



## Restriction-modification systems in *Mycoplasma* spp

Marcelo Brocchi<sup>1</sup>, Ana Tereza Ribeiro de Vasconcelos<sup>2</sup> and Arnaldo Zaha<sup>3</sup>

<sup>1</sup>*Departamento de Microbiologia e Imunologia, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, SP, Brazil.*

<sup>2</sup>*Laboratório Nacional de Computação Científica, Petrópolis, RJ, Brazil.*

<sup>3</sup>*Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil.*

### Abstract

Restriction and Modification (R-M) systems are present in all *Mycoplasma* species sequenced so far. The presence of these genes poses barriers to gene transfer and could protect the cell against phage infections. The number and types of R-M genes between different *Mycoplasma* species are variable, which is characteristic of a polymorphism. The majority of the CDSs code for Type III R-M systems and particularly for methyltransferase enzymes, which suggests that functions other than the protection against the invasion of heterologous DNA may exist. A possible function of these enzymes could be the protection against the invasion of other but similar R-M systems. In *Mycoplasma hyopneumoniae* strain J, three of the putative methyltransferase genes were clustered in a region forming a genomic island. Many R-M CDSs were mapped in the vicinity of transposable elements suggesting an association between these genes and reinforcing the idea of R-M systems as mobile selfish DNA. Also, many R-M genes present repeats within their coding sequences, indicating that their expression is under the control of phase variation mechanisms. Altogether, these data suggest that R-M systems are a remarkable characteristic of *Mycoplasma* species and are probably involved in the adaptation of these bacteria to different environmental conditions.

*Key words:* restriction-modification systems, *Mycoplasma* spp, genomes.

Received: April 4, 2006; Accepted: October 5, 2006.

### Introduction

With the sequencing of several mycoplasma genomes, an extensive analysis of differences in gene content among different species or strains has been carried out, and the possible consequences of the differences in the biology of these organisms have been discussed (Himmelreich *et al.*, 1997; Rocha and Blanchard, 2002; Vasconcelos *et al.*, 2005). Possible mechanisms regulating gene expression have also been analyzed. Since mycoplasmas possess only one sigma factor and a low number of transcription factors, other mechanisms are proposed for regulation of gene expression. Phase variation is an important mechanism for generating protein variability in several microorganisms, including mycoplasmas (Dybvig *et al.*, 1998; Gumulak-Smith *et al.*, 2001). Interestingly, short tandem repeated sequences have been observed in genes encoding proteins of the R-M system in bacteria (Ryan and Lo, 1999; Saunders *et al.*, 1998; De Bolle *et al.*, 2000).

Three main groups of restriction-modification (R-M) enzymes, called Types I, II and III, are found in bacteria. Type I R-M enzymes are multi-functional complexes composed of subunits for specificity (HsdS/S), modification (HsdM/M) and restriction (HsdR/R). The subunits do not act alone, but as complexes (M<sub>2</sub>S) or (R<sub>2</sub>M<sub>2</sub>S) that catalyze many different functions, such as recognition of and binding to a specific DNA sequence, methylation of specific adenine residues within that same sequence, DNA translocation coupled to ATP hydrolysis and, finally, DNA cleavage (Bourniquel and Bickle, 2002; Tock and Dryden, 2005). The HsdS subunit is responsible for the specificity of the Type I R-M system. The complex (M<sub>2</sub>S) is composed of two HsdM and one HsdS subunits and catalyzes the transfer of a methyl group from S-adenosyl methionine (AdoMet) to a specific adenine of the recognition site. The HsdR subunit as part of the complex R<sub>2</sub>M<sub>2</sub>S is able to cleave DNA, but also has the ATPase and DNA translocation activities characteristic of Type I restriction enzymes. The Type II R-M enzymes usually act as monomers, dimers or even tetramers. The restriction endonucleases recognize specific DNA sequences close to or at the recognition site and cleave the DNA. The methylases trans-

fer a methyl group from the donor (AdoMet) directly to double-stranded DNA and form m4C, m5C or m6A. Type III R-M enzymes are composed of two protein subunits that function either in DNA recognition and modification (Mod) or restriction (Res). To cleave DNA, both subunits are necessary, as well as an absolute requirement for ATP hydrolysis. After recognition of the DNA sequence, an ATP-dependent DNA translocation occurs, as with the Type I restriction enzyme. The Mod subunit can function independently of the Res subunit to methylate DNA to form m6A (reviewed by Bourniquel and Bickle, 2002; Tock and Dryden, 2005).

The three different types of enzymes are present in mycoplasmas, although not all species present all different types. Interestingly, phase variation was observed in some genes encoding Type I and Type III enzymes (Dybvig et al., 1998; Rocha and Blanchard, 2002). Although the importance of this mechanism for the biology of mycoplasmas is not clear, a possible relationship to virulence has been proposed for *Mycoplasma pulmonis* (Gumulak-Smith et al., 2001).

In the present work we have analyzed the R-M system in ten different species of mycoplasmas, including two different strains of *Mycoplasma hyopneumoniae*. With the exception of *Mycoplasma genitalium*, the Type I R-M system is present in all species analyzed. Type II and Type III R-M are the less frequent, being present in only four species. Variation in the number of short repeated sequences was observed in genes encoding Type I and Type III R-M systems.

### Materials and Methods

The sequence of three complete mycoplasma genomes, a pathogenic (7448) and a non-pathogenic (J) strain of the swine pathogen *M. hyopneumoniae* and a strain (53) of the avian pathogen *Mycoplasma synoviae*, was recently described (Vasconcelos et al., 2005).

