Variability of trinucleotide tandem repeats in the MgPa operon and its repetitive chromosomal elements in *Mycoplasma genitalium*

Liang Ma,^{1†} Jørgen S. Jensen,² Miriam Mancuso,¹ Ryoichi Hamasuna,³ Qiuyao Jia,¹ Chris L. McGowin¹ and David H. Martin¹

Correspondence

Liang Ma

LMa1@lsuhsc.edu or

MaL3@mail.nih.gov

¹Section of Infectious Diseases, Department of Medicine, Louisiana State University Health Sciences Center, New Orleans, Louisiana, USA

²Mycoplasma Laboratory, Statens Serum Institut, Copenhagen, Denmark

³Department of Urology, University of Occupational and Environmental Health, Yahatanishi-ku, Kitakyushu, Japan

Mycoplasma genitalium, a human pathogen associated with sexually transmitted diseases, is unique in that it has the smallest genome of any known free-living organism. Despite its small genome, 4.7 % of the total genomic sequence is devoted to making the MgPa adhesin operon (containing the MG190, MG191 and MG192 genes) and its repetitive chromosomal sequences (known as MgPars). The goals of this study were to investigate the location, organization and variability of trinucleotide tandem repeats (TTRs) in the MgPa operon and MgPars and to explore the possible mechanisms and role of TTR variations. By analysing the complete MgPa operon and complete or partial MgPar sequences in a collection of 15 geographically diverse clinical strains of *M. genitalium*, TTR sequences were identified in four regions in MG191, one region in MG192, and two or three regions in each of all nine MgPars except for MgPar 3. These TTRs were variable not only in the repeat copy number but also in the repeat unit sequence among or within strains. The key mechanisms for the TTR variations likely include recombination between MgPa and MgPars, and slipped-strand mispairing. TTR variation may represent a mechanism to maximize the variation of the MgPa operon, which is complementary to genetic variation involving segmental recombination between MgPa and MgPars, thus enhancing the organism's ability to adhere to and colonize host cells as well as evasion of the host immune system.

Received 30 January 2011

Accepted 5 October 2011

INTRODUCTION

Tandem DNA repeat sequences are found in both eukaryotic and prokaryotic genomes. Expansion and contraction of repeats may alter gene function or affect local structure of the DNA molecule or the encoded proteins (Charlesworth *et al.*, 1994; Htun & Dahlberg, 1989; Richard *et al.*, 2008). In humans a particular class of tandem DNA repeats, called trinucleotide tandem repeats (TTRs), has been shown to be associated with various neurodegenerative disorders, known as triplet repeat

expansion diseases (Di Prospero & Fischbeck, 2005). In some microbes variation in intragenic repeat number provides the functional diversity of cell-surface antigens that allows rapid adaptation to the environment and avoidance of the host immune system (van Belkum *et al.*, 1998; Verstrepen *et al.*, 2005). Since these repeats are often hypervariable between individuals of the same species, they have been widely used for genome mapping and molecular typing (Ma *et al.*, 2008; van Belkum *et al.*, 1998).

Mycoplasma genitalium, a sexually transmitted pathogen associated with genital tract diseases (Jensen, 2006), has the smallest genome of any known free-living micro-organism (Fraser *et al.*, 1995; Glass *et al.*, 2006). Despite its small genome, 4.7 % of the total genomic sequence is devoted to making the MgPa adhesin operon and its repetitive sequences (Fraser *et al.*, 1995). These repetitive sequences, known as MgPars, are not arranged in tandem but dispersed around the genome in nine loci (Fraser *et al.*, 1995; Iverson-Cabral *et al.*, 2006, 2007; Jensen, 2006; Ma *et al.*, 2007; Peterson *et al.*, 1995). We and others have

[†]Present address: Critical Care Medicine Department, Clinical Center, National Institutes of Health, Bethesda, MD 20892, USA.

Abbreviation: TTR, trinucleotide tandem repeat.

The GenBank accession numbers for the *M. genitalium* MgPa operon are FJ872584–FJ872592 and GU226196–GU226203, and those for the MgPar sequences are EF117289–EF117301, FJ872560–FJ872583 and FJ872593.

