IAI Journals.ASM.org

Extensive Variation and Rapid Shift of the MG192 Sequence in *Mycoplasma genitalium* Strains from Patients with Chronic Infection

Liang Ma,[a]* Miriam Mancuso,[a] James A. Williams,[b] Barbara Van Der Pol,[c] J. Dennis Fortenberry,[d] Qiuyao Jia,[a] Leann Myers,[e] David H. Martin[a]

Section of Infectious Diseases, Department of Medicine, Louisiana State University Health Sciences Center, New Orleans, Louisiana, USA[a]; Division of Infectious Diseases, Department of Medicine, Indiana University School of Medicine, Indianapolis, Indiana, USA[b]; Division of Infectious Diseases, Department of Medicine, University of Alabama at Birmingham, Birmingham, Alabama, USA[c]; Section of Adolescent Medicine, Department of Pediatrics, Indiana University School of Medicine, Indianapolis, Indiana, USA[d]; Department of Biostatistics and Bioinformatics, Tulane School of Public Health and Tropical Medicine, New Orleans, Louisiana, USA[e]

***Mycoplasma genitalium* causes persistent urogenital tract infection in humans. Antigenic variation of the protein encoded by the MG192 gene has been proposed as one of the mechanisms for persistence. The aims of this study were to determine MG192 sequence variation in patients with chronic *M. genitalium* infection and to analyze the sequence structural features of the MG192 gene and its encoded protein. Urogenital specimens were obtained from 13 patients who were followed for 10 days to 14 months. The variable region of the MG192 gene was PCR amplified, subcloned into plasmids, and sequenced. Sequence analysis of 220 plasmid clones yielded 97 unique MG192 variant sequences. MG192 sequence shift was identified between sequential specimens from all but one patient. Despite great variation of the MG192 gene among and within clinical specimens from different patients, MG192 sequences were more related within *M. genitalium* specimens from an individual patient than between patients. The MG192 variable region consisted of 11 discrete subvariable regions with different degrees of variability. Analysis of the two most variable regions (V4 and V6) in five sequential specimens from one patient showed that sequence changes increased over time and that most sequences were present at only one time point, suggesting immune selection. Topology analysis of the deduced MG192 protein predicted a surface-exposed membrane protein. Extensive variation of the MG192 sequence may not only change the antigenicity of the protein to allow immune evasion but also alter the mobility and adhesion ability of the organism to adapt to diverse host microenvironments, thus facilitating persistent infection.**

As a sexually transmitted human pathogen, *Mycoplasma genitalium* causes nongonococcal urethritis (NGU) in men and is associated with genital tract inflammatory diseases in women, including endometritis, cervicitis, pelvic inflammatory disease, and tubal factor infertility (reviewed in reference 1). Additionally, there are increasing numbers of studies suggesting that *M. genitalium* increases the risk of HIV-1 acquisition and/or transmission (2–4). Like other pathogenic mycoplasmas, *M. genitalium* is capable of causing chronic infections, as has been documented in cultured human endocervical epithelial cells (5), in animal models (6, 7), in men with NGU (8, 9), and in women with cervicitis (10). The mechanisms for persistence remain poorly understood.

The MgPa operon in the *M. genitalium* genome encodes three proteins, MG190 (*mgpA*), MG191 (*mgpB* or P140), and MG192 (*mgpC* or P110). The latter two are the known major adhesion proteins located on the surface of the terminal structure of *M. genitalium*, which plays a major role in the attachment of the organism to host epithelial cells (11, 12). Both proteins are highly antigenic and capable of eliciting strong antibody responses in *M. genitalium*-infected patients and experimentally infected animals (6, 12–14), suggesting an important role in pathogenesis. According to the genome sequence of *M. genitalium*, there is a single expression site for the MgPa operon while there are nine repetitive elements in the form of truncated copies of the MG191 and MG192 genes dispersed throughout the genome, which are designated MgPa repeats or MgPar sequences (15–19). It is believed that the MgPar sequences are not translated unless they are translocated into the expression site (18–20). Because *M. genitalium* is haploid, a single organism can express only a single MG191-MG192 isoform at a given time. Changing the gene sequence at the expression site results in an organism that has a different MG191-MG192 protein on its surface. The mechanism of such changes involves gene crossover and possibly also gene conversion (18, 21). It has been hypothesized that this recombination mechanism generates antigenic variation, allowing *M. genitalium* to evade the host immune response and to adapt to diverse host microenvironments, thus establishing persistent infection.

We have been particularly interested in studying the MG192 gene. Our previous studies showed that in the *M. genitalium* type strain G37, the MG192 sequence changes during *in vitro* passage as a result of recombination with MgPar sequences (18). Subsequent studies of sequential urine specimens from two *M. genitalium*-infected men with NGU identified extensive variation and rapid shift of the MG192 sequence over a 10- to 11-day follow-up period, with all sequence changes explained by recombination with MgPar sequences (18, 20). Similar MG192 variations also were

Received 28 November 2013 **Returned for modification** 18 December 2013
Accepted 23 December 2013

Published ahead of print 6 January 2014

Editor: R. P. Morrison

Address correspondence to Liang Ma, LMa1@lsuhsc.edu.

* Present address: Liang Ma, Critical Care Medicine Department, Clinical Center, National Institutes of Health, Bethesda, Maryland, USA.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/IAI.01526-13.

doi:10.1128/IAI.01526-13

TABLE 1 MG192 variants and strain typing of *M. genitalium* clinical specimens

Region and patient no.[a]	Specimen no.	Date (mo/day/yr)	MG192 sequence		Genotype	
			No. of clones sequenced	Unique sequence variant(s)[b]	MG309 STR no.	MG191 type no.[c]
LA						
64	64.0	6/10/2002	14	64.0a	10, 11, 12	4
	64.1	6/21/2002	9	64.1b	10, 11, 12	4
111	111.0	8/5/2002	5	111.0a, 111.0b, 111.0c*	13	51
	111.2	9/10/2002	5	111.2d*, 111.2e, 111.2f	13	51
137	137.0	10/30/2002	7	137.0a*	9	8
	137.2	11/25/2002	5	137.1b*	9	8
168	168.0	11/18/2002	7	168.0a*, 168.0b, 168.0c$, 168.0d	10	44
	168.1	12/3/2002	5	168.1e$, 168.1f, 168.1g*	10	44
199	199.0	1/28/2003	10	199.0a, 199.0b, 199.0c	10	4
	199.1	2/7/2003	18	199.1d, 199.1e, 199.1f	10	4
IN						
31	413	4/30/2001	7	31.0a, 31.0b, 31.0c	16	61
	688	1/14/2002	6	31.3d, 31.3e	16	61
61	269	11/20/2000	8	61.0a	17	2
	420	5/7/2001	6	61.2b, 61.2c, 61.2d, 61.2e, 61.2f	17	2
105	711	1/31/2002	6	105.0a	13	62
	952	9/30/2002	6	105.2b, 105.2c, 105.2d, 105.2e, 105.2f	13	62
126	126.0	10/23/2001	10	126.0a, 126.0b, 126.0c, 126.0d	16	4
	126.1	1/15/2002	10	126.1e, 126.1f, 126.1g, 126.1h, 126.1i, 126.1j, 126.1k	16	4
	126.2	6/26/2002	10	126.2l, 126.2 m, 126.2n, 126.2o, 126.2p	16	4
	126.3	9/17/2002	10	126.3q, 126.3r, 126.3s, 126.3t, 126.3u, 126.3v, 126.3w, 126.3x	16	4
	126.4	12/3/2002	10	126.4.1, 126.4.2, 126.4.3, 126.4.4, 126.4.5, 126.4.6, 126.4.7, 126.4.8, 126.4.9	16	4
136	961	10/7/2002	5	136.0a, 136.0b, 136.0c, 136.0d	13	63
	1050	12/30/2002	6	136.1e, 136.1f, 136.1g, 136.1h, 136.1i	13	63
165	982	10/22/2002	6	165.0a, 165.0b, 165.0c, 165.0d	12	4
	1261	7/8/2003	5	165.2e, 165.2f, 165.2g	12	4
172	1520	1/26/2004	7	172.0a	8	8
	2391	3/21/2005	6	172.2b, 172.2c, 172.2d, 172.2e, 172.2f	8	8
325	2300	2/11/2005	6	325.0a, 325.0b, 325.0c, 325.0d, 325.0e	10	8
	2769	7/29/2005	7	325.2f	10	8

