Naturally occurring horizontal gene transfer and homologous recombination in *Mycobacterium*

Elżbieta 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

Acquisition of genetic information through horizontal gene transfer (HGT) is an important evolutionary process by which micro-organisms gain novel phenotypic characteristics. In pathogenic bacteria, for example, it facilitates maintenance and enhancement of virulence and spread of drug resistance. In the genus *Mycobacterium*, to which several primary human pathogens belong, HGT has not been clearly demonstrated. The few existing reports suggesting this process are based on circumstantial evidence of similarity of sequences found in distantly related species. Here, direct evidence of HGT between strains of *Mycobacterium avium* representing two different serotypes is presented. Conflicting evolutionary histories of genes encoding elements of the glycopeptidolipid (GPL) biosynthesis pathway led to an analysis of the GPL cluster genomic sequences from four *Mycobacterium avium* strains. The sequence of *M. avium* strain 2151 appeared to be a mosaic consisting of three regions having alternating identities to either *M. avium* strains 724 or 104. Maximum-likelihood estimation of two breakpoints allowed a ~4100 bp region horizontally transferred into the strain 2151 genome to be pinpointed with confidence. The maintenance of sequence continuity at both breakpoints and the lack of insertional elements at these sites strongly suggest that the integration of foreign DNA occurred by homologous recombination. To our knowledge, this is the first report to demonstrate naturally occurring homologous recombination in *Mycobacterium*. This previously undiscovered mechanism of genetic exchange may have major implications for the understanding of *Mycobacterium* pathogenesis.

INTRODUCTION

Increase in genic content is an important evolutionary mechanism shaping prokaryotic genomes. This dynamic process, counterbalanced by DNA loss, is affected by either duplication of sequences existing in the genome or acquisition of foreign genetic material through horizontal gene transfer (HGT). Novel DNA transferred into the cell is thought to be a major source of new phenotypic characteristics, often allowing survival under strong selection regimes. In pathogenic bacteria, for example, well known cases of transfer include spread of genes conferring antibiotic resistance and virulence determinants (Morschhauser *et al.*, 2000; Ochman *et al.*, 2000; Weigel *et al.*, 2003). HGT can bring not only entirely new sequence families into the genome, but also sequences that are homologous to existing genes (Ochman, 2001), which may lead to a replacement of autochthonous sequences with the acquired copies through homologous recombination.

The genus *Mycobacterium* contains over 70 species of human and animal pathogens as well as non-pathogenic saprophytes. The major human pathogens are *Mycobacterium tuberculosis*, the causative agent of tuberculosis, *Mycobacterium leprae*, causing leprosy, and *Mycobacterium avium*, responsible for opportunistic infections, particularly in AIDS patients. It is estimated that *M. tuberculosis* infects up to one-third of the world’s population and results in 3 million deaths annually (Snider, 1994).

In *Mycobacterium* HGT has not been clearly demonstrated. Several existing reports suggesting this process are based on circumstantial evidence involving similarity in sequences found in distantly related bacterial or even eukaryotic species (Gamieldien *et al.*, 2002; Kinsella *et al.*, 2003; Le Dantec *et al.*, 2001; Poelarends *et al.*, 2000). In addition, population genetic studies indicate that, unlike other bacteria, mycobacteria seem not to exchange genetic material between individuals. Genome-wide multilocus analyses of *M. tuberculosis* and *Mycobacterium bovis* environmental samples consistently detected highly significant linkage disequilibrium, suggesting either extreme rarity or non-existence of recombination in natural settings.
This notion is consistent with early unsuccessful attempts at mutagenesis via allelic exchange (homologous recombination of native gene with an inactivated copy) in these slow-growing mycobacteria. It has been suggested that the unusual structure of the Mycobacterium recA gene, which encodes a key protein involved in recombination, DNA repair and regulation of the SOS response (Walker, 1984), may be responsible for inefficient homologous recombination in the M. tuberculosis complex (McFadden, 1996). The recA gene in these mycobacteria and in M. leprae is interrupted by an in-frame ORF encoding an intein that is removed from a precursor protein by a protein-splicing reaction (Davis et al., 1992, 1991). However, subsequent experiments showed that the intein does not affect RecA protein function and the frequency of double cross-over homologous recombination events (Papavinasasundaram et al., 1998).