Four supplementary figures are available with the online version of this paper.

demonstrated that the MgPa serve as the donor sequences that can recombine into two of the three genes encoded by the MgPa operon, MG191 (*mgpB*) and MG192 (*mgpC*) (Iverson-Cabral *et al.*, 2006, 2007; Jensen, 2006; Ma *et al.*, 2007; Peterson *et al.*, 1995). Both MgpB and MgpC proteins are antigenic and can elicit serum antibody responses in *M. genitalium* patients and experimentally inoculated animals (Clausen *et al.*, 2001; Hu *et al.*, 1987; McGowin *et al.*, 2010; Morrison-Plummer *et al.*, 1987; Svenstrup *et al.*, 2006). Recombination within the minimized genome can generate extensive variation, which presumably allows the organism to evade the host immune response and to adapt to diverse host microenvironments, thus establishing persistent infection. Recently we searched the *M. genitalium* genome and identified short tandem repeats (1–5 bp) in 18 loci throughout the genome (Ma *et al.*, 2008). Surprisingly, the majority of these repeats (15/18) are TTRs and eight of the 15 TTRs were located in the MgPa operon and MgPars. These observations raise questions as to the role of TTRs in the MgPa operon and what mechanisms are involved in TTR variation. In addition, current data on TTRs in the MgPa operon and MgPars are based on *in vitro* studies of several *M. genitalium* laboratory strains, including the type strain G37 (Iverson-Cabral *et al.*, 2006, 2007; Jensen, 2006; Ma *et al.*, 2007, 2008; Rocha & Blanchard, 2002; van Belkum *et al.*, 1998), and *in vivo* studies of clinical specimens from *M. genitalium*-infected patients (Jensen, 2006; Ma *et al.*, 2007). However, in all these studies, only small portions of the MgPa operon and MgPars were examined for TTRs. It is unknown if additional TTRs exist and what degree of TTR variation exists among clinical strains.

The goals of this study were to investigate the location, organization and variability of TTRs in the complete MgPa operon and MgPars in a collection of geographically diverse strains of *M. genitalium*, and to explore the possible mechanisms and role of TTR variations.

METHODS

***M. genitalium* specimens.** Thirteen axenic isolates of *M. genitalium* (Table 1) were obtained from our previous studies (Hamasuna *et al.*, 2007; Jensen *et al.*, 1996; Jensen, 2009). All isolates except for M2282 were cloned by standard filtration or limiting dilution cloning procedures (Hamasuna *et al.*, 2007; Jensen *et al.*, 1996). These isolates were grown in modified Friis's medium containing horse serum (Jensen *et al.*, 1996). In addition, we used two sequential urine specimens obtained from each of two men with acute urethritis in New Orleans, Louisiana (patient nos 199 and 64), who were followed for 10 and 11 days, respectively (Hjorth *et al.*, 2006; Ma & Martin, 2004; Ma *et al.*, 2008). All urine specimens had been genotyped at multiple genomic loci (Hjorth *et al.*, 2006; Ma & Martin, 2004; Ma *et al.*, 2008). Each patient presented an identical genotype profile between the two sequential specimens, indicating that each patient was infected with only a single *M. genitalium* strain. Written informed consent was obtained from both patients and the study protocol was approved by the Institutional Review Board of the Louisiana State University Health Sciences Center.

PCR and sequencing of the MgPa operon and MgPar regions.

The whole MgPa operon sequence and complete or partial sequences of nine MgPars were amplified by PCR as described elsewhere (Ma *et al.*, 2007, 2010). Proofreading *rTth* DNA polymerase (Applied Biosystems) was used to amplify DNA from all axenic isolates and AmpliTaq Gold DNA polymerase (Applied Biosystems) for DNA from all urine specimens. Initially, all PCR products were directly sequenced using internal primers. If there was any indication of mixed or ambiguous sequences present in the sequence chromatograms, the PCR products were cloned into pCR 2.1-TOPO vector (Invitrogen) and multiple plasmid clones were sequenced as described previously (Ma *et al.*, 2007, 2010). In addition, all TTR regions in the four urine specimens were sequenced after subcloning, as previous studies have found extensive variation in the MG191 and MG192 genes in clinical specimens (Ma *et al.*, 2007; Iverson-Cabral *et al.*, 2006).