Genome assembling, annotation and comparative *in silico* analyses were carried out by the Brazilian National Genome Sequencing Consortium (BRGENE) and the Southern Network for Genome Analysis (PIGS). The sequences were assembled by using PHRED/PHRAP/CONSED. The annotation was carried out by using the system for automated bacterial integrated annotation (Almeida et al., 2004). The GC content of specific genomic regions was determined by using the Artemis software (Rutherford et al., 2000).

After annotation, we systematically analyzed the genomes of these bacteria as well as the genomes of the other *Mycoplasma* species sequenced so far, for the presence of CDSs encoding putative R-M genes and categorized each CDS into R-M Types according to sequence analysis and BLASTp searches. For some sequences, BLASTp searches were also carried out using REBASE<sup>R</sup>, The Restriction Enzyme Database and REBASE Genomes.

R-M systems have been frequently associated with mobile DNA elements as insertion sequences (IS) and transposons (Kobayashi, 2001). Thus, to evaluate the possible involvement of these elements in the transfer of R-M genes between *Mycoplasma* strains or species, we also investigated the presence of CDSs in the vicinity of the R-M genes encoding transposase enzymes using the NCBI Entrez Genome database).

### Results

The presence of sequences related to enzymes of the R-M system was observed in all mycoplasma genomes analyzed (Table 1). In the Pneumoniae clade, all analyzed genomes present sequences related to the Type I R-M system. Sequences related to the Type II R-M system are present in *M. genitalium*, *Mycoplasma pneumoniae*, *Ureaplasma urealyticum*, and *Mycoplasma penetrans*. The Type III R-M system related sequences were observed only in *M. penetrans* and *U. urealyticum*. In *M. genitalium* only two sequences related to the R-M system were detected, one encoding a Type I specificity subunit (MG438) and another encoding a Type II methyltransferase (MG184), but no CDS encoding restriction enzyme was observed.

*M. pneumoniae* presents nine regions containing sequences related to the Type I R-M system, but only one containing the complete system (encoding specificity, modification and restriction subunits). However, some sequences were truncated. The other regions contain sequences related to the specificity subunit. Only one sequence related to methyltransferase of Type II R-M was observed. No sequence related to Type III was detected in the *M. pneumoniae* genome.

**Table 1** - Occurrence of sequences related to R-M systems Type I, II and III in different species of mycoplasmas. The number of different regions containing R-M system sequences is indicated. CDSs with some identity to R-M sequences but not listed in REBASE<sup>R</sup> were not included.

Clade	Species/strain	Type I	Type II	Type III
Pneumoniae	<i>M. genitalium</i> G37	1	1	No
	<i>M. pneumoniae</i> M129	9	1	No
	<i>U. urealyticum</i> serovar 3	2	1	1
	<i>M. penetrans</i> HF-2	4	5	1
	<i>M. gallisepticum</i> R	1	No	No
Hominis	<i>M. pulmonis</i> strain UAB CTIP	3	2	2
	<i>M. mobile</i> 163K	1	2	1
	<i>M. hyopneumoniae</i> J	1	3	5
	<i>M. hyopneumoniae</i> 7448	1	4	3
	<i>M. hyopneumoniae</i> 232	1	3	3
	<i>M. synoviae</i>	No	No	1
Mycoides	<i>M. mycoides</i> subsp. <i>mycoides</i> SC str. PG1	No	5	No

In *U. urealyticum* we identified a region containing sequences related to the Type I R-M system corresponding to specificity, modification, and restriction subunits, but the sequences were truncated. One copy each of Type II and Type III methyltransferase were also identified. *M. penetrans* presented a region containing a complete Type I R-M system. Four regions containing Type II methyltransferase were observed. A genomic region contained a sequence encoding Type II restriction enzyme and methyltransferase. Only one region encoding a truncated form of Type III methyltransferase was identified. *Mycoplasma gallisepticum* presented only a set of sequences related to the Type I R-M system.

*Mycoplasma mycoides* subsp. *mycoides* Sc str. PG1 of the Mycooides-Entomoplasmataceae clade presents five regions containing sequences related to Type II RM systems (Table 1). Two of them contain CDSs encoding the restriction and modification enzymes. One of these R-M complexes is similar to CDSs of *Bacillus clausii* and the other presents identity to the *Sau96I* of *Staphylococcus aureus*. The other three regions are composed of Type II methyltransferase CDSs. One of these regions is composed of two CDSs, one encoding an adenine-specific methyltransferase and the other for a cytosine-specific methyltransferase, both with identities of about 58% to *Helicobacter pylori* CDSs. The other two regions are composed of orphan CDSs, one with identity to another methyltransferase of *M. mycoides* and the other with identity to a putative methyltransferase of NAV1 phage (Tu *et al.*, 2002). Interestingly, two of the *M. mycoides* Type II RM complexes are located near mobile DNA elements as ISs, which suggests a mobility capacity for these R-M CDSs.

The Hominis clade is heterogeneous regarding the presence of R-M Types and all three Types of R-M systems are present (Table 1). *M. pulmonis* strain UAB CTIP presents three Type I, two Type II and two Type III R-M systems. The three Type I CDSs are organized in operons containing all the subunits (*hsdR*, *hsdM*, *hsdS*) of the R-M complex. *M. pulmonis* contains an *hsdS1* gene belonging to the *MpuUI* system that is regulated by a site-specific DNA inversion event (Dybvig and Voelker, 1996). The two Type II R-M sequences are represented by cytosine-specific methyltransferases. One of the regions containing the Type III R-M system is formed by three contiguous methyltransferase CDSs and a CDS encoding a restriction enzyme. Interestingly, there are several copies of the insertion sequence IS1138 (Bhugra and Dybvig, 1993) in the vicinity of these CDSs. The other region contains only one CDS encoding a putative methyltransferase.