To date, gene knock-outs using allelic exchange have been successfully achieved in both fast-growing Mycobacterium smegmatis (Husson et al., 1990) and slow-growing M. tuberculosis (Balasubramanian et al., 1996) and M. bovis BCG (Aldovini et al., 1992). Pavelka & Jacobs (1999) demonstrated that the recombination frequencies in M. smegmatis, M. tuberculosis and M. bovis were similar, suggesting that the homologous recombination machinery in fast- and slow-growing mycobacteria functions with comparable efficiency.

In another study (Krzywinska et al., 2004) we performed a phylogenetic analysis of 13 M. avium strains using as markers two DNA fragments (a 926 bp fragment of the gtfB gene and a 2150 bp region spanning rtfA-mtfC genes; cf. Fig. 1a) lying in each other’s proximity within the gene cluster responsible for biosynthesis of glycopeptidolipids (GPLs). Trees inferred from each marker had significantly incongruent topologies due to a well supported alternative placement of M. avium strain 2151. Such incongruence strongly suggests HGT involving one of the markers between the studied strains (Ragan, 2001). Here we present the results of further analysis of the available orthologous sequences of the GPL biosynthesis gene cluster from M. avium strains 2151, 104, 724 and A5, and show unequivocal evidence for HGT and homologous recombination in Mycobacterium.

**METHODS**

**Nucleotide sequences.** The nucleotide sequences of the GPL biosynthesis gene cluster were retrieved from GenBank for the following M. avium strains: A5 (accession no. AY130970), 724 (AF125999) and 2151 (AF143772). The corresponding sequence from M. avium strain 104 was obtained from The Institute for Genomic Research website (www.tigr.org).

**Sequence analysis.** The sequences were aligned using CLUSTALX (Thompson et al., 1997). The putative recombinant and parental

---

**Fig. 1.** (a) A schematic of the 5’ end of the GPL biosynthesis cluster in M. avium strain A5. Arrows show the position and orientation of individual genes, a line between the arrows represents intergenic spacers and dotted lines under the schematic denote the position of marker sequences yielding discordant phylogenies (Krzywinska et al., 2004). (b) An alignment of the 5’ end of the GPL biosynthesis cluster. The variable positions among four strains of M. avium for which the sequence was available are shown; nucleotides at positions variable between strains 104, 2151 and 724 are in bold type. Numbers above the sequences (read from top to bottom) correspond to alignment positions. Portions of identical sequences are boxed; the box around sequences identical in strains 104 and 2151 corresponds to a ~4000 bp fragment horizontally transferred to strain 2151. Contiguous lines above the alignment mark gene sequences and the portions without lines correspond to intergenic spacers. A broken line under the alignment shows the regions corresponding to the marker sequences. The insertion point of the IS1245 sequence is marked by an arrow.
sequences were identified upon inspection of variable alignment sites retrieved using MEGA 2.1 (Kumar et al., 2001). The putative recombinant sequence was subjected to a maximum-likelihood analysis using the program LARD developed by Holmes et al. (1999). The likelihood of the null hypothesis, H0, that there has been no recombination event and the likelihood of the hypothesis H1, that recombination took place, were assessed from a simple unrooted tree of a putative recombinant (strain 2151) and two ‘parental’ sequences (strains 724 and 104). The likelihood ratio test was performed using the joint likelihood of the two trees on either side of the putative breakpoint and the likelihood of a single tree was calculated on the entire alignment. To assess whether H1 is a significantly better fit to the data, the likelihood ratio of experimental data was compared to a null distribution obtained by LARD analysis of 500 sequence datasets generated by a Monte Carlo simulation of clonal evolution using Seq-Gen (Rambaut & Grassly, 1997).

