

July, 2006

Distinctive Repertoire of Contingency Genes Conferring Mutation- Based Phase Variation and Combinatorial Expression of Surface Lipoproteins in *Mycoplasma capricolum* subsp. *capricolum* of the *Mycoplasma mycoides* Phylogenetic Cluster

Kim S. Wise, *University of Missouri-Columbia*

Mark F. Foecking, *University of Missouri-Columbia*

Kerstin Roske, *University of Missouri-Columbia*

Young Jin Lee, *University of California, Davis*

Young Moo Lee, *University of California, Davis*, et al.

Distinctive Repertoire of Contingency Genes Conferring Mutation-Based Phase Variation and Combinatorial Expression of Surface Lipoproteins in *Mycoplasma capricolum* subsp. *capricolum* of the *Mycoplasma mycoides* Phylogenetic Cluster†

Kim S. Wise,^{1*} Mark F. Foecking,¹ Kerstin Röske,^{1‡} Young Jin Lee,^{3§}
Young Moo Lee,^{3¶} Anup Madan,^{4||} and Michael J. Calcutt²

Department of Molecular Microbiology and Immunology¹ and Department of Veterinary Pathobiology,² University of Missouri-Columbia, Columbia, Missouri 65212; Molecular Structure Facility, University of California Davis, Davis, California 95616³; and Institute for Systems Biology, Seattle, Washington 98103-8904⁴

Received 17 February 2006/Accepted 21 April 2006

The generation of surface variation among many divergent species of *Mollicutes* (mycoplasmas) occurs through stochastic expression patterns of diverse lipoprotein genes. The size and wide distribution of such variable gene sets in minimal (~0.6- to 1.4-Mb) mycoplasmal genomes suggest their key role in the adaptation and survival of these wall-less monoderms. Diversity through variable genes is less clearly established among phylogenetically similar mycoplasmas, such as the *Mycoplasma mycoides* cluster of ruminant pathogens, which vary widely in host range and pathobiology. Using (i) genome sequences from two members of this clade, *Mycoplasma capricolum* subsp. *capricolum* and *M. mycoides* subsp. *mycoides* small colony biotype (SC), (ii) antibodies to specific peptide determinants of predicted *M. capricolum* subsp. *capricolum* gene products, and (iii) analysis of the membrane-associated proteome of *M. capricolum* subsp. *capricolum*, a novel set of six genes (*vmcA* to *vmcF*) expressing distinct Vmc (variable *M. capricolum* subsp. *capricolum*) lipoproteins is demonstrated. These occur at two separate loci in the *M. capricolum* subsp. *capricolum* genome, which shares striking overall similarity and gene synteny with the *M. mycoides* subsp. *mycoides* SC genome. Collectively, Vmc expression is noncoordinate and combinatorial, subject to a single-unit insertion/deletion in a 5' flanking dinucleotide repeat that governs expression of each *vmc* gene. All *vmc* genes share modular regions affecting expression and membrane translocation. In contrast, *vmcA* to *vmcD* genes at one locus express surface proteins with highly structured size-variable repeating domains, whereas *vmcE* to *vmcF* genes express products with short repeats devoid of predicted structure. These genes confer a distinctive, dynamic surface architecture that may represent adaptive differences within this important group of pathogens as well as exploitable diagnostic targets.

Among the monoderms comprising the low-G+C *Firmicutes*, *Mollicutes* (termed mycoplasmas in this report) represent a clade of organisms that displays both marked genome reduction and extensive phylogenetic divergence (21, 52). Many of the >200 known mycoplasmal species (22, 50) are obligate parasites and significant pathogens of vertebrate hosts. Their dependence on host factors for survival stems in part from an absence of genes encoding critical metabolic pathways that are present in many bacteria, a paucity of regulatory networks providing classic gene regulation in response to environmental cues and limitations imposed by the lack of genes for cell wall

synthesis. In addition, and despite some examples of their growth and survival in intracellular environments (1, 5, 26, 55), mycoplasmas generally occur in niches outside host cells. This changing milieu demands an adaptive survival strategy and implies the need in these wall-less organisms for a cell surface capable of dynamic host interactions, both opportunistic as well as defensive. Although the full range of such strategies is not known, the generation of membrane surface diversity within mycoplasma populations is now appreciated as one key factor in this regard (5, 35, 61).

Mycoplasma genomes have been reported to contain from 482 to 1,037 protein-coding genes (18, 43). Despite this limited capacity, the genomes of several mycoplasma species contain distinctive sets of genes that encode diverse surface lipoproteins. Products of these genes are anchored in the single plasma membrane of the mycoplasma and often are among the most abundant membrane proteins. These lipoproteins are commonly subject to high-frequency phase-variable expression, mediated by a variety of mechanisms causing localized reversible mutations that affect transcription or the translational reading frame of individual genes (35, 61). The diversity of the coding sequences comprising these gene “families” is remarkable, whether compared within a single species or

* Corresponding author. Mailing address: University of Missouri-Columbia, Molecular Microbiology and Immunology, M616 Medical Sciences Building, One Hospital Drive, Columbia, MO 65212. Phone: (573) 882-8138. Fax: (573) 882-4287. E-mail: wisek@missouri.edu.

† Supplemental material for this article may be found at <http://j.b.asm.org/>.

‡ Present address: Institute for Medical Microbiology and Hygiene, University of Dresden, D-01307 Dresden, Germany.

§ Present address: Genome Center Proteomics Core, University of California, Davis, CA 95616.

¶ Present address: 4937 Marlborough Way, Carmichael, CA 95608.

|| Present address: 200-B Neurogenomics Research Lab, University of Iowa, Iowa City, IA 52246.

across species from widely divergent clades (5, 35, 61). Nevertheless, unifying features shared among members of a particular family are commonly found, for example, (i) highly conserved coding sequences that present a shared segment of otherwise variable surface proteins, (ii) tandemly repeating coding sequences that confer variable length (and, in some instances, function) to the surface protein, and (iii) a common sequence motif appended to structural genes that mediates mutation-driven phase variation in expression. Overall, these systems provide a means of generating extensive surface diversity in rapidly mutating mycoplasma populations (35) and are prime examples of phase variation among bacteria (48). Importantly, these features are readily apparent in mycoplasmal genome sequences.

Variable gene families are widespread among phylogenetically distant mycoplasma species, including some for which complete genome sequences are now known (4, 32, 43). In contrast, comparison of variable gene families among closely related mycoplasmas within narrowly defined clades has only been partially explored (19). These comparisons may offer key insights into the evolution of variable gene systems within highly constrained phylogenetic frameworks and under influences driving genome reduction. This approach may also identify important features in mycoplasmal surface architecture and adaptability that determine factors such as host range and pathogenicity. In addition, these products are known to be targets of the adaptive immune response and contribute to the distinctive antigenic profile of the organism. Nevertheless, the detection and serological discrimination of closely related mycoplasmal species or strains can be confounded due to overlapping sets of antigens and their variable expression (40). The genomic repertoire, pattern of expression, and mechanism of variation of genes encoding surface proteins are therefore critically important in understanding the consequences of mycoplasmal surface variation.

The "*Mycoplasma mycoides* cluster" (16, 21, 34, 52), comprising a phylogenetic clade of very closely related mycoplasmas, offers particularly attractive opportunities for understanding the dynamics of surface diversification at the genomic level. Subspecies of this group are distinguishable by detailed analysis of 16S rRNA sequence divergence, although their resolution is near the limit of this approach (28, 34). Members of the *M. mycoides* cluster include two severe pathogens that cause reportable diseases listed by the Office International des Epizooties: contagious bovine pleuropneumonia is caused by *Mycoplasma mycoides* subsp. *mycoides* small colony biotype (SC), and contagious caprine pleuropneumonia is caused by *Mycoplasma capricolum* subsp. *capripneumoniae*. Four other members of this clade represent significant pathogens of ruminants, each with a distinctive pathobiology and host range. These are *M. mycoides* subsp. *mycoides* large colony type, *M. mycoides* subsp. *capri*, *Mycoplasma* sp. bovine serogroup 7, and *M. capricolum* subsp. *capricolum*, a caprine pathogen investigated in this report. *M. capricolum* subsp. *capricolum* has also been widely studied as a model to understand the basic molecular biology of mycoplasmas and as a representative of this phylogenetic cluster (9, 23, 31, 37, 44). Notably, subspecies within this cluster show considerable and inconsistent cross-reactivity in assorted serological assays designed to differentiate these organisms (8, 10, 14, 16, 42).

This study was prompted by the completion of the genome sequence of *M. capricolum* subsp. *capricolum* (GenBank accession no. NC_007633; <http://www.ncbi.nlm.nih.gov>), which allowed its comparison to that of *M. mycoides* subsp. *mycoides* SC type strain PG1 (*M. mycoides* subsp. *mycoides* SC PG1), the only other member of the *M. mycoides* cluster for which the whole-genome sequence is currently available (53). With information from the complete *M. capricolum* subsp. *capricolum* sequence determined from the type strain (Kid) and from the partial genomic sequence (described herein) determined for a clonal isolate derived from the same source of the organism, we examined a repertoire of genes encoding a diverse family of variable surface proteins, termed Vmc (variable *Mycoplasma capricolum* subsp. *capricolum*) lipoproteins, that differed from currently reported variable genes in *M. mycoides* subsp. *mycoides* SC PG1 (33). We characterized two sets of vmc-encoded size-variable proteins with highly structured or unstructured features. We further verified that the predicted phase-variable expression of vmc genes was governed by a reversible promoter mutation mechanism analogous to that occurring in the *vmm* gene of *M. mycoides* subsp. *mycoides* SC PG1 but with greater subtlety than previously known. These results document the expanded utilization of a multigenic, phase-variable system of contingency genes in the *M. mycoides* cluster of mycoplasmas and underscore the high degree of genomic variation and plasticity that these systems bring to otherwise conserved genomes of this important group. Our results also suggest applications for diagnostics and detection that may overcome current limitations in these areas.

MATERIALS AND METHODS

Mycoplasma cultures. *Mycoplasma capricolum* subsp. *capricolum* Kid (ATCC 27343) was obtained from the American Type Culture Collection (Manassas, VA) and grown in liquid culture using modified Hayflick medium supplemented with 20% heat-inactivated horse serum as previously described (58). Cultures from the initial passage were plated on 1% Noble agar medium of the same composition, and individual colonies were picked as described previously (6, 39) to generate clonal starting populations. One randomly picked clonal derivative, mck3, was used throughout this study. This isolate was therefore derived from the same propagated culture source that had previously been used to determine the complete genome sequence of *M. capricolum* subsp. *capricolum* (GenBank accession no. NC_007633; <http://www.ncbi.nlm.nih.gov>).

Clonal lineages. Clonal lineages were derived from the mck3 population as described in detail previously (39, 46, 65). Plated populations were screened by immunoblot analyses to score the Vmc expression phenotype, using specific polyclonal antibodies (PABs) or monoclonal antibodies (MAbs) to synthetic peptides. Assurance of phenotypic switching in successive populations included the use of multiple rounds of colony isolation for a phenotype and the inclusion of lineages where sequential switching was selected to proceed through three alternative phenotypes, in addition to reciprocal switches between two phenotypes. For analysis of DNA sequence or protein expression in particular variants, cells were grown from the primary colony to approximately 10 ml, harvested and rinsed by centrifugation, and suspended in buffer for protein analysis as previously described (39) or for extraction of genomic DNA using a DNeasy tissue kit (QIAGEN, Valencia, CA).

Antibodies, immunological assays, and protein analysis. Synthetic peptides derived from the sequences of predicted Vmc lipoproteins were prepared as described previously, with a Cys residue appended at the N terminus for coupling reactions (6, 7). These included Vmc peptide A (pep A) (CEKQAKQQAENKKEELDKAEAEELKTAREN), Vmc pep B (CAAKAKLNEQLKPSDQAKQEELKKAQEAFT), Vmc pep C (CNEIDAAQNEVERAEAEELDKAKENLQKLKK), Vmc pep D (CAVKEAEKKLELDKKAQENPKKEQLLEAQ), Vmc pep E (CGSGTVKPAEGGSGTVKPAEGGSGTVKPAR), and Vmc pep F (CSKPGTGSETAPSKPGTGSETAPSKPGK). For immunization, peptides were coupled to maleimide-activated keyhole limpet hemocyanin and inoculated into BALB/c mice as described previously in detail (6). Preimmune and postnuc-

lation sera were collected and used at a 1:200 dilution, except where indicated. All procedures using animals were performed under an animal use protocol. Resulting PABs used in all immunological assays were compared to control preimmune sera, which were uniformly negative. For immunological inhibition assays, peptides were added to primary Abs at a final concentration of 10 μ M.

The enzyme-linked immunosorbent assay (ELISA) has been described in detail elsewhere (6). Briefly, peptides were conjugated with *N*-[6-(biotinamido)hexyl]-3'-(2'-pyridyldithio)propionamide (biotin-HPDP), immobilized on microtiter plates that were precoated with neutral avidin, and then incubated successively with PAB, secondary horseradish peroxidase (HP)-conjugated Ab, and the chromogenic substrate 2,2'-azino-bis(3-ethylbenz-thiazoline-6-sulfonic acid) to quantify binding that was monitored at an optical density at 410 nm (OD₄₁₀). To measure the relative binding of each PAB to pep A through pep F, saturating amounts (1:1,000 dilutions) of each PAB were incubated with a panel representing all six immobilized peptides (each PAB yielded an OD of ~1.0 with its cognate peptide). Binding to pep A through pep F was determined for each PAB as the percent relative normalized binding, calculated as the ratio of background-corrected OD values obtained with immobilized cognate peptide (normalized to 100%) versus that obtained with the other peptides. For one ELISA, a peptide with the same sequence as pep F, but with Cys appended to the C terminus, was used (SKPGTGSETAPSKPGTGSETAPSK PGKC).

Murine MAbs were constructed at the University of Missouri-Columbia Cell and Immunology Core Facility as described in detail previously (6, 7), using the ELISA as an initial screening assay, followed by other immunological assays described below. Isotypes of MAb 2H8 to VmcE [immunoglobulin G1(κ) [IgG1(κ)], 29E9 to VmcF [IgG2b(κ)], and 14H2 to VmcF [IgG1(κ)] were determined by ELISA, using HP-conjugated, isotype-specific secondary Abs (Southern Biotech, Birmingham, AL).

