751–1759.
- Sreevatsan, S., Pan, X., Stockbauer, K. E., Connell, N. D., Kreiswirth, B. N., Whittam, T. S. & Musser, J. M. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 9869–9874.
- Stead, W. W., Eisenach, K. D., Cave, M. D., Beggs, M. L., Templeton, G. L., Thoen, C. O. & Bates, J. H. (1995) *Am. J. Respir. Crit. Care Med.* **151**, 1267–1268.
- Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E., 3rd, et al. (1998) *Nature (London)* **393**, 537–544.
- Gordon, S. V., Brosch, R., Billault, A., Garnier, T., Eiglmeier, K. & Cole, S. T. (1999) *Mol. Microbiol.* **32**, 643–655.
- Behr, M. A., Wilson, M. A., Gill, W. P., Salamon, H., Schoolnik, G. K., Rane, S. & Small, P. M. (1999) *Science* **284**, 1520–1523.
- Brosch, R., Philipp, W. J., Stavropoulos, E., Colston, M. J., Cole, S. T. & Gordon, S. V. (1999) *Infect. Immun.* **67**, 5768–5774.
- Kremer, K., van Soolingen, D., Frothingham, R., Haas, W. H., Hermans, P. W., Martin, C., Palittapongarnpim, P., Plikaytis, B. B., Riley, L. W., Yakrus, M. A., et al. (1999) *J. Clin. Microbiol.* **37**, 2607–2618.
- Supply, P., Lesjean, S., Savine, E., Kremer, K., van Soolingen, D. & Locht, C. (2001) *J. Clin. Microbiol.* **39**, 3563–3571.
- van Soolingen, D., de Haas, P. E. W., Hermans, P. W. M. & van Embden, J. D. A. (1994) *Methods Enzymol.* **235**, 196–205.
- Heym, B., Honore, N., Truffot-Pernot, C., Banerjee, A., Schurra, C., Jacobs, W. R., Jr., van Embden, J. D., Grosset, J. H. & Cole, S. T. (1994) *Lancet* **344**, 293–298.
- Scorpio, A., Collins, D., Whipple, D., Cave, D., Bates, J. & Zhang, Y. (1997) *J. Clin. Microbiol.* **35**, 106–110.
- Sreevatsan, S., Escalante, P., Pan, X., Gillies, D. A., 2nd, Siddiqui, S., Khalaf, C. N., Kreiswirth, B. N., Bifani, P., Adams, L. G., Ficht, T., et al. (1996) *J. Clin. Microbiol.* **34**, 2007–2010.
- van Embden, J. D., van Gorkom, T., Kremer, K., Jansen, R., van Der Zeijst, B. A. & Schouls, L. M. (2000) *J. Bacteriol.* **182**, 2393–2401.
- van Soolingen, D., Hoogenboezem, T., de Haas, P. E., Hermans, P. W., Koedam, M. A., Teppema, K. S., Brennan, P. J., Besra, G. S., Portaels, F., Top, J., et al. (1997) *Int. J. Syst. Bacteriol.* **47**, 1236–1245.
- Papa, F., Laszlo, A., David, H. L. & Daffe, M. (1989) *Acta Leprol.* **7**, Suppl. 1, 98–101.
- Wells, A. Q. Suppl. (1937) *Lancet*, 1221.
- van Soolingen, D., van der Zanden, A. G., de Haas, P. E., Noordhoek, G. T., Kiers, A., Foudraire, N. A., Portaels, F., Kolk, A. H., Kremer, K. & van Embden, J. D. (1998) *J. Clin. Microbiol.* **36**, 1840–1845.
- Rothschild, B. M., Martin, L. D., Lev, G., Bercovier, H., Bar-Gal, G. K., Greenblatt, C., Donoghue, H., Spigelman, M. & Brittain, D. (2001) *Clin. Infect. Dis.* **33**, 305–311.
- Aranaz, A., Liebana, E., Gomez-Mampaso, E., Galan, J. C., Cousins, D., Ortega, A., Blazquez, J., Baquero, F., Mateos, A., Suarez, G. & Dominguez, L. (1999) *Int. J. Syst. Bacteriol.* **49**, 1263–1273.
- van Soolingen, D., de Haas, P. E. W., Haagsma, J., Eger, T., Hermans, P. W. M., Ritacco, V., Alito, A. & van Embden, J. D. A. (1994) *J. Clin. Microbiol.* **32**, 2425–2433.
- Samper, S., Martin, C., Pinedo, A., Rivero, A., Blazquez, J., Baquero, F., van Soolingen, D. & van Embden, J. (1997) *Aids* **11**, 1237–1242.
- Mahairas, G. G., Sabo, P. J., Hickey, M. J., Singh, D. C. & Stover, C. K. (1996) *J. Bacteriol.* **178**, 1274–1282.
- Gordon, S. V., Eiglmeier, K., Garnier, T., Brosch, R., Parkhill, J., Barrell, B., Cole, S. T. & Hewinson, R. G. (2001) *Tuberculosis* **81**, 157–163.
- Brosch, R., Gordon, S. V., Eiglmeier, K., Garnier, T., Tekaia, F., Yeramian, E. & Cole, S. T. (1999) in *Molecular Genetics of Mycobacteria*, eds Hatful, G. F. & Jacobs, W. R., Jr. (Am. Soc. Microbiol., Washington, DC), pp. 19–36.
- Radhakrishnan, I., K. M. Y., Kumar, R. A. & Mundayoor, S. (2001) *J. Clin. Microbiol.* **39**, 1683.
- Fletcher, H. A., Donoghue, H. D., Holton, J., Pap, I. & Spigelman, M. (2002) *Am. J. Phys. Anthropol.*, in press.
- Mays, S., Taylor, G. M., Legge, A. J., Young, D. B. & Turner-Walker, G. (2001) *Am. J. Phys. Anthropol.* **114**, 298–311.
- Nerlich, A. G., Haas, C. J., Zink, A., Szeimies, U. & Hagedorn, H. G. (1997) *Lancet* **350**, 1404.
- Salo, W. L., Aufderheide, A. C., Buikstra, J. & Holcomb, T. A. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 2091–2094.

