JOURNAL OF BACTERIOLOGY, Sept. 2004, p. 6332–6334 Vol. 186, No. 18
0021-9193/04/$08.00+0 DOI: 10.1128/JB.186.18.6332–6334.2004

Extensive Genomic Polymorphism within *Mycobacterium avium*†

Makeda Semret,[1] Gary Zhai,[1] Serge Mostowy,[1] Cynthia Cleto,[1] David Alexander,[1] Gerard Cangelosi,[2] Debby Cousins,[3] Desmond M. Collins,[4] Dick van Soolingen,[5] and Marcel A. Behr[1]*

McGill University Health Centre, Montreal, Canada[1]; Seattle Biomedical Research Institute, Seattle, Washington[2]; Australian Reference Laboratory for Bovine Tuberculosis, Department of Agriculture, South Perth, Australia[3]; AgResearch, Wallaceville Animal Research Centre, Upper Hutt, New Zealand[4]; and National Institute of Public Health and the Environment, Bilthoven, The Netherlands[5]

Received 27 January 2004/Accepted 27 June 2004

We have initiated comparative genomic analysis of *Mycobacterium avium* subspecies by DNA microarray, uncovering 14 large sequence polymorphisms (LSPs) comprising over 700 kb that distinguish *M. avium* subsp. *avium* from *M. avium* subsp. *paratuberculosis*. Genes predicted to encode metabolic pathways were overrepresented in the LSPs, and analysis revealed a polymorphism within the mycobactin biosynthesis operon that potentially explains the in vitro mycobactin dependence of *M. avium* subsp. *paratuberculosis*.

The *Mycobacterium avium* complex (MAC) comprises a group of closely related organisms responsible for a broad range of diseases in humans and livestock. *M. avium* subsp. *avium* causes cervical lymphadenitis in children and disseminated disease in AIDS patients, while *M. avium* subsp. *paratuberculosis* causes an inflammatory bowel disease in ruminants and possibly humans (2, 6). As MAC organisms are highly prevalent in the environment (12), their genomic complement is predicted to also reflect this lifestyle.

Recent work in mycobacterial genomics has revealed that genomic reduction through the loss of large sequence polymorphisms (LSPs) is a major contributor to genetic diversity. Studies of the *Mycobacterium tuberculosis* complex have used LSPs for inferences of phylogenetics (5, 10) and biological properties such as virulence (9, 13). Since previous DNA hybridization and sequencing studies have shown that *M. avium* subspecies are indistinguishable at the species level (14) and that they share about 98% sequence identity in coding regions (1), we hypothesized that LSPs would be important sources of genetic variability among MAC organisms.

We have annotated the sequence of *M. avium* subsp. *avium* strain 104 (provided by the Institute for Genomic Research [http://www.tigr.org]) in order to assemble a whole-genome DNA microarray representing the predicted coding sequences (details on the annotation are provided at www.molepi.mcgill.ca/MAC.htm). Seventy-base-pair-long oligonucleotide probes were designed and synthesized (MetaBion GmbH, Martinsried, Germany) for 4,158 of 4,480 predicted open reading frames (ORFs). Each probe was printed in duplicate onto microarray slides (SigmascreenTM; Sigma) by using a microarray robot (Virtek Chipwriter model SDDC2) to permit genomic DNA comparisons of *M. avium* subsp. *avium* strain 104 and the following strains: (i) *M. avium* subsp. *paratuberculosis* K10 (cow strain), (ii) *M. avium* subsp. *paratuberculosis* LN20 (sheep strain), and (iii) *M. avium* subsp. *silvaticum* 49884 (ATCC strain). Cohybridization experiments were performed by using previously published methods to screen for regions of six or more contiguous *M. avium* subsp. *avium* 104 ORFs absent from the test isolate (3); these regions were then confirmed by PCR and sequencing (10). In a second step, primers used to confirm the presence or absence of a region were used to test a panel of 43 isolates in order to determine the distribution of these LSPs across other samples.