Triton X-114 (TX-114) detergent phase fractionation of broth-grown *M. capricolum* subsp. *capricolum* was performed using standard procedures described in detail elsewhere (57). Proteins were visualized after separation by Tris-glycine sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) in 10% gels by staining them with Coomassie brilliant blue R250 (Serva, Heidelberg, Germany) or silver (Bio-Rad, Hercules, CA). Western blot analyses were performed as previously described (11). Briefly, separated proteins were transferred to nitrocellulose blots that were blocked with 10% fetal calf serum and incubated successively with primary mouse Ab, HP-conjugated secondary Ab, and 4-chloro-1-naphthol substrate in the presence of hydrogen peroxide to visualize the bound Ab. Colony immunoblot analyses were performed as described previously in detail (6, 39) by transfer of 3- to 4-day-old *M. capricolum* subsp. *capricolum* colonies to nitrocellulose filters that were then blocked and immunostained as described for Western blot analyses.

Proteomics. TX-114 phase proteins were prepared and precipitated with methanol as previously described (57). Precipitated proteins were digested either en masse in solution or in gel after their separation by SDS-PAGE. For in-solution digestion, the complex mixture of proteins was dissolved in 8 M urea–200 mM Tris buffer (pH 7.8), reduced with dithiothreitol, alkylated by iodoacetamide, and treated with trypsin (modified trypsin, sequencing grade; Promega, Madison, WI) at an enzyme-to-substrate ratio of 1:50 (wt/wt) overnight at 37°C. For in-gel digestion, the SDS-PAGE gel was cut into 40 bands that covered the entire length of the gel, and each band was digested according to a standard in-gel digestion protocol. Digested peptides were analyzed using a capillary liquid chromatography-tandem mass spectrometry (LC-MS/MS) system. They were loaded sequentially on an on-line trap column (0.25 mm by 30 mm, Magic C18AQ, 5 μ m, 100 Å; Michrom BioResources, Auburn, CA) at a flow rate of 10 μ l/min with buffer A (see below). After application and removal of salt and urea, the flow rate was decreased to 300 nl/min, and the trap column effluent was switched to a home-built fritless reverse-phase microcapillary column (0.1 mm by 180 mm, packed with Magic C18AQ, 5 μ m, 100 Å) (17). The reverse-phase separation of peptides was performed using a Paradigm MG4 LC system (Michrom BioResources) and buffers of 5% acetonitrile–0.1% formic acid (buffer A) and 80% acetonitrile–0.1% formic acid (buffer B), using a 150-min gradient (0 to 10% buffer A for 20 min, 10 to 45% for 110 min, 45 to 100% buffer B for 20 min). Peptides separated by reverse-phase chromatography were sprayed directly into an ion-trap mass spectrometer (LCQ Deca XP Plus; Finnigan, San Jose, CA). An MS survey scan was obtained for the *m/z* range of 400 to 1,400, and MS/MS spectra were acquired for the three most intense ions from the survey scan. An isolation mass window of 3.0 Da was used for the precursor ion selection, and normalized collision energy of 35% was used for the fragmentation. Dynamic exclusion for a 2-min duration was used to acquire MS/MS spectra from low-intensity ions. SEQUEST (13) analysis software (Bioworks v3.1) was used to find the best matching peptides in a translated open reading frame (ORF) database generated in Artemis from the complete *M. capricolum* subsp.

capricolum genome. DTA files (Bioworks v3.1) in ASCII format for each MS/MS spectrum were generated from the raw data for a peptide mass range of 500 to 3,500, a minimum ion count of 10, and a minimum signal of 10⁵. Peptide (parent ion) tolerance of 2.5 Da and fragment ion tolerance of 1 Da were allowed, and fixed modification of carbamidomethylation on Cys (+57 Da) and differential modification of oxidation on Met (+16 Da) were used. ProteinProphet (30) was used to evaluate peptide and protein probability. A total of 185 expressed proteins were identified by the combined approach of a shotgun proteomics (en masse) and SDS-PAGE separation with a 99% confidence level (data not shown). Among these proteins were all six predicted Vmc products. Representative peptide sequences for each (all above 90% confidence level) are shown in Results. Full data on peptides mapped to Vmc or other ORFs described in the text are included in the supplemental material.

Genomic shotgun sequencing. Genomic DNA isolated by phenol extraction from *M. capricolum* subsp. *capricolum* clone mck3 was sequenced to 8 \times redundancy, using shotgun sequencing as described elsewhere (36, 41). Six different plasmid libraries with 3- to 5-kb DNA inserts were prepared using purified chromosomal DNA. For each library, genomic DNA was randomly fragmented by sonication and end repaired with T4 DNA polymerase. DNA fragments were resolved on a 1% agarose gel, and fragments of 3 to 5 kb in length were isolated from the gel, ligated with the pUC18 vector, and introduced into competent *Escherichia coli* host XL10-Gold cells (Stratagene, LaJolla, CA). Plasmid DNA templates from 9,000 transformants (1,500 from each library) were prepared in 96-well format using an Eppendorf 5-Prime Perfect station. DNA templates were sequenced from both ends using M13 forward and reverse primers and BigDye Terminator chemistry. Sequences were resolved on PE Biosystems 3700 capillary sequencers, assembled using Phred-Phrap (15), and viewed using consed (20). Sequences spanning *vmcA* to *vmcD* and *vmcE* to *vmcF* were submitted to the National Center for Biotechnology Information (GenBank; <http://www.ncbi.nlm.nih.gov>).

PCR, DNA sequencing, and molecular cloning. All PCRs were performed with a PerkinElmer GeneAmp 2400 or model 480 thermocycler, using a primer concentration of 0.2 μ M. DNA sequencing of amplicons and engineered recombinant constructs was performed at the DNA Core Facility at the University of Missouri-Columbia, using BigDye Terminator chemistry on a PE Biosystems 3730 capillary DNA sequencer.

For analysis of the dinucleotide tract 5' of *vmcF* in clonal expression variants, 10 to 20 ng of *M. capricolum* subsp. *capricolum* genomic DNA template was used in a reaction mixture containing 200 μ M deoxynucleoside triphosphate, 5 mM Mg²⁺, and 3 U of *Taq* polymerase (Promega), with primers (5' to 3') CAAGA ATTCCATTATAACAATACTACTA and GATCTGTCTCTGGTTTCTTC. PCR incubation conditions were as follows: one cycle of 1 min at 94°C, 35 cycles of 40 s at 94°C, 40 s at 44°C, and 50 s at 72°C, and a final extension for 10 min at 72°C. The PCR amplicon was purified using the QIAquick PCR purification kit (QIAGEN) and eluted in 35 μ l of 10 mM Tris, pH 8.5 (QIAGEN buffer EB). Two microliters of this preparation was subjected to agarose gel electrophoresis to determine the size and quantity of amplicon by ethidium bromide staining and comparison to standard markers. Approximately 40 ng of the 0.4-kb amplicons was sequenced using the respective amplifying primers. Amplicons from representative expression variants were sequenced on both strands.

To confirm the chromosomal location of *vmcF* in clonal VmcF expression variants, PCR was performed using primers GAAGAAAAACAGGAACAG GATC and GTTCGTGTGGATTTTAAATGTACC and conditions similar to those used for analysis of the dinucleotide tract 5' to *vmcF*, except for the following incubation conditions: one cycle of 1 min at 94°C, 35 cycles of 40 s at 94°C, 30 s at 44°C, and 2 min at 72°C, and a final extension for 10 min at 72°C. To confirm the chromosomal location of *vmcE* in these same variants, PCR was performed using primers GACACAGAAAGGGAAACC and ACCTTCAGC TGGTTTTCTTC in a reaction mixture containing approximately 40 ng of template, 3 mM Mg²⁺, 600 μ M deoxynucleoside triphosphate, and 3 U of *Pfu* Turbo polymerase (Stratagene). The following incubation conditions were used: initial denaturation for 1 min at 94°C, 32 cycles of 30 s at 94°C, 1 min at 46°C, and 18 min at 68°C, and a final extension of 10 min at 68°C. The respective amplicons linking *vmcF* (1.68 kb) or *vmcE* (9.88 kb) to their adjacent conserved chromosomal flanking regions were purified as described above and sequenced using the respective amplifying primers. To confirm the length of the repeating region in *vmcA*, a PCR was performed using genomic template from clone mck3 and primers GGCGATGATTCTGGAAGTGG and TAGTGAATCTGATACTGTT AATAACC. The following incubation conditions were used: one cycle of 1 min at 94°C, 35 cycles of 1 min at 94°C, 1 min at 51°C, and 1 min 45 s at 72°C, and a final extension for 10 min at 72°C.

For the cloning and expression of VmcE and VmcF fusion proteins, amplicons were generated from *M. capricolum* subsp. *capricolum* mck3 genomic DNA,

using a common forward primer, TTCCCCGGGGATAGATCAAATACT GAA, with an engineered SmaI site and reverse primers CCAGGATCCGTGA GTTTTTTATCTTCC (for *vmcE*) and GGAGGATCCGTGTTAGATAAGT GAGTT (for *vmcF*), each with engineered BamHI sites. A PCR was performed using *Pfu* Turbo, with the following incubation conditions: one cycle of 94°C for 1 min, 35 cycles of 1 min at 94°C, 1 min at 53°C, and 72 s at 72°C, and a final extension for 10 min at 72°C. Amplicons were purified by electrophoresis in a low-melting-point agarose gel, recovered with the QIAquick gel purification kit (QIAGEN), and ligated with pCR4Blunt-TOPO vector using the Zero Blunt TOPO cloning kit (Invitrogen, Carlsbad, CA). Inserts from randomly selected transformants were confirmed by sequencing, using M13 forward and reverse primers. The 0.9-kb insert containing *vmcE* was also sequenced using the initial amplifying primers, in order to determine the number of repeat sequences in the coding region. A complete sequence across the insert was not achieved due to a decrease in the quality of sequence over this length. Nevertheless, 22 tandem repeats were confirmed, consistent with the 26 predicted from shotgun sequencing. Recombinant *vmcE* and *vmcF* gene sequences (lacking signal peptide sequences) were recovered from their respective constructs by SmaI-BamHI digestion, purified in 1% low-melting-point agarose gels, and ligated with XmnI-BamHI-digested, and then purified, pMALc-2 (TEV) vector (24) to create in-frame *malE::vmc* gene fusions. Transformant colonies, grown in *E. coli* DH10B and selected with ampicillin, were screened for inserts by membrane hybridization as previously described (3), using digoxigenin-labeled oligonucleotides corresponding to those used for initial amplification of the respective genes. Plasmid DNA from positive transformants was sequenced using flanking, vector-specific primers GAAATCATGCCGAACATCCCGCAG and ATCCGAATTCTGAA ATCCTTCCCT to confirm sequences. This also provided verification of the number of repeats in the recombinant *vmcF* sequence. Fusion protein was overexpressed from each recombinant construct and purified by amylose affinity chromatography as described previously (25).

Informatics. Nucleotide sequence comparisons were performed with BLASTn (without filter) through the National Center for Biotechnology Information (NCBI) resource (<http://www.ncbi.nlm.nih.gov>). Comparison of genome and regional genomic sequences utilized BLASTn output files visualized in ACT (<http://www.sanger.ac.uk/Software/ACT/>), with a default setting for the figure of 100 bp as the minimum size of the matches shown. ORF and DNA sequence analyses and ORF library construction were performed using Artemis (<http://www.sanger.ac.uk/Software/Artemis/>). Protein comparisons were performed with BLASTp through NCBI. The last search for sequence similarities was performed on 15 May 2006. Protein sequences were analyzed for secondary structure using the PELE suite of programs, and membrane topologies were analyzed using TM-HMM and TMAP; other properties are available through Biology WorkBench (<http://seqtool.sdsc.edu/CGI/BW.cgi#1>). Disordered structure predictions were performed with DISOPRED (<http://bioinf.cs.ucl.ac.uk/disopred/>) and DisEMBL (<http://dis.embl.de/>). Coiled-coil structure was analyzed with COILS (http://www.ch.embnet.org/software/COILS_form.html) and MARCOIL (<http://www.isrec.isb-sib.ch/webmarcoil/webmarcoilC1.html>), including weighted programs for charged polypeptides and a window of 28 residues.

Nucleotide sequence accession numbers. The sequences of regions spanning the *vmcA* to *vmcD* genes and the *vmcE* to *vmcF* genes in *M. capricolum* subsp. *capricolum* clonal isolate mck3 have been assigned GenBank accession numbers DQ480439 and DQ480440, respectively.

RESULTS

Chromosomal regions containing *vmc* genes. Two sources of genomic DNA sequence from *M. capricolum* subsp. *capricolum* Kid type strain (ATCC 27343) were examined in this study: (i) the recently completed genome sequence (GenBank accession no. NC_007633) determined from a propagated culture of the ATCC stock culture and (ii) a high-coverage (>8×) genome sequence draft determined from a subcloned isolate (clone mck3) of this same source, described in Materials and Methods. For purposes of comparison with other mycoplasmal genomes, the sequence from the completed *M. capricolum* subsp. *capricolum* genome was used, whereas sequence derived from the mck3 clonal isolate was used to characterize variable genes and their products expressed in this clonal population. As expected, these two sequences were nearly identical over the

~90% of the genome covered by both sequencing efforts, with minor exceptions in variable regions that are discussed below in further detail.

