Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 3410–3415, April 1996
Microbiology

Homing events in the *gyrA* gene of some mycobacteria

(GyrA/exteins/inteins/protein splicing)

HAFIDA FSIHI*, VERONIQUE VINCENT†, AND STEWART T. COLE*‡

*Unité de Génétique Moléculaire Bactérienne and †Laboratoire de Référence des Mycobactéries, Institut Pasteur, 28 rue du Dr. Roux, F-75724 Paris cedex 15, France

Communicated by Werner Arber, University of Basel, Basel, Switzerland, December 18, 1995 (received for review June 14, 1995)

ABSTRACT **The A subunit of DNA gyrase in *Mycobacterium leprae*, unlike its counterpart in *Mycobacterium tuberculosis*, is produced by protein splicing as its gene, *gyrA*, harbors a 1260-bp in-frame insertion encoding an intein, a putative homing endonuclease. Analysis of the *gyrA* locus from different mycobacterial species revealed the presence of inteins in *Mycobacterium flavescens, Mycobacterium gordonae*, and *Mycobacterium kansasii* but not in 10 other pathogenic or saprophytic mycobacteria. In all four cases where intein coding sequences were found, they were localized in the same position in *gyrA*, immediately downstream of the codon for the key active-site residue Tyr-130. The intein products were similar, but not identical, in sequence and the splice junctions displayed all the features found in other polypeptides known to be produced by protein splicing from a precursor protein. Paired motifs, found in homing endonucleases encoded by some group I RNA introns, and inteins showing endonuclease activity, were present in the *gyrA* inteins as were other intein-specific signatures. Some strains of *M. flavescens, M. gordonae*, and *M. kansasii* were shown by PCR analysis to have inteinless *gyrA* genes, in contrast to the situation in *M. leprae* where all the isolates possessed insertions in *gyrA*. Sequencing of the corresponding regions revealed that, although the GyrA protein sequence was conserved, the nucleotide sequences differed in *gyrA* genes with and without inteins, suggesting that the homing endonuclease displays sequence specificity.**

In 1990, revision of one of the dogmas of molecular biology was required when it was found that splicing can occur at the protein level and that a single gene can encode a precursor protein, a mature protein, and a free intein (for reviews, see refs. 1–4). The term intein refers to a protein sequence that is excised from the precursor protein during maturation; its coding sequence is always inserted in-frame within a protein-coding gene (5). This process has been described for the vacuolar ATPase from *Saccharomyces cerevisiae* (6) and *Candida tropicalis* (7) and the DNA polymerase of *Thermococcus litoralis* (8, 9) and *Pyrococcus* sp. (5). In mycobacteria, the first case of excision of an internal segment of a polypeptide and religation of the flanking peptides, the N-extein and C-extein, to create a functional protein was discovered in the RecA protein from *Mycobacterium tuberculosis* and later from *Mycobacterium leprae* (10–12).

Several potential biochemical mechanisms for protein splicing have been proposed and these generally involve nucleophilic attack on the peptide bond at the N-extein–intein junction by the hydroxyl (or thiol) group of the N-terminal amino acid of the C-extein (serine, threonine, or cysteine), thus giving rise to a branched polypeptide chain. Cyclization of the invariant Asn residue at the C terminus of the intein to succinimide then occurs leading to excision of the intein and ligation of the exteins (2, 9, 13–15).

While analyzing the *oriC* region of the chromosome of *M. leprae*, it became apparent that protein splicing, would be required to produce the A subunit of DNA gyrase. This topoisomerase is the primary target for quinolone drugs and consists of two A subunits and two B subunits encoded by *gyrA* and *gyrB*, respectively. The A subunit is responsible for the double-stranded breakage and reunion of DNA (nicking-closing enzyme) while the B subunit mediates energy transduction via ATP hydrolysis (16). DNA gyrase negatively supercoils closed circular DNA and thus affects the essential processes of DNA replication, recombination, and transcription.

It is shown here that an intein is located within the *M. leprae* GyrA coding sequence after codon 130 and that its product is related to inteins found in *M. tuberculosis, T. litoralis*, and *Pyrococcus*, which are believed to correspond to homing endonucleases. These enzymes mediate gene conversion by cleaving at distinct chromosomal sites into which their own coding sequences, or introns, are subsequently inserted. On analysis of the *gyrA* locus of 14 mycobacterial species, inteins were also found in the *gyrA* genes of *Mycobacterium flavescens, Mycobacterium gordonae*, and *Mycobacterium kansasii*. Further evidence in favor of homing, in the form of inteinless alleles in different strains of the same species, is presented.§

MATERIALS AND METHODS

Mycobacteria. Details of the various mycobacteria used in this study are given in Table 1.