*Mycoplasma mobile* strain 163K contains one region for Type I, two regions for Type II, and one region for Type III R-M systems. The Type I R-M complex is composed of five contiguous CDSs, one *hsdR*, two *hsdM*, and two *hsdS* organized in an interesting way. In addition to the common organization (*hsdR*, *hsdM*, *hsdS*), there are two additional

CDSs, one encoding HsdM and other for HsdS, which are localized downstream of the operon. The identity between the HsdM putative methyltransferases is high (81%), which would suggest a possible gene duplication event. However, the lower identity between the HsdS specificity subunits (55%) does not support this conclusion. The two Type II CDSs present identity to DpnI of *Streptococcus pneumoniae* (44%), to a methyltransferase of *M. penetrans* (34%), and to a eukaryotic virus of *Chlorella*, the NY2A (39%) virus (Zhang *et al.*, 1998), respectively.

The genomes of two strains of *M. hyopneumoniae* and one strain of *M. synoviae* were recently sequenced by our group (Vasconcelos *et al.*, 2005). Thus, in the following sections we will concentrate our analysis on these strains. The CDSs encoding putative R-M systems present in these strains are depicted in Table 2. The R-M systems present in *M. hyopneumoniae* strain 232 (Minion *et al.*, 2004) have been compared with the putative RM CDSs of the other two strains of *M. hyopneumoniae*.

### *M. hyopneumoniae*

*M. hyopneumoniae* strains 7448 and J present one CDS each (MHP614 and MHJ0615), which were classified in REBASE<sup>R</sup> as Type I and Type II enzymes, respectively, but which also present similarity to the cytosine-specific DNA methyltransferase gene. These strains also have one CDS each (MHP0622 and MHJ0623) with identity to Type II enzymes and also sharing identity with the DNA adenine methyltransferase enzyme. Homologous genes were also found in the genome of *M. hyopneumoniae* strain 232 (Table 2). Sequence comparisons between them revealed a high identity, varying from 97% to 99%. These CDSs present similarity to the *dcm* and *dam* genes of *Escherichia coli* (Lobner-Olesen *et al.*, 2005).

*M. hyopneumoniae* strain 7448 presents one CDS (MHP0626) which is not listed in the REBASE<sup>R</sup> but which encodes a putative Type I methyltransferase (*hsdM* homolog) with high identity to CDSs of *M. hyopneumoniae* strains J (99%) and 232 (98%). *M. hyopneumoniae* strain 7448 also presents one set of the complete Type II R-M system encoded by two contiguous CDSs (MHP0291 and MHP0292) (Table 2). These two CDSs are separated by twelve nucleotides only and present 54% and 37% of identity to putative Type II methyltransferase and restriction genes of *H. pylori* strains J99 and 26695, respectively (Alm *et al.*, 1999; Tomb *et al.*, 1997). The MHP0292 CDS also exhibits identity to a gene coding for an S subunit of *Haemophilus somnus* (data not shown). This observation would suggest that this CDS is an S subunit of Type I R-M. However, considering its similarity to the *H. pylori* Type II restriction gene and its proximity to the MHP0291 CDS, we concluded that it represents a Type II restriction gene. Another CDS (MHP0327) was found with identity to the Type II methyltransferase (Table 2). Interestingly, this CDS

**Table 2** - CDSs encoding enzymes of the restriction and modification systems in *Mycoplasma hyopneumoniae* strains J and 7448 and *M. synoviae* strain 53.

<i>Mycoplasma</i> species	Strain	R-M System	CDS Designation / Type	Best Hit (% Identity)	
<i>M. hyopneumoniae</i>	7448	Type I	MHP0626 / M*	MHJ0627 (99)	
			MHP0614 / M	<i>M. hyopneumoniae</i> 232 (97)	
		Type II	MHP0622 / M	<i>M. hyopneumoniae</i> 232 (99)	
			MHP0327 / M	MHJ0319 (98)	
			MHP0291 / M	<i>H. pylori</i> J99 (54)	
			MHP0292 / R	<i>H. pylori</i> 26695 (53)	
		Type III	MHP0388 / M	MHJ0399 (98)	
			MHP0410 / M	MHJ0423 (98)	
			MHP0386 / M*	<i>M. hyopneumoniae</i> 232 (97)	
			MHP0316 / M	<i>M. hyopneumoniae</i> 232 (94)	
		J	Type I	MHJ0627 / M*	MHP0626 (99)
				MHJ0283 / R	<i>H. pylori</i> 26695 (45)
	Type II		MHJ0623 / M	<i>M. hyopneumoniae</i> 232 (99)	
			MHJ0319 / M	MHP0327 (98)	
			MHJ0615 / M	<i>M. hyopneumoniae</i> 232 (98)	
	Type III		MHJ0399 / M	MHJ0388 (98)	
			MHJ0423 / M	<i>M. hyopneumoniae</i> 232 (97)	
			MHJ0308 / M	<i>M. hyopneumoniae</i> 232 (95)	
		MHJ0382 / M	MHJ0383 (75)		
		MHJ0383 / M	MHJ0382 (75)		
<i>M. synoviae</i>	53	Type III	MS0161	<i>M. hyopneumoniae</i> 232 (55)	

CDS, Coding sequence; M, Methyltransferase enzyme; R, Restriction enzyme; \*, CDSs not present in the REBASE but with partial identity to restriction enzymes.

presents a high identity (98%) to the MHJ0319 CDS of strain J.