We compared the two available complete genomes that currently represent the *M. mycoides* cluster: *M. mycoides* subsp. *mycoides* SC PG1 (53), GenBank accession no. NC_005364 (1,211,703 bp), and *M. capricolum* subsp. *capricolum*, GenBank accession no. NC_007633 (1,010,023 bp). As expected from their phylogenetic classification, the overall chromosomal content reflected extensive regions of high sequence similarity as well as striking similarity in gene synteny, involving conserved housekeeping genes, hypothetical proteins, and intergenic regions. Superimposed on this overall similarity were several chromosomal inversions involving segments ranging in length from <1 to >300 kb. Global differences between the chromosomes were also apparent, such as the large number of putative genes (72 ORFs) related to insertion sequence (IS)-like elements in *M. mycoides* subsp. *mycoides* SC PG1 (53) compared to those in *M. capricolum* subsp. *capricolum* (2 ORFs). Some of these differences have been presented elsewhere (K. Röske, M. J. Calcutt, J. I. Glass, A. Persson, J. Westberg, and K. S. Wise, Abstr. 104th Gen. Meet. Amer. Soc. Microbiol., abstr. G-027, 2004). Further examination at finer resolution revealed several chromosomal regions containing putative genes that were selectively present in one genome or the other. Among these genes were several that encoded predicted surface lipoproteins, having characteristic signal peptide and lipobox signatures. These differences provided the impetus to examine the *M. capricolum* subsp. *capricolum* genome in order to identify variable surface lipoproteins that were novel within the *M. mycoides* cluster.

We searched for an initial set of lipoprotein genes selectively present in *M. capricolum* subsp. *capricolum* with features similar to those of other known mycoplasma variable lipoprotein gene families and, in particular, the prototypic *vmm* gene in *M. mycoides* subsp. *mycoides* SC PG1 (33). The *vmm* gene is subject to phase-variable expression that is associated with insertion/deletion (indel) mutations that occur in a region of dinucleotide (TA) repeats lying between a putative -10 and -35 box located 5' of the *vmm* structural gene. Transcripts of *vmm* and the Vmm lipoprotein translation product are documented to be expressed in spontaneous mutants having a TA repeat number of 10. In contrast, mutants containing 12 TA repeats do not express Vmm (33). This "poly(TA)" motif is one of the distinctive variable-number tandem repeat (VNTR) sequences found associated with phase-variable expression of surface proteins in assorted mycoplasma systems (35, 61) and is the only one characterized in the *M. mycoides* cluster. One of our initial search criteria was the presence of an appropriately placed poly(TA) tract in the noncoding 5' flanking region of candidate lipoprotein genes, emphasizing those that most closely resembled the *vmm* gene. An additional criterion was the presence of in-frame tandem repeat sequences in the coding region of candidate genes. Such repeats are a hallmark of some other variable gene families reported for mycoplasmas (35, 61) and can generate translation products of variable length on the surface of the organism, through indels that occur in repeated sequences (62).

Through this process, we identified two loci in the *M. capricolum* subsp. *capricolum* chromosome that contained six novel

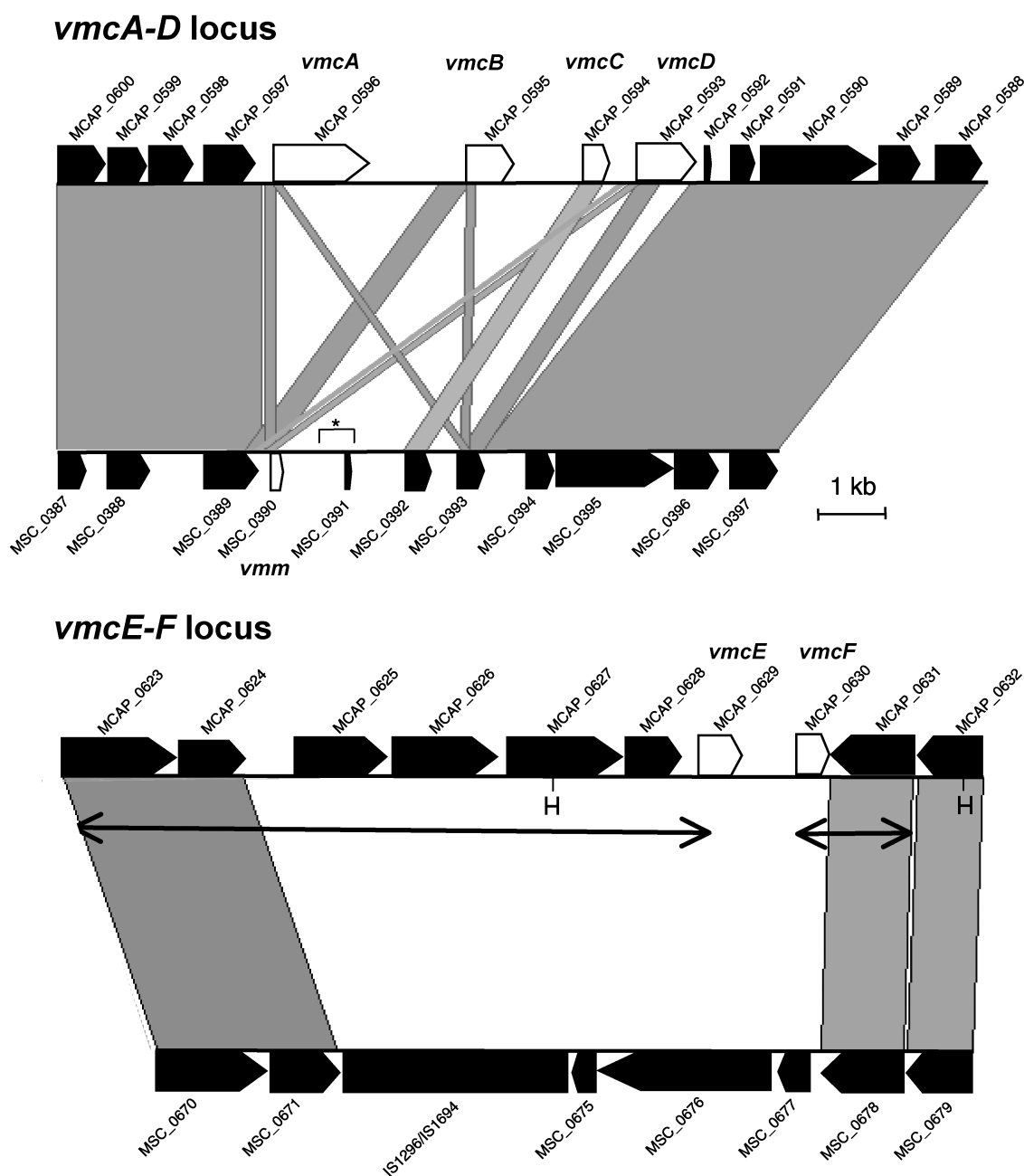


FIG. 1. Comparison of genomic regions containing *vmc* genes of *M. capricolum* subsp. *capricolum* with corresponding regions of the *M. mycoides* subsp. *mycoides* SC PG1 genome. Chromosomal segments (horizontal black lines) of *M. capricolum* subsp. *capricolum* (top) and *M. mycoides* subsp. *mycoides* SC PG1 (bottom) were compared by BLASTn and visualized using ACT with a setting of 100 bp as the minimum size of the matches shown (gray shading), as described in Materials and Methods. Genes are indicated by arrows and labeled as annotated in GenBank for *M. capricolum* subsp. *capricolum* (accession no. NC_007633) or *M. mycoides* subsp. *mycoides* SC PG1 (accession no. NC_005364). Arrows without fill indicate genes shown previously (33) or in the current study to express translation products, including six *vmc* genes in *M. capricolum* subsp. *capricolum* and the *vmm* gene of *M. mycoides* subsp. *mycoides* SC PG1. The bracket with the asterisk in one *M. mycoides* subsp. *mycoides* SC PG1 locus indicates the location of an IS-associated DNA sequence (nucleotides [nt] 443543 through 444153 in GenBank accession no. NC_005364) that is precisely reiterated at another location in that chromosome. Horizontal bidirectional arrows in the *vmcE* to *vmcF* locus indicate the location of PCR amplicons used to establish the stable location of *vmc* genes in clonal expression variants. In the same locus, HindIII sites (H) that create a restriction fragment described in the text are shown. The sequences of these two *M. capricolum* subsp. *capricolum* loci correspond to coordinates in GenBank accession no. NC_007633 (*vmcA* to *vmcD* locus, nt 700569 through 714671; *vmcE* to *vmcF* locus, nt 741987 through 756176). Sequences spanning *vmc* genes in the equivalent loci from *M. capricolum* subsp. *capricolum* clone mck3 are deposited under separate entries in GenBank (see Materials and Methods). These include a region containing *vmcA* to *vmcD* corresponding to nt 711635 through 704793 in GenBank accession no. NC_007633 and a region containing *vmcE* to *vmcF* corresponding to nt 751560 through 753820 in GenBank accession no. NC_007633. These sequences are nearly the same as those represented in the figure, with the exception that different numbers of repeats occur in repetitive sequences associated with some *vmc* genes.

ORFs representing a potential family of genes encoding variable lipoproteins of *M. capricolum* subsp. *capricolum*. These loci are compared in Fig. 1 to the analogous regions in the *M. mycoides* subsp. *mycoides* SC PG1 chromosome, using a BLASTn comparison at intermediate stringency to depict regions of high similarity. These putative genes of *M. capricolum* subsp. *capricolum* encode predicted products termed Vmc lipoproteins and comprise the *vmc* gene family investigated in this report. Four of these (*vmcA* to *vmcD*) occur at one locus, and two (*vmcE* to *vmcF*) occur at another. These loci are separated by ~40 kb of intervening sequence comprised predominantly of genes, including multiple housekeeping genes, with sequences and synteny highly similar to those of genes in the corresponding region of the *M. mycoides* subsp. *mycoides* SC PG1 genome. No characteristics suggested a genome "island" encompassing both loci. Within each locus, the *vmc* genes were adjacent, similarly oriented, and separated by relatively large intergenic regions ranging from 397 to 1,452 bp. No Rho-independent terminator motifs were associated with *vmc* genes. The G+C content of intergenic regions (19 to 20%) was lower than the whole-genome average (23.8%), whereas those of the *vmc* coding regions (*vmcA* to *vmcD*, 31 to 34%; *vmcE* to *vmcF*, 43 to 45%) were markedly elevated by comparison. Each of the two loci in *M. capricolum* subsp. *capricolum* is flanked by genes with sequences and synteny that are highly similar to those of their counterparts in *M. mycoides* subsp. *mycoides* SC PG1 (portions of the flanking regions are shown in Fig. 1).

The *vmcA* to *vmcD* locus is flanked at one end, 5' of *vmc* genes, by housekeeping genes MCAP_0600 and MCAP_0599 as well as genes for hypothetical proteins MCAP_0598 and MCAP_0597 (the latter forming the immediate boundary of the locus). All of these genes have counterparts in *M. mycoides* subsp. *mycoides* SC PG1, although the ortholog of MCAP_0598, lying between MSC_0388 and MSC_0389, is not annotated in GenBank. The other end of this locus, 3' of *vmc* genes, is also flanked by a sequence that is highly conserved in *M. mycoides* subsp. *mycoides* SC PG1. The boundary begins upstream of MCAP_0592 and extends through immediately adjacent genes that encode components of a phosphotransferase system (MCAP_0591 and MCAP_0590), ribulose phosphate epimerase (MCAP_0589), and a surface lipoprotein (MCAP_588, containing a verified authentic frameshift) that has ~90% sequence identity to an immunogenic surface lipoprotein encoded by the ortholog MSC_0397 of *M. mycoides* subsp. *mycoides* SC PG1 (27). In contrast to the conserved flanking regions, the sequence within this locus, including *vmc* genes (MCAP_0596 through MCAP_0593), shows distinctive features. With the exception of limited regions of sequence showing similarity between the two genomes, this genomic segment generally represents sequence that is selectively present in the *M. capricolum* subsp. *capricolum* genome (based on further comparison of the whole genomes) (data not shown). The corresponding locus of *M. mycoides* subsp. *mycoides* SC PG1 contains sequence most of which is absent from *M. capricolum* subsp. *capricolum*. Interestingly, one segment in this locus (Fig. 1) is precisely reiterated elsewhere in the *M. mycoides* subsp. *mycoides* SC PG1 genome and is related to the mobility element ISMmyl (54), prevalent in this genome but absent from *M. capricolum* subsp. *capricolum*. Despite the dis-

tinctive features at this locus, in either genome, closer comparison showed that regions of some *vmc* genes have significant DNA sequence similarity to counterpart regions in the *M. mycoides* subsp. *mycoides* SC PG1 genome and that some of these same regions are redundant in the *M. capricolum* subsp. *capricolum* genome. A key focus in this regard is the *vmm* gene, located at the boundary of the locus in the same position as *vmcA* in the *M. capricolum* subsp. *capricolum* genome. Particularly striking is a region of sequence within and adjoining the *vmm* gene (MSC_0390) of *M. mycoides* subsp. *mycoides* SC PG1 that corresponds to similar regions in multiple *vmc* genes. This relationship is also apparent in a whole-genome comparison that additionally reveals similar conserved regions in *vmcE* and *vmcF* genes (data not shown). Another example of local similarity is the presence of limited DNA sequences within *vmcC* and *vmcD* coding regions that correspond to counterpart regions in MSC_0392 and MSC_0393, respectively. The significance of these local similarities is considered in the following section.

The *vmcE* to *vmcF* locus is flanked by sequences that are highly conserved between the *M. capricolum* subsp. *capricolum* and *M. mycoides* subsp. *mycoides* SC PG1 genome (Fig. 1). The immediate 3' flanking region contains hypothetical proteins MCAP_0623 and MCAP_0624 that have corresponding orthologs in *M. mycoides* subsp. *mycoides* SC PG1. The 5' flanking region contains conserved housekeeping genes *pgk* (MCAP_0631) and *gap* (MCAP_0632) that also occur in both genomes. Within the locus are *vmcE* and *vmcF*, again separated by a relatively long intergenic region. The remaining sequence in this locus is absent from the *M. mycoides* subsp. *mycoides* SC PG1 genome. Within the locus is encoded a set of four adjacent ORFs (MCAP_0625 through MCAP_0628), each encoding a protein with membrane-spanning domains but with no known function. In a reciprocal manner, the analogous locus in the *M. mycoides* subsp. *mycoides* SC PG1 genome contains regions of sequence that are absent from *M. capricolum* subsp. *capricolum*, including a segment reflecting a complex insertional event involving the prevalent IS1296 and IS1694 elements and a segment encoding three hypothetical proteins. Overall, the *vmc* gene family of *M. capricolum* subsp. *capricolum* is distributed between two small regions of the chromosome that contain distinctive gene sets, compared in this binary manner with that of *M. mycoides* subsp. *mycoides* SC PG1. In addition, local segments of *vmc*-associated sequences are similar to those of the *vmm* gene of *M. mycoides* subsp. *mycoides* SC PG1.