The sequence data involved in this study included the complete MgPa operon of all 15 strains (GenBank accession numbers FJ872584–FJ872592 and GU226196–GU226203), the complete sequence of all nine MgPars in the two sequential specimens from patient no. 199 (GenBank accession numbers EF117293–EF117301), the complete sequence of MgPars 3, 6, 8 and 9 and the TTR-containing regions of other MgPars in the two sequential specimens of patient no. 64 (GenBank accession numbers FJ872560–FJ872569 and FJ872593), the 5'-portion of MgPar 1 (containing MG191-EF1 and MG191-EF2 homologous regions) in all five Danish isolates (GenBank accession numbers FJ872570–FJ872574), and the MG192-L homologous region in MgPars 2, 8 and 9 in two Danish isolates M2321 and M2341 (GenBank accession numbers FJ872575–FJ872583). Because there was a high cost to sequence all nine MgPars of each strain and because studies of selected strains had allowed us to identify the TTR loci and their variability, we did not analyse all nine MgPars in all strains involved in this study.

Possibility of generation of TTR variation during PCR amplification.

To investigate the possibility of deletion or insertion of TTR units during PCR amplification, a plasmid clone containing the G37 MG192 repeat region (Ma *et al.*, 2007) was amplified by PCR using primers 5346F and 227529R (Ma *et al.*, 2007) and conditions described below. In addition, the four urine samples were amplified in at least two independent PCRs followed by separate cloning and sequencing for each of the following regions: (1) TTRs in the MG192-L region in all four samples (using primers 5346F and 227529); and (2) the MG191-EF region (using primers 2833F and 4146R) and the MG191-F region (using primers 3942F and 4146R) in samples 64.0 and 64.1. The sequences of all primers have been reported elsewhere (Ma *et al.*, 2007, 2010). The PCR mixture (50 µl) contained 10 ng plasmid DNA or 2 µl DNA extract from urine specimens, 0.5 µM each primer, 0.2 mM dNTPs, 1 × PCR buffer (10 mM Tris/HCl, pH 8.3; 50 mM KCl; 2.5 mM MgCl₂) and 2.5 U AmpliTaq Gold DNA polymerase (Applied Biosystems). Amplification was performed in a PTC-100 programmable thermal controller (MJ Research) with a touchdown cycling protocol as follows: 95 °C for 8 min, then 10 cycles of 1 min at 95 °C, 2 min at 65 °C (with a decrease by 1.5 °C every cycle to reach 50 °C in the last cycle) and 3 min at 72 °C, followed by 30 cycles of 1 min at 95 °C, 2 min at 50 °C and 3 min at 72 °C. Each experiment included a negative control without template DNA. The PCR product was cloned into the pCR 2.1-TOPO vector (Invitrogen) and plasmid clones were sequenced as described previously (Ma *et al.*, 2007).

RESULTS

Location and variability of TTRs in the MgPa operon

Comparison of the complete MgPa operon sequence in 15 strains and the G37 type strain revealed TTRs in four

Table 1. Variation of TTRs in the MgPa operon among *M. genitalium* axenic isolates and urine specimens

