Phylogeny of *Mycobacterium avium* strains inferred from glycopeptidolipid biosynthesis pathway genes

Elzbieta Krzywinska,† Jaroslaw Krzywinski and Jeffrey S. Schorey

Department of Biological Sciences, Center for Tropical Disease Research and Training, University of Notre Dame, 130 Galvin Life Science Center, Notre Dame, IN 46556, USA

The *Mycobacterium avium* complex (MAC) encompasses two species, *M. avium* and *Mycobacterium intracellulare*, which are opportunistic pathogens of humans and animals. The standard method of MAC strain differentiation is serotyping based on a variation in the antigenic glycopeptidolipid (GPL) composition. To elucidate the relationships among *M. avium* serotypes a phylogenetic analysis of 13 reference and clinical *M. avium* strains from 8 serotypes was performed using as markers two genomic regions (890 bp of the *gltB* gene and 2150 bp spanning the *rtfA–mtfC* genes) which are associated with the strains’ serological properties. Strains belonging to three other known *M. avium* serotypes were not included in the phylogeny inference due to apparent lack of the marker sequences in their genomes, as revealed by PCR and Southern blot analysis. These studies suggest that serotypes prevalent in AIDS patients have multiple origins. In trees inferred from both markers, serotype 1 strains, known to have the simplest and shortest GPLs among all other serotypes, were polyphyletic. Likewise, comparisons of the inferred phylogenies with the molecular typing results imply that the existing tools used in epidemiological studies may be poor estimators of *M. avium* strain relatedness. Additionally, trees inferred from each marker had significantly incongruent topologies due to a well supported alternative placement of strain 2151, suggesting a complex evolutionary history of this genomic region.

INTRODUCTION

The *Mycobacterium avium* complex (MAC) is recognized as an important group of bacteria causing disease in man and animals. MAC consists of two principal species, *M. avium* and *Mycobacterium intracellulare*, containing morphologically and biochemically indistinguishable organisms, and many strains not assigned to either species (Wayne & Sramek, 1992). Widespread reports on the isolation of MAC from patients with AIDS and from patients without predisposing conditions, increases the need for a reliable strain identification system to obtain information on the epidemiology and virulence of MAC. The standard technique used for MAC strain differentiation has been serologic typing based on the sugar residue composition of surface glycopeptidolipids (GPLs). Of 28 described MAC serotypes, 11 have been assigned to *M. avium* (serotypes 1–6, 8–11 and 21) and 11 to *M. intracellulare* (serotypes 7, 12–20 and 25), based on DNA and antibody probes and HPLC analysis (Saito et al., 1990). Serotyping analysis showed that MAC infections in AIDS patients were almost exclusively due to *M. avium* belonging predominantly to serotypes 1, 4, 6 and 8; however, the prevalence of isolated serotypes varied depending on geographic location (Julander et al., 1996; Yakrus & Good, 1990). This raises the question whether these serotypes form a related subset of strains.

Two forms of GPLs are present in MAC, the serotype-specific GPL (ssGPL) and non-serotype specific GPL (nsGPL) (for a review, see Chatterjee & Khoo, 2001). All GPLs have in common an N-acylated lipotetrapeptide core that bears a rhamnosylated alanyl C terminus. In both GPL forms, a single 6-deoxytalose (6-dTal) unit is attached to the D-allo-Thr. In case of the ssGPL, 6-dTal is further glycosylated to yield a haptenic oligosaccharide that varies in composition among different serotypes. The biosynthetic pathway leading to the formation of the GPLs is poorly understood. Recently, Eckstein et al. (2003) proposed a pathway leading to the formation of serotype-specific GPL of serotype 2. However, the exact function of the genes potentially involved remains unknown, with one exception. It has been experimentally demonstrated that the *rtfA* gene encodes rhamnosyltransferase that catalyses the addition of rhamnose to 6-dTal (Eckstein et al., 1998;
Maslow et al., 2003). Glycosylation performed by rhamnosyltransferase is assumed to be the first step in the conversion of the non-specific GPLs to serotype-specific GPLs.