Features of *vmc* genes and predicted translation products.

The features of *vmc* genes of *M. capricolum* subsp. *capricolum* and the predicted properties of their annotated translation products are shown in Fig. 2, which also shows the assembled sequences from the mck3 clonal isolate that we used for further experimentation. All *vmc* genes encode prolipoproteins with very similar signal peptide sequences, terminating with a standard mycoplasmal lipobox (AVA ↓ C for VmcA to VmcD; VVA ↓ C for VmcE to VmcF). After prolipoprotein acylation, signal peptide processing, and translocation, each mature Vmc surface lipoprotein is predicted to contain a diacylglycerol-Cys residue at the N terminus, anchored in the single limiting plasma membrane of the organism. Each mature Vmc represents a unique sequence among the set of deduced Vmc pro-

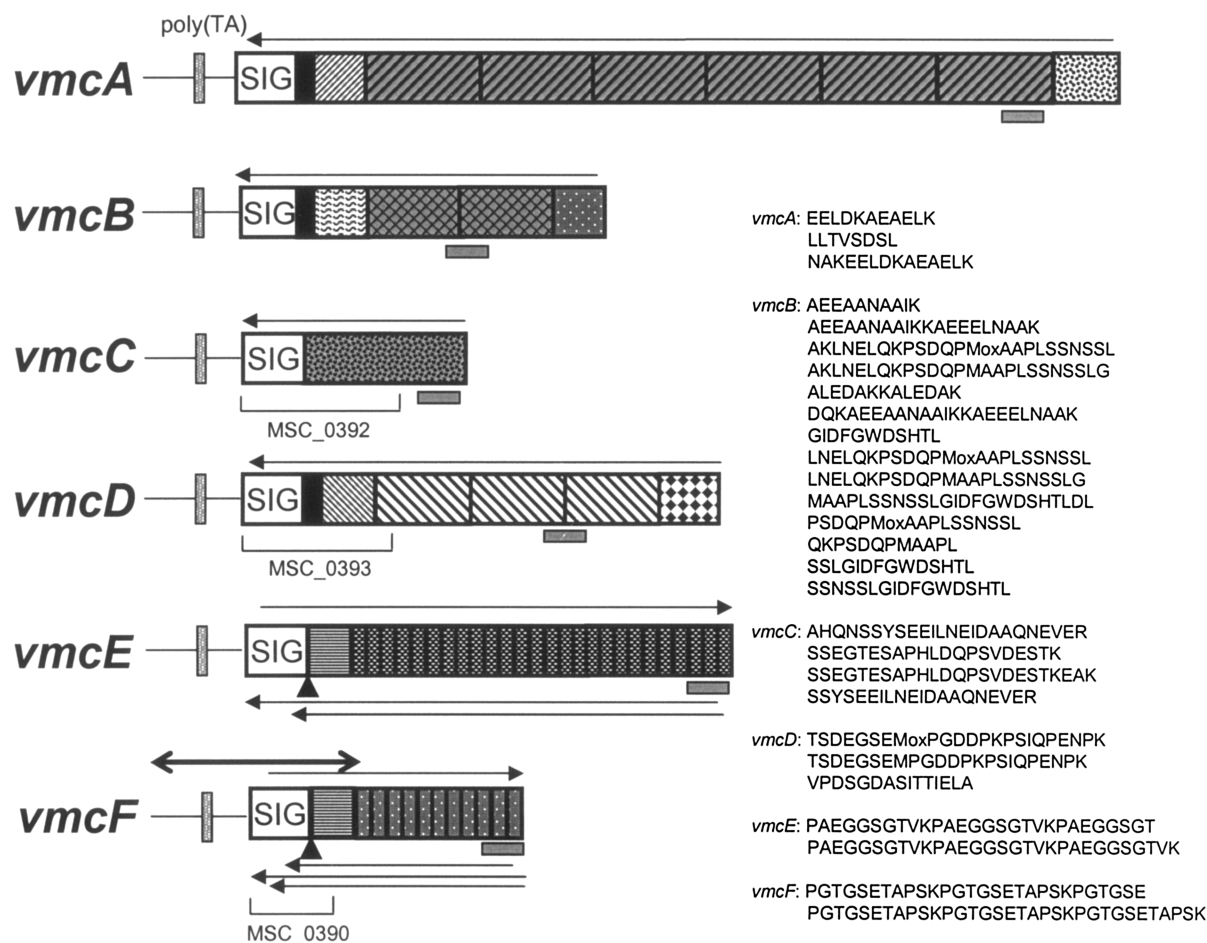


FIG. 2. Features of *vmc* genes and their predicted translation products. Sequence features associated with *vmc* genes in the *M. capricolum* subsp. *capricolum* clonal isolate mck3 are shown. The large rectangles represent the predicted, annotated polypeptide sequences with the following features indicated: a highly conserved signal peptide sequence (SIG) ending in a typical lipobox for acylation and signal cleavage, unique tandem repeat sequences in the coding regions of different Vmc products, and additional regions of shared or unique sequences encoded by particular genes. Regions having similar sequences are shown as boxes with a common fill pattern. Distinctive fill patterns denote dissimilar sequences. Unidirectional arrows indicate alternative, overlapping ORFs that occur in the DNA sequence. Shaded bars below each gene indicate predicted residues specific for each Vmc that were synthesized as peptides to generate PABs. The shaded box to the left of each *vmc* gene represents the position of the poly(TA) tract located within a conserved 5' flanking sequence adjoining each gene. The bidirectional arrow above *vmcF* indicates the location of a PCR amplicon used to characterize the poly(TA) tract from clonal variants in VmcF expression. The arrowheads below *vmcE* and *vmcF* indicate the location of an oligonucleotide sequence, discussed in the text, that is predicted to hybridize selectively with these two genes. Brackets indicate regions of high sequence similarity between a Vmc sequence and a corresponding ORF in the *M. mycoides* subsp. *mycoides* SC PG1 genome (labeled as in Fig. 1). On the right, representative peptides of the membrane-associated proteome generated from TX-114 phase proteins of *M. capricolum* subsp. *capricolum* clone mck3 and assigned by LC-MS/MS to specific Vmc translation products are indicated, as described in Materials and Methods. Mox indicates the presence of oxidized methionine residues.

teins. None of the predicted Vmc ORFs contains Trp codons. Several other ORFs overlap *vmc* gene coding regions (Fig. 2) and contain both UGA and UGG Trp codons, characteristic of mycoplasma genes (60). These were not annotated. Tandemly repeating in-frame DNA sequences occur in *vmc* coding regions and encode corresponding tandem repeat sequences in Vmc surface polypeptides. An exception is *vmcC*, which lacks such repeats but was included in our analysis due to similarities in other features. Both the sequence and the length of the tandemly repeated unit in each Vmc product are unique. Longer repeat units occur in VmcA (80 amino acids [aa]), VmcB (69 aa), and VmcD (68 aa), whereas products of the other *vmc* locus, VmcE (10 aa) and VmcF (11 aa), contain shorter

repeat units analogous to those reported for some mycoplasma variable gene families (35, 61). Assembled shotgun sequencing data from the uncloned parent culture (GenBank accession no. NC_007633) revealed two differences from the mck3 clone: *vmcA* encoded 5 tandem repeats instead of 6, and *vmcE* encoded 16 repeats instead of 26. The numbers of repeats within *vmcA* and *vmcE* for the mck3 clonal isolate shown in Fig. 2 were verified by generating respective amplicons of the predicted size and sequence, using opposing primers to unique sequences flanking the repeat regions (data not shown). Together, these results suggest the possible dynamic variation in Vmc size based on indels in the repeat regions.

Products encoded by *vmc* genes from the two loci differ

markedly in the organization of their sequences and their predicted structures. Repeat sequences in VmcA, VmcB, and VmcD (and the C-terminal half of VmcC) coincide precisely with regions of marked coiled-coil structure. These highly structured domains are interrupted with short (3- to 5-residue) stretches of random coil sequence that could provide flexibility between redundant modules. Other regions of VmcA to VmcD adjacent to coiled-coil modules are nonredundant and comprise unstructured regions predicted to have random coil conformation. These regions contain both shared sequences and sequences unique for a particular product. For example, mature products VmcA, VmcB, and VmcD each contain a nonrepeating region near the C terminus comprising a distinctive protein sequence (Fig. 2). Unstructured regions in the N-terminal portions of these mature proteins also contain unique sequences, but the first 10 to 12 N-terminal residues, adjacent to their cleaved signal peptides, are highly conserved and distinguish this group from other Vmc products. Overall, the Vmc products of this locus are organized with similarly structured modules of repetitive regions affixed to more-flexible regions. Potential orthologs of VmcC and VmcD occur in *M. mycoides* subsp. *mycoides* SC PG1 (Fig. 2), although the limited regions of sequence similarity do not include the highly structured portions of Vmc products.

VmcE and VmcF products encoded in the other locus are organized quite differently. These mature lipoproteins share a nonrepeating N-terminal portion, and distinctive small tandem repeats extend to the C terminus of each protein. The entire sequence of each mature product is predicted to have random coil conformation without secondary structure. The common N-terminal sequence reflects sequence similarity at the DNA level and defines a clear boundary at which the C-terminal sequences of these products diverge. Interestingly, this common region encoded by *vmcE* and *vmcF* is shared in part with the *vmm* gene of *M. mycoides* subsp. *mycoides* SC PG1 (MSC_0390; refer to Fig. 2 for region of sequence similarity), even though the corresponding genes lie in a different locus from that of *vmm* (Fig. 1). More specifically, the first 11 N-terminal residues of the mature VmcE and VmcF polypeptides are identical to those of Vmm. This similarity further explains the previously reported (33) evidence of a "vmm-like gene" in *M. capricolum* subsp. *capricolum*, based on identification in Southern blot analyses of a HindIII genomic fragment hybridizing with a probe specific for this region (Fig. 2). The reported fragment is compatible with the 5,552-bp fragment bearing both *vmcE* and *vmcF* indicated in Fig. 1. Due to divergence in the region of the probe, *vmcA* to *vmcD* gene sequences would not be predicted to hybridize.

All six *vmc* genes of *M. capricolum* subsp. *capricolum* shared sequence features with the *vmm* gene of *M. mycoides* subsp. *mycoides* SC PG1 relevant to their possible phase-variable expression (33). Each *vmc* coding region adjoins a highly conserved 5' flanking sequence with similarity to that of *vmm*, containing -10 and -35 sequence motifs separated by a signature poly(TA) tract (see Fig. 5B for examples). The genome sequence assembled from the mck3 clonal isolate identified poly(TA) tracts ranging from 10 to 16 TA repeats among the *vmc* genes. Due to the possible presence of variant genomes having indel polymorphisms at this VNTR, expression patterns were not predicted but are addressed experimentally in a later

section. Overall, Vmc sequences deduced from the *M. capricolum* subsp. *capricolum* genome predicted that (i) Vmc proteins may present distinct polypeptide products on the cell surface, (ii) complete Vmc products are predicted to occur selectively in *M. capricolum* subsp. *capricolum*, (iii) two distinct structural classes of Vmc product are encoded in corresponding loci of the *M. capricolum* subsp. *capricolum* genome, and (iv) *vmc* genes in both loci are potentially subject to mutations conferring phase-variable patterns of expression for each Vmc product.

Identification of authentic Vmc translation products. To verify and further characterize the predicted Vmc lipoproteins in *M. capricolum* subsp. *capricolum*, murine PABs were raised to synthetic peptides that represented specific portions of each predicted *vmc* coding region (Vmc pep A through Vmc pep F, each containing 28 to 30 predominantly hydrophilic residues). The locations of these sequences (Fig. 2) were selected with various strategies and reflected regions comprising redundant sequences within structured (VmcA to VmcD) or nonstructured (VmcE to VmcF) domains. PAB raised to each Vmc peptide bound with high specificity to its cognate antigen but not to unrelated peptides in the series, as determined in an ELISA using immobilized Vmc pep A through Vmc pep F as targets to measure PAB binding (Fig. 3A). In two cases, highly specific MABs to Vmc pep E (2H8) and to Vmc pep F (29E9 and 14H2) were subsequently derived from immunized animals and were also shown to be highly specific in this ELISA (data not shown). We then used these specific PAB or MAB reagents in two assays to monitor expression of authentic Vmc products in the clonal mck3 population of *M. capricolum* subsp. *capricolum*. We chose this population, understanding that phase-variable expression might fortuitously result in only a few mycoplasmas elaborating a particular Vmc but also because analysis of a clonal derivative offered immediate insight into the degree of variation. Colony immunoblots of this rigorously cloned population showed specific staining with PABs to each of four products: VmcB, VmcD, VmcE, and VmcF. Each PAB revealed striking variation in the staining of individual colonies (Fig. 3B), characteristic of the phase-variable expression of surface antigens occurring at high frequencies (38, 57). At the population level, the patterns of staining among the individual PABs also differed considerably (data not shown). For example, PAB to Vmc pep D stained a small fraction (<0.005) of colonies plated, whereas in the same population, colonies stained by other PABs were far more prevalent (0.80 to 0.92). These results suggested that each of the four distinct Vmc surface antigens identified might reflect an individual expression pattern. No staining was observed with PABs to peptides corresponding to VmcA or VmcC.