Specimen	Origin	Reference	Repeat unit-copy no.*				
			MG191-C	MG191-EF1	MG191-EF2	MG191-F	MG192-L
G37†	UK	Tully <i>et al.</i> (1981)	TCT-3	TCT-2	AGT-1	AGT-7	AGT-11
Axenic isolates							
M30	UK	Tully <i>et al.</i> (1981)	TCT-3	TCT-1	AGT-2	AGT-6	AGT-7
M2282	Denmark	Jensen <i>et al.</i> (1996)	TCT-3	TCT-1	AGT-1	AGT-6 ² , 7 ¹	AGT-7
M2288	Denmark	Jensen <i>et al.</i> (1996)	TCT-7	TCT-4 ⁵	AGT-6 ⁴ , 8 ¹	AGT-10 ¹ , 11 ⁴	AGT-5 ² , 6 ²
M2300	Denmark	Jensen <i>et al.</i> (1996)	TCT-3	None	AGT-1	AGT-5 ² , 6 ¹	AGT-6 ¹ , 8 ¹
M2321	Denmark	Jensen <i>et al.</i> (1996)	TCT-10	TCT-1	AGT-5	AGT-7 ¹ , 8 ³	TCT-11 ³ , 12 ¹
M2341	Denmark	Jensen <i>et al.</i> (1996)	TCT-7	TCT-4	AGT-1	AGT-7 ²	TCT-7
M6257	Sweden	Jensen <i>et al.</i> (2008)	TCT-5	TCT-3	AGT-2	AGT-8	AGT-13
M6280	Sweden	Jensen <i>et al.</i> (2008)	TCT-7	TCT-8	AGT-4	AGT-7	TCT-5
M6282	Japan	Hamasuna <i>et al.</i> (2007)	TCT-6	TCT-1	AGT-1	AGT-6	AGT-7
M6283	Japan	Hamasuna <i>et al.</i> (2007)	TCT-5	TCT-2	AGT-8	AGT-6	AGT-4
M6284	Japan	Hamasuna <i>et al.</i> (2007)	TCT-3	TCT-1	AGT-1	AGT-6	AGT-7
M6285	Sweden	Hamasuna <i>et al.</i> (2005)	TCT-5	TCT-1	AGT-6	AGT-10	AGT-7
M6286	Sweden	Hamasuna <i>et al.</i> (2005)	TCT-9	TCT-1	AGT-5	AGT-6	AGT-9
Urine specimens‡							
199.0	USA	Ma <i>et al.</i> (2007)	TCT-6 ¹ , 7 ⁴	None ⁷	AGT-4 ⁷	AGT-6 ⁶	AGT-6 ¹⁰
199.1	USA	Ma <i>et al.</i> (2007)	TCT-6 ² , 7 ³	None ⁷	AGT-4 ⁷	AGT-6 ¹ , 7 ⁵	AGT-6 ¹⁸
64.0	USA	Ma <i>et al.</i> (2008)	TCT-6	None ⁶	AGT-1 ⁴ , 2 ⁸	AGT-9 ¹ , 10 ¹ , 11 ⁹ , 12 ⁴	AGT-5 ¹ , 6 ¹⁰ , 7 ³
64.1	USA	Ma <i>et al.</i> (2008)	TCT-6	None ⁷	AGT-2 ¹²	AGT-8 ¹ , 9 ³ , 10 ⁶ , 11 ¹¹ , 12 ¹	AGT-6 ¹ , 7 ⁸

*Superscript numbers represent the number of plasmid clones analysed for respective repeats. For example, 'AGT-6², 7¹' in the MG191-F locus of strain M2282 indicates six and seven copies of the AGT repeat present in two plasmid clones and one plasmid clone, respectively. Repeat units not shown with a superscript number were determined by direct sequencing.

†From the published genome sequence of the type strain G37 (GenBank no. NC_000908).

‡Specimens 199.0 and 199.1 were from patient no. 199 on his first visit and a 10 day follow-up visit, respectively; specimens 64.0 and 64.1 were from patient no. 64 on his first visit and an 11 day follow-up visit, respectively.

regions of the MG191 gene (designated region MG191-C, MG191-EF1, MG191-EF2 and MG191-F, respectively) and one region of the MG192 gene (designated MG192-L) (Fig. 1). All of these TTRs showed variation in the repeat copy number among and within strains except for the TTR in the MG191-EF1 region, which showed variation only among strains, not within strains (Table 1; see also Supplementary Fig. S1, available with the online version of this paper). The TTR unit was either TCT or AGT and was consistent for each of the regions except the MG192-L region, where some strains had AGT repeat units and others had TCT repeats (Table 1, Supplementary Fig. S1E). The repeat units in the MG191-EF1 region may also include ACT and TCC as degenerative repeat units (Supplementary Fig. S1B). Mixed alleles were observed in four TTR loci in four axenic isolates from Denmark and in urine specimens of both patients from New Orleans (Table 1). In all samples containing mixed TTR alleles, the number of repeats varied by only one or two repeat units.

Location and variability of TTRs in MgPars

We analysed all nine complete MgPar sequences in the G37 type strain and urine specimens from patient 199, and

additional complete or partial MgPars in the five Danish isolates and urine specimens from patient 64 (Ma *et al.*, 2010). TTRs were identified in all MgPars except for MgPar 3. Each MgPar (except for MgPar 3) contained two TTR loci in the region homologous to the MG191-EF1 and MG191-EF2 regions. One additional TTR locus was present in the MG192-L homologous region of MgPars 2, 8 and 9. In total, 19 TTR loci were identified within MgPars (indicated by ovals in Fig. 1). These TTRs were variable not only in the repeat copy number but also in the repeat unit sequence. The most predominant repeat units were TCT/ACT in MG191-EF1 homologous regions (Supplementary Fig. S2A, B), and AGT in MG191-EF2 homologous regions (Supplementary Fig. S2C, D) and MG192-L homologous regions (Table 2). The TTR pattern in MG191-EF1 homologous regions was complicated by the presence of mixed repeat units, primarily consisting of TCT and ACT, and, less frequently, TCC and TCA. All such repeat units encode either serine (AGT, TCT, TCC and TCA) or threonine (ACT). A mixture of TCT and ACT was also present in the MG192-L homologous region of MgPar 8 (Table 2). TTRs in the MG192-L homologous region of MgPars 2, 8 and 9 showed variation in the repeat copy number and in the repeat unit sequence among and within strains (Table 2, Supplementary Fig. S3).

