New York, ed. 7, 1995), pp. 1853–1885. Abetalipoproteinemia is an autosomal recessive disease in which plasma VLDL and LDL are virtually absent. Affected people have fat malabsorption and have triglyceride accumulation in enterocytes and the liver. Secondary to a vitamin E deficiency, affected people may have spinocerebellar ataxia, peripheral neuropathy, degenerative pigmentary retinopathy, and ceroid myopathy.
7. M. F. Linton, R. V. Farese Jr., S. G. Young, *J. Lipid Res.* **34**, 521 (1993).
8. C. J. Glueck, P. S. Gartside, M. J. Mellies, P. M. Steiner, *Trans. Assoc. Am. Physicians* **90**, 184 (1977).
9. H. Jamil *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 11991 (1996).
10. R. Borensztajn, M. Rone, T. Kotlar, *Biochem. J.* **156**, 539 (1976).
11. R. M. Lawrence, S. A. Biller, O. M. Fryszman, M. A. Poss, *Synthesis* **1997**, 553 (1997).
12. Plasma chemistries for treatments with vehicle and 1, 3, and 6 mg/kg doses of compound **9** were as follows: ALT (U/liter), 73 ± 3 (mean ± SEM),65 ± 5, 75 ± 5, and 71 ± 2, respectively; AST (U/liter), 58 ± 2, 45 ± 3, 58 ± 2, and 59 ± 3; alkaline phosphatase (U/liter), 131 ± 11, 137 ± 1, 135 ± 1, and 131 ± 3; lactic dehydrogenase (U/liter), 156 ± 5, 136 ± 3, 150 ± 3, and 139 ± 5; amylase (U/liter), 1400 ± 300, 1600 ± 100, 1500 ± 200, and 1700 ± 500; gamma glutamyltransferase (U/liter), 2.3 ± 0.7, 2.3 ± 0.2, 1.8 ± 0.4, and 2.2 ± 0.5; glucose (mg/dl), 230 ± 7, 235 ± 8, 194 ± 22, and 199 ± 15; creatinine (mg/dl), 0.4 ± 0, 0.4 ± 0, 0.4 ± 0, and 0.4 ± 0; blood urea nitrogen (mg/dl), 14 ± 1, 14 ± 1, 15 ± 1, and 18 ± 2; total protein (g/dl), 5.4 ± 0.1, 5.5 ± 0.1, 5.7 ± 0.2, and 5.2 ± 0.5; and albumin (g/dl), 3.7 ± 0.1, 3.6 ± 0.1, 3.6 ± 0.1, and 3.6 ± 0.1.
13. The milligrams of triglyceride measured per gram of wet weight liver for treatments with vehicle and 1, 3, and 6 mg/kg doses of compound **9** were as follows: 9.1 ± 0.6 (mean ± SEM), 11 ± 0.3, 18 ± 1.8, and 24 ± 0.4, respectively, at 1 week; 8 ± 0.4, 11 ± 1.1, 12 ± 0.8, and 19 ± 3.4 at 2 weeks; and 9 ± 0.5, 11 ± 1.5, 23 ± 3.5, and 19 ± 1.0 at 3 weeks.
14. S. R. Cairns and T. J. Peters, *Clin. Sci.* **65**, 645 (1983).
15. In an independent 3-week study, liver weight measured as percent of body weight was 11% higher in hamsters that were treated with a 6 mg/kg dose of compound **9** (4.1 ± 0, mean ± SEM) than in vehicle-treated hamsters (3.7 ± 0.1).
16. T. Kita, M. S. Brown, Y. Watanabe, J. L. Goldstein, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 2268 (1981).
17. The ALT and AST levels of drug-treated rabbits were 54.0 ± 7.4 and 34.6 ± 3.5 U/liter (mean ± SEM) in comparison to vehicle-treated rabbit levels of 48.3 ± 5.2 and 28.8 ± 12.5 U/liter.

18 June 1998; accepted 23 September 1998

Genome Sequence of an Obligate Intracellular Pathogen of Humans: *Chlamydia trachomatis*

Richard S. Stephens,* Sue Kalman, Claudia Lammel, Jun Fan, Rekha Marathe, L. Aravind, Wayne Mitchell, Lynn Olinger, Roman L. Tatusov, Qixun Zhao, Eugene V. Koonin, Ronald W. Davis

Analysis of the 1,042,519–base pair *Chlamydia trachomatis* genome revealed unexpected features related to the complex biology of chlamydiae. Although chlamydiae lack many biosynthetic capabilities, they retain functions for performing key steps and interconversions of metabolites obtained from their mammalian host cells. Numerous potential virulence-associated proteins also were characterized. Several eukaryotic chromatin-associated domain proteins were identified, suggesting a eukaryotic-like mechanism for chlamydial nucleoid condensation and decondensation. The phylogenetic mosaic of chlamydial genes, including a large number of genes with phylogenetic origins from eukaryotes, implies a complex evolution for adaptation to obligate intracellular parasitism.

Chlamydia are bacterial pathogens whose representatives are widely distributed in nature, and *C. trachomatis* causes several human diseases of medical significance. Ocular infection leads to trachoma, a leading cause of preventable blindness. Of all infectious diseases reported to U.S. state health departments and the U.S. Centers for Disease Control and Prevention, chlamydial genital tract infections are the most common, and infection of the genital tract often results in pelvic inflammatory disease, ectopic pregnancy, chronic pelvic pain, epididymitis, and infant pneumoniae (*1*). *Chlamydia trachomatis* genital tract infections may also significantly increase the risk for HIV infection (*2*).

Chlamydiae are deeply separated from other eubacteria and represent one of the kingdom-level branches of the phylogenetic tree (*3*). After invasion of eukaryotic cells, chlamydiae grow within an intracellular vacuole, called an inclusion, that does not fuse with lysosomes. Microbiologically, *Chlamydia* are characterized by a developmental cycle involving a metabolically inactive infectious developmental form called the elementary body (EB) that, after entry into the target host cell, differentiates into a metabolically active developmental form called the reticulate body (RB). At approximately 20 hours postinfection and after multiple divisions by binary fission, the RB differentiates into the EB developmental stage and infectious EBs are released to initiate new rounds of infection. Chlamydial physiology, structure, developmental biology, and genetics are poorly understood. The limited and obligate intracellular growth of chlamydiae and the lack of any direct or indirect genetic methods for their study has restricted the development of biological and molecular understanding of these unusual organisms (*4*).