Microarray comparisons revealed 14 LSPs (LSP1 to LSP14) ranging in length from 21 to 197 kb (Table 1) and encompassing 572 genes (see Table SA in the supplemental material). Combined, these LSPs comprise 727 kb and represent 13.5% of the *M. avium* subsp. *avium* 104 genome. This remarkable diversity far exceeds the genomic variability described among *M. tuberculosis* complex isolates, estimated to be 1.7% of the genome (9, 11). Moreover, the MAC diversity documented here must be considered a minimum estimate, as only very large LSPs uncovered from comparisons of just four clinical isolates were studied. Through the study of isolates from broader sampling frames and diverse environments, one would expect even greater genomic variability to be revealed.

The exact sizes and locations of the LSPs, the subspecies from which they are missing, and the key features of each LSP are shown in Table 1. Seven of the LSPs revealed are simple genomic deletions or insertions compared to the reference strain *M. avium* subsp. *avium* 104. The other seven LSPs involve a more complex combination of insertion and deletion events. This complexity indicates that the genome of MAC organisms is the product of both vertical inheritance, as seen in the *M. tuberculosis* complex, and horizontal acquisition of DNA. Although plasmids have been described for *M. avium* isolates, the reference strain *M. avium* subsp. *avium* 104 does not contain a plasmid, indicating that the genomic variability described here involves chromosomal DNA.

In terms of predicted gene function based on homology searches, genes encoding proteins involved in information

* Corresponding author. Mailing address: Division of Infectious Diseases and Medical Microbiology, A5-156, Montreal General Hospital, 1650 Cedar Ave., Montreal, Quebec H3G 1A4, Canada. Phone: (514) 934-1934, ext. 42815. Fax: (514) 934-8423. E-mail: marcel.behr@mcgill.ca.

† Supplemental material for this article may be found at http://jb.asm.org/.

TABLE 1. LSP characteristics and distribution across *M. avium* subspecies

Sequence[a]	Start[b]	End[b]	Number of ORFs	Key features (predicted functions)	Presence in[c]:		
					M. avium subsp. *avium*	*M. avium* subsp. *silvaticum*	*M. avium* subsp. *paratuberculosis*
LSP1*	2,549,110	2,728,236	160	*mce3* operon (intermediary and lipid metabolism)	+/−	−	−
LSP2	3,917,471	3,939,509	17	Possible prophage (unknown)	+/−	−	−
LSP3*	254,272	294,378	7	Probable prophage	+/−	−	−
LSP4*	1,795,197	1,992,429	170	Intermediary and lipid metabolism (mycobactin synthesis)	+/−	−	−
LSP5*	746,437	794,502	14	Probable prophage	+/−	−	−
LSP6	5,173,499	5,270,803	84	Hydrogen metabolism (unknown)	+/−	+	−
LSP7	462,328	493,802	25	Transposable elements (unknown)	+/−	−	−
LSP8	5,122,380	5,132,388	12	Protease-encoding operon (regulation)	+	+	−
LSP9*	3,394,920	3,414,585	22	Glycopeptidolipid cluster	+/−	+/−	−
LSP10*	2,220,300	2,241,562	14	Probable prophage	+/−	−	−
LSP11	4,674,473	4,682,256	7	Part of *mce2* operon	+	+	+/−
LSP12*	665,425	675,801	8	Transposable elements (unknown)	+/−	+/−	−
LSP13	1,443,886	1,463,442	14	Transposable elements (heavy metal transport)	+/−	+	−
LSP14	1,418,088	1,441,399	18	Transposable elements (unknown)	+	+	−

[a] LSPs marked with asterisks are results of complex insertion-deletion events. For LSP1, LSP4, and LSP9, coordinates are provided for *M. avium* subsp. *paratuberculosis* K10 only, as the LSP could not be precisely mapped by PCR across all isolates. LSP3, LSP5, and LSP12 are replaced by insertion-like elements in *M. avium* subsp. *paratuberculosis.*

