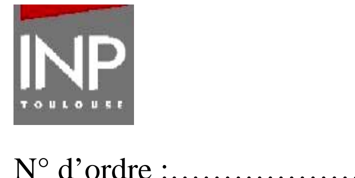


N° d'ordre :………………

THESE

Présentée par

Woubit SALAH

Pour obtenir

LE TITRE DE DOCTEUR DE L'INSTITUT NATIONAL POLYTECHNIQUE DE TOULOUSE
PhD

UTILISATION D'UNE SEQUENCE COMPLETE DU GENOME D'UNE SOUCHE DE MYCOPLASMA MYCOIDES SUBSP. MYCOIDES LC POUR LA MISE AU POINT DE TESTS DE DIAGNOSTIC, APPLICATION A L'EXPRESSION DE GENES HETEROLOGUES

Ecole doctorale: SEVAB

Spécialité: SCIENCE VETERINAIRE

Soutenance le 25 Janvier 2008, devant le jury composé de:

Henri Dabernat	Président
Colette Saillard	Rapporteur
Dominique Le Grand	Rapporteur
Xavier Berthelot	Directeur de thèse
François Poumarat	Membre
Alain Blanchard	Membre
Christine Citti	Membre
François Thiaucourt	Membre

H. Dabernat : Laboratoire de Bactériologie Hygiène, Institut Fédératif de Biologie 330 avenue de Grande Bretagne - TSA 30021 ; 31059 Toulouse Cédex 9 (F)
Tél. +33 (0)5 67 69 04 17; Fax +33 (0)5 67 69 04 83 ; E.mail : dabernat.h@chu-toulouse.fr

C. Saillard : Laboratoire de Biologie Cellulaire et Moléculaire, UMR 1090 IBVM, INRA BP 81, 71, Avenue Edouard Bourlaux, 33883 Villenave d'Ornon Cedex, France
Tel: + 33 (0)5 57 12 23 62; E-mail: saillard@bordeaux.inra.fr

D. Le Grand : UMR Mycoplasmoses des Ruminants AFSSA-ENVL ; Pathologie du Bétail ; Ecole Nationale Vétérinaire de Lyon ; 1, Avenue Bourgelat - 69280 Marcy l'Etoile
Tel: +33 (0)4.78.87.26.05 / Fax: +33 (0)4.78.87.26.06 ; E-Mail: d.legrand@vet-lyon.fr

X. Berthelot : Departement Elevage et Produits et Unite Mixte de Recherches INRA-ENVT 1225 "Interactions Hôtes-Agents Pathogènes" (IHAP) ; Ecole Nationale Veterinaire, 23 chemin des Capelles, B.P. 87614, F-31076 Toulouse cedex 3
Tel +33 (0)5 61 19 38 57 ; Fax +33 (0)5 61 19 32 73; E.mail : x.berthelot@envt.fr

F. Poumarat : Unité "Mycoplasmologie" AFSSA site de Lyon ; 31 av. Tony-Garnier 69364 LYON cedex 07 ;
Tel : +33 (0)4 78 69 68 31 ; Fax : 33 (0)4 78 61 91 45 E.mail : f.poumarat@lyon.afssa.fr

A. Blanchard : UMR 1090 GDPP ; INRA, IBVM, Centre de Recherche de Bordeaux 71 avenue Edouard Bourlaux BP 81 33883 Villenave D'Ornon Cedex France
Tel +33 (0)5 57 12 23 93; Fax: +33 (0)5 57 12 23 69; E.mail : ablancha@bordeaux.inra.fr

C. Citti : UMR 1225 INRA-ENVT Ecole Nationale Vétérinaire 23 Chemin des Capelles BP 87614 ; 31076 Toulouse Cedex 3 France
Tel : +33 (0)5 61193856; Fax : +33 (0)5 61193273 ; E.mail: c.citti@envt.fr

F. Thiaucourt :CIRAD, UPR 15 « Contrôle des maladies » TA A-15/G Campus de Baillarguet 34398 Montpellier cedex 5
Tel : +33 (0)467593723 ; Fax : +33 (0)467593798 ; E.mail : thiaucourt@cirad.fr

TABLE OF CONTENTS

ACKNOWLEDGEMENT……………………………………………………………………I
LIST OF PUBLICATIONS………………………………………………………………III
RESUME………………………………………………………………………………IV
SUMMARY……………………………………………………………………………VI
LIST OF FIGURES……………………………………………………………………VIII
LIST OF TABLES……………………………………………………………………XI
LIST OF ABBREVIATIONS…………………………………………………………XII

GENERAL INTRODUCTION……………………………………………………………1

CHAPTER I

Introduction to *Mycoplasma,* the *M. mycoides* cluster and Genome sequencing

1.1 The Mycoplasmas……………………………………………………………………4
1.1.1 General description and habitat ……………………………………………………4
1.1.2 Genome characteristics and phylogeny………………………………………………5
1.1.3 Pathogenicity in Mycoplasmas ………………………………………………………7
1.1.3.1 Adhesion to host cells……………………………………………………………7
1.1.3.2 Invasion …………………………………………………………………………7
1.1.3.3 Antigenic variation ………………………………………………………………7
1.2 The Mycoplasma mycoides cluster ……………………………………………………8
1.2.1 Diseases associated with the *M. mycoides* cluster …………………………………11
1.2.1.1 CBPP and CCPP…………………………………………………………………11
1.2.1.2 Contagious agalactia……………………………………………………………11
1.2.1.3 Other disease conditions…………………………………………………………13
1.2.2 Pathogenicity of members of the *M. mycoides* cluster……………………………13
1.2.2.1 Variable surface protein…………………………………………………………13
1.2.2.2 Metabolic byproduct……………………………………………………………14
1.2.2.3 Capsular polysaccharide …………………………………………………………14
1.3 Genome sequencing approach in Mycoplasmas ………………………………………15
1.3.1 Genome sequencing……………………………………………………………………15
1.3.1.1 Genome sequencing strategies……………………………………………………16
1.3.1.2 Sequencing techniques …………………………………………………………18
1.3.1.3 Genomic sequence analysis………………………………………………………20
1.4 Genome sequencing of *M. mycoides* subsp. *mycoides* LC………………………………21
1.5 Annotation of *M. mycoides* subsp. *mycoides* LC genome sequence……………………21

CHAPTER II

Genome size estimation of MmmLC strain 95010-C1 and analysis of presence of repeated genes

1 INTRODUCTION 23

1.1 Genome size estimation 23

1.2 DNA Hybridization 24

2 MATERIALS AND METHODS 26

2.1 Strains 26

2.2 DNA plug preparation 26

2.3 Restriction Enzyme Analysis 26

2.3.1 Choice of restriction enzymes and digestion for PFGE 26

2.3.2 Choice of restriction enzyme and digestion for Southern blotting 27

2.4 PFGE Electrophoresis conditions 27

2.5 DNA probes for Southern blot hybridization test 27

2.6 Southern blotting 28

3 RESULTS AND DISCUSSION 30

3.1 Genome size estimation of strain 95010-C1 30

3.1.1 Mycoplasma strains analyzed 32

3.1.1 1 Mcc, *M. bovis* and *M. agalactiae* 32

3.1.1.2 MmmSC strains 32

3.1.1.3 MmmLC strains 34

3.2 Assessment of the presence of repeated gene sequences 35

3.2.1 Gene lppQ (ApalI digest) 35

3.2.2 Gene mgtA (XbaI digest) 35

3.2.2.1 mgtA gene in strain PG1[T] 35

3.2.2.2 mgtA gene in strain 8740-Rita 36

3.2.2.3 mgtA gene in MmmLC strains 36

3.2.2.3.1 XbaI enzyme activity in 95010-C1 37

3.2.2.3.2 Dam methylase activity between MmmLC and MmmSC 37

3.2.3 Gene phnC (EcoRI digest) 39

3.2.3 1 phnC gene in strain PG1[T] 39

3.2.3 2 phnC gene in strain 8740-Rita 39

3.2.3 3 phnC gene in MmmLC strains 39

3.2.4 Gene galE (EcoR*V digest*) 40

3.2.4.1 galE gene in strain PG1[T] 40

3.2.4.2 galE gene in strain 8740-Rita 40

3.2.4.3 galE gene in MmmLC strains 40
3.2.5 Gene epsG (EcoRV digest) 41
3.2.5.1 epsG gene in strain PG1[T] 41
3.2.5.2 epsG gene in 8740-Rita 42
3.2.5.3 epsG gene in MmmLC strains 42
4 CONCLUSIONS AND DISCUSSION 44

CHAPTER III

Application of whole genome sequence for the development of specific diagnostic tools

1 INTRODUCTION 48
1.1 Identification of members of the *M. mycoides* cluster 48
1.2 The Arginine deiminase pathway 49
1.3 The "ADI operon" within the *M. mycoides* cluster 51
Paper I 54
Paper II 73
Paper III 87
2 CONCLUSION 100

CHAPTER IV

Assessment of previously described and Identification of novel insertion sequence elements

1 Insertion sequence elements 102
1.1 Structure of IS-elements 102
1.1.2 Transposition of IS-elements 103
1.1.2 Distribution of IS-elements 106
1.1.2.1 IS-elements in the M. mycoides cluster 107
1.1.2.2 IS-elements of M. mycoides subsp. mycoides LC strain 95010-C1 109
Paper IV 112

CHAPTER V

Transformation of MmmLC and expression of H-protein from peste des petits ruminant's virus structural proteins

1 INTRODUCTION 130
1.1 Peste des petits ruminants (PPR) 131
1.2 Morbilliviruses 132
1.2.1 Structure of Morbilliviruses 132
1.2.2 H-protein 134
2 MATERIALS AND METHODS 135

2.1 Heterologous gene 135
2.1.1 Codon usage analysis 135
2.2 Promoter sequence signal analysis 135
2.3 Plasmids 135
2.4 Choice of restriction enzymes for cloning 135
2.5 PCR conditions 136
2.6 Cloning procedure 137
2.7 Plasmid propagation and amplification 137
2.7.1 Plasmid Mini-preps and Maxi-preps 137
2.8 Mycoplasma transformation 137
2.9 Immunoblotting 138
3 RESULTS AND DISCUSSION 139
3.1 Codon usage in the PPR-H gene 139
3.2 Choice of promoter sequences 139
3.3 Choice of restriction enzymes for cloning purposes 143
3.4 Promoter and H-gene ligations in an intermediate vector 145
3.5 Preparation of pMYCO1 for cloning 148
3.6 PR-CHP-H-gene construct in ΔpMYCO1 149
3.7 Validation of the PR-CHP-H-gene insertion sequence in pWS 150
3.8 Obtention of the final construct, pWS1 153
3.9 Verification of the H-gene sequence in plasmid pWS1 153
3.10 Mycoplasma transformation using pWS1 156
3.10.1 Transformation efficiency 157
3.10.2 Stability of pWS1 in MmmLC cells 158
3.11 Sequencing of the PR-CHP-H-gene fragment from transformed mycoplasma 158
3.12 H-protein expression analysis 159
4 CONCLUSION AND PERSPECTIVES 160
GENERAL CONCLUSION 162
REFERENCES 166
ANNEXES 178

ACKNOWLEDGEMENTS

I would like to thank the French Ministry of Foreign Affairs and French Embassy in Addis Ababa, Ethiopia for financing my PhD program. I also would like to thank Prof Philip Dorchies for all the contributions he has made for this scholarship program. In this context many thanks to my former colleagues from the Faculty of Veterinary Medicine, Prof Getachew and Dr. Bayleyegn for allowing me to continue further my studies.

I am particularly grateful to my supervisor François Thiaucourt for taking me as his student and for the constant support, reliable advice, and constructive criticism and providing directions all the time of my research and writing of this thesis.

I am also highly thankful to Professor Xavier Berthelot for accepting to be my Director of the thesis. I am very grateful to Prof. Henri Dabernat, Dr. Alain Blanchard Dr. François Poumarat, Dr. Christine Citti, Dr. Dominique Legrand and Dr. Colette Saillard for their invaluable time dedication to be part of my Jury and accepting to be my rapporters.

I am more than grateful to CIRAD, UPR15 for welcoming and supporting me all the year through my stay. My special thanks to Lucía Manso-Silván for being always there, sharing her knowledge, and her writing skills and diligence for the corrections of my manuscript, which helped me relieve many writing blocks. Sophie Lorenzon made me discover the first step in molecular biology technique, best friend and who made me realize how much I valued the moments I shared with her and her constant help in every step of my life during my stay in France. Armelle Peyraud who complements our bacteriology section and I thank her for all the contributions she made in sample preparations of my first time works.

My heart felt gratitude also goes to all members of the Virology Department CIRAD-UPR15, who have helped me go through the last part of my thesis work, especially I would like to mention Genevieve Libeau, for her dedications to answer every bits of my questions and material provision, Cecile Minet, Olivier Kwiatek and Vincent Michaud and Patricia Gil for their experience sharing in the world of molecular cloning.

I am also thankful for those who have been there to lend a hand, and I appreciate their constant encouragement, with no special order Flo, Denise, Nadege, Renata A., Isabelle

Leplomb, Mme. Seban, Laurence Dedieu, Philippe T, Isabelle P, Lolette, Nadin and many more……. I am grateful to every one who have participated directly or indirectly for the fruitful out come of my thesis work.

ከሩቅም:ሆነ:ከቅርብ:ሆነዉ:በሃሳብ:ለረዱኝ:ለቤተሰቦቼ:በሙሉ:ምስጋና:አቀርባለሁ::
My beloved mother Abuye, Emi, Munna, Titi, Mamme……

LIST OF PUBLICATIONS

1. **S. Woubit**, Manso-Silván L., Thiaucourt, F., ISMmy*2* a Novel IS*3*-Type Insertion Sequence found in *M. mycoides* subsp. *mycoides* LC [Article submitted to the journal of Research in Microbiology].

2. **Woubit, S**., Manso-Silván, L., Lorenzon, S., Gaurivaud, P., Poumarat, F., Pellet, M.P., Singh, V.P., Thiaucourt, F. A PCR for the detection of mycoplasmas belonging to the *Mycoplasma mycoides* cluster: Application to the diagnosis of contagious agalactia. Mol Cell Probes (**2007**) 21, 391-399.

3. **Woubit, S**., Lorenzon S., Peyraud A., Manso-Silván L., Thiaucourt F. A specific PCR for the identification of *M. capricolum* subsp. *capripneumoniae,* the causative agent of contagious caprine pleuropneumonia (CCPP), Veterinary Microbiology (**2004**) 104 (1-2): 125-32.

4. Peyraud, A., **S. Woubit**, J.B. Poveda, C.De la Fe, P. Mercier, F. Thiaucourt. A specific PCR for the rapid detection of *Mycoplasma putrefaciens*, one of the agents of the contagious agalactia syndrome of goats. Molecular and cellular probes .17 (**2003**) 289-294.

5 Bayleyegn M., **Woubit S**., Ahmed M. Antimicrobial resistance pattern of salmonella serotypes isolated from apparently healthy slaughtered camels *(Camelus dromedaries*) in eastern Ethiopia. Berl. Munich. Tierarztl. Wschr. 117, 39-45 (**2004**).

6. B. Molla, A. Mohammed and **Woubit S**. *Salmonella* prevalence and distribution of serotypes in apparently healthy slaughtered camels *(Camelus dromedaries)* in eastern Ethiopia. Tropical animal health and production, 36 (**2004**) 451-458.

7. Bayleyegn Molla, Daniel Alemayehu, **Woubit Salah**. Sources and distribution of Salmonella serotypes isolated from food animals, slaughterhouse personnel and retail meat products in Ethiopia: 1997-2002. Ethiop. J. Health Dev.**2003**; 17(1):63-70.

8. **Woubit S**, Bayleyegn M, P. Bonnet, S. Jean-Baptiste. Camel (*Camelus dromedarius*) Mastitis in Borena Lowland Pastoral Area, South-western Ethiopia. Revue élev. Méd. Vét. Pays trop., **2001**, 54 (3-4): 207-212.

9. B. Tibaijuka, B. Molla, G. Hildebrandit, J. Kleer and **Woubit S**. Antimicrobial resistance to salmonella isolated from retail raw chicken meat and giblets in Ethiopia. Bull. Anim. Hlth. Prod. Afr. (**2002**), 50, 86-96.

POSTERS

1. **Woubit S.**, Manso-Silván L., P. Gaurivaud, V.P. Singh, Thiaucourt F. PCR based identification of "*Mycoplasma mycoides* cluster", application to the diagnosis of contagious agalactiae (**2006**) [Poster presented for the 16th international organizations for mycoplasmology, IOM, Cambridge UK]

2. **Woubit S.**, Lorenzon S., Peyraud A., Manso-Silván L., Thiaucourt F. A specific PCR for the identification of *M. capricolum* subsp. *capripneumoniae*, the causative agent of contagious caprine pleuropneumonia (CCPP) (**2004)** [Poster presented for the 15th international congress for mycoplasmology. IOM, Athens, Georgia USA]

RESUME

Les mycoplasmes sont des bactéries dépourvues de paroi qui dérivent de bactéries Gram plus. Il existe un groupe d'espèces, appelé le «Groupe mycoides» qui rassemble des espèces pathogènes pour les ruminants bien qu'il soit phylogénétiquement proche d'espèces pathogènes de plante. Leur génome est très réduit, environ un million de paires de bases et peu riche en G + C, environ 25 %. Parmi ce groupe, *Mycoplasma mycoides* subsp. *mycoides* LC (MmmLC) est un agent responsable d'agalaxie contagieuse chez les chèvres. Il est très proche de *Mycoplasma mycoides* subsp. *mycoides* SC (MmmSC) qui est l'agent responsable de la péripneumonie contagieuse bovine et dont le génome de la souche de référence était disponible. En raison de cette proximité nous avons décidé de séquencer complètement le génome d'une souche de MmmLC afin de pouvoir réaliser des études de génomique comparative. Préalablement au séquençage et à l'assemblage, nous avons évalué la taille du génome avec la technique d'électrophorèse en champs pulsé. Nous avons pu ensuite contrôler l'assemblage obtenu en comparant les données expérimentales «*in-silico*» avec les résultats d'électrophorèse. De plus nous avons réalisé des «Southern blots» afin de vérifier si les séquences dupliquées chez MmmSC l'étaient également chez MmmLC. La comparaison avec les génomes complets déjà disponibles pour des souches du «Groupe mycoides» a permis d'identifier un locus intéressant pour développer des PCR spécifiques. Ce locus comprend des gènes ou des fragments de gènes appartenant chez certaines bactéries à un opéron de «voie de déimination de l'arginine». Le nombre ainsi que l'agencement et la séquence de ces gènes varie d'une espèce à l'autre au sein du «Groupe mycoides». Il a ainsi été possible de développer une PCR spécifique pour *Mycoplasma capricolum* subsp. *capripneumoniae*, l'agent de la pleuropneumonie contagieuse caprine en amplifiant un fragment du gène *arc*D et une PCR spécifique de *M. putrefaciens*, un agent d'agalaxie contagieuse, en amplifiant un fragment du gène *arc*B. Pour la détection de l'ensemble des espèces du «Groupe mycoides» nous avons choisi un gène plus conservé, *glk*, situé en aval de l'opéron. L'annotation du génome de MmmLC a également permis d'identifier des séquences d'insertion. L'une d'entre elles, appartenant à la famille IS*3*, n'avait pas encore été décrite et a été appelée ISMmy*2*. Elle est présente chez certaines espèces du «Groupe mycoides» mais pas chez toutes les souches. Un variant de cette IS existe chez des espèces proches du «Groupe mycoides» et il en existe une copie non fonctionnelle chez MmmSC. Enfin nous avons voulu évaluer les capacités de MmmLC à exprimer des antigènes hétérologues dans le but ultime d'en faire un vecteur d'expression vaccinal. C'est pourquoi

nous avons choisi un gène d'intérêt vétérinaire majeur, le gène H du virus de la peste des petits ruminants.

SUMMARY

Mycoplasmas are the smallest bacteria without a cell wall derived from Gram positive bacteria. A group of mycoplasma known as the "*Mycoplasma mycoides* cluster" is composed of five subspecies and an unassigned group of strains known for their pathogenicity in ruminant hosts. Phylogenetically, this cluster is found to be closely related to species of mycoplasma plant pathogens. Mycoplasmas have a reduced genome size of about 1 Mbp, characterized by a low GC content of about 25 %. Among members of the *Mycoplasma mycoides* cluster, *Mycoplasma mycoides* subsp. *mycoides* large colony biotype (MmmLC) is one of the agents responsible for contagious agalactia in goats. This organism is closely related to *Mycoplasma mycoides* subsp. *mycoides* small colony biotype, the causative agent of contagious bovine pleuropneumonia (CBPP), for which the whole genome sequence is available. Because of the close relationship of these two species we have decided to sequence the genome of an MmmLC strain for comparative genomics. Before whole genome sequencing and assembly, we have estimated the genome size of MmmLC using pulse field gel electrophoresis (PFGE). Data generated from this initial study have permitted us to verify the genome assembly by comparing *in-silico* profiles. In addition the preliminary analysis included DNA hybridization tests to verify the presence of duplicated genes in MmmLC as that of the genome of MmmSC. Comparative genomics made from the available whole genome sequence data of species within the *M. mycoides* cluster has permitted the identification of target genes, which were used for the development of specific PCR tests. The target genes chosen included genes of the "arginine deiminase operon", in most bacteria genes of this operon code for enzymes involved in the degradation of arginine to produce energy. The number of these genes as well as their organization within the operon found to vary between members of the *M. mycoides* cluster. From this operon *arc*D has been used to develop a specific PCR for the identification of *Mycoplasma capricolum* subsp. *capripneumoniae*, the causative agent of contagious caprine pleuropneumonia (CCPP), and *arc*B has been used for the development of specific PCR for the identification of *M. putrefaciens*, another causative agent of the contagious agalactia syndrome. The *glk* gene, flanking the operon on the 3' end, was found to be highly conserved among all members of the *M. mycoides* cluster and was used for the design of specific primers able to detect all members of *M. mycoides* cluster. Furthermore, annotation of the genome sequence of MmmLC allowed the discovery of two new insertion sequence elements. One of these two insertion sequence elements was found in higher copy in the genome and belongs to the IS*3*

family. This insertion sequence was not described in any other mycoplasma species or bacteria, was given a new name: ISMmy*2*. It was also found in some species of the *M. mycoides* cluster, but not in all the strains under these species. Interestingly, a non-functional vestige of ISMmy*2* was also found in the MmmSC genome. Copies of this ISMmy*2* were also found in species closely related to the *M. mycoides* cluster. Finally, we have tried to evaluate the capacity of MmmLC to be transformed and to express a heterologous gene with the ultimate aim to create a multivalent vaccine. For this aim we have chosen the H-gene of peste des petits ruminant virus of veterinary health importance.

LIST OF FIGURES

CHAPTER I

Figure I. 1 Electron micrograph of thin-sectioned mycoplasma cells. 4
Figure I. 2 Phylogenetic tree of the Mollicutes constructed based on the 16S rDNA sequence. 6
Figure I. 3 Morphology of a typical "fried-egg" mycoplasma colony 8
Figure I. 4 Phylogenetic tree obtained from distance analysis of five concatenated protein coding sequences (fusA, glpQ, gyrB, lepA, rpoB). 10
Figure I. 5 Strategies used to obtain whole genome sequences. 17
Figure I. 6 Diagram showing steps taken in Pyrosequencing. 20

CHAPTER II

Figure II. 1 Pulsed-field profiles of MluI digested genomic DNA of mycoplasma strains. 30
Figure II. 2 Pulsed-field profiles of BglI digested genomic DNA of mycoplasma strains. 31
Figure II. 3 Pulsed-field profiles of NcoI digest of genomic DNA of mycoplasma strains. 31
Figure II. 4 Southern hybridization of XbaI digested DNA using mgtA specific DNA probe. 36
Figure II. 5 Restriction enzyme analysis of MmmLC and MmmSC strains. 38
Figure II. 6 Southern hybridization of EcoRI digested DNA using phnC specific DNA probe. 40
Figure II. 7 Southern hybridization of EcoRV digested DNA using galE specific DNA probe. 41
Figure II. 8 Southern hybridization of EcoRV digested DNA using epsG specific DNA probe. 42

CHAPTER III

Figure III. 1 Arginine Deiminase Pathway. 51
Figure III. 2 Organization of the arginine deiminase "ADI-operon" in the representative strains of the *M. mycoides* cluster. 52

Paper 1

1、Figure 1. Schematic representation of the complete glk gene sequence alignment of three mycoplasmas belonging to the *M. mycoides* cluster 61
1、Figure 2. Distance tree calculated from the alignment of a 380-bp-fragment of the glk gene of six strains representing the *M. mycoides* cluster. 62
1、Figure 3. A 428 bp amplicon was obtained from all strains belonging to the *M. mycoides* cluster 63
1、Figure 4 Organization of the arginine deiminase (ADI) operon within the *M. mycoides* cluster. ..67

Paper 2

2. Fig 1. PCR results obtained with primer pair ArcBMccpF and ArcBMccpR with strains listed in Table 1......79

2. Fig 2. Multiple alignment of a 523 bp-long partial sequence from the ArcB gene (AlignX of VectorNTI 8).......80

2. Fig 3. Alignment of the sequences homologous to primers Mput1 (a) and Mput2 (b)..................81

2. Fig 4. PCR results obtained with primer pair Mput1 and Mput2 with strains listed in Table 1 82

Paper 3

3. Fig. 1 Diagram showing the organization of the ADI operon in *M. capricolum* subsp. *capripneumoniae* strain GL100.------ 93

3. Fig. 2 PCR results obtained with the primer pair Mccp-spec-F/R. ------ 94

3. Fig. 3 Agarose gel electrophoresis of ClaI-cleaved specific PCR products from *M. capricolum* subsp. *capripneumoniae* strains.------ 95

CHAPTER IV

Figure IV. 1 Structure of a typical IS-element. 103

Figure IV. 2 The structure of the central intermediate formed by a cut-and-paste transposase........104

Figure IV. 3 Cut-and-paste movement of IS transposition. 105

Figure IV. 4 Copy and paste transposition......106

Figure IV. 5 Southern hybridization of IS*1296* DNA probe to *Hind*III digested genomic DNA. 110

Paper 4

4. Fig. 1. Schematic representation of ISMmy2 structure and organization……..………………..126

4. Fig. 2. Alignment of the putative transposase region containing the DDE-K motif of ISMmy*2* and consensus sequence of the IS*3* family……………………………………………....127

4. Fig. 3. Genomic map of MmmLC strain 95010-C1 showing the positions of insertion sequences ISMmy2-A to-F……………………………...…………………………………......128

4. Fig. 4 Identification of multiple copies of ISMmy*2* in different mycoplasma strains…………......129

CHAPTER V

Figure V. 1 Schematic representation of the structure of Morbillivirus..........133
Figure V. 2 Promoter sequence of CHP of locus tag MLC_3560 from the genome sequence of MmmLC strain 95010-C1 141
Figure V. 3 Promoter sequence of spiralin gene Acc no. AF012877..........142
Figure V. 4 Flow diagram representing the cloning strategy used for final plasmid pWS1 construction. 144
Figure V. 5 PCR amplifications of H-gene and promoter sequences..........145
Figure V. 6 PCR amplification of ligation product of promoter sequences with H-gene. 146
Figure V. 7 pCR®2.1-TOPO® vector containing inserts PR-CHP-H-gene and PR-spiralin-H-gene. 147
Figure V. 8 Double digest of pMYCO1 by enzymes AatII and NgoMIV. 148
Figure V. 9 PCR amplification of pWS from E. coli transformed colonies..........149
Figure V. 10 In-silico pWS construct containing PR-CHP-H-gene sequence. 150
Figure V. 11 Sequence alignment of the actual and the in-silico generated pWS sequences. 152
Figure V. 12 Site directed mutagenesis of pWS yielding plasmid pWS1..........153
Figure V. 13 Amino acid sequence alignment of H-gene from PPRV and H-gene from plasmid pWS1..........155
Figure V. 14 pWS1-transformed mycoplasma colonies after 14 days of incubation..........156
Figure V. 15 Amplification of plasmid pWS1 from transformed mycoplasma cells..........158
Figure V. 16 Dot-immunoblotting test of transformed mycoplasma cells C1-C13. 159

LIST OF TABLES

CHAPTER II

Table II. 1 Primers designed from the PG1^{T} genome sequence for the preparation of specific DNA probes 28

Table II. 2 Estimated DNA sizes of mycoplasmas strains. 33

CHAPTER III

Paper 1

1. Table 1.List of mycoplasma strains used for initial validation of the specific PCR at CIRAD.....57

1. Table 2.Validation of glk-PCR from field milk samples. 64

Paper 2

2. Table. 1 Strains used for the initial validation of the specific PCR for *Mycoplasma putrefaciens*.76

2. Table. 2 Oligonucleotide primers used for the search of a specific PCR identification of *Mycoplasma putrefaciens* 77

Paper 3

3. Table 1 List of mycoplasma strains used in the study....90

CHAPTER IV

Paper 4

4. Table 1 Mycoplasma strains used to study the distribution of the insertion sequence element ISMmy2....123

4. Table 2 Primers used for ISMmy2 amplification and sequencing....124

4. Table 3 Homologies between copy ISMmy2-A, other copies within MmmLC strain 95010 genome and with ISMmy2 or its variant in closely related mycoplasma species. Asterisks indicate junctions between the insertion sequence and target site sequence....125

CHAPTER V

Table V. 1 List of primers used for amplification, cloning and sequencing of promoters and foreign gene 136

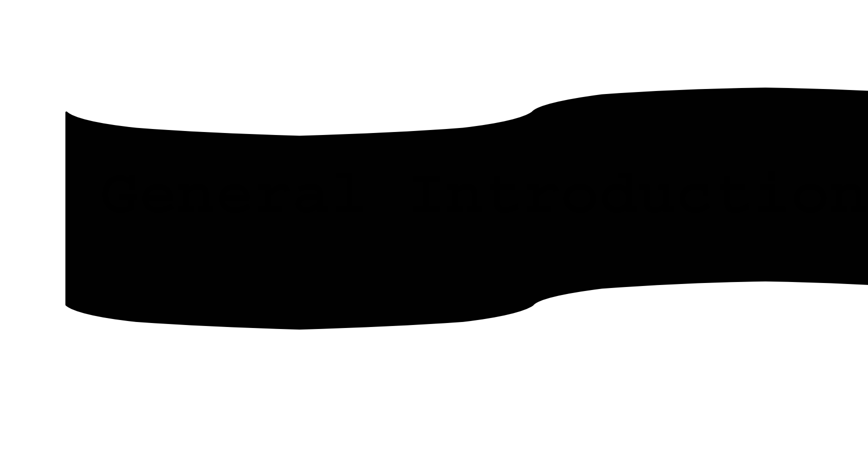
Table V. 2 BPROM promoter sequence analysis of genes from the genome sequence of MmmLC140

LIST OF ABBREVIATIONS

aa	amino acid
Acc. No	accession number
ADI	arginine deiminase
AFSSA	French Agency for Food Safety
Amp^R	ampiciline resistance
ATP	Adenosin triphosphate
BAC	bacterial artificial chromosome
BLAST	Basic Local Alignment Search Tool
bp	base pair
BPROM	bacterial promoter prediction program
CAAT	contig assembly and annotation tool
CBPP	contagious bovine pleuropneumonia
CCPP	contagious caprine pleuropneumonia
CDS	coding sequence
cDNA	complementary DNA
c-ELISA	competitive enzyme linked immuno sorbent assay
cfu	colony forming unit
CHEF	contour-clamped homogeneous electric field
CIRAD	Centre International de Recherche Agronomique pour le Developement
*col*E1	*Escherichia coli* origin of replication
CPS	capsular polysaccharide
Da, kDa	Dalton, kilo Dalton
DAB	3-3' diamino-benzidine
dATP	deoxy adenine triphosphate
dCTP	deoxy cytosine triphosphate
dGTP	deoxy guanine triphosphate
DHAP	dihydroxyacetone phosphate
Dig	digoxygenine
DNA	Deoxyribonucleic acid
dNTP	deoxy nuclotide triphosphate
DR	direct repeat
dTTP	deoxy thiamine triphosphate
dUTP	deoxy uracil triphosphate
ECL	Enhanced ChemiLuminescence
EDTA	ethylene diamine tetra-acetic acid
EMBL	European Molecular Biology Laboratory
F1-origin	viral origin of replication
FAO	world food organization
g	acceleration due to gravity
gcua	Graphical codon usage analysis
GFP	Green fluorescence protein
H	haemagglutinin
hr	hour
ICE	integrative and conjugative elements
IPF	individual protein file
IR	inverted repeat

IRL	inverted repeat left
IRR	inverted repeat right
IS	Insertion sequence
IVRI	Indian veterinary research institute
Kan^R	Kanamycin resistance
Kbp	kilo base pair
LAMP	Loop-mediated isothermal amplification
M	Molar
MAKePS	mastitis arthritis keratitis pneumonia and septicaemia
Mbg 7	*Mycoplasma* species bovine serogroup 7
Mbp	mega base pair
Mcc	*Mycoplasma capricolum* subsp. *capricolum*
Mccp	*Mycoplasma capricolum* subsp. c*apripneumoniae*
MLC	Mycoides large colony
MLST	Multi locus sequence typing
Mmc	*Mycoplasma mycoides* subsp. *capri*
MmmLC	*Mycoplasma mycoides* subsp. *mycoides* large colony
MmmSC	*Mycoplasma mycoides* subsp. *mycoides* small colony
mRNA	messenger ribonucleic acid
MSC	Mycoides small colony
NVI	National veterinary institute
OIE	Office International des Epizooties
ORF	open reading frame
*ori*C	chromosomal origin of replication
OTC	ornithine transcarbamylase
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDV	phocine distemper virus
PEG	polyethylene glycol
PFGE	pulse field gel electrophoresis
PPLO	pleuropneumonia like organisms
PPR	peste des petits ruminants
PPRV	peste des petits ruminants virus
PR	promoter
PR-CHP	promoter conserved hypothetical protein
QPCR	Quantitative Polymerase Chain Reaction
RBS	ribosomal binding site
rDNA	ribosomal Deoxyribonucleic acid
RPV	rinderpest virus
rRNA	ribosomal ribonucleic acid
SDS	Sodium dodecyl sulfate
SLAM	signalling lymphocyte activation molecule
SNPs	single nucleotide polymorphism
SSC	sodium chloride and sodium citrate solution
TBE	tris borate EDTA
TE	trisEDTA

TetM	tetracycline resistance
Tpn	Transposon
Trp	Tryptophan
U	unit
UV	ultra violet
V	volts



GENERAL INTRODUCTION

Ethiopia, my home land, is located in eastern Africa, and has the largest highland area of the African continent. Ethiopia's highland topography is rugged and complex. The central part of the country is mostly high plateau, at least 1500 m above sea level with peaks rising to more than 4000 m, and is dissected by gorges and broad valleys. This plateau culminates in the East in a coastal plain spreading to the Red Sea, and in the West in the White Nile Valley plain on the Sudanese border. Agriculture is the mainstay of the economy; it provides a livelihood for 90 per cent of the population living by large in the countryside. Ethiopia's livestock population is one of the largest in Africa, with more than 40 million head of cattle. Livestock contributes the major agricultural outputs supplying the power to cultivate virtually all of Ethiopia's crop land because mechanization is not well developed and ploughing depends on ox traction. It also provides a source of milk, meat, manure, hides and skin and an important source of foreign exchange for the country.

Despite the importance of animal production for the national economy, livestock diseases and poor nutrition remain the major constraints and inflict heavy losses on the sub-sector. Apart from high mortality rates, diseases in livestock affect fertility, growth rate and traction power output. Rinderpest was previously the most devastating infectious disease in cattle but it is now eradicated. However, other diseases remain severely limiting livestock production. Among many, some can be highlighted: contagious bovine pleuropneumonia (CBPP), contagious caprine pleuropneumonia (CCPP), peste des petits ruminants (PPR), sheep and goat pox, blue tongue, fowl pox, pasteurellosis, anthrax, blackleg, haemorrhagic septicaemia, bovine tuberculosis, brucellosis, foot and mouth disease (FMD) and internal and external parasites. Effective disease control is then a major prerequisite in enhancing livestock productivity.

Disease control programmes in Ethiopia involved the establishment of the National veterinary Institute (NVI) in 1964, devoted to disease diagnostics and vaccine production, and the foundation of the first faculty of veterinary medicine (FVM) in 1979, both located at Debre Zeit, 50 km south of the capital city, Addis Ababa. Today the main task of the NVI is to produce veterinary vaccines to fulfil the national needs but also to export these vaccines, mostly in Africa. It produces vaccines for PPR, FMD, sheep and goat pox, fowl typhoid, African horse sickness (AHS), anthrax, Newcastle disease, CBPP, CCPP, Bovine Pasteurellosis, lumpy skin disease (LSD), etc. This vaccine plant produces over 40 million

doses of vaccines every year. Many of these vaccines are independently controlled by the PANVAC Institution (Pan African Vaccine Control), also located in Debre Zeit but depending directly from the African Union.