Authentic translation products containing the epitopes recognized by the specific Abs were identified by probing Western blots of proteins derived from *M. capricolum* subsp. *capricolum* clone mck3. These were first fractionated to separate proteins that were soluble in an aqueous phase from those amphiphilic membrane proteins (including lipoproteins) that selectively partition into the TX-114 detergent phase (57). As shown for many mycoplasma species, the majority of proteins in *M. capricolum* subsp. *capricolum* fractionated into the aqueous phase with fewer proteins apparent in the TX-114 phase, even when measured by sensitive staining with silver (Fig. 3C, left

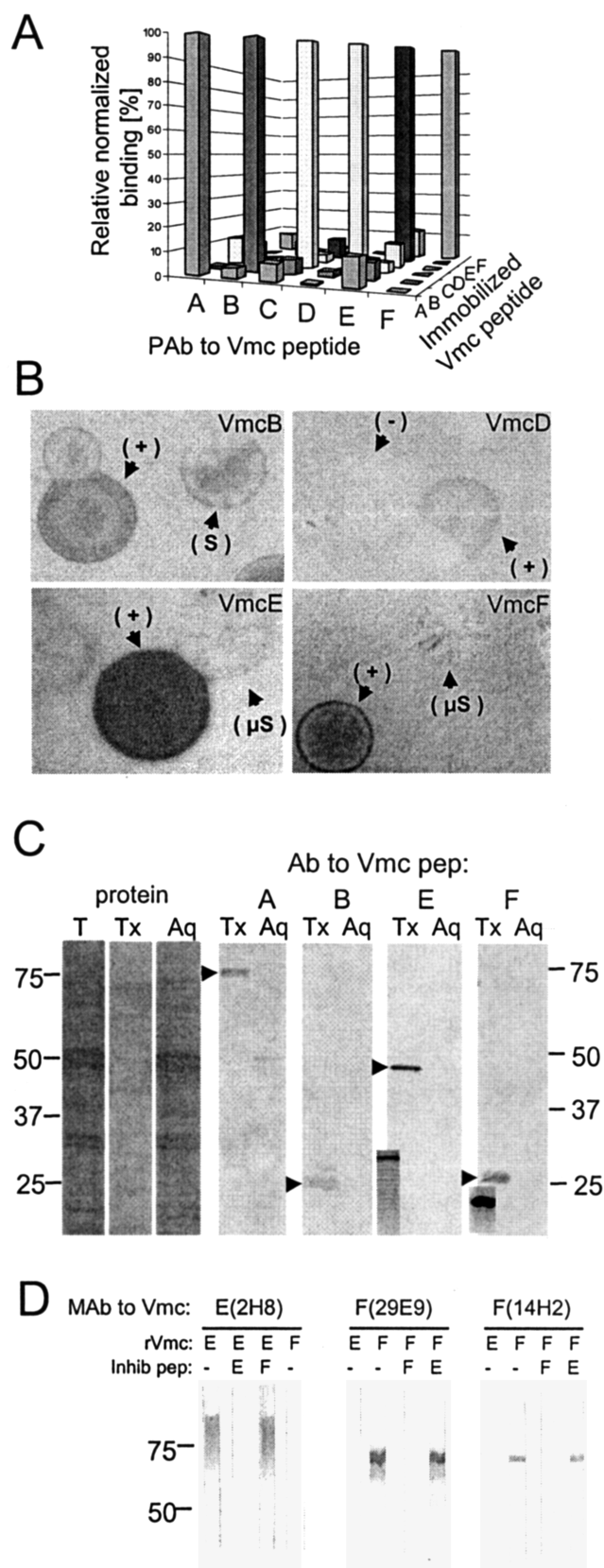


FIG. 3. Immunological characterization of Vmc proteins. (A) Specificity of binding of PABs to synthetic peptides representing unique se-

quences of VmcA to VmcF products. Binding values (relative binding as a percentage of normalized OD obtained with the cognate binding partners) are indicated for each PAB to synthetic Vmc peptides A to F, assayed by ELISA on a panel of all six immobilized peptides (A to F) as described in Materials and Methods. Bars are variously shaded to aid visualization. (B) Immunostaining of colony lifts of the mck3 clonal population of *M. capricolum* subsp. *capricolum*. Staining patterns of specific PABs to Vmc peptides representing VmcB, VmcD, VmcE, and VmcF are shown. Distinctive phenotypic patterns obtained with the PABs are indicated, including colonies showing fully positive (+), completely negative (-), microsectored (μS), or more fully sectorized (S) immunostaining. Colonies range in diameter from 0.5 to 1.0 mm. (C) Detection of Vmc translation products with specific Abs. Total proteins (T) or those from the TX-114 (Tx) or aqueous (Aq) phase of detergent-fractionated cells from *M. capricolum* subsp. *capricolum* clone mck3 were separated by SDS-PAGE and stained with silver (left panels) or transferred to Western blots (right panels) and immunostained with PABs to peptides representing VmcA or VmcB or with MAbs to peptides representing VmcE (2H8) or VmcF (29E9). Insets for VmcE and VmcF indicate the immunostaining patterns of these respective products in heavily loaded whole-cell preparations. TX-114 phase proteins representing VmcA, VmcB, VmcE, and VmcF are indicated by arrowheads. Molecular mass standards (in kDa) are indicated next to the panels. (D) Western blots of recombinant VmcE (rVmcE) or VmcF (rVmcF) immunostained with MAb 2H8 (to VmcE), 29E9 (to VmcF), or 14H2 (to VmcF) in the presence or absence of 10 μM soluble peptide representing VmcE (pep E) or VmcF (pep F). Inhib pep, inhibitory peptide.

panels). Western blotting with Abs to Vmc peptides revealed four distinct Vmc products that partitioned selectively into the TX-114 phase: VmcA, VmcB, VmcE, and VmcF (Fig. 3C). Because each reagent recognized a specific peptide sequence encoded by the appropriate ORF and identified a distinct product with the partitioning characteristics of mature lipoproteins, our results formally confirmed the expression of the annotated VmcB, VmcE, and VmcF ORFs shown in Fig. 2 but failed to identify a product associated with VmcD, possibly due to its underrepresentation in the mck3 clonal population. The clear identification of VmcA in Western blots, but not in colony immunoblots of the same population, was surprising. One possible explanation is that the cognate epitope sequence of VmcA is inaccessible when presented in a native coiled-coil configuration on the cell surface as in colony immunoblots but is exposed under denaturing conditions leading to Western blot analysis. Collectively, our immunological reagents identified five of the six predicted Vmc products. The inability to identify VmcC in either assay may stem from one of several factors, such as epitope inaccessibility or a promoter configuration that precludes expression in the majority of the population.

As an entirely independent method to determine the expression of Vmc protein products as well as other predicted membrane proteins, we explored the use of a proteomics approach, LC-MS/MS, to identify products comprising the "membrane-associated proteome" of *M. capricolum* subsp. *capricolum*. For this analysis, TX-114 phase proteins were subjected to trypsin digestion, either en masse after precipitation with methanol and in-solution digestion or after SDS-PAGE separation and excision of adjacent regions in the gel spanning the entire spectrum of sizes, as described in Materials and Methods. Tryptic peptides were then analyzed by nanoflow capillary LC-

MS/MS (17), and ~30,000 tandem mass spectra were scanned against a library of the annotated ORFs encoded in the *M. capricolum* subsp. *capricolum* genome. Among several putative products predicted to be present in the TX-114 phase population, peptides unambiguously assignable to each Vmc product, including VmcC, were identified (Fig. 2). These results confirmed and extended our immunological data, demonstrating that the predicted translation products were expressed from the annotated ORFs of all six *vmc* genes. This sensitive technique, although not quantitative, was a critical adjunct to our immunological methods and provides a direct means of identifying mycoplasmal membrane surface protein expression in cases where identification of antigenic entities may be hampered by their physical state, low abundance, or variable expression.

Antigenic structure and surface architecture of Vmc lipoproteins. Unusual characteristics of both structured and unstructured Vmc proteins were apparent during their analysis by standard SDS-PAGE analysis. Although VmcB migrated with a measured relative molecular mass (M_r) similar to that predicted from its deduced sequence, VmcA showed a discrepancy between its M_r (74 kDa) and calculated mass (60.7 kDa). It is possible that the extensive structure imposed by the large coiled-coil region of this product, compared to a much smaller structured region in VmcB, could account for this anomaly. Also striking were the characteristics of VmcE and VmcF products expressed in *M. capricolum* subsp. *capricolum* and their similarity to properties observed for variable lipoproteins comprising short tandem repeats in other mycoplasma species (56, 61). Both lipoproteins showed characteristic ladder patterns in more-heavily loaded Western blots of whole organisms (Fig. 3C, insets); all members of these ladderized sets partitioned into the TX-114 phase fraction (data not shown). For some mycoplasmal variable surface proteins (62), such patterns have been rigorously shown to result from indel mutations in tandem repetitive sequences within coding regions that generate heterogeneity in the number of repeats expressed by different progeny in a propagating population. Typically, a prevalent size of protein dominates in the cell population, with fewer members expressing products of different size. We did not segregate VmcE or VmcF size variants to formally distinguish this model from an alternative possibility that size variants represent precise degradation products of proteolysis. Nevertheless, our observation that size variant *vmcE* genes occurred in clonal versus parental populations of *M. capricolum* subsp. *capricolum* is compatible with the former model. These two mature lipoproteins also show a marked discrepancy between their M_r s determined from SDS-PAGE (VmcE, 46 kDa; VmcF, 26 kDa) and the masses predicted from DNA sequence derived from the *mck3* clone (VmcE, 26.6 kDa; VmcF, 14.7 kDa). For each product, this feature is also reflected in the size interval between members of its ladderized set, showing an analogous discrepancy between the measured (larger) and calculated (smaller) differences in mass between size variants proposed to differ by integral numbers of repeats (data not shown).

To further characterize the VmcE and VmcF polypeptides and the interactions of their repetitive sequences with cognate peptide-directed MAbs, recombinant plasmids encoding fusion proteins (rVmcE and rVmcF) were constructed and expressed

in *E. coli*. These represented *malE::vmc* translational fusions expressing maltose binding protein (~41 kDa) joined at its C terminus to the full-length mature VmcE or VmcF lipoprotein sequence, devoid of a signal peptide or a lipid modification site. Both fusion proteins were readily soluble in aqueous buffers and were purified by affinity chromatography. The number of repeats in the native genes was preserved in each of these recombinant products, as determined by sequencing amplicons generated from recombinant plasmid templates as described in Materials and Methods. In Western blot analysis, MAbs to VmcE (2H8) or VmcF (29E9 and 14H2) specifically bound to the corresponding rVmc product (Fig. 3D). The relative molecular masses measured in SDS-PAGE for rVmcE (~90 kDa) and rVmcF (~68 kDa) were greater than the masses predicted from their corresponding deduced sequences (rVmcE, 72.5 kDa; rVmcF, 60.6 kDa), consistent with the discrepancy observed for each of the native products. This result argued against postranslational adducts in the mycoplasma as a basis for the abnormal migration of VmcE or VmcF in SDS-PAGE and suggested that the unusual behavior was a feature of the polypeptide sequence per se. The predicted absence of structure in these products further discounted a structural basis for the effect. A final possibility is that the acidic nature of these products (pI 5.3 or 5.4, respectively) may prevent SDS binding during PAGE. This property may also explain their poor protein staining properties, a feature also noted with similar variable proteins, such as Vlp (39).

With MAbs to VmcE and VmcF, we could examine the binding of each specific reagent to its cognate epitope in multiple contexts. Binding to each corresponding rVmc polypeptide was completely and specifically inhibited in the presence of the soluble peptide initially used for immunization and Ab screening by ELISA. Additional studies with MAbs to VmcF showed that a synthetic peptide with identical sequence to Vmc pep F, but coupled through a Cys residue appended to the C terminus rather than the N terminus, bound as well as the original peptide to either MAb when used as an immobilized target in ELISA. These results confirmed that the repeat sequence was sufficient to mediate MAb binding as a soluble peptide or in a recombinant or authentic translation product containing many tandem repeats. By extension, we determined that binding of PABs and MAbs to all Vmc proteins detected by Western or colony immunoblotting was similarly inhibited, selectively, by their cognate synthetic peptides (data not shown). Overall, these studies show that the detection system we developed for Vmc proteins can be defined by epitopes accessible on soluble polypeptides.

Phase-variable and combinatorial expression of Vmc products. An important consequence predicted by the conserved 5' flanking sequence of each *vmc* gene is that each gene (and corresponding Vmc lipoprotein) may be independently subject to phase-variable expression through indel mutations that alter the poly(TA) tract. The colony immunostaining patterns that we initially observed in the clonal *mck3* population by using specific Ab to Vmc proteins (Fig. 3B) also supported this possibility. To more rigorously assess the pattern of phase-variable expression of Vmc proteins and to formally define the underlying mutational mechanism, we characterized several clonal populations varying in expression of VmcF, using MAb to this product as a precise tool to identify and score popula-