DNA Sequencing and Computer Techniques. Cosmid B1770 carrying the *gyrA* gene has been sequenced in the framework of the *M. leprae* genome sequencing project by the shotgun technique using an automated DNA sequencer (Applied Biosystems, model 373A) and the Staden software package as described (17). The nucleotide sequences of the *M. kansasii, M. flavescens*, and *M. gordonae gyrA* regions were obtained using custom-made primers and fluorescent dye-labeled terminators.

PCR Amplification and TA Cloning. For PCR analysis of mycobacterial *gyrA* genes, primers H49 (5′-AGGTTGTGCGGCGGGATATTGGT-3′) and H50 (5′-TTCCGCCCGGACCGCAGCCACG-3′) were synthesized on the basis of conserved regions around the homing site of the *Mle gyrA* intein and used in standard PCR conditions (17). The 1665-bp PCR product obtained after amplification of the *gyrA* region surrounding the putative homing site from the *M. flavescens, M. gordonae*, and *M. kansasii* strains was cloned in the T vector pGEM-T (Promega). To avoid artifacts due to amplification with *Taq* polymerase, DNA from at least three independently isolated clones was sequenced.

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.

‡To whom reprint requests should be addressed.

§The sequences reported in this paper have been deposited in the GenBank data base [accession nos. Z68206 (*Mle gyrA* intein), Z68209 (*Mpe gyrA* intein), Z68208 (*Mgo gyrA* intein), and Z68207 (*Mka gyrA* intein)].

Table 1. Mycobacterial strains used and presence/absence (+/−) of intein coding sequences in *gyrA*

Strain	Reference number	Presence of intein
M. leprae		
G	88056	+
Ma	92002	+
NC	90049	+
P	89033	+
Se	88063	+
M. flavescens		
Fla0	930991	+
Fla4	950450	+
Fla5	940758	−
M. gordonae		
Gor0	930835	+
Gor1	941295	−
Gor2	941268	−
Gor3	941270	−
Gor4	941334	−
Gor5	941446	−
Gor12	950025	+
M. kansasii		
Kans0	930908	+
Kans6	941295	+
Kans7	941386	−
Kans8	941419	−
Kans9	941443	+
Kans10	950051	−
Kans18	950077	−
Kans19	950211	−
Kans21	950479	+
M. tuberculosis	930095	−
M. bovis BCG	Pasteur strain	−
M. avium	930994	−
M. smegmatis	930587	−
M. marinum	930732	−
M. chelonae	930836	−
M. fortuitum	930944	−
M. simiae	930759	−
M. szulgai	950112	−
M. vaccae	NCTC11659	−

RESULTS

The *gyrA* Gene of *M. leprae* Harbors an Intein. The *M. leprae gyrA* gene is ≈1.3 kb longer than its counterparts from *M. tuberculosis* and *Streptomyces coelicolor.* At the N-terminal end of the GyrA protein, where mutations associated with quinolone resistance are generally found, there is extensive similarity over the first 130 amino acid residues (Fig. 1) that terminates at Tyr-130. In *Escherichia coli*, the corresponding residue Tyr-122 becomes covalently bound to DNA when the enzyme breaks the phosphodiester bonds of DNA (20). The C-terminal domain of the *M. leprae* GyrA protein, comprising residues 551–1273 (723 amino acids), is nearly identical in sequence and in size to the region encompassing residues 130–838 (709 amino acids) of the *M. tuberculosis* protein.

The *M. leprae* GyrA sequence is thus predicted to contain a long, nonhomologous insert (residues 131–550) that is not present in other GyrA proteins. Using the BLAST program (21), significant similarity was found with four proteins: the intein from the *M. tuberculosis* RecA protein (20% identity in 423-aa overlap), the PI-TLI I endonuclease found in the DNA polymerase of *T. litoralis* (21.5% identity in 163-aa overlap), the endonuclease found in the Vent DNA polymerase of *Pyrococcus* sp. (24.7% identity in 97-aa overlap), and a putative intein from an unidentified *M. leprae* protein [Pps1 (22); 25% identity in 52-aa overlap]. Given these similarities and the presence of the amino acid residues that are essential for splicing to occur (see below), it is clear that the *M. leprae gyrA* gene harbors an intein coding sequence that may correspond to a homing endonuclease.

How Many Copies of the Intein Are Present in *M. leprae?* On examination of the codon usage in *gyrA* gene, significant deviation from the mean was observed in the region encoding the intein (Fig. 2*A*). Furthermore, a striking difference in the dG+dC content was detected between the intein sequence (47.7%) and the *gyrA* extein coding sequences (54.2%), which were more typical of *M. leprae* (56%). Taken together, these observations suggested that the intein coding sequence might be mobile and of foreign origin. To establish whether additional copies were present on the chromosome of *M. leprae*, Southern blots were hybridized at high stringency with an appropriate probe. In each case, single bands corresponding to the *gyrA* locus were detected (data not shown). Similar experiments were performed at low stringency with different mycobacterial DNAs but no signals were obtained, suggesting that either the intein coding sequence was not present in other species or that its nucleotide sequence differed extensively.