The sequenced chlamydial genome consists of a 1,042,519–base pair chromosome (58.7% A+T) and a 7493–base pair plasmid (sequence and annotation available at http://chlamydia-www.berkeley.edu:4231 and GenBank under accession number AE001273). Analysis of the chlamydial genome resulted in the identification of 894 likely protein-coding genes (*5*). Similarity searching permitted the inferred functional assignment of 604 (68%) encoded proteins, and 35 (4%) were similar to hypothetical proteins deposited for other bacteria. The remaining 255 (28%) predicted proteins were not similar to other sequences deposited in GenBank. Clustering by sequence similarity showed that 256 chlamydial proteins (29%) belong to 58 families of similar genes within the genome (paralogs), a fraction similar to other bacteria with relatively small genomes such as the mycoplasmas and *Haemophilus influenzae* (*6*). A list of the results of analysis of the predicted chlamydial proteins classified in accord with the functional systems in this bacterium and a linear map of genes are available (*5*). The most prominent findings are presented below.

Counterparts of enzymes characterized in other bacteria (orthologs) were identified in *C. trachomatis* to account for the minimal requirements for DNA replication, repair, transcription, and translation. DNA repair and recombination systems were extensively represented in the chlamydial genome, indicating that chlamydiae have considerable recombination capabilities. There are also two predicted DNA he-

R. S. Stephens, C. Lammel, W. Mitchell are with the Program in Infectious Diseases, University of California, Berkeley, CA 94720, USA, and the Francis I. Proctor Foundation, University of California, San Francisco, CA 94143, USA. S. Kalman, J. Fan, R. Marathe, R. W. Davis are at the DNA Sequencing and Technology Center, Stanford University, Stanford, CA 94305, USA. L. Aravind, R. L. Tatusov, E. V. Koonin are at the National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA. L. Olinger and Q. Zhao are with the Francis I. Proctor Foundation, University of California, San Francisco, CA 94143, USA.

*To whom correspondence should be addressed. E-mail: ctgenome@socrates.berkeley.edu

licases of the Swi2/Snf2 family. The conservation of at least one member of this family in all bacteria indicates the importance of these proteins in global nucleoid structure, transcription regulation, or DNA repair. Overall, the chlamydial repair systems appear to be more complete than those identified for the minimal mycoplasmal genomes, although the absence of oxidative damage repair proteins (orthologs of MutT or MutM) that have been detected in all other bacteria is notable. MutH, involved with Dam-directed mismatch repair, was not found. An unusual feature of the chlamydial transcription apparatus is the large size of the elongation factor GreA (CT636). Chlamydial GreA contains a large (~450 amino acids) uncharacterized COOH-terminal domain that has previously only been detected in GreA proteins of spirochetes.

Aminoacyl–transfer RNA (tRNA) synthetase genes, including the unique fusion of the two subunit genes of glycyl-tRNA synthetase (*7*), were represented except those for asparagine and glutamine. The $tRNA^{Asn}$ and $tRNA^{Gln}$ are presumably derived by transamidation of aspartate- and glutamate-charged tRNA as shown for some other eubacteria and Archaea (*8*). The operon encoding the three subunits of glutamyl-tRNA amidotransferase (GatABC-CT002-CT004), which is conserved in all bacteria whose genomes have been sequenced except for *Escherichia coli* and *H. influenzae*, was identified in *Chlamydia* (*9*). Other genes for translation machinery components include two identical ribosomal RNA operons, a complete set of ribosomal proteins, genes for a typical bacterial complement of translation factors, and RNA modification enzymes, including five pseudouridylate synthetases and nine RNA methylases.

The view that emerges for central metabolic pathways of *C. trachomatis* portrays an aerobic organism with glutamate as the likely primary carbon source supplemented by glucose and 2-oxoglutarate, each likely playing different roles depending on the chlamydial developmental stage. Chlamydiae have an apparently intact glycolytic pathway, except that a gene for fructose-1,6-diphosphate aldolase was not identified. In principle, this could be circumvented via the pentose-phosphate pathway and the hexose monophosphate shunt, which appeared to be intact; however, an aldolase was predicted (CT215) that is highly conserved in other bacteria and Archaea and may function in glycolysis. Genes for the Entner-Doudoroff pathway enzymes were not found. The tricarboxylic acid (TCA) cycle was incomplete, because genes for citrate synthase, aconitase, and isocitrate dehydrogenase were not identified; nevertheless, a cyclic pathway still may exist via the aspartate shunt. The TCA cycle could be fed by uptake of exogenous glutamate and transamidation (AspC/TyrB), as there was no glutamate dehydrogenase, or by uptake of 2-oxoglutarate. These possibilities are supported by the finding of genes encoding glutamate and dicarboxylate transporters. Genes encoding glyoxylate-bypass enzymes were not present, suggesting that chlamydiae cannot use acetate or fatty acids as carbon sources. Consistent with the observed production and massive accumulation of glycogen, *C. trachomatis* has a complete glycogen synthesis and degradation system supporting a central role for glucose or glucose derivatives as a primary carbon source for some developmental stage in chlamydial biology. Degradation of glycogen may play a role to initiate growth early in the developmental cycle or to supply energy for differentiation of RB to EB following disassociation from the inclusion membrane.

Genes encoding essential functions in aerobic respiration were well represented in the chlamydial genome. Electron flux to ubiquinone could be supported by pyruvate, succinate, glycerol-3-phosphate, and NADH (reduced nicotinamide adenine dinucleotide) dehydrogenase. The chlamydial genes encoding NADH-ubiquinone oxidoreductase subunits are orthologs of the Na^+-translocating family found in moderately halophilic bacteria, Gram-negative marine bacteria, and the human pathogen *H. influenzae* (*10*). The oxidoreductase complex may couple Na^+ transport with energy production. The pathway for porphyrin biosynthesis was intact, supporting a central role for oxidative catalysis and an electron transport system. The presence of genes encoding CydA and CydB suggests the capacity to transfer H^+ to reduce oxygen, which may serve to energize the membrane and scavenge oxygen. Protection from toxic oxygen intermediates from aerobic respiration is likely mediated by superoxide dismutase and thioredoxin peroxidase. These enzymes also may be involved in protection of chlamydiae from cellular oxidants generated within the host cell.