[a] LA, Louisiana; IN, Indiana.
[b] Sequences identical in specimens from the same patients are indicated by the same symbols (* or $).
[c] According to the numbering systems of Hjorth et al. (27) and Cazanave et al. (47).

observed *in vitro* by other investigators in the G37 strain as well as in sequential cervical specimens from a woman infected by *M. genitalium* (21). In our recent studies of the complete MgPa operon in 13 *M. genitalium* axenic isolates, both MG192 and MG191 genes showed extensive interstrain variation (20). However, the intrastrain variation of the MG192 or MG191 gene (observed in only 5 of the 13 axenic isolates) appeared to be less extensive than that reported in clinical specimens, which is expected since all of these isolates except for one had been cloned by standard filtration or limiting dilution and passed *in vitro* a few times. Thus, the isolates used in the *in vitro* sequencing studies probably do not represent all sequence variants present in the original clinical specimens. These studies show that sequence variation of the MG192 and MG191 genes occurs slowly during serial *in vitro* passage, and thus analysis of cultured clinical specimens may not reflect the *in vivo* events. So far, *in vivo* studies of the MG192 or MG191 sequence variation have been limited to clinical specimens from the three patients described above. There is a lack of information about the relationships of MG192 variants among different *M. genitalium* strains.

The primary goal of this study was to determine the MG192 sequence variation in a collection of 29 clinical specimens from 13 patients with chronic *M. genitalium* infection. We analyzed the dynamic changes and phylogenetic relationships of MG192 variants in these specimens over time and explored the MG192 sequence structural features of these changes.

(Part of this work was presented at the 108th General Meeting of the American Society of Microbiology, Boston, MA, 1 to 5 June 2008.)

MATERIALS AND METHODS

***M. genitalium* clinical specimens.** Two to five sequential urogenital specimens were obtained from 13 *M. genitalium*-infected patients who were followed for as few as 10 days and up to 14 months (Table 1). The five study subjects from New Orleans, LA, were men with symptomatic NGU who attended an urban sexually transmitted disease (STD) clinic and were treated with doxycycline at the time the initial specimen included in this study was obtained (22, 23). As this study demonstrated, doxycycline is usually ineffective for the treatment of *M. genitalium* infections. Archived specimens from Indianapolis, IN, were from eight subjects from a cohort of sexually active adolescent girls who were enrolled in a long-term study

of risk factors for incident sexually transmitted infections (STIs). They were not treated with antibiotics during the course of this study (24). Genomic DNA in all specimens was extracted by use of a High Pure PCR Template Preparation Kit (Roche Diagnostic Corporation, Indianapolis, IN) as previously described (22). Informed consent was obtained from patients, and study protocols were approved by the Louisiana State University Health Sciences Center and the Indiana University School of Medicine Institutional Review Boards.

***M. genitalium* strain typing.** To differentiate *M. genitalium* strains among the clinical specimens and to investigate the possibility of coinfection with two or more strains, all specimens were subjected to genotyping based on the variable numbers of short tandem repeats (STRs) in the putative lipoprotein gene MG309 (25, 26) and single-nucleotide polymorphisms (SNPs) in the MG191 conserved AB region (27). Both loci are present in a single copy in the genome and have been shown to provide excellent discriminatory power for *M. genitalium* strain typing.

PCR amplification and sequencing of the MG192 gene. From each specimen, we amplified the entire variable region of the MG192 gene, designated the JKLM region (approximately 1.4 kb). To obtain DNA in sufficient quality and quantity for sequence analysis, we performed a primary PCR for all samples and a secondary nested PCR for those samples for which the initial PCR products were not visible or were very faint on agarose gels. All primers used have been reported elsewhere (18, 28), with primers 5346F and 227567R for the primary PCR and primers MG192A plus 227529R for the nested PCR. These primers are located in MG192 conserved regions with no homology to any of the MgPar sequences, ensuring that only the MG192 gene would be amplified. The primary PCR was performed with AmpliTaq Gold DNA polymerase (Applied Biosystems) and a touchdown protocol as described previously (29). The nested PCR was performed with high-fidelity *Pfu* DNA polymerase (Stratagene). PCR products were subcloned into the pCR2.1-Topo vector by use of a TOPO TA cloning kit (Invitrogen, Carlsbad, CA).

To rule out the possibility of PCR-related artifacts, nine specimens (including two from patient 64, two from patient 199, and five from patient 126) were amplified in two independent PCRs, followed by separate cloning and sequencing as described previously (18, 20, 29). For each of these specimens, a total of 9 to 18 clones were obtained (Table 1). Significant differences in the distribution of sequence variants between different PCR runs were not found (18, 20, 29). Therefore, all other specimens were each amplified in only one PCR followed by one cloning experiment, which yielded a maximum of 5 to 9 clones per specimen (Table 1).

DNA sequencing and sequence analysis. DNA sequencing was carried out commercially by Macrogen, Inc. (Seoul, South Korea), using the dideoxy chain termination reaction method. Nucleotide sequences were analyzed by use of the Sequencher software (version 4.10.1; Gene Codes Corporation, Ann Arbor, MI) and MacVector, version 12.6 (MacVector, Inc.). A "sequence variant" represents a unique sequence present in one or more plasmid clones obtained from the specimen studied. Phylogenetic analysis was performed using MEGA5 (30). Membrane topology analysis was conducted using the TMpred, TMHMM, TOPCONS, TopPred, HMMTOP, and SCAMPI programs, each of which is available from the World Wide Web. Antigenicity was predicted *in silico* using the protein analysis tools of the MacVector software. Pearson's correlation coefficient was used to assess the association between the time of sample collection and the sequence changes or diversity scores. A significant P value means that as the time increases, the sequence changes increase. With an n of 5 (time points), only very strong correlations are statistically significant.

Nucleotide sequence accession numbers. Nucleotide sequences of all unique MG192 variants identified in this study have been deposited in the GenBank under accession numbers JX857881 through JX857971.