Supporting information for Brosch *et al.* (March 12, 2002) *Proc. Natl. Acad. Sci. USA*,

10.1073/pnas.052548299

Table 1: RD, RvD and TbD1 regions

Region absent from BCG	Gene	Size kb	Internal primer pair	Flanking primers or 2 nd internal* primer pair
RD1	Rv3871-Rv3879c	9.5	RD1in-Rv3878F GTC AGC CAA GTC AGG CTA CC	RD1-flank.left GAA ACA GTC CCC AGC AGG T
			RD1in-Rv3878R CAA CGT TGT GGT TGT TGA GG	RD1-flank.right TTC AAC GGG TTA CTG CGA AT
RD2	Rv1978-Rv1988	10.8	RD2-Rv1979.int.F TAT AGC TCT CGG CAG GTT CC	RD2-flank.F CTC GAC CGC GAC GAT GTG C
			RD2-Rv1979.int.R ATC GGC ATC TAT GTC GGT GT	RD2-flank.R CCT CGT TGT CAC CGC GTA TG
RD3*	Rv1573-Rv1586c	9.2	RD3-Rv1586.int.F TTA TCT TGG CGT TGA CGA TG	RD3-int-REP.F CTG ACG TCG TTG TCG AGG TA*
			RD3-Rv1586.int.R CAT ATA AGG GTG CCC GCT AC	RD3-int-REP.R GTA CCC CCA GGC GAT CTT*
RD4	Rv1505c-Rv1516c	12.7	RD4-Rv1516.int.F CAA GGG GTA TGA GGT TCA CG	RD4-flank.F CTC GTC GAA GGC CAC TAA AG
			RD4-Rv1516.int.R CGG TGA TTC GTG ATT GAA CA	RD4-flank.R AAG GCG AAC AGA TTC AGC AT
RD5*	Rv2346c-Rv2353c	9.0	RD5A-Rv2348.int.F AAT CAC GCT GCT GCT ACT CC	RD5B-plcA.int.F CAA GTT GGG TCT GGT CGA AT
			RD5A-Rv2348.int.R GTG CTT TTG CCT CTT GGT C	RD5B-plcA.int.R GCT ACC CAA GGT CTC CTG GT
RD6*	Rv3425-Rv3428c	4.9	RD6-IS1532F CAG CTG GTG AGT TCA AAT GC	ND
			RD6-IS1532R CTC CCG ACA CCT GTT CGT	ND
RD7	Rv1964-Rv1977	12.7	RD7-Rv1976.int.F TGG ATT GTC GAC GGT ATG AA	RD7-flank.F GGT AAT CGT GGC CGA CAA G
			RD7-Rv1976.int.R GGT CGA TAA GGT CAC GGA AC	RD7-flank.R CAG CTC TTC CCC TCT CGA C
RD8	<i>ephA-lpqG</i>	5.9	RD8-ephA.F GGT GTG ATT TGG TGA GAC GAT G	RD8-flank.F CAA TCA GGG CTG TGC TAA CC
			RD8-ephA.R AGT TCC TCC TGA CTA ATC CAG GC	RD8-flank.R CGA CAG TTG TGC GTA CTG GT
RD9	<i>cobL</i> -Rv2075	2.0	RD9-intF CGA TGG TCA ACA CCA CTA CG	RD9-flankF GTG TAG GTC AGC CCC ATC C
			RD9-intR CTG GAC CTC GAT GAC CAC TC	RD9-flankR GCC CAA CAG CTC GAC ATC
RD10	Rv0221-Rv0223	1.9	RD10-intF GTA ACC GCT TCA CCG GAA T	RD10-flankF CTG CAA CCA TCC GGT ACA C
			RD10-intR GTC AAC TCC ACG GAA AGA CC	RD10-flankR GTC ATG AAC GCC GGA CAG