Chlamydiae are thought to be "energy parasites" because they import adenosine triphosphate (ATP) from their host cell (*11*). The presence of two genes encoding proteins orthologous to rickettsial and organellar adenosine diphosphate (ADP)/ATP translocases supports this hypothesis; however, the genome sequence analysis supports some, perhaps limited, capacity for substrate-level phosphorylation by phosphoglycerate kinase, pyruvate kinase, and succinate thiokinase. Another pathway of ATP biosynthesis may be via the redox chain discussed above. Chlamydiae contain the V (vacuolar) adenosine triphosphatase (ATPase) operon typical of Archaea and also found in scattered bacterial groups such as spirochetes, *Thermus* spp., and *Enterococcus hirae* (*12*). Whether the chlamydial V-type archaean-like ATPase serves only to energize the chlamydial membrane or can produce ATP remains to be experimentally determined. Like the NADH-ubiquinone oxidoreductase, the chlamydial ATPase also may be a Na^+ pump by analogy to the enterococcal ATPase. In addition to the V-ATPase operon, the chlamydial genome contains genes for two flagellar-type ATPases (CT669 and CT717), which are likely to be involved in proton-coupled transport. The presence of all these genes potentially involved in ATP synthesis suggests that chlamydiae are not strict ATP auxotrophs, at least for part of their developmental cycle.

The genome revealed few genes involved in amino acid biosynthesis. Genes encoding amino acid pathway proteins are probably used for purposes other than de novo amino acid biosynthesis, because the pathways are substantially incomplete. Although two aminotransferases (AspC and TyrB) and serine hydroxymethyltransferase (GlyA) may serve to provide terminal conversions for aspartate, phenylalanine, tyrosine, and glycine, no biosynthetic capacity was found to supply the immediate precursor substrates. The more complete pathways, such as the aromatic pathway, may be used for the synthesis of the branch-point compound chorismate and for tRNA modification, and aspartate family pathway enzymes could support the synthesis of diaminopimelate necessary for peptidoglycan biosynthesis. A tryptophan biosynthesis operon (*trpA, trpB*, *trpR*) and *trpC* were present, although no homologs were identified for *trpD* or *trpE*. The presence of a gene encoding the tryptophan repressor suggests that chlamydiae respond to changes in availability of tryptophan. Chlamydial strains differ in their requirements for tryptophan as exemplified by different sensitivities to interferon-γ (IFN-γ)–mediated growth inhibition (*13*). IFN-γ induces the tryptophan-catabolizing enzyme indoleamine 2,3-dioxygenase that limits host cell intracellular pools of tryptophan. Although it is not expected that the tryptophan degradation pathway would produce large amounts of precursor substrates such as anthranilate, the ability of these strains to grow despite cellular degradation of tryptophan and the presence of these tryptophan genes in the genome is conspicuous. It is possible that *C. trachomatis* produces functional equivalents of TrpD or TrpE that are not recognized on the basis of primary sequence conservation. Nevertheless, this likely represents an adaptive virulence determinant for intracellular survival, persistence, and transmission of *C. trachomatis*.

The extensive number of genes for enzymes of fatty acid and phospholipid biosynthesis suggests that chlamydiae synthesize fatty acids, phosphatidylethanolamine, and phosphatidylglycerol de novo, whereas cardiolipin phospholipid synthesis is uncertain (*14*). No gene homolog was found that encodes β-hydroxydeconoyl–acyl carrier protein dehydrase (*fabA*), an essential enzyme for the production of unsaturated fatty acids,

nor was a homolog found for *cfa*, whose product modifies unsaturated fatty acids to produce cyclopropane fatty acids. Cofactor biosynthesis pathways such as those for ubiquinone, folate, and biotin appeared largely incomplete, and chlamydiae probably depend on host-derived intermediates.

How chlamydiae obtain required nutrients from within the intracellular vacuolar inclusion that is apparently not permeable to small compounds (*15*) is unknown. The genome sequence revealed numerous orthologs involved in eubacterial membrane transport systems, including 13 ABC transporters primarily associated with amino acid and oligopeptide transport and Na^+ and H^+ amino acid symporters. In addition, permeases for magnesium, phosphate, nitrate, and sulfate were identified. *Chlamydia* contains a highly conserved cation transport operon orthologous to the *Treponema pallidum troABCD* operon (*12*). The relative preponderance of amino acid and peptide transporters is consistent with the limited amino acid biosynthetic capability of chlamydiae. Another transport operon identified in the chlamydial genome encodes TolB, TolQ, and ExbD-like proteins. In proteobacteria, these proteins are involved in energy coupling of active membrane transport through TonB (*16*); however, no TonB ortholog was identified, suggesting that chlamydiae perform these transport functions distinct from proteobacteria. The presence of porins and membrane transport components indicate that chlamydiae use outer and inner membrane transport systems for most of their biochemical requirements.

During infection and for developmental differentiation, chlamydiae have the capacity to sense cues in their intracellular environment and respond by transcriptional activation of specific gene families. Such global regulation is likely involved in triggering differentiation from EB to RB and RB to EB. The finding of two alternative σ factors, σ^{28} and σ^{54}, suggests one or both are implicated in this process. The σ^{28} is responsible for initiation of transcription of flagellar biosynthesis genes in a variety of organisms and sporulation genes in *Streptomyces coelicolor*. The σ^{54} initiates transcription of genes involved in nitrogen-related pathways, especially glutamine synthase, in a wide variety of organisms. No homolog for a glutamine synthase gene was identified in the chlamydial genome. The σ^{54} initiates transcription by interaction with an activator ATPase (*E. coli* AtoC ortholog) that is regulated by its phosphorylation state mediated by the AtoS-like histidine kinase encoded in the same operon. Because this represents the only two-component signal transduction system identified in the chlamydial genome, it likely plays a key role in developmental stage–specific processes. In addition to AtoS, RbsW is the only other histidine kinase identified in *Chlamydia*. *Chlamydia* encodes the σ-factor regulatory system that consists of the RsbW-like single-domain histidine kinase, two RsbV orthologs, and a RsbU-like protein phosphatase. The presence of this system suggests that the regulation of the alternative σ factors occurs in a pathway similar to that in *Bacillus subtilis* and may have a role in initiating stages of the chlamydial developmental cycle or monitoring chlamydial ATP status and heat-shock response as shown for *B. subtilis* (*17*).