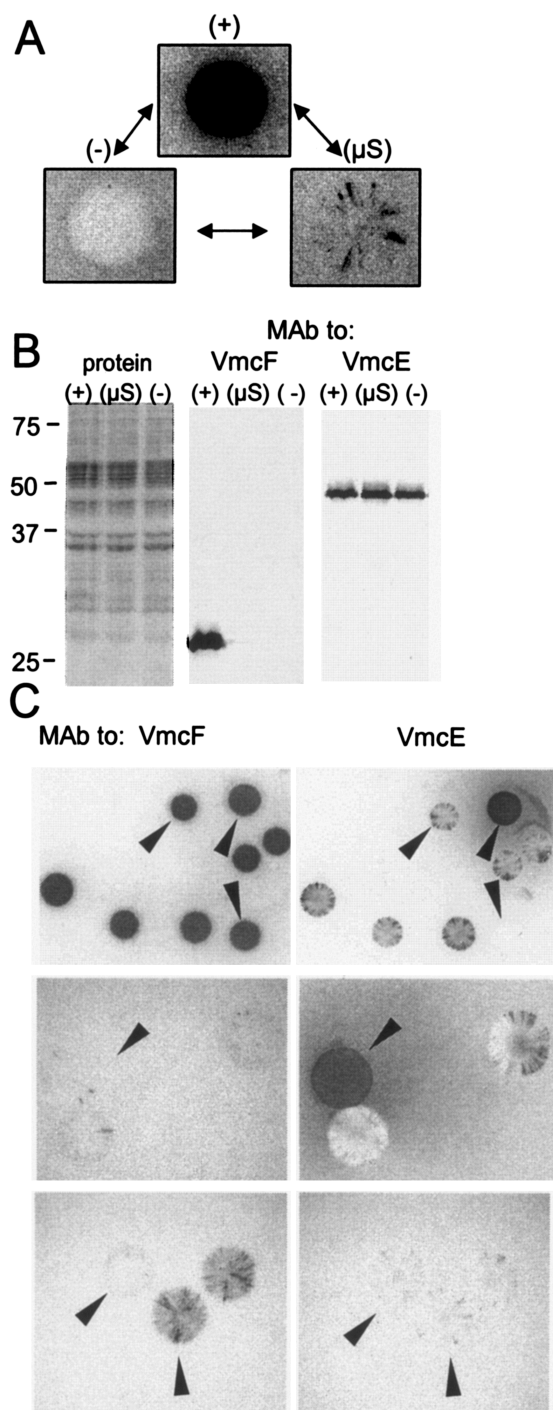


FIG. 4. Variable expression of VmcE and VmcF lipoproteins in clonal populations of *M. capricolum* subsp. *capricolum*. (A) Phenotypes identified by immunostaining colony lifts with MAb 14H2 to VmcF. Three distinct staining patterns define the positive (+), negative (-), and microsectored (μ S) phenotypes of individual colonies. Bidirectional arrows summarize the phase-variable interconversion of these heritable phenotypes (also described in the legend to Fig. 5). (B) Western blot analysis of VmcF expression in clonal populations having the VmcF⁺, VmcF⁻, or VmcF ^{μ S} phenotype. Normalized loads of whole protein were subjected to SDS-PAGE and stained with Coomassie blue R250 (left) or transferred to Western blots and immunostained with MAb 29E9 to VmcF (middle) or MAb 2H8 to VmcE (right). Populations representing each phenotype are from the lineages shown

in Fig. 5A and are designated Ic (+), IIc (μ S), and IIIc (-) in that figure. (C) Expression patterns of VmcF and VmcE within the identical *M. capricolum* subsp. *capricolum* clonal population. Replicate colony lifts from clonal populations, derived from either Ic (+) or Id (-) (Fig. 5A), were immunostained with MAb 14H2 to VmcF (left) or MAb 2H8 to VmcE (right). Arrowheads at the same locations in the right and left panels indicate examples of colonies in the population that show different combinatorial expression patterns for VmcE and VmcF. Fields for comparison were selected to illustrate these differences.

tions for subsequent analysis. In initial studies, using *mck3* and its derivative clonal populations, three distinct and consistent phenotypes in colony immunostaining were observed. Individual colonies either were stained uniformly, showed no staining, or displayed a generally negative phenotype but with small sectors of staining in peripheral populations of the colony, termed "microsectors" (μ S), as illustrated by the examples shown in Fig. 4A. We then used the MAb to identify and isolate well-separated colonies representing each of the three phenotypes. Isolates were subsequently plated following minimal replication, and the selected populations were scored and further subcloned to isolate each phenotype. Whereas the selected phenotype predominated in subcloned populations, all plated populations contained a small fraction (0.001 to 0.05) of colonies displaying each of the two nonselected phenotypes (see Fig. 5A for examples). These results indicated that the phenotypes were heritable and could interconvert, as previously demonstrated for variable phenotypes in other mycoplasma systems (38, 39).

Having segregated populations that predominantly expressed each phenotype, we formally determined by Western blot analysis whether these populations per se expressed different levels of the VmcF translation product (Fig. 4B). Normalized loads of whole-cell preparations from each population revealed striking differences in VmcF expression that correlated with the phenotype in colony immunoblots. The μ S population demonstrated only trace amounts of VmcF, consistent with the notion that a very small proportion of positive cells was observed in these populations. Importantly, these results confirmed that the phenotypes in colony immunoblots reflected variable expression of the translation product rather than the variable "masking" of epitopes on a constitutive product, a potentially confounding phenomenon that has been documented for other mycoplasma systems (40, 46, 65).

We then extended this analysis to determine whether variable expression of VmcF was independent of expression of other members of the Vmc family by monitoring VmcE as an alternative partner for comparison. VmcE was chosen due to the proximity of the corresponding *vmcE* and *vmcF* genes and the availability of a specific MAb to monitor expression. We first assessed the same clones showing variable VmcF expression (Fig. 4B) for the expression of VmcE. The three clones expressing variable amounts of VmcF by Western blot analyses all expressed VmcE. In one case, for the clone selected for the VmcF⁻ phenotype (Fig. 4B, right lanes), the result formally showed that VmcE can be expressed without expression of VmcF (a result also consistent with the absence of transcriptional read-through from the *vmcE* gene). Similarly, another

clonal population selected for the VmcF⁺ phenotype (Fig. 4B, left lanes) appeared to express both products in combination. To verify this possibility and to confirm generally that variable expression of VmcE and VmcF is noncoordinate, we used a second strategy wherein replicate lifts of the same plated clonal population were stained with MAb to VmcE or MAb to VmcF. The high resolution and specificity in these immunoblots revealed dramatic differences in the expression patterns of VmcF and VmcE, both in the plated clonal populations and within single colonies (Fig. 4C). Examples of coexpression, selective expression of either product, and distinct microsectoring patterns within single colonies clearly indicated that expression of these two products was noncoordinate. It is noteworthy that the expression patterns observed for VmcE also showed the three expression phenotypes characterized for VmcF (Fig. 4C). Finally, to determine the predicted independence of switching of other Vmc proteins, we established lineages selected for phase-variable expression of VmcB⁺ and VmcB⁻ by using our specific PAb (data not shown). Variants of either VmcB phenotype showed noncoordinate switching of VmcE and VmcF products. Overall, these results demonstrated that the phenotypic switches in expression of Vmc gene products were independent, thereby providing a strong argument that a stochastic mutational event drives the variable expression of each product.

A hallmark of mutation-based phase-variable systems is the reversibility of the phenotypic change. Using rigorously cloned populations derived from the mck3 clone, we identified several independent examples showing conversion of any one of the three VmcF phenotypes (Fig. 4A) to both of the other phenotypes (data not shown). Proof of phase-variable switching requires the demonstration of switches in lineages representing successive progeny. To strictly establish the reversible phase variation of all three phenotypes and to gain access to the underlying mutational basis for this variation, we derived and examined multiple clonal lineages selected for variable expression of VmcF. These were selected to represent subsequent switches in populations from one phenotype to another and ultimately represented all possible switches among the three phenotypes. The three clonal lineages (I, II, and III) used in this analysis are shown in Fig. 5A. Through these lineages, we could formally establish that each phenotype defined for VmcF expression was reversible and that all three patterns were interchangeable in a phase-variable manner, as shown in Fig. 4A.

A single dinucleotide indel governs phase-variable *vmc* gene expression. Our cloned lineages comprised critical switched populations that were highly enriched in a particular VmcF phenotype. These provided a source of genomic templates that could be interrogated to determine the status of poly(TA) tracts 5' of the *vmcF* gene corresponding to each phenotype and the change of the poly(TA) regions associated with phase transitions in the lineage. We used specific primers that amplified part of the coding region and the 5' flanking region of *vmcF* (Fig. 2) to determine sequences from several phase variants, including all of those depicted in Fig. 5A. Precise and consistent differences in the poly(TA) tracts associated with each phenotype were observed (examples of sequence chromatograms of these tracts are shown in Fig. 5B). These differences were confirmed by sequencing both strands of selected variants representing each phenotype. Using the convention

previously described for *vmm* variants of *M. mycoides* subsp. *mycoides* SC PG1 (33), all VmcF⁺ variants in the lineages contained 10 TA repeats, all VmcF⁻ variants contained 12 TA repeats, and all microsectored variants contained 11 TA repeats. Because these populations represented formal lineages of progeny undergoing phase variation, these correlations were compelling in linking the length of the TA region with expression of the VmcF product and, by inference, the *vmcF* gene. The correlation of a specific poly(TA) repeat number with each phenotype was further supported by determining the sequences from several additional, individually cloned populations that did not formally represent switching lineages. For each phenotype, the numbers of TA repeats determined per clone assayed were as follows: VmcF⁺, $n = 10$ for all of the 6 clones assayed; VmcF^{u/s}, $n = 11$ for all of the 10 clones assayed; and VmcF⁻, $n = 12$ for 7 (and $n = 13$ for 1) of 8 clones assayed. To rule out other possible mechanisms involving rearrangement of *vmcE* and *vmcF* genes during these switches, amplicons linking each gene to the respective conserved flanking region of the *vmcE* to *vmcF* locus (Fig. 1) were generated and partially sequenced. In each of the multiple populations queried (Fig. 5A), representing the three variant phenotypes, the genomic configuration of the genes was unchanged.

This strict correlation allowed us to refine the parameters that were initially proposed to explain expression variants of Vmm in *M. mycoides* subsp. *mycoides* SC PG1 (33). poly(TA) tracts in the *vmm* gene of 10 and 12 TA repeats, respectively, correlated with Vmm⁺ and Vmm⁻ variants, precisely corresponding to the poly(TA) tracts observed in VmcF⁺ and VmcF⁻ variants. However, our discovery that populations generating microsectored colonies invariably contain poly(TA) tracts of 11 TA repeats suggests a more subtle level of control on *vmc* (and perhaps *vmm*) gene expression, dictated by an indel of only 1 TA repeat. The detailed immunostaining pattern within microsectored colonies indicated that the founder organism initially plated, and its early progeny, has a Vmc⁻ expression phenotype and that only later in colony formation do organisms of the Vmc⁺ phenotype appear at the periphery. We interpret this to indicate that 11 TA repeats correlate with a Vmc⁻ configuration for gene expression and that the phenomenon of microsectoring results from a high rate of switching to the Vmc⁺ configuration (presumably 10 TA repeats) for a portion of the subsequent progeny within colonies. We further suggest that the frequency of switching from 11 to 10 TA repeats may exceed that of switching from 12 to 10 TA repeats; this is consistent with the pattern seen for colonies switching directly between Vmc⁻ and Vmc⁺, where a lower frequency for indels involving 2 TA repeats would be less likely to generate mixed populations within single colonies. While this model would also predict the occurrence in Vmc⁺ colonies of variants that switch from 10 to 11 TA repeats at very high frequency, the resultant Vmc⁻ populations at the periphery of Vmc⁺ colonies may simply not be resolved by our staining method. Similarly, it may be very difficult to observe predicted high-frequency switches from 11 to 12 TA repeats in microsectored colonies, since the latter would have a Vmc⁻ phenotype, undetectable in this context. It is possible that these more subtle populations were represented in sequence chromatograms of amplicons generated from clonal populations (Fig. 5B), wherein low-amplitude signals consistent with a mixture

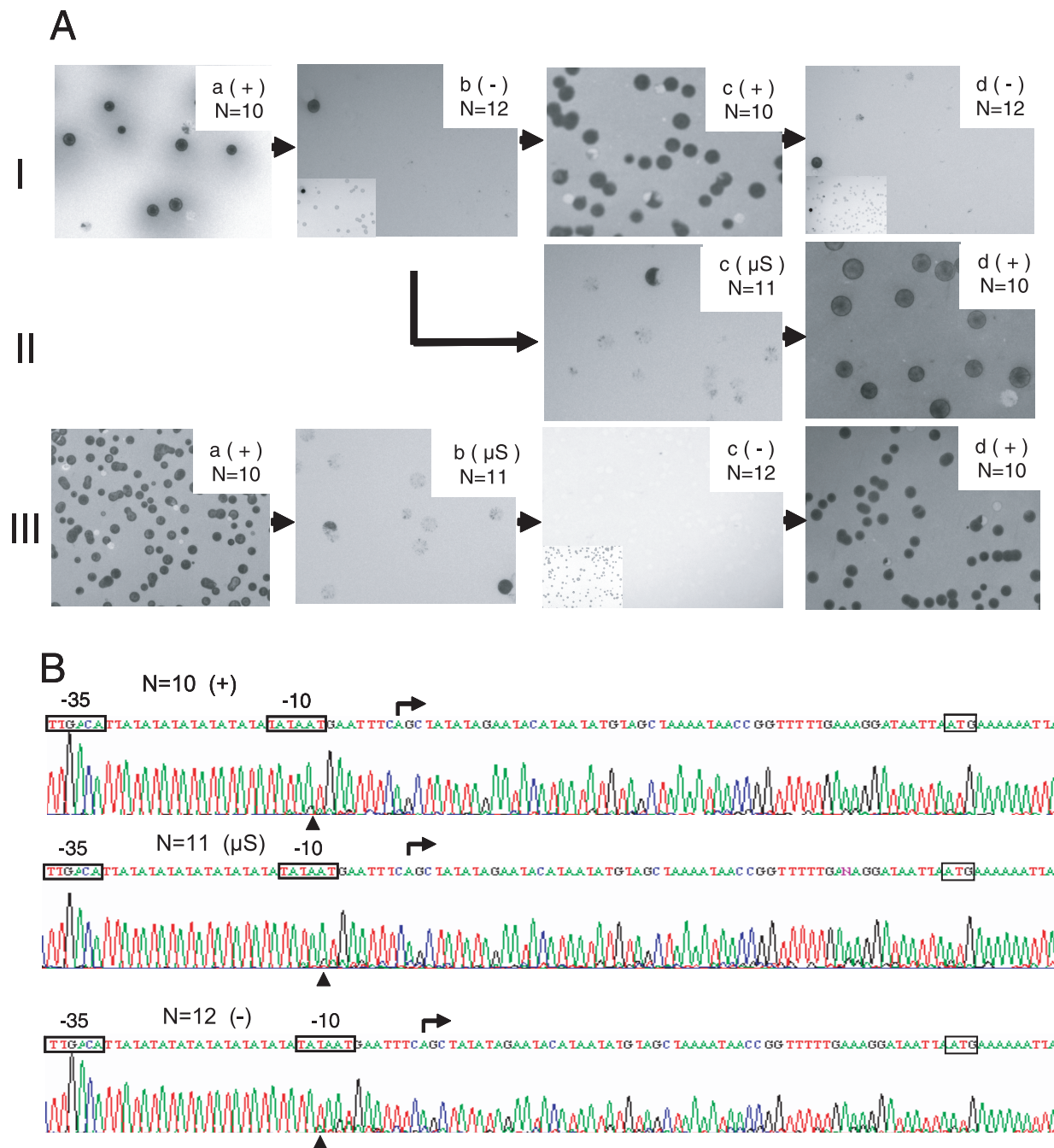


FIG. 5. Clonal lineage analysis of VmcF phase-variable expression and corresponding *vmcF* promoter region mutations. (A) Three independent lineages (I, II, and III) derived from the *M. capricolum* subsp. *capricolum* mck3 clonal population are shown. Each represents a different progression of switches in propagated progeny populations, involving the $VmcF^+$, $VmcF^-$, and $VmcF^{\mu S}$ phenotypes. Each arrow connecting panels indicates the selection, clonal purification, and monitoring of the phenotypic variant shown. All panels show colony immunoblots stained with MAb 14H2 to VmcF. Fields showing organisms of the predominant phenotype were chosen, where possible, to include minor spontaneous variants with alternative phenotypes. In panels showing the $VmcF^-$ phenotype after immunostaining, insets represent the same field counterstained with Ponceau S to visualize the colonies present. For each population, the number of TA repeats determined to be in the 5' flanking region of the *vmcF* gene is indicated (10, 11, or 12), using a previously described convention for numbering (33). (B) Upstream sequences flanking the *vmcF* gene in clonal variants with distinct VmcF phenotypes. Representative sequence chromatograms of amplicons generated from a $VmcF^+$ variant (III d), a $VmcF^{\mu S}$ variant (III b), and a $VmcF^-$ variant (III c) are shown. The variable number of TA repeats (N), between the -35 and -10 motifs (boxed), is indicated for each variant. The proposed transcription start, based on sequence similarity and position relative to those determined for the *vmm* gene (33), is indicated by arrows. The ATG start codon for VmcF translation is boxed. Arrowheads indicate the positions in the chromatograms at which minor secondary sequences appear as low-amplitude peaks following the poly(TA) tracts.