Although seroagglutination has been the most frequently employed method for differentiating MAC strains, up to one-third of all isolates could not be serotyped due to autoagglutination or a failure to react with any serum (Tsang et al., 1992). This has promoted the development of DNA-based tools to differentiate MAC strains and to analyse strain relatedness. Currently available differentiation methods include DNA–rRNA hybridization with commercial probes (Accuprobe; GenProbe), nucleotide sequence analysis of the 16S–23S rDNA internal transcribed spacer (ITS) (Frothingham & Wilson, 1993), the 16S rRNA (Rogall et al., 1990) and hsp65 (Swanson et al., 1997) genes, as well as IS1245 RFLP typing (Guerrero et al., 1995; van Soolingen et al., 1998) and analysis of restriction polymorphism of a PCR-amplified fragment of the hsp65 gene (Telenti et al., 1993). However, different genotypic assays often yield inconsistent results and they do not correlate well with serotyping (Wasem et al., 1991), suggesting different evolutionary forces acting upon individual markers or differences in their evolutionary histories. Therefore, markers based directly on genes involved in serotype specificity would be more illuminating.

In addition, the evolutionary relationships within M. avium are very poorly understood. Phylogeny of MAC has been studied using 16S–23S rDNA ITS and hsp65 sequences as markers; however, because of extremely low sequence variation the relationships among M. avium strains were nearly completely unresolved (Frothingham & Wilson, 1993; Swanson et al., 1997). Until now, only three sequence variants differing at two alignment positions have been identified in the hsp65 gene (Leao et al., 1999; Swanson et al., 1997). The ITS locus was slightly more variable because it allowed for differentiation of seven distinct M. avium genotypes referred to as sequevars (Frothingham & Wilson, 1993; Mijs et al., 2002). Further phylogenetic studies implementing alternative markers are necessary to help answer fundamental questions related to M. avium epidemiology and pathogenicity, such as why certain serotypes are more prevalent in AIDS patients.

Previously we found relatively high sequence variation within the 5’ region of the GPL biosynthesis cluster among four M. avium strains belonging to three serotypes (Krzywinska & Schorey, 2003). In addition, gene organization within this region was highly conserved among all four strains and possibly conserved among other M. avium strains, suggesting its potential value for phylogeny reconstruction. Here we employ selected GPL genes as markers to determine phylogenetic relationships among M. avium serotypes, and compare the inferred trees with serotype designations and information obtained from markers used in earlier studies. Our analysis allowed us to gain new perspective on the utility of the existing M. avium typing tools and on the mechanisms of M. avium evolution, involving likely secondary gain of serotypic characteristics and horizontal gene transfer.

METHODS

Mycobacterial strains. MAC strains used in this study, together with their serotypes, IS1245 RFLP patterns and sequevars are listed in Table 1. MAC reference strains were obtained as frozen broth cultures. Other mycobacterial strains have been maintained in our laboratory as described previously (Bohlool et al., 2001). M. intracellulare was included in the study as an outgroup in the phylogenetic analysis.