The faculty of veterinary medicine is a more recent Institution and I graduated from this Faculty in 2000. I was then employed as assistant lecturer in the Department of Microbiology, Faculty of Veterinary Medicine Addis Ababa University. In September 2002, the staff members of my department and one of my DVM external examiners Prof. Philippe DORCHIES (ENVT, Toulouse, France) have accorded me their support to pursue my studies.

The Ethio-French collaboration in the veterinary field had been initiated a long time ago. A formal agreement had been signed in 1962 between Haile Selassie, King of Ethiopia and President de Gaulle. At first, the agreement included sending French veterinarians in the countryside to perform the vaccination campaigns, which later evolved to vaccine production at the NVI, Debre Zeit. More recently French collaboration involved in teaching at the FVM or sending expert missions on specific areas such as quality assurance in the meat sector or development of new veterinary faculties. Ethiopian veterinary research institutions and NVI are also part of laboratory network financed by the French Ministry of Foreign Affairs (LABOVET), which links 5 veterinary laboratories in Africa with CIRAD in Montpellier, France. One of the goals of this project is to reduce the technology gap that has a tendency to expand between developed and developing countries. It is clear that the most recent biotechnology tools are more and more complex, requiring machines which are always more expensive and become obsolete rapidly. Hence participating to modern biology research requires establishing networks.

It is within this global context that I came to CIRAD, Montpellier in September 2002. I first had to learn French language at the University Montpellier III, and then I started my PhD studies through INP and SEVAB at Toulouse. I was hosted by UPR15 of CIRAD (Control of exotic and emerging animal diseases) and more particularly by the mycoplasma team which focuses its studies on important mycoplasma diseases of ruminants that are limiting factors in developing countries but also may be threats for the European livestock sector.

This thesis has been divided into five chapters:

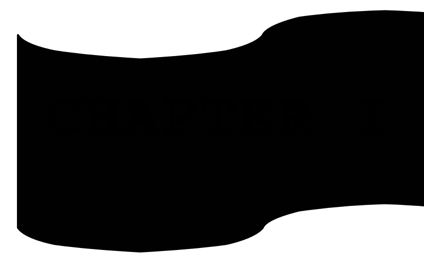
Chapter I. General features of mycoplasmas and the *M. mycoides* cluster in particular, and genome sequencing. Included in this first chapter are the objectives of this thesis.

Chapter II. Estimation of genome size of strain 95010-C1 MmmLC, and evaluation of the presence of repeated genes, which were identified in the previously sequenced MmmSC type strain $PG1^T$.

Chapter III. Application of MmmLC strain 95010-C1 whole genome sequence information for choice of target gene in the design of specific diagnostic tools. This chapter includes three published articles.

Chapter IV. Introduction to insertion sequence elements (IS-elements), assessment of previously described IS-elements from species of the *M. mycoides* cluster in MmmLC strain 95010-C1. This chapter comprises a submitted article.

Chapter V. Description of steps followed in the construction of foreign gene (H-gene) containing plasmid, transformation of MmmLC strain 95010-C1 and expression of H-protein from peste des petits ruminants virus (PPRV).



Introduction to *Mycoplasma*, the *M. mycoides* cluster and Genome sequencing

1.1 The Mycoplasmas

1.1.1 General description and habitat

Mycoplasma is a trivial name for a group of microorganisms that belongs to the class *Mollicutes* (Maniloff, 1992). The name *Mollicutes*, from Latin words *mollis* “soft” and *cutis* “skin” refers to their characteristic lack of a cell wall (Razin et al., 1998). The trivial terms “mycoplasmas” or “mollicutes” have been interchangeably used to denote any species included in *Mollicutes*. Together with *Mycoplasma* the class mollicutes consists of eight other genera: *Ureaplasma*, *Entomoplasma*, *Mesoplasma*, *Spiroplasma*, *Acholeplasma*, *Anaeroplasma, Asteroleplasma* and *Phytoplasma*. Mycoplasmas are the smallest and simplest self-replicating organisms, being built of a plasma membrane, ribosomes, and a circular double-stranded DNA molecule (Figure I.1, Razin, 1997). The lack of a cell wall has not prevented mycoplasmas from flourishing in a wide range of hosts mammals, birds, reptiles, arthropods, plants and fish (Dybvig and Voelker, 1996). The primary habitats of human and animal mycoplasmas are the mucous surfaces of the respiratory, gastrointestinal and urogenital tracts, the eyes, the mammary glands and the joints. The obligatory anaerobic *Anaeroplasmas* have so far been found in the bovine and ovine rumen only. The *Spiroplasmas* and *Phytoplasmas* are widespread in plants but can be found also in the gut, hemocele, and salivary glands of arthropods (Razin et al., 1998).

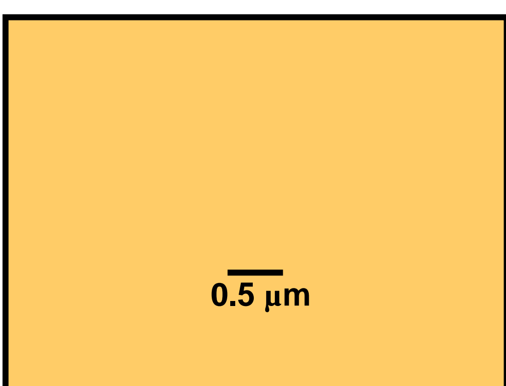


Figure I. 1 Electron micrograph of thin-sectioned mycoplasma cells.

Cells are bounded by a plasma-membrane. The cytoplasm contains thin threads representing sectioned chromosome and dark granules representing ribosomes (Adapted from Shmuel Razin, 1996 Courtesy of RM Cole, Bethesda, Maryland).

These organisms have long resisted detailed analysis because of complex nutritional requirements, poor growth yields and due to paucity of useful genetic tools. So far however, about 190 mycoplasma species have been described under the class *Mollicutes* (Rottem, 2003).

1.1.2 Genome characteristics and phylogeny

The Mycoplasmas have evolved from more conventional progenitors in the *Firmicutes* taxon by a process of massive genome reduction (Glass et al., 2006); their genome size ranges from 0.58-2.2 Mbp. They have been shown to be closely related to gram positive bacteria with a low G + C content in their genome. The reassignment, in most mollicutes of UGA from a stop codon to a tryptophan codon, a feature found in mitochondria, is the apparent outcome of codon reassignment under strong A+T pressure. Not all mollicutes do share this property; the phylogenetically early *Acholeplasma* and *Phytoplasma* use the conventional UGG codon for tryptophan retaining UGA as a stop codon (Razin et al., 1998). Mycoplasma genomes have a limited coding capacity, and as a consequence, they lack many enzymatic pathways characteristic of many bacteria. For example, they lack the pathways for the biosynthesis of purines, a functional tricarboxylic acid cycle and a cytochrome mediated electron transport system (Dybvig and Voelker, 1996; Maniloff, 1996). Mycoplasmas genome size illustrates extreme biological gene economy, imposing complex nutritional requirements for replications and survival. This implies a dependence on external supplies for biosynthetic precursors including amino acids, nucleotides, fatty acids and sterols. The nature of the selective pressure for repeated genome reduction during mycoplasma phylogeny is not known. These features make mycoplasmas suitable research objects in the determination of the minimal gene set required for independent life and to be used as a model in the creation of artificial cells devoid of all non-essential genes (Glass et al., 2006).

Mycoplasma genomes possess only 1-3 rRNA operons. Analysis made by restriction enzyme and hybridization with rDNA probes revealed many mycoplasmas of ruminant hosts to possess only two rRNA operons (*rrnA* and *rrnB*) (Christiansen and Erno, 1990). Based on the sequence analysis of the 16S rRNA genes, mycoplasmas have been divided into five phylogenetic groups including the *Hominis*, the *Pneumoniae*, the *Spiroplasma*, the *Anaeroplasma* and the *Asteroplasma* groups (Figure I. 2, Weisburg et al., 1989). These groups have been further subdivided in to sixteen clusters (Dybvig and Voelker, 1996;

Gasparich et al., 2004; Pettersson et al., 1996; Razin et al., 1998). Some species of importance in veterinary medicine are found within the phylogenetic *Spiroplasma* group in the so-called *Mycoplasma mycoides* cluster, encircled in blue.

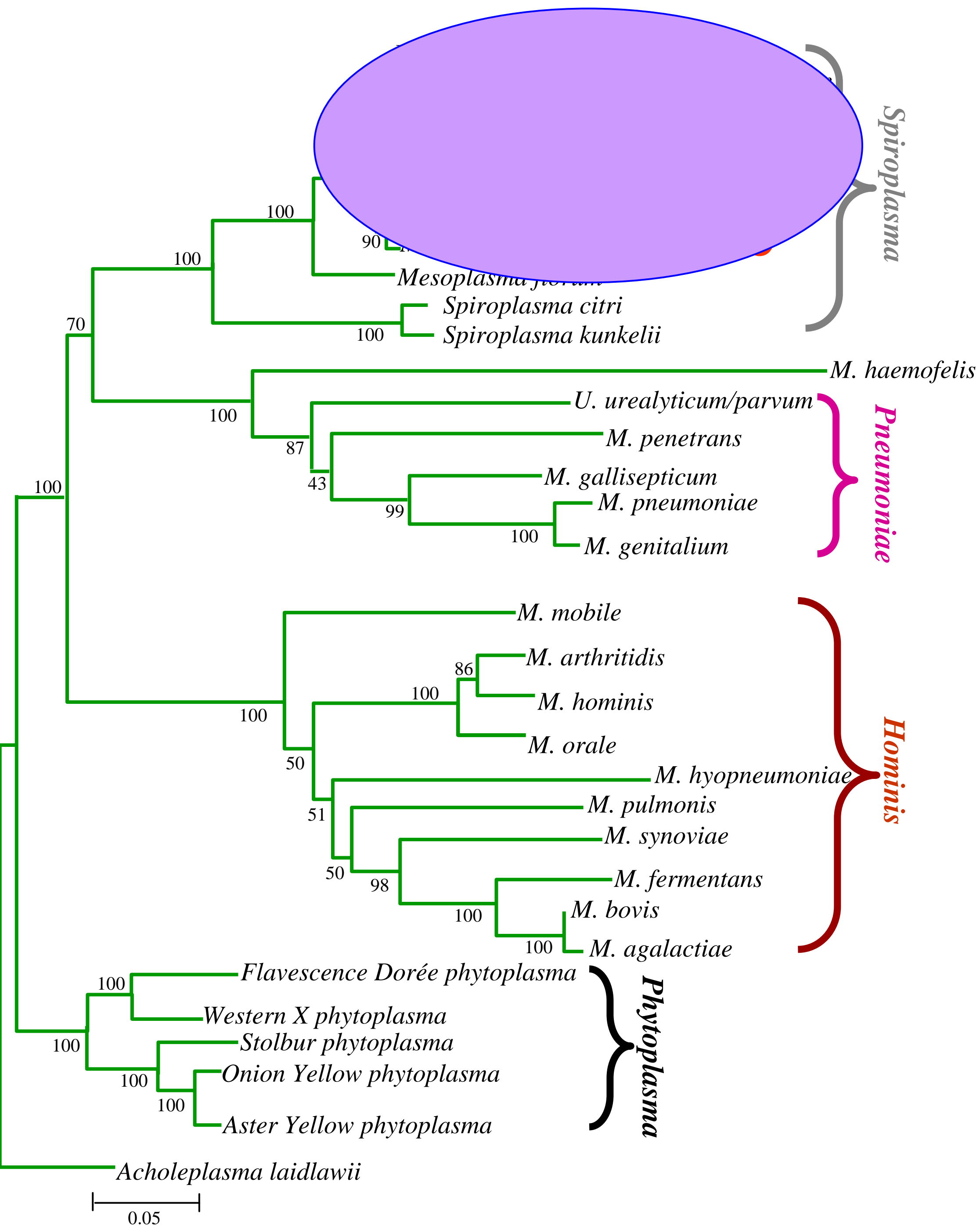


Figure I. 2 Phylogenetic tree of the Mollicutes constructed based on the 16S rDNA sequence. The six species of the *M. mycoides* cluster are encircled in blue. Among them, MmmLC (spotted in red) is the species of interest for the present study (Adapted from MolliGen Web site http://cbi.labri.fr/outils/molligen/).

1.1.3 Pathogenicity in Mycoplasmas

In most pathogenic bacteria pathogenicity is determined by toxin production, cytolysis and a capacity to invade a host cell. In mycoplasmas however pathogenicity mechanisms remain largely elusive. The lesions found in mycoplasma infections in animals and humans are more suggestive of damage due to host immune and inflammatory responses rather than direct toxic effects by the mycoplasma cell components (Razin et al., 1998). Despite the various mycoplasmal virulence factors that have been described, there appears to be no clear casual relationship between these factors and mycoplasma pathogenicity.

1.1.3.1 Adhesion to host cells

Most animal and human mycoplasmas adhere tenaciously to the epithelial linings of the respiratory or urogenital tract, rarely invading tissues. Hence, they may be considered surface parasites. Adherence to the host tissues is a prerequisite for colonization and infection (Razin and Jacobs, 1992). Adherence to host tissue has been extensively studied in *M. pneumoniae*. A virulent *M. pneumoniae* strain has a structure called tip organelle containing densely clustered P1 and P30 proteins. These proteins function both as an attachment and as a leading end in gliding motility (Rottem, 2003). The tip-mediated adherence is more complex and requires additional accessory proteins such as P40, P90 and proteins HMW1-HMW3. This has been elucidated by lack of adherence capacity to the host cell by mutants possessing only P1 and P30 (Razin and Jacobs, 1992).

1.1.3.2 Invasion

Mycoplasma species such as *M. penetrans* and *M. gallisepticum* were shown to invade HeLa cells and chicken embryo fibroblast cells. The exact signals generated by host cells to invasive mycoplasmas have yet to be investigated (Rottem, 2003).

1.1.3.3 Antigenic variation

The other important phenomena in mycoplasmas pathogenicity is the ability to circumvent hosts immunity also referred as phenotypic plasticity. The common way to achieve phenotypic plasticity in mycoplasmas is by “antigenic variation”. In several species of mycoplasmas the rates of antigenic variation have been estimated to be 10^{-2} to 10^{-4} cell per generation (Dybvig and Voelker, 1996). Variable surface proteins in mycoplasmas can be encoded by multiple gene families or a single gene. *Vlp* genes in *M. hyorhinis, vsa* genes in *M. pulmonis*, *vsp* genes in *M. bovis* and *vpma* genes in *M. agalactia* are some examples of

variable surface proteins encoded by multiple gene families. Examples of single gene encoded variable surface proteins include the *vaa* gene in *M. hominis* and the *pvpA* gene in *M. gallisepticum* (Citti et al., 2005).

1.2 The Mycoplasma mycoides cluster

The *M. mycoides* cluster consists of six pathogenic *Mycoplasma* subspecies and an unassigned group of strains causing mild to severe disease in ruminant hosts. These are MmmLC, MmmSC, *M. mycoides* subsp. *capri* (Mmc), *M. capricolum* subsp. *capricolum* (Mcc), *M. capricolum* subsp. *capripneumoniae* (Mccp), and *Mycoplasma* species bovine serogroup seven of Leach (Mbg7) (Cottew et al., 1987). By growth inhibition and immunofluorescence tests Mmc isolates from goats were found to be serologically distinct from *M. mycoides* subsp. *mycoides* isolates from cattle. Some mycoplasma isolates from goats however, were found to be serologically indistinguishable from the *M. mycoides* subsp. *mycoides* (Pettersson et al., 1996). Isolates of *M. mycoides* subsp. *mycoides* from cattle were found to differ in several physiological and biochemical features from *M. mycoides* subsp. *mycoides* isolates from goats. Some strains of *Mycoplasma mycoides* subsp. *mycoides*, mostly isolated from goats, grow to greater turbidity in broth and form larger colonies on solid medium than does isolates from cattle. These isolates also digest casein, liquefy coagulated serum actively and survive longer at 45 °C and are referred to as LC (large colony) strains. Therefore these isolates from goats and cattle were designated as MmmLC and MmmSC types respectively (Cottew and Yeats, 1978). Figure I. 3 shows characteristic colony morphology of mycoplasma on agar plates.

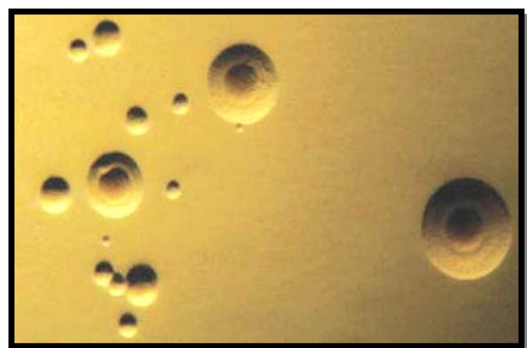


Figure I. 3 Morphology of a typical "fried-egg" mycoplasma colony

Upon culturing both Mccp and MmmSC tend to produce a filamentous growth. This is unlike the other three members of the *M. mycoides* cluster: MmmLC, Mmc, and Mbg7, which grow rapidly producing relatively heavy turbidity and large colonies (DaMassa et al., 1992).

Phylogenetically the analysis of the 16S rRNA genes from MmmLC and Mmc have considered these two as a single entity (Pettersson et al., 1996), they are also classed as a single entity by sequence analysis of a putative membrane protein (Thiaucourt et al., 2000), based on the analysis of *rpoB* gene (Vilei et al., 2006) and when analysing a set of concatenated sequences from housekeeping genes *fusA*, *glpQ*, *gyrB*, *lepA* and *rpoB* (Manso-Silvan et al., 2007). The analysis of the 16S rRNA also categorized Mbg7 with Mcc and Mccp (Pettersson et al., 1996). The close relatedness of Mbg7 to Mcc and Mccp has also been reinforced by the recent phylogenetic analysis based on a set of concatenated sequences from housekeeping genes *fusA*, *glpQ*, *gyrB*, *lepA* and *rpoB* (Manso-Silván et al., 2007).

Phylogenetic classification based on 16S rRNA placed *M. putrefaciens* with in the same phylogenetic group (*Spiroplasma*) as that of the *M. mycoides* cluster (Weisburg et al., 1989). Latter on, sequencing of the 16S rRNA from two saprophytic species *M. cottewii* and *M. yeatsii* (DaMassa et al., 1994) revealed a close similarity of *M. putrefaciens* with these species, with a similarity of 99.7 % with *M. cottewii* and 98.9 % with *M. yeatsii* (Heldtander et al., 1998). These species however, should not be regarded as members of the *M. mycoides* cluster on the basis of serological, biochemical features (Pettersson et al., 1996) and based on the phylogenetic tree derived from distance analysis of five protein coding sequences (Figure I. 4, Manso-Silván et al., 2007).

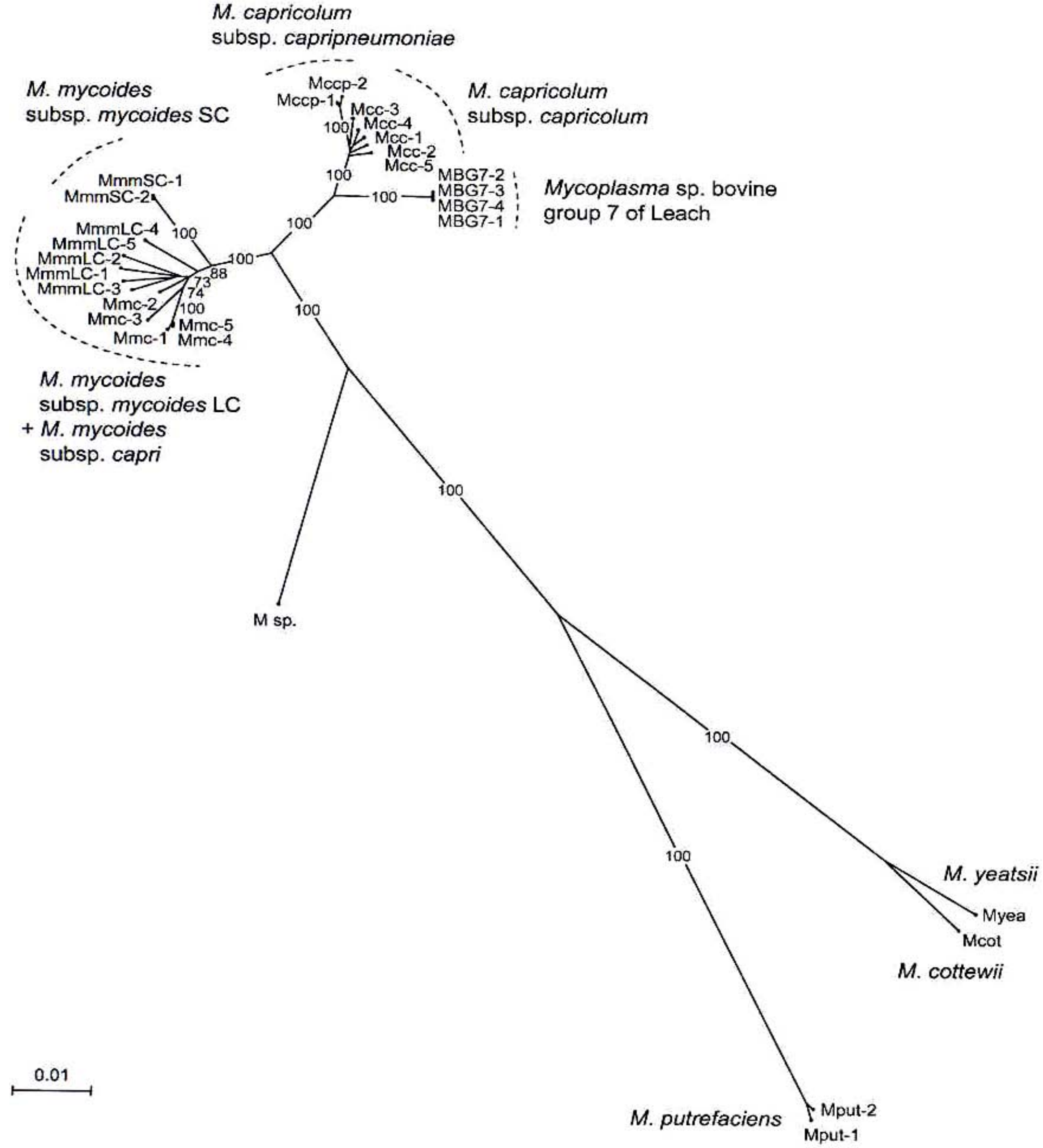


Figure I. 4 Phylogenetic tree obtained from distance analysis of five concatenated protein coding sequences (fusA, glpQ, gyrB, lepA, rpoB).

The tree was constructed using the neighbour-joining algorithm. *M. cottewii*, *M. yeatsii* and *M. putrefaciens* were clearly separated from the *M. mycoides* cluster it also shows the division of *M. mycoides* into two subclusters the *M. mycoides* subcluster containing MmmSC, MmmLC, Mmc and the *M. capricolum* subcluster containing Mcc, Mccp, Mbg7 (Manso-Silván et al., 2007).

1.2.1 Diseases associated with the M. mycoides cluster

1.2.1.1 CBPP and CCPP

MmmSC and Mccp are the causative agents of contagious bovine pleuropneumonia (CBPP) and contagious caprine pleuropneumonia (CCPP), respectively. These two diseases are of great concern in Africa and Asia causing important economic losses. They are classed by the World Organization of Animal Health (OIE) as notifiable animal diseases. Symptoms of CBPP range from hyperacute through acute to chronic forms. The causative agent induces lesions of pleuropneumonia in acute cases and the formation of pulmonary "sequestra" in chronic cases (Bergonier and Thiaucourt, 2003). Nearly 60 years after its description by Nocard and Roux in 1898 the causative agent was identified as a mycoplasma and given the present name *M. mycoides* subsp. *mycoides* (Edward and Freundt, 1956). Strain PG1^T of unknown origin was considered as a type strain of *Mycoplasma mycoides* subsp. *mycoides*.

Similarly, CCPP is a highly contagious disease of goats characterized by fever, severe respiratory distress and high mortality. The causative agent earlier called *Mycoplasma* species type F38 was first isolated from lungs of goat with CCPP in Kenya (MacOwan and Minette, 1976). Later on, based on DNA-DNA relatedness, the type strain F38^T and like strains were found similar to Mcc (Bonnet et al., 1993). However, the two mycoplasmas have markedly different growth, cultural characteristics and biochemical features. Finally, F38^T-like strains were grouped under a new subspecies: Mccp (Leach et al., 1993). The highly fastidious nature of Mccp strains has brought confusion regarding the exact causative agent of CCPP. Two other members of the *M. mycoides* cluster, MmmLC and Mmc, have been for sometime wrongly implicated in the etiology of the disease because of the pleuropneumonia that may be observed in small ruminants resembling CCPP.

1.2.1.2 Contagious agalactia

From the six members of this cluster, MmmLC, Mmc and Mcc are described as causative agents of contagious agalactia syndrome. Contagious agalactia is an infectious disease of sheep and goats known for nearly 200 years. The clinical disease was first described by Metaxa in Italy in 1816 and was given the name contagious agalactia by Brusasco in 1871 (Madanat et al., 2001). At present, contagious agalactia occurs in most countries with intensive and extensive production of sheep and goats (i.e., in the Mediterranean region

and the Balkan peninsula in Europe, in western Asia and in northern, central and eastern Africa (Bergonier et al., 1997)).

The major causal agent of the disease, *M. agalactiae*, was isolated by Bridre and Donatien in 1923 as the second known mycoplasma species. *M. agalactiae* is regarded, particularly in sheep, as the “classical” aetiological agent of contagious agalactia (Bergonier 1997). However, similar clinical and pathological features can be produced in small ruminants by mycoplasma species of the *M. mycoides* cluster (Bergonier et al., 1997; Nicholas et al., 1996). In goats, a disease with nearly identical clinical manifestations is also caused by *M. putrefaciens*. The consensus of the working group on contagious agalactia of the EC COST Action 826 (European Cooperation in the field of Scientific and Technical Research) on ruminant mycoplasmosis (met at Toulouse, France, 1999) was considered four mycoplasmas as causal agents of contagious agalactia. This decision has not taken into account Mmc, the closest relative of MmmLC.

In the most recent survey for mycoplasma species in contagious agalactia, MmmLC was found to be the most frequently isolated aetiological agent in Gran Canaria, Spain (De la Fe et al., 2005). MmmLC has the widest geographical distribution among ruminant mycoplasma. This organism is present in all the continents, including South America. It is reported where small ruminants are kept and wherever contagious agalactia and caprine pleuropneumonia are present (Bergonier et al., 1997). MmmLC is probably under-reported due to the lack of diagnostic facilities for mycoplasma diseases in many countries. MmmLC is mostly confined to goats but has occasionally been isolated from sheep with balanoposthitis and vulvovaginitis (Trichard et al., 1993) and from cattle (Perreau and Bind, 1981). Clinical cases usually occur sporadically, but the disease may persist and spread slowly within a herd (Bergonier et al., 1997).

Although contagious agalactia does not induce high mortality; morbidity in a herd may reach 30-60 %. Reduction or complete cessation of milk production and abortions in pregnant animals is responsible for economic losses. Severe cases of the disease in herds may result in the death of lambs and kids up to 40-70 %. In countries where sheep and goat dairy products play important roles as food components and export commodities, contagious agalactia is a serious veterinary public health problem (Madanat et al., 2001).

Usually the disease is manifested by mastitis, arthritis, pleurisy, pneumonia, keratoconjunctivitis and septicaemia. These clinical signs were the basis for the acronym MAKePS (Thiaucourt and Bolske, 1996). Septicaemia outbreak has been observed in kids exposed to large doses of milk from mastitic does excreting MmmLC organism. The disease is usually fatal with necropsy changes characterized by fibrinopurulent polyarthritis, embolic pneumonia, renal infarcts and thrombo-embolic lesions in various tissues resembling disseminated intravascular coagulation. Animals surviving the septicaemia may develop chronic destructive arthritis in one or more joints. Bacteraemia is common, particularly for MmmLC and Mcc and could account for the isolation of the organism from sites where it is only transiently present (Rosendal, 1993).

1.2.1.3 Other disease conditions

The sixth member of the *M. mycoides* cluster, Mbg7 is associated with polyarthritis, mastitis and pneumonia in cattle. More recently Hum and colleagues reported a severe outbreak of polyarthritis, mastitis and abortion affecting about 120 cattle in large dairy operation system in Australia, where by Mbg7 was identified as a causative agent (Hum et al., 2000).

1.2.2 Pathogenicity of members of the M. mycoides cluster

Little is known on the mechanism of pathogenicity with in the species of the *M. mycoides* cluster. MmmSC biotype the causative agent of CBPP has been the most extensively studied mycoplasma species of the *M. mycoides* cluster. Even if some hypotheses have been described, as for other mycoplasma species, the mechanism of pathogenicity in this cluster remains elusive.

1.2.2.1 Variable surface protein

A variable surface protein Vmm has been identified and characterized in MmmSC biotype. The *vmm* gene encodes a lipoprotein precursor of 59 aa, where the mature protein was predicted to be 36 aa and was anchored to the membrane by only the lipid moiety, as no transmembrane region could be identified. The protein was found to undergo reversible phase variation at a frequency of 9×10^{-4} to 5×10^{-5} per cell per generation. Vmm-like genes were also found in the other three members of the *M. mycoides* cluster: Mcc, Mccp, Mbg7 and in *M. putrefaciens* (Persson et al., 2002). The *vmm* gene in MmmSC is an example of a single gene encoded variable surface protein (Citti et al., 2005). A recent

whole genome sequence of Mcc type strain California kid^{T} has revealed genes encoding a diverse family of variable surface proteins termed the Vmc system. The *vmc* genes of Mcc present not only an alternative surface structure but also a system that permits high-frequency phase variable expression as well as structural variation (Wise et al., 2006).

1.2.2.2 Metabolic byproduct

In more recent investigations it has been shown glycerol metabolism in MmmSC strains release hydrogen peroxide (H_2O_2) as a by product, resulting in disruption of host cell integrity. The initial hypothesis was based on the fact that virulent MmmSC African strains possessed an active ATP-binding cassette (ABC) transport system for the utilization of glycerol, which is metabolized to dihydroxyacetone-phosphate (DHAP) releasing H_2O_2, while European strains lacked part of the glycerol uptake genes due to deletion and are less virulent (Vilei and Frey, 2001). H_2O_2 in MmmSC is proved to be produced by a membrane located enzyme L-α-glycerophosphte oxidase (GlpO) that is involved in glycerol metabolism (Pilo et al., 2005).

1.2.2.3 Capsular polysaccharide

An important surface antigen and pathogenicity factor in MmmSC is the capsular polysaccharide (CPS), previously known as galactan. Injection of purified CPS to cattle produced severe respiratory collapse and even death (Buttery et al., 1976; Cottew, 1979). The anti-CPS titre was related to the ability of antisera to inhibit MmmSC growth *in vitro*; conversely, strains producing high levels of CPS were found to be the most resistant to growth inhibition tests (Waite and March, 2002). CPS is a high molecular weight polymer composed predominantly or exclusively of galactose subunits. It is unclear why a polygalactan should exhibit such effects in cattle, but the unusual 1-6 β linked structure of polymer may play some role. Studies by (Waite and March, 2001) have suggested the presence of small amounts of mannose and fructose in CPS extracts, which might also contribute to pathogenicity by binding CPS to host cell receptors. Whether these sugars are an integral component of CPS or represent an additional or contaminating polysaccharide, is unknown (Waite and March, 2002).

1.3 Genome sequencing approach in Mycoplasmas

Among bacterial species, with their small genome size, mycoplasmas play a central role in the history of genome sequencing. Since 1995 mycoplasma genome sequencing has advanced to a stage that we have now seventeen completed genome sequences of mycoplasmas. These are *M. agalactiae*, *M. capricolum* subsp. *capricolum, M. mycoides* subsp. *mycoides* SC, *Mesoplasma florum*, *Ureaplasma urealyticum/ parvum*, *M. penetrans*, *M. galisepticum*, *M. pneumoniae*, *M. genitalium*, *M. mobile*, *M. hyopneumoniae* strains 232, 7448 and J, *M. pulmonis*, *M. synoviae*, Onion yellow phytoplasma and Aster yellow witches-broom phytoplasma. The genomes of so far sequenced mycoplasmas are accessible on line in MolliGen, a database dedicated to the exploration and the comparison of mollicutes genome http://cbi.labri.fr/outils/molligen (Barre et al., 2004), and a dozen more are under sequencing phase (http://www.tigr.org/tdb/index.html).

Although there are a number of investigations to understand the exact mechanisms of pathogenicity in mycoplasmas, their fastidious nature and lack of genetic systems to effectively dissect their structure and function hampered progress towards understanding their molecular biology. In very closely related species, such as the members of the *M. mycoides* cluster, genome sequencing and comparative genome analysis is imperative to identify genes potentially linked to virulence. In this context, the whole genome sequence of an MmmLC may contribute to the research on the molecular mechanism of pathogenicity in MmmLC and in the *M. mycoides* cluster at large. To date, there are different means's of achieving bacterial whole genome sequence.

1.3.1 Genome sequencing

Genome sequencing is changing the landscape of modern biology as being a new approach to the study of genes and their functions. The genome of a laboratory strain of *Haemophilus influenzae* was the first bacterium to be sequenced and was followed by an isolate of *Mycoplasma genitalium* (Fleischmann et al., 1995; Fraser et al., 1995). Although the lists of completed bacterial genome sequences are dominated by pathogenic species; over 400 bacterial genome sequences are currently available in public databases representing species as well as multiple strains of the same species (http://www.tigr.org/tdb/index.html). Genome sequencing in bacteria is providing novel insights into the intricacies of pathogen-host interactions and co-evolution. These genomes are also significant because they mark the beginning of an important trend in the

sequencing of closely related genomes, including the sequencing of more than one strain from a single pathogenic species. The comparison of the genomes of pathogenic and non-pathogenic organisms will most probably provide an insight into what makes certain bacterial strains and species pathogenic (Field et al., 1999). To date quite a number of institutions are involved in large scale projects of whole genome sequence; the goal of most projects is a finished contiguous DNA sequence of a bacterium chromosome. Some websites are trying to allow access to information regarding complete and ongoing genome projects (genome sequence institutions data from Gold Genomes Online Database v.2) (http://www.genomesonline.org).

1.3.1.1 Genome sequencing strategies

Genome-sequencing frequently uses three types of strategies, the ordered-clone approach (primer-walking), the random-sequencing approach (shotgun sequencing) and a mixed strategy of both approaches (Figure I. 5).

The ordered-clone strategy uses a large-insert library to construct a map of overlapping clones covering the whole genome from which selected clones are sequenced by primer-walking technique to obtain the whole-genome sequence. This strategy uses several methods including restriction fingerprinting and hybridization mapping (Frangeul et al., 1999). Restriction fingerprinting is a procedure for clone comparison based on the matching of characteristic restriction fragment sets (Gregory et al., 1997) and hybridization, which is a simple and rapid method for identifying stretches of homologous DNA. With this technique a large number of clones can be analysed and ordered concurrently. Nevertheless, the primer synthesis and purification procedure will take a day and consequently slow down the process relative to shotgun strategy. The sequence redundancy will normally be 2-3 times the genome size with the primer-walking strategy. Consequently the accuracy of the consensus sequence will be lower than the shotgun approach and the assembly phase will also take longer time (Westberg, 2003).

Currently however, the most widely used strategy for the sequencing of a microbial genome is whole-genome shotgun sequencing; this strategy does not require preliminary data (such as a map) before the sequencing phase (Frangeul et al., 1999).