RD11	Rv2645-Rv2659c	11.0	RD11-Rv2646F CGG CAG CTA GAC GAC CTC	RD11-fla-F TCA CAT AGG GGC TGC GAT AG
			RD11-Rv2646R AAC GTG CTG CGA TAG GTT TT	RD11-fla-R AGA GGA ACC TTT CGG TGG TT
RD12	sseC-Rv3121	2.8	RD12-Rv3120.int.F GAA ATA CGA GTG CGC TGA CC	RD12-flank.F GCC ATC AAC GTC AAG AAC CT
			RD12-Rv3120.int.R CTC TGA ACC ATC GGT GTC G	RD12-flank.R CGG CCA GGT AAC AAG GAG T
RD13	Rv1255c-Rv1257c	3.0	RD13intF GGA TGT CAC TCG GAA CGG CA	RD13-flank.F CGA TGG TGT TTC TTG GTG AG
			RD13intR CAC CGG GCT GAT CGA GCG A	RD13-flank.R GGA TCG GCT CAG TGA ATA CC
RD14	Rv1765c-Rv1773c	9.0	RD14-Rv1769.int.F GTG GAG CAC CTT GAC CTG AT	RD14-flankF TTG ATT CGC CAA CAA CTG AA
			RD14-Rv1769.int.R CGT CGA ATA CGA GTC GAA CA	RD14-flankR GGG CTG GTT AGT GTC GAT TC
Region missing from <i>M. tuberculosis</i> H37Rv				
RvD1*		5.0	RvD1-int1F AGC GCG TCG AAC ACC GGC	RvD1-int2.F GAG CCA CTC CGA TGT TGA CT
			RvD1-int1R CCT GAA TCC GCG CAA TTC CAT	RvD1-int2.R CAC GCG AAC CCT ACC TAC AT
RvD2*	plcD	5.1	RvD2-int1F GTT CTC CTG TCG AAC CTC CA	RvD2-int2F GGA CGG TGA CGG TAT TTG TC
			RvD2-int1R ACT TCA CCG GTT TCA TCT CG	RvD2-int2R TCG CCA ACT TCT ATG GAC CT
RvD3		1.0	RvD3-intF ATC GAT CAG GTC GTC AAT GC	RvD3-flank.F AAA CCA TGC AGC GTC TGC CA
			RvD3-intR ACG CCA CCA TCA AGA TCC	RvD3-flankR GCG TTT CTG CGT CTG GTT GA
RvD4*	PPE gene	0.8	RvD4-intF-PPE GGT TGC CAA CGT TAC CGA TGC	ND
			RvD4-intR-PPE CCG GTG GTG GTG GCG GCT	ND
RvD5	moa	4.0	RvD5intF GGG TTC ACG TTC ATT ACT GTT C	RvD5-flankF CCC ATC GTG GTC GTT CAC C
			RvD5intR CCT GCG CTT ATC TCT AGC GG	RvD5-flankR GTA CCC GCA CCA CCT GCT G
TbD1	mmpS6 mmpL6	2.1	TbD1intS.F CGT TCA ACC CCA AAC AGG TA	TbD1fla1-F CTA CCT CAT CTT CCG GTC CA
			TbD1intS.R AAT CGA ACT CGT GGA ACA CC	TbD1fla1-R CAT AGA TCC CGG ACA TGG TG