Fig. 1. Distribution of TTRs in the MG191 and MG192 genes and nine MgPars of *M. genitalium*. Divisions A to M in the MG191 and MG192 genes are indicated by vertical lines in the map at the top, representing the restriction fragments described previously (Dallo & Baseman, 1991; Peterson *et al.*, 1995). The four hypervariable regions B, EF, G and JKLM are highlighted in different colours (Ma *et al.*, 2010). Regions homologous between MG191/MG192 and MgPars are indicated in identical colours. The hatched boxes represent intervening sequences that are unusually A-T rich and contain stop codons. The numbers bordering each segment of each of the MgPars refer to the nucleotide positions in G37 full-length MgPars as described elsewhere (Iverson-Cabral *et al.*, 2006; Ma *et al.*, 2010). The line length in the diagrams is not always proportional to the number of nucleotides involved due to the presence of minor deletions/insertions. The TTR regions are indicated by dark black ovals, with their location designated MG191-C, MG-EF1, Mg191-EF2, MG191-F and MG191-L.

Possibility of generation of TTR variation during PCR amplification

To investigate the possibility that the observed TTR variability was due to errors during PCR amplification, we first analysed 15 plasmid clones generated from PCR amplification of the G37 MG192-L repeat region. No variation in the TTR sequence or copy number was observed. Further, we performed replicate PCR amplifications of the MG192-L, MG191-EF and MG191-F regions in the four patient urine specimens followed by cloning and sequence analysis. However, significant differences in the distribution of sequence variants between different PCR runs were not found (Ma *et al.*, 2007 and data not shown). In addition, no TTR copy number variation was seen in up to 18 plasmid clones of the MG191-EF2 or MG192-L region in samples 199.0 or 199.1, or the MG191-EF2 region in sample 64.1 (Table 1). These data strongly suggest that artefactual slipped-strand mispairing events did not occur during PCR amplification.

DISCUSSION

In the present study we identified abundant TTRs, which were present in five regions of the MgPa operon and 19 regions of MgPars. Only five of them, including two in the MgPa operon (MG191-F and MG192-L regions) and three in MgPars (MG192-L homologous region in each of MgPars 2, 8 and 9), have been described in previous publications (Iverson-Cabral *et al.*, 2006, 2007; Jensen, 2006; Ma *et al.*, 2007, 2008; Rocha & Blanchard, 2002; van Belkum *et al.*, 1998). These TTRs were variable not only in the repeat copy number but also in the repeat unit sequence.

The TTR variations observed in this study could be due to the following mechanisms.

Firstly, slipped-strand mispairing is likely the primary mechanism for the variation of TTRs present in the MG191-C and MG191-F regions (Fig. 1 and Supplementary Fig. S1A, D) because (1) both regions are present in a single copy in the genome and their flanking regions have no

Table 2. TTR variation in MG192-L homologous regions of MgPars 2, 8 and 9

Specimen	Repeat unit ^{copy number} _x		
	MgPar 2	MgPar 8	MgPar 9
G37†	AGT ¹⁶	AGT ¹⁰	AGT ⁹
199.0	AGT ⁴⁻⁶	TCT ⁹⁻¹¹ ACT ⁶⁻⁸	AGT ⁵
199.1	AGT ⁴⁻⁵	TCT ⁷⁻¹¹ ACT ⁶⁻⁷	AGT ⁵
64.0	AGT ⁷⁻¹¹	AGT ⁷	AGT ⁶⁻⁸
64.1	AGT ⁸⁻¹¹	AGT ⁷	AGT ⁷⁻⁸
M2321	AGT ⁹ ACT ¹	AGT ⁶ ACT ¹	AGT ⁶ ACT ¹
M2341	AGT ⁸	AGT ⁶	AGT ⁶

*Both urine specimens from each of patients 199 and 64 were analysed by sequencing plasmid clones while the two Danish isolates were studied by direct sequencing.