[b] Start and End columns show the distance in base pairs from the start codon of *dnaA.*

[c] Twenty isolates each of *M. avium* subsp. *avium* and *M. avium* subsp. *paratuberculosis* and three isolates of *M. avium* subsp. *silvaticum* were tested. +, LSP consistently present; −, LSP consistently absent; +/−, variable LSP presence.

pathways and proteins of the PE/PPE family were highly conserved among tested strains (0.7 and 0.6% of missing genes, respectively). Considerable diversity within the latter group has been observed in *M. tuberculosis*, where PE/PPE elements are proposed to be an important source of antigenic variation (4). The surprising lack of diversity in *M. avium* subspecies was further confirmed by in silico comparisons of *M. avium* subsp. *avium* 104 to the recently sequenced *M. avium* subsp. *paratuberculosis* K10 (GenBank accession number NC_002944). At the other extreme, genes of unknown function and those predicted to encode proteins involved in lipid metabolism and intermediary metabolism were overrepresented in the LSPs (19.3, 18, and 20.1% of missing genes, respectively). The absence of these genes in the more pathogenic *M. avium* subsp. *paratuberculosis* suggests a greater role for these genes in survival in the environment than in the intracellular milieu. Another highly variable group comprised genes designated mammalian cell entry (*mce*) genes, a group of genes thought to be involved in host cell invasion and hence virulence. *M. avium* subsp. *avium* contains 66 such genes distributed in nine operonic clusters. Of these, 21 (32%) were polymorphic among tested strains. Specifically, one of the two homologs of the *mce3* operon of *M. avium* subsp. *avium* 104 was missing from *M. avium* subsp. *paratuberculosis* and *M. avium* subsp. *silvaticum*, and four of the six genes belonging to the single *mce2* operon were lost in at least one *M. avium* subsp. *paratuberculosis* strain (LN20). The loss of *mce2* and *mce3* genes in the more pathogenic *M. avium* subsp. *paratuberculosis* isolates along with the deletion of *mce3* from virulent *Mycobacterium bovis* (8) together challenge the assignment of these *mce* operons to the category of virulence elements. In contrast, the *mce1* operon, which in *M. tuberculosis* has been associated with a more virulent phenotype (15), was conserved in *M. avium* subsp. *paratuberculosis* and *M. avium* subsp. *silvaticum*.

Orthologs of the mycobactin synthesis operon (*mbtABCD EFGHIJ*) of *M. tuberculosis* were found in *M. avium* subsp. *avium* 104. In *M. avium* subsp. *avium* 104, *mbtJ* is separated from *mbtA* by a large sequence of 197 kb, corresponding to LSP4. In *M. avium* subsp. *paratuberculosis* K10, LSP4 has been replaced by a 19-kb insert which truncates the 1,724-bp *mbtA* gene at position 1081. As MbtA is responsible for an early event in mycobactin synthesis (7), disruption of *mbtA* would predictably impair mycobactin synthesis at its inception and potentially explains the strict dependence of *M. avium* subsp. *paratuberculosis* on this siderophore for in vitro growth.

In conclusion, our results reveal remarkable genomic diversity within the MAC. Further characterization of the LSPs and their distribution across more isolates may suggest reasons for the host species specificities and pathogenic potentials of the *M. avium* subspecies and provide further insight into their complex evolutionary history.

This work was supported by a grant from the Natural Science and Engineering Research Council (grant number GEN2282399). M.S. is a recipient of the CIDS/CIHR/Bayer Healthcare fellowship award and is currently funded by the Fonds de la Recherche en Santé du Québec (FRSQ). M.A.B. is a New Investigator of the Canadian Institutes of Health Research. None of the authors have a conflict of interest or any commercial association that may pose a conflict of interest.

We acknowledge M. Bernstein, M. Kirtsman, M. Katz-Lavigne, D. Shersher, and D. Livingston-Rosanoff for their contributions to this project and thank L. Mutharia, J. Bannantine, B. Brooks, C. Inderlied, H. Huchzermeyer, G. de Lisle, and F. Saxegaard for supplying isolates.