Analysis of the *gyrA* Locus in Different Mycobacteria. To investigate the presence of intein sequences in the *gyrA* genes of other mycobacterial species, a PCR strategy was devised that employed primers to conserved sequences around the intein insertion site in the *gyrA* gene from *M. leprae.* Mycobacteria with *gyrA* genes like that of *M. leprae* should yield a PCR product of 1665 bp, whereas those like *M. tuberculosis* should generate smaller fragments of around 350 bp (Fig. 2*A*). Fourteen different mycobacterial species (Table 1), including *M. tuberculosis, M. avium*, and *M. smegmatis*, were subjected to PCR analysis and representative results are presented in Fig. 2*B.* Only the PCR products obtained from *M. flavescens, M. gordonae*, and *M. kansasii* were as large as that from *M. leprae*, thus suggesting the presence of an intein coding sequence in the *gyrA* gene of these organisms. Hybridization experiments with *gyrA*-specific probes confirmed the identity of the PCR products (data not shown).

Inteins from *M. kansasii, M. flavescens*, and *M. gordonae.* To establish whether the intein coding sequences were related to that of *M. leprae* and if it were inserted at the same position in *gyrA*, the ≈1665 bp PCR products from *M. kansasii, M. flavescens*, and *M. gordonae* were cloned and sequenced, and the primary structures of the inteins were deduced. In all three species, the intein coding sequence was found to be inserted in-frame after the codon corresponding to Tyr-130 of *M. leprae.* However, as expected from the hybridization studies, the nucleotide sequences differed extensively and identity values of 66–74% were found on pairwise comparison (Table 2). In all three mycobacterial species the dG+dC content of the intein coding sequence of 59–61% was significantly lower than that of the genome (66–67%), as previously observed with *M. leprae.* Sequence divergence was also seen in the intein primary structures (Table 2), although many of the differences correspond to conservative substitutions. The key amino acid residues at the N and C termini of the intein, which catalyze protein splicing, were all conserved between the GyrA inteins and those from other mycobacteria, archaebacteria, and yeast (Fig. 3). Likewise, all of the inteins contain a pair of highly related dodecapeptide motifs (26, 27), also referred to as the LAGLI-DADG signature, that are known to be of functional importance in intron-encoded nucleases (1, 28). In mycobacterial inteins, these are separated by 80–86 residues (Fig. 3). In a very recent study (22), Pietrokovski identified an additional set of intein-specific motifs, blocks A–G, and these are all present in the four GyrA-inteins described here (Fig. 1; data not shown).

```
                1                                                   50                                            *                        * **  *Δ
M. leprae       ....MTDITL PPGDGSIQRV EPVDIQQEMQ RSYIDYAMSV IVGRALPEVR  DGLKPVHRRV LYAMLDSGFR PDRSHAKSAR SVAETMGNYH PHGDASIYDT
M. tuberculosis ....MTDTTL PP DDSLDRI EPVDIEQEMQ RSYIDYAMSV IVGRALPEVR  DGLKPVHRRV LYAMFDSGFR PDRSHAKSAR SVAETMGNYH PHGDASIYDS
S. coelicolor   MTDENTPVTT PEGDALAMRV EPVGLETEMQ RSYLDYAMSV IVSRALPVVR  DGLKPVHRRV LYAMYDGGYR PERGFYKCAR VVGDVMGNYH PHGDSSIYDA

                101                  *                  #          150  A                                                      200
M. leprae       LVRMAQPWSL RYPLVDGQGN FGSPGNDPPA AMRYCVSGNS LVRLLFGKSI  RIGDIVTGAQ FNSDNPIDLK VLDRHGNPVV ADYLFHSGEH QTYTVRTTEG
M. tuberculosis LVRMAQPWSL RYPLVDGQGN FGSPGNDPPA AMRY...... ..........  .......... .......... .......... .......... ..........
S. coelicolor   LVRLAQPWSM RMPLVDSNGN FGSPGNDPAA AMRY...... ..........  .......... .......... .......... .......... ..........

                201   B                                        250                 C                                      300
M. leprae       YEITGTSNHP LLCLVNVGGI PTLLWKLIGE IRSGDYVVLQ RIPPVEFGPA  DWYSTMEALL FGAFISGGFV SQDHAGFNSL DRDYFTMVVN AYDTVVGGLR
M. tuberculosis .......... .......... .......... .......... ..........  .......... .......... .......... .......... ..........
S. coelicolor   .......... .......... .......... .......... ..........  .......... .......... .......... .......... ..........

                301                                         D  350                 E                                      400
M. leprae       CISSRITVSG STLLELDVYN LIEFKKTRLS GLCGQRSADK LVPDWLWHSP  STVKRAFLQA LFEGEGFSSI LSRNIIEISY STLSERLAAD VQQMLLEFGV
M. tuberculosis .......... .......... .......... .......... ..........  .......... .......... .......... .......... ..........
S. coelicolor   .......... .......... .......... .......... ..........  .......... .......... .......... .......... ..........

                401                                            450                                                        500
M. leprae       VSERYCHTVN EYKVVIANRA QVEMFFTQVG FGVTKQAKLI RDVVSMSPCV  GMDINCVPGL ATFIRKHCDN RWVEEDSFNQ HNVDCVQHWH HHSAEIVGHI
M. tuberculosis .......... .......... .......... .......... ..........  .......... .......... .......... .......... ..........
S. coelicolor   .......... .......... .......... .......... ..........  .......... .......... .......... .......... ..........

                501                                       F        550  G                                                     600
M. leprae       ADPDIRAIVT DLTDGRFYYA RVASVTDTGI QPVFSLHVDT EDHSFLTNG   FISHNTEARLT PLAMEMLREI DEETVDFISN YDGRVQEPMV LPSRFPNLLA
M. tuberculosis .......... .......... .......... .......... .........   .....TEARLT PLAMEMLREI DEETVDFIPN YDGRVQEPTV LPSRFPNLLA
S. coelicolor   .......... .......... .......... .......... .........   .....TECKMA PLSMEMVRDI DEETVDFTDN YDGRSQEPTV LPARFPNLLI

                601                                            650                                                        700
M. leprae       NGSGGIAVGM ATNIPPHNLY ELADAVFWCL ENHDADEETM LVAVMERVKG  PDFPTAGLIV GSQGIADAYK TGRGSIRIRG VVEVEEDSRG RTSLVITELP
M. tuberculosis NGSGGIAVGM ATNIPPHNLR ELADAVFWAL ENHDADEEET LAAVMGRVKG  PDFPTAGLIV GSQGTADAYK TGRGSIRMRG VVEVEEDSRG RTSLVITELP
S. coelicolor   NGSAGIAVGM ATNIPPHNLR EVAAGAQWYL ENYEASHEEL LDALIERIKG  PDFPTGALVV GRKGIEEAYR TGRGSITMRA VVEVEE.IQN RQCLVVTELP
```