RESULTS AND DISCUSSION

The flood of novel sequence information from an increasing number of organisms opens unprecedented opportunities for comparative studies allowing new insights into the fundamental mechanisms of evolutionary change. One of the results of such studies is a growing recognition that HGT is an important evolutionary force acting upon bacterial genomes. However, despite numerous reports in which HGT has been invoked, the actual transfer events are rarely observed (Ochman, 2001).

Our recent phylogenetic analysis of M. avium strains using two closely linked markers, gtfB and rtfA-mtfC from the GPL biosynthesis gene cluster, unexpectedly revealed a potential horizontal transfer of one of the marker sequences between the strains belonging to different serotypes (Krzywinska et al., 2004). The notion of transfer was based on a well-supported alternative position of one of the strains (i.e. 2151) in the trees inferred from each marker individually. Inspection of the sequence substitution patterns from all studied strains revealed that within the rtfA-mtfC marker, strain 2151 shared identical sequence with strain 724, whereas within the gtfB gene fragment it was identical to five other strains, including 104, but substantially different from strain 724.

To further explore the transfer hypothesis and identify a potential sequence breakpoint, we retrieved and analysed all M. avium genomic sequences that exist in the publicly available databases which corresponded to the GPL cluster. These included sequences from strains 104, 724, 2151 and A5. We focused our study on the highly conserved 5’ region of the GPL cluster stretching between the mtfB and gtfB genes (Krzywinska & Schorey, 2003). The alignment of the four 10624 bp sequences was unambiguous. The pattern of the variable alignment positions within that region provided clear evidence for a mosaic structure within the strain 2151 genomic sequence (Fig. 1). Its 5’ half, containing the rtfA-mtfC marker, was identical to the sequence from strain 724. A potential breakpoint was located within the intergenic spacer between mtfC and mtfD genes, at which point the subsequent sequence of strain 2151 shared identity with strain 104. Remarkably, another potential breakpoint was identified at the end of the aligned region, within the gtfB gene. Downstream from the gtfB marker sequence, the pattern of distribution of nucleotide substitutions reverted to the one observed within the 5’ half of the alignment. To substantiate our hypothesis of gene transfer we performed sequence analysis using the LARD maximum-likelihood method (Holmes et al., 1999), in which the likelihoods of the null hypothesis, H0 (i.e. there has been no recombination event at the analysed regions) and of the alternative hypothesis, H1, were calculated. This method allowed us to identify the two breakpoint regions within the alignment, which were found to be before positions 6219 and 10340. Monte Carlo simulation used to assess whether H1 is a significantly better fit to the data (i.e. that the likelihood ratio of H1 to H0 is greater than what we would expect by chance) revealed that the log likelihood ratio of the real data (recombination event involving region spanning positions 6219–10340) was significantly greater than for any of the simulated dataset (73.4 vs maximum of 15.09). A mosaic structure of the aligned region with blocks of sequences having unambiguously different evolutionary affiliations and identification of both breakpoints allowed us to pinpoint with confidence an entire region horizontally transferred into strain 2151. The maintenance of sequence continuity (lack of indels) and the absence of mobile elements within both breakpoints clearly points to homologous recombination as a mechanism of the foreign DNA integration into the strain 2151 genome.

It is worth noting that within the recombination region an insertion sequence IS1245 was embedded, although the significance of this finding remains unknown. Insertion sequence (IS) elements are known to modify a bacterial genome by mediating mobility of DNA fragments within an individual. The widespread occurrence and high degree of similarity between IS elements in different Mycobacterium species suggest that they spread across members of the genus through horizontal transfer (Gordon et al., 1999; Howard et al., 2002). However, evidence from the analysis of the M. tuberculosis H37Rv genome shows no apparent link between potential IS transfer and introduction of novel genes into the bacillus (Gordon et al., 1999).