of indel mutations occurred following poly(TA) tracts (irrespective of the strand sequenced).

The pattern of poly(TA) mutations we observed and the limitation in defining the resultant subtle phenotypes precluded any meaningful assignment of switch frequencies to *vmc* gene expression. Populations scored for VmcE and VmcF expression displayed the microsectored phenotype, indicating switch frequencies higher than that quantifiable by these methods. This phenotype was less prevalent in lineages expressing variants of VmcB, which we monitored in colony immunoblot analyses with our specific PAb (data not shown). These showed positive, negative, and traditionally sectored phenotypes (Fig. 3B), all consistent with classic high-frequency phase variation systems in mycoplasmas (39, 57) but with a frequency below that of the hypervariable VmcE or VmcF. From our detailed analysis of the *vmcF* gene, we conclude that all *vmc* genes are likely to operate in a manner analogous to that of the *vmm* gene, through a spontaneous indel mutation in the poly(TA) tract that independently governs the expression of each gene. It is quite possible that more-subtle differences occur between these two systems or even between the sets of *vmc* genes in each locus of *M. capricolum* subsp. *capricolum*. For example, although the sequences 5' of *vmcF* and *vmm* coding regions are nearly identical, discrepancies occur at critical positions. The transcriptional start identified (33) for *vmm* (beginning 8 nt downstream of the -10 box with ATCT...) differs from the analogous region for *vmcF* (AGCT...), shown in Fig. 5B. In addition, the reported Shine-Dalgarno sequence of *vmm* (AGGAG) differs from that in the corresponding position in *vmcE* and *vmcF* (AGGAT), whereas this motif in each of the *vmcA* to *vmcD* genes is identical to that in *vmm*. These differences could confer properties to specific genes that affect aspects of variable expression, such as the frequencies of switching and hierarchical efficiencies of transcription or translation.

DISCUSSION

The *vmc* genes of *M. capricolum* subsp. *capricolum* present not only a repertoire of alternative surface structures for *M. capricolum* subsp. *capricolum* but also a system that permits their high-frequency phase-variable and combinatorial expression as well as structural variation. Consequently, any *M. capricolum* subsp. *capricolum* population must now be viewed as a highly complex set of cells reflecting the stochastic generation of diversity during propagation. Nevertheless, the structure of *vmc* genes and the nature of the associated mutations appear to limit this variation to patterns involving a set of stably encoded protein sequences, in contrast to the generation of nearly unlimited sequence divergence through gene conversion or other mechanisms (48). Moreover, the clustering and separate locations of *vmc* genes encoding two structurally distinct sets of lipoproteins underscore the possibility that analogous or possibly unique gene sets may occur among mycoplasmal genomes within the *M. mycoides* cluster. Only two genes, with limited regional similarity to *vmc* genes, occur in *M. mycoides* subsp. *mycoides* SC PG1, and one of these is translationally interrupted. Other *vmc* genes are unique. These gene sets therefore confer distinctive, variable cell surface architectures to the respective subspecies within this clade. Exploration of other mycoplasmas within this cluster will more fully define the

contribution of such genes to diversity within their highly conserved genomic frameworks and the degree to which specific sets of genes may reflect the known subspecies, their unique host niches, or their pathobiology. As a binary comparison between two completed genome sequences, our studies did not reveal a mechanism by which these sequences have evolved or have been mobilized to reside in their current genomic contexts. Although lateral gene transfer offers one intriguing possibility, *vmc* genes are not physically associated with integrative conjugative elements or IS elements that occur in the *M. capricolum* subsp. *capricolum* genome (3), nor does the occurrence of prolific IS element-related sequences in the *M. mycoides* subsp. *mycoides* SC PG1 locus containing *vmm* offer any concrete evidence for mobilization. It is perhaps noteworthy that the 5' flanking regions of *vmcA* and *vmm* represent the 5' boundary of their respective loci and that these two chromosomal regions first diverge within these two genes. The nature of any gene pool from which *vmc* or analogous gene sets emerged remains obscure, but the biased G+C compositions of coding and intergenic regions might signal an ultimate source outside the current chromosomal reservoirs of this phylogenetic cluster.

The diversity represented by the surface-exposed variable sequences encoded by *vmc* (and *vmc*-like) genes appears to be displayed using two highly conserved, functional sequence "platforms": a common N-terminal region for membrane translocation of the prolipoprotein and a common 5' flanking region determining phase-variable gene expression. This modular structure obscures any useful classification of these genes into paralogous or orthologous "families" defined by standard algorithms that measure sequence similarity (which are further skewed by the presence of unique repetitive coding regions of variable length). Nevertheless, it is noteworthy that reiteration of these two common motifs among *vmc* and *vmc*-like genes could provide templates for possible recombinational pathways leading to gene duplication or rearrangement over time. The individual or combined modular features of the *vmc* genes reported here were easily recognizable. Nevertheless, we consider this to be only the provisional set of variable genes in *M. capricolum* subsp. *capricolum* for which we have demonstrated both expression and the nature of variation. We observed additional genes in this genome that encode lipoproteins with features suggesting possible variation. Moreover, translation products of some but not all of these were detected among the 185 proteins that we identified in the membrane-associated proteome of *M. capricolum* subsp. *capricolum*. Examples include lipoproteins (i) with tandem peptide repeats but no 5' flanking regions containing VNTRs (e.g., MCAP_0583 and MCAP_0585, the former detected in the proteome), (ii) without tandem repeats but with a poly(TA) tract, in a context different from that of *vmc* genes (e.g., MCAP_0431 through MCAP_0433, not detected in the proteome), and (iii) with tandem repeats and a poly(TA) tract, where the repeats comprise a widely distributed family of true paralogous and orthologous sequences (e.g., MCAP_0268 and MCAP_0269, the former detected in the proteome). These observations raise the possibility that other surface lipoproteins encoded in *M. capricolum* subsp. *capricolum*, or in other subspecies of the *M. mycoides* cluster, could be subject to variation in expression or structure.

Although the functions of most mycoplasmal variable surface lipoproteins or any advantages conferred by their phase-variable expression are still largely unknown, we speculate that Vmc lipoproteins could provide both protective and opportunistic capabilities to *M. capricolum* subsp. *capricolum*. Since the first instance of such a variable gene set in mycoplasmas was reported, it has long been questioned whether such finite sets of coding sequence provide sufficient diversity for mycoplasmas to evade the immune system by variation of antigenic structure (47, 56, 59, 61). However, at least one protective effect, based on size variation, has been demonstrated for analogous proteins comprising small tandem repeats. Versions of Vlp lipoproteins of *Mycoplasma hyorhinis* or Vsa lipoproteins of *Mycoplasma pulmonis* that contain extensive regions of such repeats protect these organisms, respectively, from immunological targeting by growth-inhibiting host antibody (6) or complement (45), compared to shorter versions of these lipoproteins with fewer repeats. This protective effect may be general, since it occurs irrespective of the protein sequence of the particular Vlp or Vsa expressed and among species in divergent mycoplasmal clades. In this regard, the unusual features shared among Vlp, Vsa, and the VmcE and VmcF lipoproteins of *M. capricolum* subsp. *capricolum* may point to a unifying role as surface coat proteins in these wall-less prokaryotes. Although the actual structure of these proteins has yet to be assessed experimentally, their lack of predicted structure could reflect one of several key biological roles for intrinsically unstructured proteins (12). An additional, opportunistic role can be imagined for Vmc proteins, based on the demonstration that phase-variable adhesion-associated lipoproteins containing repeated, often coiled-coil regions occur in diverse mycoplasma species (2, 25, 51, 63, 64), although any such role in *M. capricolum* subsp. *capricolum* is entirely speculative. Despite limited understanding of variable gene families in mycoplasmas, their distinctive repertoires and capacity for variable expression suggest an overall function that is tailored by selection to the specific host or in-host niches populated by a particular species or subspecies of these obligate host parasites.

The distinctive sequences comprising Vmc proteins, and their possible counterparts within the *M. mycoides* cluster, offer potentially useful targets for diagnostic applications. Components of various serological specificities are recognized within the *M. mycoides* cluster, such as orthologs of the surface protein LppB (49), MSC_0519 and MCAP_0451 (expressed in the *M. capricolum* subsp. *capricolum* proteome), and orthologs of LppA (29), MSC_0240 and MCAP_0034 (expressed in the *M. capricolum* subsp. *capricolum* proteome), or the immunogen MSC_0397 (the ortholog MCAP_0588 is not expressed due to an authentic frameshift). In contrast, the unique sequences of vmc gene products (i) are present as combinatorial arrays in virtually any population of *M. capricolum* subsp. *capricolum*, (ii) are amplified to very high copy number per organism due to their abundance and reiterated epitopic structure, (iii) can be configured as native lipoproteins or soluble recombinant or synthetic polypeptides, adaptable to a variety of multiplexed array formats, and (iv) may be released from cells as targets spatially separated from the organisms, as is apparent in clones expressing VmcE⁺ or VmcF⁺ phenotypes that show striking zones of epitope-bearing products radiating from colony immunoblots specifically stained with MAbs (Fig. 4C and Fig. 5A,

panel Ia). Although the recognition of Vmc proteins during infection of the natural host has not been examined, their use as targets for direct identification and discrimination of this subspecies is promising.

ACKNOWLEDGMENTS

This work was supported by USDA ARS Project 1940-32000-039-00D and in part by a University of Missouri Research Board grant to M.J.C.

We thank Anja Persson, Joakim Westberg, and John Glass for providing genome sequences prior to public release and Kurt Downing and Marlyn Burks for technical assistance. For peptide synthesis and many other contributions, we thank Jeong Hyok Im, to whose memory this paper is dedicated.

REFERENCES

1. Baseman, J. B., M. Lange, N. L. Criscimagna, J. A. Giron, and C. A. Thomas. 1995. Interplay between mycoplasmas and host target cells. *Microb. Pathog.* **19**:105–116.
2. Boesen, T., J. Emmersen, A. Baczyńska, S. Birkelund, and G. Christiansen. 2004. The vaa locus of *Mycoplasma hominis* contains a divergent genetic islet encoding a putative membrane protein. *BMC Microbiol.* **4**:37.
3. Calcutt, M. J., M. S. Lewis, and K. S. Wise. 2002. Molecular genetic analysis of ICEF, an integrative conjugal element that is present as a repetitive sequence in the chromosome of *Mycoplasma fermentans* PG18. *J. Bacteriol.* **184**:6929–6941.
4. Chambaud, I., R. Heilig, S. Ferris, V. Barbe, D. Samson, F. Galisson, I. Moszer, K. Dybvig, H. Wroblewski, A. Viari, E. Rocha, and A. Blanchard. 2001. The complete genome sequence of the murine respiratory pathogen *Mycoplasma pulmonis*. *Nucleic Acids Res.* **29**:2145–2153.
5. Citti, C., G. F. Browning, and R. Rosengarten. 2005. Phenotypic diversity and cell invasion in host subversion by pathogenic mycoplasmas, p. 439–483. In A. Blanchard and G. Browning (ed.), *Mycoplasmas, molecular biology pathogenicity and strategies for control*. The Cromwell Press, Trowbridge, Wiltshire, United Kingdom.
6. Citti, C., M. F. Kim, and K. S. Wise. 1997. Elongated versions of Vlp surface lipoproteins protect *Mycoplasma hyorhinis* escape variants from growth-inhibiting host antibodies. *Infect. Immun.* **65**:1773–1785.
7. Cleavinger, C. M., M. F. Kim, J. H. Im, and K. S. Wise. 1995. Identification of mycoplasma membrane proteins by systematic *TnphoA* mutagenesis of a recombinant library. *Mol. Microbiol.* **18**:283–293.
8. Cottew, G. S., and F. R. Yeats. 1978. Subdivision of *Mycoplasma mycoides* subsp. *mycoides* from cattle and goats into two types. *Aust. Vet. J.* **54**:293–296.
9. Dahl, C. E., and J. S. Dahl. 1984. Phospholipids as acyl donors to membrane proteins of *Mycoplasma capricolum*. *J. Biol. Chem.* **259**:10771–10776.
10. Davis, G., and W. C. S. Read. 1968. A modification of the growth-inhibition test and its use for detecting *Mycoplasma mycoides* var. *mycoides*. *J. Hyg.* **66**:319–324.
11. Davis, K. L., and K. S. Wise. 2002. Site-specific proteolysis of the MALP-404 lipoprotein determines the release of a soluble selective lipoprotein-associated motif-containing fragment and alteration of the surface phenotype of *Mycoplasma fermentans*. *Infect. Immun.* **70**:1129–1135.
12. Dyson, H. J., and P. E. Wright. 2005. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **6**:197–208.
13. Eng, J. K., A. L. McCormack, and J. R. I. Yates. 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**:976–989.
14. Erno, H. 1983. *Mycoplasma* related to *Mycoplasma mycoides* subsp. *mycoides*, p. 33–39. In S. H. Hall (ed.), *The diagnosis of contagious bovine pleuropneumonia and other infections with Mycoplasma mycoides* subsp. *mycoides*. Commission of the European Communities, Brussels, Belgium.
15. Ewing, B., and P. Green. 1998. Base-calling of automated sequencer traces using Phred. II. Error probabilities. *Genome Res.* **8**:186–194.
16. Freundt, E. A., and D. G. Edward. 1979. Classification and taxonomy, p. 1–41. In M. F. Barile and S. Razin (ed.), *The mycoplasmas*. Academic Press, Inc., New York, N.Y.
17. Gatlin, C. L., G. R. Kleemann, L. G. Hays, A. J. Link, and J. R. I. Yates. 1998. Protein identification at the low femtomole level from silver-stained gels using a new fritless electrospray interface for liquid chromatography—microspray and nanospray mass spectrometry. *Anal. Biochem.* **263**:93–101.
18. Glass, J. I., N. Assad-Garcia, N. Alperovich, S. Yoosheph, M. R. Lewis, M. Maruf, C. A. Hutchison III, H. O. Smith, and J. C. Venter. 2006. Essential genes of a minimal bacterium. *Proc. Natl. Acad. Sci. USA* **103**:425–430.
19. Glew, M. D., L. Papazisi, F. Poumarat, D. Bergeron, R. Rosengarten, and C. Citti. 2000. Characterization of a multigene family undergoing high-frequency DNA rearrangements and coding for abundant variable surface proteins in *Mycoplasma agalactiae*. *Infect. Immun.* **68**:4539–4548.