FIG. 1. Sequence alignment between the GyrA proteins from *M. leprae, M. tuberculosis*, and *S. coelicolor.* Only the first 700 residues are shown. The position of the Tyr-122 residue (according to the *E. coli* numbering) is indicated by # as well as the residues identified in *M. tuberculosis* (18), *Mycobacterium avium*, and *Mycobacterium smegmatis* (19) associated with mutations conferring quinolone-resistance (*). Residue Ser-95 which, in the *M. tuberculosis* complex, displays a natural polymorphism is indicated (△).

Are Inteins Always Associated with *gyrA* in Strains of *M. leprae, M. kansasii, M. flavescens*, and *M. gordonae*? To establish whether the intein coding sequence was always present in *gyrA*, PCR was performed with genomic DNA from different strains of *M. leprae, M. kansasii, M. flavescens*, and *M. gordonae.* In contrast to the situation in *M. leprae*, where the intein was detected in all five isolates tested, it did not appear to be present in five strains of *M. kansasii* of nine tested, one strain of *M. flavescens* of three tested, and five strains of *M. gordonae* of seven tested; typical results illustrating this dichotomy are presented in Fig. 4 and are summarized in Table 1.

Consequently, PCR products from the inteinless *gyrA* genes were cloned and sequenced, and the nucleotide sequences aligned with those from other mycobacterial *gyrA* genes (Fig. 5). Interestingly, in all *gyrA* genes that did not contain the intein coding sequence, the Tyr codon corresponding to the site of insertion was TAC, whereas in three of the four species