<i>katG</i> , <i>gyrA</i> , <i>oxyR</i> ['] , <i>pncA</i> and <i>mmpL6</i>	PCR	and sequencing primers
<i>katG</i> ⁴⁶³	<i>katG</i> -2154,225-PCR-F CTA CCA GCA CCG TCA TCT CA	<i>katG</i> -2154,872-SEQ-R ACA AGC TGA TCC ACC GAG AC
	<i>katG</i> -2155,157-PCR-R AGG TCG TAT GGA CGAACA CC	
<i>gyrA</i> ⁹⁵	<i>gyrA</i> -7,127-PCR-F GTT CGT GTG TTG CGT CAA GT	<i>gyrA</i> -7,461F CGG GTG CTC TAT GCA ATG TT
	<i>gyrA</i> - 8,312-PCR-R CAG CTG GGT GTG CTT GTA AA	
<i>oxyR</i> ²⁸⁵	<i>oxyR</i> 2725,559F TAT GCG ATC AGG CGT ACT TG	<i>oxyR</i> -2726,024-SEQ-R CAA AGC AGT GGT TCA GCA GT
	<i>oxyR</i> -2726,024-PCR-R CAA AGC AGT GGT TCA GCA GT	
<i>pncA</i> ⁵⁷	<i>pncA</i> -2288,678-PCR-F ATC AGG AGC TGC AAA CCA AC	<i>pncA</i> - 2289,319-SEQ-R GGC GTC ATG GAC CCT ATA TC
	<i>pncA</i> - 2289,319-PCR-R GGC GTC ATG GAC CCT ATA TC	
<i>mmpL6</i> ⁵⁵¹	<i>mmpL</i> -seq5F GTA TCA GAG GGA CCG AGC AG	<i>mmpL</i> -seq5F GTA TCA GAG GGA CCG AGC AG
	TBD1fla1-R CAT AGA TCC CGG ACA TGG TG	

The RD nomenclature used in this table is based on that used by Brosch *et al.* (1) and differs from that proposed by Behr *et al.* (2). Primer sequences are shown in 5' →3' direction.

More information on the exact location of the deleted regions can be obtained at <http://www.pasteur.fr/recherche/unites/Lgmb/>.

* Regions where a second pair of internal primers was used rather than flanking primers, because of flanking repetitive regions, and/or mobile genetic elements.

References

1. Brosch, R., Gordon, S. V., Eiglmeier, K., Garnier, T., Tekaia, F, Yeramian, E., & Cole, S.T. (2000) in *Molecular Genetics of Mycobacteria*, eds. Hatfull, G. F., & Jacobs, W. R., Jr. (Am. Soc. Microbiol., Washington, DC), pp. 19-36.
2. Behr, M. A., Wilson, M. A., Gill, W. P., Salamon, H., Schoolnik, G. K., Rane, S. & Small, P. M. (1999) *Science* **284**, 1520-1523.

EDITORS' CHOICE

edited by Gilbert Chin

GEOPHYSICS

Fault, Heal Thyself

Large earthquakes are known to recur on major faults, but before the next earthquake can happen, the previous rupture must first heal. Determining the amount of time for the new fractures (generated by shear faulting during the main shock) to disappear and for the tectonic stress cracks to reappear would facilitate hazard assessment along active faults.

Tadokoro and Ando used shear wave anisotropy analysis to show that the Nojima fault

in Japan, site of the 1995 Hyogo-ken Nanbu earthquake (moment magnitude = 6.9), has healed in just 33 months. This rapid healing suggests that communities may need to remain on alert for further seismicity even soon after a recent event. — LR

Geophys. Res. Lett. **29**, 10.1029/2001GL013644 (2002).

CLIMATE SCIENCE

Mercury at Dawn

The toxic chemical mercury is extremely mobile in its gaseous elemental form and is transported globally due to the

long atmospheric residence time of Hg^0 . High concentrations of mercury have been found in the Arctic, but its accumulation at those latitudes cannot be explained by the same depositional processes used to account for the presence of organic toxins in the same areas, because Hg^0 does not condense at the ambient atmospheric temperatures. Therefore, some other mechanism must be responsible.