PCR amplification and DNA sequencing of GPL cluster genes. Three primer pairs were designed based on the nucleotide sequence alignment of the GPL biosynthesis cluster from M. avium strains A5 (GenBank accession no. AY130970), 724 (AF125999), 2151 (AF143772) and 104 [The Institute for Genomic Research (TIGR)], http://www.tigr.org. Primers rraf (5’-GCTGTGGCAAGTTATGGG-3’) and rhar (5’-CGGCTGGAGTTATCGGG-3’) were used to amplify a fragment of the rhamnosyltransferase A gene corresponding to nucleotide positions 2310–3482 in the A5 strain GPL cluster sequence; primers gtBF (5’-CGTGTGCTGTAGCTTCTTT-3’) and gtBR (5’-TGCACACGCTCAATCGGT-3’) were used to amplify a fragment of the glycosyltransferase B gene spanning positions 6332–7557; and primers mtRF (5’-CCTCCGAAGATCCTGGC-3’) and mtRCR (5’-TCGGATCTCTCAGTCCTCC-3’) were used to amplify a fragment spanning positions 3446–4479, encompassing the 3’ end of the rhamnosyltransferase A gene, an intergenic region and most of the methyltransferase C gene. Selection of these genomic regions as markers was dictated by their higher sequence variability relative to other aligned portions of the cluster. The PCR template was prepared by centrifugation of a 200 μl frozen broth culture or glycerol stock of mycobacteria, resuspension in 50 μl sterile TE and boiling for 10 min. The lysates were divided into aliquots and stored at −20 °C. PCR reactions were performed using the Advantage-GC cDNA Polymerase Mix (Clontech), in a final volume of 50 μl containing 1 μl crude lysate, 10 mM dNTP mix and 50 pmol each primer. PCR thermal conditions consisted of an initial denaturation step of 2 min at 94 °C, followed by 30 cycles at 94 °C for 30 s, 58–59 °C for 40 s, 72 °C for 1 min, and a final extension at 72 °C for 10 min. PCR products were excised from a gel, purified with QIAEX II Gel Extraction Kit (Qiagen) and, with one exception (see Results and Discussion), sequenced directly from both strands using forward, reverse and internal primers. Sequencing was carried out on an ABI 3700 sequencer using ABI BigDye terminator chemistry and the sequences were edited using Sequence Navigator (Perkin-Elmer). To confirm the identity of the sequences, amplification and sequencing were performed twice on eight strains selected at random.

Southern blot hybridization. Mycobacterial genomic DNA was isolated (Hunter et al., 1982) and 5 μg was digested with BanHI, separated by electrophoresis on a 0.6% agarose gel and capillary transferred to positively charged nylon membranes using a standard protocol (Sambrook et al., 1989). Purified PCR products from M. avium 101, labelled with digoxigenin using the DIG High Prime DNA Labelling Kit (Boehringer Mannheim), were used as probes. Membranes were hybridized overnight at 68 °C with the probes and signal detection was performed according to the manufacturer’s protocol. Post-hybridization washes were performed under high stringency (2 × SSC, 0.1% SDS at room temperature; 0.1 × SSC, 0.1% SDS at 68 °C) or low stringency (2 × SSC, 0.1% SDS at room temperature; 1 × SSC, 0.1% SDS at 55 °C) conditions. Before reprobing, membranes were stripped according to the manufacturer’s protocol.
DNA fingerprinting. DNA fingerprinting was performed by restriction fragment length polymorphism (RFLP) typing using the insertion sequence IS1245 as a probe, according to van Soolingen et al. (1998).

ITS sequencing and sequevar determination. The ITS region was amplified as described by De Smet et al. (1995) and sequenced. To determine the sequevar of each strain, the ITS sequences were compared with the published sequevar sequences (Frothingham & Wilson, 1993; Mijs et al., 2002) following alignment using CLUSTALX (Thompson et al., 1997).

Sequence analysis. Maximum-parsimony (MP) and maximum-likelihood (ML) phylogenetic analyses of CLUSTALX-aligned sequences were carried out with PAUP 4.0b10 (Swofford, 2001), using heuristic searches and TBR branch-swapping. The analyses were done by stepwise random addition of sequences with 1,000 (MP) or 100 (ML) replications; confidence in the inferred topologies was estimated by bootstrapping, with 500 (MP) or 100 (ML) pseudoreplicates, each with 10 random additions of sequences. The F81+I substitution model used in likelihood analyses was selected for both datasets using MODELTEST 3.06 (Posada & Crandall, 1998). The presence of conflict between the datasets was evaluated by using the incongruence length difference (ILD) test (Farris et al., 1995), and the significance of differences between the most likely tree and the tree with the alternative topology was evaluated in the Shimodaira & Hasegawa (1999) test, both implemented in PAUP.