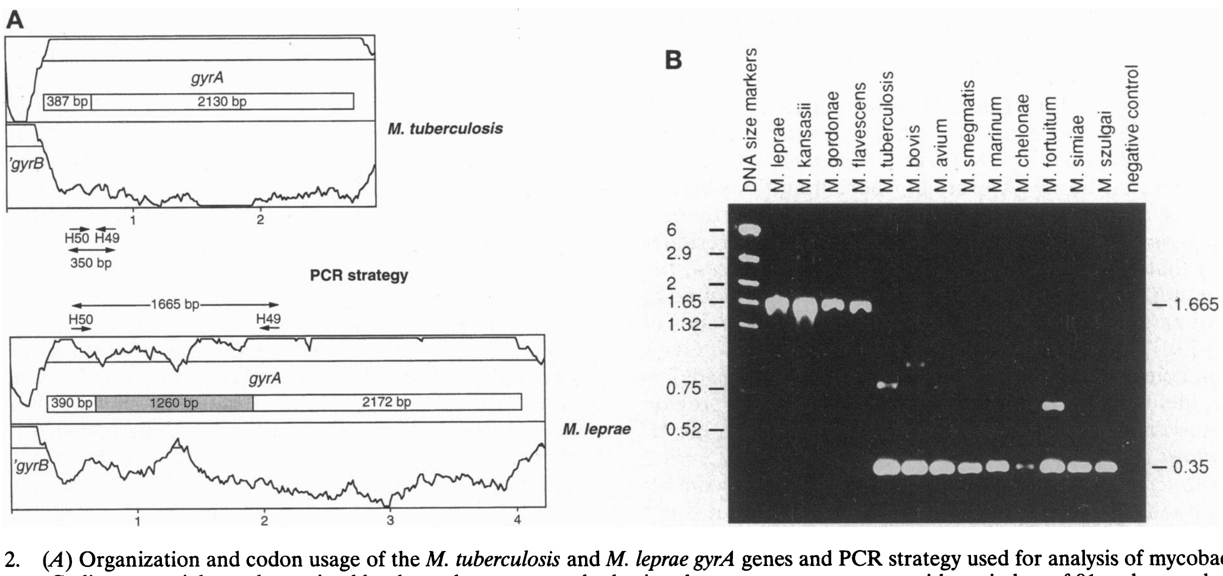

FIG. 2. (*A*) Organization and codon usage of the *M. tuberculosis* and *M. leprae gyrA* genes and PCR strategy used for analysis of mycobacterial *gyrA* genes. Coding potential was determined by the codon usage method using the FRAMESCAN program with a window of 91 codons and a table of known *M. leprae* codons (23, 24). Only two of the three possible reading frames, corresponding to *'gyrB* and *gyrA*, are shown and the scale is in kb. When the codon usage is typical of mycobacteria a high-scoring plateau is obtained—e.g., above *gyrA* in *M. tuberculosis*—deviations from the norm result in troughs and this is readily apparent in the 1260-bp segment corresponding to the intein coding sequence in *M. leprae.* The midpoint value in the open reading frame is denoted by the line. Primers H49 and H50 used for amplification and the length of the PCR product, depending on the absence or presence of an intein in the *gyrA* gene, are indicated. (*B*) PCR analysis of genomic DNAs from different mycobacterial species (indicated at top). Sizes of markers in kb are shown.

Table 2. Pairwise identifies (%) between intein sequences at DNA and protein levels determined by FASTA (25)

	Protein			
	M. leprae	*M. flavescens*	*M. gordonae*	*M. kansasii*
DNA				
M. leprae	100	64.6	65.0	63.6
M. flavescens	66.2	100	74.6	73.9
M. gordonae	67.2	70.7	100	77.6
M. kansasii	66.8	71.9	73.8	100

haboring inteins this was TAT. In the exception, *M. flavescens* Fla0, there were 2 base differences between strains with and without the intervening sequence within 12-bp stretches either side of the insertion site in *gyrA* (indicated in Fig. 5).

The finding that some strains of a given species have GyrA inteins while others do not raised the possibility that if sufficient isolates of an apparently inteinless species were examined, intein coding sequences might be detected in *gyrA*. This was appraised by screening 40 recent clinical isolated of *M. tuberculosis* by PCR but in all cases intein-free *gyrA* genes were detected (data not shown).

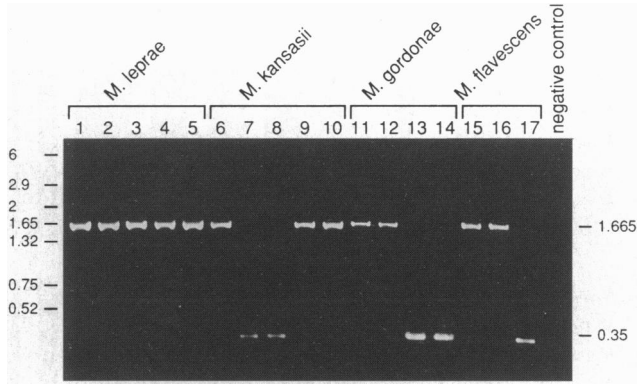

FIG. 4. PCR analysis of genomic DNAs from different strains of various mycobacterial species (indicated at top). Lanes: 1–5, *M. leprae* isolates, G, Ma, NC, P, and Se (Table 1), respectively; 6–10, *M. kansasii* strains Kans0, Kans7, Kans8, Kans9, and Kans21, respectively; 11–14, *M. gordonae* strains Gor0, Gor12, Gor1, and Gor2, respectively; 15–17, *M. flavescens* strains, Fla0, Fla4, and Fla5, respectively. Strains with *gyrA* genes harboring inteins gave PCR fragments of 1.665 kb and those with inteinless genes gave fragments of 0.35 kb. Sizes of markers in kb are indicated.