RESULTS

***M. genitalium* strain typing.** As shown in Table 1, the genotypes at MG309 STRs and MG191 SNPs differed among the specimens from the 13 patients studied. The sequential specimens from each patient showed identical genotypes at both MG309 STRs and MG191 SNPs, thus indicating that each patient was infected with a single *M. genitalium* strain.

Extensive variation of MG192 among and within clinical strains. We sequenced a total of 220 plasmid clones for 29 clinical specimens (5 to 18 clones/specimen) from 13 patients and obtained 97 unique MG192 variant sequences. The variant sequences from two patients (patients 64 and 199) were described in our previous reports (18, 20). Only eight specimens showed homogenous sequences (identical sequences from multiple plasmid clones) while all others showed heterogeneous sequences with two to nine variant sequences per specimen (Table 1). No variants were shared between any two patients. Compared to the G37 type strain MG192 sequence, all variants contained apparent base substitutions, insertions, and/or deletions (8.6 to 15.8% difference), which resulted in significant changes in predicted amino acid sequences (4.2 to 11.7% difference). Despite this sequence difference, the intact reading frame was preserved in all variants.

The MG192 variable region is composed of 11 discrete subvariable regions with different variabilities. Alignment of all MG192 variant sequences revealed 11 discrete subvariable regions (designated V1 to V11), each flanked by highly conserved regions. These subvariable regions exhibited different degrees of sequence variation within and among clinical specimens, with V1 and V9 being the least variable and V4 and V6 the most variable (Fig. 1 and Table 2). The V10 region consisted primarily of variation in the size of the polyserine tract as a result of changes in the copy number of the AGT triplet repeats (29).

Transmembrane topology prediction of the deduced MG192 protein. We used six computational programs to predict transmembrane topology for the deduced full-length MG192 protein sequence in the type strain G37. Two transmembrane helices (TM1 and TM3) were consistently recognized by all six programs while another transmembrane helix (TM2) was recognized by only two programs (TMpred and TopPred) (Fig. 1 and Table 3). The central region (including the entire variable region) was predicted to be located outside the cell membrane, and both termini were predicted to be inside the membrane by all three algorithms capable of predicting surface orientation (TMHMM, TOPCONS, and SCAMPI).

Prediction of antigenic sites of the deduced MG192 protein. According to *in silico* antigenicity prediction using three commonly used methods, antigenic sites were present in all of the 11 subvariable regions of the deduced MG192 protein by at least two prediction methods (see Fig. S1 in the supplemental material). The antigenicity of these regions was further supported by the high surface probability predicted in all regions except for the V1 region (see Fig. S1 in the supplemental material). There were six subvariable regions (including V2, V3, V4, V5, V8, and V11) that were consistently shown to be antigenic by all three prediction methods and with high surface probability. Of note is the observation that the most variable region V4 showed strong antigenicity and high surface probability by all prediction methods.

MG192 sequence changes over time. Except for the two sequential specimens from one patient (number 137), which showed an identical sequence, two or more MG192 variants were detected from at least one time point in the other 12 patients (Table 1). Each MG192 variant from the 12 patients was detected at only one time point except for one variant in patient 111 and

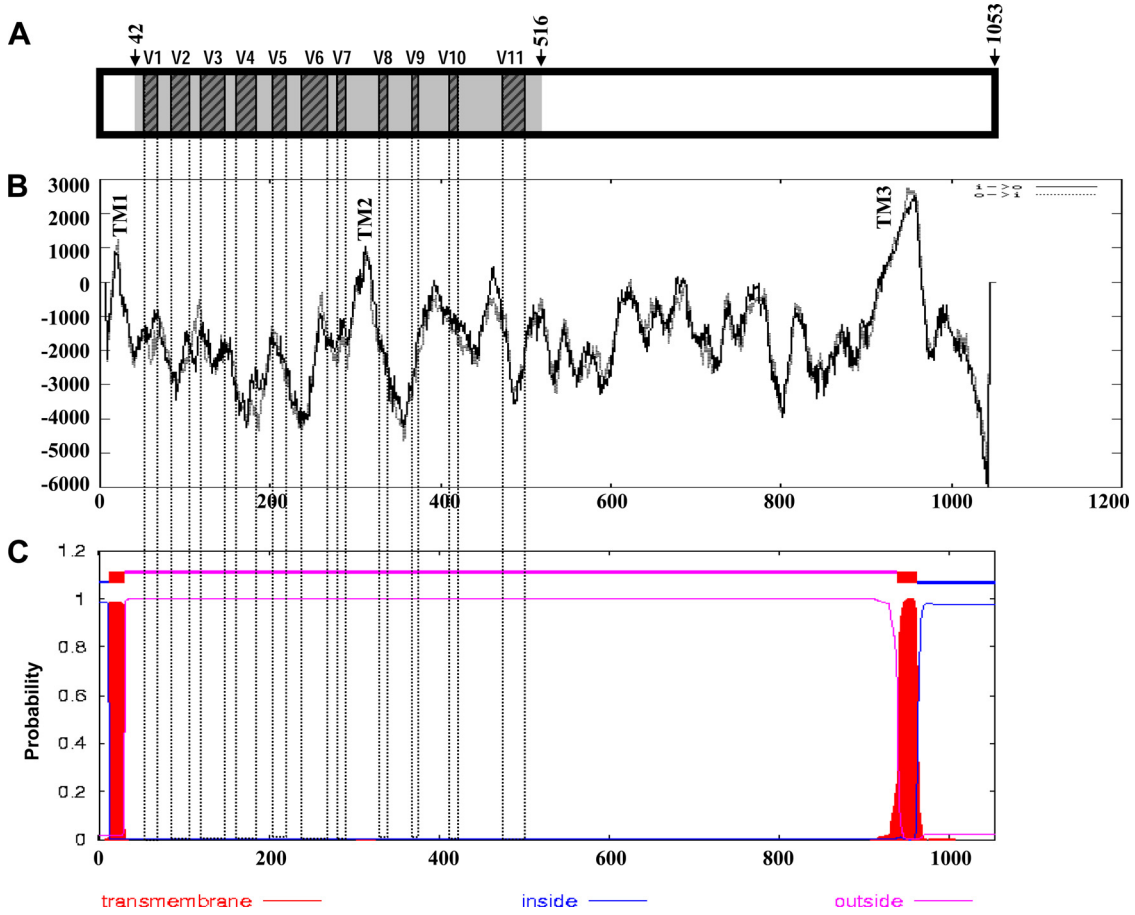

FIG 1 Sequence architecture of the MG192 protein. (A) Schematic drawing of the MG192 protein sequence based on the type strain G37. The region highlighted in gray (residues 42 to 516) represents the variable region, which is further divided into 11 subvariable regions, named V1 through V11 (indicated by hatched boxes). The degree of variation of each subvariable region is shown in Table 2. (B) Transmembrane prediction by the TMpred program. Positive values in the graph depict the probability of transmembrane helices. Three predicted transmembrane regions (TM1, TM2, and TM3) are indicated. (C) Transmembrane prediction by the TMHMM program. Red bars indicate transmembrane domains, blue lines indicate intracellular helices, and magenta lines indicate extracellular helices. Dotted vertical lines indicate borders for the subvariable regions.