Lindberg *et al.* measured concentrations of atmospheric mercury at Point Barrow, Alaska (71° N), from 1998 to 2001 and found

that rapid photochemical oxidation of boundary-layer Hg^0 after polar sunrise causes deposition of oxidized gaseous mercury. Depletion of atmospheric Hg^0 began within a few days of polar sunrise and continued until snowmelt, and was correlated with variations in ozone concentration during these months. They attribute this behavior to rapid in-situ oxidation of gaseous Hg^0 by the same photochemically active halogen species involved in surface O_3 destruction: the halogen radicals Br and Cl and the halogen oxide radicals BrO and ClO. Studies of bioaccumulation of Hg in seabirds and marine mammals suggest that this depositional process has operated only for the past few decades and could be the result of changes in Arctic climate that have increased atmospheric transport of photooxidants and local production of reactive halogens. — HJS

Environ. Sci. Technol. **36**, 1245 (2002).

BIOCHEMISTRY

Pausing Signals

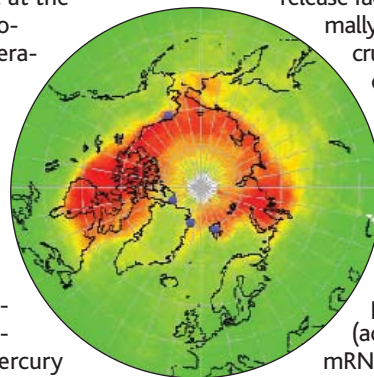
The curious story of tmRNA has taken an intriguing turn. When a translating ribosome reaches the end of an aberrantly truncated messenger RNA (mRNA), it stalls and becomes trapped, waiting in vain for the ribosome

release factors that normally would be recruited by the stop codon. Enter tmRNA, which first binds to the ribosome (acting as a transfer RNA), and then encodes a short peptide segment (acting as a mRNA) that is attached to the end of the truncated protein and tags it for

degradation by intracellular proteases. Hayes *et al.* have found that tmRNA also gets involved when the ribosome slows down as a consequence of encountering a rarely used arginine codon just in front of the stop codon. Apparently, this pause results in about 20% of correctly synthesized protein (in this instance, RbsK) being tagged and subsequently destroyed. How can this wastage be beneficial?

Nakatogawa and Ito, in a pair of papers examining the peculiar fate of the secretion monitor SecM, may have an answer. At first glance, it appears that SecM is made simply to be exported to the periplasm and degraded. A closer look has shown that a proline-containing segment near to the COOH-terminus of SecM interacts with the exit tunnel of the ribosome and serves to slow translation. This pause allows time for a secondary structural element encompassing the Shine-Dalgarno sequence in the SecA gene (which lies immedi-

CONTINUED ON PAGE 2329



Elevated BrO concentrations (red) in April 2000.

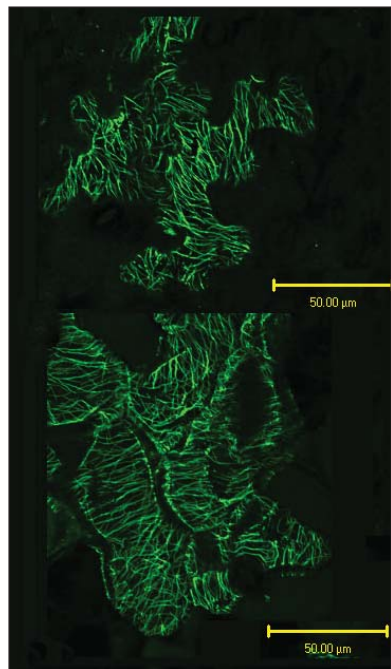
PLANT BIOLOGY

The Long and the Short of It

Making leaves is key to a plant's development, and the shape and size of leaves are carefully controlled by genetic pathways. Kim *et al.* and Folkers *et al.* have studied the role of the gene *ANGUSTIFOLIA* (*AN*) in leaf morphogenesis. Mutations in the *AN* gene yield leaves that are narrower and longer than normal. The *AN* protein is related to carboxyl-terminal binding protein (CtBP), which is a transcriptional regulator involved in differentia-

tion pathways in mammals. In developing leaf cells in *an* plants, the orientation of cortical microtubules was disrupted: Microtubules were generally aligned parallel to the width of the leaf instead of being arranged more or less randomly. This restricted distribution would be expected to constrain leaf cell expansion. Folkers *et al.* also observed a genetic and physical interaction between *AN* and a kinesin-like, microtubule-based molecular motor known as ZWICHEL. The differentiation-specific regulation of microtubule distribution in this system may provide a paradigm for morphogenesis pathways in other systems. — SMH

EMBO J. **21**, 1267; 1280 (2002).