DISCUSSION

In some mycobacteria the GyrA subunit of DNA gyrase appears to be made by means of a protein splicing event in which an intein is excised from the precursor protein in a reaction catalyzed by critical amino acid residues at the extein–intein junctions. A protein of the size expected of mature GyrA (≈94 kDa) has been detected by means of immunoblotting experiments (data not shown) in extracts of cultivable mycobacteria, both with or without intein-coding sequences in *gyrA*. Given the highly conserved structural features and sequence homologies with intron-encoded nucleases and homing endonucleases, it is reasonable to propose that the *gyrA*-associated intein should also behave as a homing endonuclease. There are striking parallels between the inteins in *gyrA* and the PI-*Sce* I endonuclease, which is produced by protein splicing from the vacuolar H^+-ATPase subunit precursor encoded by *VMA1/TFP1* in *S. cerevisiae*. PI-*Sce* I has been shown to behave as a homing endonuclease that mediates meiotic gene conversion leading to insertion of its own coding sequence (29). A closely related enzyme is encoded by the intervening sequence located at exactly the same site in the *VMA1* gene of *C. tropicalis* (7). In some yeast strains, this homing site remains unoccupied.

Among the mycobacteria, the putative homing site in *gyrA* appears to be permanently occupied in *M. leprae*, occasionally occupied in *M. kansasii, M. flavescens*, and *M. gordonae*, and probably unoccupied in the other species examined. This may indicate that the intein coding sequence has been acquired independently by each species since divergence from their common ancestor or, alternatively, that it was lost by others. Indirect evidence for the former is available as in all four cases the base composition and codon usage of the intein coding sequence is not representative of mycobacteria, thus suggesting horizontal transfer. Differences in the nucleotide sequences of the putative homing sites (Fig. 5) might also explain why some, but not all, strains of the same species had acquired the intein coding sequence since the corresponding homing endonuclease could exhibit sequence specificity. Inteinless alleles have also been reported for other genes encoding spliced proteins (5). Homing endonucleases generally recognize extended sites of 15–40 bp, which would occur only once in a bacterial genome (30); this is consistent with the finding that only one copy of the intein was present in *M. leprae*.

If one assumes that the ancestral mycobacterium had a *gyrA* gene with an intein, it is difficult to envisage how the intein coding sequence could have been lost by the majority of the species, although it should be remembered that mycobacteria are haploid. Consequently, in contrast to the well-documented situation in yeast where homing endonuclease-mediated gene conversion occurs at high frequency (29), loss of the intein coding sequence would not be corrected by the second allele. The present analysis does not exclude the possibility that intein coding sequences may be inserted elsewhere in the *gyrA* genes

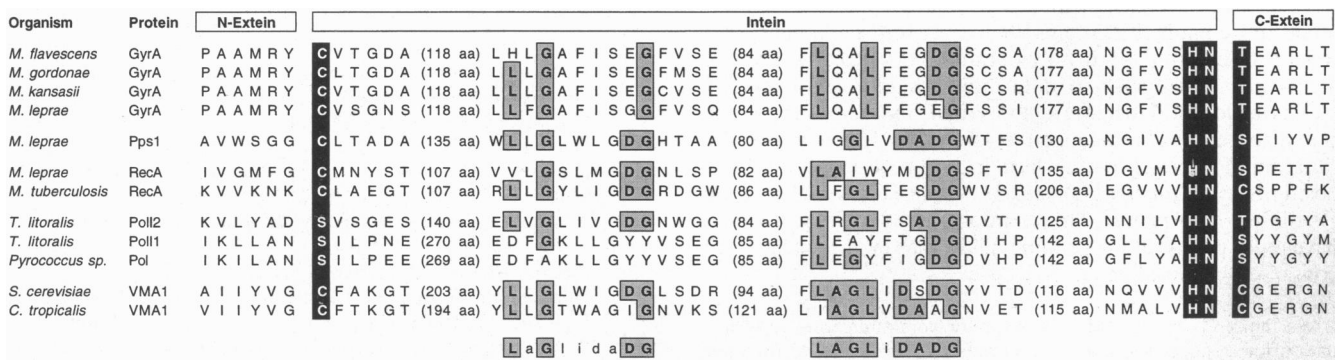

FIG. 3. Schematic sequence alignment of key regions from polypeptides known, or suspected, to be produced by protein splicing showing the positions of the N-extein, intein, C-extein. Distances between conserved features are shown and the paired motifs associated with intron-encoded nucleases and homing endonucleases (26, 27) are indicated. Critical residues at the intein–extein junctions are shown as white on black, whereas conserved residues of the LAGLI-DADG motif and Pietrokovski's intein signatures are boxed (22).