TABLE 2 Sequence variation of MG 192 subvariable regions based on sequence analysis of 97 variant sequences from 29 *M. genitalium* clinical specimens

Subvariable region	Location (aa)[a]	Length of subvariable region[b]	Frequency of amino acid change[c]
V1	52–67	16	0.0052
V2	83–106	20–27	0.0095–0.0129
V3	119–146	28	0.0107
V4	170–192	21–23	0.0215–0.0236
V5	202–217	16	0.0064
V6	237–267	28–34	0.0131–0.0158
V7	277–287	11–12	0.0069–0.0075
V8	328–336	9	0.0080
V9	363–370	8	0.0052
V10	409–419	4–12 (polyserine)	0.0112–0.0335
V11	471–497	27–29	0.0124–0.0130

[a] Relative to the predicted MG192 amino acid sequence of the type strain G37 (GenBank accession no. NC_000908). aa, amino acids.
[b] Number of amino acids.
[c] Calculated by dividing the number of unique sequences for each region by the total number of MG192 variants (97) and then by the number of amino acids in each region.

two variants in patient 168 which were detected at two sequential time points. In three patients (numbers 61, 105, and 172), a single MG192 sequence was detected at the first visit, and five MG192 variants were detected at the second visit. These data show that MG192 is changing and suggest that variation might be accumu-

TABLE 3 Topology analysis of deduced MG192 protein by different computational programs

Program	Transmembrane region[a]			Surface orientation[a]	
	TM1	TM2	TM3	Inside	Outside
TMpred	14–30	302–325	939–965	NA[c]	NA
TopPred	11–31	306–326	947–967	NA	NA
HMMTOP	12–30	NP[b]	943–967	NA	NA
TMHMM	13–30	NP	939–961	1–12, 962–1053	31–938
TOPCONS	12–32	NP	946–966	1–11, 967–1052	33–945
SCAMPI	12–32	NP	947–967	1–11, 968–1052	33–946

[a] Numbers represent the amino acid location of each protein segment based on the predicted MG192 amino acid sequence of the type strain G37 (GenBank accession no. NC_000908).
[b] NP, not present.
[c] NA, not available.

TABLE 4 MG192 sequences of *M. genitalium* specimens collected from five time points from patient 126

Specimen no.	Sampling date (mo/day/yr)	No. of variants[a]	Avg no. of nucleotide changes (±SD)[b,c]	Avg no. of amino acid changes (±SD)[b,d]
126.0	10/23/2001	4	34.3 ± 24.3	12.3 ± 8.5
126.1	1/15/2002	7	49.5 ± 25.1	16.8 ± 8.4
126.2	6/26/2002	5	91.2 ± 12.0	34.4 ± 4.7
126.3	9/17/2002	8	103 ± 18.9	37.7 ± 7.6
126.4	12/3/2002	9	83.4 ± 25.9	30.2 ± 10.7

[a] Pearson's correlation coefficient $r = 0.77$, $P = 0.1294$.
[b] Relative to the MG192 sequence of the variant 126.0a (GenBank accession no. JX857911), which is the most predominant variant from the first specimen 126.0.
[c] Pearson's correlation coefficient $r = 0.88$, $P = 0.0465$.
[d] Pearson's correlation coefficient $r = 0.87$, $P = 0.0547$.

lating over time. To study this issue more closely, we examined the entire MG192 variable region in five sequential specimens obtained from one patient (number 126) who was followed for a 14-month period. Different numbers of variants were detected by sequencing 10 plasmid clones from each time point (Table 1). Quantitative assessment of nucleotide and predicted amino acid changes revealed progressive increases in MG192 sequence variation over the entire variable region (Table 4). Further analysis of each subvariable region found that, generally, sequence variation progressively increased over the 14-month observation period (see Table S1 in the supplemental material). Additionally, as V4 and V6 are the most variable regions, we compared these two regions in the five sequential specimens from patient 126 and found that a only a few sequences were present at multiple time points, while most sequences were unique at each time point (Fig. 2; see also Fig. S2 in the supplemental material).

Phylogenetic relationships of the MG192 variants. Phylogenetic analysis by the neighbor-joining method with either nucleotide or deduced amino acid sequences showed that MG192 variants from the same patients clustered together in the same branches except for one variant from patient 172, which was only distantly related to the other variants from this patient. This clustering pattern was supported by strong bootstrap values of at least 98% and 82%, based on nucleotide and deduced amino acid sequences, respectively. The only exceptions are the variants from patient 126, which grouped together with a bootstrap value of 64% and 27% based on nucleotide and deduced amino acid sequences, respectively. The phylogenetic tree based on deduced amino acid sequences is shown in Fig. 3. We also constructed phylogenetic trees by the maximum-likelihood method and unweighted pair group method with arithmetic mean. These analyses yielded clustering patterns very similar to those obtained by the neighbor-joining method (data not shown). The clustering patterns were not associated with geographic location or sample date (data not shown).

DISCUSSION

Antigenic variation is one of the most common and effective strategies of immune evasion found in a variety of human pathogens, including bacteria, fungi, and parasites (31). Despite its small genome, *M. genitalium* also has the ability to generate extensive variation of the major surface proteins encoded by the MgPa operon, as demonstrated in this and previous studies (16–21). Variation in these genes is hypothesized to contribute to *M. genitalium* immune evasion and persistence of infection. The genetic architecture of the MgPa operon in *M. genitalium* is most similar to that of the P1 operon in *Mycoplasma pneumoniae*, the closest known relative to *M. genitalium*. However, the P1 operon undergoes only limited variation (32–34). It is of interest that the MgPa variation system closely resembles the extensively studied TprK variation system of *Treponema pallidum*, another sexually transmitted pathogen (35–37). In both systems, there is a single expression site per genome, but there are multiple donor sites that can be recombined into the expression site. Each donor site contains only a portion of the gene to be expressed; within the variable region of the expressed gene, there are multiple discrete subvariable regions (7 in Tprk and 11 in MG192 as described below), each flanked by highly conserved regions. However, in contrast to the TprK system in which each subvariable region contains interspersed tetranucleotide repeats, the subvariable regions of MG192 do not contain any repeats except for V10, which contains tandem trinucleotide repeats. The concept that MG192 and MG191 play essential roles in the *in vivo* survival of *M. genitalium* is supported by their exposure to the environment external to the host cell, strong recognition by serum antibodies in infected patients (13, 14), and involvement in cell adhesion (11), as well as the complete preservation of the intact reading frame in the approximately 120 MG192 and MG191 variant sequences described in the present study and in previous reports (16, 18–21).

Based on various computational analyses, the MG192 protein contains an N-terminal transmembrane domain, which may represent a cleavable signal sequence, and a C-terminal transmembrane domain, which may serve to anchor the protein into the cell membrane (Fig. 1). The entire variable region is predicted to be located external to the cell membrane. Each of the 11 subvariable regions is predicted to be antigenic (see Fig. S1 in the supplemental material). These topology analyses provide further evidence of the MG192 protein as a surface-exposed membrane protein. Variation of the MG192 sequence may change not only the antigenicity of the protein but also the mobility and adhesion properties of the organism.