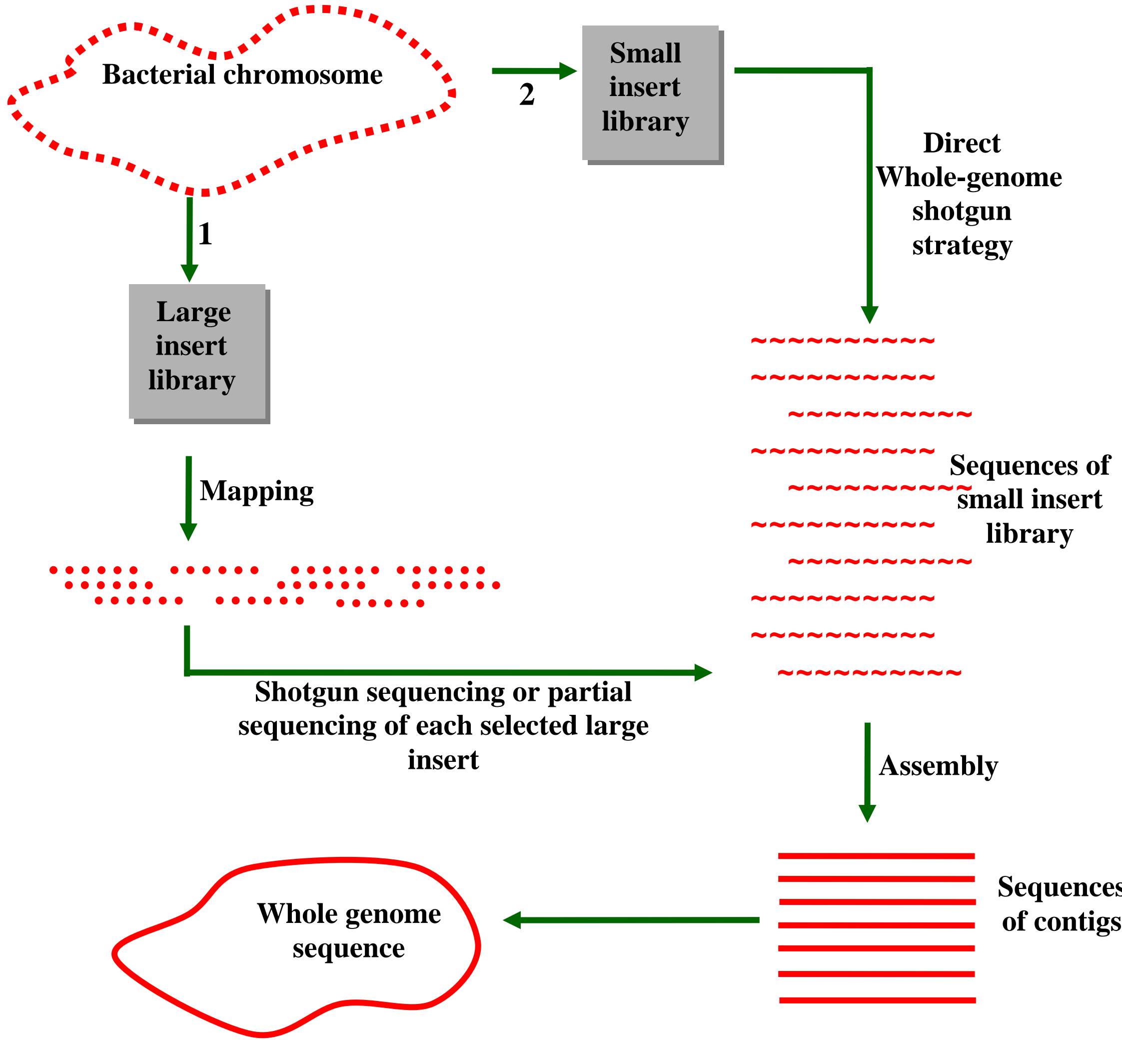


Figure I. 5 Strategies used to obtain whole genome sequences.

1=the ordered-clone approach with the help of a map constructed from a large-insert and 2=the direct shotgun approach with the utilization of a small-insert library (Frangeul et al., 1999).

In case of whole genome shotgun sequencing, a large number of clones from libraries representative of the whole genome are sequenced and assembled into contigs. The contigs are then joined together to obtain the whole genome sequence in a single contig using a variety of methods such as specific PCR products, or cloned inserts that span each gap (Frangeul et al., 1999). This procedure is mainly suited for large sequencing facilities, where large contigs are produced relatively fast. However, most prokaryotic genomes are more or less redundant in their sequence and are often containing large repetitive sequences with high sequence similarity, which will make the assembly of the shotgun

sequences impossible. Accordingly, most genome sequencing projects will never be completed by using solely shotgun sequencing strategy (Westberg, 2003).

There is not a single sequencing strategy that is optimal for all microbial sequencing projects as different genomes have different features of difficulties. For instance, the genomes differ in size, G+C content and in the number of repetitive regions, the organisms are more or less difficult to cultivate and the size of the sequencing facilities is varying a lot. Therefore, a mixed approach of the above two strategies has been a strategy of choice in genome sequencing. Typically, there is an initial random sequencing phase, which is followed by a direct sequencing approach that completes the genome sequence. In general the genome project is entering the directed phase when the random phase is not producing unique sequences in a sufficiently high speed, normally at 8-10 fold genome coverage.

1.3.1.2 Sequencing techniques

The pioneer DNA sequencing technique is the chain termination technique of Sanger (Sanger et al., 1977). This method produces a nested set of DNA fragments by extending a primer hybridized to a DNA template and interrupting the process with nucleotide terminators, chemically altered "dideoxy" bases to terminate newly synthesized DNA fragments at specific bases (either A, C, T, or G). These fragments are then size-separated, and the DNA sequence can be read. The second most used sequencing method is the chemical degradation method most similar to Sanger method. In this technique the nested set of fragments are generated by cleavage at specific nucleotides with different chemicals (Maxam and Gilbert, 1977) this method however, is no more in use due to the hazardous chemicals. The most recently developed sequencing technique is pyrosequencing (Ronaghi et al., 1998), the principle of this technique is explained in more detail below.

Pyrosequencing (454 Life Sciences) is a rapid shot-gun genome sequencing that comprises two systems: GS20 released in 2005 and GSFLX released in 2007, reading 100 bp at a time and up to 300 bp respectively. These systems support simultaneous sequencing of samples from a variety of starting materials, including genomic DNA, PCR products, BACs and cDNA. In this technique the shotgun fragmentation of the genome is followed by preparation of single stranded template DNA (sstDNA) library conjugated with the so called “A” and “B” adaptors. The double-stranded DNA fragments are blunt ended and phosphorylated. Short adaptors, “A” and “B”, are then ligated onto the ends of the

fragments. The "B" adaptor contains a 5' biotin tag that enables immobilization of the library onto streptavidin coated beads. The "4 base key" is a short sequence of 4 nucleotides included in the adaptor used by the system software for base calling and to recognize legitimate library reads. Four types of products are generated from the ligation. Products containing the biotinylated "B" adaptors are captured by streptavidin coated magnetic beads. Products with two "A" adaptors do not bind to the bead and are washed away. The double stranded products bound to the bead are denatured. The non biotinylated strands containing an "A" and "B" adaptor are recovered. The recovered strands are used as a sstDNA library.

The sstDNA library is immobilized onto capture beads by hybridization to complementary primers attached to the capture beads. The beads containing a library fragment carry one sstDNA molecule. Beads containing sstDNA emulsified with amplification reagents in a water-in-oil mixture. Each bead is captured within its own microreactor where PCR amplification occurs; resulting in beads immobilized clonally amplified DNA fragments called emulsion PCR (emPCR). In order to start the sequencing, sstDNA template is hybridized to a sequencing primer and incubated with the enzymes DNA polymerase, ATP sulfurylase, luciferase and apyrase, and with the substrates adenosine 5' phosphosulfate (APS) and luciferin. This incubation mix then layered onto PicoTiterPlate and the plates centrifuged to deposit the beads into the wells and are fixed in place by packing beads. Finally the loaded PicoTiterPlate device is placed into the GS20 or GSFLX instrument. Nucleotides are then flowed sequentially in a fixed order, TACG, across the PicoTiterPlate device during which hundreds of thousands of beads with millions of copies of DNA are sequenced in parallel (www.roche-applied-science.com). This system can generate 100 million nucleotide data in a 7 hour run with a single machine. DNA polymerase incorporates the correct, complementary dNTPs onto the template. This incorporation releases pyrophosphate (PPi) stoichiometrically. ATP sulfurylase quantitatively converts PPi to ATP in the presence of adenosine 5´ phosphosulfate. This ATP acts as fuel to the luciferase-mediated conversion of luciferin to oxyluciferin that generates visible light in amounts that are proportional to the amount of ATP. The light produced in the luciferase-catalyzed reaction is detected by a charge coupled device (CCD) camera and this can be analyzed in a program. Each light signal is proportional to the number of nucleotides incorporated. Unincorporated nucleotides and ATP are degraded by the enzyme apyrase (Figure I. 6).

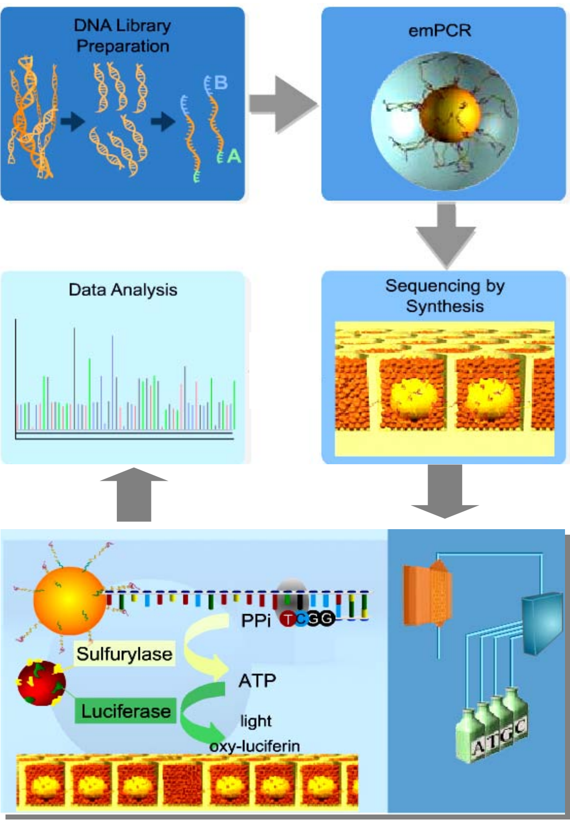


Figure I. 6 Diagram showing steps taken in Pyrosequencing.
Preparation of single stranded DNA ligated with A and B adapters, Emulsion based clonal amplification of single stranded DNA molecule by emPCR, followed by sequencing by synthesis where DNA polymerization produces light corresponding to the incorporated nucleotides (Roche applied science).

The limitation of this technique is that the lengths of individual reads of DNA sequence are in the neighborhood of 250 nucleotides, shorter than the 800-1000 obtainable with Sanger sequencing. This can make the process of genome assembly more difficult, particularly for sequences containing a large amount of repetitive DNA. As of 2007, Pyrosequencing is commonly used for re-sequencing of closely related strains or sequencing of genomes for which the sequence of a close relative is already available. The already available genome sequence will be used as a DNA back-bone and allows the detection of SNPs.

1.3.1.3 Genomic sequence analysis

The increasing numbers of sequenced genomes have led to the development of different annotation tools (Glasner et al., 2006; McNeil et al., 2007; Overbeek et al., 2005;

Romualdi et al., 2005; Selengut et al., 2007; Tcherepanov et al., 2006; Viswanath et al., 2007; Zhu et al., 2007). The common procedure is to identify as many genes encoding proteins, tRNAs and rRNAs as possible there after assigning them a function. There are a number of programs that find protein-coding genes such as GeneMark, Glimmer and Orpheus. The normal procedure is to look for an open reading frame (ORF) above a certain size. The probability to find a gene in an ORF is lower for shorter ORFs. Another difficulty is to find the correct 5' boundary of the gene when the gene contains multiple start codons. Coding regions have a more organized nucleotide pattern frequency than non-coding regions (Frangeul et al., 1999). The ribosomal binding site (RBS) is one of those important signals for the identification of genes in a DNA sequence, since almost every bacterial mRNA (messenger RNA) has an RBS, to the polypeptidical product to be produced. The RBS is a region where the ribosome binds to an mRNA to begin the translation of mRNA into a protein (Oliveira et al., 2004).

1.4 Genome sequencing of *M. mycoides* subsp. *mycoides* LC

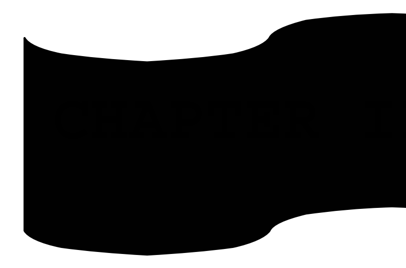
During the course of this research work, whole genome sequencing of MmmLC strain 95010-C1 was performed using whole genome shotgun strategy at Genoscope (Evry, France). MmmLC strain 95010-C1 was isolated at CIRAD in 1995 from a clinical case of goat mastitis occurring in France. The genetic material was extracted from a clone culture using classical phenol: chloroform extraction. Sequencing was carried out using 3 plasmid libraries. 7,000 inserts of 3 Kb were cloned into the multi-copy plasmid pcDNA2.1 (Invitrogen, USA). 2,500 inserts of 10 Kb fragment sizes were cloned into the small copy number plasmid pCNS. Finally, about 800 inserts of 20-25 Kb were cloned into the pBBc vector. A total of 24,000 readings were performed to obtain total 12-fold sequence coverage. The finishing step was also performed at Genoscope.

1.5 Annotation of *M. mycoides* subsp. *mycoides* LC genome sequence

As part of my PhD research activities, I participated in the annotation of the genome sequence of strain 95010-C1. Annotation was done in collaboration with INRA Bordeaux, using the Contigs-Assembly and Annotation Tool-Box (CAAT-box) platform (Frangeul et al., 2004). This software was developed for the computational part of a genome project where the sequence is obtained by a shotgun strategy. CAAT-Box allows annotation of contigs to start during the finishing phase of sequencing projects. The DNA sequence in

CAAT-box is segmented into sequence fragments 500 nucleotides before and 200 nucleotides after each identified ORF. These segments are also known as individual protein files (IPFs). Each IPF is identified by a unique number and is linked to the assembly by the identification of the contig on which it is located and is by its position on the contig (Frangeul et al., 2004). Certain IPFs which do not seem to contain a coding sequence were automatically marked as "FALSEORF". The CAAT-box platform (http://cbi.labri.fr/outils/CAAT-Box/MYCOLC/IPFutil.html) incorporates a number of features that facilitate the annotation process of each IPF. This platform is used to determine CDSs with potential start codons AUG, UUG and GUG and potential stop codons UAG and UAA, including the most probable ribosome binding sites, using the BLAST and Gene mark programs (Altschul et al., 1990; Isono et al., 1994). Three different databases were used in the annotation process: SwissProt (http://www.ebi.ac.uk/swissprot/index.html), (http://www.ebi.ac.uk/embl/index.html) TREMBLE and MolliGen (http://cbi.labri.fr/outils/molligen). In addition to these three databases, the platform includes the tools "InterProScan" and "PrositeScan" for domain detection and TMHMM for trans-membrane segment prediction (Sirand-Pugnet et al., 2007). During the annotation process proteins were considered to be similar when the identity exceeded 40 %. Predicted proteins with lower or local similarities to previously characterized proteins were annotated as hypothetical proteins.

The objective of this thesis was to use comparative genomics within the *M. mycoides* cluster for the development of specific and robust molecular diagnostic tools for the identification of several members of the *M. mycoides* cluster. The work then advanced to the study of functional genomics, primarily to evaluate the capacity of an MmmLC strain under whole genome sequencing to be transformed and to express foreign genes. All along these studies, genome sequencing and annotation of MmmLC strain 95010-C1 has been a baseline study guide.



Genome size estimation of MmmLC strain 95010-C1 and analysis of presence of repeated genes

1 INTRODUCTION

1.1 Genome size estimation

While the whole genome sequence of MmmLC strain 95010-C1 was underway in Genoscope (Evry, France), we have undertaken this part of the study with two main objectives: (1) estimation of the genome size of strain 95010-C1 and (2) validation of the accuracy of genome sequence assembly using data from size estimation and DNA hybridization tests.

Pulse field gel electrophoresis (PFGE) was used for genome size estimation of MmmLC strain 95010-C1. PFGE is one of the most important techniques developed for genome size estimation and physical genomic mapping of bacterial DNA (Birren et al., 1989b). Gel electrophoresis is a method of choice for size fractioning of DNA in analytical biochemistry. Depending on the intended application a number of systems have been used to facilitate the separation of different fragment sizes. Conventional agarose gel separates DNA using steady fields and it is commonly used to separate molecules from a few base pairs to about 20 Kbp, while PFGE is used to separate molecules beyond this size range, up to 10 Mbp. In PFGE, the electric field is periodically alternated in two directions and DNA separation depends on the way the molecules reorient through the gel in response to the changing electric field (Gurrieri et al., 1999; Wrestler et al., 1996).

The application of PFGE to genome size estimation has provided a much more accurate and labor-saving procedure than the previously used renaturation kinetics method and has resulted in a wealth of genome size data (Bové, 1993). Genome size in mollicutes is variable not only within the same genus but sometimes among strains of the same species. One of the reasons for this variability is the frequent occurrence in these genomes of repetitive elements, consisting of segments of protein genes differing in size and number, or insertion sequence (IS) elements (Razin et al., 1998). Genome size among mollicutes varies from less than 600 Kbp in *M. genitalium* to over 2.2 Mbp in *Spiroplasma*, with values overlapping between genera (Bové, 1993).

Beside genome size estimation, PFGE has been used for more than two decades as a tool for molecular typing of strains and it has shown to be discriminatory and reproducible (Chen et al., 2005). This technique allows discrimination of strains differing by the presence of insertions or deletions. It can also detect recombination events if they include

an enzyme recognition site. For example, profiles showing a single band difference were observed for all MmmSC strains from Tanzania as compared to PG1^T (Kusiluka et al., 2001a). PFGE fragments generated by rare-cutting restriction endonucleases have permitted molecular typing of *M. bovis* (McAuliffe et al., 2004) and MmmLC biotype (Tardy et al., 2007). The drawbacks of typing with PFGE technique are that not all isolates are typeable using this technique, that it requires culturing, and that the procedure is difficult to use routinely (Marois et al., 2001). Beside strain typing, PFGE together with DNA hybridization, may be implemented for the localization of certain genes, and in the estimation of the copy number of these genes in the genome (Blank and Stemke, 2000; Ye et al., 1992).

In the present study, DNA fragments, generated by the PFGE technique, were used to estimate the genome size of MmmLC strain 95010-C1. As a control, the fully sequenced strain PG1^T was included to compare estimated fragment sizes with real data by comparing PFGE results and *in-silico* data. MmmSC strains included all along the study together with type strain PG1^T were 8740-Rita (whole genome sequencing underway) and vaccine strain T1/44. Strain Y-goat^T of the MmmLC biotype was also included. Additional strains California kid^T of Mcc, strain L of *M. bovis* and strain PG2^T of *M. agalactia* have also been analyzed.

1.2 DNA Hybridization

Although certain genome projects do not encounter significant problems during their finishing phase, others have been quite difficult to finish. The reason for this may be the presence of unclonable genes and repeated regions (Frangeul et al., 1999). For instance, during the sequencing of strain PG1^T the presence of high copy numbers of insertion sequence elements, the large size of long repeats and high similarities between their copies caused problem in genomic assembly (Westberg, 2003). Upon completion, the PG1^T genome revealed the presence of large repeated genomic segments comprising 29 % of the total genome size and was found to possess the highest density of insertion sequence elements known to bacteria (Westberg et al., 2004). Genome analysis of sequenced strain PG1^T of MmmSC biotype showed four regions containing long repeats of 24 Kbp, 13 Kbp, 12 Kbp that were duplicated twice and 8 Kbp duplicated three times (Bischof et al., 2006).

In order to assess gene duplication in the genome of MmmLC strain 95010-C1 we have analyzed the following genes: *lppQ*, coding for lipoprotein Q and found within the 24 Kbp long repeat, *mgtA*, coding for enzyme magnesium transport ATpase, found within the 13 Kbp long repeat, *phnC*, coding for alkylophosphonate ABC transporter, present in the 12 Kbp long repeat, and finally *galE* and *epsG*, coding for UDP-glucose 4-epimerase and glycosyltransferase respectively and located within the 8 Kbp long repeat.

One way to fingerprint DNA is by doing a Southern blot. A Southern blot is a method routinely used in molecular biology to check for the presence of a specific sequence in a DNA sample. Southern blotting combines agarose gel electrophoresis for the size separation of DNA with methods to transfer the size separated DNA to a filter membrane for probe hybridization. The method is named after its inventor, the British biologist Edwin Southern (Southern, 1975).

In this part of the study Southern blot was used to asses the presence of repeated genes in MmmLC strain 95010-C1. As it was done for the physical genome mapping, additional strains Y-goat^T from MmmLC, strains PG1^T and 8740-Rita from MmmSC were analysed simultaneously. DNA probes of the genes listed were used in order to verify if MmmLC possessed multiple copies of the selected genes. These data were then used to validate whole genome sequence results for strain 95010-C1.

2 MATERIALS AND METHODS

2.1 Strains

Mycoplasma strains PG1^T, 8740-Rita (MmmSC biotype) and 95010-C1, Y-goat^T (MmmLC biotype) were used for both PFGE and Southern blot techniques. However, for the analysis by PFGE additional strains T1/44 (MmmSC), California kid^T (Mcc), PG2^T (*M. agalactiae*) and L_2 (*M. bovis*) were included in the study.

2.2 DNA plug preparation

Mycoplasmas were cultivated in modified Hayflick media (PPLO broth without crystal violet 21 g/l, de-complemented horse serum 20 %, fresh yeast extract 10 %, glucose 0.2 % and sodium pyruvate 0.4 %) (Thiaucourt et al., 1992). Cultures at stationary phase (40 ml) were harvested by spinning at 13000 g for 30 min at +4°C. This was then followed by washing of the mycoplasma cells three times with 1XPBS and twice with 10 % sucrose, centrifugation at 12000 g for 25 min at +4°C. Cells were re-suspended in 500 µl of 250 mM EDTA pH8, 20.6 % sucrose. One millimeter-thick agarose plugs (using Plug Molds, Bio-Rad) were made with a 1:1 mixture of 1 % low melting agarose (Bio-Rad) prepared in 25 mM EDTA pH 8, 10.3 % sucrose and the cell suspension. The plugs were then incubated in digestion buffer (10 mM Tris-HCl, 0.5 M EDTA, 1 % lauroyl sarcosine, 0.5 mg/ml proteinase K) for 48 hrs at 50°C.

2.3 Restriction Enzyme Analysis

2.3.1 Choice of restriction enzymes and digestion for PFGE

In-silico enzymatic digestion of the PG1^T genome (http://www.*in-silico*.com) and data obtained from previous studies (Kusiluka et al., 2001b; Pyle et al., 1990) were used as a reference in the selection of restriction enzymes *BamH*I, *Bgl*I, *Mlu*I and *Nco*I (New England Bio labs). For digestion, agarose plugs were dialyzed overnight by placing them in a 50 ml conical tube containing TE (10 mM Tris, pH 7.4, 1 mM EDTA) and washed three times at room temperature with slow agitation. DNA-agarose plugs were equilibrated for 1 hr in 160 µl 1X restriction buffer (New England Bio labs), 60 Units of enzyme were mixed in the same volume of fresh 1X restriction buffer and digestion reactions proceeded for 16 hrs at 37°C. For each enzyme assay a mock digestion was included (i.e. incubation of DNA-plugs in buffer mix without addition of restriction enzyme).

2.3.2 Choice of restriction enzymes and digestion for Southern blotting

Restriction enzymes used for the analysis of duplicated genes were chosen with the help of Vector NTI software. A large sequence containing a duplicated zone was extracted from the PG1^T genome sequence and analyzed for non-cutting enzymes of the genes of interest (*lppQ, phnC*, *epsG*, *mgtA,* and *galE*). Enzymes *Apal*I, *Eco*RV, *Eco*RI, and *Xba*I (New England Biolabs) were chosen for the digestion of genomic DNA from MmmLC strains 95010-C1 and Y-goat^T and MmmSC strains PG1^T and 8740-Rita. Restriction enzyme digests were performed using standard procedures as recommended by the manufacturer (New England Biolabs).

2.4 PFGE Electrophoresis conditions

The digested fragments in plugs were separated on 1.2 % pulse field certified agarose (Bio Rad) gels; running conditions were adjusted according to fragment sizes. The CHEF-DR II (Bio-Rad, Richmond, CA, USA) that uses two homogenous electric fields fixed at 120° reorientation angle was used for migration. Because of the highly uniform field it generates this system is supposed to give better gel runs. For digests using *BamH*I, gels were run at 6 Vcm^{-1} for 16 hr with the pulse ramp time varied from 0.1 to 28 sec in 0.5XTBE buffer (45 mM Tris, 45 mM borate, 1 mM EDTA, pH 8.2) at 14°C. A lambda ladder and low range PFG marker (New England Biolabs) were used as molecular weight markers. In the case of *Mlu*I and *Nco*I digests, the same parameter was used except for the switching time that varied from 1 sec to 42 sec. The gels were stained with 1 µg/ml ethidium bromide for 15 min, de-stained with water for 15 min and the DNA bands were visualized using a UV light.

2.5 DNA probes for Southern blot hybridization test

Digoxigenin-labeled specific DNA probes were prepared for each of the genes to be tested *lppQ*, *phnC*, *epsG*, *galE* and *mgtA*. Primer pairs for this purpose were designed from corresponding genes of the genome sequence of PG1^T and named accordingly as LppQ-F/R, phnc-F/R, epsG-F/R, galE-F/R, and mgtA-F/R (Table II. 1). DNA probes of 758, 821, 716, 601, 747bp-long were prepared for genes *epsG*, *galE, mgtA, phnC* and *lppQ* respectively. Digoxigenin (DIG) PCR labelling was performed according to the manufacturers instructions (Roche Diagnostics GmbH, Mannheim, Germany). The reaction volume was fixed at 100 µl with: 10 µl of 10 X Taq buffer (Roche), a 1:6 ratio of dig11dUTP: dTTP i.e. 30 µM dig-11-dUTP, 180 µM dTTP, 200 µM dATP, 200 µM dGTP

and 200 μM dCTP (Roche Diagnostics), 0.4 μM of each of the above primers, 5.25 U of Taq long expand DNA polymerase (Roche Diagnostics) and 1 μl of a 20 ng/μl DNA template (DNA from type strain PG1^T). For all the DNA probes the PCR conditions were consisted of an initial denaturation step at 94°C for 2 min followed by 30 cycles of 94°C for 20 sec, 52°C for 20 sec and 68°C for 30 sec.

Table II. 1 Primers designed from the PG1^T genome sequence for the preparation of specific DNA probes

Primer name	Sequence 5'-3'	Locus Tags	Fragment (bp)	Tm (°C)
LppQ-F	CGATTCTAAAATTATAAAAGGTGAACT			52
LppQ-R	TCATGAGCACCATTAAACATACTATC	MSC_1021	747	51
epsG-F	TAATCACGGATGAAGATACATTCG			52
epsG-R	CAAGTATAACTAATGGTGTTGTTGTAAAA	MSC_0973	758	53
galE-F	GAGCAATGTCTATATTTAGTTTATAACCTA			50
galE-R	AAGCCATGTAGCTGAAATTATTAATA	MSC_0971	821	51
mgtA-F	TGCAAATCCCATAAAGATTAATG			51
mgtA-R	GTGGAAGTTTTGATAGTATTAAAGATAATC	MSC_0868	716	51
phnC-F	GACCACCAGAAAGTTCACTTACTT			51
phnC-R	ACACTGGATATAACAAATTTATGCC	MSC_0803	601	50

2.6 Southern blotting

Digested genomic DNAs from strains 95010-C1, Y-goat^T, PG1^T and 8740-Rita were separated on a 0.7 % agarose gel and migrated overnight at 18V in 1 X TEA buffer. DNA fragments on a gel were depurinated for 15 min (0.5 M HCl) and then denatured for 1 hr in 0.5 M NaOH /1.5 M NaCl. Before transfer to a membrane the gel was kept for 1hr in neutralization solution (Tris-HCl 0.5 M / 1.5 M NaCl). The above three steps depurination, denaturation and neutralization were conducted under slow agitation, and the gel was rinsed in between each step. The DNA from the gel was transferred to a highly positively charged Nylon Hybond N + membrane (Amersham Pharmacia Biotech). Passive transfer of denatured DNA fragments to the nylon membranes were performed using 20XSSC (Annex 3) solution over a period of 16-18 hrs. Following transfer, the membranes were placed DNA facing upwards on a 2XSSC soaked Whatman 3 mm paper and fixed for 3 min under UV light (Amersham UV cross-linker). Immediately after, the membranes were rinsed in double distilled water and kept at room temperature to dry. A pre-

hybridization was performed for 2 hr before the hybridization. Membranes were hybridized overnight at 50°C with 800 ng digoxigenin marked DNA probe in 1X Blocking reagent (Roche Diagnostics). After overnight hybridization the membranes were first washed in low stringency solution 2X SSC/ 0.1%SDS for 15 min at room temperature, followed by a high stringency solution 0.5X SSC / 0.1%SDS at 65°C for 15 min. Membranes were then blocked for 30 min in 1X blocking solution, and placed in a new 1X blocking reagent containing a conjugate anti-Digoxigenin-POD (Peroxidase), Fab fragments diluted to a 1/8000 for another 30 min (Roche Diagnostics). These steps were conducted under slow agitation. Then membranes were washed 3 times in 1X PBS solution each time for about 10 min. The non-radioactive ECL chemiluminescence detection solution was prepared by mixing two detection reagents immediately before use (Amersham, Life science). Detection reagent 1 contains H_2O_2, the substrate for peroxidise, and detection reagent 2 contains luminol, which produces a blue light upon oxidation. This reaction was revealed by Hyperfilm-ECL (Amersham, Life science) in dark room with a red light.

3 RESULTS AND DISCUSSION

3.1 Genome size estimation of strain 95010-C1

The genome size of MmmLC strain 95010-C1 was estimated using the PFGE technique. Rarely cutting enzymes *BamH*I, *Bgl*I, *Mlu*I, and *Nco*I were used to generate large fragments of the genome of strain 95010-C1, Y-goat^{T} (MmmLC), PG1^{T}, vaccine strain T1/44 and 8740-Rita (MmmSC strains), California kid^{T} (Mcc), L (*M. bovis*) and PG2^{T} (*M. agalactiae*).

Most of the DNA fragments generated by enzyme *BamH*I were shorter than 97 Kbp and were unsuitable for genome size analysis (data not shown). The digestion of DNA plugs using the enzymes *Mlu*I (Figure II. 1), *Bgl*I (Figure II. 2) and *Nco*I (Figure II. 3) gave different band patterns for all species analyzed and also for different strains of the same species: Y-goat^{T} and 95010-C1 (MmmLC); T1/44, PG1^{T} and 8740-Rita (MmmSC).

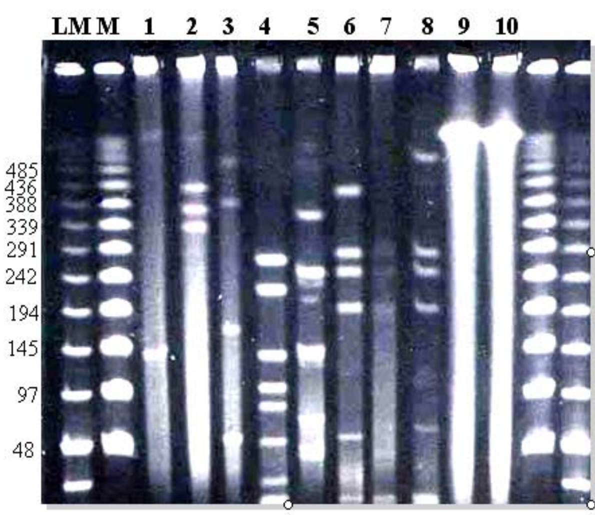


Figure II. 1 Pulsed-field profiles of *Mlu*I digested genomic DNA of *mycoplasma* strains.

Lanes LM and M: Lambda DNA molecular size marker in Kb; lane 1: California kid^{T} (*M. capricolum* subsp. *capricolum*); lanes 2 and 3: Y-goat^{T} and 95010-C1 (*M. mycoides* subsp. *mycoides* LC); lanes 4 and 5: PG2^{T} and L2 (*M. agalactiae* and *M. bovis* respectively) lanes 6-8: T1/44, 8740-Rita and PG1^{T} (*M. mycoides* subsp. *mycoides* SC); lanes 9 and 10: Mock Digestion of PG1^{T} genomic DNA and undigested PG1^{T} DNA.

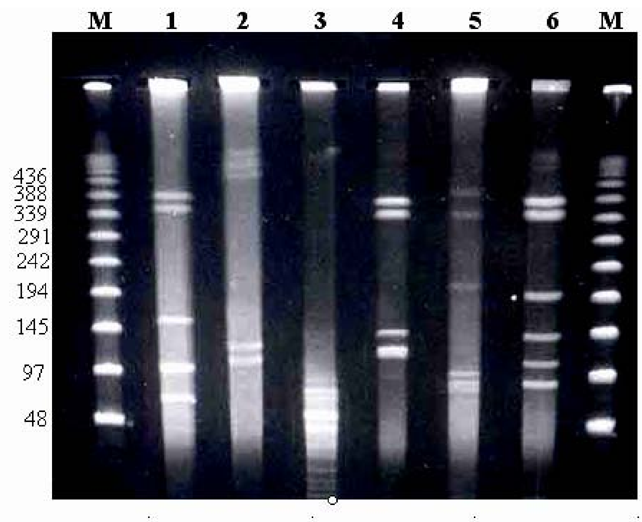


Figure II. 2 Pulsed-field profiles of *Bgl*I digested genomic DNA of *mycoplasma* strains.

Lanes M: Lambda DNA molecular size marker; lanes 1 and 2: Y-goat^T and 95010-C1 (*M. mycoides* subsp. *mycoides* LC); lane 3: L2 (*M. bovis*); lanes 4-6: T1/44, 8740-Rita and PG1^T (*M. mycoides* SC).

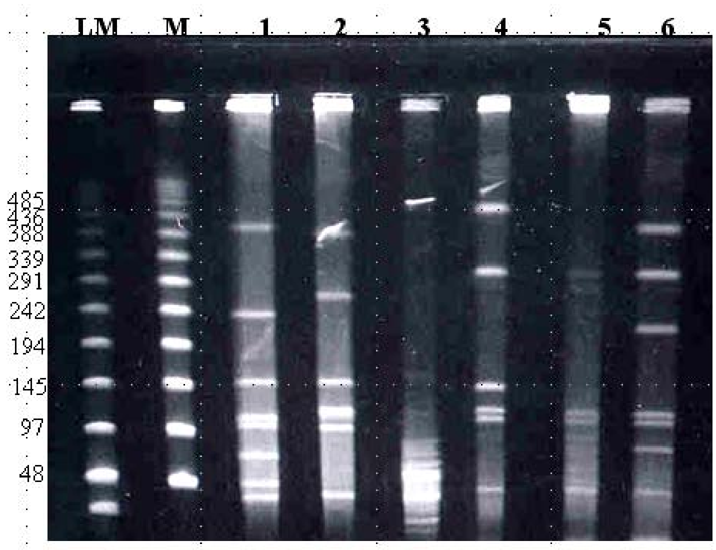


Figure II. 3 Pulsed-field profiles of *Nco*I digest of genomic DNA of *mycoplasma* strains.

Lanes LM and M: Lambda DNA molecular size markers; lanes 1 and 2: Y-goat^T and 95010-C1 (*M. mycoides* subsp. *mycoides* LC); lane 3: L2 (*M. bovis*); lanes 4-6: T1/44, 8740-Rita and PG1^T (*M. mycoides* SC).

3.1.1 Mycoplasma strains analyzed

3.1.1 1 Mcc, M. bovis *and* M. agalactiae

Strain California kid^T tested with the *Mlu*I (Figure II.1) revealed only two fragments. One bigger fragment of over 727.5 Kbp that was beyond the margin of the molecular weight marker (M) (Biolabs, England), which made fragment size estimation difficult, and one smaller fragment below 145 Kbp. *In-silico* digestion of the sequenced genome of strain California kid^T (NC_007633) using enzyme *Mlu*I generated two fragments of 874 and 136 Kbp that corresponded with the bands observed by PFGE. The profiles obtained with enzymes *Bgl*I (Figure II.2) and *Nco*I (Figure II.3) for strain L2 (*M. bovis*) were not satisfactory. Apparently, multiple restriction sites for these two enzymes are present in the *M. bovis* genome, resulting in smaller fragments and, consequently, unsuitable profiles for PFGE analysis. Strain PG2^T (*M. agalactiae*) was only analyzed with *Mlu*I (Figure II.1), providing a distinct band pattern.

3.1.1.2 MmmSC strains

*Mlu*I yielded six fragments for the PG1^T strain tested (Figure II.1). The number of fragments obtained corresponds with the *in-silico* digest of the published PG1^T genome. However, in our PFGE result the genome size was estimated as 1304 Kbp; 92 Kbp greater than the published PG1^T sequence (Table II.2).