RESULTS AND DISCUSSION

The strains in the MAC have gained importance because of the high prevalence of disseminated MAC infection in patients with AIDS and the increasing number of infections in non-AIDS patients. A significant number of studies have been initiated to understand the epidemiology of these mycobacteria; nevertheless, the origin of MAC infections in humans remains a matter of speculation. It has been suggested that MAC infections in AIDS patients may be caused by a limited number of related M. avium strains (Ritacco et al., 1998; Yakrus & Good, 1990), almost invariably belonging to serotypes 1, 4, 6 and 8 (Julander et al., 1996; Yakrus & Good, 1990), but other evidence indicates that they do not represent a single clonal lineage (Feizabadi et al., 1996). Here we address the issue of phylogenetic relationships among M. avium serotypes by using, as markers, genes associated with the strains’ serological properties.

DNA amplification, gene sequences and alignment

A total of 19 MAC strains belonging to 13 serotypes were included in this study (Table 1). Data from four strains

Table 1. MAC strains

<table>
<thead>
<tr>
<th>Strain</th>
<th>Serotype</th>
<th>IS1245 RFLP*</th>
<th>Sequevar (ITS)</th>
<th>rtfA-mtfC and gtfB spacer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>M. avium</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B-92</td>
<td>1</td>
<td>mb</td>
<td>Mav-A</td>
<td>This study</td>
</tr>
<tr>
<td>104</td>
<td>1</td>
<td>mb†</td>
<td>Mav-B‡</td>
<td>TIGR</td>
</tr>
<tr>
<td>101</td>
<td>1</td>
<td>mb†</td>
<td>Mav-B‡</td>
<td>This study</td>
</tr>
<tr>
<td>724</td>
<td>2</td>
<td>Bird-type</td>
<td>Mav-A</td>
<td>GenBank AF125999</td>
</tr>
<tr>
<td>2151</td>
<td>2</td>
<td>mb†</td>
<td>Mav-B‡</td>
<td>GenBank AF143772</td>
</tr>
<tr>
<td>6195</td>
<td>3</td>
<td>Bird-type</td>
<td>Mav-A</td>
<td>This study</td>
</tr>
<tr>
<td>13528-1079</td>
<td>4</td>
<td>mb</td>
<td>Mav-B</td>
<td>This study</td>
</tr>
<tr>
<td>Sparrow 185</td>
<td>4</td>
<td>mb</td>
<td>Mav-A</td>
<td>This study</td>
</tr>
<tr>
<td>A5</td>
<td>4</td>
<td>1 band†</td>
<td>Mav-B‡</td>
<td>GenBank AF130970</td>
</tr>
<tr>
<td>25546-759</td>
<td>5</td>
<td>ND</td>
<td>Mav-B</td>
<td>ND</td>
</tr>
<tr>
<td>34540-Wales</td>
<td>6</td>
<td>mb</td>
<td>Mav-A</td>
<td>This study</td>
</tr>
<tr>
<td>SJB#2</td>
<td>8</td>
<td>mb</td>
<td>Mav-B</td>
<td>This study</td>
</tr>
<tr>
<td>1602-1965</td>
<td>10</td>
<td>ND</td>
<td>Mav-B</td>
<td>ND</td>
</tr>
<tr>
<td>Borne</td>
<td>10</td>
<td>ND</td>
<td>Mav-B</td>
<td>ND</td>
</tr>
<tr>
<td>14186-1424</td>
<td>11</td>
<td>ND</td>
<td>Mav-A</td>
<td>ND</td>
</tr>
<tr>
<td>2993</td>
<td>21</td>
<td>mb</td>
<td>Mav-B</td>
<td>This study</td>
</tr>
<tr>
<td>M. intracellulare</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>157 Manten</td>
<td>7</td>
<td>Min-A</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>5509-Borstel</td>
<td>13</td>
<td>Min-A</td>
<td>This study</td>
<td></td>
</tr>
</tbody>
</table>

ND, Not determined.
*mb, Multibanded patterns.
†RFLP pattern determined in this study. Other patterns from Guerrero et al. (1995) and Ritacco et al. (1998).
‡Sequevar determined in this study. Other sequevar from Frothingham & Wilson (1993).

http://mic.sgmjournals.org 1701
were derived from GenBank (M. avium 724, accession no. AF125999; M. avium 2151, AF143772; M. avium A5, AJ130970) and TIGR (M. avium 104) databases. Sequences from other strains, except the five PCR-negative strains, were obtained in this study (see below). The final dataset used in phylogenetic analysis consisted of 13 strains of M. avium and one strain of M. intracellulare (Table 1).