Atypical for most prokaryotes, *Chlamydia* encodes three serine/threonine protein kinases and a PP2C-type protein phosphatase (CT259). Homologs for *ptsI* and *ptsH* encoding enzyme I phosphotransferase and Hpr phosphocarrier protein of the phosphoenolpyruvate:carbohydrate phosphotransferase system (PTS) were identified. *Chlamydia* encodes two PTS IIA proteins (PtsN orthologs) that, in *E. coli*, are involved in linking the regulation of carbohydrate and nitrogen metabolism (*18*). A novel finding was that one of the chlamydial PtsN proteins (CT290) contains an additional helix-turn-helix DNA-binding domain, suggesting it functions to transduce a signal to directly regulate transcription. The remarkable absence of PTS genes for carbohydrate binding and transport, also recently noted for *T. pallidum* (*12*), does not support a role for PtsN in a transport system, but alternatively suggests that it is involved in sensing, possibly at the phosphoenolpyruvate level. Two additional candidates for signal transduction functions are a cyclic AMP–binding protein (CT235) and a forkhead homology–associated domain–containing protein (CT664) similar to those found in *Mycobacteria*.

It has been proposed that chlamydiae lack

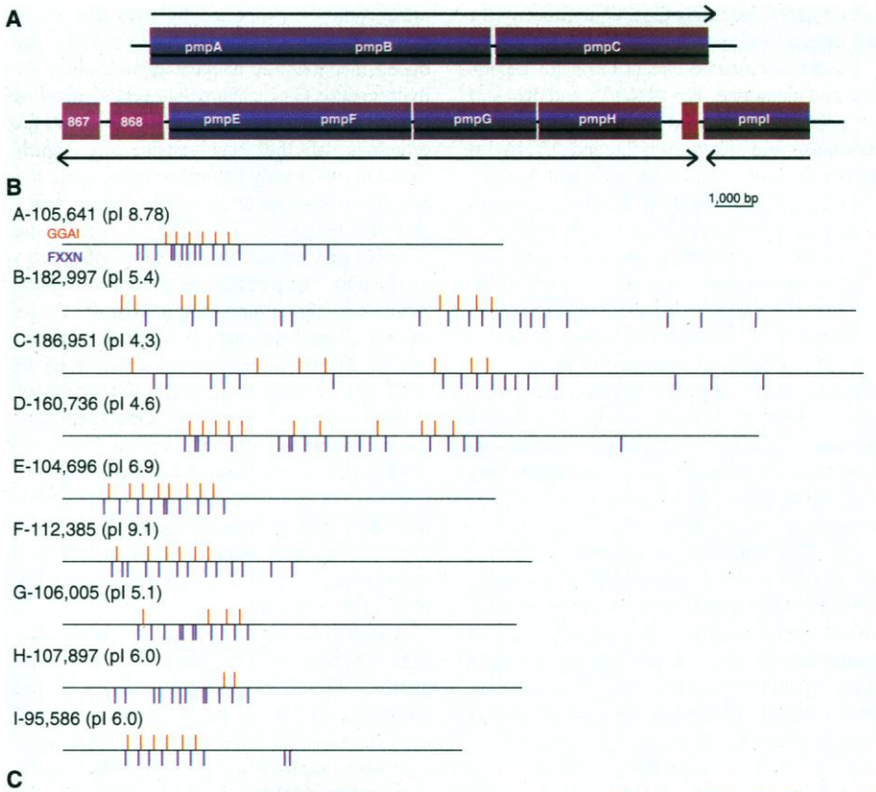

Fig. 1. Gene clusters encoding Pmp outer membrane proteins and proteases implicated in pathogenesis. (**A**) Gene organization of the two clusters of *pmp* genes and the predicted adenovirus-like protease genes (CT867 and CT868). Arrows indicate the coding strand direction and not necessarily operon transcription. (**B**) Predicted molecular mass and pI for each Pmp is shown. Arrangement of tetrameric amino acid repeats in the Pmp outer membrane proteins: GGAI, upper red lines; FXXN, lower blue lines. (**C**) Conserved motifs containing predicted catalytic residues in adenovirus-like proteases (*38*). The consensus, also shown by highlighting, includes amino acid residues conserved in at least 85% of the aligned sequences. Numbers indicate the distance from the protein NH_2-terminus. GenBank gene identification numbers are indicated for each comparative sequence.

peptidoglycan because muramic acid has not been biochemically detected or is detected in relatively small amounts (*19*). Unexpectedly, genes encoding proteins for the entire pathway for peptidoglycan synthesis and membrane assembly and recycling were present in the chlamydial genome. A gene encoding D-Ala:D-Ala ligase (*ddlA*) was found as a 3′ gene fusion with *murC* that encodes L-Ala adding enzyme; however, glutamine and alanine racemases were not identified. Chlamydiae are susceptible to D-cycloserine (*20*), which presumably acts to inhibit an unidentified D-Ala racemase or D-Ala incorporation into peptidoglycan. The latter possibility appears most likely; thus, the amount of D–amino acids available could represent a rate-limiting step in peptidoglycan biosynthesis. The complete composition of this biosynthetic pathway strongly supports an alternative hypothesis that chlamydiae synthesize peptidoglycan or a peptidoglycan-like component, albeit in small amounts and whose function may be different from that of other bacteria.

Relatively little is known regarding the chlamydial outer membrane other than the presence of lipopolysaccharide, the major outer membrane protein (MOMP), and envelope-associated cysteine-rich proteins. The chlamydial MOMP and cysteine-rich proteins are highly disulfide cross-linked in the EB form of the organism and are thought to confer resistance to osmotic stress (*21*). Five genes for disulfide bond isomerases were identified that are probably essential in the outer membrane organization and reorganization of EB and RB cysteine-rich proteins. A notable finding was genes possibly encoding three new outer membrane proteins. One was homologous to chlamydial MOMP (CT713) and is thus implicated as an additional porin; two others were homologous to outer membrane proteins from *Neisseria* sp. (CT241) and *Synechocystis* sp. (CT415).