Four CDSs encoding Type III methyltransferases (*mod* genes) were found in the strain 7448 genome (Table 2). Three of these CDSs (MHP0386, MHP0388 and MHP0410) were clustered in the genome. However, the CDS MHP0316 was found away from the other Type III genes. The MHP0386 CDS is not listed in REBASE. The MHP0316 CDS present high identity with an ORF found in the *M. hyopneumoniae* strain 232 genome (94% identity) and with CDSs found in *M. synoviae* (57% identity) and *M. hyopneumoniae* strain J (55% identity). CDSs with high identity to the MHP0410 were also found in the *M. hyopneumoniae* strains J and 232 (98 and 92%, respectively). CDSs with high similarity to MHP0386 and MHP0388 are present in *M. hyopneumoniae* strain 232. Interestingly, MHP0388 is probably involved in the integration of a Genomic Island (GI) specific for the *M. hyopneumoniae* strain J. This region is described below. This conclusion is supported by the observed high identity between MHP0388 and MHJ0399 (98%). The genome of strain 7448 is devoid of Type I and Type III restriction genes.

*M. hyopneumoniae* strain J contains one CDS encoding a Type I specificity subunit (MHJ0283) represented by a truncated and probably not functional gene, and seven

CDSs encoding methylases, one Type I (MHJ0627), one Type II (MHJ0319) and five Type III (MHJ308, MHJ0382, MHJ0383, MHJ0399, MHJ0423), in addition to the Type II methyltransferases (MHJ0615, MHJ0623) discussed previously (Table 2). The MHJ0627 CDS is not listed in REBASE. Some of these CDSs (MHJ0319, MHJ0399, MHJ0627) exhibit high identity to CDSs present in strains 7448 and 232. These genes are mapped in similar positions, indicating that they were acquired prior to the divergence of these strains from a common ancestor. MHJ0382, MHJ383 and MHJ0399 CDSs are clustered in a specific genomic region of *M. hyopneumoniae* strain J forming a Genomic Island (GI). This region will be described further in this text.

All CDSs encoding Type III R-M methyltransferases present 5'-AG repeats of lengths ranging from 10 to 22 (data not shown). One of the CDSs (MH12674) presents a block of imperfect AG repeats. With the exception of one copy, all others present a frameshift at the region encoding the amino-terminal portion of the methyltransferase, probably making impossible the production of a functional methyltransferase.

### *M. synoviae*

The *M. synoviae* strain 53 presents only a CDS (MS0161) encoding a methyltransferase of the Type III R-M system (Table 2). No repetition was found in this

gene. This strain also contains a putative cytosine-specific methyltransferase (MS0440).

### R-M systems and transposases

Many of the CDSs encoding R-M enzymes are mapped near transposase genes (Table 3). The transposase gene (MHS0166) found in the vicinity of the putative Type II Methylase of *M. synoviae* exhibits similarity to IS1630, a transposable element of *Mycoplasma fermentans* that belongs to the IS30 family of ISs (Calcutt *et al.*, 1999). A transposable element of the IS4 family called ISMhp1 (GenBank AF272977) was found close to many R-M genes in both *M. hyopneumoniae* strains. Strain J also contains ISs with similarity to the transposable elements IS1138 and IS1221 of the IS3 family (Zheng and McIntosh, 1995) in the vicinity of some R-M genes.

### *M. hyopneumoniae* strain J Genomic Island (MhJGI)

The *M. hyopneumoniae* strain J contains a specific region absent from strains 7448 and 232, as determined by the alignment of their genomes. This region is called MhJGI for “*Mycoplasma hyopneumoniae* (strain) J Genomic Island”. There are seventeen annotated CDSs in this GI but an accurate analysis of this region revealed the presence of mutations creating intragenic stop codons or frame shifts, which indicated that some individual CDSs are indeed parts of the same gene. Based on this observation, this GI is made up of thirteen CDSs encompassing 16 kb of the bacterial genome (Table 4). The comparison of this island with the *M. hyopneumoniae* 7448 genome reveals that it is flanked by the prolyl-tRNA synthetase (*proS*) gene (MHJ0381) at one end and by a Type III methylase (MHJ0399) CDS at the other end. This last CDS exhibits a high identity to a CDS (MHP0388) found in the genome of strain 7448, as pointed out previously. Interestingly, this GI is also composed of two other Type III methylases (Table 4) of the strain J genome. In this genomic region, there are also CDSs encoding transposase genes (MHJ0385, MHJ0386, MHJ0387,

MHJ0390, MHJ0391, and MHJ0392) with similarity to the IS1138 insertion sequence found in *M. pulmonis* (Chambaud *et al.*, 2001). However, these IS3-like elements appear to be not functional since their coding sequence is disrupted by stop codons and frameshifts. There are two CDSs (MHJ0397 and MHJ0398) that probably represent parts of the same CDS encoding a subtilisin-like serine protease with similarity to a CDS of *M. hyopneumoniae* 232. Other CDSs (MHJ0394 and MHJ0395) exhibit similarity to membrane proteins found in erythrocytes infected by *Plasmodium yoelii* (PFEMP3) or *Plasmodium falciparum* (3D7) parasites. There is also a hypothetical CDS (MHJ0389) coding for ABC transporter proteins characterized by the presence of ATP/GTP binding motifs and a CDS (MHJ0384) with motifs (DEXHc and DEXDc) found in helicase proteins (Table 4). The GC content (24.42%) of this region, including the transposase genes, is lower than that observed for the overall genome of *M. hyopneumoniae* (28%), suggesting that it was acquired by lateral gene transfer from another microorganism.