In all M. avium PCR-positive strains amplification with gtfB primers yielded a 926 bp fragment of the glycosyltransferase B gene, with rtfA primers yielding a 1151 bp fragment of the rhamnosyltransferase A gene and with mtfC primers yielding a 1032–1034 bp fragment, consisting of the 3’ end (161 bp) of the rhamnosyltransferase A gene, an intergenic spacer (101–103 bp) and the 5’ fragment (770 bp) of the methyltransferase C gene. Use of the rtfA and mtfC primer pairs, amplifying slightly overlapping genomic regions, allowed us to obtain a 2150 bp fragment of contiguous genomic sequence, referred to as rtfA-mtfC. In M. intracellulare serotype 13 (strain 5509-Borstel) the only specific product was obtained with the rtfA primers after lowering the annealing temperature and cloning. Interestingly, the positive clones contained, apart from a corresponding sequence in M. avium, an entire 3’ end of the rtfA gene, an intergenic region and a 139 bp fragment of the mtfC gene. Inspection of the obtained sequence revealed that the annealing of the rtfAR primer was prevented by four mismatches with its target site. Instead, the amplification was effected solely by the rtfAF primer, which, not counting the target site within the rtfA gene, annealed in a reverse orientation within the 5’ end of the mtfC gene.

Despite efforts involving changes in PCR conditions and using isolated genomic DNA instead of crude bacterial lysates as templates, amplification of markers from M. avium serotypes 5 (strain 25546-759), 10 (strains 1602-1965 and Borne) and 11 (strain 14186-1424), and from M. intracellulare strain 157 Manten was unsuccessful. With the exception of a product from M. intracellulare strain 5509-Borstel amplified using rtfA primers, lowering annealing temperature led to the amplification of only non-specific products. PCR failure in such cases can be attributed to mismatches between primers and their target sites or lack of the target sequences in the genomes. Southern blot analysis of the genomic DNA from PCR-negative strains was performed to determine the cause of the amplification failure (data not shown). Under high stringency conditions, PCR-amplified fragments of rtfA, mtfC and gtfB genes used as probes yielded a signal only for M. intracellulare serotypes 7 (strain 157 Manten) and 13 (strain 5509-Borstel), although in both cases the probes hybridized to different molecular mass fragments compared to M. avium serotype 21 DNA used as a positive control. This indicated that the lack of PCR products resulted from modified primer target sites in 157 Manten and 5509-Borstel strains. No hybridization signal in the PCR-negative M. avium strains suggested that either their genomes lack the DNA fragments homologous to the probes or low similarity to the probes prevented hybridization. To allow detection of more divergent sequences, the same blots were reprobed at low stringency. This approach appeared uninformative, because in each strain, including the control, several DNA fragments showed cross-reactivity with the probes, consistent with the presence of genes similar to the marker sequences within the M. avium genome. Because the rtfA locus in M. intracellulare strain 5509-Borstel was detected at high stringency, despite substantial divergence from the rtfA probe (see below), we conclude that the marker sequences are absent from the PCR-negative M. avium strains 25546-759 (serotype 5), 1602-1965 (serotype 10), Borne (serotype 10) and 14186-1424 (serotype 11). Evidence suggests that these strains may be closely related. IS1245-based RFLP typing clustered serotype 5 and 10 strains into a fairly homogeneous group, which differed from serotype 11 strains in their banding pattern (Ritacco et al., 1998). However, GPLs of serotype 10 and 11 have been shown to be identical by GC-MS and ELISA (Denner et al., 1992). Another distinct characteristic of the ssGPLs that is shared by all three PCR-negative M. avium serotypes is a unique component that allows attachment of haptenic oligosaccharide to the N-acylpeptide core (Chatterjee & Khoo, 2001). We suggest that the lack of genes used as markers in the present study reflects an altered genetic basis for the ssGPL biosynthesis in these strains.