A family of paralogous genes was identified that are related to recently identified genes encoding a 98-kD protein in *C. psittaci* (*22*). A total of nine genes (named *pmp*) encoding related proteins was identified. The gene family was located in two clusters with one gene being separate (Fig. 1A). For each of the predicted proteins, the COOH-terminal residue was phenylalanine, and some of this family contained predicted cleavable signal peptide leader sequences (PmpC, PmpD, PmpE, and PmpI). These attributes suggest they are outer membrane proteins. Each was a different size and was quite dissimilar in sequence (9 to 42% amino acid identity), but all were found to be related and shared tetrapeptide repeat motifs organized in the NH_2-terminal half of the protein (Fig. 1B). The fact that chlamydiae with such a small genome would maintain these paralogs suggests an important role in structural, functional, or antigenic polymorphism. The *pmpE* and *pmpF* genes are followed by a tandem pair of paralogous genes encoding predicted thiol proteases related to a structurally characterized adenovirus protease (Fig. 1C) (*23*). Both proteases are predicted to be integral membrane proteins.

A complete set of genes orthologous to type III secretion systems was found that is an essential virulence determinant for a variety of Gram-negative animal and plant pathogens (*24*). Genes for this system are typically linked in "pathogenicity islands" containing an A+T-rich signature and have been distributed among enterobacteriaceae by horizontal transfer (*24*). For chlamydiae, these orthologs were found in three genomic locations and lacked an identifiable A+T-rich signature within an A+T-rich background. One gene cluster is represented by homologs to *Yersinia* spp. genes *yscJ*, *yscL*, *yscR*, *yscS*, and *yscT*. A second gene cluster includes six adjacent genes encoding proteins on the same strand. This second cluster contains orthologs to *yscN*, *fliN*, and *yscC* that flank two orphan genes and one gene encoding a serine/threonine protein kinase (CT673). Adjacent to this cluster were orthologs to general secretion proteins GspDEF and LcrH. A third gene cluster encodes orthologs to secretion proteins YscU and SycE and the low-calcium response proteins LcrD and LcrE recently identified in *C. psittaci* (*25*). The probable function of type III secretion in chlamydiae is to modify host cell processes that may be necessary for host cell invasion, remodeling of the inclusion membrane, or affecting host cell regulatory pathways. The apparent linkage to a gene encoding a protein kinase suggests that chlamydiae may deliver a protein kinase to modify functions of the target host cell.

To date, only one chlamydial protein has been reported that is exclusively associated with the host cell vacuolar membrane. *Chlamydia psittaci* produces a protein, IncA, that is present in the vacuolar inclusion membrane and is phosphorylated (*26*). A *C. trachomatis* homolog to IncA (CT119) was identified in addition to two *C. psittaci* genes whose protein products have recently been associated with the

Table 1. Horizontally transferred genes in *Chlamydia trachomatis*.

Chlamydial gene	Support for horizontal transfer
Topoisomerase I, COOH-terminal domain (CT643), and stand-alone SWIB protein (CT460)	The α-helical SWIB domain detected only in eukaryotic chromatin-associated proteins.
SET domain protein (CT737)	SET domain found only in numerous eukaryotic chromatin-associated proteins.
Methionyl-tRNA synthetase (Met-RS) (CT032)	Chlamydial Met-RS groups with eukaryotic ones in phylogenetic analysis. A C2C2 "little finger" domain is conserved in *Chlamydia*, Archaea, and eukaryotes, to the exclusion of the other bacteria. An accessory, COOH-terminal RNA-binding domain is present in bacterial Met-RS, but not in *Chlamydia*, Archaea, or eukaryotes.
Pyrophosphate-dependent phosphofructokinase 1 and 2 (CT205, CT207)	Much greater similarity to eukaryotic than to bacterial orthologs, with the exception of the spirochetes.
Glucose-1-phosphate adenylyltransferase (CT489)	Supported by phylogenetic analysis (*39*); likely chloroplast origin in plants.
Glycogen phosphorylase (CT248)	Supported by phylogenetic analysis (*39*).
UDP-glucose pyrophosphorylase (CT715)	No easily detectable bacterial homologs.
Adenine nucleotide translocases 1 and 2 (CT495, CT065)	ATP/ADP translocases of this family are detectable only in *Rickettsia* and in plants.
Glycerol-3-phosphate acyltransferase (CT807)	So far detected only in plants.
β-ketoacyl-ACP synthase (CT770)	Supported by phylogenetic analysis (*39*).
Enoyl-ACP reductase (CT104)	Much greater similarity to eukaryotic than to bacterial orthologs; see phylogenetic analysis (Fig. 3).
Biotin-protein ligase (CT035)	So far detectable only in two yeast species.
V-type ATPase operon (CT304-310)	Grouping with spirochetes to the exclusion of Archaea is apparent.
Adenovirus type thiol protease (CT867, CT868)	Poorly conserved but common in eukaryotes; among prokaryotes, detected only in *E. coli*.
Thioredoxin peroxidase (CT603)	Supported by phylogenetic analysis (*39*).
Superoxide dismutase (CT294)	Supported by phylogenetic analysis (*39*).
Putative metal-dependent hydrolase (CT386)	So far detected only in yeast and *C. elegans*.
Uncharacterized, conserved protein (CT472)	So far detected only in yeast and *C. elegans*.
Tryptophan operon (CT169-171)	Tryptophan repressor has been found only in Proteobacteria, suggesting that the whole operon has been acquired by horizontal transfer.
Predicted ATPase (CT402)	So far found only in Proteobacteria.

inclusion membrane. These two proteins in *C. trachomatis* appear to form an operon along with a third, uncharacterized protein that contains predicted transmembrane helices (CT232, CT233, and CT234). This suggests that the proteins may function in inclusion membrane remodeling and transport.

A provocative finding is the identification of a family of six paralogous chlamydial proteins belonging to a superfamily of enzymes (named HKD superfamily after the conserved catalytic triad and present in each chlamydial paralog) that includes phospholipase-D and phospholipid synthases as well as some nucleases (*27*). Four of these genes (*CT154, CT155, CT157, and CT158* form a cluster, whereas the other two genes (*CT084* and *CT284*) are found separately. The chlamydial HKD proteins show relatively weak similarity to other members of the superfamily (Fig. 2), precluding a confident prediction of the specificity of these enzymes. The products of the clustered genes may be functionally distinct from the stand-alone gene products, especially as the latter (CT084 and CT284) have confidently predicted signal peptides and may be secreted. Elucidation of the function of the HKD enzymes may shed light on chlamydial-directed modification of host cell phospholipids (*14, 28*).