†From the published genome sequence of the type strain G37 GenBank no. NC_000908.

homology to any of the nine MgPar sequences and thus are not expected to undergo recombination with MgPars; and (2) TTR variation in these two regions, particularly in the MG191-F region, was commonly present as mixed sequences in the same specimen, reflecting stepwise addition or deletion of one repeat unit (Table 1), which is typically the result of the slipped-strand mispairing mechanism in other bacteria (Levinson & Gutman, 1987).

Secondly, homologous recombination is likely responsible for much of the TTR variation in the MG191 and M192 repeat regions (MG191-EF1, MG191-EF2 and MG192-L) as well as in some MgPars, as a result of recombination between MG191 or M192 and MgPars that contain different numbers of repeat units or between MgPars, as has been noted in previous studies (Ma *et al.*, 2007). This possibility is supported by the presence of long stretches of constant sequences flanking the TTR region between MG191 or MG192 and MgPars as exemplified in Supplementary Fig. S3 for the TTR variation in the MG192-L region in specimens from patient 64. These constant sequences could serve as anchors for recombination reactions. The MG192 variants containing six or seven AGT units might have resulted from recombination of the MG192 variant containing five AGT units with MgPar 2, 8 or 9 containing six or seven AGT units. Importantly, such recombination may occur alone or together with slipped-strand mispairing.

Thirdly, site-specific recombination and additional, unidentified mechanisms may be involved in the TTR variation of some regions of the MgPa operon and MgPars. For example, in the MG192-L TTR region AGT repeats were found in all strains tested except for three strains (M2321, M2341 and M6280) which contained TCT repeats (Supplementary Fig. S1). Such variation obviously cannot be explained by slipped-strand mispairing. Furthermore, we sequenced the MG192-L homologous region in MgPars

2, 8 and 9 of strains M2321 and M2341 (other MgPars have no homology to this region) and did not find TCT repeats (Table 2), suggesting that the occurrence of TCT repeat tracts in these strains did not result from homologous recombination with MgPars. Given that the TTRs in the MG191-C and MG191-EF1 regions are exclusively made up of TCT repeats (Supplementary Fig. S1), we speculate a possibility of translocation of the TCT repeats between different TTR regions by a mechanism independent of homologous recombination. Another example is a likely translocation of the TCTⁿ-ACTⁿ repeats from MgPar 9 (MG191-EF homologous region) to MgPar 8 (MG192-L homologous region) in the patient 199 specimens (Supplementary Fig. S4). Together, these two examples suggest that other mechanisms for generation of TTR variability may exist, including site-specific recombination or recombination with additional, unidentified MgPar regions. Further insight into the mechanisms of TTR and MgPar variability will likely be revealed by complete genome sequencing of additional *M. genitalium* strains.

Given the small size of the *M. genitalium* genome, the maintenance of abundant TTRs in the MgPa operon and MgPars is likely to be of functional importance. It is worthy of note that, despite the abundance of TTRs in MgPa and MgPars, we did not find other tandem repeats containing repeat units other than multiples of three nucleotides (such as dinucleotides, tetranucleotides, pentanucleotides, etc., whose copy number changes would lead to abrogation of the reading frame and disruption of protein translation). The TTR units present in the MgPa operon are primarily composed of either AGT or TCT as shown in this study (Table 1). Other units, including TCC, TCA and ACT, are also present in the MG191-EF1 region of the MgPa operon and in its homologous region in MgPars (Supplementary Fig. S2), which are likely to recombine into the MgPa operon (Iverson-Cabral *et al.*, 2006, 2007; Jensen, 2006; Ma *et al.*, 2007; Peterson *et al.*, 1995). Interestingly, all such repeat units encode either serine (TCT, TCC and TCA) or threonine (ACT). The strict preservation of Ser/Thr residues in five regions of the MgPa operon suggests an important role of these amino acids in maintaining the structure of the MG191 and MG192 proteins encoded by this operon. Changes in the repeat copy number of TTRs in MgPa operon do not interrupt the reading frame but result in heterogeneity in the size of the poly(Ser/Thr) tract. The functional significance of the Ser/Thr-rich repetitive motifs in the MgPa operon is unknown. They could presumably serve as flexible spacer regions to optimize protein interactions, as has been proposed for polyserine domains in other organisms (Hall *et al.*, 1989; Howard *et al.*, 2004). They may also provide sites for protein modifications such as glycosylation and phosphorylation as has been found in studies of Ser/Thr-rich proteins in yeast (Bowen & Wheals, 2006). In addition, variation in the size of the poly(Ser/Thr) tract may increase antigenicity of the MG191 and MG192 proteins (Iverson-Cabral *et al.*, 2006, 2007; Jensen, 2006; Ma *et al.*, 2007, 2010; Peterson *et al.*, 1995).