*Bgl*I yielded six fragments similar to what was observed in *in-silico* digest of strain PG1^T (Figure II.2). The total genome size estimated from this enzyme was 1247 Kbp; 35 Kbp higher than the actual genome size (Table II.2).

*Nco*I yielded seven fragments in strain PG1^T (Figure II.3). Six restriction fragments were determined *in-silico* for this enzyme. The calculated band of 80 Kbp was double the size of the last band observed *in-silico* (Table II.2). In addition to this, the PFGE profile revealed the presence of a 30 Kbp DNA fragment that was absent in the *in-silico* digest of the PG1^T genome. The sum of the total fragment sizes generated by *Nco*I was 1251 Kbp; 39 Kbp higher than the sequenced PG1^T genome. Systematically, we have estimated a higher genome size for strain PG1^T by PFGE as compared to the published 1211 Kbp. The average genome size obtained with the three enzymes was calculated as 1267 Kbp. This may be explained by differences in the stock of PG1^T strains analysed (presumably due to multiple passages in different labs).

Table II. 2 Estimated DNA sizes of mycoplasmas strains.

Y-goat^T and 95010-C1 (*M. mycoides* subsp. *mycoides* LC); PG1^T, T1/44 (*M. mycoides* subsp. *mycoides* SC) obtained by PFGE following digestion with *Mlu*I, *Nco*I and *Bgl*I. *In-silico* profiles obtained from the sequenced genomes PG1^T and 95010-C1 are also displayed.

List of Fragment Size in Kbp obtained from three Enzymes							
No. Fragments	**Enzyme**	***In- silico*** **PG1^T**	**PG1^T**	**T1/44**	**Y-goat^T**	**95010-C1**	***In- silico*** **95010-C1**
1	*Mlu*I	430.957	514	420	424	509	511.012
2		272.863	280	280	363	390	385.921
3		245.026	242	242	338	168	153.062
4		191.381	194	194		50*	52.136
5		57.443	60	50			51.868
6		14.033	14	14			
Size from *Mlu*I	ΔPG1=92	1211.703	1304	1200	1125	1117	1153.999
1	*Bgl*I	364.71	380	380	380		491.573
2		332.898	343	343	345		445.782
3		191.246	194	140	157		115.913
4		135.789	138	118	97*		100.731
5		105.105	112	115	60		
6		81.955	80				
Size from *Bgl*I	ΔPG1=35	1211.703	1247	1096	1039		1153.999
1	*Nco*I	402.055	390	445	400	370	373.445
2		309.031	318	318	238	266	254.081
3		219.246	210	140	145	145	147.086
4		122.581	118	118	110	118	116.689
5		113.922	105	105	102	115	112.882
6		44.868	80	30	60	97	105.537
7			30		48	35	33.576
8					28	-	6.757
9						-	3.946
Size from *Nco*I	ΔPG1=39	1211.703	1251	1156	1131	1146	1153.999

* Represents double fragments

Based on data generated from all the three enzymes, the genome size of MmmSC strain T1/44 was estimated to be smaller than that of the PG1^T strain analysed. This difference was observed to be 104 Kbp when estimated from fragments generated by enzyme *Mlu*I, 151 Kbp from fragments generated by enzyme *Bgl*I, and 95 Kbp from fragments generated by enzyme *Nco*I. The average genome size estimation from fragments obtained by the three enzymes was calculated as 1151 Kbp (Table II. 2). This average size estimation however is 60 Kbp lower than the published PG1^T sequence.

Strain 8740-Rita gave a relatively distinct profile with *Bgl*I; the weak or non distinguishable DNA pattern obtained with enzymes *Nco*I and *Mlu*I resulted in difficulty to determine its approximate genome size.

3.1.1.3 MmmLC strains

*Mlu*I gave three fragments in strain Y-goat^T and four fragments in strain 95010-C1 (Figure II. 1). The total sum of the molecular weights generated by this enzyme gave genome size estimations of 1125 and 1117 Kbp for strains Y-goat and 95010-C1 respectively. The *in-silico* digest of the whole genome sequence of strain 95010-C1 using *Mlu*I generated five fragments corresponding to sizes 511 Kbp, 385 Kbp, 153 Kbp, 52 Kbp and 51 Kbp. In the PFGE experiment the corresponding sizes of fragments generated by this enzyme were 509 Kbp, 390 Kbp, 168 Kbp and 50 Kbp, with the latter fragment being more intense than the preceding three fragments. The intensity of the 50 Kbp fragment in PFGE may represent the overlapping fragments of 52 Kbp and 51 Kbp predicted *in-silico* (Table II. 2).

*Bgl*I resulted in partial digestion of the 95010-C1 genomic DNA (Figure II. 2). In order to avoid bias, fragments obtained with this enzyme were not considered for genome size estimation. However, five distinct fragments were obtained by digestion of Y-goat^T with this enzyme, providing a genome size estimation of 1039 Kbp, which was lower than that obtained with *Mlu*I. The higher intensity of the fifth fragment at 97 Kbp (Figure II. 2) may reflect the presence of an additional fragment in the Y-goat^T genome, which would then bring the genome size to 1136 Kbp. This value is closer to the size estimated with *Mlu*I (Table II. 2).

*Nco*I generated eight fragments for strain Y-goat^T and seven for 95010-C1 (Figure II. 3). For strain Y-goat^T the sum of the molecular weights of the generated fragments resulted in

an average genome size estimation of 1131 Kbp (Table II. 2). This estimation was close to that obtained with *Mlu*I. For strain 95010-C1, seven fragments were generated by PFGE of sizes 370 Kbp, 266 Kbp, 145 Kbp, 118 Kbp, 115 Kbp, 97 Kbp and 35 Kbp (Figure II. 3) resulting in a genome size estimation of 1146 Kbp. However, *in-silico* digestion of this same strain with *Nco*I predicted nine fragments of 373 Kbp, 254 Kbp, 147 Kbp, 116 Kbp, 112 Kbp, 105 Kbp, 33 Kbp, 7 Kbp and 4 Kbp. In our PFGE result we were not able to detect the latter two fragments, as the electrophoretic migration conditions were set to separate larger fragments. These smaller fragments must have migrated away in the electrophoresis buffer. Surprisingly, the other seven fragments provided similar sizes as compared to the *in-silico* predictions (Table II.2).

3.2 Assessment of the presence of repeated gene sequences

The Southern blot technique was used to evaluate the presence of duplicated genes in the genome of MmmLC strain 95010-C1. In type strain PG1^T, genes *lppQ*, *phnC*, *mgtA*, *galE* and *epsG* are found in multiple copies. Together with strain 95010-C1, type strain Y-goat^T of MmmLC and strain 8740-Rita of MmmSC were analyzed simultaneously.

3.2.1 Gene lppQ (ApaII digest)

The DNA probe prepared for gene *lppQ* did not hybridized with any band from the MmmLC genomes of strains Y-goat^T and 95010-C1 (data not shown). On the contrary, this probe hybridized with two bands from the PG1^T DNA and with a single band from MmmSC strain 8740-Rita. These results confirmed the previous findings of Bischof et al., (2006) that showed PG1^T to be the only MmmSC strain known to posses a duplicated 24 Kbp fragment containing the *lppQ* gene.

3.2.2 Gene mgtA (XbaI digest)

3.2.2.1 mgtA gene in strain PG1^T

The mgtA probe hybridized strongly with two PG1^T *Xba*I fragments with apparent molecular sizes of 13700 bp and 8000 bp (Figure II. 4). *In-silico* analysis of the PG1^T genome revealed that the observed signals corresponded to *Xba*I fragments containing locus tags MSC_0868 (13302 bp) and MSC_0881 (7827 bp). In the PG1^T genome a third locus (MSC_0907) has been annotated as *mgtA*, though this locus shares an amino acid identity of only 30% with the preceding two copies of *mgtA*.

3.2.2.2 mgtA gene in strain 8740-Rita

The same mgtA DNA probe hybridized to a single band in MmmSC strain 8740-Rita, but with a different band size of around 9500 bp. This result revealed that the *mgtA* gene is not duplicated in 8740-Rita.

3.2.2.3 mgtA gene in MmmLC strains

It was not possible to analyse *mgt*A in strain 95010-C1 using enzyme *Xba*I, because this enzyme did not cut this strain's DNA (Figure II. 4). However, BLAST analysis and annotation of the whole genome sequence of strain 95010-C1 allowed to identify locus tag MLC_7930 containing the *mgtA* gene. The mgtA DNA probe hybridized strongly with only one band of around 5600 bp in strain Y-goat^T, showing that *mgtA* is neither duplicated in this strain.

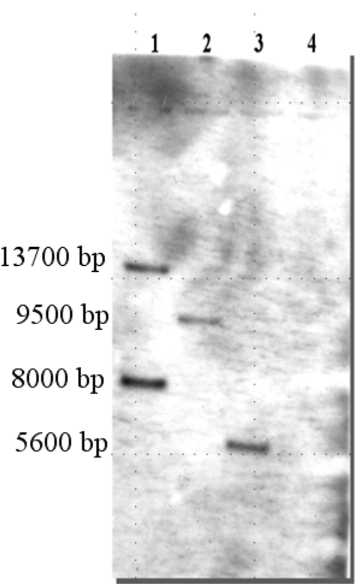


Figure II. 4 Southern hybridization of *Xba*I digested DNA using *mgtA* specific DNA probe.

Lanes 1 and 2: MmmSC strains PG1^T and 8740-Rita; lane 3 and 4: MmmLC strains Y-goat and 95010-C1. The absence of bands for strain 95010-C1 was attributed to the inability of *Xba*I to cut this strain's DNA.

3.2.2.3.1 XbaI enzyme activity in 95010-C1

In order to explain the mechanism behind the absence of *Xba*I enzymatic activity we assessed if *Xba*I belonged to the list of enzymes sensitive to Dam methylase activity. From this analysis we have found out that the restriction recognition site of *Xba*I may partially be overlapped by Dam Methylase Site 'GATC'. This enzyme has a cutting site (T↓CTAG↑A), which may be blocked by Dam methylation only if this recognition site is preceded by nucleotide sequences GA or followed by TC. *In-silico Xba*I restriction of the 95010-C1 genome revealed a number of sites not partially overlapped by Dam methylase. Therefore, Dam methylation cannot explain the complete absence of digestion of 95010-C1 by enzyme *Xba*I. There must be another mechanism at stake that is not present in strains PG1^T, Y-goat^T or 8740-Rita.

3.2.2.3.2 Dam methylase activity between MmmLC and MmmSC

To verify the presence of Dam methylase activity in the 95010-C1 genome we have chosen three enzymes having a 'GATC' recognition site: *Mbo*I, *Dpn*I and *Dpn*II. Enzymes *Mbo*I and *Dpn*II do not cut if adenine is methylated in the recognition site; on the contrary, *Dpn*I requires the adenine to be methylated for cleavage. MmmLC and MmmSC digests revealed that 95010-C1 and Y-goat^T could not digested by either *Mbo*I or *Dpn*II. However, these enzymes were able to digest the two MmmSC strains tested PG1^T and 8740-Rita (Figure II. 5). *Dpn*I, which recognizes a methylated site, was able to digest the genomes of both MmmLC strains but not those from MmmSC. In conclusion, there is a Dam methylase activity in the two MmmLC strains tested but not in MmmSC.

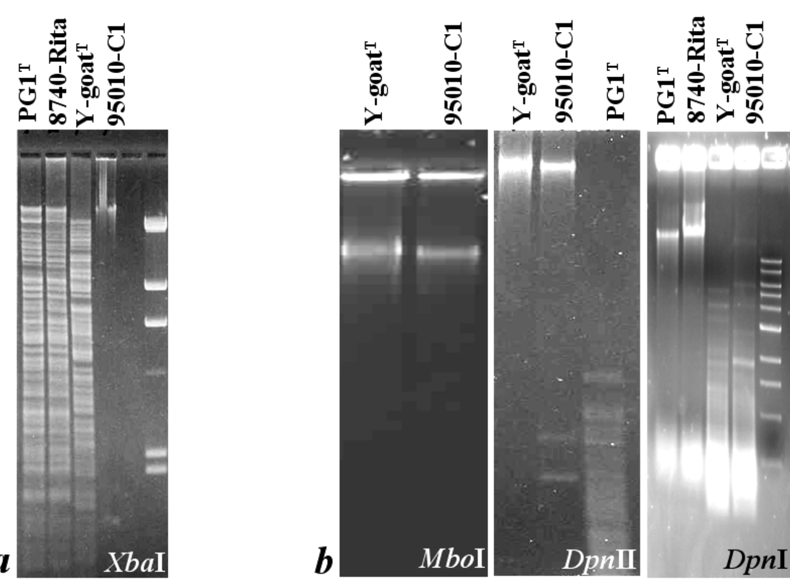


Figure II. 5 Restriction enzyme analysis of MmmLC and MmmSC strains.

a) *Xba*I did not digest 95010-C1 DNA b) Dam methylase activity of MmmLC and MmmSC genomes evidenced when using enzymes *Mbo*I, *Dpn*II and *Dpn*I. Lower visible fragments of *Dpn*II digestion from strain 95010-C1 are nick and supercoiled forms of its natural plasmid.

The 95010-C1 genome annotation has enabled the identification of seven genes that tentatively code for methylases. Identified in locus tags MLC_5410 (Tetrapyrrole methylase family protein), MLC_4870 (rRNA methylase), MLC_1980 (DNA methylase), MLC_1620 (DNA modification methylase), MLC_8270 (tRNA/rRNA methyltransferase), MLC_7790 (Adenine-specific DNA methylase) and MLC_7290 (Putative C5 methylase). BLAST analyses were made for the above four underlined putative DNA methylases. CDSs from loci MLC_1980 and MLC_7290 have revealed strong similarity of 83 % and 97 % of with their homolog in MmmSC, MSC_0219 and MSC_0780 respectively. Locus containing the Adenine-specific DNA methylase MLC_7790 was neither identified in MmmSC strain PG1^T nor in other sequenced mycoplasma genomes. However, this gene existed with 100 % homology in the recently sequenced MmmLC strain GM12b. The closest relative of this gene in other bacterial species has been found in *Haemophilus haemolyticus*, with a CDS similarity of 63.4 %. The fourth locus tag we analysed was MLC_1620, containing a CDS for DNA modification methylase. We have found its homolog MSC_0186 in MmmSC strain PG1^T, possessing a longer sequence (i.e.: 662

amino acids). Amino acid sequence alignment revealed only the N-terminal 390 aa sequence being similar to MLC_1620. From other bacterial species, a similarity of 57 % was observed in a 396 aa sequence CDS of *Streptococcus pneumoniae*, encoding an *Xba*I methylase.

3.2.3 Gene phnC (EcoRI digest)

3.2.3 1 phnC gene in strain PG1^T

The phnC DNA-probe hybridized strongly with two PG1^T *Eco*RI fragments with apparent molecular sizes of 14200 bp and 5000 bp. A very weak signal appeared with a lower band of 3400 bp (Figure II. 6). Through *in-silico* analysis, the strongest signal corresponded to fragments containing MSC_0789 (14277 bp) and MSC_0803 (5102 bp). The weaker signal may correspond to fragment containing MSC_0078 that has a similar annotation (*phnC*) but which had a similarity of only 59 % with the other two *phnC* genes.

3.2.3 2 phnC gene in strain 8740-Rita

The phnC DNA probe also hybridized with two *Eco*RI bands of strain 8740-Rita DNA, though of different sizes. This shows that *phnC* is also duplicated in the genome of this strain but mutation may have altered the digestion pattern.

3.2.3 3 phnC gene in MmmLC strains

The phnC probe hybridized strongly with only one band at around 4900 bp showing that *phn*C is not duplicated in strain 95010-C1. BLAST analysis and annotation allowed the identification of MLC_7320 as the *phn*C ortholog. By comparison three bands of weaker intensities were observed with strain Y-goat^T, which may suggest either lower sequence similarity or presence of a cutting site within the *phnC* ortholog in Y-goat^T (Figure II. 6).

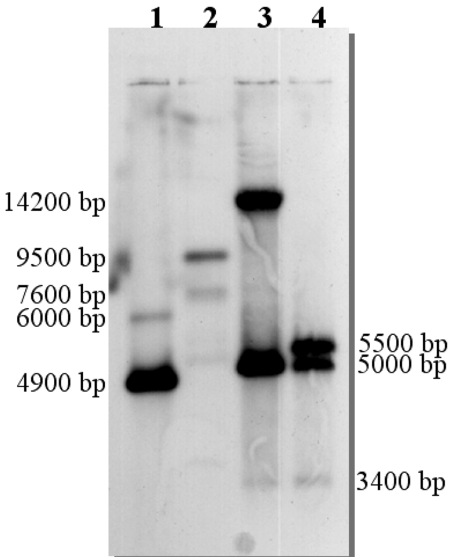


Figure II. 6 Southern hybridization of *EcoR*I digested DNA using *phnC* specific DNA probe. Lanes 1 and 2: MmmLC strains 95010-C1 and Y-goat^T; lane 3 and 4: MmmSC strains PG^T and 8740-Rita. This gene appears to exist in two copies in strain 95010-C1, three or 4 copies in Y-goat and three copies in both MmmSC strains.

3.2.4 Gene galE (EcoR*V digest*)

3.2.4.1 galE gene in strain PG1^T

The galE DNA-probe hybridized strongly with three PG1^T *EcoR*V fragments with apparent molecular sizes of 13280 bp, 9600 bp and 8200 bp (Figure II. 7). Through *in-silico* analysis the three strong signals corresponded to *EcoR*V fragments containing MSC_0985 (13307 bp), MSC_0978 (9669 bp) and MSC_0971 (7732 bp).

3.2.4.2 galE gene in strain 8740-Rita

By comparison with 8740-Rita, the same probe hybridized with three bands of the same size. This shows that gene *galE* is also triplicate in this strain (Figure II. 7).

3.2.4.3 galE gene in MmmLC strains

The galE DNA-probe hybridized strongly with only one band at around 3000 bp showing that *galE* is present in a single copy in strain 95010-C1. BLAST analysis and annotation allowed the identification of MLC_8190 as the *galE* ortholog. By comparison, one band of

high intensity at around 3000 bp and a second band of weaker intensity at around 3500 bp were observed with strain Y-goat^T. The weak intensity of the higher band in Y-goat^T could represent lower sequence similarity of the second copy *galE* ortholog (Figure II. 7).

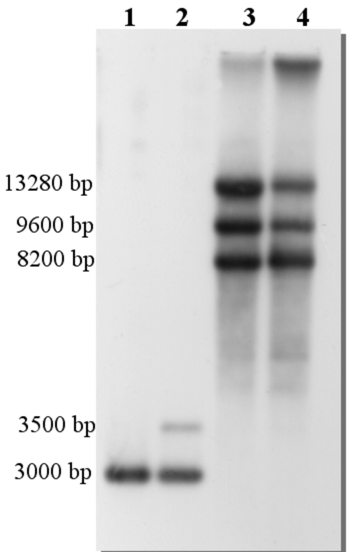


Figure II. 7 Southern hybridization of *EcoR*V digested DNA using *galE* specific DNA probe. Lanes 1 and 2: MmmLC strains 95010-C1 and Y-goat ^{T}; lanes 3 and 4: MmmSC strains PG1^T and Rita, this gene seems to exist in one copy in strain 95010-C1, two copies in Y-goat ^{T} and in three copies in both MmmSC strains.

3.2.5 Gene epsG (EcoRV digest)

3.2.5.1 epsG gene in strain PG1^T

The epsG DNA probe hybridized strongly with four PG1^T *EcoR*V fragments with apparent molecular sizes of 16000 bp, 13500 bp, 9800 bp and 7800 bp. Through *in-silico* analysis the strong signals from the four bands corresponded to fragments containing MSC_0108 (15984 bp), MSC_0987 and MSC_0993 (13307 bp), MSC_0980 (9669 bp) and a fourth fragment of 7732 bp corresponding to MSC_0973. The fragment at 13307 bp was found to contain two *eps*G copies, revealing the presence of five copies of this gene in the PG1^T genome (Figure II. 8).

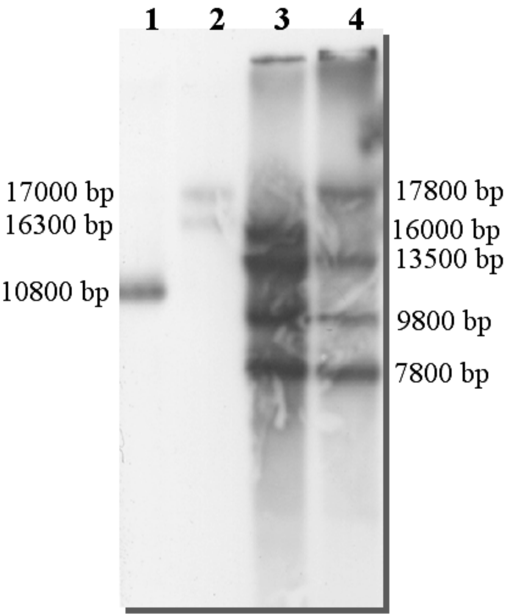


Figure II. 8 Southern hybridization of *EcoR*V digested DNA using *epsG* specific DNA probe. Lanes 1 and 2: MmmLC strains 95010-C1 and Y-goat^T; lanes 3 and 4: MmmSC strains PG1^T and Rita. This gene seems to exist in one copy in strain 95010-C1, two copies in strain Y-goat^T and both MmmSC strains revealed four hybridizing bands.

3.2.5.2 epsG gene in 8740-Rita

By comparison, the same DNA probe hybridized with four bands in strain 8740-Rita. Three bands observed at the same size as those from PG1^T: 13500 bp, 9800 bp and 7800 bp. The first band however was located at around 17000 bp, this could refer mutation in the restriction site of the genome of strain 8740-Rita resulted in alerted digestion pattern.

3.2.5.3 epsG gene in MmmLC strains

The epsG DNA probe hybridized strongly with only a single band at 10800 bp in strain 95010-C1, showing presence of a single *epsG* gene. BLAST analysis and annotation allowed identification of two locus MLC_8180 and MLC_1020 as the *eps*G orthologs. Locus MLC_1020 was localized in an *EcoR*V fragment of 10636 bp and locus MLC_8180 was found in a 1062 bp *EcoR*V fragment. The 5'end of MLC_8180 was also found to be truncated by restriction enzyme *EcoR*V. This may explain the absence of hybridization of the epsG DNA probe with the gene copy in locus MLC_8180. By comparison, two bands

of weak intensities were observed with strain Y-goat^T, corresponding to 17000 and 16300 bp. The weaker intensity in this strain could suggest lower sequence similarity (Figure II. 8).

4 CONCLUSIONS AND DISCUSSION

The finishing phase of the whole genome sequence of strain 95010-C1 provided a total genome size of 1154 Kbp and a natural plasmid of 1800 bp. The genome size has corresponded to that obtained by PFGE genome size estimation using enzyme *Nco*I (i.e. 1146 Kbp with a variation of 0.69 %). The estimation obtained by PFGE was used in the final assembly and validation of the whole genome sequence of strain 95010-C1. Although the PFGE profiles obtained with *Nco*I and *Mlu*I have not provided precise genome size estimation, both enzymes provided interesting guidelines for the finishing phase. The presence of two integrative and conjugative elements (ICEs) of sizes 28 Kbp and 30 Kbp brought difficulty to whole genome sequence assembly. *In-silico* genome analysis of 95010-C1 has revealed the two ICEs to be located within 254 and 105 Kbp *Nco*I fragments. The corresponding *Nco*I fragments from the PFGE profiles were 266 Kbp, which contained the 28 Kbp ICE, and 97 Kbp, containing the 30 Kbp ICE.

The drawbacks we came across during genome size estimation of 95010-C1 using PFGE profiles were, first, when fragments of equal molecular weights were generated that could not be distinguished (i.e.: *Mlu*I DNA fragments of 52 and 51 Kbp). When the high intensity of the 50 Kbp fragment generated by enzyme *Mlu*I was considered as overlapping of two fragments of approximate sizes, then the genome size estimation (i.e. 1167 Kbp) was closer to the *in-silico* value with a difference of 13 Kbp. The other drawback encountered was the failure to detect smaller bands. For instance, the 7 and 3 Kbp fragments generated by the enzyme *Nco*I were not detected on the gel.

The analysis made by Pyle and colleagues (1990) after restriction digestion by enzymes *Bgl*I, *Xho*I, *Sal*I and *Sma*I; and Pyle and Finch (1988) restriction digestion by enzymes *Apal*I, *BamH*I, *Bgl*I, *BssH*I, *Sal*I, *Sma*I, *Xho*I demonstrated the average genome size for PG1^T to be 1280 Kbp. This estimation is very close to the average genome size estimation obtained in our study from the three enzymes (i.e. 1267 Kbp). The estimation made by Pyle and colleagues and that presented in our study were respectively 68 and 55 Kbp greater than the published PG1^T genome size data (Westberg et al., 2004). One reason that was previously mentioned may be the difference in the stock of strain PG1^T used in these analyses. On the other hand differences may also be due to possible errors during fragment size estimation. Precision in measuring the exact size of each DNA fragments had been problematic, especially when DNA fragments lied at an intermediate position between two

molecular weight markers. To overcome this bias in size estimation, advanced softwares for precise molecular weight measuring are required.

Different attempts to prepare MmmSC strain 8740-Rita's DNA failed to give clear-cut PFGE profile in our study. Hollis and colleagues compared PFGE and ribotyping of bacterial isolates. After reiterated attempts to prepare DNA they concluded that some strains were non-typeable with the PFGE technique (Hollis et al., 1999). In some cases the number of non-typeable PFGE strains can be significant, masking the discrimination ability of this technique. This phenomenon was noticed when (Marois et al., 2001) found non-typeable *M. synoviae* strains. Non typeable *M. bovis* strains have also been observed by McAuliffe et al. (2004). Contrarily to strain 8740-Rita, T1/44 gave a clear-cut profile upon PFGE, which helped in estimating the genome size of this strain. In this study, the average genome size difference between T1/44 and sequenced PG1^T genome was found to around 60 Kbp. Bischof et al. (2006) analyzed the different repeated zones observed in the PG1^T genome and found absence of the 24 Kbp DNA fragment duplication in strain T1/44. The variation obtained from the present implies that the differences between strains T1/44 and PG1^T must relate to additional DNA sequences, besides the 24 Kbp DNA fragment.

PG1^T has a larger genome size both by *in-silico* and PFGE analysis as compared to the two MmmLC strains analyzed. An explanation for the greater genome size of this strain may be the presence of large numbers of IS elements as well as repeated zones containing multiple gene copies. MmmLC strain Y-goat^T was estimated to have an average genome size of 1200 Kbp by Pyle and Finch (1990). In our study the same number of fragments were obtained by digestion with the commonly used enzyme *Bgl*I, although size estimations in the present study using the three enzymes *Bgl*I, *Mlu*I and *Nco*I were lower than the results obtained by Pyle and Finch. In a recent study of the prevalence of MmmLC isolates from the ear canal of healthy goats, *Nco*I was found the most suitable enzyme generating interpretable PFGE profiles (Tardy et al., 2007). The six to nine fragments generated by this enzyme corresponded to the fragments generated from our MmmLC strains tested. The PFGE profiles obtained for these two MmmLC strains using all the three enzymes permitted differentiation of the two strains. However, the majority of MmmLC strains tested by Tardy and colleagues (2007) gave similar, if not identical PFGE patterns with the exception of three isolates. Variation in PFGE pattern may be associated with recombination events including enzyme recognition site, point mutations occurring within

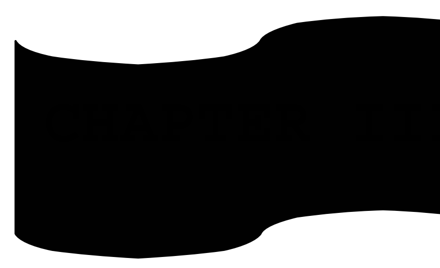
the cleavage sites or, as in the case of strain 95010-C1, could be the presence of ICE elements. The *Nco*I fragment containing the 28 Kbp ICE element in strain 95010-C1 (i.e. 266 Kbp) corresponded to a 238 Kbp fragment obtained in strain Y-goat^T. Similarly, the 97 Kbp DNA fragment containing the 30 Kbp ICE element corresponded to a 60 Kbp DNA fragment in strain Y-goat^T. These findings imply absence of the ICE elements in the corresponding fragments in strain Y-goat^T.

Southern hybridization studies conducted in order to assess the presence of repeated DNA sequences in the 95010-C1 genome revealed clear differences between MmmSC and MmmLC biotypes. Multiple copies of the genes *lppQ*, *epsG*, *galE*, *phnC* and *mgtA* were observed in MmmSC strain PG1^T and 8740-Rita, while these genes were often found in a single copy in MmmLC strains Y-goat^T and 95010-C1. These studies also confirmed the results obtained by Bischof et al. (2006) showing that the 24 Kbp duplicated DNA fragment was characteristic of MmmSC strain PG1^T. As for the 28 MmmSC strains studied by Bischof and colleagues, strain 8740-Rita possessed a single copy of the 24 Kbp and 13 Kbp repeated zones, a duplication of the 12 Kbp fragment and a triplicate of the 8 Kbp zone.

Variation in the activity of Dam-methylation was observed between MmmLC and MmmSC in our study. This was in agreement with previous findings by Bergemann and colleagues. These researchers showed that MmmSC, Mcc, Mbg7 and some strains of Mmc did not exhibit adenine methylation activity. However, all the MmmLC and Mccp strains tested showed adenine methylation at the ‘GATC’ cleavage site. MmmLC strain Y-goat^T DNA was found to be resistant to restriction digestion by *Mbo*I and *Dpn*II, enzymes that are inhibited by methylation of adenine at their cleavage site ‘GATC’. However, this strain was susceptible to digestion by *Dpn*I (Bergemann et al., 1990). Restriction endonucleases *Dpn*I and *Dpn*II are produced by two distinct strains of *Diplococcus pneumoniae*. The two enzymes share the same recognition site and show complementary specificity with respect to adenine methylation. *Dpn*II cleaves at the unmodified sequence 5'-GATC-3' whereas *Dpn*I requires the site to be methylated (Lacks and Greenberg, 1977). In this study, Dam methylation activity was also demonstrated in strain 95010-C1. Locus MLC_7790 containing adenine specific DNA methylase enzyme may play a role in the activity of Dam methylation in MmmLC strains tested.

The absence of restriction digestion by *Xba*I in this strain, however, cannot be explained by Dam methylation. On the other hand, BLAST analysis revealed that CDS MLC_1620, encoding a DNA modification methylase, exhibit some similarity with an *Xba*I methylase of *Streptococcus pneumoniae*. This methylase may be responsible for the modification that inhibited *Xba*I enzyme activity in MmmLC strain 95010-C.

The isolation of the naturally occurring plasmid from MmmLC strain 95010-C1, 1800 bp corresponds to those previously obtained cryptic plasmids from MmmLC strain GM12 with a size range of 1.7-1.9 Kbp (Bergemann et al., 1989; King and Dybvig, 1992). Similar to the previously discovered plasmids this new plasmid will be of future relevant for the development of alternative shuttle vectors.



Application of whole genome sequence for the development of specific diagnostic tools

1 INTRODUCTION

1.1 Identification of members of the *M. mycoides* cluster

The six members of the *M. mycoides* cluster are known to produce various disease of economic importance in cattle sheep and goats. Differential diagnostic tests are often difficult to develop due to shared common antigens (Howard and Taylor, 1985). When using conventional serological methods for mycoplasma identification, cross reactive antibodies often hamper the identification of pathogenic agents of the *M. mycoides* cluster thereby complicating interpretation (Cottew and Yeats, 1978; Rurangirwa et al., 2000). This phenomenon was observed between Mbg7 and Mccp (Thiaucourt, 2002) and between Mbg7 and Mcc (Bolske, 1988). As a result rapid identification of causative agent of CBPP and CCPP, MmmSC and Mccp have been achieved only through monoclonal antibodies and PCR (Bascunana et al., 1994; Bashiruddin et al., 1994; Brocchi et al., 1993; Thiaucourt et al., 1994). However, some reports on variable surface proteins of mycoplasmas indicate that ambiguous results may occur when monospecific antibodies are used in diagnostics based on antigen detection (Rosengarten et al., 1994; Rosengarten and Yogev, 1996). Recently identification of mycoplasmas belonging to the *M. mycoides* cluster was found to be hampered by the occurrence of field strains, antigenically intermediate between some serotypes and species. This has been demonstrated by atypical cross reaction schemes upon membrane filtration dot immunoblot tests using polyclonal anti-sera from members of the *M. mycoides* cluster (Le Grand et al., 2004). DNA probes have been used to produce more specific diagnostic procedures (Dedieu et al., 1992; Taylor et al., 1992a, b) problems with these tests could be non-specific binding of probes to non-target sites with poor reproducibility and difficult interpretation.

In recent years PCR has replaced traditional diagnostic tests for the identification of members of the *M. mycoides* cluster. DNA amplification techniques offer a promising identification system by avoiding variability that hinder serological methods (Le Grand et al., 2004). So far a number of PCR tests have been developed for the rapid identification of species of the *M. mycoides* cluster. Most of the PCR systems developed until recently and before the flourishing of genomic sequence of mycoplasmas were based on CAP-21 sequence fragment encoding notably for ribosomal proteins *rpsL* and *rpsG*. This fragment has been used for the design of specific primers for MmmSC (Bashiruddin et al., 1994), *M. mycoides* cluster and *M. putrefaciens* (Hotzel et al., 1996; Rodriguez et al., 1997). The other most widely used target for specific PCR tests is the 16S rRNA gene, this gene has

been used for specific PCR for the identification of Mccp (Bascunana et al., 1994; Bolske et al., 1996) and MmmSC (Persson et al., 1999). In addition to these two a gene that codes for lipoprotein, p72 gene has been used for the design of specific primers for MmmSC (Miserez et al., 1997) and Mbg7 (Frey et al., 1998) detection. Finally gene *lppA* has been used for the detection of Mccp, Mcc, MmmLC and Mmc (Monnerat et al., 1999a; Monnerat et al., 1999b). A non coding intergenic sequence between MSC_0390 and MSC_0391 from the genome sequence of MmmSC strain PG1^T has also been employed in the design of specific PCR primers for MmmSC (Dedieu et al., 1994). A recent evaluation of the above mentioned PCR tests have revealed all PCR tests based on lipoprotein genes were not strictly specific (Le Grand et al., 2004). With respect to MmmSC identification, PCR tests by Dedieu et al. (1994) and Bashiruddin at al. (1994) remain reliable tests as these two have been used extensively in our laboratory for the diagnosis of CBPP.

Many mycoplasma species take up arginine from the host and degrade it to produce ATP (Schimke and Barile, 1963). The non fermenting mycoplasmas and a small group of fermenting mycoplasma, hydrolyse arginine producing ammonia with an increased pH. This change in the pH can be detected by a pH indicator (Barile, 1983). In mycoplasma besides fermentation of sugars to lactate and oxidation of lactate or pyruvate to acetate plus carbon dioxide; the metabolism of L-arginine to L-ornithine is one of the recognized mechanisms of energy generation.

Among species of the *M. mycoides* cluster, Mcc has a definite but delayed ability to hydrolyse arginine. An enzyme activity test revealed this species to posses the enzyme ornithine transcarbamylase (DaMassa et al., 1992). The closely related species of the *M. mycoides* cluster, *M. putrefacien*s posses a variable activity towards arginine hydrolysis test. Mycoplasma species closely related to *M. putrefaciens*, *M. yeatsii* and *M. auris* were found to hydrolyse arginine (DaMassa et al., 1994).

1.2 The Arginine deiminase pathway

Multiple pathways for arginine degradation have been described in micro-organisms and, occasionally several of them are simultaneously present in the same organism (Zuniga et al., 2002). Among these pathways the arginine deiminase pathway (ADI) pathway is the most widespread anaerobic route by which arginine is degraded to produce energy (i.e. one mol of ATP per mol of arginine consumed). This pathway was first named in 1940 by

Hills, who described the conversion of arginine to citrulline in *Streptococcus* (Schimke et al., 1966). It was found that arginine did constitute a major energy source for mycoplasmas isolated from the contaminated cell cultures, because of their finding of ornithine as a major metabolic product in growing cells. Mycoplasmas from contaminated cell cultures demonstrated a high activity of all three enzymes of the arginine deiminase pathway (Barile et al., 1966). Three enzymes are involved in this pathway. Enzyme arginine deiminase (ADI, EC 3.5.3.6) encoded by gene *arcA* catalyses the first step of the pathway, the deimination of arginine yielding citrulline and ammonia (Figure III. 1). The second enzyme (catabolic) ornithine transcarbamylase (cOTCase, EC 2.1.3.3) encoded by gene *arcB* is closely related to anabolic OTCase. Anabolic OTCase is involved in arginine biosynthesis and catalyses the reaction opposite to that catalysed by the cOTCase, namely the formation of citrulline from ornithine and carbamoyl phosphate. The third enzyme, carbamate kinase (CK, EC 2.7.2.2) encoded by gene *arcC* catalyses the hydrolysis of carbamoyl phosphate to CO_2 and NH_3. The phosphate group of carbamoylphosphate is used to phosphorylate one ADP to ATP (Spano et al., 2004). Gene *arcD* encodes a membrane bound protein that is necessary for the uptake of arginine and the excretion of ornithine (D'Hooghe et al., 1997). Apart from these genes encoding the three catalytic activities additional genes are involved associated with ADI gene cluster. These genes are involved in transcription regulation (Zuniga et al., 2002, Figure III.1).