Correct identification of ancient HGT events is difficult and is associated with a high degree of uncertainty. Many of the reports claiming such ancient transfer based on sequence similarity or phylogenetic distribution of sequences are likely to be false positives, and at least one such case apparently concerns mycobacteria. It has been proposed that 19 genes found in the M. tuberculosis genome, and implicated to have a role in mycobacterial pathogenesis, are of eukaryotic origin (Gamieldien et al., 2002). However, this finding was shown to be an artifact of insufficient taxon sampling, when the genes were reanalysed in the context of a larger dataset (Kinsella & McInerney, 2003). Recently Kinsella et al. (2003) suggested that a close phylogenetic affiliation of genes involved in fatty acid
biosynthesis in mycobacteria and proteobacteria, two
distantly related groups of organisms, is a signature of
HGT. Although plausible, considering a broad spectrum
of taxa used for the analysis, their hypothesis leaves a tinge
of suspicion, because the codon usage bias in the puta-
tively transferred genes was typical of Mycobacterium genes.
For this reason, ancient gene duplications and subsequent
retention of sequences in unrelated lineages can serve as
an alternative hypothesis explaining the pattern observed in
that study.

Apparently, recent gene transfer has been implicated in
the evolution of the catabolic pathway for the man-made
chemical 1,2-dibromoethane, found in Mycobacterium
sp. strain GP1 (Poelarends et al., 2000). This highly toxic
compound is degraded by the haloalkane dehalogenase
(dhaA) gene product. A highly conserved dhaA gene was
also found in phylogenetically distant Pseudomonas pavo-
naeae and Rhodococcus rhodochrous (Kulakova et al., 1997),
organisms isolated, like Mycobacterium sp. strain GP1,
different strongly contaminated environmental samples. Remarkably, in Rhodococcus rhodochrous the dhaA
region was located on the autotransmissible plasmid
pRTL1 (Kulakova et al., 1997). Another case of potentially
recent HGT was suggested by the sequence analysis of the
linear plasmid pCLP from Mycobacterium celatum, which
revealed loci with high nucleotide sequence identity to loci
on the M. tuberculosis chromosome (Le Dantec et al., 2001).
Moreover, the acquisition of a plasmid encoding myco-
lactone, a key toxin responsible for skin tissue destruction,
was likely to be a recent event, resulting in the emergence
of Mycobacterium ulcerans as a causative agent of Buruli
ulcer (Stinear et al., 2004). Whereas the mycolactone gene
cluster in M. ulcerans remained on a plasmid, the other
two examples of presumably recent HGT implied incor-
poration of the transferred genes into the genome by either
excise-integrative recombination or integrase-dependent
acquisition. Our study is the first to report naturally
occurring transfer and incorporation by homologous
recombination of the transferred DNA into the Myco-
bacterium chromosome. The transfer most likely occurred
very recently, since the recombination event has not been
observed by subsequent nucleotide substitutions within
the replaced region, as shown by its comparison with the
parental sequence. Recent successful allelic exchange experi-
ments confirmed the feasibility of homologous recombina-
tion in M. avium (Maslow et al., 2003; Otero et al., 2003;
Krzywinska & others, unpublished). In two of those studies
a gene knock-out has been achieved in the rtfA and mtfD
genes located within the GPL cluster (Maslow et al., 2003;
Krzywinska & others, unpublished).