The genome sequence confirms investigations demonstrating that chlamydiae lack the ability to synthesize purine and pyrimidine nucleotides de novo except for synthesis of cytosine triphosphate (*29*). Genes for enzymes required for the synthesis and interconversion of each of the deoxyribonucleotides were present, with the exception of the thymidylate synthase gene, supports evidence that deoxythymidine monophosphate is required from the host cell (*29*). Systematic examination of the list of highly conserved bacterial families (COGs) resulted in the identification of representatives of 30 COGs that have been previously found in all completely sequenced bacterial genomes and are missing in *C. trachomatis* (*30*). The most conspicuous observations were the absence of FtsZ, a guanosine triphosphatase required for septum formation in all bacteria and Archaea; *S*-adenosylmethionine (SAM) synthetase (MetK), which is encoded in all bacterial and eukaryotic genomes; and phosphorybosylpyrophosphate (PRPP) synthase (PrsA), an apparently ubiquitous central enzyme of nucleotide biosynthesis. Also absent were hexose-biphosphate aldolase, adenine and guanine phosphoribosyltransferases (Apt/Gpt), sugar kinases (such as FruK/PfkB), peptide methionine sulfoxide reductase (MsrA), and guanosine pyrophosphohydrolase and synthase (SpoT/RelA). The absence of FtsZ suggests that the cell division mechanisms in chlamydiae may show major mechanistic differences from other prokaryotes. The absence of such key metabolic enzymes as MetK and PrsA may be explained by the availability of relatively unstable metabolic intermediates as SAM and PRPP in an intracellular but not extracellular environment.

Homologs were not identified for proteins involved with transformation and competence to acquire exogenous DNA, nor were phage- or transposon-like homologs identified. Likewise, no genes were found that encode DNA restriction endonuclease or modification systems. These data are consistent with the concept that the intracellular and intravacuolar niche in which modern chlamydiae grow has isolated these organisms from frequent genetic exchange with other organisms (*31*).

The *C. trachomatis* genome sequence contains striking features that support its atypical phylogenetic classification, suggesting a role for horizontal gene transfer both from bacterial ancestors as well as from eukaryotic hosts to chlamydiae. Phylogenetic analysis (*32*) of a number of housekeeping genes such as DNA and RNA polymerase subunits, translation factors, and others supports the position of *Chlamydia* as a distinct bacterial lineage, without clear affinities with other major taxa. Examination of the taxonomic distribution of homologs (*33*), as well as more detailed phylogenetic tree analysis, suggests horizontal gene transfer as the origin of numerous chlamydial genes (Table 1). Acquisition from other bacterial ancestors is exemplified by the NADH-ubiquinone oxidoreductase and type III secretion operons. In both cases, and apparently unlike in other bacteria, the operon organization has been only partially retained and the genes have significantly diverged, suggesting ancient acquisition. Such acquisitions may have occurred before *Chlamydia*-like organisms became host-cell–dependent because they were likely required for effective intracellular interaction with eukaryotic cells. Perhaps directly related to their natural history as intracellular parasites, gene origins from eukaryotes appear to be more common in chlamydiae than in other sequenced genomes of bacterial pathogens, with 35 protein coding sequences phylogenetically associated with eukaryotic homologs (*33*).

Some examples of acquired eukaryotic genes are convincingly indicated by phylogenetic analysis. Horizontally transferred genes are involved even in housekeeping functions of chlamydiae, as illustrated by the examples of methionyl-tRNA synthetase (somewhat weaker evidence was seen for isoleucyl-tRNA synthetase), phosphofructokinases, and a lipid metabolism enzyme (Table 1 and Fig. 3). Among the chlamydial proteins of apparent eukaryotic origin, particularly notable are two that consist almost entirely of domains previously identified in multidomain, eukary-

							*		*				*																			
PLD1.2_Hs_2627323	459	Y	L	W	A	H	H	E	K	L	V	I	I	D	Q	S	V	A	F	V	G	G	I	D	L	A	Y	G	R	486
PLD1.1_Hs_2627323	891	E	L	I	Y	V	H	S	K	L	L	A	I	D	D	N	T	V	I	I	G	S	A	N	I	N	D	R	S	918
Y190_Hp_2313277	162	I	K	K	R	M	H	N	K	L	F	I	V	D	N	F	A	V	I	I	G	G	R	N	I	G	D	N	Y	189
VK04.1_VACCV_335358	108	Y	S	R	V	N	H	A	K	Y	M	V	T	D	-	K	T	A	Y	I	G	T	S	N	W	T	G	N	Y	137
VK04.2_VACCV_335358	320	L	G	G	V	L	H	T	K	F	W	I	S	D	N	T	H	I	Y	L	G	S	A	N	M	D	W	R	S	346
PSS.1_Ec_147389	134	A	L	G	V	L	H	F	K	G	F	I	I	D	D	S	V	L	Y	S	G	A	S	L	N	D	V	Y	L	170
PSS.2_Ec_147389	353	D	D	N	T	Y	H	L	K	G	M	W	V	D	D	K	W	M	L	I	T	G	N	N	L	N	P	R	A	380
CT154.1_Ct	136	G	R	T	L	Q	H	K	K	T	M	L	A	D	F	Q	T	V	V	T	G	S	A	N	Y	T	D	L	S	163
CT154.2_Ct	282	T	E	G	V	L	H	T	K	I	C	C	I	D	N	K	T	L	I	F	G	S	A	N	W	S	G	A	G	309
CT155.1_Ct	119	G	C	S	L	L	H	R	K	T	L	L	I	D	N	N	I	V	V	T	G	T	A	N	Y	T	E	A	S	146
CT155.2_Ct	211	C	E	G	I	L	H	T	K	V	C	C	I	D	S	S	T	L	I	I	G	S	V	N	W	S	R	G	G	238
CT157.1_Ct	195	P	Q	I	C	L	H	K	K	T	T	L	I	D	N	Q	L	T	I	I	G	T	A	N	Y	T	K	S	S	222
CT157.2_Ct	341	P	Q	I	C	L	H	K	K	T	T	L	I	D	N	Q	L	T	I	I	G	T	A	N	Y	T	K	S	S	368
CT158.1_Ct	10	N	Q	A	R	L	H	C	K	N	I	V	V	D	G	S	L	V	I	T	G	S	A	N	F	S	D	A	A	37
CT158.2_Ct	157	E	N	A	L	L	H	C	K	V	G	L	I	D	T	N	L	L	I	T	G	S	A	N	W	T	V	R	G	184
CT084.1_Ct	126	H	R	K	L	M	H	Q	K	T	M	A	I	D	G	E	L	A	W	I	G	S	A	N	F	T	L	A	S	153
CT084.2_Ct	272	T	P	Y	Q	L	H	H	K	F	G	I	F	D	K	K	T	L	I	T	G	S	V	N	W	S	E	N	G	299
CT284.1_Ct	137	N	V	T	E	S	H	T	K	L	S	I	V	D	G	K	Y	I	F	I	G	G	S	N	L	E	D	L	Q	164
CT284.2_Ct	386	A	N	T	Q	L	H	K	K	C	M	L	V	D	D	H	I	L	V	I	G	S	Y	N	F	G	K	K	S	413
Consensus(80%)		.	.	.	.	.	H	.	K	.	.	h	.	D	.	.	.	h	h	.	G	s	.	s	h	s	.	.	.		