In summary, we have identified abundant TTRs in the MgPa operon and MgPars, which are variable in the repeat copy number and/or in the repeat unit sequence. There are multiple potential mechanisms involved in TTR variation. Extensive TTR variation may confer maximal variation to the MgPa operon and enhance antigenic variation, thus optimizing the organism's adhesion, colonization and immune evasion.

ACKNOWLEDGEMENTS

This work was supported by the US Department of Defense (grant W81XWH-08-1-0676), a Gulf South Sexually Transmitted Infections/Topical Microbicide Cooperative Research Center grant from the NIH-NIAID (5 U19 AI061972), 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. Part of this work was presented at the 17th Congress of the International Organization for Mycoplasmaology, 6–11 July 2008, Tianjin, China.

REFERENCES

- Bowen, S. & Wheals, A. E. (2006). Ser/Thr-rich domains are associated with genetic variation and morphogenesis in *Saccharomyces cerevisiae*. *Yeast* **23**, 633–640.
- Charlesworth, B., Sniegowski, P. & Stephan, W. (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **371**, 215–220.
- Clausen, H. F., Fedder, J., Drasbek, M., Nielsen, P. K., Toft, B., Ingerslev, H. J., Birkelund, S. & Christiansen, G. (2001). Serological investigation of *Mycoplasma genitalium* in infertile women. *Hum Reprod* **16**, 1866–1874.
- Dallo, S. F. & Baseman, J. B. (1991). Adhesin gene of *Mycoplasma genitalium* exists as multiple copies. *Microb Pathog* **10**, 475–480.
- Di Prospero, N. A. & Fischbeck, K. H. (2005). Therapeutics development for triplet repeat expansion diseases. *Nat Rev Genet* **6**, 756–765.
- Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., Bult, C. J., Kerlavage, A. R., Sutton, G. & other authors (1995). The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**, 397–403.
- Glass, J. I., Assad-Garcia, N., Alperovich, N., Yooseph, S., Lewis, M. R., Maruf, M., Hutchison, C. A., III, Smith, H. O. & Venter, J. C. (2006). Essential genes of a minimal bacterium. *Proc Natl Acad Sci U S A* **103**, 425–430.
- Hall, J., Hazlewood, G. P., Huskisson, N. S., Durrant, A. J. & Gilbert, H. J. (1989). Conserved serine-rich sequences in xylanase and cellulase from *Pseudomonas fluorescens* subspecies *cellulosa*: internal signal sequence and unusual protein processing. *Mol Microbiol* **3**, 1211–1219.
- Hamasuna, R., Osada, Y. & Jensen, J. S. (2005). Antibiotic susceptibility testing of *Mycoplasma genitalium* by TaqMan 5' nuclease real-time PCR. *Antimicrob Agents Chemother* **49**, 4993–4998.
- Hamasuna, R., Osada, Y. & Jensen, J. S. (2007). Isolation of *Mycoplasma genitalium* from first-void urine specimens by coculture with Vero cells. *J Clin Microbiol* **45**, 847–850.
- Hjorth, S. V., Björnelius, E., Lidbrink, P., Falk, L., Dohn, B., Berthelsen, L., Ma, L., Martin, D. H. & Jensen, J. S. (2006). Sequence-based typing of *Mycoplasma genitalium* reveals sexual transmission. *J Clin Microbiol* **44**, 2078–2083.
- Howard, M. B., Ekborg, N. A., Taylor, L. E., Hutcheson, S. W. & Weiner, R. M. (2004). Identification and analysis of polyserine linker domains in prokaryotic proteins with emphasis on the marine bacterium *Microbulbifer degradans*. *Protein Sci* **13**, 1422–1425.
- Htun, H. & Dahlberg, J. E. (1989). Topology and formation of triple-stranded H-DNA. *Science* **243**, 1571–1576.
- Hu, P. C., Schaper, U., Collier, A. M., Clyde, W. A., Jr, Horikawa, M., Huang, Y. S. & Barile, M. F. (1987). A *Mycoplasma genitalium* protein resembling the *Mycoplasma pneumoniae* attachment protein. *Infect Immun* **55**, 1126–1131.