Regulation of the arginine deiminase pathway has been extensively studied in *Pseudomonas aeroginosa*, where the pathway is encoded by four genes organized in an operon: *arcDABC* and expression of the *arc* genes is up regulated by the Anr protein (Gamper et al., 1991). In some other bacteria such as *Bacillus lichenformis* this operon is organized as *arcABDC* followed by *argR*; in the presence of arginine *argR* is both a repressor of the anabolic ornithine carbamoyltransferase and an activator of the arginine deiminase pathway (Maghnouj et al., 1998). In *M. pneumoniae* these genes are organized as *arcABC*, with no such detected regulatory genes (Zuniga et al., 2002).

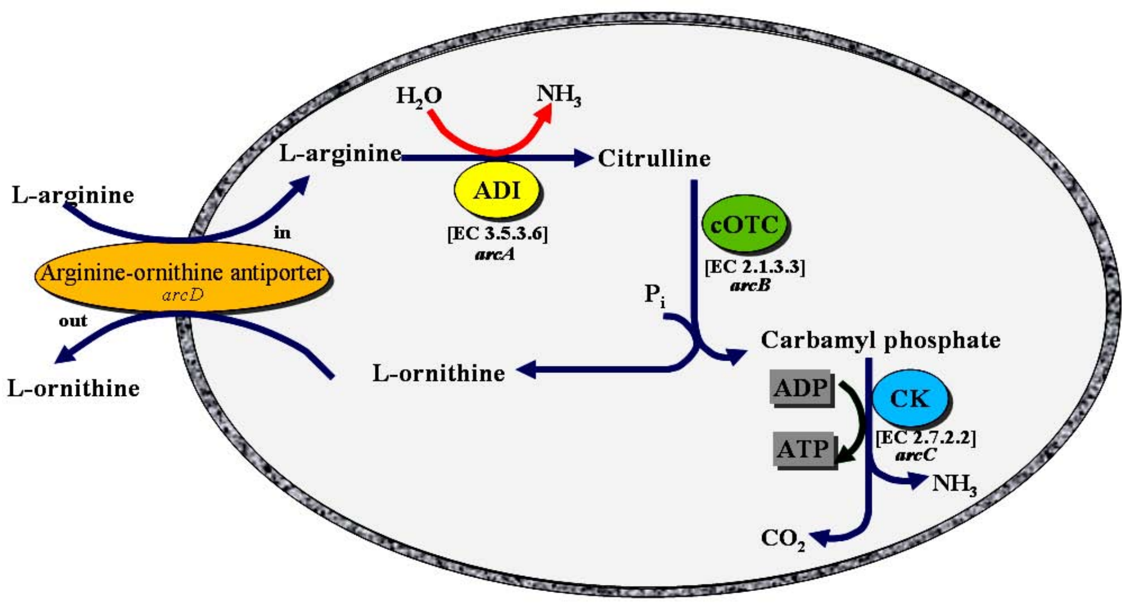


Figure III. 1 Arginine Deiminase Pathway.

The three important enzymes Arginine deiminase (ADI), ornithine transcarbamylase (OTC) and carbamate kinase (CK) are encoded by genes *arcA*, *arcB* and *arcC* respectively. The third important membrane protein arginine ornithine anti-porter is encoded by *arcD*. This pathway involves degradation of arginine as a means of energy generation. Adapted from Spano et al. (2004).

1.3 The "ADI operon" within the *M. mycoides* cluster

At the beginning of this study we had identified arginine deiminase operon-like sequence from some of the representative strains of species of the *M. mycoides* cluster. Flanking genes were then obtained from the genome sequences of type strain Mcc, California kid^T (GenBank/ EMBL Accession No. NC_007633) and MmmSC, PG1^T (Westberg et al., 2004). The identification of the flanking genes has allowed amplifying the ADI operon-like sequence from the rest of the representative strains. It was observed that in almost all the strains the ADI operon-like sequence was flanked by genes *mgtA* (magnesium transport ATpase) at 5'end and *glk* (glucokinase) at the 3' end. In the five type strains studied the arginine deiminase like operon was organized in specific fashion between *M. mycoides* and *M. capricolum* subspecies. It was revealed to be organized as *arcBDC* in *M. mycoides* and *arcABD* in *M. capricolum* subspecies excluding type strain PG50^T of Mbg7 (Figure III. 2).

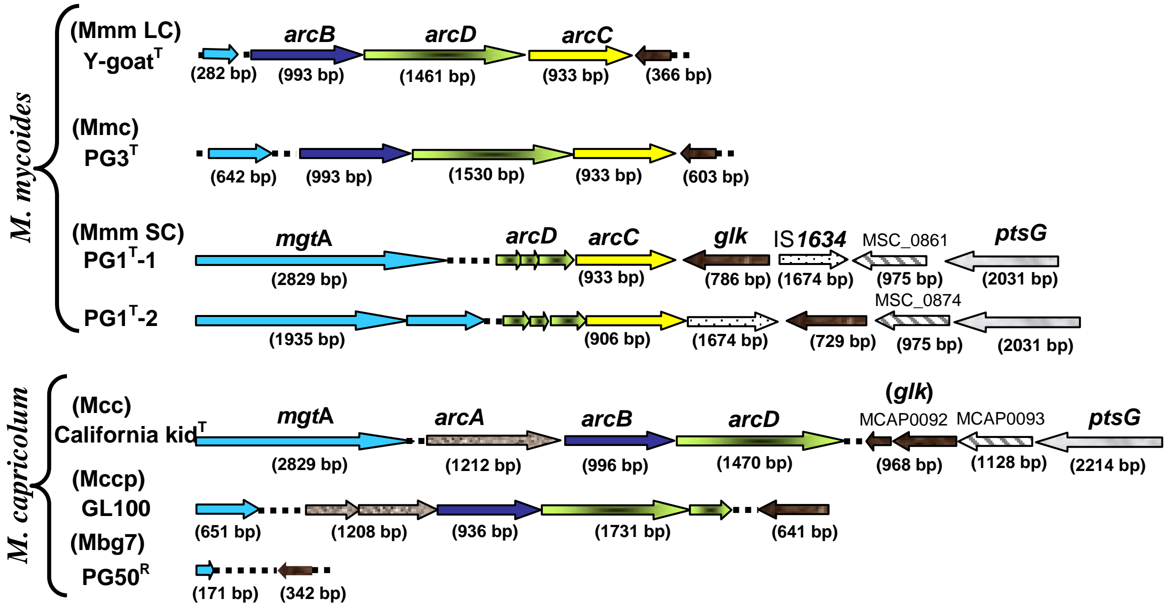


Figure III. 2 Organization of the arginine deiminase "ADI-operon" in the representative strains of the *M. mycoides* cluster.

The operon was found to be organized as *arcBDC* in the *M. mycoides* subspecies and as *arcABD* in the *M. capricolum* subspecies. In all the type strains this operon is flanked by genes *mgtA* and *glk*, excluding the type strain PG50 of Mbg7.

At first, we considered this variation as a good basis for the design of specific primers that could differentiate the *M. mycoides* cluster into the *M. mycoides* subspecies and the *M. capricolum* subspecies. This first attempt was designing primers from *arc*A and *arc*D for all *capricolum* subspecies and from *arc*B and *arc*C for all *mycoides* subspecies. However the whole genome sequence data from MmmLC strain 95010-C1 has revealed existence of intra species variation on the organization of this operon. Unlike Y-goat^{T}, in strain 95010-C1 *arcB* is not followed by *arcDC*. In this species *arcB* is the only gene flanked by genes *mgtA* and *glk*. This observation together with the absence of the ADI genes from sequenced locus of reference strain PG50^{R} did not allowed us to proceed with this approach (i.e. designing specific diagnostic tools categorizing the two subspecies of the *M. mycoides* cluster). Therefore our initial attempt was shifted to the use of *glk* gene, one of the genes flanking the ADI operon within the *M. mycoides* cluster as a target gene for the design of specific PCR.

In the present study the genes of “ADI operon” of the *M. mycoides* cluster were used for the design of specific PCR system used for the detection and identification of members of the *M. mycoides* cluster, and to *M. putrefaciens* species.

Paper I describes the choice of *glk* gene, flanking gene of the “ADI operon” in the design of specific PCR for the detection of members of the *M. mycoides* cluster and its application in the diagnosis of the contagious agalactia syndrome.

Paper II describes the use of *arcB* gene, coding for the enzyme ornithine transcarbamoylase transferase (OTC), for the design of specific primers used for the identification of *M. putrefaciens*, a species closely related to the *M. mycoides* cluster and a causative agent of the contagious agalactia syndrome.

Paper III describes specific PCR for the identification of Mccp, the causative agent of the economically most important disease CCPP. The arginine-ornithine antiporter encoded by the gene *arcD* was used for the design of specific primer pair for the identification of Mccp.

Paper I

Molecular and Cellular Probes
Volume 21, Issues 5-6, October-December 2007, Pages 391-399
Copyright © 2007 Elsevier Ltd All rights reserved.

A PCR for the detection of mycoplasmas belonging to the *Mycoplasma mycoides* cluster: Application to the diagnosis of contagious agalactia

S. Woubit[a], L. Manso-Silván[a], S. Lorenzon[a], P. Gaurivaud[b], F. Poumarat[b], M. P. Pellet[c], V.P. Singh[d] and F. Thiaucourt[a,]

[a]CIRAD UPR 15 Contrôle des maladies animales exotiques et émergentes, TA A-15/G Campus International de Baillarguet, 34398 Montpellier Cedex 5, France
[b]UMR Mycoplasmoses des Ruminants, AFSSA Site de Lyon, 31 avenue Tony Garnier, 6936 4 Lyon Cedex 07, France
[c]AFSSA Niort 60 rue de Pied-de-Fond BP3081, 79012 Niort Cedex, France
[d]IVRI, National Referral Lab on Mycoplasma, Indian Veterinary Research Institute, Izatnagar 243122, Uttar Pradesh, India
Received 25 October 2006; accepted 18 May 2007. Available online 7 June 2007.

Abstract

Contagious agalactia is a mycoplasmal infection caused by *Mycoplasma agalactiae*, *Mycoplasma mycoides* subsp. *mycoides* LC, *M. mycoides* subsp. *capri*, *Mycoplasma capricolum* subsp. *capricolum* and *Mycoplasma putrefaciens*. Identification of the causative organisms is usually performed by isolation and classical biochemical and serological tests, though this is a lengthy and cumbersome process for mycoplasmas. Specific PCR assays have been developed for the identification of *Mycoplasma agalactiae* and *M. putrefaciens*. For members of the *M. mycoides* cluster existing PCR tests are based on the amplification of highly conserved genes coding for ribosomal proteins, hence a possibility of cross-reactions. The gene *glk*, coding for a glucokinase, that is found in this cluster is very distantly related to any other bacterial glucokinase described so far. It was therefore chosen as target to design a new PCR test. The validation was performed independently in three laboratories in France and India using over 100 mycoplasma strains of various geographical origins. All strains belonging to the *M. mycoides* cluster were detected by amplification of the expected PCR product (428 bp) while no amplification was obtained from *M. agalactiae* strains. Our results demonstrate the universality of this

PCR in spite of the great heterogeneity found within this cluster. This new tool may be of great help for the implementation of control measures directed towards contagious agalactia.

Keywords: Contagious agalactia; *Mycoplasma mycoides* cluster; Glucokinase; PCR

1. Introduction

Contagious agalactia is a disease of sheep and goats characterized by prominent lesions of mastitis, arthritis, keratitis, pneumonia and septicaemia (MAKePS). Because of its economic importance, this disease is listed as a notifiable disease by the OIE (World Organization for Animal Health). Losses are mainly caused by mastitis in lactating ewes, though mortality may also occur, particularly in goats affected by arthritis and pneumonia [1]. The disease has a worldwide distribution. Originally, it was associated with *M. agalactiae*, the main causative agent in sheep that is particularly frequent in the Mediterranean basin. However, the members of the *M. mycoides* cluster *M. mycoides* subsp. *mycoides* large colony biotype (Mmm LC), *M. mycoides* subsp. *capri* (Mmc) and *M. capricolum* subsp. *capricolum* (Mcc) are repeatedly mentioned as causative agents of contagious agalactia in goats [2], [3], [4] and [5]. *M. putrefaciens* is also considered a potential etiologic agent, although ocular lesion have not been described so far [5] and [6].

The *M. mycoides* cluster comprises six pathogens causing mild to severe disease in ruminants: Mmm LC, Mmc, Mcc, *M.* sp. bovine group 7 of Leach (Mbg7), *M. mycoides* subsp. *mycoides* small colony biotype (Mmm SC) and *M. capricolum* subsp. *capripneumoniae* (Mccp). The latter two are the causative agents of contagious bovine and caprine pleuropneumonia (CBPP and CCPP, respectively) [7], [8] and [9]. Organisms within this cluster share many antigenic and genotypic features and some are difficult to differentiate by conventional techniques [10]. The common traits exhibited by these mycoplasmas have often confused both diagnosis and taxonomy [8]. Mmm LC and Mmc were found to be very closely related with 99.9% sequence identity of their 16S rRNA genes [11]. Similar analyses based on 16S rRNA [12] and 16S–23S intergenic spacer region [13] have shown that *M. putrefaciens* was closely related to the *M. mycoides* cluster [12]. Other mycoplasma species such as *Mycoplasma cottewii* and *Mycoplasma yeatsii*

[14] have been shown to be closely related to *M. putrefaciens* [15]. However, the latter three species are still considered not to belong to the *M. mycoides* cluster.

Because of the economic importance of contagious agalactia, some countries have implemented prophylactic measures to control and, whenever possible, to eradicate the disease. This is the case of the Southwest of France, where ewe's milk cheese has great economic importance. Up to now, the control of the disease is based on the eradication of *M. agalactiae* infection, with its detection playing a key role in this strategy. The presence of this mycoplasma in milk tanks is regularly checked and a rapid identification allows the implementation of control measures in the infected herds. Control strategies regarding the other mycoplasmas involved in contagious agalactia are hampered by the lack of detection tools.

To date, there are specific PCR tests for the direct identification of *M. agalactiae* [16] and [17] and *M. putrefaciens* [3]. PCR assay for rapid detection of the mycoplasmas of the *M. mycoides* cluster causing contagious agalactia have already been described [18], [19] and [20]. However, these tests are based on the same sequence, CAP21, originally described by Bashiruddin et al. [18]. The CAP21 sequence encompasses two highly conserved genes coding for ribosomal proteins *rpsL* and *rpsG*, hence a very high probability of cross-reactions. Therefore, the aim of the present study was to design a PCR test for the rapid identification of all members of the *M. mycoides* cluster based on a more suitable target. A locus containing the glucokinase (*glk*) gene, a flanking gene of the arginine deiminase (ADI) operon, was chosen as a target to select a specific primer pair.

2. Materials and methods

2.1. Strains

A representative number of strains (*N*=46) of species belonging to the *M. mycoides* cluster, of the closely related species *M. putrefaciens*, *M. cottewii* and *M. yeatsii*, as well as the ruminant pathogens *Mycoplasma auris*, *Mycoplasma bovis* and *M. agalactiae*, were used for the initial validation of the assay at CIRAD, France (Table 1). Strains originating from various continents were included in order to ensure the universality of the assay. All were cultivated in modified Hayflick media containing sodium pyruvate and glucose [21] in a

high-security L_3 laboratory. A subsequent validation of the PCR test was performed at two other laboratories: representative samples of 68 strains of French origin were used at AFSSA, whereas the validation at IVRI (Izatnagar, India) was performed on 12 strains. In total, the validation included 130 strains: Mmm LC (31), Mmc (5), Mmm SC (5), Mcc (25), Mccp (6), Mbg7 (8), *M. putrefaciens* (14), *M. cottewii* (1), *M. yeatsii* (11), *M. agalactiae* (16), *M. bovis* (6), *M. auris* (1) and *Mycoplasma gallinarum* (1).

Table 1 List of mycoplasma strains used for initial validation of the specific PCR at CIRAD

Strain	Origin	Supplier institute
Mycoplasma mycoides subsp. *mycoides* LC		
Y-goat^T	Australia	Type strain
95010-C1	France	CIRAD
9298 (CM508)	Tanzania	NVI-S, Uppsala, Sweden. Dr. Bölske
9050-143	Ivory Coast	LPA, Bingerville, Ivory Coast. Dr. Domenech
9186 (CPR27)	Cameroon	LANAVET, Garoua, Cameroon. Dr. Martrenchar
99055	France	AFSSA-N, Niort, France. Dr. Mercier
2002-055 (VPNC2)	India	IVRI, Izatnagar, India. Dr. Srivastava and Dr. Singh
WK354	Switzerland	Bern U, University of Bern, Switzerland. Prof. Nicolet
55507-1	Germany	TiHo, Hannover, Germany. Dr. Schmidt
Kombolcha	Ethiopia	NVI-E, Debre Zeit, Ethiopia. Dr. Fikré
7730	France	CIRAD
9096 (C9415)	Nigeria	U. Nigeria, Nsukka, Nigeria. Dr. Shoyinka
8065	France	CIRAD
Mycoplasma mycoides subsp. *capri*		
PG3^T	Turkey	Type strain
2002-054 (VP9L)	India	IVRI, Izatnagar, India. Dr. Srivastava and Dr. Singh
L	France	CIRAD
9139 (11/91)	Turkey	NVI-E, Debre Zeit, Ethiopia. Dr. Fikré
Mycoplasma mycoides subsp. *mycoides* SC		
PG1^T	Unknown	Type strain
8740-Rita	Cameroon	LANAVET, Garoua, Cameroon. Dr. Martrenchar
T1/44	Tanzania	PANVAC, Addis Ababa, Ethiopia. Dr. Litamoi
Mycoplasma capricolum subsp. *capricolum*		
California kid-CIRAD	USA	Type strain
97058 (O12)	Morocco	IAV, Rabat, Morocco. Dr. Berrada
8086-1	France	CIRAD

IPX	France	CIRAD
2002-053 (VP286)	India	IVRI, Izatnagar, India. Dr. Srivastava and Dr. Singh
2002-057 (RLM23C)	India	IVRI, Izatnagar, India. Dr. Srivastava and Dr. Singh
4146	France	AFSSA-L, Lyon, France. Dr. Poumarat
Vienne	France	CIRAD
E-570	UK	NVI-S, Uppsala, Sweden. Dr. Bölske.
8601-2703-C3	Portugal	LNIV, Lisbon, Portugal. Dr. Machado
8601-50	Portugal	LNIV, Lisbon, Portugal. Dr. Machado
Mycoplasma capricoulum subsp. *capripneumoniae*		
F38	Kenya	Type strain
Gabes	Tunisia	CIRAD
94156 (438P)	Chad	LRVZ, Farcha, Chad. Dr. Hendrikx
9081(487P)	Oman	CIRAD
95043	Niger	CIRAD
9231 (Abomsa)	Ethiopia	NVI-E, Debre Zeit, Ethiopia. Dr. Fikré
Mycoplasma species bovine serogroup 7		
PG50^R	Australia	Reference strain
Calf 1	Nigeria	NVI-S
9733	India	BGVV, Jena, Germany. Dr. Sachse
Mycoplasma putrefaciens		
KS1	USA	Type strain
Mycoplasma cottewii		
VIS	Australia	Type strain
Mycoplasma yeatsii		
GIH	Australia	Type strain
Mycoplasma auris		
UIA	Australia	Type strain
Mycoplasma bovis		
8891 (5/2)	Turkey	MRI, Edimburgh, UK. Dr. Jones
97138	France	
97027	Germany	Hannover. Dr. Martin
95035 (C58)	Cameroon	LANAVET
Mycoplasma agalactiae		
2002-052 (VP15L)	India	IVRI, Izatnagar, India. Dr. Srivastava and Dr. Singh

2.2. PCR primer selection

The *glk* gene, flanking the ADI operon in the *M. mycoides* cluster, was chosen as target to select a specific primer pair. The *glk* sequences of the *M. mycoides* cluster reference strains $PG1^T$ and California kid^T were retrieved from the genomic sequence data published in GenBank (BX293980 and CP000123). In addition, the *glk* sequence of an Mmm LC strain (95010-C1) was obtained from a whole genome sequencing project (Genoscope, France). The AlignX software of Vector NTI Suite (InforMax) was used for sequence alignment. The degenerate primers glk-myc-2F (5′-TGCACTTGGTGAATATA**R**AAAAGG-3′) and glk-myc-2R (5′-GGATCTAAAGC**R**TGTATTA**R**TAATG-3′) were chosen within the most conserved regions of the *glk* gene.

2.3. PCR amplification and sequencing

The PCR conditions were similar to those set by Lorenzon et al. [22]. DNA template was produced with a slight modification of that described by Miserez et al. [23]. Briefly, a pellet from 1 ml culture was re-suspended in 50 µl of distilled water, and lysed for 1 h at 60 °C in 100 µl of lysis buffer (100 mM Tris-HCl, pH 8.5; 0.05% Tween 20; 0.24 mg ml^{-1} proteinase K). After inactivation of proteinase K at 95 °C for 10 min, 1 µl of the lysed mycoplasma cells diluted 1/10 in distilled water was added as template to the PCR mix. The reaction volume was fixed at 50 µl with: 1×Taq buffer (Qiagen), 1.5 mM $MgCl_2$, 0.3 mM dATP and dTTP, 0.15 mM dCTP and dGTP (dNTPs from Roche), 0.4 µM of each primer glk-myc-2F and glk-myc-2R, 2.5 U of Taq polymerase (Qiagen) and 1 µl of DNA template. The PCR reactions consisted in an initial denaturation step at 94 °C for 2 min, followed by 30 cycles of denaturation at 94 °C for 15 s, annealing at 50 °C for 15 s and extension at 72 °C for 15 s and a final extension step at 72 °C for 5 min in a Perkin-Elmer Gene Amp PCR system 2700 (Applied Biosystems). PCR-amplified fragments from the reference strains $PG50^R$ (Mbg7) and $PG3^T$ (Mmc) and from an Mccp strain (Gabes) were sent for sequencing to Genome express, Meylan, France.

The use of higher proportions of dATP and dTTP in the PCR mix for mycoplasma PCR is justified given the low GC % in mycoplasma (25% in the case of the *M. mycoides* cluster). This is particularly important when amplifying large DNA fragments or when decreasing the quantity of dNTP in the reaction. In the case of this glk-PCR at the defined dNTP concentration, a common dNTP pre-mix will give comparable results.

For the secondary validation at AFSSA, the same reagents and procedure were applied using a Gene Cycler Thermal Cycler (Bio-Rad). Likewise, at IVRI (India) the same PCR conditions were validated using Taq polymerase (Fermentas), dNTP mix (Sigma-Aldrich) and a PCR thermocycler Mastercycler personal (Eppendorf, North America).

2.4. Primary field validation on milk samples

Twenty-one goat milk samples were collected from milk tanks in the Poitou-Charente region in France, where mycoplasma infections are frequent. These samples were sent to CIRAD, where 0.5 ml of milk was seeded in 4.5 ml of standard mycoplasma broth and incubated for 5 days at 37 °C for an initial enrichment. The presence of mycoplasmas was then tested by performing the "glk-PCR" and plating a drop of culture onto mycoplasma agar medium. Positive cultures were purity checked and isolated mycoplasma strains identified by standard techniques.

2.5. Evaluation of the sensitivity of glk-PCR

A 24-h culture of an Mmm LC strain (95010-C1) at the exponential phase of growth was used to evaluate the sensitivity of this PCR test. Ten-fold dilutions of this culture in mycoplasma medium were used both for the titration of the culture, by seeding 20 µl drops onto solid medium, and for the evaluation of PCR sensitivity, treating these dilutions as separate samples.

3. Results

3.1. Sequence alignment and design of a specific primer pair

The complete *glk* sequences available from the sequenced genomes of three mycoplasmas belonging to the *M. mycoides* cluster were aligned: two copies of this gene from strain PG1^T of Mmm SC (786 and 788 bp) and one copy from strains California kid^T of Mcc (788 bp) and 95010-C1 of Mmm LC (786 bp). These sequences showed 87.2% overall identity (Fig. 1). The 101 polymorphic positions were rather uniformly distributed along the sequences, though certain positions were relatively conserved. This allowed the selection of adequate primers with a limited number of degenerate bases (one for glk-myc-2F and two for glk-myc-2R), which allowed the efficient PCR amplification of a 428-bp-long fragment.

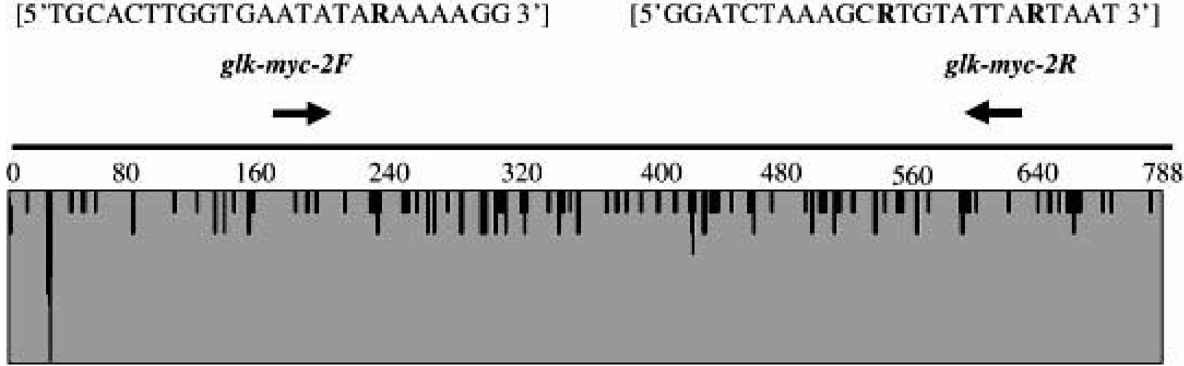


Figure 1. Schematic representation of the complete *glk* gene sequence alignment of three mycoplasmas belonging to the *M. mycoides* cluster.

The two *glk* copies of strain PG1^T (*M. mycoides* subsp. *mycoides* small colony) 788 and 786 bp, and the single *glk* sequences of strains California kid^T (*M. capricolum* subsp. *capricolum*) 788 bp and 95010-C1 (*M. mycoides* subsp. *mycoides* LC) 786 bp were retrieved from available genomic data and aligned for sequence comparison. Vertical black lines along the shaded region represent residual polymorphic positions. The relatively conserved regions in which the forward and reverse primers (glk-myc-2F and glk-myc-2R) were designed are indicated by arrows.

Sequences from the other species of the *M. mycoides* cluster encompassing the genes located between *mgtA* and *ptsG* or *glk* obtained in this study were deposited at GenBank: Mmm LC strain Y-goat^T (EF529697), Mmm LC strain 95010-C1 (EF529696), Mmc strain PG3^T (EF529698) and Mbg7 strain PG50^R (EF529699).

A second alignment (380-bp-long sequences) was performed, also including the sequences of strains Gabes (Mccp) PG50^R (Mbg7) and PG3^T (Mmc) obtained after PCR amplification. The alignment allowed the comparison of all species of the *M. mycoides* cluster and construction of a phylogenetic tree using the Neighbor Joining method of Saitou and Nei (Vector NTI Suite) on a fragment of the *glk* gene (Fig. 2). The overall identity of the seven sequences analysed (83.2%) fell slightly below the value obtained by comparison of the complete *glk* gene from three strains. The sequences of strains PG50^R, Gabes and California kid^T were 96.8% identical, with PG50^R being more closely related to the *M. capricolum* subspecies. Sequences from PG3^T and 95010-C1 differed only by eight

bases and grouped together with the two *glk* copies of PG1^T, with an overall identity of 91.6%.

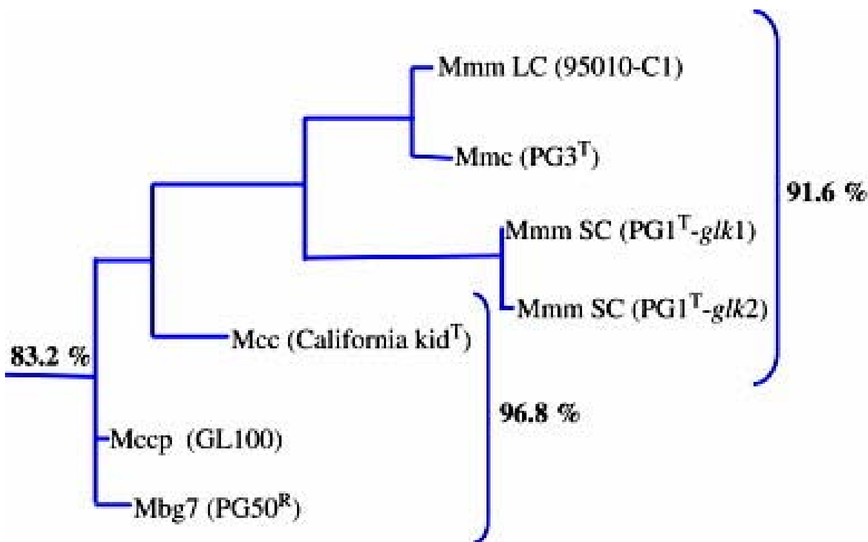


Figure 2 Distance tree calculated from the alignment of a 380-bp-fragment of the *glk* gene of six strains representing the *M. mycoides* cluster. Percentages of identity for each group of strains are shown at the right of the parenthesis, with the overall identity indicated at the root of the tree.

Besides, it was found that the *glk* gene exists in two copies (MSC_0863 and MSC_0875) in the Mmm SC-type strain PG1^T (BX293980) and this gene was annotated as a pseudogene (MCAP0092) in the Mcc-type strain California kid^T (CP000123). Sequence alignment of the pseudogene with sequences from other Mcc strains (California kid-CIRAD, IPX and 8086-1) revealed a deletion of one A at position 524 that resulted in a frame-shift in the sequence of the fully sequenced strain California kid^T.

3.2. Validation of the PCR test

The validation of the PCR assay gave consistent results in three independent laboratories. A DNA fragment of the expected size (428 bp) was obtained by analysis of 80 strains belonging to the *M. mycoides* cluster. No amplification was obtained with any of the heterologous species such as *M. agalactiae*, *M. bovis*, *M. cottewii*, *M. yeatsii*, *M. auris* and *M. gallinarum* (N=35) (Fig. 3). On the other hand, 3 out of the 14 *M. putrefaciens* strains tested yielded an amplified product by PCR. The band obtained from strain KS1^T was of similar size and equal intensity than those obtained from strains of the *M. mycoides* cluster. Two other *M. putrefaciens* strains yielded very faint bands of the same size.

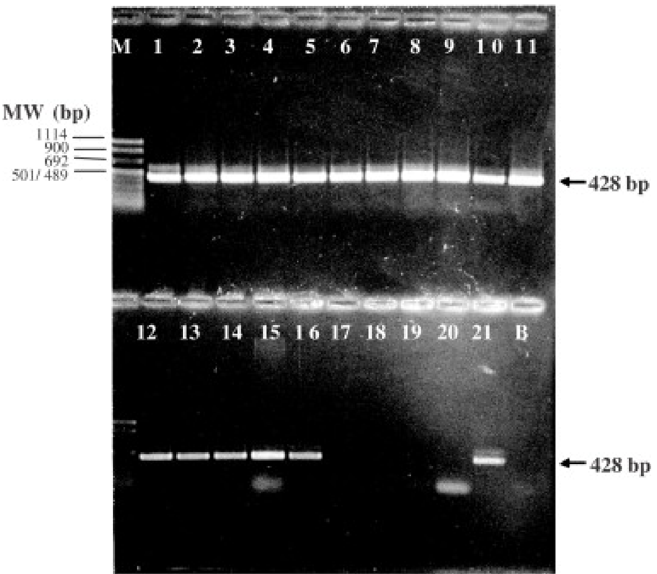


Figure 3 A 428 bp amplicon was obtained from all strains belonging to the *M. mycoides* cluster**.**

From top to bottom and from left to right M: molecular weight marker VIII (Roche); lanes 1-4: *Mycoplasma mycoides* subsp. *mycoides* LC strains Kombolcha, 99045, Y-goat^{T} and 8065; lanes 5-7: *M. mycoides* subsp. *capri* strains PG3^{T}, 9139 (11/91) and L; lanes 8 and 9: *M. mycoides* subsp. *mycoides* SC strains T1/44 and PG1^{T}; lanes 10-12: *M. capricolum* subsp. *capricolum* strains 2002-057, California kid^{T} and IPX; lanes 13 and 14 to *M. capricolum* subsp. *capripneumoniae* strains 9231-Abomsa and 94156 (438P); lanes 15 and 16: *M.* sp. bovine group 7 of Leach strains PG50^{R} and Calf-1; lanes 17 and 18: *M. putrefaciens* strains 11174 and Tours 2; lane 19: *M. yeatsii* strain GIH^{T}; lane 20: *M. cottewii* strain VIS^{T}; lane 21: positive control Mmm LC strain 95010-C1; and B: blank, deionised H_2O.

3.3. Preliminary validation on milk samples

Out of the 21 milk samples tested, 8 were found positive by glk-PCR (Table 2). Pure mycoplasma cultures were obtained from six of them, while for the other two mycoplasmas, cultures could not be isolated because of heavy contamination with other bacteria. Six of the isolated mycoplasma strains were identified as *M. capricolum* subsp. *capricolum* for milk sample 3 and *M. mycoides* subsp. *mycoides* LC for the rest of the isolates.

Table 2. Validation of glk-PCR from field milk samples.

Milk sample	Culture	glk-PCR	Species identification
1	B	–	
2	B	–	
3	M	Positive	*Mycoplasma capricolum* subsp. *capricolum*
4	B	–	
5	M+B	Positive	Not possible due to contamination
6	B	–	
7	M	Positive	*Mycoplasma mycoides* subsp. *mycoides* LC
8	B	–	
9	B	–	
10	B	–	
11	B	–	
12	B	–	
13	M	Positive	*Mycoplasma mycoides* subsp. *mycoides* LC
14	B	–	
15	M	Positive	*Mycoplasma mycoides* subsp. *mycoides* LC
16	B	–	
17	B	–	
18	M+B	Positive	Not possible due to contamination
19	B	–	
20	M	Positive	*Mycoplasma mycoides* subsp. *mycoides* LC
21	M+B	Positive	*Mycoplasma mycoides* subsp. *mycoides* LC

M, pure mycoplasmas colonies; B, classical bacterial colonies; and M+B, mixed bacterial and mycoplasmal colonies.

3.4. Estimation of PCR sensitivity

A positive PCR result was obtained up to dilution 10^{-6} and the viable cell count at this dilution was 30 colonies (CFU) per 20 µl. Hence, the sensitivity of the PCR can be estimated at 3.2 log CFU/ml, according to the formula: log CFU/ml=log (colony count/20 µl) + 1.7.

4. Discussion and conclusion

For effective control of ruminant mycoplasmoses, rapid and specific detection is of primordial importance and PCR is the method of choice to this aim. Specific PCR assays have already been designed for the rapid detection of Mmm SC, the agent of CBPP, and Mccp, the agent of CCPP. The task was relatively simple for these two diseases, since they are both caused by a single pathogen. The main difficulty in these two cases was due to the close proximity of the etiologic organisms to many other related mycoplasma species. In the case of contagious agalactia, the situation is more complex because this disease may be caused by different mycoplasma species: *M. agalactiae*, Mmm LC, Mmc, Mcc and *M. putrefaciens*. The rapid detection and identification of *M. agalactiae* can be obtained by various PCR assays that have been extensively validated [16] and [17]. Furthermore, when detecting *M. agalactiae* in small ruminants, the risk of cross-reactions with other mycoplasma species causing contagious agalactia is very limited, as this species belongs to a very distant phylogenetic group [24] and [25]. The closest proximity is to the species *M. bovis*, which is very rarely isolated from small ruminants.