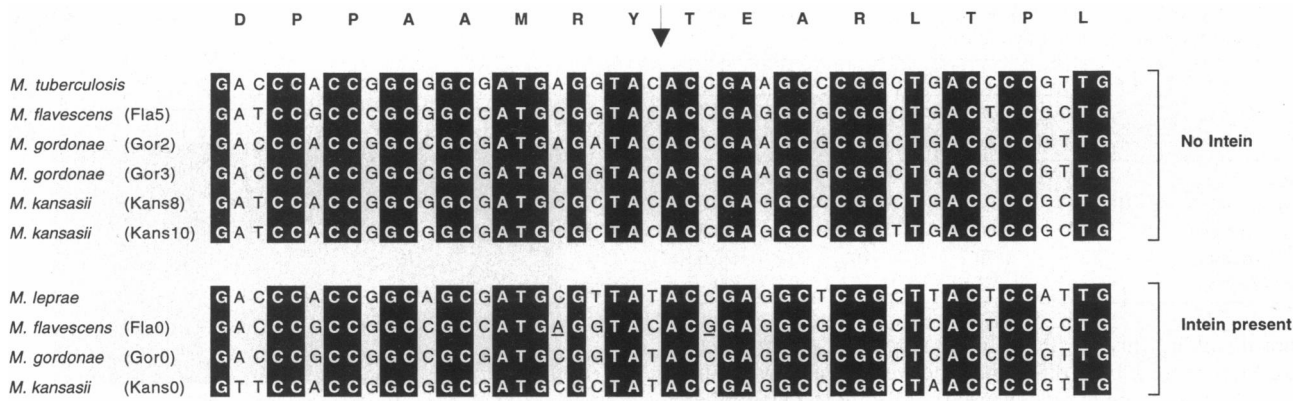

FIG. 5. Alignment of nucleotide sequence of the putative homing sites in mycobacterial *gyrA* genes. Sequences were from alleles from *gyrA* genes with (+) or without (−) inteins from *M. leprae, M. flavescens, M. gordonae, M. kansasii*, and *M. tuberculosis.* Universally conserved nucleotides are shaded. Deduced GyrA protein sequence is shown above and arrow indicates intein insertion site. Two nucleotides that differ in the respective *gyrA* alleles of *M. flavescens* strains are underlined.

of some mycobacterial strains, as described for *recA* (12), for only a critical segment of the gene was screened. However, this was excluded in the case of *M. kansasii* as PCR experiments using intein-specific primers only yielded products of the expected size from intein-plus strains Kans0 and Kans6 (Table 1), thereby indicating that this intein coding sequence was present neither in *gyrA* nor elsewhere in the genome (data not shown).

Undoubtedly, the most intriguing feature of this study is the location of the putative homing site adjacent to, and partially overlapping, the codon for Tyr-130. This tyrosine residue corresponds to Tyr-122, the residue involved in strand cleavage, covalent attachment to DNA, and rejoining by the *E. coli* GyrA protein (20). There is a strong, albeit indirect, argument in favor of protein splicing being required to liberate a mature GyrA protein since the presence of an additional domain of 420 amino acids, near the active site, would almost certainly inhibit both the folding and the catalytic activity of the enzyme. Based on their findings with the precursor of the RecA protein of *M. leprae*, which negatively complements mature multimeric RecA in some situations, Colston and Davis proposed that splicing may be conditional and involve regulation by environmental factors or accessory molecules (11, 12). It is conceivable that unspliced GyrA could also have a negative dominant effect on DNA gyrase, a multisubunit enzyme. Furthermore, it is indeed noteworthy that inteins appear to be associated predominantly with genes encoding proteins, like RecA (12), GyrA, DNA polymerase (8), and possibly DnaB (22), that are involved in DNA metabolism.

It has also been hypothesized that inteins could play a role in intracellular survival or pathogenesis of *M. tuberculosis* and *M. leprae* (12). With the finding that nonpathogenic mycobacteria, such as *M. gordonae* and *M. flavescens*, possess inteins, it seems more likely that inteins are merely selfish elements that can be transferred both horizontally and vertically. Experiments are in progress to determine whether all four *gyrA* inteins have endonuclease activity and, as seems likely, are isoschizomers that recognize the same site. This is unlike the situation of the *recA* intein where insertion occurs at different positions in *M. tuberculosis* and *M. leprae* (12). Finally, if the *gyrA* intein does prove to have endonuclease activity, it would have great potential for genome analysis as *gyrA* is such a well conserved bacterial gene.

We thank H. Takiff and L. Salazar for communicating sequence data prior to publication, and A. Varnerot and N. Honoré for their help with manipulating the mycobacterial strains. We are also indebted to O. Lefebvre for encouragement and help with the figures. This work received financial support from the United Nations Development Program/World Bank/World Health Organization Special Programme for Research and Training in Tropical Diseases and additional funds from the Association Française Raoul Follereau, the Groupement d'Intérêt Publique–Groupement de Recherches et d'Etudes sur les Génomes (GIP-GREG), and the Institut Pasteur.