Fig. 2. Unique expansion of HKD superfamily phospholipase-D–like enzymes in chlamydiae. Designations as in Fig. 1C.

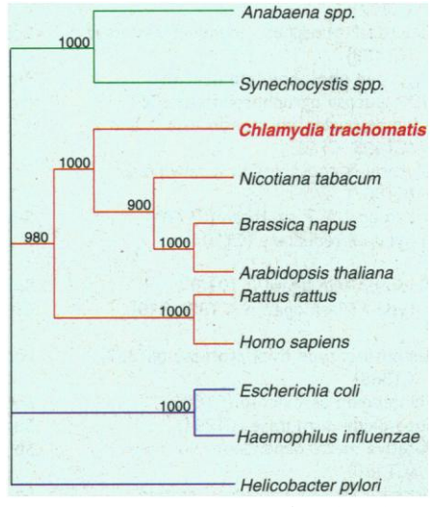

Fig. 3. Phylogeny of chlamydial enoyl–acyl carrier protein reductase as an example of horizontal transfer. Tree was constructed using the KITSCH program (*32*). Numbers at forks indicate the number of bootstrap replications, out of 1000, in which the given node was observed.

otic chromatin-associated factors. These two domains, SET and SWIB, have not been identified outside eukaryotes. Most remarkably, a diverged copy of the SWIB domain was found fused to the COOH-terminus of the chlamydial topoisomerase I. It appears that SWIB and SET domains, together with the Swi/Snf2 family of helicases (CT555 and CT708), may participate in the chromatin condensation-decondensation that is characteristic of the chlamydial developmental cycle (*34*). Moreover, these findings suggest that there are mechanistic similarities between these processes in chlamydiae and eukaryotic chromatin dynamics.

Most of the "eukaryotic" proteins of chlamydia do not appear to be specifically related to their animal homologs. Unexpectedly, they tend to group with plant proteins in phylogenetic analyses (Fig. 3). A clue as to their previous niche may be suggested by the observation of *Chlamydia*-like endosymbionts of *Acanthamoebae* (*35*). The phylogenetic position of *Acanthamoebae* is near the original branching point of the eukaryotes, and grouping with plants cannot be ruled out (*36*). It is possible that the evolution of chlamydiae as intracellular parasites started with an opportunistic interaction with amoebal hosts, and the protochlamydiae became amoebal parasites or symbionts for a period long enough to acquire the "plant-like" genes, whose origin may be amoebal. Thus, the emergence of host-dependent protochlamydiae may have started in single-cell amoebic hosts before moving to multicellular invertebrate hosts, perhaps then diverging into the major chlamydial genera prior to adapting to vertebrate hosts (*37*).