### Discussion

Bacteria have developed many biological strategies to live in different environments. The presence of Restriction-Modification systems (R-M) is considered one of these strategies (Murray, 2002). The genome of many species of bacteria sequenced so far is homologous to R-M genes, which are associated with important biological functions (reviewed by Bujnicki, 2001; Kobayashi, 2001).

R-M systems are present in virtually all mycoplasma species so far sequenced (Table 1). The presence of these genes poses barriers to gene transfer and can protect the cell against bacteriophage (phage) infections (reviewed by Kobayashi, 2001). Indeed, a variety of phages were described to infect mycoplasma species (reviewed by Razin *et al.*, 1998). Therefore, the expression of restriction enzymes by mycoplasmas, at least in some cases, could be important for resistance to phage infection. R-M systems may also be

**Table 3** - R-M CDSs of *M. hyopneumoniae* strains J and 7448 and *M. synoviae* strain 53 linked to transposable elements.

Restriction-modification		Distance (bp)	Transposase		Reference
CDS	Type		CDS	Type	
MS0161	III	3497	MS0166	IS1630	Calcutt <i>et al.</i> , 1999
MHJ0308	III	2975	MHJ0310	ISMhp1	GenBank AF272977
MHJ0319	II	8740	MHJ0329	ISMhp1	GenBank AF272977
MHJ0382	III	5042	MHJ385/386	IS1138	Bhugra and Dybvig, 1993
MHJ0383	III	3263	MHJ385/386	IS1138	Bhugra and Dybvig, 1993
MHJ0399	III	5623	MHJ0391	IS1221	Zheng and McIntosh, 1995
MHJ0627	I	176	MHJ0626	ISMhp1	GenBank AF272977
MHP0316	III	6910	MHP0323	ISMhp1	GenBank AF272977
MHP0626	I	176	MHP625	ISMhp1	GenBank AF272977

CDS, coding sequence; bp, base pair.

**Table 4** - The *Mycoplasma hyopneumoniae* strain J Genomic Island (MHJGI).

CDS	GC (%)	Designation	Hits found
MHJ0381	30.90	Known protein	Prolyl-tRNA synthetase of <i>M. pulmonis</i>
MHJ0382	27.26	Known protein	Type III Restriction-Modification System: Methylase ( <i>M. pulmonis</i> )
MHJ0383	25.87	Known protein	Type III Restriction-Modification System: Methylase ( <i>M. pulmonis</i> )
MHJ0384	23.80	Known protein	ATP/GTP-binding protein: Helicase domains ( <i>U. urealyticum</i> )
MHJ0385	23.25	Known protein	Putative transposase for IS1138 element ( <i>M. pulmonis</i> )
MHJ0386	30.07	Known protein	Putative partial transposase for IS3 element ( <i>Streptococcus pneumoniae</i> )
MHJ0387	24.93	Known protein	Putative partial transposase for IS1138 element ( <i>M. pulmonis</i> )
MHJ0388	21.11	Hypothetical	Hypothetical protein of <i>M. hyopneumoniae</i>
MHJ0389	22.40	Hypothetical	ABC transporter with ATP/GTP-binding site motif A (P-loop)
MHJ0390	23.25	Known protein	Putative transposase for IS1138 element ( <i>M. pulmonis</i> )
MHJ0391	25.91	Known protein	Putative partial transposase for IS1138 element ( <i>M. pulmonis</i> )
MHJ0392	24.89	Known protein	Putative partial transposase for IS1221 element ( <i>M. hyopneumoniae</i> )
MHJ0393	20.55	Hypothetical	Hypothetical protein of <i>M. hyopneumoniae</i>
MHJ0394	20.69	Hypothetical	Hypothetical protein of <i>Plasmodium falciparum</i> 3D7
MHJ0395	23.87	Hypothetical	Erythrocyte membrane protein PFEMP3 <i>Plasmodium yoelii yoelii</i>
MHJ0396	21.95	Hypothetical	Hypothetical protein of <i>P. yoelii yoelii</i>
MHJ0397	27.75	Hypothetical	Subtilisin-like serine protease ( <i>M. hyopneumoniae</i> )
MHJ0398	22.85	C h p	Subtilisin-like serine protease ( <i>M. hyopneumoniae</i> and <i>M. pulmonis</i> )
MHJ0399	27.71	Known protein	Type III Restriction-Modification System: Methylase ( <i>M. hyopneumoniae</i> )

CDS, coding sequence; GC (%), percentage of GC content; C h p, Conserved hypothetical protein.

an effective barrier to gene transfer because any piece of non-methylated DNA is restricted (Dybvig and Voelker, 1996; Razin *et al.*, 1998). However, the majority of the R-M homologous genes found in the genome of *M. hyopneumoniae* are not restriction endonucleases but are CDSs encoding putative methylase proteins. The same situation was observed in the other species as well. *M. synoviae* presents only one CDS encoding a methylase. Thus, these CDSs must have biological functions other than protection against phage infection. Probably, the major role of these orphan methylases is to protect against the invasion of heterologous R-M systems encoding restriction enzymes that recognize the same or a very similar DNA sequence modified by the resident methylases.

Type III R-M sequences are common in the Hominis clade but are rare in the Pneumoniae clade and in *M. mycoides* as well (Table 1). The biological significance of this difference is unknown but could reflect the evolutionary history of these different clades.