In this study, we identified a total of 97 unique MG192 variant sequences in clinical specimens from 13 patients with *M. genitalium* infection and found great MG192 variation among and within patient specimens, consistent with previous studies of small numbers of specimens (20, 21). Strikingly, no variants were shared between any two patients. Genotyping confirmed that each patient was infected with a single *M. genitalium* strain, and thus the MG192 variations observed in these patients are not likely the result of coinfections with multiple strains. Phylogenetic analysis demonstrated that MG192 sequences within *M. genitalium* specimens from an individual patient were more related than sequences between patients (Fig. 3), and therefore the MG192 sequence variability is more limited within an *M. genitalium* strain than between strains. This phenomenon is very similar to the findings for the TprK gene of *T. pallidum* (37). The observations suggest that each patient was infected by only one *M. genitalium* strain and that MG192 variants in individuals were derived from a single or at least limited number of variants in the original inoculum. As it has been confirmed that MG192 variation results from recombination with MgPar sequences (18, 20, 21), the presence of a unique set of MG192 variants in each strain is consistent with the notion that each strain may have a unique set of MgPar sequences

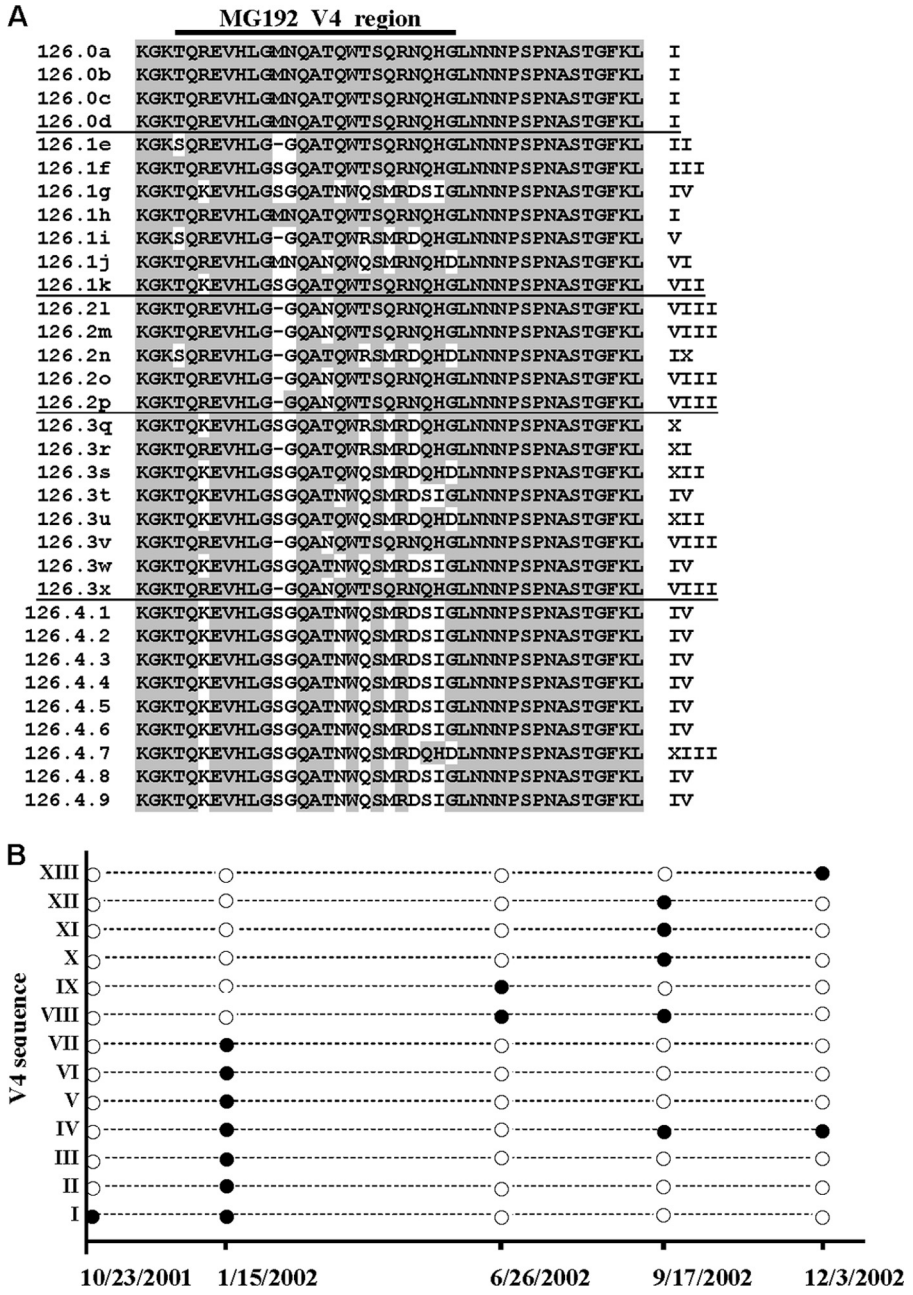

FIG 2 Sequence shift in the MG192 subvariable region V4 in sequential specimens from patient 126. (A) Alignment of the deduced amino acid sequences of MG192 variants. The label on the left side is for the full-length variant sequences obtained. Roman numbers I through XIII on the right side represent unique sequence types only for the V4 region shown. Amino acids identical to the variant 126.0a sequence are highlighted with shading. (B) Distribution of the MG192 V4 region sequences over time. Roman numbers I through XIII on the *y* axis correspond to the sequences I through XIII in the MG192 V4 region shown in panel A. The presence and absence of the sequences are indicated by the filled and empty circles, respectively. The sampling dates are indicated at the bottom as month/day/year.

that have evolved independently in different strains (20). The significant variation of MG192 between different strains theoretically would impair the host's ability to develop an effective immune response against new infection with a different strain, thus contributing to chronic infection.

Comparison of MG192 variant sequences in this study revealed 11 discrete subvariable regions. The finding of different degrees of variability in these subvariable regions suggests that these regions had different intrinsic rates of sequence change or may have been under different levels of selective pressure. The high frequency of variation in the V4 and V6 regions suggests that these regions of the protein may be especially susceptible to attack by the host immune system. Interestingly, previous studies of serial *in vitro* passages of the *M. genitalium* type strain G37 has also found sequence changes in these two regions as well as in other subvariable regions (18, 21). Perhaps the spontaneous changes confer replica-