References and Notes

1. *Morb. Mortal. Wkly Rep.* **46**, 193 (1997).
2. M. Laga, N. Nzila, J. Goeman, *AIDS* (suppl. 1) **5**, S55 (1991).
3. N. R. Pace, *Science* **276**, 734 (1997); W. G. Weisburg, T. P. Hatch, C. R. Woese, *J. Bacteriol.* **167**, 570 (1986). *Chlamydia* are classified in the order Chlamydiales, with one family (Chlamydiaceae) that contains a single genus [J. W. Moulder *et al.*, Chlamydia, *Bergey's Manual of Systematic Bacteriology*, N. R. Krieg, Ed. (Williams & Wilkins, Baltimore, MD, 1984), vol. 1, p. 729].
4. R. S. Stephens, *Infect. Agents Dis.* **1**, 279 (1992).
5. Genome sequencing and annotation methodologies are available at the *Science* Web site (www.sciencemag.org/feature/data/982604.shl)
6. R. L. Tatusov, E. V. Koonin, D. J. Lipman, *Science* **278**, 631 (1997); E. V. Koonin *et al.*, *Mol. Microbiol.* **25**, 619 (1997); S. E. Brenner *et al.*, *Nature* **378**, 140 (1995).
7. E. A. Wagar *et al.*, *J. Bacteriol.* **177**, 5179 (1995).
8. M. Ibba, A. W. Curnow, D. Soll, *Trends Biochem. Sci.* **22**, 39 (1997).
9. A. W. Curnow *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 11819 (1997).
10. M. Hayashi, Y. Nakayama, T. Unemoto, *FEBS Lett.* **381**, 174 (1996).
11. J. W. Moulder, *Microbiol. Rev.* **55**, 143 (1991); T. P. Hatch, E. Al-Hossainy, J. A. Silverman, *J. Bacteriol.* **150**, 662 (1982).
12. C. M. Fraser *et al.*, *Science* **281**, 375 (1998); C. M. Fraser *et al.*, *ibid.* **270**, 397 (1995); K. Yokoyama *et al.*, *J. Biol. Chem.* **269**, 12248 (1994); K. Takase *et al.*, *ibid.*, p. 11037.
13. W. L. Beatty *et al.*, *Infect. Immun.* **62**, 3705 (1994).
14. J. L. Wylie, G. M. Hatch, G. McClarty, *J. Bacteriol.* **179**, 7233 (1997).
15. R. A. Heinzen and T. Hackstadt, *Infect. Immun.* **65**, 1088 (1997).
16. T. E. Letain and K. Postle, *Mol. Microbiol.* **24**, 271 (1997).
17. A. Dufour and W. G. Haldenwang, *J. Bacteriol.* **176**, 1813 (1994); U. Voelker, A. Dufour, W. G. Haldenwang, *ibid.* **177**, 114 (1995).
18. B. S. Powell *et al.*, *J. Biol. Chem.* **270**, 4822 (1995).
19. A. Fox *et al.*, *Infect. Immun.* **58**, 835 (1990); H. Su, Y. X. Zhang, R. Li, *Kexue Tongbao* **30**, 695 (1985).
20. J. W. Moulder, *Microbiol. Rev.* **55**, 143 (1991).
21. W. J. Newhall and R. B. Jones, *J. Bacteriol.* **154**, 998 (1983); T. P. Hatch, M. Miceli, J. E. Sublett, *ibid.* **165**, 379 (1986).
22. D. Longbottom *et al.*, *FEMS Microbiol. Lett.* **142**, 277 (1996).
23. A. Webster and G. Kemp, *J. Gen. Virol.* **74**, 1415 (1993).
24. C. J. Hueck, *Microbiol. Mol. Biol. Rev.* **62**, 379 (1998).
25. R. C. Hsia, *et al.*, *Mol. Microbiol.* **25**, 351 (1997).
26. D. D. Rockey *et al.*, *ibid.* **24**, 217 (1997).
27. C. P. Ponting and I. D. Kerr, *Protein Sci.* **5**, 914 (1996); E. V. Koonin, *Trends Biochem. Sci.* **21**, 242 (1996).
28. G. M. Hatch and G. McClarty, *Infect. Immun.* **66**, 3727 (1998); T. Hackstadt, M. A. Scidmore, D. D. Rockey, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 4877 (1995).
29. G. Tipples and G. McClarty, *Mol. Microbiol.* **8**, 1105 (1993); *J. Biol. Chem.* **270**, 7908 (1995); T. P. Hatch, *J. Bacteriol.* **125**, 706 (1976).
30. R. L. Tatusov, E. V. Koonin, D. J. Lipman, *Science* **278**, 631 (1997).
31. E. A. Wagar, S. Safarians, M. Pang, *FEMS Microbiol. Lett.* **77**, 161 (1992).
32. Phylogenetic trees were constructed from multiple alignments of protein sequences using the neighbor-joining method and the Fitch-Margoliash method as implemented in the KITSCH program, with 1000 bootstrap replications [J. Felsenstein, *Methods Enzymol.* **266**, 418 (1996)].
33. Taxonomic distribution of database hits was evaluated using the TAX_COLLECTOR program from the SEALS package [D. R. Walker and E. V. Koonin, *Ismb* **5**, 333 (1997)]. The analysis detected 35 chlamydial proteins, for which the similarity to eukaryotic homologs was significantly greater than that to bacterial homologs (e-values differed by at least a factor of 100). For more than 20 of these, the eukaryotic or archaeal origin was unequivocally supported by phylogenetic tree analysis. In comparable analyses of the genomes of other pathogenic bacteria such as spirochetes or *Helicobacter pylori*, only three or four genes were linked to eukaryotic origins (L. Aravind, unpublished observation).
34. The chlamydial nucleoid structure differs significantly between the RB and EB developmental stages, wherein it is highly condensed in the EB stage [J. W. Costerton, L. Poffenroth, J. C. Wilt, N. Kordova, *Can. J. Bacteriol.* **22**, 16 (1976)]. Nucleoid condensation is associated with EB stage–specific proteins, so-called histone-like proteins [E. A. Wagar and R. S. Stephens, *Infect. Immun.* **56**, 1678 (1988); T. Hackstadt, W. Baehr, Y. Ying, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 3937 (1991)], that modulate superhelical density in EBs [M. V. Solbrig, M. L. Wong, R. S. Stephens, *Mol. Microbiol.* **4**, 1535 (1990)] and when expressed in *E. coli* [C. E. d. Barry, S. F. Hayes, T. Hackstadt, *Science* **256**, 377 (1992)].
35. R. Amann *et al.*, *Appl. Environ. Microbiol.* **63**, 115 (1997).
36. M. L. Sogin, H. J. Elwood, J. H. Gunderson, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 1383 (1986).
37. J. W. Moulder, in *Microbiology of Chlamydiae*, A. L. Barron, Ed. (CRC Press, Boca Raton, FL, 1988), pp. 3–19.
38. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. X indicates any residue.
39. E. V. Koonin, L. Aravind, R. L. Tatusov, data not shown.

40. We thank K. Koshiyama, D. R. Walker, E. Chung, C. Komp, S. Mirthipati, and especially, T. Brettin and F. Dietrich for their contributions to this project. S. Kalman led the DNA sequencing phase of this project. Supported by National Institute of Allergy and Infectious Diseases grant AI 39258.

18 May 1998; accepted 27 August 1998

Attenuation of Virulence by Disruption of the *Mycobacterium tuberculosis erp* Gene

François-Xavier Berthet,* Micheline Lagranderie, Pierre Gounon, Christine Laurent-Winter, Danielle Ensergueix, Pierre Chavarot, Françoise Thouron, Eddie Maranghi, Vladimir Pelicic,† Denis Portnoï, Gilles Marchal, Brigitte Gicquel‡

The virulence of the mycobacteria that cause tuberculosis depends on their ability to multiply in mammalian hosts. Disruption of the bacterial *erp* gene, which encodes the exported repetitive protein, impaired multiplication of *M. tuberculosis* and *M. bovis* Bacille Calmette-Guérin in cultured macrophages and mice. Reintroduction of *erp* into the mutants restored their ability to multiply. These results indicate that *erp* contributes to the virulence of *M. tuberculosis*.

Mycobacterium tuberculosis, *M. bovis*, and *M. africanum* cause tuberculosis, an infectious disease killing more than 3 million people per year worldwide (*1*). These bacteria constitute the *M. tuberculosis* complex (MTC), which also includes the tuberculosis vaccine strain *M. bovis* Bacille Calmette-Guérin (BCG) and the murine pathogen *M. microti*. Members of the MTC multiply within phagocytic cells in a specialized vacuolar compartment called the phagosome (*2*). Phagosomes containing mycobacteria do not acidify (*3*), and they escape fusion with lysosomes (*4*). Remodeling of the phagosome architecture by these pathogenic bacteria is thought to be critical for their intracellular multiplication