The concatenated nucleotide sequences of both markers with primer sequences removed produced an unambiguous 3007 character-long alignment encompassing 309 variable and 82 parsimony informative positions. There were four indels, 1–3 bp long, present in the intergenic spacer region. The alignments of both markers are available from the NCBI website via Popset. Across all studied strains the numbers of variable and informative sites were, respectively, 218 and 44 for rtfA, 17 (plus 7 indel sites) and 5 (plus 3 indel sites) for the intergenic spacer, 55 and 19 for mtfC, and 19 and 14 for gtfB. There were 50 (11 in M. avium) non-synonymous (causing amino acid change) nucleotide substitutions in rtfA, 2 (both in M. avium) in mtfC and 6 in gtfB. Proportions of polymorphic sites varied between the studied genes. Maximum pairwise sequence divergence within the ingroup reached 1.4 % for the gtfB gene, 3.6 % for the rtfA gene and 4.0 % for the mtfC gene. The divergence of M. avium strains from M. intracellulare within the available region was much greater, reaching ~15 % for each pairwise comparison.

The 13 M. avium strains, for which sequences were obtained, were represented by 9 different genotypes, where a unique combination of two markers defined a particular genotype. Strains 101, 2993, 104, 13528-1079 and SJB#2 shared the same genotype.

**Phylogeny**

The analysis of the rtfA-mtfC region resulted in trees with basically identical topology regardless of the inference method applied (Fig. 1). Three clades (named here A, B and
C) could be distinguished within M. avium, with strains 724, 2151, 6195, B-92 and 17584-286 (clade C) forming a sister group to the remaining strains, among which Sparrow 185 (clade B) emerged as a distinct basal branch. This topology was well supported by bootstrap under MP, but support under ML was lower, especially for the clustering of clades A and B.

Apart from being less resolved, trees based on the gtfB gene fragment differed from the rtfA-mtfC trees in the position of strain 2151 within clade A rather than clade C. This discrepancy was well supported by bootstrap. The other difference concerned strain A5, which, instead of being placed within clade A, was inferred as a sister to (Sparrow 185 + remaining strains from clade A). The position of strain A5 in the trees, however, was based on 4 out of 14 informative positions present in the gtfB alignment and the signal from the remaining 10 informative positions was not in conflict with the topology inferred from the rtfA-mtfC marker. It is likely that the characters supporting the A5 location in the gtfB tree in those four positions do not have a single origin, but have arisen by independent substitutions or other mechanisms not involving vertical inheritance.

A phylogenetic analysis of both markers combined was not performed because the two datasets were significantly incongruent according to the ILD test (P < 0·001).

In the present dataset, three serotypes were represented by more than one strain. Although limited, this sample allowed us to address the question of whether strains with a given serotype have a clonal origin. Our results suggest that this is likely not to be the case, at least for the serotype 1 strains, here inferred as polyphyletic, i.e. not sharing the most recent common ancestor. As mentioned earlier, serotypic differences in M. avium are caused by a variation in the ssGPL composition which is controlled by a complex glycosylation pathway. Most of the molecular mechanisms and involvement of individual gene products in this pathway have yet to be elucidated. However, it has been suggested that during ssGPL biosynthesis the sugars are added singly and sequentially to the lipopeptide core (Eckstein et al., 1998). This hypothesis may be fundamental for the understanding of the polyphyly of serotype 1 strains observed in this study. The oligosaccharide chain in the ssGPL of serotype 1, consisting of a disaccharide unit observed in this study. The oligosaccharide chain in the ssGPL of serotype 1, consisting of a disaccharide unit 

\[
\text{Rha}_2 \text{Rha}_2, \text{Rha}_2 \text{Rha}_2, \text{Rha}_2 \text{Rha}_2, \text{Rha}_2 \text{Rha}_2
\]