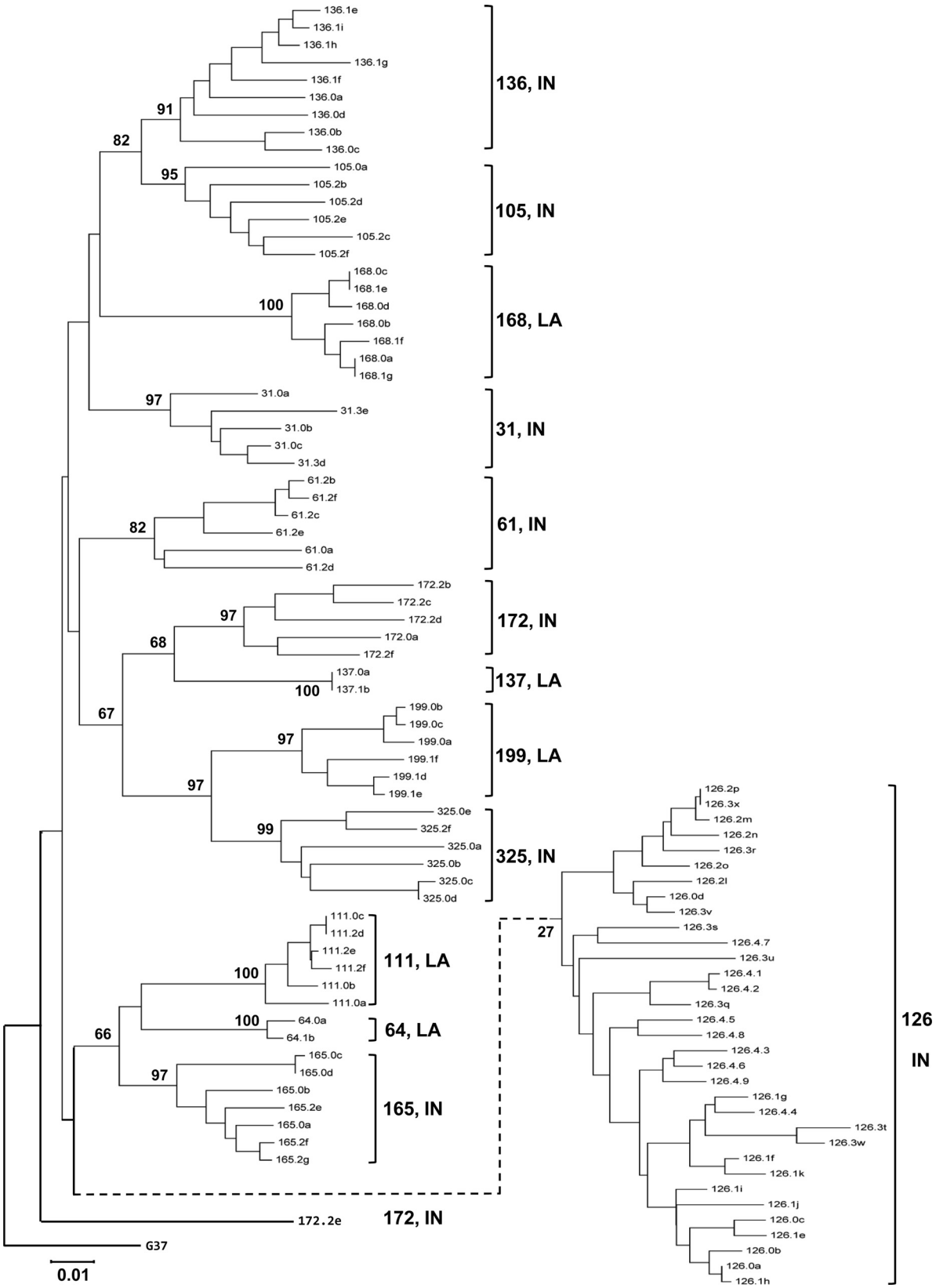

FIG 3 Phylogenetic analysis of the MG192 variant sequences in clinical specimens from 13 *M. genitalium*-infected patients. The tree was constructed using the neighbor-joining method and deduced amino acid sequences of the MG192 gene and rooted by using the type strain G37. Labels on the right end of each line correspond to the MG192 variants in Table 1, with patient identification numbers and sources indicated outside the right bracket. LA, Louisiana; IN, Indiana. Bootstrap values (percentages) indicated at the nodes were determined from 1,000 replicates. Only bootstrap values of >60% for the branches are shown except for the branch (27%) containing the variants from patient 126. The branch for all variants from patient 126 was separated at the bottom right to reduce the height of the image.

tive advantages to the organism and may be necessary to produce escape variants under selection pressure from the host immune system as has been suggested from studies of other bacteria (38, 39). In the current study, analysis of the V4 and V6 regions in five sequential specimens obtained from one patient showed increased sequence diversity over the 14-month observation period (Fig. 2). Only 3 of 13 V4 variants were present at multiple time points, and this was also true for only 3 of 10 V6 variants. The presence of unique variants at only one time point suggests immune selection during infection. The persistence of a variant (Fig. 2, variant IV, for example; see also Fig. S2, variant A, in the supplemental material) over an observation period of about 1 year or longer is likely due to the lack of antigenicity of the V4 or V6 sequence in this variant. Given the similarities of the MgPa and TprK systems as described above, it would be of interest to investigate if the MG192 subvariable regions are targeted by the humoral response while the conserved flanking regions are targeted by T cell responses, as has been demonstrated for the TprK gene (40). Additionally, given that the MG192 protein together with the MG191 protein has been shown to play a role not only in cellular adhesion and terminal organelle development but also in regulating cell division (11, 41), it is likely that variation in the amino acid sequence also may occur in response to *in vivo* environmental changes and/or the availability of certain human cell types, thus optimizing cell adherence, motility, and cell division.

A limitation of the study lies in the method used to identify MG192 variants, which was based on sequencing plasmid clones following PCR amplification of clinical specimens, as has been described in previous studies of the MG192 and MG191 genes (16, 18, 20, 21) as well as of the variable antigen genes in other pathogens (42, 43). This method is relatively insensitive for the detection of minority sequence populations. To overcome this limitation and improve the efficiency of variant identification, high-throughput next-generation sequencing (NGS) technologies would appear to be the answer. However, the short read length and high data volume from NGS technologies create problems for assembling highly repetitive sequences (44). In our genome sequencing studies of *M. genitalium* (45) and *Pneumocystis* (L. Ma et al., unpublished data; also http://www.broadinstitute.org/annotation/genome/Pneumocystis_group.2/MultiHome.html), we found that it is impossible to assemble highly repetitive sequences (>1.5 kb) using short reads generated by the 454 and Illumina platforms. One potential alternative approach is the newly developed PacBio SMRT sequencing technology, which is able to generate sequence reads longer than 3 kb (46). However, currently this technology is not widely available or affordable and requires sophisticated bioinformatics analysis, especially in relation to the correction of the high error rate of sequence reads inherent in this technology (46).

ACKNOWLEDGMENTS

This work was supported by a U.S. Department of Defense grant (W81XWH-08-1-0676), by a Gulf South Sexually Transmitted Infections/Topical Microbicide Cooperative Research Center grant from the NIH NIAID (5 U19 AI061972), and by the Louisiana Vaccine Center and the South Louisiana Institute for Infectious Disease Research sponsored by the Louisiana Board of Regents (grant 149752505J).

We thank Mary Welch and Judy Burnett for technical assistance and Christopher McGowin for helpful discussion.