20. Gordon, D., C. Abajian, and P. Green. 1998. Consed: a graphical tool for sequence finishing. *Genome Res.* **8**:195–202.
21. Johansson, K.-E., and B. Pettersson. 2002. Taxonomy of mollicutes, p. 1–29. In S. Razin and R. Herrmann (ed.), *Molecular biology and pathogenicity of mycoplasmas*. Kluwer Academic/Plenum Publishers, London, United Kingdom.
22. Jones, D., F. Rainey, K. Schliefer, and J. G. Tully. 2006. *Bergey's manual of systematic bacteriology*. Springer-Verlag, New York, N.Y.
23. Lartigue, C., A. Blanchard, J. Renaudin, F. Thiaucourt, and P. Sirand-Pugnet. 2003. Host specificity of mollicutes *oriC* plasmids: functional analysis of replication origin. *Nucleic Acids Res.* **31**:6610–6618.
24. Leigh, S. A. 2000. Functional analysis of the *Mycoplasma fermentans* P29 adhesin. Ph.D. dissertation. University of Missouri-Columbia, Columbia, Mo.
25. Leigh, S. A., and K. S. Wise. 2002. Identification and functional mapping of the *Mycoplasma fermentans* P29 adhesin. *Infect. Immun.* **70**:4925–4935.
26. Lo, S.-C., M. S. Dawson, D. M. Wong, P. B. Newton III, M. A. Sonoda, W. F. Engler, R. Y. H. Wang, J. W. K. Shih, H. J. Alter, and D. J. Wear. 1989. Identification of *Mycoplasma cognitus* infection in patients with AIDS: an immunohistochemical, *in situ* hybridization and ultrastructural study. *Am. J. Trop. Med. Hyg.* **41**:601–616.
27. March, J. B., C. D. Jepson, J. R. Clark, M. Totsika, and M. J. Calcutt. 2006. Phage library screening for the rapid identification and *in vivo* testing of candidate genes for a DNA vaccine against *Mycoplasma mycoides* subsp. *mycoides* small colony biotype. *Infect. Immun.* **74**:167–174.
28. McAuliffe, L., R. J. Ellis, J. R. Lawes, R. D. Ayling, and R. A. J. Nicholas. 2005. 16S rDNA PCR and denaturing gradient gel electrophoresis: a single generic test for detecting and differentiating *Mycoplasma* species. *J. Med. Microbiol.* **54**:731–739.
29. Monnerat, M.-P., F. Thiaucourt, J. Nicolet, and J. Frey. 1999. Comparative analysis of the *lppA* locus in *Mycoplasma capricolum* subsp. *capricolum* and *Mycoplasma capricolum* subsp. *capripneumoniae*. *Vet. Microbiol.* **69**:157–172.
30. Nesvizhskii, A. I., A. Keller, E. Kolker, and R. Aebersold. 2003. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **75**:4646–4658.
31. Oba, T., Y. Andachi, A. Muto, and S. Osawa. 1991. CGG: an unassigned or nonsense codon in *Mycoplasma capricolum*. *Proc. Natl. Acad. Sci. USA* **88**:921–925.
32. Papazisi, L., T. S. Gorton, G. Kutish, P. F. Markham, G. F. Browning, D. K. Nguyen, S. Swartzell, A. Madan, G. Mahairas, and S. J. Geary. 2003. The complete genome sequence of the avian pathogen *Mycoplasma gallisepticum* strain R_{low}. *Microbiology* **149**:2307–2316.
33. Persson, A., K. Jacobsson, L. Frykberg, K.-E. Johansson, and F. Poumarat. 2002. Variable surface protein Vmm of *Mycoplasma mycoides* subsp. *mycoides* small colony type. *J. Bacteriol.* **184**:3712–3722.
34. Pettersson, B., T. Leitner, M. Ronaghi, G. Bölske, M. Uhlén, and K. E. Johansson. 1996. Phylogeny of the *Mycoplasma mycoides* cluster as determined by sequence analysis of the 16S rRNA genes from the two rRNA operons. *J. Bacteriol.* **178**:4131–4142.
35. Razin, S., D. Yogeve, and Y. Naot. 1998. Molecular biology and pathogenicity of mycoplasmas. *Microbiol. Mol. Biol. Rev.* **62**:1094–1156.
36. Roach, J. C., C. Boysen, K. Wang, and L. Hood. 1995. Pairwise end sequencing: a unified approach to genomic mapping and sequencing. *Genomics* **26**:345–353.
37. Romano, N., M. H. Shirvan, and S. Rottem. 1986. Changes in membrane lipid composition of *Mycoplasma capricolum* affect the cell volume. *J. Bacteriol.* **167**:1089–1091.
38. Rosengarten, R., and K. S. Wise. 1990. Phenotypic switching in mycoplasmas: phase variation of diverse surface lipoproteins. *Science* **247**:315–318.
39. Rosengarten, R., and K. S. Wise. 1991. The Vlp system of *Mycoplasma hyorhinis*: combinatorial expression of distinct size variant lipoproteins generating high-frequency surface antigenic variation. *J. Bacteriol.* **173**:4782–4793.
40. Rosengarten, R., and D. Yogeve. 1996. Variant colony surface antigenic phenotypes within mycoplasma strain populations: implications for species identification and strain standardization. *J. Clin. Microbiol.* **34**:149–158.
41. Rowen, L., and B. F. Koop. 1994. *Zen and the art of the large-scale genomic sequencing*. Academic Press Limited, London, United Kingdom.
42. Rurangirwa, F. R., P. S. Shompole, A. N. Wambugu, and T. C. McGuire. 2000. Monoclonal antibody differentiation of *Mycoplasma mycoides* subsp. *mycoides* small-colony strains causing contagious bovine pleuropneumonia from less important large-colony strains. *Clin. Diagn. Lab. Immunol.* **7**:519–521.
43. Sasaki, Y., J. Ishikawa, A. Yamashita, K. Oshima, T. Kenri, K. Furuya, C. Yoshino, A. Horino, T. Shiba, T. Sasaki, and M. Hattori. 2002. The complete genomic sequence of *Mycoplasma penetrans*, an intracellular bacterial pathogen in humans. *Nucleic Acids Res.* **30**:5293–5300.
44. Seto, S., and M. Miyata. 1999. Partitioning, movement, and positioning of nucleoids in *Mycoplasma capricolum*. *J. Bacteriol.* **181**:6073–6080.
45. Simmons, W. L., A. M. Denison, and K. Dybvig. 2004. Resistance of *Mycoplasma pulmonis* to complement lysis is dependent on the number of Vsa tandem repeats: shield hypothesis. *Infect. Immun.* **72**:6846–6851.
46. Theiss, P. M., M. F. Kim, and K. S. Wise. 1993. Differential protein expression and surface presentation generate high-frequency antigenic variation in *Mycoplasma fermentans*. *Infect. Immun.* **61**:5123–5128.
47. Van der Woude, M. W. 2006. Re-examining the role and random nature of phase variation. *FEMS Microbiol. Lett.* **254**:190–197.
48. Van der Woude, M. W., and A. J. Baumler. 2004. Phase and antigenic variation in bacteria. *Clin. Microbiol. Rev.* **17**:581–611.
49. Vilei, E. M., E. M. Abdo, J. Nicolet, A. Botelho, R. Gonçalves, and J. Frey. 2000. Genomic and antigenic differences between the European and African/Australian clusters of *Mycoplasma mycoides* subsp. *mycoides* SC. *Microbiology* **146**:477–486.
50. Volokhov, D. V., J. George, S. X. Liu, P. Ikonomi, C. Anderson, and V. Chizhikov. 2006. Sequencing of the intergenic 16S-23S rRNA spacer (ITS) region of *Mollicutes* species and their identification using microarray-based assay and DNA sequencing. *Appl. Microbiol. Biotechnol.* [Online.] doi: 10.1007/s00253-005-0280-7.
51. Washburn, L. R., K. E. Weaver, E. J. Weaver, W. Donelan, and S. Al-Shebul. 1998. Molecular characterization of *Mycoplasma arthritidis* variable surface protein MAA2. *Infect. Immun.* **66**:2576–2586.
52. Weisburg, W. G., J. G. Tully, D. L. Rose, J. P. Petzel, H. Oyaizu, D. Yang, L. Mandelco, J. Sechrest, T. G. Lawrence, J. Van Etten, J. Maniloff, and C. R. Woese. 1989. A phylogenetic analysis of the mycoplasmas: basis for their classification. *J. Bacteriol.* **171**:6455–6467.
53. Westberg, J., A. Persson, A. Holmberg, A. Goesmann, J. Lundberg, K.-E. Johansson, B. Pettersson, and M. Uhlén. 2004. The genome sequence of *Mycoplasma mycoides* subsp. *mycoides* SC type strain PG1^T, the causative agent of contagious bovine pleuropneumonia (CBPP). *Genome Res.* **14**:221–227.
54. Westberg, J., A. Persson, B. Pettersson, M. Uhlén, and K.-E. Johansson. 2002. ISMmy1, a novel insertion sequence of *Mycoplasma mycoides* subsp. *mycoides* small colony type. *FEMS Microbiol. Lett.* **208**:207–213.
55. Winner, F., R. Rosengarten, and C. Citti. 2000. *In vitro* cell invasion of *Mycoplasma gallisepticum*. *Infect. Immun.* **68**:4238–4244.
56. Wise, K. S. 1993. Adaptive surface variation in mycoplasmas. *Trends Microbiol.* **1**:59–63.
57. Wise, K. S., M. F. Kim, and R. Watson-McKown. 1995. Variant membrane proteins, p. 227–241. In S. Razin and J. G. Tully (ed.), *Molecular and diagnostic procedures in mycoplasma*. Academic Press, New York, N.Y.
58. Wise, K. S., and R. K. Watson. 1983. *Mycoplasma hyorhinis* GDL surface protein antigen p120 defined by monoclonal antibody. *Infect. Immun.* **41**:1332–1339.
59. Wise, K. S., D. Yogeve, and R. Rosengarten. 1992. Antigenic variation, p. 473–490. In J. Maniloff, R. N. McElhaney, L. R. Finch, and J. B. Baseman (ed.), *Mycoplasmas: molecular biology and pathogenesis*. American Society for Microbiology, Washington, D.C.
60. Yamao, F., A. Muto, Y. Kawauchi, M. Iwami, S. Iwagami, Y. Azumi, and S. Osawa. 1985. UGA is read as tryptophan in *Mycoplasma capricolum*. *Proc. Natl. Acad. Sci. USA* **82**:2306–2309.
61. Yogeve, D., G. F. Browning, and K. S. Wise. 2002. Genetic mechanisms of surface variation, p. 417–443. In S. Razin and R. Herrmann (ed.), *Molecular biology and pathogenicity of mycoplasmas*. Kluwer Academic/Plenum Publishers, London, United Kingdom.
62. Yogeve, D., R. Rosengarten, R. Watson-McKown, and K. S. Wise. 1991. Molecular basis of *Mycoplasma* surface antigenic variation: a novel set of divergent genes undergo spontaneous mutation of periodic coding regions and 5' regulatory sequences. *EMBO J.* **10**:4069–4079.
63. Zhang, Q., and K. S. Wise. 1996. Molecular basis of size and antigenic variation of a *Mycoplasma hominis* adhesin encoded by divergent *vaa* genes. *Infect. Immun.* **64**:2737–2744.
64. Zhang, Q., and K. S. Wise. 1997. Localized reversible frameshift mutation in an adhesin gene confers a phase-variable adherence phenotype in mycoplasma. *Mol. Microbiol.* **25**:859–869.
65. Zhang, Q., and K. S. Wise. 2001. Coupled phase-variable expression and epitope masking of selective surface lipoproteins increase surface phenotypic diversity in *Mycoplasma hominis*. *Infect. Immun.* **69**:5177–5181.