Interestingly, various lines of evidence associate R-M genes with mobile DNA elements (Kobayashi *et al.*, 1999, reviewed by Kobayashi, 2001). Their horizontal transfer between distant groups of bacteria has been suggested by sequence comparison (Malone *et al.*, 1995) and by analysis of their codon usage and GC content (Nobusato *et al.*, 2000a). For instance, RM genes were found in a region of the *H. pylori* genome that is characterized by the presence of strain-specific genes and translocation events forming

the so-called plasticity zone (Alm *et al.*, 1999; Salama *et al.*, 2000). Based on these and other observations, some authors have considered R-M genes as selfish DNA units. Once transferred to and established in a cell clone, the R-M gene complex could be difficult to eliminate because its loss leads to cell death by the action of the restriction enzyme (Kobayashi, 2001). In *M. hyopneumoniae* and also in other species, such as *M. mycoides* and *M. pulmonis*, a considerable number of CDSs encoding methylases were mapped near transposable elements, reinforcing this association (Table 3). Our data suggest the mobility of R-M complex as a common event in the evolution of *M. hyopneumoniae*.

There is also considerable evidence that R-M complexes are involved in bacterial genome rearrangements. These rearrangements are a consequence of the generation of DNA fragments originated by the action of restriction enzymes in heterologous non-methylated DNA. Once in the cell, these DNA fragments could be the target of host enzymes that recombine them in the recipient genome (Chinen *et al.*, 2000; Nobusato *et al.*, 2000b; Kobayashi, 2001).

The MhJGI region has characteristics that support both these events, *e.g.*, DNA rearrangements and DNA mobility. There are at least two transposable elements in this region, suggesting that its acquisition could have been mediated by transposition. However, these IS3-like transposable elements do not appear to be functional since

their coding sequences are disrupted by stop codons or frameshift mutations. Notwithstanding, we can not rule out that these elements were functional when this island was acquired by strain J.

Sequence comparisons indicated that the MhJGI region was inserted between a CDS encoding a Type III methylase (MHJ0399) and the *proS* gene (MHJ0381) in the genome of the strain J ancestor. This observation raises the possibility that the MhJGI region was acquired by homologous recombination, as suggested by the presence of a homolog CDS in strain 7448 mapping in a similar genomic location. However, as discussed above, the presence of transposase genes in this region suggests an active process of transference, probably mediated by transposition. In the light of the involvement of R-M complexes with the DNA recombination process, we hypothesize that the R-M genes present in this island also could have had an active role in the process of insertion/recombination of this island in the host genome. So far, this is only a speculative idea. The GC content of this island is lower than that of the *M. hyopneumoniae* genome, which reinforces the heterologous origin of this region.

The biological function, if any, of the MhJGI region is also speculative. In addition to transposons and methylases, this island is composed of CDSs encoding putative ABC transporters, a possible DNA helicase, and putative membrane proteins. It is possible that this island encodes a transport or secretion apparatus, but if this is the case, its organization is different from the other types of bacterial secretion systems (reviewed by Remaut and Waksman, 2004). The R-M systems in MhJGI are incomplete because the restriction enzyme counterparts are lacking. The existence of orphan methylases is linked to mechanisms of protection against the invasion of heterologous R-M systems by means of conjugation, transformation, transduction, or conjugative transposition (reviewed by Kobayashi, 2001). This is probably the major function of the MhJGI region.

A remarkable characteristic of the Type III methylase CDSs in *M. hyopneumoniae* is the presence of sequence repeats. Simple repetitive sequences described in intergenic or coding sequences of several bacterial genomes can be involved in the on-off switching of associated gene products, generating phase, or antigenic variation. The high mutation rates in these repetitive sequences is due to slippage within the DNA repeats during replication, resulting in changes in the number of repeats (Henderson *et al.*, 1999). Phase variation has been described in genes involved in LPS biosynthesis, cell-surface-associated proteins and DNA restriction-modification systems (Saunders *et al.*, 1998), and can occur at the transcriptional or the translational level (de Vries *et al.* 2002). This kind of repeat indicates that these genes are possibly regulated by phase variation. Phase variation in the Type III R-M system has been observed in several bacterial species, including *M. pulmonis* (De Bolle *et al.*, 2000; de Vries *et al.*, 2002; Rocha and

Blanchard, 2002; Seib *et al.*, 2002) and also in *M. hyopneumoniae* (this paper).

The possible importance of phase variation in R-M systems for *M. hyopneumoniae* biology remains to be elucidated. The possible relationship of R-M systems to virulence has been described in *H. pylori* (Figueiredo *et al.*, 2000) and *M. pulmonis* (Gumulak-Smith *et al.*, 2001). The aforementioned authors demonstrated that differences in selective pressures in animal tissues affect the R-M gene activity in *M. pulmonis*. These same authors proposed that variations in R-M activity are important for the survival of mycoplasmas within the host. In the same way, the presence of repeats in Type III R-M CDSs suggests dynamism of R-M activity in *M. hyopneumoniae*, which is probably associated with the adaptation of this bacterium to different conditions found within the host.

The presence of CDSs homologous to the *dam* gene in *M. hyopneumoniae* strains and *dcm* gene in both *M. hyopneumoniae* and *M. synoviae* species points to methylation as an epigenetic event controlling several aspects of their biology, as is the case for other bacterial species (reviewed by Lobner-Olsen *et al.*, 2005). Meanwhile, further experimental analysis is needed to elucidate the biological role of these genes in mycoplasmas. This could be accomplished by the construction of isogenic mutants where those genes were deleted, followed by phenotypic analysis.