REFERENCES

1. **Taylor-Robinson D, Jensen JS.** 2011. *Mycoplasma genitalium*: from chrysalis to multicolored butterfly. Clin. Microbiol. Rev. **24:**498–514. http://dx.doi.org/10.1128/CMR.00006-11.
2. **Kapiga SH, Sam NE, Mlay J, Aboud S, Ballard RC, Shao JF, Larsen U.** 2006. The epidemiology of HIV-1 infection in northern Tanzania: results from a community-based study. AIDS Care **18:**379–387. http://dx.doi.org/10.1080/09540120500465012.
3. **Mavedzenge SN, Van Der Pol B, Weiss HA, Kwok C, Mambo F, Chipato T, Van der Straten A, Salata R, Morrison C.** 2012. The association between *Mycoplasma genitalium* and HIV-1 acquisition in African women. AIDS **26:**617–624. http://dx.doi.org/10.1097/QAD.0b013e32834ff690.
4. **Perez G, Skurnick JH, Denny TN, Stephens R, Kennedy CA, Regivick N, Nahmias A, Lee FK, Lo SC, Wang RY, Weiss SH, Louria DB.** 1998. Herpes simplex type II and *Mycoplasma genitalium* as risk factors for heterosexual HIV transmission: report from the heterosexual HIV transmission study. Int. J. Infect. Dis. **3:**5–11. http://dx.doi.org/10.1016/S1201-9712(98)90088-1.
5. **McGowin CL, Annan RS, Quayle AJ, Greene SJ, Ma L, Mancuso MM, Adegboye D, Martin DH.** 2012. Persistent *Mycoplasma genitalium* infection of human endocervical epithelial cells elicits chronic inflammatory cytokine secretion. Infect. Immun. **80:**3842–3849. http://dx.doi.org/10.1128/IAI.00819-12.
6. **McGowin CL, Spagnuolo RA, Pyles RB.** 2010. *Mycoplasma genitalium* rapidly disseminates to the upper reproductive tracts and knees of female mice following vaginal inoculation. Infect. Immun. **78:**726–736. http://dx.doi.org/10.1128/IAI.00840-09.
7. **Taylor-Robinson D, Furr PM, Tully JG, Barile MF, Moller BR.** 1987. Animal models of *Mycoplasma genitalium* urogenital infection. Isr. J. Med. Sci. **23:**561–564.
8. **Taylor-Robinson D, Gilroy CB, Thomas BJ, Hay PE.** 2004. *Mycoplasma genitalium* in chronic non-gonococcal urethritis. Int. J. STD AIDS **15:**21–25. http://dx.doi.org/10.1258/095646204322637209.
9. **Wikstrom A, Jensen JS.** 2006. *Mycoplasma genitalium*: a common cause of persistent urethritis among men treated with doxycycline. Sex. Transm. Infect. **82:**276–279. http://dx.doi.org/10.1136/sti.2005.018598.
10. **Bradshaw CS, Chen MY, Fairley CK.** 2008. Persistence of *Mycoplasma genitalium* following azithromycin therapy. PLoS One **3:**e3618. http://dx.doi.org/10.1371/journal.pone.0003618.
11. **Burgos R, Pich OQ, Ferrer-Navarro M, Baseman JB, Querol E, Pinol J.** 2006. *Mycoplasma genitalium* P140 and P110 cytadhesins are reciprocally stabilized and required for cell adhesion and terminal-organelle development. J. Bacteriol. **188:**8627–8637. http://dx.doi.org/10.1128/JB.00978-06.
12. **Hu PC, Schaper U, Collier AM, Clyde WA, Jr, Horikawa M, Huang YS, Barile MF.** 1987. A *Mycoplasma genitalium* protein resembling the *Mycoplasma pneumoniae* attachment protein. Infect. Immun. **55:**1126–1131.
13. **Baseman JB, Cagle M, Korte JE, Herrera C, Rasmussen WG, Baseman JG, Shain R, Piper JM.** 2004. Diagnostic assessment of *Mycoplasma genitalium* in culture-positive women. J. Clin. Microbiol. **42:**203–211. http://dx.doi.org/10.1128/JCM.42.1.203-211.2004.
14. **Svenstrup HF, Jensen JS, Gevaert K, Birkelund S, Christiansen G.** 2006. Identification and characterization of immunogenic proteins of *Mycoplasma genitalium*. Clin. Vaccine Immunol. **13:**913–922. http://dx.doi.org/10.1128/CVI.00048-06.
15. **Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, Bult CJ, Kerlavage AR, Sutton G, Kelley JM, Fritchman RD, Weidman JF, Small KV, Sandusky M, Fuhrmann J, Nguyen D, Utterback TR, Saudek DM, Phillips CA, Merrick JM, Tomb JF, Dougherty BA, Bott KF, Hu PC, Lucier TS, Peterson SN, Smith HO, Hutchison CA, III, Venter JC.** 1995. The minimal gene complement of *Mycoplasma genitalium*. Science **270:**397–403. http://dx.doi.org/10.1126/science.270.5235.397.
16. **Iverson-Cabral SL, Astete SG, Cohen CR, Rocha EP, Totten PA.** 2006. Intrastrain heterogeneity of the *mgpB* gene in *Mycoplasma genitalium* is extensive in vitro and in vivo and suggests that variation is generated via recombination with repetitive chromosomal sequences. Infect. Immun. **74:**3715–3726. http://dx.doi.org/10.1128/IAI.00239-06.
17. **Jensen JS.** 2006. *Mycoplasma genitaliu*m infections. Diagnosis, clinical aspects, and pathogenesis. Dan. Med. Bull. **53:**1–27.
18. **Ma L, Jensen JS, Myers L, Burnett J, Welch M, Jia Q, Martin DH.** 2007. *Mycoplasma genitalium*: an efficient strategy to generate genetic variation