The detection and identification of the other agents causing contagious agalactia is complicated by the fact that they all belong to, or are very closely related to, the *M. mycoides* cluster. This cluster was initially characterized by conventional methods [8] and more recently by 16S rRNA phylogenetic analysis [11]. However, whatever the technique used, establishing a precise classification and taxonomy for this cluster is hampered by a number of difficulties. The different mycoplasmas belonging to this cluster share many genetic and phenotypic features. Furthermore, the subspecies *M. mycoides* and *M. capricolum* are characterized by high intraspecies variability. This may be the reason why the subspecies Mccp has been named only recently, in 1993 [9], whereas the group of strains referred to as Mbg7 remains unassigned. As a consequence, it is difficult to design diagnostic tests that are both universal and specific. An additional complication may eventually arise from the detection of intermediary strains of difficult classification. This is to be expected, particularly when an exhaustive validation including large numbers of strains of diverse geographical origin is performed [26].

The initial objective of this study was to obtain specific PCR assays for each of the mycoplasma species of the *M. mycoides* cluster causing contagious agalactia. To this aim, the ADI operon was identified as a potential target. This operon had already been used to

design specific PCR assays for the detection of *M. putrefaciens* [3] and Mccp [27]. The organization of the ADI operon seemed to be quite species specific, at least when comparisons where limited to the reference strains and Mccp strain GL100 (accession number AY529662). When the sequences from other strains were included in the analysis it was noted that the organization of the operon varied also within each species, which precluded its use for the design of specific PCR assays (Fig. 4). The *glk* gene, flanking the ADI operon, was then chosen as an alternative target. This gene codes for the enzyme glucokinase (EC 2.7.1.2), a group of enzymes found in invertebrates and microorganisms, highly specific for glucose metabolism [28]. In spite of the apparent universality of this enzyme, it has only been found in some mycoplasma species belonging to the *M. mycoides* cluster, and *Mycoplasma mobile*, *Mycoplasma hyopneumoniae* and *Mycoplasma pulmonis* (http://cbi.labri.fr/outils/molligen/). In the Mcc type strain, which has been fully sequenced, *glk* appears as a pseudogene. However, sequences obtained from other Mcc strains at CIRAD, including a stock of the type strain California kid^T, showed that the gene may be functional in the *capri*colum species. This emphasizes the fact that different stocks of the same strain may differ, a possible consequence of multiple *in vitro* passages. Furthermore, a blast analysis (blastn and blastp performed at the NCBI, http://www.ncbi.nlm.nih.gov/blast/ on 09/05/2007, BLASTP 2.2.16 [25/03/2007]) against all non-redundant nucleotide and protein sequence databases for bacteria showed that the *glk* sequences found in the *M. mycoides* cluster are quite unique. The closest relative was a sequence from *Spiroplasma citri*, which is a plant pathogen belonging to the same 16S rRNA phylogenetic group. Still, the amino-acid similarity was limited to 55%. Glucokinase sequences from the *M. mycoides* cluster seemed to be more closely related to the *glk* sequences of Gram-positive bacteria that possess an amino-acid motif: CXCGXXGCXE, which is involved in the enzymatic activity [28]. However, this motif is not entirely conserved in the *M. mycoides* cluster, lacking the G at position 7 (CNCGLN*N*CIE), which highlights the peculiarity of this group of bacteria.

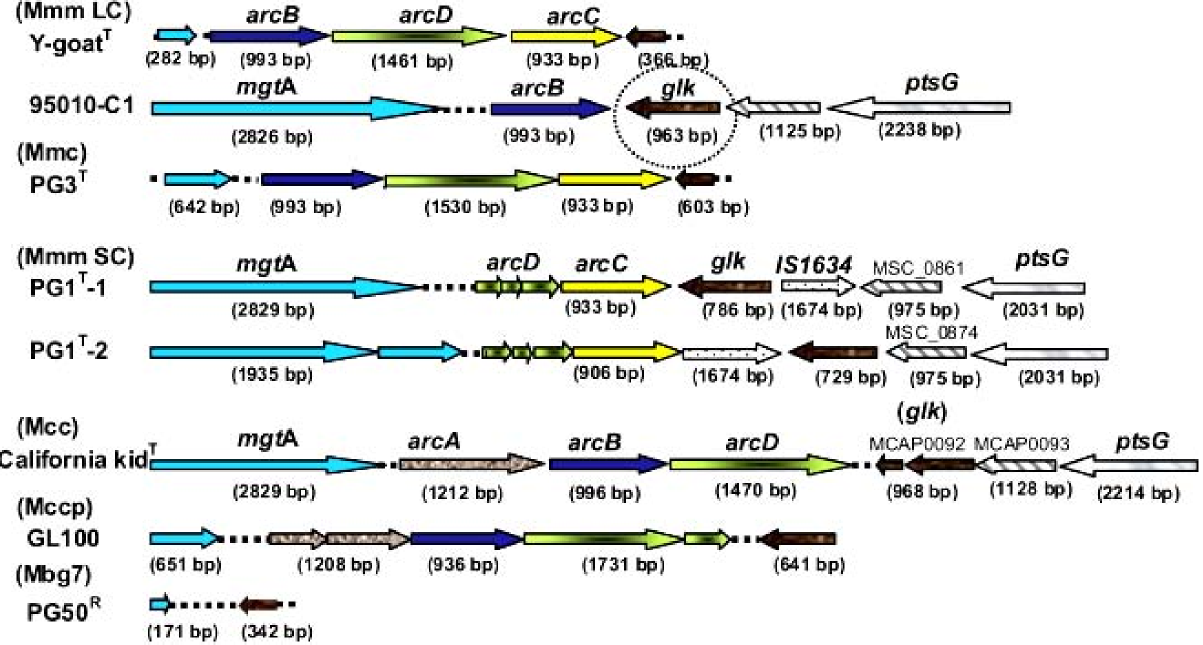


Figure 4 Organization of the arginine deiminase (ADI) operon within the *M. mycoides* cluster. The ADI operon in this cluster is flanked by the glucokinase gene (*glk*) and the magnesium transport ATPase (*mgt*A). The organization of the operon is *arc*B-*arc*D-*arc*C in the subspecies of *M. mycoides* and *arc*A-*arc*B-*arc*D in the subspecies of *M. capricolum*. The genome sequence of the type strain PG1^T revealed that this operon and flanking genes are duplicated. Note that in this particular locus the ADI operon is absent in the reference strain PG50^R.

The primer pair designed on the *glk* gene allowed the amplification of the expected DNA fragment from all the *M. mycoides* cluster strains tested, showing that this gene is well conserved within the cluster and proving the universality of the test. The tree inferred from the sequence alignment of the amplified products was in agreement with previous results. It showed that strain PG50^R (Mbg7) was more closely related to *M. capricolum* than to the *M. mycoides* subspecies. This had already been shown by analysis of 16S–23S rRNA intergenic sequences [13], as well as with a gene of unknown function [29]. In addition, it also confirmed the very close relationship existing between Mmm LC and Mmc, two subspecies that should certainly be grouped into a single entity [11], [30] and [31].

Some cross-reactivity was observed with a few strains belonging to the species *M. putrefaciens*. Cross-reactions have already been observed in the assessment of different PCR tests used for routine identification of species of the *M. mycoides* cluster and *M. putrefaciens*. An "*M. putrefaciens* specific" PCR yielded positive results with members of the *M. mycoides* cluster [32]. This was also observed by Persson et al. [33] when

lowering the annealing temperature or using substantial amount of template, which demonstrated once more the close relationship existing between these species. The proximity of these two bacterial entities needs to be assessed more precisely. Based on 16S rRNA sequence comparisons *M. putrefaciens* was considered to be related to the *M. mycoides* cluster, though the number of nucleotide positions specific to *M. putrefaciens* should suffice to consider it as a separate species [11]. 16S–23S intergenic sequence comparisons also showed that *M. putrefaciens* was related, though somewhat distant to the *M. mycoides* cluster. From a practical point of view, however, the few cross-reactions seen between these species should not be of consequence, as a specific PCR assay is already available for the rapid detection of *M. putrefaciens* [2]. More importantly, this new PCR assay does not yield any cross-reactions with any of the *M. agalactiae* strains tested.

The routine procedure to detect mycoplasmas in milk tanks in field diagnostic laboratories is to seed the milk in adequate culture medium and wait for 5–6 days before checking for mycoplasma growth by streaking this enriched culture onto solid medium to isolate mycoplasma colonies and identify them. Applying the glk-PCR before culture allowed us to detect eight positive milk samples, while culture and identification was achieved only for six of them. This has proved the interest of performing PCR before culture, as the "classical" isolation and identification techniques were much slower and were also hampered by the presence of bacterial contaminants in two cases. The sensitivity of this PCR was evaluated at 3.2 log CFU/ml. In our case, the number of CFU should almost be equivalent to the number of mycoplasmas, as we chose an easy growing strain in the exponential phase of growth. This ensured that there were no dead bacteria that could have biased the estimation by artificially increasing the sensitivity of the PCR. This sensitivity is in accordance with other PCR tests. When detecting mycoplasmas in enrichment broth, sensitivity is not a real issue, as a visible turbidity corresponds to approximately 8 log CFU/ml.

In conclusion, this new PCR can be used for the detection of mycoplasmas of the *M. mycoides* cluster causing contagious agalactia in small ruminants. When used in combination with other PCR assays able to detect *M. putrefaciens* and *M. agalactiae* [16] and [17], it should provide a means to detect all the mycoplasma agents involved. The use of multiple PCR assays may be cumbersome, but the necessity arises from the number of mycoplasma species involved in contagious agalactia. This new assay offers the great

advantage of yielding a specific amplification and, therefore, it does not necessitate any further manipulations such as restriction enzyme analysis or sequencing, which may not be practical for routine veterinary laboratories. The validation performed in three different laboratories demonstrated that it is a robust technique that can be applied successfully in different environmental conditions. This new tool may be of great help for the implementation of control measures directed towards contagious agalactia.

Acknowledgments

We are thankful to the French Ministry of Foreign Affairs and the French Embassy in Addis Ababa for financing Dr. Woubit Salah's sojourn at CIRAD-UPR 15, "Control of exotic and emerging animal diseases". We are also grateful to Valerie Barbe, Genoscope, Ivry, and to Dr. Alain Blanchard and his colleagues at INRA, Bordeaux, for genome sequencing and annotation facilities of *M. mycoides* subsp. *mycoides* LC strain 95010-C1.

References

[1] Nicolas RAJ. Contagious agalactia. In: OIE manual of diagnostic tests and vaccines for terrestrial animals (mammals, birds and bees), vol 2, Paris, 2004. p. 607–14.

[2] A.C. Johnstone and C.M. King, Bovine and ovine mycoplasmal arthritis and synovitis (abstract), *NZ Vet J* **51** (2003), p. 46.

[3] A. Peyraud, S. Woubit, J.B. Poveda, C. De la Fe, P. Mercier and F. Thiaucourt, A specific PCR for the detection of *Mycoplasma putrefaciens*, one of the agents of the contagious agalactia syndrome of goats, *Mol Cell Probes* **17** (2003), pp. 289–294.

[4] F. Rodriguez, S. Kennedy, T.D. Bryson, A. Fernandez, J.L. Rodriguez and H.J. Ball, An immunohistochemical method of detecting Mycoplasma species antigens by use of monoclonal antibodies on paraffin sections of pneumonic bovine and caprine lungs, *Zentralbl Veterinarmed B* **43** (1996), pp. 429–438.

[5] F. Thiaucourt and G. Bolske, Contagious caprine pleuropneumonia and other pulmonary mycoplasmoses of sheep and goats, *Rev Sci Tech* **15** (1996), pp. 1397–1414.

[6] J.B. Bashiruddin, P. De Santis, E. Varga and L. Stipkovits, Confirmation of the presence of *Mycoplasma bovis* in Hungarian cattle with pneumonia resembling pleuropneumonia, *Vet Rec* **148** (2001), pp. 743–746.

[7] F. Bonnet, C. Saillard, J.M. Bove, R.H. Leach, D.L. Rose and G.S. Cottew *et al.*, DNA relatedness between field isolates of *Mycoplasma* F38 group, the agent of contagious caprine pleuropneumonia, and strains of *Mycoplasma capricolum*, *Int J Syst Bacteriol* **43** (1993), pp. 597–602.

[8] G.S. Cottew, A. Breard, A.J. DaMassa, H. Erno, R.H. Leach and P.C. Lefevre *et al.*, Taxonomy of the *Mycoplasma mycoides* cluster, IS*r J Med Sci* **23** (1987), pp. 632–635.

[9] R.H. Leach, H. Erno and K.J. MacOwan, Proposal for designation of F38-type caprine mycoplasmas as *Mycoplasma capricolum* subsp. *capripneumoniae* subsp. nov. and consequent obligatory relegation of strains currently classified as *M. capricolum* (Tully, Barile, Edward, Theodore, and Erno 1974) to an additional new subspecies, *M. capricolum* subsp. *capricolum* subsp. nov, *Int J Syst Bacteriol* **43** (1993), pp. 603–605.

[10] B. Pettersson, G. Bölske, F. Thiaucourt, M. Uhlen and K.E. Johansson, Molecular evolution of *Mycoplasma capricolum* subsp. *capripneumoniae* strains, based on polymorphisms in the 16S rRNA genes, *J Bacteriol* **180** (1998), pp. 2350–2358.

[11] B. Pettersson, T. Leitner, M. Ronaghi, G. Bolske, M. Uhlen and K.E. Johansson, Phylogeny of the *Mycoplasma mycoides* cluster as determined by sequence analysis of the 16S rRNA genes from the two rRNA operons, *J Bacteriol* **178** (1996), pp. 4131–4142.

[12] W.G. Weisburg, J.G. Tully, D.L. Rose, J.P. Petzel, H. Oyaizu and D. Yang *et al.*, A phylogenetic analysis of the mycoplasmas: basis for their classification, *J Bacteriol* **171** (1989), pp. 6455–6467.

[13] R. Harasawa, H. Hotzel and K. Sachse, Comparison of the 16S–23S rRNA intergenic spacer regions among strains of the *Mycoplasma mycoides* cluster, and reassessment of the taxonomic position of *Mycoplasma* sp. bovine group 7, *Int J Syst Evol Microbiol* **50** (Pt 3) (2000), pp. 1325–1329.

[14] A.J. DaMassa, J.G. Tully, D.L. Rose, D. Pitcher, R.H. Leach and G.S. Cottew, *Mycoplasma auris* sp. nov., *Mycoplasma cottewii* sp. nov., and *Mycoplasma yeatsii* sp. nov., new sterol-requiring mollicutes from the external ear canals of goats, *Int J Syst Bacteriol* **44** (1994), pp. 479–484.

[15] M. Heldtander, B. Pettersson, J.G. Tully and K.E. Johansson, Sequences of the 16S rRNA genes and phylogeny of the goat mycoplasmas *Mycoplasma adleri*, *Mycoplasma auris*, *Mycoplasma cottewii* and *Mycoplasma yeatsii*, *Int J Syst Bacteriol* **48** (Pt 1) (1998), pp. 263–268.

[16] J.B. Bashiruddin, J. Frey, M.H. Konigsson, K.E. Johansson, H. Hotzel and R. Diller *et al.*, Evaluation of PCR systems for the identification and differentiation of *Mycoplasma agalactiae* and *Mycoplasma bovis*: a collaborative trial, *Vet J* **169** (2005), pp. 268–275.

[17] S. Subramaniam, D. Bergonier, F. Poumarat, S. Capaul, Y. Schlatter and J. Nicolet *et al.*, Species identification of *Mycoplasma bovis* and *Mycoplasma agalactiae* based on the *uvr*C genes by PCR, *Mol Cell Probes* **12** (1998), pp. 161–169.

[18] J.B. Bashiruddin, T.K. Taylor and A.R. Gould, A PCR-based test for the specific identification of *Mycoplasma mycoides* subspecies *mycoides* SC, *J Vet Diagn Invest* **6** (1994), pp. 428–434.

[19] G. Greco, M. Corrente, V. Martella, A. Pratelli and D. Buonavoglia, A multiplex-PCR for the diagnosis of contagious agalactia of sheep and goats, *Mol Cell Probes* **15** (2001), pp. 21–25.

[20] J.L. Rodriguez, R.W. Ermel, T.P. Kenny, D.L. Brooks and A.J. DaMassa, Polymerase chain reaction and restriction endonuclease digestion for selected members of the "*Mycoplasma mycoides* cluster" and *Mycoplasma putrefaciens*, *J Vet Diagn Invest* **9** (1997), pp. 186–190.

[21] F. Thiaucourt, C. Guérin, V. Mady and P.C. Lefèvre, Diagnostic de la pleuropneumonie contagieuse caprine: améliorations récentes, *Rev Sci Tech Off Int Epiz* **11** (1992), pp. 859–865.

[22] S. Lorenzon, H. Wesonga, L. Ygesu, T. Tekleghiorgis, Y. Maikano and M. Angaya *et al.*, Genetic evolution of *Mycoplasma capricolum* subsp. *capripneumoniae* strains and molecular epidemiology of contagious caprine pleuropneumonia by sequencing of locus H2, *Vet Microbiol* **85** (2002), pp. 111–123.

[23] R. Miserez, T. Pilloud, X. Cheng, J. Nicolet, C. Griot and J. Frey, Development of a sensitive nested PCR method for the specific detection of *Mycoplasma mycoides* subsp. *mycoides* SC, *Mol Cell Probes* **11** (1997), pp. 103–111.

[24] J.G. Mattsson, B. Guss and K.E. Johansson, The phylogeny of *Mycoplasma bovis* as determined by sequence analysis of the 16S rRNA gene, *FEMS Microbiol Lett* **115** (1994), pp. 325–328.

[25] B. Pettersson, M. Uhlen and K.E. Johansson, Phylogeny of some mycoplasmas from ruminants based on 16S rRNA sequences and definition of a new cluster within the hominis group, *Int J Syst Bacteriol* **46** (1996), pp. 1093–1098.

[26] F. Thiaucourt, S. Lorenzon, A. David and A. Breard, Phylogeny of the *Mycoplasma mycoides* cluster as shown by sequencing of a putative membrane protein gene, *Vet Microbiol* **72** (2000), pp. 251–268.

[27] S. Woubit, S. Lorenzon, A. Peyraud, L. Manso-Silvan and F. Thiaucourt, A specific PCR for the identification of *Mycoplasma capricolum* subsp. *capripneumoniae*, the causative agent of contagious caprine pleuropneumonia (CCPP), *Vet Microbiol* **104** (2004), pp. 125–132.

[28] L.R. Mesak, F.M. Mesak and M.K. Dahl, *Bacillus subtil*IS GlcK activity requires cysteines within a motif that discriminates microbial glucokinases into two lineages, *BMC Microbiol* **4** (2004), p. 6.

[29] H. Hotzel, K. Sachse and H. Pfutzner, A PCR scheme for differentiation of organisms belonging to the *Mycoplasma mycoides* cluster, *Vet Microbiol* **49** (1996), pp. 31–43.

[30] M. Costas, R.H. Leach and D.L. Mitchelmore, Numerical analysis of PAGE protein patterns and the taxonomic relationships within the '*Mycoplasma mycoides* cluster', *J Gen Microbiol* **133** (Pt 12) (1987), pp. 3319–3329.

[31] M.P. Monnerat, F. Thiaucourt, J.B. Poveda, J. Nicolet and J. Frey, Genetic and serological analysis of lipoprotein LppA in *Mycoplasma mycoides* subsp. *mycoides* LC and *Mycoplasma mycoides* subsp *capri*, *Clin Diagn Lab Immunol* **6** (1999), pp. 224–230.

[32] D. Le Grand, E. Saras, D. Blond, M. Solsona and F. Poumarat, Assessment of PCR for routine identification of species of the *Mycoplasma mycoides* cluster in ruminants, *Vet Res* **35** (2004), pp. 635–649.

[33] A. Persson, B. Pettersson, G. Bolske and K.E. Johansson, Diagnosis of contagious bovine pleuropneumonia by PCR-laser- induced fluorescence and PCR-restriction endonuclease analysis based on the 16S rRNA genes of *Mycoplasma mycoides* subsp. *mycoides* SC, *J Clin Microbiol* **37** (1999), pp. 3815–3821.

Paper II

Molecular and Cellular Probes
Volume 17, Issue 6, December 2003, Pages 289-294
View Record in Scopus, Cited By in Scopus (6)
Copyright © 2003 Elsevier Ltd. All rights reserved.

A specific PCR for the detection of *Mycoplasma putrefaciens*, one of the agents of the contagious agalactia syndrome of goats

A. Peyraud[a], S. Woubit[b, 1], J. B. Poveda[c], C. De la Fe[c], P. Mercier[d] and F. Thiaucourt[a]

[a] CIRAD EMVT Sante Animale, Animal Health Program, TA30/G Campus International de Baillarguet, 34398, Montpellier Cedex 5, France
[b] Faculty of Veterinary Medicine, Debre Zeit, Ethiopia
[c] Epidemiology and Preventive Medicine, Veterinary Faculty, University of Las Palmas, Trasmontaña s/n., 35416 Arucas, Las Palmas, Spain
[d] AFSSA France 60 rue Pied-de-Fond, BP 3081, 79012, Niort Cedex, France , Accepted 29 July 2003; Available online 24 September 2003.

Abstract

Mycoplasma putrefaciens is listed as one of the etiologic agents of the contagious agalactia syndrome by the world organisation for animal health. This species has been characterized only recently, 1974, and the number of outbreaks caused by this microorganism so far is very scarce. It induces mastitis in infected goats although other symptoms such as arthritis in adults and septicaemia in kids are also frequently described. Up to now, the identification of *M. putrefaciens* relied on classical isolation and identification techniques which present a number of limitations. Specific primers for PCR have been designed based on sequence comparisons of the *Arc*B gene among the '*Mycoplasma mycoides* cluster' and related species such as *Mycoplasma cottewii* and *Mycoplasma yeatsii*. Sequence alignments confirmed the taxonomic position of *M. putrefaciens*, which is related to the '*M. mycoides* cluster' but also very close to *M. yeatsii*. The polymorphism observed amongst the different *Arc*B sequences allowed the determination of a primer pair yielding a specific amplification of a 316 bp-long DNA fragment by PCR. This PCR was validated in two different laboratories with a variety of mycoplasma strains isolated from goats. This new PCR technique will be very useful for a quicker determination of *M. putrefaciens* strains as well as a better understanding of the prevalence of *M. putrefaciens* infections.

Author Keywords: *Mycoplasma putrefaciens*; Specific PCR; Contagious agalactia

1. Introduction

Mycoplasma putrefaciens (Mput) is listed as one of the etiologic agents of the contagious agalactia syndrome by the world organisation for animal health [1] but it has received little attention up to now.

The reference strain for this species, KS1, was first isolated by HE Adler in 1954 from the same goat infected with the California goat strain that was deposited in the ATCC under the name 'California kid', becoming the reference strain for *Mycoplasma capricolum* subsp. *capricolum* (Mcc) [2]. However, subsequent studies showed that the KS1 strain differed markedly from the California kid strain. It had a higher G+C%, 28.9% compared to 25.5%, it did not hydrolize arginine, did not liquefy coagulated serum and showed little serological cross-reactions by the metabolic inhibition test. Therefore, it was classified as a new species and was denominated *M. putrefaciens*, owing to its putrid odour in broth and agar cultures [2].

The taxonomic position of *M. putrefaciens* was not clearly 'established in 1987', when Cottew defined the so-called '*M. mycoides* cluster' [3]. Comparison of 16S rRNA genes showed that *M. putrefaciens* was closely related to the members of the '*M. mycoides* cluster' [4], although the closest relationship existed with the newly described species *Mycoplasma cottewii* and *Mycoplasma yeatsii* [5]. The relative proximity of *M. putrefaciens* to the '*M. mycoides* cluster' was also confirmed by comparing the 16S–23S intergenic spacer region [6]. In contrast, no serological cross-reaction was observed between the LppA lipoprotein found in the members of the '*M. mycoides* cluster' and that of the KS1 strain [7].

Few mycoplasma isolates have been characterised as *M. putrefaciens* worldwide. One of the first isolates was obtained in 1972 by Perreau in central France but it was not identified until a few years later [8], following the description of the new species *M. putrefaciens*. Another isolate was obtained in 1979 in California [9]. Two other strains were isolated in France in 1985 [10], although in different regions than the original French isolate of 1972. In 1987 it was isolated again in California [11]. More recently, these strains were isolated in the Canary Islands from kids with polyarthritis [12]. Interestingly, *M. putrefaciens* has also been detected in Australia, although the strain was isolated from the ear canal of

healthy goats that harboured other mycoplasma species at the same time in the same location [13].

Early descriptions of *M. putrefaciens* infections recorded mastitis as the only symptom [9], a feature that distinguished it from the members of the '*M. mycoides* cluster' infecting goats such as *M. capricolum* subsp. *capricolum* (Mcc), *M. mycoides* subsp. *mycoides* LC (MmmLC) and *M. mycoides* subsp. *capri* (Mmc). The latter are the causative agents of what has been called the 'MAKePS' syndrome, (mastitis, arthritis, keratitis, pneumonia and septicaemia), which can be observed in adult goats and kids [14]. The experimental inoculation of *M. putrefaciens* into the udder of goats resulted in excretion of the organism in high titres, soon followed by reduced lactation or agalactia [15]. Neither fever nor transmission to the other udder was observed. Subsequent descriptions of *M. putrefaciens* infections differed markedly. Although agalactia was the main symptom observed in lactating goats, arthritis was also recorded in kids and young goats. Death of kids resulting from septicaemia was also recorded [10]. Similar observations were made in California in 1987, where adults and kids suffered from acute arthritis with fibrinopurulent discharge. *M. putrefaciens* was also isolated from other organs such as brain, kidneys, lymph nodes, uterus and from urine samples [11 and 16]. Polyarthritis in kids was also the main symptom observed in an outbreak occurring in the Canary Islands, also associated with general symptoms such as lameness, depression and anorexia, with none of the animals being pyrexic [12]. Therefore, *M. putrefaciens* may certainly be considered one of the mycoplasma agents causing the MAKePS syndrome, although ocular symptoms have not been described so far. The virulence of *M. putrefaciens* seems to be lower than that of other mycoplasma species such as *M. capricolum* subsp. *capripneumoniae* (Mccp), Mcc or MmmLC [17]. Interestingly, *M. putrefaciens* was isolated in association with these other mycoplasma species in a single herd in which enzootic pneumonia had been observed [18]. In recent years *M. putrefaciens* strains, as well as other mycoplasmas of the '*M. mycoides* cluster', have been regularly isolated from goat herds in the Charente Poitou region in Western France. The increased incidence of *M. putrefaciens* infections, added to the scarcity of specific molecular typing tools available for this organism, lead us to look for a specific PCR method enabling its rapid detection and identification.

The arginine deiminase pathway is the most widespread anaerobic route for arginine degradation and constitutes a major source of energy for several microorganisms. This

pathway comprises three reactions catalysed by arginine deiminase (ADI, EC 3.5.3.6), ornithine transcarbamylase (OTC, EC 2.1.3.3) and carbamate kinase (CK, EC 2.7.2.2). Of these three enzymes, OTC is absent only in some obligate parasitic bacteria. Its widespread presence in bacteria and its apparent polymorphism [19] lead to the selection of this gene as a target for a specific PCR.

2. Materials and methods

2.1. Strains used

An initial search for specific primers was performed using the reference strain for *M. putrefaciens* (KS1) and a limited set of strains belonging to the '*M. mycoides* cluster', as well as closely related species such as *M. yeatsii* and *M. cottewii*. Strains able to hydrolize arginine, such as *Mycoplasma arginini* and *M. alkalescens*, were also included in this preliminary step. All strains are listed in Table 1.

Table. 1 Strains used for the initial validation of the specific PCR for *Mycoplasma putrefaciens.*

Strain	Species	Origin	ArcB-PCR²
F38	Mccp	Kenya	Pos
Vienne	Mcc	France	Pos
L	Mmc	France	Pos
7302	Mmm LC	France	Pos
PG50	M.sp.Gr7	Australia	-
KS1	M.put	USA	Pos
GM623	*M. auris*	USA	Neg
GM612	*M. cottewii*	USA	Neg
GM624	*M. yeatsii*	USA	Pos
G23	*M. arginini*		Neg
PG51	*M. alkalescens*		Neg

Mccp: *Mycoplasma capricolum* subsp. *capripneumoniae*; Mcc: *Mycoplasma capricolum* subsp. *capricolum*; Mmc: *Mycoplasma mycoides* subsp. *capri*; MmmLC: *Mycoplasma mycoides* subsp. *mycoides* LC; M. sp. Gr7: *Mycoplasma* sp. Group 7 of Leach; M. put: *Mycoplasma putrefaciens.*

2.2. Target for PCR

The *Arc*B gene, coding for the ornithine transcarbamylase, was selected as a potential target for PCR. It was identified and sequenced in a parallel study concerning *M. capricolum* subsp. *capripneumoniae* (Mccp). Its complete sequence was obtained from an Mccp (F38) genomic DNA library and deposited in Genebank under the Acc Number: AY282502.

Homologous OTC protein sequences from various bacteria were retrieved from the SWISS-PROT database and used to search for conserved domains using the AlignX program from Vector NTI 8 suite (Informax). The sequences from *R. etli* (O31018), *P. aeruginosa* (P08308), *H. influenzae* (P44770), *B. licheniformis* (O86132), *S. pyogenes* (Q8P052), *Halobacterium* sp. (Q42296), *C. perfringens* (Q46169) and *M. pneumoniae* (P75473) were analysed. A first primer pair was then determined from the Mccp *Arc*B sequence so that the primers matched two conserved domains in species closely related to Mccp, *C. perfringens* and *M. pneumoniae*. Primers were also chosen as to allow a correct amplification of the targeted DNA. Their sequence is given in Table 2.

Table. 2 Oligonucleotide primers used for the search of a specific PCR identification of *Mycoplasma putrefaciens*

Name	Position	Sequence
ArcBMccpF	301-326 on AY282502	5'-ATGTATGATGCTATTGAATTTAGAGG-3'
ArcBMccpR	820-845 on AY282502	5'-TCAGTGTTTAAATCATGAAATGATGG-3'
Mput1		5'-AAATTGTTGAAAAATTAGCGCGAC-3'
Mput2		5'-CATATCATCAACTAGATTAATAGTAGCACC-3'

AY282502 is the Genebank accession number for the *Arc*B sequence of *Mycoplasma capricolum* subsp. *capripneumoniae*.

2.3. Amplification with ArcBMccp primers

The primer pair ArcBMccpF and ArcBMccpR was used to amplify DNA from the strains listed in Table 1. The PCR conditions were similar to those described by Lorenzon et al. [20]. Briefly, the template DNA consisted of mycoplasma cells (1 ml of culture) pelleted by centrifugation and lysed in 100 µl of lysis buffer (100 mM Tris–HCl pH 8.5, 0.05% Tween 20, 0.24 mg/ml proteinase K) for one hour at 60 °C. After inactivation at 95 °C for 10 min, the template consisted of 8 µl of the lysed mycoplasmas diluted one in ten in

distilled water. The amplification consisted of 30 cycles at 94 °C for 30 s, 40 °C for 30 s and 72 °C for 30 s using Taq DNA polymerase (Qiagen, France) according to the manufacturer's instructions in a GeneAmp 2400 (Perkin Elmer) DNA thermal cycler.

2.4. Sequencing

The PCR amplification products were analysed by gel electrophoresis on 1.5% (w/v) agarose gels and visualised after staining with ethidium bromide using a UV transilluminator. Amplified products yielding a unique band were then sent to Genomexpress (Meylan, France) for DNA sequence analysis.

2.5. Alignments

After correction, the DNA sequences were aligned using the AlignX software of the Vector NTI 8 program in order to search for possible polymorphisms amongst the '*M. mycoides* cluster' and closely related strains. Primers were then designed according to polymorphism observed between the sequence obtained with the KS1 strain and all the others, in order to allow specific amplification of *M. putrefaciens* DNA. Primers were denominated Mput1 and Mput2; sequences are given in Table 2.

2.6. Validation of the PCR system

The validation of the new primer pair was carried out in two steps. The first step was performed at the CIRAD by testing a wide selection of strains of various origins. Ten *M. putrefaciens* strains of French origin, as well as strains belonging to various heterologous species, MmmLC: 5 strains, Mmc: 44 strains, Mcc: 8 strains, Mccp: 5 strains were included. The PCR conditions were similar to those described for the primer pair ArcBMccp F and R, with the exception that the program consisted of 35 cycles of 94 °C for 15 s, 52 °C for 15 s and 72 °C for 15 s. The second step of the validation was performed at the University of Las Palmas, using a different subset of strains originating from the Canary Islands: *M. putrefaciens* (2 strains), MmmLC (12 strains), *M. agalactiae* (8 strains), Mmc (3 strains), Mcc (1 strain from Canary Islands and one from Murcia, mainland Spain) and *M. arginini* (5 strains). The only difference in the second validation was that Taq polymerase was obtained from Bio Line, using a Mastercycler Gradient (Eppendorf).

3. Results

3.1. PCR with primer pair ArcBMccp F and R

Most strains belonging to the '*M. mycoides* cluster' tested yielded a unique DNA fragment of the expected size after amplification with this primer pair. Among the '*M. mycoides* cluster', strains Pg1 (MmmSC) and Pg50 (M sp Gr7) did not. Other strains such as KS1 and GM624, belonging to *M. putrefaciens* and *M. yeatsii*, respectively, also yielded a PCR amplicon of similar size, whereas no amplification was obtained with the remaining strains (Fig. 1). The absence of amplification from the other strains was not due to a technical problem, since a control PCR performed with universal primers amplifying the 16S rRNA gene [21] yielded a positive result with all the samples tested (data not shown).

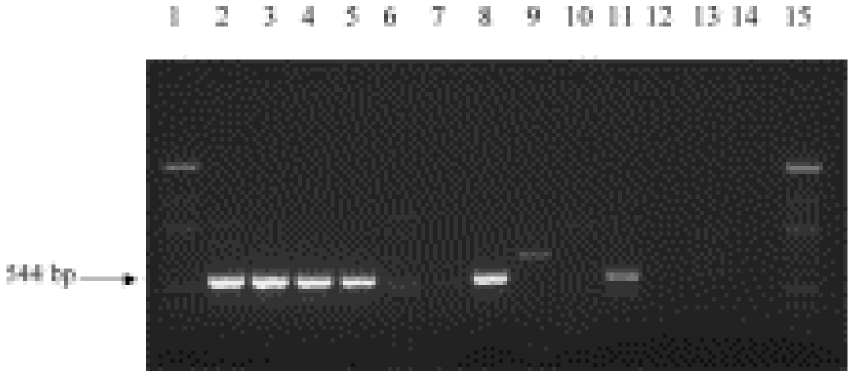


Fig 1. PCR results obtained with primer pair ArcBMccpF and ArcBMccpR with strains listed in Table 1.

Lane 1: MW, 2: Mccp, 3: Mcc, 4: Mmc, 5: MmmLC, 6: MmmSC, 7: M. sp. Gr7, 8: Mput, 9: *M. auris*, 10: *M. cottewii*, 11: *M. yeatsii*, 12: *M. arginini*, 13: *M. alkalescens*, 14: Negative control, 15: MW (Mccp: *Mycoplasma capricolum* subsp. *capripneumoniae*; Mcc: *M. capricolum* subsp. *capricolum*; Mmc: *Mycoplasma mycoides* subsp. *capri*; MmmLC: *M. mycoides* subsp. *mycoides* LC; M. sp. Gr7: *Mycoplasma* sp. Group 7 of Leach; M. put: *Mycoplasma putrefaciens*).

3.2. Alignment of *Arc*B partial sequences

Alignment of the six 523 bp-long partial *Arc*B sequences revealed an identity limited to 74.2%. The relative proximity of the different sequences is shown in Fig. 2. The closest relative of the Mccp sequence was its homologous from the subspecies *capricolum*, varying only in six bases (98.9% identity between the two). Sequences from MmmLC and Mmc were also very closely related, varying only in nine bases (98.3% identity). The *M. putrefaciens* sequence was the most distant, sharing an identity of only 86.2% with its closest relative, the sequence of *M. yeatsii*.