Recently, Xia (2003) suggested that the common ancestry of three mycoplasma species (*M. pulmonis*, *M. pneumoniae* and *M. genitalium*) has CpG-specific DNA methyltransferases that were lost in *M. pneumoniae* and *M. genitalium*. This author suggested that the absence of CpG-specific methyltransferases explained the higher percentage of CpG dinucleotides in these two species when compared with *M. pulmonis*. The presence of cytosine-specific DNA methyltransferase genes in *M. hyopneumoniae* strains is consistent with this hypothesis since we recently demonstrated that this species is phylogenetically related to *M. pulmonis* and more distantly associated with *M. pneumoniae* and *M. genitalium* (Vasconcelos *et al.*, 2005).

One remarkable characteristic of mycoplasma species is the lack of genome synteny, despite the presence of genes with high similarity (Dybvig and Voelker, 1996; Razin *et al.*, 1998; Rocha and Blanchard, 2002). This picture is suggestive of a dynamic genome where recombination plays an important role in the adaptability of this organism to its parasite mode of life. The presence of unique regions in the genome of *M. hyopneumoniae* suggests that lateral gene transfer is also important for the biology of this organism. The proximity of some R-M genes and transposable elements reinforces the proposition of the R-M systems as mobile, selfish DNA. Mycoplasmas are heterogeneous regarding the number of R-M genes present, which could reflect the development of different genomic structures and different routes of adaptation. The presence

of repeats in intergenic or intragenic R-M sequences suggests a dynamic process of gene regulation and expression, which could also be important for the adaptation of mycoplasmas to different environmental conditions. Altogether, these data suggest that R-M systems are a common characteristic of mycoplasma species, and that they probably have important roles in the genome organization and architecture, as well as in the adaptability of these bacteria to different niches found in their respective hosts.

## Acknowledgments

The strategic vision and enthusiastic support of the present and former staff members of the Ministério da Ciência e Tecnologia (MCT)/Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) is gratefully acknowledged. This work was undertaken by the Brazilian National Genome Program (Southern Network for Genome Analysis and Brazilian National Genome Project Consortium) with funding provided by MCT/CNPq and SCT/FAPERGS (RS).

## References

- Alm RA, Ling LS, Moir DT, King BL, Brown ED, Doig PC, Smith DR, Noonan B, Guild BC, deJonge BL, *et al.* (1999) Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* 397:176-180.
- Almeida LGP, Paixão R, Souza RC, Costa GC, Barrientos FJA, Santos MT, Almeida DF and Vasconcelos AT (2004). A system for automated bacterial (genome) integrated annotation-SABIA. *Bioinformatics* 20:2832-2833.
- Bhugra B and Dybvig K (1993) Identification and characterization of IS1138, a transposable element from *Mycoplasma pulmonis* that belongs to the IS3 family. *Mol Microbiol* 7:577-584.
- Bourniquel AA and Bickle TA (2002) Complex restriction enzymes: NTP-driven molecular motors. *Biochimie* 84:1047-1059.
- Bujnicki JM (2001) Understanding the evolution of restriction-modification systems: Clues from sequence and structure comparisons. *Acta Biochim Pol* 48:935-967.
- Calcutt MJ, Lavrrar JL and Wise KS (1999) IS1630 of *Mycoplasma fermentans*, a novel IS30-type insertion element that targets and duplicates inverted repeats of variable length and sequence during insertion. *J Bacteriol* 181:7597-7607.
- Chambaud I, Heilig R, Ferris S, Barbe V, Samson D, Galisson F, Moszer I, Dybvig K, Wroblewski H, Viari A, *et al.* (2001) The complete genome sequence of the murine respiratory pathogen *Mycoplasma pulmonis*. *Nucleic Acids Res* 29:2145-2153.
- Chinen A, Uchiyama I and Kobayashi I (2000) Comparison between *Pyrococcus horikoshii* and *Pyrococcus abyssi* genome sequences reveals linkage of restriction-modification genes with large genome polymorphisms. *Gene* 259:109-121.
- De Bolle X, Bayliss CD, Field D, van de Ven T, Saunders NJ, Hood DW and Moxon ER (2000) The length of a tetra-nucleotide repeat tract in *Haemophilus influenzae* determines the phase variation rate of a gene with homology to type III DNA methyltransferases. *Mol Microbiol* 35:211-222. Erratum in: *Mol Microbiol* 46:293.
- de Vries N, Duinsbergen D, Kuipers EJ, Pot RG, Wiesenekker P, Penn CW, van Vliet AH, Vandenbroucke-Grauls CM and Kusters JG (2002) Transcriptional phase variation of a type III restriction-modification system in *Helicobacter pylori*. *J Bacteriol* 184:6615-6623.
- Dybvig K and Voelker LL (1996) Molecular biology of mycoplasmas. *Annu Rev Microbiol* 50:25-57.
- Dybvig K, Sitaraman R and French CT (1998) A family of phase-variable restriction enzymes with differing specificities generated by high-frequency gene rearrangements. *Proc Natl Acad Sci USA* 95:13923-13928.
- Figueiredo C, Quint WG, Sanna R, Sablon E, Donahue JP, Xu Q, Miller GG, Peek RM Jr, Blaser MJ and van Doorn LJ (2000) Genetic organization and heterogeneity of the *iceA* locus of *Helicobacter pylori*. *Gene* 246:59-68.
- Gumulak-Smith J, Teachman A, Tu AH, Simecka JW, Lindsey JR and Dybvig K (2001) Variations in the surface proteins and restriction enzyme systems of *Mycoplasma pulmonis* in the respiratory tract of infected rats. *Mol Microbiol* 40:1037-1044.
- Henderson IR, Owen P and Nataro JP (1999) Molecular switches - The ON and OFF of bacterial phase variation. *Mol Microbiol* 33:919-932.
- Himmelreich R, Plagens H, Hilbert H, Reiner B and Herrmann R (1997) Comparative analysis of the genomes of the bacteria *Mycoplasma pneumoniae* and *Mycoplasma genitalium*. *Nucleic Acids Res* 25:701-712.
- Kobayashi I (2001) Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Res* 29:3742-3756.
- Kobayashi I, Nobusato A, Kobayashi-Takahashi N and Uchiyama I (1999) Shaping the genome-restriction-modification systems as mobile genetic elements. *Curr Opin Genet Dev* 9:649-656.
- Lobner-Olesen A, Skovgaard O and Marinus MG (2005) Dam methylation: Coordinating cellular processes. *Curr Opin Microbiol* 8:154-160.
- Malone T, Blumenthal RM and Cheng X (1995) Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. *J Mol Biol* 253:618-632.
- Minion FC, Lefkowitz EJ, Madsen ML, Cleary BJ, Swartzell SM and Mahairas GG (2004) The genome sequence of *Mycoplasma hyopneumoniae* strain 232, the agent of swine mycoplasmosis. *J Bacteriol* 186:7123-7133.
- Murray NE (2002) Immigration control of DNA in bacteria: Self versus non-self. *Microbiology* 148:3-20.
- Nobusato A, Uchiyama I and Kobayashi I (2000a) Diversity of restriction-modification gene homologues in *Helicobacter pylori*. *Gene* 259:89-98.
- Nobusato A, Uchiyama I, Ohashi S and Kobayashi I (2000b) Insertion with long target duplication: A mechanism for gene mobility suggested from comparison of two related bacterial genomes. *Gene* 259:99-108.
- Razin S, Yogeve D and Naot Y (1998) Molecular biology and pathogenicity of mycoplasmas. *Microbiol Mol Biol Rev* 62:1094-1156.