Unlike the random pattern in wild-type leaves (upper), microtubules (green) in *an* cells (lower) run across the leaf.

ately downstream from the *SecM* gene) to unfold. Furthermore, this pause only transpires when there is not enough SecA protein to bind to the NH₂-terminal portion of SecM, which acts to overcome the frictional pull of the exit tunnel. Thus, SecM regulates the cellular levels of SecA (which mediates general protein export to the periplasm) by signaling to the ribosome: If more SecA is needed, slow down so that the way ahead can be cleared of obstacles; if there is enough SecA, finish making SecM and do something else. The close spacing of bacterial genes—the start site of *rbsR* is just three nucleotides away from the end of *rbsK*—prompts the proposal that rare arginine codons may offer another mechanism for coordinate gene expression. — GJC

Proc. Natl. Acad. Sci. U.S.A. **99**, 3440 (2002);
Cell **108**, 629 (2002); *Mol. Cell* **7**, 185 (2001).

CHEMISTRY

Like an Enzyme

Ideally, chemical synthesis would be performed catalytically at low temperature and pressure, with few side products and without polluting solvents. But in reality, aerobic oxidations are particularly tricky because catalysts themselves often become oxidized, limiting their lifetime. Bench *et al.* have overcome this problem by synthesizing a homogeneous catalyst that contains a heme-like metal center protected within a Teflon-coated sterically restrictive cavity. This pseudoenzyme complex catalyzes the oxidative formation of carbon-phosphorus double bonds, as demonstrated by the coupling of phosphanes with acetone to produce synthetically utile ylides (Wittig reagents) at ambient conditions in air. No decomposition of the catalyst was observed after numerous catalytic cycles. With further modification, this new class of catalysts should be applicable to other substrates, replacing less efficient stoichiometric syntheses. — JU

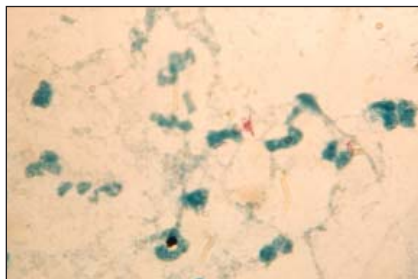
Angew. Chem. Int. Ed. **41**, 750 (2002).

MICROBIOLOGY

No Coughing in the Cattle Shed

Tradition has it that *Mycobacterium tuberculosis*, the causative agent of tuberculosis, evolved from *M. bovis*, which displays a broad host spectrum and is thought to have colonized human beings as our ancestors domesticated cattle. Brosch *et al.* used deletion analysis to locate 20 variable regions found among the otherwise very similar (99.9%) genomes of the tubercle group bacilli. They discovered that the polymorphisms have not

evolved independently but appear to have resulted from a few ancient events in ancestral strains. A corollary finding is that these regions can serve as diagnostic markers for different strains; one in particular, called TbD1, is specific to *M. tuberculosis* and dis-



Clumps of *M. tuberculosis* (red) surrounded by human cells (blue).

criminate between ancient and modern pandemic strains of the pathogen, suggesting that the ancestral strain was African or Asian. Progressive mapping of the deletions indicated that *M. bovis* evolved from *M. tuberculosis* and that its genome is smaller than that of *M. tuberculosis*. Previously, Cole *et al.* had shown that considerable gene decay has occurred in a related, specialist human pathogen, *M. leprae*, which since its divergence from their last common ancestor appears to have lost 2000 genes. — CA

Proc. Natl. Acad. Sci. U.S.A. **99**, 3684 (2002);
Nature **409**, 1007 (2001).

APPLIED PHYSICS

Seeing Beneath the Surface

The presence of defects in optical-electronic materials can detrimentally affect the output efficiency of light-emitting diodes. Pinpointing the locations of defects is therefore important for characterizing semiconductor crystals and is a first step toward countering their effects or improving the synthetic process. Illuminating defects with short-wavelength light and observing the subsequent photoluminescence is the routine analytical method; however, many of the defects are buried deep below the crystal surface and out of reach of short-wavelength photons, which are absorbed in the surface region. Kawata *et al.* use weakly absorbed long-wavelength, or sub-bandgap, light to illuminate their semiconductor sample, but focus the beam to a sharp spot within the sample so that the defects can be excited via two-photon absorption. The ability to control where the beam is focused provides the ability to create a three-dimensional image of the defects throughout the sample. — ISO

Opt. Lett. **27**, 297 (2002).