- Iverson-Cabral, S. L., Astete, S. G., Cohen, C. R., Rocha, E. P. & Totten, P. A. (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.
- Iverson-Cabral, S. L., Astete, S. G., Cohen, C. R. & Totten, P. A. (2007). *mgpB* and *mgpC* sequence diversity in *Mycoplasma genitalium* is generated by segmental reciprocal recombination with repetitive chromosomal sequences. *Mol Microbiol* **66**, 55–73.
- Jensen, J. S. (2006). *Mycoplasma genitalium* infections. Diagnosis, clinical aspects, and pathogenesis. *Dan Med Bull* **53**, 1–27.
- Jensen, J. S. (2009). Single-dose azithromycin treatment for *Mycoplasma genitalium*-positive urethritis: best but not good enough. *Clin Infect Dis* **48**, 1655–1656.
- Jensen, J. S., Hansen, H. T. & Lind, K. (1996). Isolation of *Mycoplasma genitalium* strains from the male urethra. *J Clin Microbiol* **34**, 286–291.
- Jensen, J. S., Bradshaw, C. S., Tabrizi, S. N., Fairley, C. K. & Hamasuna, R. (2008). Azithromycin treatment failure in *Mycoplasma genitalium*-positive patients with nongonococcal urethritis is associated with induced macrolide resistance. *Clin Infect Dis* **47**, 1546–1553.
- Levinson, G. & Gutman, G. A. (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol* **4**, 203–221.
- Ma, L. & Martin, D. H. (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.
- Ma, L., Jensen, J. S., Myers, L., Burnett, J., Welch, M., Jia, Q. & Martin, D. H. (2007). *Mycoplasma genitalium*: an efficient strategy to generate genetic variation from a minimal genome. *Mol Microbiol* **66**, 220–236.
- Ma, L., Taylor, S., Jensen, J. S., Myers, L., Lillis, R. & Martin, D. H. (2008). Short tandem repeat sequences in the *Mycoplasma genitalium* genome and their use in a multilocus genotyping system. *BMC Microbiol* **8**, 130.
- Ma, L., Jensen, J. S., Mancuso, M., Hamasuna, R., Jia, Q., McGowin, C. L. & Martin, D. H. (2010). Genetic variation in the complete MgPa operon and its repetitive chromosomal elements in clinical strains of *Mycoplasma genitalium*. *PLoS ONE* **5**, e15660.
- McGowin, C. L., Spagnuolo, R. A. & Pyles, R. B. (2010). *Mycoplasma genitalium* rapidly disseminates to the upper reproductive tracts and knees of female mice following vaginal inoculation. *Infect Immun* **78**, 726–736.
- Morrison-Plummer, J., Lazzell, A. & Baseman, J. B. (1987). Shared epitopes between *Mycoplasma pneumoniae* major adhesin protein P1 and a 140-kilodalton protein of *Mycoplasma genitalium*. *Infect Immun* **55**, 49–56.
- Peterson, S. N., Bailey, C. C., Jensen, J. S., Borre, M. B., King, E. S., Bott, K. F. & Hutchison, C. A., 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.

Richard, G. F., Kerrest, A. & Dujon, B. (2008). Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiol Mol Biol Rev* **72**, 686–727.

Rocha, E. P. & Blanchard, A. (2002). Genomic repeats, genome plasticity and the dynamics of *Mycoplasma* evolution. *Nucleic Acids Res* **30**, 2031–2042.

Svenstrup, H. F., Jensen, J. S., Gevaert, K., Birkelund, S. & Christiansen, G. (2006). Identification and characterization of

immunogenic proteins of *Mycoplasma genitalium*. *Clin Vaccine Immunol* **13**, 913–922.

Tully, J. G., Cole, R. M., Taylor-Robinson, D. & Rose, D. L. (1981). A newly discovered mycoplasma in the human urogenital tract. *Lancet* **317**, 1288–1291.

van Belkum, A., Scherer, S., van Alphen, L. & Verbrugh, H. (1998). Short-sequence DNA repeats in prokaryotic genomes. *Microbiol Mol Biol Rev* **62**, 275–293.

Verstrepen, K. J., Jansen, A., Lewitter, F. & Fink, G. R. (2005). Intragenic tandem repeats generate functional variability. *Nat Genet* **37**, 986–990.