- Remaut H and Waksman G (2004) Structural biology of bacterial pathogenesis. *Curr Opin Struct Biol* 14:161-170.
- Rocha EP and Blanchard A (2002) Genomic repeats, genome plasticity and the dynamics of Mycoplasma evolution. *Nucleic Acids Res* 30:2031-2042.
- Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA and Barrell B (2000) Artemis: Sequence visualization and annotation. *Bioinformatics* 16:944-945.
- Ryan KA and Lo RY (1999) Characterization of a CACAG pentanucleotide repeat in *Pasteurella haemolytica* and its possible role in modulation of a novel type III restriction-modification system. *Nucleic Acids Res* 27:1505-1511.
- Salama N, Guillemin K, McDaniel TK, Sherlock G, Tompkins L and Falkow S (2000) A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. *Proc Natl Acad Sci USA* 97:14668-14673.
- Saunders NJ, Peden JF, Hood DW and Moxon ER (1998) Simple sequence repeats in the *Helicobacter pylori* genome. *Mol Microbiol* 27:1091-1098.
- Seib KL, Peak IR and Jennings MP (2002) Phase variable restriction-modification systems in *Moraxella catarrhalis*. *FEMS Immunol Med Microbiol* 32:159-165.
- Tock MR and Dryden TF (2005) The biology of restriction and anti-restriction. *Curr Opin Microbiol* 8:466-472.
- Tomb JF, White O, Kerlavage AR, Clayton RA, Sutton GG, Fleischmann RD, Ketchum KA, Klenk HP, Gill S, Dougherty BA, *et al.* (1997) The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* 388:539-547.
- Tu AH, Lindsey JR, Schoeb TR, Elgavish A, Yu H and Dybvig K (2002) Role of bacteriophage MAV1 as a mycoplasmal virulence factor for the development of arthritis in mice and rats. *J Infect Dis* 186:432-435.
- Vasconcelos AT, Ferreira HB, Bizarro CV, Bonatto SL, Carvalho MO, Pinto PM, Almeida DF, Almeida LG, Almeida R, Alves-Filho L, *et al.* (2005) Swine and poultry pathogens: The complete genome sequences of two strains of *Mycoplasma hyopneumoniae* and a strain of *Mycoplasma synoviae*. *J Bacteriol* 187:5568-5577.
- Xia X (2003) DNA methylation and mycoplasma genomes. *J Mol Evol* 57:S21-S28.
- Zhang Y, Nelson M, Nietfeldt J, Xia Y, Burbank D, Ropp S and Van Etten JL (1998) Chlorella virus NY-2A encodes at least 12 DNA endonuclease/methyltransferase genes. *Virology* 240:366-75.
- Zheng J and McIntosh MA (1995) Characterization of IS1221 from *Mycoplasma hyorhinis*: Expression of its putative transposase in *Escherichia coli* incorporates a ribosomal frameshift mechanism. *Mol Microbiol* 16:669-685.

### Internet Resources

- Entrez Genome database and Blast search tools: <http://www.ncbi.nlm.nih.gov> (verified 29/09/2006).
- PHRED/PHRAP/CONSED software for sequence assembly: <http://www.phrap.org> (verified 29/09/2006).
- Restriction Enzyme Database: <http://rebase.neb.com/rebase/rebase.html> (verified 29/09/2006).
- REBASE Genomes: <http://tools.neb.com/~vincze/genomes/>.

*Associate Editor: Darcy F. de Almeida*