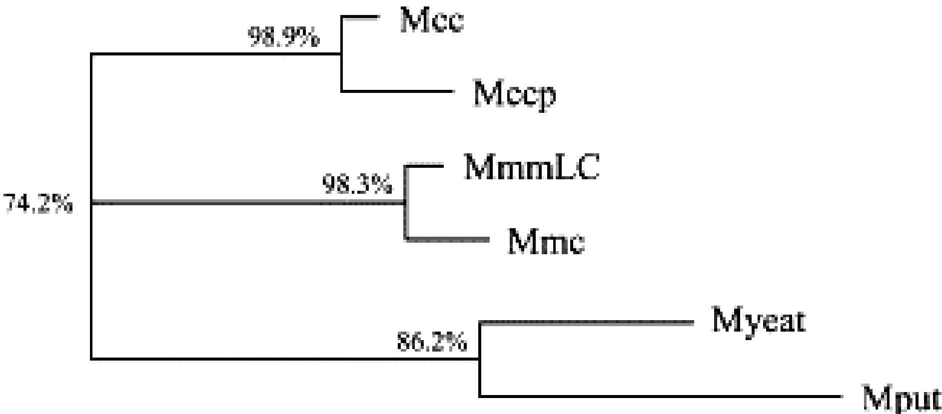


Fig 2. Multiple alignment of a 523 bp-long partial sequence from the *Arc*B gene (AlignX of VectorNTI 8). The percentage given for each fork indicates the identity shared by the sequences at the right of the forks.

3.3. Selection of specific primers for *M. putrefaciens*

The regions showing greatest polymorphism were selected for the design of specific primers. Primer Mput1 conferred the specificity to the PCR, bearing six sites strictly specific to the *M. putrefaciens* sequence. It is noteworthy that four of these sites were found at the 3′ end of the primer, allowing specific amplification (Fig. 3(a)).

(a)

```
Mcc[1]  : GTGTTGTTGAAGATCTTGCAAAAT
Mccp    : GTGTTGTTGAAGATCTTGCAAAAT
Mmc     : CTGTTGTTGAAGAGCTAGCAAAAT
MmmLC   : CTGTTGTTGAAGAGTTAGCAAAAT
Myeat   : AAGTTGCTGAAGAATTAGCAAAAT
M.put   : AAATTGTTGAAAAATTAGCGCGAC
```

(b)

```
Mcc     : CATGTCTTCAACAAATGAAATTGTAGCATT
Mccp    : CATGTCTTCAACAAATGAAATTGTAGCATT
Mmc     : CATATCTTCAGTGAATGAAATAATAGCATT
MmmLC   : CATATCTTCAGTAAATGAAATAATAGCATT
Myeat   : CATATCTTCAACTAATGTAATTGTTGCACC
Mput    : CATATCATCAACTAGATTAATAGTAGCACC
```

Fig 3. Alignment of the sequences homologous to primers Mput1 (a) and Mput2 (b). Specific nucleotides to *M. putrefaciens* are underlined. Sequences specific to *M. putrefaciens* and *M. yeatsii* are highlighted.

The sequence from primer Mput2 was partially shared by *M. putrefaciens* and *M. yeatsii*, but two bases at the 3′ end formed a CC clamp that differed from the sequences of the other species of the '*M. mycoides* cluster' (Fig. 3(b)).

3.4. Validation of the *M. putrefaciens* specific PCR

All *M. putrefaciens* strains tested were identified by this PCR producing products of the expected size, 316 bp (Fig. 4). In contrast, no amplification was obtained from any of the heterologous strains tested.

Fig 4. PCR results obtained with primer pair Mput1 and Mput2 with strains listed in Table 1

Lane 1: MW, 2: Mccp, 3: Mcc, 4: Mmc, 5: MmmLC, 6: MmmSC, 7: M sp. Gr 7, 8: Mput, 9: *M. auris*, 10: *M. cottewii*, 11: *M. yeatsii*, 12: MW (Mccp: *Mycoplasma capricolum* subsp. *capripneumoniae*; Mcc: *M. capricolum* subsp. *capricolum*; Mmc: *Mycoplasma mycoides* subsp. *capri*; MmmLC: *M. mycoides* subsp. *mycoides* LC; M. sp. Gr7: *Mycoplasma* sp. Group 7 of Leach; M. put: *Mycoplasma putrefaciens*).

4. Discussion and conclusion

The arginine deiminase pathway exists in many bacteria, including mycoplasmas, and analogous genes have already been identified in the *M. pneumoniae* genome [22]. This pathway may also be active in various other mycoplasma species, since many of them are able to hydrolize arginine, which constitutes the basis for one of the few biochemical tests available for the differentiation of mycoplasma species. Amongst the '*M. mycoides* cluster', only *M. capricolum* seems to possess an effective ADI pathway [3], although the reaction takes a long time (2–3 weeks), compared to some other species that yield a positive result within one day. Eventually, the anaerobic nature of this pathway in Mcc is evidenced by the change of colour, which appears first at the bottom of the tube. The *Arc*B sequences, coding for OTC, may also be present in mycoplasmas that are not positive by the arginine hydrolysis test, as it was found when studying the genome of Mccp, the agent of contagious caprine pleuropneumonia. *Arc*B is also present in other members of the '*M. mycoides* cluster' as shown by the PCR results with the non-specific primer pair ArcBMccp F and R. This gene may also be present in those species yielding a negative result by this PCR, given that negative results may be due to DNA polymorphism rather

than absence of the gene. Testing this hypothesis was outside the scope of this study. Comparative studies of the ADI operon within the '*M. mycoides* cluster' may eventually lead to potentially interesting results for designing new specific diagnostic tests or for evaluating phylogenetic relationships.

Alignment of the partial *Arc*B sequences obtained from six different strains of the '*M. mycoides* cluster' and related species showed the relatively high polymorphism existing within these sequences, which shared an identity of only 74.2 %. This contrasted markedly with comparisons made between the 16S rRNA sequences, for which the polymorphism was very limited, and between the 16S–23S rRNA intergenic spacer regions [6], where polymorphism was also less important. Nevertheless, the comparison of these three regions gave consistent results, showing that *M. putrefaciens* is related to the '*M. mycoides* cluster', although obviously distinct from it. The results obtained from the *Arc*B sequence comparisons also confirmed that *M. yeatsii* is a very close relative of *M. putrefaciens* [4]. Interestingly, *M. yeatsii* has rarely been isolated and so far only from the external ear canal of goats [23 and 24]. Its pathogenicity for goats and especially its ability to induce mastitis may be worth investigating in view of its genetic proximity to *M. putrefaciens*.

Up to now the identification of *M. putrefaciens* required the isolation of the agent, its purification and identification by conventional methods, such as biochemical analysis or growth inhibition test. The only PCR technique available for the identification of *M. putrefaciens* was derived from the CAP-21 probe [25]. However, this technique was based on a non-specific amplification, followed by enzymatic digestion of the amplified product, a technique that may pose difficulties for diagnostic laboratories. Furthermore, this technique should be re-evaluated in view of the close relationship existing between *M. putrefaciens*, *M. cottewii* and *M. yeatsii*.

The new PCR technique described here, using primers Mput1 and Mput2, offers the advantage of being strictly specific for *M. putrefaciens* and, if needed, the identity of the amplified product may be confirmed by enzymatic digestion or by sequencing. All heterologous mycoplasma strains tested yielded only negative results, although they were chosen mostly amongst the closely related '*M. mycoides* cluster' cluster. This PCR was also universal, since it was able to yield a positive result from all the *M. putrefaciens* strains tested, in spite of their various origins: California, France, Canary Islands. This test

offers new opportunities for epidemiological studies of the ‘contagious agalactia syndrome’, which may be caused by various mycoplasma species including *M. putrefaciens* [1].

Acknowledgements

The help of Lucia Manso-Silvan for reading and correcting the manuscript was greatly appreciated.

References

[1] P.C. Lefèvre. *Contagious agalactia In: Manual of standards for diagnostic tests and vaccines, OIE 12 rue de Prony, 75017 Par*is*, France* (2000) p. 490–6; chapter 2.4.3.

[2] J.G. Tully, M.F. Barile, D.G. Edward, T.S. Theodore and H. Erno, Characterization of some caprine mycoplasmas, with proposals for new species, mycoplasma *capri*colum and *Mycoplasma putrefaciens*. *J Gen Microbiol* **85** (1974), pp. 102–120.

[3] G.S. Cottew, A. Breard, A.J. DaMassa, H. Erno, R.H. Leach, P.C. Lefevre, A.W. Rodwell and G.R. Smith, Taxonomy of the *Mycoplasma mycoides* cluster. Is*rael J Med Sci* **23** (1987), pp. 632–635.

[4] M. Heldtander, B. Pettersson, J.G. Tully and K.E. Johansson, Sequences of the 16S rRNA genes and phylogeny of the goat mycoplasmas *Mycoplasma adleri*, *Mycoplasma auris*, *Mycoplasma cottewii* and *Mycoplasma yeatsii*. *Int J Syst Bacteriol* **48** (1998), pp. 263–268.

[5] A.J. DaMassa, J.G. Tully, D.L. Rose, D. Pitcher, R.H. Leach and G.S. Cottew, *Mycoplasma auris* sp. nov, *Mycoplasma cottewii* sp. nov., and *Mycoplasma yeatsii* sp. nov., new sterol-requiring mollicutes from the external ear canals of goats. *Int J Syst Bacteriol* **44** (1994), pp. 479–484.

[6] R. Harasawa, H. Hotzel and K. Sachse, Comparison of the 16S–23S rRNA intergenic spacer regions among strains of the *Mycoplasma mycoides* cluster, and reassessment of the taxonomic position of *Mycoplasma* sp. bovine group 7. *Int J Syst Evol Microbiol* **50** (2000), pp. 1325–1329.

[7] J. Frey, X. Cheng, M.P. Monnerat, E.M. Abdo, M. Krawinkler, G. Bölske and J. Nicolet, Genetic and serological analysis of the immunogenic 67-kDa lipoprotein of *Mycoplasma* sp. bovine group 7. *Res Microbiol* **149** (1998), pp. 55–64.

[8] Erno: personal communication cited by Cottew GS, Caprine–ovine mycoplasmas. In: , J.G. Tully and R.S. Whitcomb, Editors, *The mycoplasmas Human and animal mycoplasmas* **vol. II**, Academic Press, New York (1979), p. 125.

[9] H.E. Adler, A.J. DaMassa and D.L. Brooks, Caprine mycoplasmosis: *Mycoplasma putrefaciens*, a new cause of mastitis in goats. *Am J Vet Res* **41** (1980), pp. 1677–1679.

[10] G. Gaillard-Perrin, D.P. Picavet and G. Perrin, Isolement de *Mycoplasma putrefaciens* dans deux troupeaux de chèvres présentant des symptômes d'agalactie. *Revue de Médecine Vétérinaire* **137** (1985), pp. 67–70.

[11]. A.J. DaMasa, D.L. Brooks, C.A. Holmberg and A.I. Moe, Caprine mycoplasmosis: an outbreak of mastitis and arthritis requiring destruction of 700 goats. *Vet Rec* **120** (1987), pp. 409–413.

[12] J.L. Rodriguez, J.B. Poveda, C. Gutierrez, B. Acosta and A. Fernandez, Polyarthritis in kids associated with *Mycoplasma putrefaciens*. *Vet Rec* **135** (1994), pp. 406–407.

[13] G.S. Cottew and F.R. Yeats, Occurrence of mycoplasmas in clinically normal goats. *Aust Vet J* **57** (1981), pp. 52–53.

[14] F. Thiaucourt and G. Bölske, Contagious caprine pleuropneumonia and other pulmonary mycoplasmoses of sheep and goats. *Revue Scientifique et technique de l' Office international des Epizooties* **15** (1996), pp. 1397–1414.

[15] D.L. Brooks, A.J. DaMasa and H.E. Adler, Caprine mycoplasmosis: immune response in goats to *Mycoplasma putrefaciens* after intramammary inoculation. *Am J Vet Res* **42** (1981), pp. 1898–1900.

[16] A.J. DaMasa, P.S. Wackenell and D.L. Brooks, Mycoplasmas of goats and sheep. *J Vet Diag Invest* **4** (1992), pp. 101–113.

[17] D. Bergonier and F. Poumarat, Agalactie contagieuse des petits ruminants: epidémiologie, diagnostic et contrôle. *Revue Scientifique et technique de l'Office international des Epizooties* **15** (1996), pp. 1431–1475.

[18] J. Oros, A. Fernandez, J.L. Rodriguez, F. Rodriguez and J.B. Poveda, Bacteria associated with enzootic pneumonia in goats. *Zentralblat für Veterinärmedizin B.* **44** (1997), pp. 99–104.

[19] M. Zuniga, G. Perez and F. Gonzalez-Candelas, Evolution of arginine deiminase (ADI) pathway genes. *Mol Phylogenet Evol* **25** (2002), pp. 429–444.

[20] S. Lorenzon, A. David, M. Nadew, H. Wesonga and F. Thiaucourt, Specific PCR identification of the T1 vaccine strains for contagious bovine pleuropneumonia. *Mol Cell Probes* **14** (2000), p. 20510.

[21] B. Pettersson, T. Leitner, M. Ronaghi, G. Bölske, M. Uhlén and K.E. Johansson, Phylogeny of the *Mycoplasma mycoides* cluster as determined by sequence analysis of the 16S rRNA genes from the two rRNA operons. *J Bacteriol* **178** (1996), pp. 4131–4142.

[22] R. Himmelreich, H. Hilbert, H. Plagens, E. Pirkl, B.C. Li and R. Herrmann, Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucleic Acids Res* **24** (1996), pp. 4420–4449.

[23] G.S. Cottew and F.R. Yeats, Mycoplasmas and mites in the ears of clinically normal goats. *Aust Vet J* **59** (1982), pp. 77–81.

[24] A.J. DaMasa and D.L. Brooks, The external ear canal of goats and other animals as a mycoplasma habitat. *Small Ruminant Res* **4** (1991), pp. 85–93.

[25] J.L. Rodriguez, R.W. Ermel, T.P. Kenny, D.L. Brooks and A.J. DaMassa, Polymerase chain reaction and restriction endonuclease digestion for selected members of the *Mycoplasma mycoides* cluster and *Mycoplasma putrefaciens*. *J Vet Diag Invest* **9** (1997), pp. 186–190.

Paper III

Veterinary Microbiology
Volume 104, Issues 1-2, 30 November 2004, Pages 125-132
View Record in ScopusCited By in Scopus (2)
Copyright © 2004 Elsevier B.V. All rights reserved.

Short communication

A specific PCR for the identification of *Mycoplasma capricolum* subsp. *capripneumoniae*, the causative agent of contagious caprine pleuropneumonia (CCPP)

S. Woubit[a, b], S. Lorenzon[a], A. Peyraud[a], L. Manso-Silván[a] and F. Thiaucourt[a,]

[a]CIRAD EMVT Sante Animale, Animal Health Program, TA30/G Campus International de Baillarguet, 34398 Montpellier Cedex 5, France
[b]Faculty of Veterinary Medicine, Addis Ababa University, PO Box 34, Debre Zeit, Ethiopia
Received 2 April 2004; revised 11 August 2004; accepted 19 August 2004. Available online 2 November 2004.

Abstract

Contagious caprine pleuropneumonia is a severe infectious disease of goats in Africa and the Middle East. It is caused by a fastidious mycoplasma, *Mycoplasma capricolum* subsp. *capripneumoniae*, a member of the "*M. mycoides* cluster". Members of this cluster share genomic and antigenic features, which result in common biochemical and serological properties, complicating species identification. Two species of this cluster, *M. mycoides* subsp. *capri* and *M. mycoides* subsp. *mycoides* large colony biotype, are very often isolated from clinical cases resembling contagious caprine pleuropneumonia. Furthermore, in the laboratory, *M. capricolum* subsp. *capripneumoniae* can be easily confused with the closely related *capricolum* subspecies. Considering these constraints and the scarcity of available methods for identification, a specific polymerase chain reaction was developed. A DNA fragment of 7109 bp containing genes coding for the arginine deiminase pathway (ADI) was chosen as target sequence for the selection of a specific primer pair. The full ADI operon from *M. capricolum* subsp. *capripneumoniae* strain GL100 was sequenced. Polymorphism within this locus was analyzed by comparison with the sequence from the closely related IPX strain (*M. capricolum* subsp. *capricolum*). It varied from 0.6 % to 3.5 %. The highest divergence was found in a region coding for *arcD*. Therefore, this gene

was chosen as target for the specific amplification of a 316 bp-long DNA fragment. The specificity of this PCR was validated on 14 *M. capricolum* subsp. *capripneumoniae* strains and 27 heterologous strains belonging to the "*M. mycoides* cluster" and *M. putrefaciens*. This new PCR will be a valuable tool for the surveillance of contagious caprine pleuropneumonia.

Keywords: Contagious caprine pleuropneumonia; Specific PCR; *Mycoplasma capricolum* subsp. *capripneumoniae*; ADI pathway

1. Introduction

Contagious caprine pleuropneumonia (CCPP) is a severe infectious disease of goats caused by *Mycoplasma capricolum* subsp. *capripneumoniae* (Mccp), which occurs in many countries of Africa and Asia where the total goat population is over 500 million (Rurangirwa and Kinyili, 2000). It is a disease of major economic relevance characterized by high morbidity and mortality. It represents a potential threat to many disease-free countries. As a result, it is classified as a List B disease by the OIE.

The exact distribution of CCPP is not known and there are very few official confirmations of outbreaks. The first reason is that, from a clinical point of view, CCPP can be confused with a number of diseases inducing similar respiratory signs in goats, such as Peste des Petits Ruminants or pasteurellosis. Also, amongst mycoplasma species that induce various syndromes: Mastitis, Arthritis, Keratitis, Pneumonia and Septicaemia (MAKePS) (Thiaucourt and Bolske, 1996), some peculiar strains may have a specific tropism for the lung. The second reason is that Mccp is one of the most fastidious mycoplasmas to be grown in vitro. As a result, isolation trials are often unsuccessful, especially if the conservation of the clinical sample has not been adequate. In addition, other mycoplasma species, such as *M. ovipneumoniae*, may be isolated from CCPP cases, although they are in small number in the sample, simply because they grow faster and more easily. Once isolated, Mccp strains may also be difficult to identify as this subspecies belongs to the so-called "*M. mycoides* cluster", which includes six species, subspecies or groups of strains: *M. mycoides* subsp. *mycoides* LC (MmmLC), *M. mycoides* subsp. *capri* (Mmc), *M. mycoides* subsp. *mycoides* SC (MmmSC), *M. capricolum* subsp. *capricolum* (Mcc)

M. capricolum subsp. *capripneumoniae* (Mccp) and *Mycoplasma* sp. bovine group 7 (Mbg7), sharing many genomic as well as antigenic properties (Cottew et al., 1987). For example, cross-reactions may occur in many serological typing methods, such as the growth inhibition test, with strains belonging to Mbg7 (ter Laak, 1991). Cross reactions are also very often observed between these two species when looking for specific monoclonal antibodies (Thiaucourt et al., 1994 and Rurangirwa et al., 1997). Because of these difficulties, the direct detection of Mccp in clinical material may be a very useful alternative for the confirmation of CCPP outbreaks. This detection can be achieved by immunoperoxydase techniques from pleural fluid blotted on nitrocellulose using hyperimmune sera (Thiaucourt et al., 1992) or monoclonal antibodies. However, these techniques have the disadvantage of a relatively low sensitivity. A higher sensitivity can be achieved with PCR and, up to now, there are only two PCR-based detection techniques described for the identification of Mccp. Both assays are based on an initial "*M. mycoides* cluster"-specific PCR amplification of a fragment of the 16S rRNA (Bascuñana et al., 1994) or the CAP21 locus (Hotzel et al., 1996). The specificity for Mccp identification is then obtained by *Pst*I cleavage, detecting a point mutation in one of the two rRNA operons for Mccp (Bascuñana et al., 1994), or by a nested PCR using a second set of primers (Hotzel et al., 1996). The first PCR assay has been extensively used and validated, although it has some possible drawbacks. Firstly, it may not be able to detect Mccp when it is mixed with other members of the "*M. mycoides* cluster" as the first PCR is not Mccp specific. Secondly, the use of restriction endonuclease analysis requires additional time and may lead to false positive results in the case of a partial digestion of the product, as the presence of an undigested rRNA fragment is theoretically typical for Mccp. Thirdly; the existence of Mccp strains that do not possess the specific single nucleotide mutation is always to be feared. As for the second PCR assay, it has not really been validated for Mccp identification, as the original paper dealt mainly with the distinction between MmmSC and its close relatives. The aim of this work was, therefore, to develop a new PCR assay that would not have the aforementioned drawbacks and would allow a one-step specific detection of Mccp strains.

2. Materials and methods

Representative strains from each member of the "*M. mycoides* cluster" were selected according to their geographical distribution. All the strains used in this study, with reference to their origin, are listed in Table 1. These strains were cultivated in modified Hayflick medium (PPLO broth without crystal violet (21 g/l), 20 % de-complemented horse serum, 10 % fresh yeast extract, 0.2 % glucose, 0.4 % sodium pyruvate) in a high security (L_3) laboratory.

Table 1 List of mycoplasma strains used in the study.

CIRAD ref. no. (original ID)	Origin	Species	Received from
8789	Chad	Mccp	
8891 (7/2)	Oman	Mccp	Dr. Jones, MRI, Edinburgh, GB
9081 (487)	Oman	Mccp	CVRL, Rumais
91106 (C550-1)	Dubaï	Mccp	
92138 (CLP)	Ethiopia	Mccp	Dr. Laikemariam, NVI, Debre Zeit
9231 (Abomsa)	Ethiopia	Mccp	Dr. Laikemariam, NVI, Debre Zeit
94156 (438p)	Chad	Mccp	Dr. Angaya, N'Djamena
95043	Niger	Mccp	Dr. Maikano LABOCEL, Niamey
98113 (M 74/93)	Uganda	Mccp	Dr. Bölske NVI Uppsala, Sweden
99108 (P1)	Eritrea	Mccp	Dr. Tesfaalem T, Asmara
F38	Kenya	Mccp	
Gabes 5p	Tunisia	Mccp	
Gabes 96p	Tunisia	Mccp	
GL100	Tunisia	Mccp	
7714	France	Mcc	
7759 (585)	France	Mcc	
8069	France	Mcc	
8086	France	Mcc	
8086-1	France	Mcc	
8110	France	Mcc	
8601-2-03	Portugal	Mcc	Dr. Regalla, Lisbon
8601-50	Portugal	Mcc	Dr. Regalla, Lisbon
9125 (E-570)	UK	Mcc	Dr. Bölske NVI Uppsala, Sweden
970058	Ethiopia	Mcc	Dr. Laikemariam, NVI, Debre Zeit

C-Kid	USA	Mcc	
Courtenay	France	Mcc	
G189	France	Mcc	
IPX	France	Mcc	
7730	France	MmmLC	
8065	France	MmmLC	
9298-CN508	Tanzania	MmmLC	
99055	France	MmmLC	
YG	Australia	MmmLC	
9139	Turkey	Mmc	
PG-3	Turkey	Mmc	
C11	Chad	MmmSC	
PG-1	Not Known	MmmSC	
KS1	USA	*M. putrefaciens*	
Tours 2	France	*M. putrefaciens*	
PG-50	Australia	Mbg7	
97012 (Eldoret)	Kenya	*M. ovipneumoniae*	Dr. Wesonga, KARI, Muguga

Abbreviations: *M. capricolum* subsp. *capripneumoniae* (Mccp), *M. capricolum* subsp. *capricolum* (Mcc), *Mycoplasma mycoides* subsp. *mycoides* LC (MmmLC), *M. mycoides* subsp. *mycoides* SC (MmmSC), *M. mycoides* subsp. *capri* (Mmc), *Mycoplasma* sp. bovine group 7 (Mbg7).

Samples for PCR were prepared as described by Miserez et al. (1997): 3 ml culture were centrifuged at 12,000 × *g* for 20 min at 4 °C. The pellets were then washed once in PBS and re-suspended in 300 µl of lysis buffer (100 mM Tris/HCl pH 8.5, 0.05% Tween 20, and 0.24 mg/ml proteinase K). Following incubation for 1 h at 60 °C, proteinase K was inactivated at 95 °C for 10 °min. The heat treatment ensured that no live mycoplasmas remained in the samples, which could then be examined in a conventional L_2 laboratory. For PCR reactions, these samples were diluted 1/10 in sterile distilled water.

For the selection of target genes, genomic libraries were created in pBluescript (Stratagene, CA, USA), using Mccp DNA (strain GL100) cleaved with *Cla*I, *Eco*RI, *Hin*dIII and *Pst*I, respectively. The inserts were amplified and dig-labelled by PCR, and then probed by hybridization with mycoplasma DNA from various strains belonging to the "*M. mycoides* cluster". Inserts that yielded a positive signal limited to the *capricolum* species were retained and sequenced (GENOME express, Meylan, France). Two of these sequences were assembled (Vector NTI "Contig Express", Informax, USA) and formed a 3915 bp long contig. This sequence was analysed by pBLAST (http://www.ncbi.nlm.nih.gov) and also compared with the whole genome sequence of strain California Kid (C. kid), the Mcc reference strain (courtesy of Mike Calcutt, University of Missouri, USA). This allowed the identification of the flanking genes of the initial contig and the subsequent sequencing of a 7109 bp long DNA fragment for comparison between Mccp (strain GL100) and Mcc: strain C. kid for the totality of the fragment and strain IPX for partial sequences.

PCR was carried out in a Perkin-Elmer GeneAmp 2400 PCR system. The specific amplification was performed in a 50 µl-final volume obtained by mixing 34 µl of distilled water, 0.5 µl of dNTP (150 µM for dCTP and dGTP, 300 µM for dATP and dTTP), 3 µl of $MgCl_2$ (1.5 mM), 5 µl of 10× Taq Buffer (Qiagen), 1 µl of each primer (0.4 µM of Mccp-spe-F: 5′-ATCATTTTTAATCCCTTCAAG-3′ and Mccp-spe-R, 5’-TACTATGAGTAATTATAATATATGCAA-3′), 0.5 µl of Taq polymerase (1 unit, Qiagen) and 5 µl of the sample. PCR conditions consisted of an initial denaturation step of 2 min at 94 °C, followed by 35 cycles of 30 s at 94 °C, 15 s at 47 °C and 15 s at 72 °C and a final extension step of 5 min at 72 °C.

This PCR was validated by testing all strains listed in Table 1. Specific Mccp PCR products were analyzed without further purification using *Cla*I according to the manufacturer's instructions (Promega). In addition, all mycoplasma samples were further analysed by a non-specific PCR assay that amplifies a DNA fragment from all mycoplasma cultures using primers chosen on the 16S rRNA genes (van Kuppeveld et al., 1994).

3. Results

The 7109 bp-long sequence, obtained from strain GL100, was deposited at GenBank (accession number AY529462).

The pBLAST analysis on all the ORFs identified along this sequence allowed the identification of a putative arginine deiminase (ADI) pathway in the *capricolum* species. It is composed of the putative *arcA*, *arcB* and *arcD* genes that share maximum homologies with genes of *Clostridium perfringens*, *Vibrio cholerae* and *Haemophilus influenzae* respectively. This ADI pathway is flanked upstream by a putative Mg^{2+} transport ATPase and downstream by a putative glucokinase. Interestingly, *arcA* and *arcD* are apparently disrupted by frameshift mutations in the GL100 sequence but neither in the C. kid nor in the IPX sequences (Fig. 1).

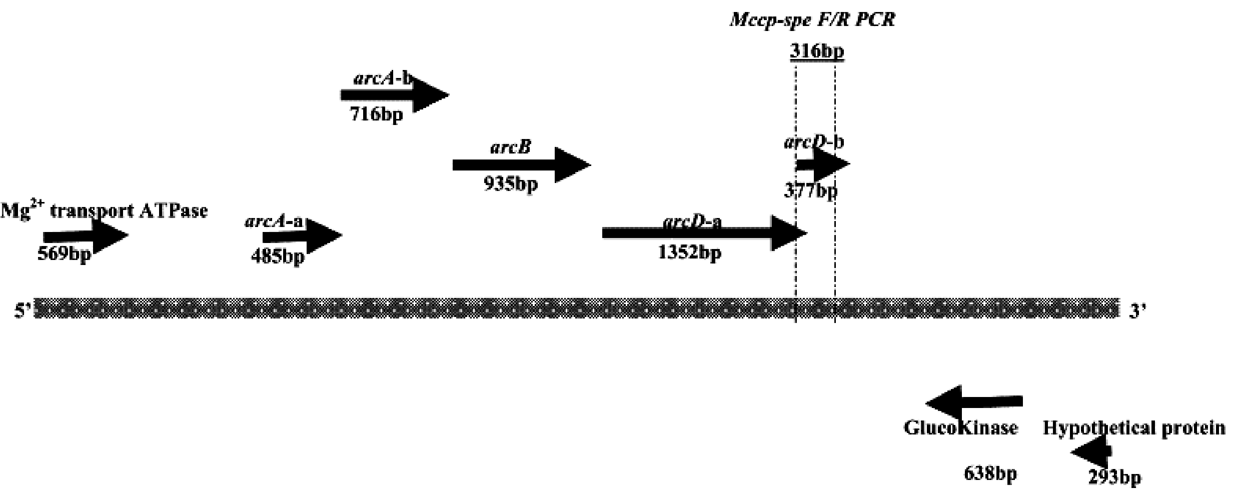


Fig. 1 Diagram showing the organization of the ADI operon in M. capricolum subsp. capripneumoniae strain GL100.

The operon is flanked by genes coding for Mg^{2+} transport ATPase and Glucokinase. Genes *arcA* and *arcD* present frameshift mutations leading to two ORFs for each of them. The vertical broken line indicates the target site used for the selection of the specific primer pair Mccp-spec-F/R.

Alignments of the DNA sequences from GL100 and C. kid strains allowed the detection of some polymorphisms consisting in a deletion in the C. kid sequence upstream *arcA*, as well as a 9 bp-long deletion in the GL100 sequence at the 3′ end of the *arc*D putative gene. In

addition, scattered mutations occurred along the whole sequence, with frequencies varying from 0.6% in *arcB* to 3.4% in *arcD*.

The partial sequences obtained with strain IPX suggested that the 9 bp deletion was possibly specific to Mccp strains and, accordingly, a primer pair was chosen on the GL100 sequence so that the 3′ end of the primers matched polymorphic sites. In particular, primer Mccp-spe-R was chosen so that hybridization to template DNA from strain C. kid yielded a 4 bp-long mismatch at its 3′ end because of the deletion existing in the GL100 sequence.

All tested samples yielded a positive amplification with the non-specific PCR (van Kuppeveld et al., 1994) (data not shown), whereas only Mccp strains yielded a positive specific amplification by PCR using the Mccp-spe-F/R primer pair (Fig. 2). The amplified product was of the expected size (316 bp) and digestion by *Cla*I also yielded the two expected bands (190 and 126 bp long, respectively (Fig. 3). This digestion results confirmed the identity of the amplified product. Among the other strains tested, only strain YG (MmmLC) yielded a faint amplification of a smaller fragment (Fig. 2).

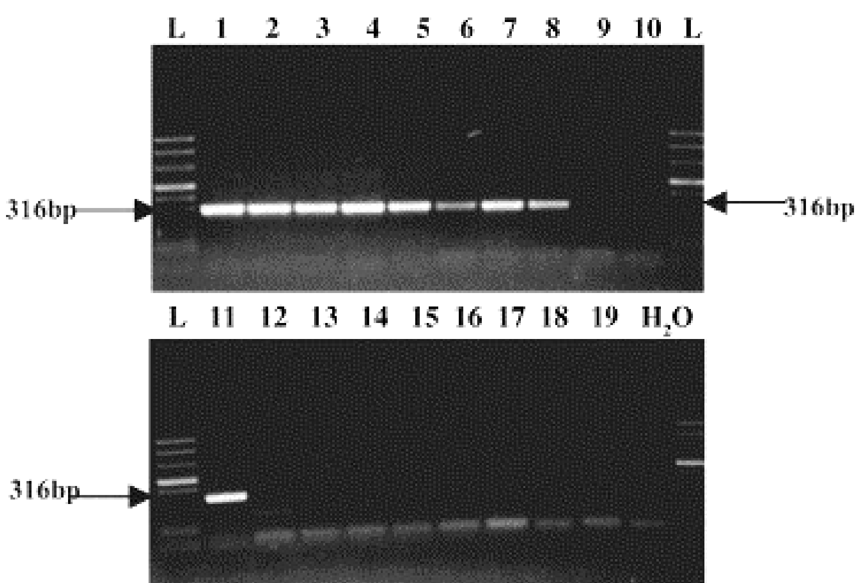


Fig. 2 PCR results obtained with the primer pair Mccp-spec-F/R.

A specific 316 bp-long band was only obtained with *M. capricolum* subsp. *capripneumoniae* (Mccp) strains. Lane L: molecular weight marker; Mccp strains in lane 1: 95043, lane 2: 8789, lane 3: 8891, lane 4: 98113, lane 5: F38, lane 6: 92138, lane 7: 94156-438p, lane 8: 91106-C550-1, lanes 9 and 10: *M. capricolum* subsp. *capricolum* strains IPX and C. kid, respectively, lane 11:

GL100, lanes 12 and 13: M. *mycoides* subsp. *mycoides* LC type YG and 8065 CN 508, lanes 14 and 15: *M. mycoides* subsp. *capri* strains PG-3 and 9139, lane 16: *M. mycoides* subsp. *mycoides* sc type PG-1, lanes 17 and 18: *M. putrefaciens* strains KS1 and Tours, lane 19: Mbg 7 strain PG-50 and lane 20: H_2O.

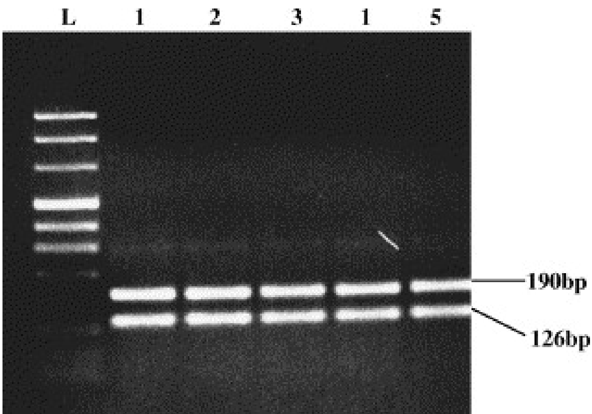


Fig. 3 Agarose gel electrophoresis of *ClaI*-cleaved specific PCR products from *M. capricolum* subsp. *capripneumoniae* strains.

From left to right lane L: molecular weight marker, lane 1: 95043, lane 2: 8789, lane 3: 8891, lane 4: 98113 and lane 5: F38.

4. Discussion and conclusions

The ADI pathway is a route for the anaerobic catabolism of arginine (D'hooghe et al., 1997) that is found in many bacteria (Barcelona-Andrés et al., 2002). The genes for the three pathway enzymes *arcA*, *arcB* and *arcC*, code for arginine deiminase (ADI), ornithine transcarbamylase (OTC) and carbamate kinase (CK), respectively. These genes are usually clustered in an operon that may also include, *arcD*, which codes for a membrane-bound protein and *arcR,* which codes for a repressor (D'hooghe et al., 1997). However, the organization of the operon and the number of genes present vary widely from one species to another (Zúñiga et al., 2002). In the case of Mccp, both *arcR* and *arcC* are missing in the DNA fragment sequenced in this study. However, other mycoplasma species do not display the same organization for this operon. *ArcD* is not found in any of the six completely sequenced mycoplasma genomes: *M. genitalium*, *M. pneumoniae*, *M. pulmonis*,

M. penetrans and *M. gallisepticum* (http://cbi.labri.fr/outils/molligen). In *M. penetrans* the ADI operon is organized as *arcA–arcB–arcC*–AApermease and in *M. pneumoniae*, the sequence of genes is *arcA–argI–arcC*. Furthermore, only *arcA* is found in the *M. gallisepticum* genome.

In Mccp, two genes of the ADI pathway showed frame-shift mutations, a deletion of a T at position 1940 bp (*arcA*) and an insertion of an A at position 4963 bp (*arcD*), which should prevent the correct translation of these genes. This may account for the absence of arginine catabolism in Mccp strains, observed by arginine hydrolysis testing (Bonnet et al., 1993).

The *ArcB* gene has been used previously to design a specific PCR for the identification of *M. putrefaciens* (Peyraud et al., 2003), taking advantage of the high polymorphism observed in this sequence between "*M. mycoides* cluster" and related species. The DNA polymorphism observed on this gene between Mccp and the closest relative Mcc was very limited (0.6%) and did not allow the selection of specific PCR primers. This result also confirmed the very close similarity between these two subspecies. A greater polymorphism was evidenced on the *arcD* gene and the deletion observed on this gene in Mccp strain GL100 allowed the selection of a specific primer pair.

Great care has been taken for the choice of strains included in the validation. For universality, 14 Mccp strains of various geographical origins were included. This ensured that strains belonging to all clusters, defined in a previous molecular epidemiology study (Lorenzon et al., 2002), were included. In this previous study, one strain was found to be very peculiar (91106-C550-1) as one pseudo-gene was completely deleted. This strain was correctly identified with the new PCR technique.