Several mechanisms have been proposed for the exchange
of chromosomal DNA between bacteria by homologous
recombination as a consequence of conjugation, transduc-
tion and transformation. However, there is little informa-
tion regarding the significance of these mechanisms in
Mycobacterium. It has been shown that mycobacterial
plasmids can replicate within most Mycobacterium species,
so they can theoretically be spread horizontally, promoting
gene transfer between mycobacteria (Kirby et al., 2002;
Le Dantec et al., 2001). Several plasmids were found in the
M. avium complex (MAC) (Hellyer et al., 1991; Meissner &
Falkingham, 1986), among them plasmid pVT2, which is
thought to be conjugal (Kirby et al., 2002), and raises the
possibility for conjugal transfer in MAC – a mechanism
known to occur between M. smegmatis strains (Parsons
et al., 1998) and between M. smegmatis and other bacterial
species (Gormley & Davies, 1991). The widespread occur-
rence of related plasmids from M. avium, Mycobacterium
intracellulare and Mycobacterium scrofulaceum (Jucker &
Falkingham, 1990) suggests that the plasmids have the
ability to transfer between hosts in the environment,
where MAC strains are ubiquitous. Transfer can also
occur via a transformation process. Spontaneous plasmid
transformation was discovered in M. smegmatis (Bhatt
et al., 2002) and natural competence for transformation by
chromosomal DNA was reported in M. avium (Tsukamura
et al., 1960). Moreover, DNA fragments can be transferred
between bacterial cells in a transduction process mediated
by bacteriophages, over 250 of which have been described
from Mycobacterium (Hatfull & Jacobs, 1994). Genomic
characterization of 14 mycobacteriophages revealed within
their genomes over 50 genes not associated with phage
growth (Pedulla et al., 2003). Intriguingly, two of these
phages contained a close homologue of a gene found in
M. leprae and M. tuberculosis, which encodes the antigen
Lsr2, a strong stimulator of the immune response, hinting
at a possible role of phages in mycobacterial virulence. It
is not clear if mycobacteriophages transmit such genes to
bacterial genomes (Pedulla et al., 2003); however, experi-
ments in which mycobacteriophages have been used to
create systems for gene delivery in mycobacteria (Bardarov
et al., 1997) show the feasibility of such a process.

Within pathogenic species, recombination may have impor-
tant ramifications regarding adaptation to host, response to
control or tracing of strains for epidemiological purposes
(Ochman, 2001). The presence of different gene delivery
mechanisms and functional homologous recombination
machinery in various Mycobacterium species raises the
possibility of naturally occurring transfer and recombi-
nation not only in M. avium, but also in other members of
the genus. Moreover, the evidence presented in this study
may contribute to the understanding of the mechanisms
of drug resistance development in Mycobacterium. In
M. tuberculosis all antibiotic resistance is chromosomally
mediated and results from point mutations in different
genomes (depending on the antibiotic), rather than from
acquisition of new genetic elements encoding antibiotic-
altering enzymes (Musser, 1995). The rate of spontaneous
mutations in vitro is low in M. tuberculosis, relative to the
range noticed in other bacteria (David & Newman, 1971).
This contrasts with a high frequency of emergence of drug
resistance in tuberculosis patients (Espinal et al., 2000).
Likewise, multiple drug resistance, which occurs in 1–3 %
of total tuberculosis isolates, is inexplicably high. It has been shown that DnaE2 polymerase, which is responsible for inducible mutagenesis in M. tuberculosis, can contribute directly to the appearance of resistance in vivo (Boshoff et al., 2003). Homologous recombination brings a new potential explanation for the generation of multiple drug resistance. Although HGT has not been shown to be involved in the generation of drug resistance, it is possible that the mutated allele conferring resistance to a given antibiotic in one Mycobacterium can be transferred to another. If this isolate, containing a different resistance phenotype, incorporated the mutated allele into its genome through homologous recombination, one could obtain a multidrug-resistant strain. Simultaneous infection of patients by two different strains of M. tuberculosis (Braden et al., 2001) or M. avium (von Reyn et al., 1995) also creates the opportunity for such a DNA exchange.

In summary, we have unequivocally shown naturally occurring homologous recombination within the GPL biosynthesis cluster of M. avium 2151. At present, it is not yet known how often recombination of highly selected loci contributes to pathogenesis or the generation of multidrug resistance in mycobacteria.

ACKNOWLEDGEMENTS

We thank the Institute for Genomic Research for providing sequences of M. avium strain 104. This work was supported through a grant from the United States Department of Agriculture/CSREES.

REFERENCES