REFERENCES

1. **Bannantine, J. P., Q. Zhang, L. L. Li, and V. Kapur.** 2003. Genomic homogeneity between *Mycobacterium avium* subsp. *avium* and *Mycobacterium avium* subsp. *paratuberculosis* belies their divergent growth rates. BMC Microbiol. **3:**10.
2. **Behr, M. A., M. Semret, A. Poon, and E. Schurr.** 2004. Crohn's disease, mycobacteria, and NOD2. Lancet Infect. Dis. **4:**136–137.
3. **Behr, M. A., M. A. Wilson, W. P. Gill, H. Salamon, G. K. Schoolnik, S. Rane, and P. M. Small.** 1999. Comparative genomics of BCG vaccines by whole-genome DNA microarray. Science **284:**1520–1523.
4. **Brennan, M. J., and G. Delogu.** 2002. The PE multigene family: a 'molecular mantra' for mycobacteria. Trends Microbiol. **10:**246–249.
5. **Brosch, R., 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. 2002. A new evolutionary scenario for the Mycobacterium tuberculosis complex. Proc. Natl. Acad. Sci. USA **99:**3684–3689.

6. **Bull, T. J., E. J. McMinn, K. Sidi-Boumedine, A. Skull, D. Durkin, P. Neild, G. Rhodes, R. Pickup, and J. Hermon-Taylor.** 2003. Detection and verification of *Mycobacterium avium* subsp. *paratuberculosis* in fresh ileocolonic mucosal biopsy specimens from individuals with and without Crohn's disease. J. Clin. Microbiol. **41:**2915–2923.
7. **Crosa, J. H., and C. T. Walsh.** 2002. Genetics and assembly line enzymology of siderophore biosynthesis in bacteria. Microbiol. Mol. Biol. Rev. **66:**223–249.
8. **Gordon, S. V., R. Brosch, A. Billault, T. Garnier, K. Eiglmeier, and S. T. Cole.** 1999. Identification of variable regions in the genomes of tubercle bacilli using bacterial artificial chromosome arrays. Mol. Microbiol. **32:**643–655.
9. **Kato-Maeda, M., J. T. Rhee, T. R. Gingeras, H. Salamon, J. Drenkow, N. Smittipat, and P. M. Small.** 2001. Comparing genomes within the species Mycobacterium tuberculosis. Genome Res. **11:**547–554.
10. **Mostowy, S., D. Cousins, J. Brinkman, A. Aranaz, and M. A. Behr.** 2002. Genomic deletions suggest a phylogeny for the Mycobacterium tuberculosis complex. J. Infect. Dis. **186:**74–80.
11. **Mostowy, S., A. G. Tsolaki, P. M. Small, and M. A. Behr.** 2003. The in vitro evolution of BCG vaccines. Vaccine **21:**4270–4274.
12. **Primm, T. P., C. A. Lucero, and J. O. Falkinham III.** 2004. Health impacts of environmental mycobacteria. Clin. Microbiol. Rev. **17:**98–106.
13. **Pym, A. S., P. Brodin, R. Brosch, M. Huerre, and S. T. Cole.** 2002. Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines Mycobacterium bovis BCG and Mycobacterium microti. Mol. Microbiol. **46:**709–717.
14. **Saxegaard, F., I. Baess, and E. Jantzen.** 1988. Characterization of clinical isolates of Mycobacterium paratuberculosis by DNA-DNA hybridization and cellular fatty acid analysis. APMIS **96:**497–502.
15. **Shimono, N., L. Morici, N. Casali, S. Cantrell, B. Sidders, S. Ehrt, and L. W. Riley.** 2003. Hypervirulent mutant of Mycobacterium tuberculosis resulting from disruption of the mce1 operon. Proc. Natl. Acad. Sci. USA **100:**15918–15923.