1. Shub, D. A. & Goodrich-Blair, H. (1992) *Cell* **71,** 183–186.
2. Wallace, C. J. A. (1993) *Protein Sci.* **2,** 697–705.
3. Doolittle, R. F. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 5379–5381.
4. Colston, M. J. & Davis, E. O. (1994) *Mol. Microbiol.* **12,** 359–363.
5. Perler, F. B., Davis, E. O., Dean, G. E., Gimble, F. S., Jack, W. E., Neff, N., Noren, C. J., Thoner, J. & Belfort, M. (1994) *Nucleic Acids Res.* **22,** 1125–1127.
6. Hirata, R., Ohsumi, Y., Nakano, A., Kawasaki, H., Suzuki, K. & Anraku, Y. (1990) *J. Biol. Chem.* **265,** 6726–6733.
7. Gu, H. H., Xu, J., Gallagher, M. & Dean, G. E. (1993) *J. Biol. Chem.* **268,** 7372–7381.
8. Perler, F. B., Comb, D. G., Jack, W. E., Moran, L. S., Qiang, B., Kucera, R. B., Brenner, J., Slatko, B. E., Nwankwo, D. O., Hempstead, S. K., Carlow, C. K. S. & Jannasch, H. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 5577–5581.
9. Hodges, R. A., Perler, F. B., Noren, C. J. & Jack, W. E. (1992) *Nucleic Acids Res.* **20,** 6153–6157.
10. Davis, E. O., Sedgwick, S. G. & Colston, M. J. (1991) *J. Bacteriol.* **173,** 5653–5662.
11. Davis, E. O., Jenner, P. J., Brooks, P. C., Colston, J. M. & Sedgwick, S. G. (1992) *Cell* **71,** 201–210.
12. Davis, E. O., Thangaraj, H. S., Brooks, P. C. & Colston, J. M. (1994) *EMBO J.* **13,** 699–703.
13. Clarke, N. E. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 11084–11088.
14. Xu, M., Southworth, M. W., Mersha, F. B., Hornstra, L. J. & Perler, F. B. (1993) *Cell* **75,** 1371–1377.
15. Xu, M., Comb, D. G., Paulus, H., Noren, C. J., Shao, Y. & Perler, F. B. (1994) *EMBO J.* **13,** 5517–5522.
16. Reece, R. J. & Maxwell, A. (1991) *CRC Crit. Rev. Biochem. Mol. Biol.* **26,** 335–375.
17. Fsihi, H. & Cole, S. T. (1995) *Mol. Microbiol.* **16,** 909–919.
18. Takiff, H. E., Salazar, L., Guerrero, C., Philipp, W., Huang, W. M., Kreisworth, B., Cole, S. T., Jacobs, W. R. & Telenti, A. (1994) *Antimicrob. Agents Chemother.* **38,** 773–780.
19. Cambau, E., Sougakoff, W. & Jarlier, V. (1994) *FEMS Microbiol. Lett.* **116,** 49–54.
20. Horowitz, D. S. & Wang, J. C. (1987) *J. Biol. Chem.* **262,** 5339–5344.
21. Altschul, S., Gish, W., Miller, W., Myers, E. & Lipman, D. (1990) *J. Mol. Biol.* **215,** 403–410.
22. Pietrokovski, S. (1994) *Protein Sci.* **3,** 2340–2350.
23. Honoré, N., Bergh, S., Chanteau, S., Doucet-Populaire, F., Eiglmeier, K., Garnier, T., Georges, C., Launois, P., Limpaiboon, P., Newton, S., Nyang, K., del Portillo, P., Ramesh, G. K., Reddy, T., Riedel, J. P., Sittisombut, N., Wu-Hunter, S. & Cole, S. T. (1993) *Mol. Microbiol.* **7,** 207–214.
24. Staden, R. & McLachlan, A. D. (1982) *Nucleic Acids Res.* **10,** 151–156.

25. Pearson, W. & Lipman, D. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 2444–2448.
26. Cummings, D. J., Michel, F. & McNally, K. L. (1989) *Curr. Genet.* **16,** 381–406.
27. Waring, R. B., Davies, R. W., Scazzochio, C. & Brown, T. A. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 6332–6336.
28. Lazowska, J., Claisse, M., Gargouri, A., Kotylak, Z., Spyridakis, A. & Slominski, P. P. (1989) *J. Mol. Biol.* **205,** 275–289.
29. Gimble, F. S. & Thorner, J. (1992) *Nature (London)* **357,** 301–306.
30. Mueller, J. E., Byrk, M., Loizos, N. & Belfort, M. (1993) in *Nucleases*, eds. Linn, S. M., Lloyd, R. S. & Roberts, R. J. (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed., pp. 111–143.