from a minimal genome. Mol. Microbiol. **66:**220–236. http://dx.doi.org/10.1111/j.1365-2958.2007.05911.x.
19. **Peterson SN, Bailey CC, Jensen JS, Borre MB, King ES, Bott KF, Hutchison CA, III.** 1995. Characterization of repetitive DNA in the *Mycoplasma genitalium* genome: possible role in the generation of antigenic variation. Proc. Natl. Acad. Sci. U. S. A. **92:**11829–11833. http://dx.doi.org/10.1073/pnas.92.25.11829.
20. **Ma L, Jensen JS, Mancuso M, Hamasuna R, Jia Q, McGowin CL, Martin DH.** 2010. Genetic variation in the complete MgPa operon and its repetitive chromosomal elements in clinical strains of *Mycoplasma genitalium*. PLoS One **5:**e15660. http://dx.doi.org/10.1371/journal.pone.0015660.
21. **Iverson-Cabral SL, Astete SG, Cohen CR, Totten PA.** 2007. *mgpB* and *mgpC* sequence diversity in *Mycoplasma genitalium* is generated by segmental reciprocal recombination with repetitive chromosomal sequences. Mol. Microbiol. **66:**55–73. http://dx.doi.org/10.1111/j.1365-2958.2007.05898.x.
22. **Mena L, Wang X, Mroczkowski TF, Martin DH.** 2002. *Mycoplasma genitalium* infections in asymptomatic men and men with urethritis attending a sexually transmitted diseases clinic in New Orleans. Clin. Infect. Dis. **35:**1167–1173. http://dx.doi.org/10.1086/343829.
23. **Mena LA, Mroczkowski TF, Nsuami M, Martin DH.** 2009. A randomized comparison of azithromycin and doxycycline for the treatment of *Mycoplasma genitalium*-positive urethritis in men. Clin. Infect. Dis. **48:**1649–1654. http://dx.doi.org/10.1086/599033.
24. **Tosh AK, Van Der Pol B, Fortenberry JD, Williams JA, Katz BP, Batteiger BE, Orr DP.** 2007. *Mycoplasma genitalium* among adolescent women and their partners. J. Adolesc. Health **40:**412–417. http://dx.doi.org/10.1016/j.jadohealth.2006.12.005.
25. **Ma L, Martin DH.** 2004. Single-nucleotide polymorphisms in the rRNA operon and variable numbers of tandem repeats in the lipoprotein gene among *Mycoplasma genitalium* strains from clinical specimens. J. Clin. Microbiol. **42:**4876–4878. http://dx.doi.org/10.1128/JCM.42.10.4876-4878.2004.
26. **Ma L, Taylor S, Jensen JS, Myers L, Lillis R, Martin DH.** 2008. Short tandem repeat sequences in the *Mycoplasma genitalium* genome and their use in a multilocus genotyping system. BMC Microbiol. **8:**130. http://dx.doi.org/10.1186/1471-2180-8-130.
27. **Hjorth SV, Bjornelius E, Lidbrink P, Falk L, Dohn B, Berthelsen L, Ma L, Martin DH, Jensen JS.** 2006. Sequence-based typing of *Mycoplasma genitalium* reveals sexual transmission. J. Clin. Microbiol. **44:**2078–2083. http://dx.doi.org/10.1128/JCM.00003-06.
28. **Musatovova O, Herrera C, Baseman JB.** 2006. Proximal region of the gene encoding cytadherence-related protein permits molecular typing of *Mycoplasma genitalium* clinical strains by PCR-restriction fragment length polymorphism. J. Clin. Microbiol. **44:**598–603. http://dx.doi.org/10.1128/JCM.44.2.598-603.2006.
29. **Ma L, Jensen JS, Mancuso M, Hamasuna R, Jia Q, McGowin CL, Martin DH.** 2012. Variability of trinucleotide tandem repeats in the MgPa operon and its repetitive chromosomal elements in *Mycoplasma genitalium*. J. Med. Microbiol. **61:**191–197. http://dx.doi.org/10.1099/jmm.0.030858-0.
30. **Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S.** 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. **28:**2731–2739. http://dx.doi.org/10.1093/molbev/msr121.
31. **Deitsch KW, Lukehart SA, Stringer JR.** 2009. Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens. Nat. Rev. Microbiol. **7:**493–503. http://dx.doi.org/10.1038/nrmicro2145.
32. **Kenri T, Taniguchi R, Sasaki Y, Okazaki N, Narita M, Izumikawa K, Umetsu M, Sasaki T.** 1999. Identification of a new variable sequence in the P1 cytadhesin gene of *Mycoplasma pneumoniae*: evidence for the generation of antigenic variation by DNA recombination between repetitive sequences. Infect. Immun. **67:**4557–4562.
33. **Ruland K, Himmelreich R, Herrmann R.** 1994. Sequence divergence in the ORF6 gene of *Mycoplasma pneumonia*. J. Bacteriol. **176:**5202–5209.
34. **Zhao F, Cao B, Li J, Song S, Tao X, Yin Y, He L, Zhang J.** 2011. Sequence analysis of the p1 adhesin gene of *Mycoplasma pneumoniae* in clinical isolates collected in Beijing in 2008 to 2009. J. Clin. Microbiol. **49:**3000–3003. http://dx.doi.org/10.1128/JCM.00105-11.
35. **Centurion-Lara A, LaFond RE, Hevner K, Godornes C, Molini BJ, Van Voorhis WC, Lukehart SA.** 2004. Gene conversion: a mechanism for generation of heterogeneity in the *tprK* gene of *Treponema pallidum* during infection. Mol. Microbiol. **52:**1579–1596. http://dx.doi.org/10.1111/j.1365-2958.2004.04086.x.
36. **Giacani L, Molini BJ, Kim EY, Godornes BC, Leader BT, Tantalo LC, Centurion-Lara A, Lukehart SA.** 2010. Antigenic variation in *Treponema pallidum*: TprK sequence diversity accumulates in response to immune pressure during experimental syphilis. J. Immunol. **184:**3822–3829. http://dx.doi.org/10.4049/jimmunol.0902788.
37. **LaFond RE, Centurion-Lara A, Godornes C, Rompalo AM, Van Voorhis WC, Lukehart SA.** 2003. Sequence diversity of *Treponema pallidum* subsp. *pallidum tprK* in human syphilis lesions and rabbit-propagated isolates. J. Bacteriol. **185:**6262–6268. http://dx.doi.org/10.1128/JB.185.21.6262-6268.2003.
38. **Borst P.** 1991. Molecular genetics of antigenic variation. Immunol. Today **12:**A29–A33. http://dx.doi.org/10.1016/S0167-5699(05)80009-X.
39. **Criss AK, Kline KA, Seifert HS.** 2005. The frequency and rate of pilin antigenic variation in *Neisseria gonorrhoeae*. Mol. Microbiol. **58:**510–519. http://dx.doi.org/10.1111/j.1365-2958.2005.04838.x.
40. **Morgan CA, Molini BJ, Lukehart SA, Van Voorhis WC.** 2002. Segregation of B and T cell epitopes of *Treponema pallidum* repeat protein K to variable and conserved regions during experimental syphilis infection. J. Immunol. **169:**952–957.
41. **Pich OQ, Burgos R, Querol E, Pinol J.** 2009. P110 and P140 cytadherence-related proteins are negative effectors of terminal organelle duplication in *Mycoplasma genitalium*. PLoS One **4:**e7452. http://dx.doi.org/10.1371/journal.pone.0007452.
42. **Centurion-Lara A, Giacani L, Godornes C, Molini BJ, Brinck Reid T, Lukehart SA.** 2013. Fine analysis of genetic diversity of the *tpr* gene family among treponemal species, subspecies and strains. PLoS Negl. Trop. Dis. **7:**e2222. http://dx.doi.org/10.1371/journal.pntd.0002222.
43. **Kutty G, Maldarelli F, Achaz G, Kovacs JA.** 2008. Variation in the major surface glycoprotein genes in *Pneumocystis jirovecii*. J. Infect. Dis. **198:**741–749. http://dx.doi.org/10.1086/590433.
44. **Treangen TJ, Salzberg SL.** 2012. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat. Rev. Genet. **13:**36–46. http://dx.doi.org/10.1038/nrg3117.
45. **McGowin CL, Ma L, Jensen JS, Mancuso MM, Hamasuna R, Adegboye D, Martin DH.** 2012. Draft genome sequences of four axenic *Mycoplasma genitalium* strains isolated from Denmark, Japan, and Australia. J. Bacteriol. **194:**6010–6011. http://dx.doi.org/10.1128/JB.01478-12.
46. **Jiao XL, Zheng X, Ma L, Kutty G, Gogineni E, Sun Q, Sherman BT, Hu XJ, Jones K, Raley C, Tran B, Munroe DJ, Stephens R, Liang D, Imamichi T, Kovacs JA, Lempicki RA, Huang DW.** 2013. A benchmark study on error assessment and quality control of CCS reads derived from the PacBio RS. J. Data Mining Genomics Proteomics **4:**136. http://dx.doi.org/10.4172/2153-0602.1000136.
47. **Cazanave C, Charron A, Renaudin H, Bebear C.** 2012. Method comparison for molecular typing of French and Tunisian *Mycoplasma genitalium*-positive specimens. J. Med. Microbiol. **61:**500–506. http://dx.doi.org/10.1099/jmm.0.037721-0.