is the simplest and shortest among all other serotypes (Table 2). With the exception of serotypes 5, 10 and 11, which are distinguished by the 3-O-Me-\(\alpha\)-L-Rhap-(1→2)-L-dTal component, this unit is present in all serotypes, serving as a core to which the next sugar is added during oligosaccharide chain elongation (Chatterjee & Khoo, 2001). Insertional inactivation by a mobile element or deletion of the gene responsible for the addition of that sugar would lead to the formation of the serotype 1 GPL, regardless of the presence of genes encoding enzymes necessary for further modification of the GPL. This would result in a phylogenetic pattern of serotype 1 having multiple origins. Feasibility of such a scenario is supported by an observation of a spontaneous deletion of the GPL genomic region mediated by IS1601 that caused complete loss of GPL in strain 2151 (Eckstein et al., 2000). It is also supported by a proposed GPL biosynthesis pathway model in which a serotype 1-specific GPL is regarded as an intermediate for the synthesis of the serotype 2 GPL (Eckstein et al., 2003). Moreover, this scenario is consistent with genetic heterogeneity of serotype 1 strains revealed by IS1245-based RFLP typing (this study; Ricacco et al., 1998). Clustering of strains belonging to serotypes 2, 3 and 9 in clade C correlates with the very similar structure of their oligosaccharides (Table 2), strongly suggesting a common origin of these strains. From this perspective, strain B-92, belonging to serotype 1, seems not to fit its nested position within clade C. It is worth noting that strains belonging to

http://mic.sgmjournals.org

1703
serotypes 8 and 21, having very similar GPL structure, were clustered in clade A.

Is there a correlation between our results and the sequence data used in earlier phylogenetic studies? For the comparison we used the ITS locus, which was shown to have higher discriminatory power among _M. avium_ strains compared to the _hsp65_ gene (Frothingham & Wilson, 1993; Mijs et al., 2002). The correlation of the ITS sequence variants with our marker phylogenies was evaluated by mapping each strain’s sequevar designation on trees inferred in this study (Fig. 1). Sequevar data were obtained either by sequencing (strains 101, 104, 2151 and A5) or from the literature (Frothingham & Wilson, 1993). With the exception of strain 17584-286, corresponding to sequevar Mav-D, all of the analysed strains belonged to sequevars Mav-A or Mav-B. There was only rough correspondence between sequevars and inferred clades; sequevars were not monophyletic in either tree. This finding, however, may have no biological significance considering the fact that Mav-A is differentiated from Mav-B by only one mutational change, which could have occurred independently in different strains belonging to a given sequevar. The notion of independent acquisition of the same sequevar is supported by discrepancies between Mav-A sequevar designations and IS1245 RFLP patterns (Fig. 1).

IS1245 RFLP typing has become a frequently used molecular tool for distinguishing isolates within the MAC (Guerrero et al., 1995; van Soolingen et al., 1998). Studies have shown that human _M. avium_ isolates almost invariably displayed polymorphic multibanded IS1245 RFLP patterns (Guerrero et al., 1995; Ritacco et al., 1998), similar to isolates from pigs (Komijn et al., 1999). In contrast, bird isolates were found to possess identical three-banded patterns (Ritacco et al., 1998). Mijs et al. (2002) proposed the designation of these bird-type _M. avium_ isolates [additionally characterized by (1) being unable to grow at 24 and 45 °C, (2) belonging to Mav-A ITS sequevar and (3) being associated with serotypes 1, 2 and 3] as a separate, evolutionarily conserved taxon. The limited number of _M. avium_ serotypes that cause infections in humans and similarity in their genotypes manifested by multibanded

### Table 2. Comparison of the structure of known serotype-specific GPLs from _M. avium_ serotypes used in present study

<table>
<thead>
<tr>
<th>Serotype</th>
<th>Structure of GPL</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>x-L-Rhap-(1→2)-L-dTal</td>
<td>McNeil et al. (1989)</td>
</tr>
<tr>
<td>2</td>
<td>Ac-2,3-di-O-Me-z-L-Fucp-(1→3)-x-L-Rhap-(1→2)-L-dTal</td>
<td>Riviere &amp; Puzo (1992)</td>
</tr>
<tr>
<td>3</td>
<td>2,3-di-O-Me-z-L-Fucp-(1→4)-β-D-GlcPA-(1→4)-2,3-di-O-Me-z-L-Fucp-(1→3)-x-L-Rhap-(1→2)-L-dTal</td>
<td>Khoo et al. (1999)</td>
</tr>
<tr>
<td>4</td>
<td>4-O-Ac-2,3-di-O-Me-z-L-Fucp-(1→4)-β-D-GlcPA-(1→4)-2,3-di-O-Me-z-L-Fucp-(1→3)-x-L-Rhap-(1→2)-L-dTal</td>
<td>Hough &amp; Theobald (1963)</td>
</tr>
<tr>
<td>5</td>
<td>4-O-Me-z-L-Rhap-(1→4)-2-O-Me-z-L-Fucp-(1→3)-x-L-Rhap-(1→2)-L-dTal</td>
<td>Aspinall et al. (1991)</td>
</tr>
<tr>
<td>6</td>
<td>4,6-O-(1-carboxyethylidene)-3-O-Me-β-D-Glc-(1→3)-x-L-Rhap-(1→2)-L-dTal</td>
<td>McNeil et al. (1989)</td>
</tr>
<tr>
<td>7</td>
<td>4,6-O-(1-carboxyethylidene)-β-D-Glc-(1→3)-x-L-Rhap-(1→2)-L-dTal</td>
<td>Khoo et al. (1999)</td>
</tr>
</tbody>
</table>

Alternative placement of strain 2151 in the trees inferred with two analysed markers begs an explanation. Significant discrepancy between trees (Shimodaira-Hasegawa test, _P_ < 0.001) hints at horizontal gene transfer as a likely cause (Philippe & Douady, 2003). Further evidence for the involvement of gene transfer comes from our analysis of a larger genomic region encompassing the GPL gene cluster (Krzywinska et al., 2004). There we show that a short chromosome region encompassing a _gtfB_ gene fragment used for the phylogenetic analysis has been transferred between _M. avium_ strains. Because the _rtfA-mtfC_ marker was outside of the transferred region, we hypothesize that _rtfA-mtfC_ provides a better estimation of _M. avium_ phylogeny than _gtfB_. However, further studies of MAC phylogeny should utilize other new markers, possibly outside the GPL cluster. Despite a considerable amount of phylogenetic information, a prerequisite of a good phylogenetic marker, _rtfA-mtfC_ cannot be used for tracing the history of the entire group because it was not present in the genomes of all strains studied. It is conceivable that the same will apply to other genes from the cluster, when a range of strains, representing all serotypes, are considered. Moreover, this study clearly shows that phylogenetic inference, to be meaningful, should be based on more than one marker.

Our analysis suggests that serotypes prevalent in AIDS patients have multiple origins. Further study involving a larger number of serotypes and different genomic regions would demonstrate if our hypotheses hold. Here we have proposed a simple scenario for a secondary gain of the serotype 1 GPL that would result in lack of concordance between serotype delineation and the distribution of

**References**

particular genotypic and phenotypic characters in strains belonging to serotype 1. It cannot be excluded that a similar lack of concordance concerning serotypes 2 and 3 (Komijn et al., 1999; Ritacco et al., 1998) is a consequence of horizontal transfer. However, answering this issue would require further genomic studies involving a larger number of representative strains.

ACKNOWLEDGEMENTS

We thank Andrea Cooper for providing M. avium 724 and 2151, Kathleen Eisenach for providing M. avium AT and Eric Brown for providing M. avium 104 and 101. We acknowledge Colorado State University and NIH for providing MAC reference strains (NIAID contract AI-175220 entitled 'TB Research Materials and Vaccine Testing Program') and the Institute for Genomic Research for providing sequences of M. avium strain 104. We thank Nora J. Besansky for access to computing and Frank Collins for use of the sequencing facility. We also thank Hope Hollocher for her careful reading of the manuscript and helpful suggestions. This work was supported by a grant from the United States Department of Agriculture, CSREES.

REFERENCES