The reliability of the validation of PCR specificity was ensured by the inclusion of a representative number (27) of strains from the "*M. mycoides* cluster", with a greater emphasis on Mcc strains, the closest relatives to Mccp.

This new PCR will allow a more reliable identification of Mccp strains and, hopefully, a better assessment of the distribution and economical impact of CCPP, as the isolation of Mccp is very cumbersome. Up to now, a single validated PCR system was available (Bascuñana et al., 1994). It has already been very useful, notably because PCR can be performed on dried samples. This is a major advantage for developing countries that lack

proper infrastructures to send samples rapidly and under a cold chain. However, a single PCR technique might not be sufficient, as it does not allow any confirmation based on different genes. Such a confirmation is necessary as PCR may yield false positive results due to contaminations and also because 16S rRNA sequences in Mccp strains have been shown to be quite polymorphic. The major advantage of the new PCR technique is that it gives a specific amplification that should allow the detection of Mccp strains even in mixed cultures or in samples containing multiple mycoplasma species. In goat lung samples, the presence of multiple mycoplasma species is a common feature and, sometimes, multiple species of the "*M. mycoides* cluster" can be isolated. After the first in vitro passage, any contaminant may become the major population in the culture and only a sensitive technique such as PCR may be able to detect it. In addition, the advantage of a specific amplification is that the identity of the amplified product can then be verified by *Cla*I digestion or by sequencing. Getting PCR positive results in some samples will also stimulate new trials to isolate Mccp strains, especially in countries that have not yet confirmed the presence of CCPP. This could be done by harvesting new samples, such as fresh pleural fluid from an acute case, or by cloning mycoplasma cultures that may contain multiple species. Detecting CCPP by PCR might be the only way to get a better distribution map for this disease, which remains largely unrecognized in Africa and Asia, in spite of many suspicious clinical cases. In disease-free countries the detection of Mccp by a direct specific amplification method will be of great help in case of an emergency. At present, it is unlikely that CCPP would be easily recognized if imported in disease-free countries. Goats may receive antibiotic treatments that will hamper bacteriological isolations and the growth of Mccp is so fastidious that it is unlikely that a laboratory will succeed in isolating it using routine procedures. Hence, this new PCR should play a major role in the surveillance of CCPP as a complementary tool to clinical and post-mortem inspections and to mycoplasma isolation trials. The PCR from Bascuñana based on 16S rRNA gene amplification can be used in parallel in order to get a confirmation of identification.

Acknowledgements

We are grateful to the French Ministry of Foreign Affairs and the French Embassy in Addis Ababa that have financed Woubit Salah's stay at CIRAD-EMVT Animal Health

Program. We are also extremely grateful to M. Calcutt, who kindly provided us with the sequence of the ADI operon and the flanking genes for strain C. kid (Mcc).

References

B. Barcelona-Andrés, M. Alberto and R. Vicente, Gene structure, organization, expression, and potential regulatory mechanisms of the arginine catabolism in *Enterococcus faecalis*, *J. Bacteriol.* **184** (2002), pp. 6289–6300.

C.R. Bascuñana, J.G. Mattsson, G. Bolske and K.E. Johansson, Characterization of the 16S rRNA genes from *Mycoplasma* sp. strain F38 and development of an identification system based on PCR, *J. Bacteriol.* **176** (1994), pp. 2577–2586.

F. Bonnet, C. Saillard, J.M. Bove, R.H. Leach, D.L. Rose, G.S. Cottew and J.G. Tully, DNA relatedness between field isolated of *Mycoplasma capricolum*, *Int. J. Syst. Bacteriol.* **43** (1993), pp. 597–602.

G.S. Cottew, A. Breard, A.J. Damassa, H. Ernø, R.H. Leach, P.C. Lefèvre, A.W. Rodwell and G.R. Smith, Taxonomy of the *Mycoplasma mycoides* cluster, IS*r. J. Med. Sci.* **23** (1987), pp. 632–635.

I. D'hooghe, C.V. Wauven, J. Michiels, C. Tricot, D.W. Petra, J. Vanderleyden and V. Stalon, The arginine deiminase pathway in *Rhizobium etli*: DNA sequence analysis and functional study of the arcABC Genes, *J. Bacteriol.* **179** (1997), pp. 7403–7409.

H. Hotzel, K. Sachse and H. Pfützner, A PCR scheme for differentiation of organisms belonging to the *Mycoplasma mycoides* cluster, *Vet. Microbiol.* **49** (1996), pp. 31–43.

S. Lorenzon, H. Wesonga, L. Ygesu, T. Tekleghiorgis, Y. Maikano, M. Angaya, P. Hendrikx and F. Thiaucourt, Genetic evolution of *Mycoplasma capricolum* subsp. *capripneumoniae* strains and molecular epidemiology of contagious caprine pleuropneumonia by sequencing of locus H2, *Vet. Microbiol.* **85** (2002), pp. 111–123.

R. Miserez, T. Pilloud, X. Cheng, J. Nicolet, C. Griot and J. Frey, Development of a sensitive nested PCR method for the detection of *Mycoplasma mycoides* subsp. *mycoides* SC, *Mol. Cell. Probes* **11** (1997), pp. 103–111.

A. Peyraud, S. Woubit, J.B. Poveda, C. De la Fe, P. Mercier and F. Thiaucourt, A specific PCR for the detection of *Mycoplasma putrefaciens*, one of the agents of the contagious agalactia syndrome of goats, *Mol. Cell. Probes* **17** (2003), pp. 289–294.

F.R. Rurangirwa, P.S. Shompole, A.N. Wambugu, S.M. Kihara and T.C. McGuire, Monoclonal antibody E8-18 identifies an integral membrane surface protein unique to *Mycoplasma capricolum* subsp. *capripneumoniae*, *Clin. Diagn. Lab. Immunol.* **4** (1997), pp. 615–619.

F.R. Rurangirwa and J.H. Kinyili, Contagious caprine pleuropneumonia, *Manual of Standards for Diagnostic Tests and Vaccines*, Office International des Epizooties (2000).

E. ter Laak, Identification of Mycoplasma F38 biotype, *Vet. Rec.* **129** (1991), p. 295.

F. Thiaucourt, C. Guerin, V. Mady and P.C. Lefèvre, Diagnosis of caprine contagious pleuropneumonia: recent improvements, *Rev. Sci. Technol.* **11** (1992), pp. 859–865.

F. Thiaucourt, G. Bölske, G. Libeau, C. Le Goff and P.C. Lefèvre, The use of monoclonal antibodies in the diagnosis of contagious caprine pleuropneumonia (CCPP), *Vet. Microbiol.* **41** (1994), pp. 191–203.

F. Thiaucourt and G. Bölske, Contagious caprine pleuropneumonia and other pulmonary mycoplasmoses of sheep and goats, *Rev. Sci. Technol.* **15** (1996), pp. 1397–1414.

F.J.M. van Kuppeveld, K.E. Johansson, J.M.D. Galama, J. Kissing, G. Bölske, J.T.M. van der Logt and W.J.G. Melchers, Detection of *Mycoplasma* contamination in cell cultures by a mycoplasma group-specific PCR, *Appl. Environ. Microbiol.* **60** (1994), pp. 149–152.

M. Zúñiga, G. Perez and F. Gonzalez-Candelas, Evolution of arginine deiminase (ADI) pathway genes, *Mol. Phylogenet. Evol.* **25** (2002), pp. 429–444.

2 CONCLUSION

In the present study specific PCR approaches were developed for the detection of mycoplasmas of the *M. mycoides* cluster (Woubit et al., 2007). Identification of Mccp, the aetiological agent of contagious caprine pleuropneumonia (CCPP) (Woubit et al., 2004) and *M. putrefaciens*, one of the causative agent of contagious agalactia syndrome (Peyraud et al., 2003).

These PCR diagnostic tools have several advantages over a number of other assays previously available: 1) As PCR-based assays, they do not require cultivation of the respective causative agents, allowing for direct detection of the causative agent from clinical cases albeit the presence of multiple mycoplasmas; 2) These PCR assays will play an important role in reporting the right mycoplasma in disease outbreaks. Consequently, this should permit to understand the exact geographical distribution of causative agents involved. For instance, CCPP outbreaks have been under-reported due to the frequent isolation of Mmc / MmmLC or *M. ovipneumoniae*, which is a common contaminant found in small ruminant samples.

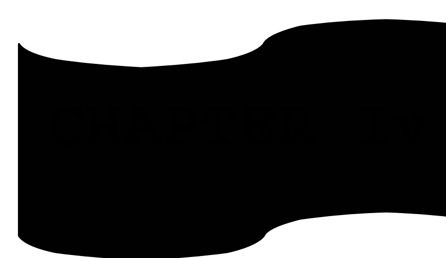
The importance of these diagnostic PCR tests has been illustrated in a recent epidemic of contagious agalactia in a disease endemic area. In contagious agalactiae surveys conducted in Gran Canaria (De la Fe et al., 2005) and in Lanzarote (De la Fe et al., 2007), *M. putrefaciens* PCR (Peyraud et al., 2003) was an invaluable tool towards identification of the causative agents, ruling out the presence of *M. putrefaciens*. These researchers showed the endemic nature of agalactia syndrome in these regions to be basically due to MmmLC / Mmc and *M. agalactia*. Although *M. putrefaciens* was not isolated in this studies, this species had been observed in the past in Gran Canary and Lanzarote (Gil et al., 1999), suggesting a sporadic distribution of this agent. These observations emphasize the importance of the present specific PCRs (Woubit et al, 2007; Peyraud et al., 2003) in the surveillance of this syndrome.

CCPP is one of the most important mycoplasmal diseases causing persistent ravaging of animal population in Africa and Asia. It is found in the list of OIE diseases subjected to international control measures. In the case of a recent outbreak of CCPP in captive wild ungulates at Al-Wabara wildlife preservation state of Qatar (Arif et al., 2007), Mccp specific PCR (Woubit et al., 2004) was used in the confirmation of the disease outbreak.

The specific PCR for the identification of Mccp has also been used in the confirmation of CCPP outbreak in the Thrace region of Turkey (Ozdemir et al., 2006).

Furthermore, the identification of specific sequences may be used for the development of highly specific, sensitive and rapid techniques such as real time PCR and Loop-mediated isothermal amplification (LAMP), for the specific identification of all members of the *M. mycoides* cluster. The specific primers for Mccp have already been applied in real time PCR assays and found to be more sensitive than the conventional PCR assay used.

In any case, these diagnostic tools will play a great role in the advancement of animal disease diagnosis and effective control measures.



Assessment of previously described and Identification of novel insertion sequence elements

1 Insertion sequence elements

Insertion sequence (IS) elements are loosely defined as small (<2.5 Kbp) cryptic DNA segments with a simple genetic organization encoding only proteins for their transposition and capable of inserting at multiple sites in a target molecule (Mahillon and Chandler, 1998). IS-elements are thought to be one of the major players in the plasticity of prokaryote genomes. More than 600 IS have been identified to date of which 385 are unique to 171 bacterial and archaeal species, falling to about 20 distinct groups. This number represents only a few of the many predicted from the ongoing microbial genome projects (Mahillon et al., 1999). IS-elements like all mobile genetic elements move or rearrange neighbouring DNA sequences of the host genome. Genetic phenomena associated with transposition of IS elements are spontaneous, associated with deletion or inversion of adjacent DNA segments, as well as activation of the transcription of flanking genes (Galas and Chandler, 1981). These activities can lead to assembly of gene clusters with specialized functions such as multiple antibacterial activities; virulence or symbiotic functions, or new catabolic pathways (Mahillon et al., 1999), mobile elements that carry antibiotic resistance genes are the major factors underlying the widespread dissemination of antibiotic resistance in bacterial population (Bruce et al., 2002).

1.1 Structure of IS-elements

Insertions sequences are genetically compact DNA fragments encoding functions involved in their mobility. They encode functions required in *cis*, which are recombinationally active DNA sequences defining the ends of the element. They also encode for the enzyme transposase (Tpase), which recognizes and process these ends. ISs are classified into about 20 families on the bases of various shared features. An IS family is defined as a group of ISs with related transposase, strong similarities of the catalytic site and conservation of organization such as disposition of their open reading frames (ORFs), length and similarity of the terminal inverted repeats (IRs) (Siguier et al., 2006) and the characteristic number of base pairs in the target DNA, which they duplicate upon insertion (Mahillon et al., 1999). The catalytic site of the transposase is also known as the DDE motif named for the three amino acids Asp, Asp and Glu that play a role in co-ordinating divalent metal ions involved in the chemical reactions required for transposition (Mahillon and Chandler, 1998). The Tpase is generally encoded by single or two open reading frames, and stretches nearly the entire length of the element (Figure IV. 1, Mahillon and Chandler, 1998). An IS element therefore displays a characteristic structure in which its ends are identified by the

inverted terminal repeats (IR) while the adjacent ends of the flanking host DNA are identified by the short repeats also known as direct repeats (DR) (Lewin, 2000).

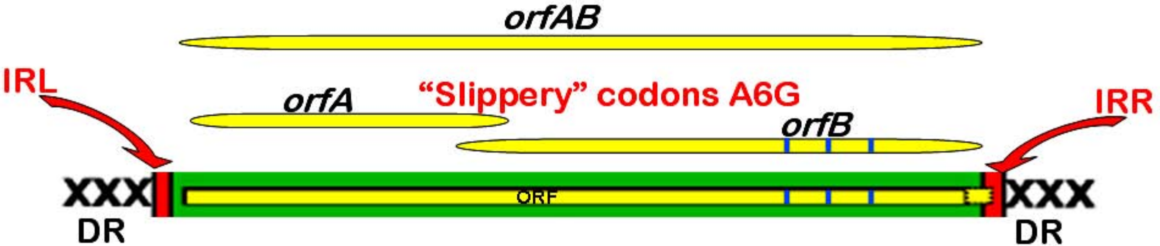


Figure IV. 1 Structure of a typical IS-element

Characterized by the presence of a single or double ORF (yellow), direct repeats of target sequence (black) and inverted repeats left and right of the sequence marking the ends of insertion sequence element IRL and IRR (red). Usually in IS3 family member insertion sequence elements have two consecutive overlapping open reading frames indicated as orfA and orfB, and are arranged in reading phase 0 and -1. ORFAB is produced by a programmed translational frameshifting. This involves a group of "slippery" codons (in this case A6G) on which the ribosome slides back for a single nucleotide. The DDE catalytic domains are shown by blue stippled lines (Adapted from IS-Finder Information).

1.1.2 Transposition of IS-elements

Most IS-elements move only very rarely (once in 10^5 cell generations for many elements in bacteria), and for this reason it is often difficult to distinguish them from non-mobile parts of the chromosome. In most cases, it is not known what suddenly triggers their movement (Bruce et al., 2002). Experiments show that IRs, which can be as short as 20 nucleotides, plays an important role for the DNA between them to be transposed by the particular transposase enzyme associated with the element.

The two types of ways an IS-element transposes in the genome are, cut-and-paste and copy-and-paste replication systems described in Figure IV.3 and Figure IV. 4. The cut-and-paste movement or non-replicative transposition of IS-element from one chromosomal site to another begins when the transposase brings the two inverted DNA sequences together, forming a DNA loop as shown in Figure IV. 2.

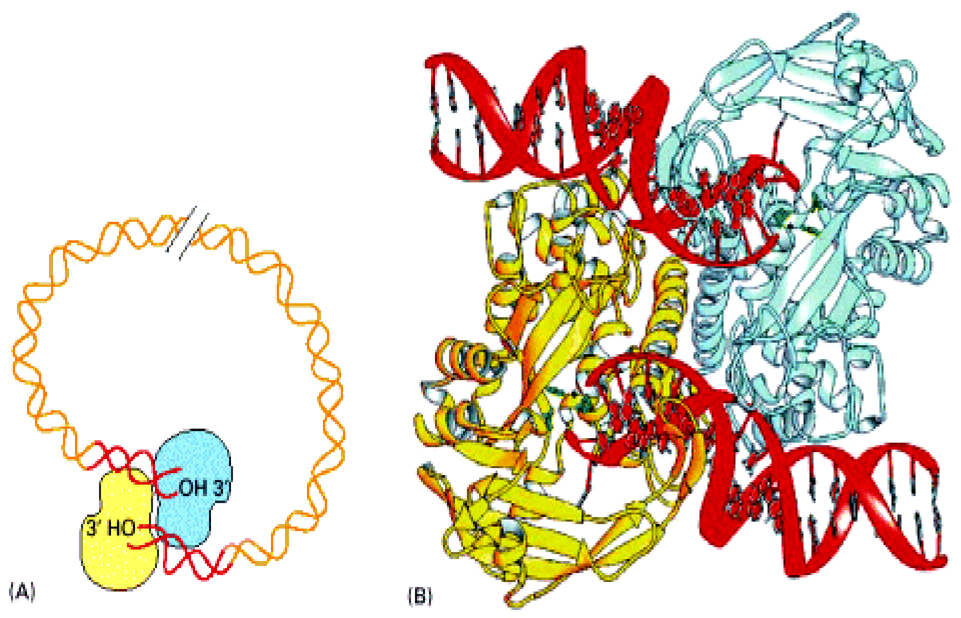


Figure IV. 2 The structure of the central intermediate formed by a cut-and-paste transposase *(A)* Schematic view of the overall structure. *(B)* The detailed structure of a transposase holding the two DNA ends, whose 3′-OH groups are poised to attack a target chromosome. (B, from. D.R. Davies et al., *Science* 289:77–85, 2000. © AAAS, Bruce et al., 2002)

Insertion into the target chromosome, catalyzed by the transposase, occurs at a random site through the creation of staggered breaks in the target chromosome *(red arrowheads)* (Figure IV.3). As a result, the insertion site is marked by a short direct repeat of the target DNA sequence, as shown. Although the break in the donor chromosome *(green)* is resealed, the breakage-and-repair process often alters the DNA sequence, causing a mutation at the original site of the excised transposable element (Galas and Chandler, 1981; Lewin, 2002; Turlan and Chandler, 1995).

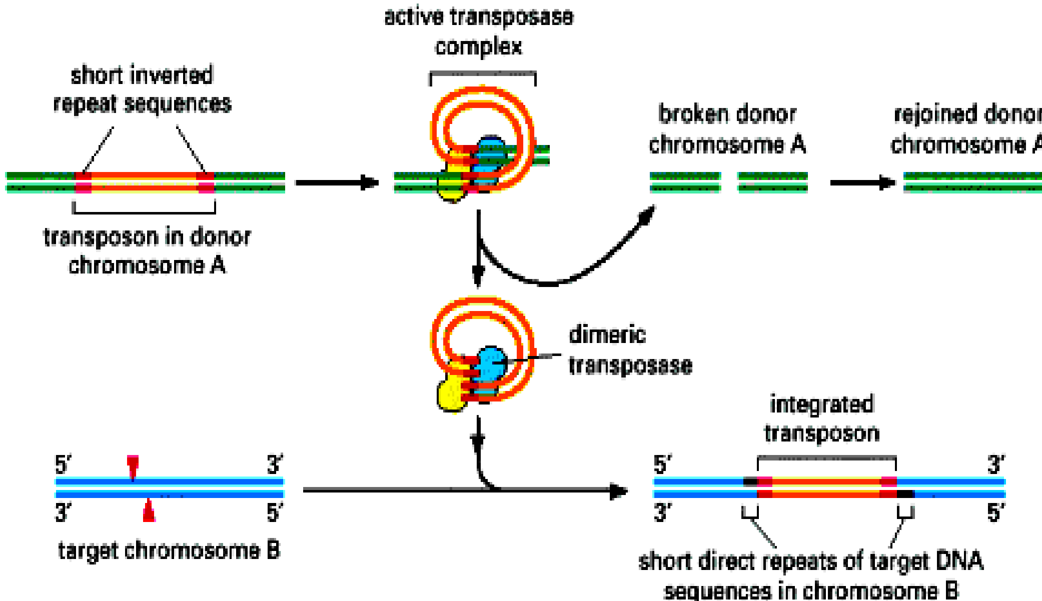


Figure IV. 3 Cut-and-paste movement of IS transposition.

This begins when the transposase brings the two inverted DNA sequences together, forming a DNA loop. Insertion into the target chromosome, catalyzed by the transposase, occurs at a random site through the creation of staggered breaks in the target chromosome *(red arrowheads).* As a result, the insertion site is marked by a short direct repeat of the target DNA sequence, as shown. Although the break in the donor chromosome *(green)* is resealed, the breakage-and-repair process often alters the DNA sequence, causing a mutation at the original site of the excised transposable element (not shown) (Bruce et al., 2002).

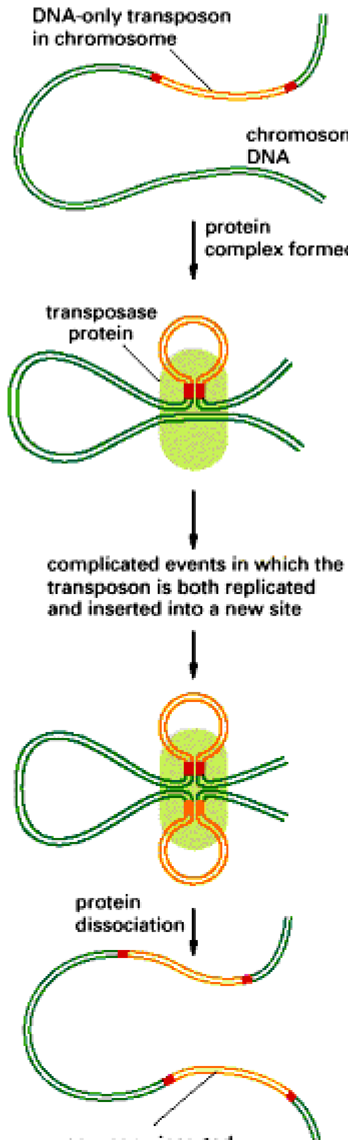


Figure IV. 4 Copy and paste transposition.

Here the DNA sequence of insertion sequence element is copied by DNA replication. The end products are an insertion sequence element that is identical to the original donor in to the target DNA molecule.

1.1.2 Distribution of IS-elements

IS-elements found in the genomes of many different bacteria and mycoplasma at a multiplicity of between a few and several hundred per genome. In bacteria they are very frequently found as part of natural plasmids. Insertion sequences provide a valuable source of experimental material for the study of gene expression, recombination and repair events, population dynamics, and horizontal transmission of genes (Frey, 2003). The distribution of many insertion sequences within and between various bacterial species has often been investigated as part of the initial characterisation of a new element, usually by simple Southern hybridization. Although useful in "typing" strains, much of the data remains

purely descriptive. However, few systematic attempts have been made to determine the dynamics of insertion sequences within bacterial populations in a controlled manner (Mahillon and Chandler, 1998).

In mycoplasma species it is common to observe repetitive elements with features similar to the IS*3* family of insertion sequences. The most recent data from IS-finder database (http://www-is.biotoul.fr/is.html) of IS elements presents the discovery of more than 20 different IS elements in mycoplasmas including mycoplasmas of the *M. mycoides* cluster (Bhugra and Dybvig, 1993; Calcutt et al., 1999; Frey et al., 1995; Peterson et al., 1995; Vilei et al., 1999; Westberg et al., 2002; Zheng and McIntosh, 1995), most of these elements have not been shown to transpose within the chromosome (Dybvig and Volker, 1996). The exception is IS*1138* from *Mycoplasma pulmonis*, which is observed to transpose actively (Bhugra and Dybvig, 1993).

1.1.2.1 IS-elements in the M. mycoides cluster

The first known IS element to be discovered from the members of the *M. mycoides* cluster was IS*1296* a member of the IS3 family. The insertion sequence IS*1296* was identified from MmmSC in 18-19 copies, the agent of contagious bovine pleuropneumonia (Frey et al., 1995). In this same study Frey and colleagues found that this IS-element also existed in other members of the *M. mycoides* cluster, MmmLC and Mbg7, so far there is no evidence of its presence in other species than the *M. mycoides* cluster. In spite of its high A+T content and use of mycoplasma specific codon UGA, it shows high structural similarities to the IS*150* of *E. coli*. In IS*1296* the 1483 bp sequence length is characterized by two partially overlapping ORFs, ORFA and ORFB and a terminal 30 bp IRs (Frey et al., 1995). Owing to the low copy number of RNA genes (*rrn*), ribotyping is not applicable to mycoplasma, IS-elements however found to offer practical application in strain typing. For the first time IS*1296* have been used in the differentiation of MmmSC isolates from European to those from Africa and Australia, southern hybridization pattern of African and Australian strains showed four closely related patterns which belong to a separate cluster (Cheng et al., 1995).

The other insertion sequence named IS*1634* was initially discovered while studying the DNA segments proximal to copies of IS*1296* in MmmSC species; up on southern blot

hybridization Vilei and colleagues found 30 copies of this IS-element within the genome of MmmSC. The amino acid sequence of the putative transposase is more closely related to IS*1549* of *Mycobacterium smegmatis*. So far this insertion sequence is not found in any other members of the *M. mycoides* cluster. But ISMbov*3* from *M. bovis* was discovered to be very closely related to IS*1634* (Thomas et al., 2005). IS*1634* belongs to the IS*4* family member of insertion sequences with a total size of 1872 bp containing a single ORF and characterized by 13 bp IRs (Vilei et al., 1999). Like IS*1296*, IS*1634* have been used for MmmSC strain characterization and differentiation of strains geographical origin (March et al., 2000; Vilei et al., 1999).

An additional IS-element has been discovered while genome sequencing of MmmSC strain PG1^T (Westberg et al., 2004), this new insertion sequence, ISMmy*1* existed in eight copies with one copy being truncated (Westberg et al., 2002). It has a total length of 1670 bp, contains a single ORF and posses a 30 bp terminal IR. Like IS*1634*, the putative transposase of ISMmy*1* contains the DDE motif that resembles the IS*4* family. It was suggested that MmmSC probably acquired this IS element by horizontal gene transfer from that of *M. bovis*. Though different MmmSC strains gave similar hybridization pattern for this IS, it was found to be a useful tool in the identification of vaccinal strain T1Sr49 due to its unique hybridization pattern (Westberg et al., 2002).

Analysis of whole genome sequence of MmmSC type strain PG1^T revealed presence of 60 copies of IS*1634* where two copies are split by IS element and one is truncated. Twenty eight copies of IS*1296* where four copies are interrupted by IS-elements and seven truncated copies (Westberg, 2003). Specific DNA probes from IS-elements ISMmy*1* and IS*1634* have recently been used to profile *M. bovis* field isolates in UK. These two insertion sequences, IS*1634* and ISMmy*1* show similarities of 96 % and 98 % with their homologue from *M. bovis*, identified as IS*Mbov3* and IS*Mbov2* respectively (Miles et al., 2005). This shows Insertion sequence elements have evolved divergently in to different species of mycoplasmas. The high relatedness of these two IS elements between *M. bovis* and MmmSC, could be the result of horizontal transfer of these elements during co-infection of same bovine host (Thomas et al., 2005).

The other fully sequenced genome of type strain California kid^T possesed locus MCAP_0848, which has been identified IS*1296EH* and two other loci MCAP_0187,

MCAP_0315 as putative transposase of IS*3* families (GenBank/ EMBL Accession No. NC_007633). Although it is not characterized, IS*1296EH* from California kid^T should be a new insertion sequence element as it has presented homology of only 58 % with IS*1296* ORFA.

Besides the known presence of insertion sequence IS*1296* in MmmLC type strain Y-goat^T, whole genome sequence of MmmLC strain 95010-C1 has revealed the presence of new insertion sequence elements.

1.1.2.2 IS-elements of M. mycoides subsp. mycoides LC strain 95010-C1

Prior to whole genome sequencing of strain 95010-C1, presence of formerly discovered insertion sequence elements were analysed using the southern hybridization. DNA-probes from IS*1296*, IS*1634* and ISMmy*1* hybridized to the *Hind*III digested total DNA of 95010-C1 and Y-goat^T and strains PG1^T and 8740-Rita of MmmSC. The result revealed the presence 7 copies of IS*1296* in 95010-C1; with bands at around 4 Kbp and 6.5 Kbp being more intense than the rest of the five bands (Figure IV. 5). The differences in band intensity perhaps reveal the presence of additional copies of this insertion sequence element. Strain Y-goat^T has revealed the presence of 5 copies of IS*1296*. The southern hybridization has also revealed difference in IS*1296* copies between the two MmmSC strains, strain PG1^T seems to possess one more copy than strain 8740-Rita, shown by an arrow head in Figure IV. 5.

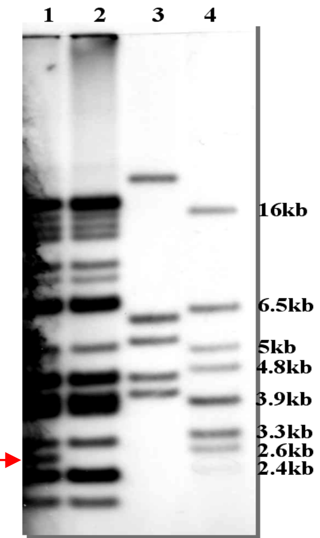


Figure IV. 5 Southern hybridization of IS*1296* DNA probe to *Hind*III digested genomic DNA**.** Lanes 1 & 2 corresponding to PG1^T & 8740-Rita from MmmSC strains, lanes 3 & 4 corresponding to Y-goat^T & 95010-C1 of MmmLC strains.

Insertion sequence ISMmy*1* was observed in 6 copies in strain 95010-C1 (data not shown) and no signal was detected for IS*1634* (Data not shown).

Upon annotation of the genome sequence of strain 95010-C1, the exact copy numbers of IS*1296* and ISMmy*1* were found to be 9 and 6 copies respectively. The result of the southern blot correlates with what was obtained with genome sequence result. From the whole genome sequence data *Hind*III digested IS*1296* containing fragments were obtained at 11887 bp, 6403 bp, 6400 bp, 5115 bp, 4619 bp, 3888 bp, and 3250 bp. The bands of higher intensity in our result at 6.5 Kbp and 3.9 Kbp can be explained by overlapping of DNA fragment at 6.5 Kbp (6403 bp and 6400 bp) and presence of two copies of IS*1296* within the 3.9 Kbp fragment (3888 bp). This observation revealed that IS*1296* is the most abundant insertion sequence element in MmmLC strains studied. In addition to this, in the course of whole genome annotation of strain 95010-C1 MmmLC, we discovered the presence of two novels insertion sequence elements of the IS*3* family. These two insertion sequences were named ISMmy*2* and ISMmy*3* as suggested by the curators of the IS

database (http://www-is.biotoul.fr/). The description of ISMmy*3* is out of the scope of this thesis.

The objective of this part of the study is characterization of new IS-element, ISMmy*2* and determination of its distribution, which may be of additional tool of future relevant in typing of species of the *M. mycoides* cluster. Therefore, the fourth article of this thesis characterizes ISMmy*2* one of the novel insertion sequence element and tries to explore its presence in other mycoplasmas of the *M. mycoides* cluster including closely related mycoplasma species.

Paper IV

ISMmy*2*
a Novel IS3-Type Insertion Sequence found in *Mycoplasma mycoides* subsp. *mycoides* LC.

Woubit Salah
Lucía Manso-Silván
François Thiaucourt

CIRAD-Bios, UPR15 “control of exotic and emerging animal diseases”. TA-A15/G Campus International de Baillarguet, 34398, Montpellier cedex 5, France

Abstract

A new insertion sequence, ISMmy*2*, has been identified in the caprine pathogen *Mycoplasma mycoides* subsp. *mycoides* LC (MmmLC). It shows structural and functional similarities to IS*1138* of *Mycoplasma pulmonis*, although their sequences differ significantly. ISMmy*2* consists of 1373 bp with two 24 bp terminal inverted repeats. It contains a single open reading frame encoding a product of 426 amino acids including a transposase domain and displaying a DDE motif. ISMmy*2* is present in 6 copies in the completely sequenced genome of MmmLC strain 95010-C1, where it has apparently generated trinucleotide direct repeats at the target sites. However, this insertion sequence is neither present in all MmmLC strains nor exclusive to this biotype. ISMmy2 has been found in other members of the *Mycoplasma mycoides* cluster such as *M. capricolum* (96 % amino acid similarity) and in the related species *M. cottewii* and *M. yeatsii* (86 % amino acid similarity). A remnant of ISMmy2 has also been found in *Mycoplasma mycoides* subsp. *mycoides* SC (MmmSC), where it is present as a single copy devoid of transposase activity.

Keywords: Insertion sequence; ISMmy2; *Mycoplasma mycoides*;

1 Introduction

The class Mollicutes groups bacteria that lack a cell wall. Within this class, the genus Mycoplasma contains more than 120 species that are found in animals and humans as commensals or parasites. They are notably characterized by a small genome size (0.5-1.3Mb) with a low G+C content, constituting the smallest self-replicating organisms [16]. Phylogenetic studies based on 16S rRNA sequences showed that these bacteria evolved from Gram positive bacteria such as *Bacillus*, *Lactobacillus* and *Clostridium* by reductive evolution. Genome size diminution in Mollicutes was a result of gene loss rather than gene size reduction [2] and was accompanied by specific adaptation to an ecological niche, such as the respiratory or urogenital tract, joints, etc, serving as a supply for the nutrients that they were no longer able to synthesize. As a consequence, these are extremely fastidious organisms requiring rich, complex media for in vitro growth, whereas some of them remain uncultivable.

Within the Mycoplasma genus there is a group called the *Mycoplasma mycoides* cluster gathering six species that are all pathogenic for animals [4]. This group is particularly important for veterinarians, as it includes two pathogens responsible for notifiable animal diseases of great economic importance that threat international trade. These diseases are known as Contagious Bovine Pleuropneumonia (CBPP) [15], caused by *M. mycoides* subsp. *mycoides* SC (MmmSC), and Contagious caprine pleuropneumonia (CCPP) [19], caused by *Mycoplasma capricolum* susbp. *capripneumoniae* (Mccp). Surprisingly however, from a phylogenetic point of view, the *M. mycoides* cluster belongs to the *Spiroplasma* group, which comprises species that parasite insects and plants. This relatedness was demonstrated by 16S rDNA sequence similarity [23] and, more recently, by comparison of whole genome information and inference from multiple sets of concatenated core housekeeping genes [11]. Concatenated housekeeping gene sequence comparisons have also allowed the construction of a robust phylogenetic tree, assigning a phylogenetic position to all the members of this cluster, including the yet unassigned "group 7 of Leach" strains [10]. Within the *M. mycoides* cluster, MmmSC was the first for which a complete genome sequence was made available [24]. The unravelling of the 1,211 Kb-long sequence of the type strain PG1^T was in some way disappointing, as it did not allow the identification of clear virulence factors such as toxins or adhesins. On the other hand, it revealed that more than 13 % of the genome consisted of three kinds of Insertion Sequence (IS) elements: IS*1296* [6], IS*1634* [22] and ISMmy*1* [25], which were present in

28, 60 and 9 copies respectively. IS elements are mobile DNA segments of less than 2.5 kb that share the same structure [9]. They contain usually one or two open reading frames coding for a transposase and are terminated by short inverted repeats that serve as recognition and cleavage sites for the transposase [8]. The MmmSC sequence was found to be the most IS-dense genome known in spite of its reduced size. Complete genome sequencing of the Large Colony biotype of *M. mycoides* subsp. *mycoides* (MmmLC) was undertaken in our laboratory aiming to unravel the relationship existing between the two organisms and to understand why such phylogenetically close relatives differ in pathogenicity and physiology. Comparison of MmmLC and MmmSC coding sequences allowed the identification MmmLC-specific sequences.

Upon annotation of the MmmLC genome sequence, one of these specific coding sequences was identified as an IS element by similarity with IS*1138* of *M. pulmonis* [1]. The present study aims to characterize this IS and its distribution within the *M. mycoides* cluster and related species.

2. Materials and Methods

2.1. Mycoplasma strains and culture conditions

The mycoplasma strains used in this study (Table 1) were selected to represent a wide geographical distribution. The MmmLC sequenced strain 95010-C1 was isolated from a clinical case of mastitis in a French goat, where it was obtained as a pure culture. Strains were cultured in modified Hayflick broth (PPLO broth without crystal violet 21 g l^{-1}, 20 % horse serum de-complemented for 1 hr at 56°C, 5 % fresh yeast extract, 0.2 % glucose, 0.4 % sodium pyruvate) [14, 20] at 37 °C, 5 % CO_2 in a containment level 3 laboratory and were harvested at the late exponential phase of growth. DNAs were extracted from stationary phase using phenol: chlorophorm: isoamyl alcohol (Sigma Aldrich).

2.2. Identification of ISMmy*2*

ISMmy*2* was discovered while annotating the genome of MmmLC strain 95010-C1 (Unpublished data). The MmmLC genome was sequenced by Genoscope (Evry, France). Once supercontigs were available, sequences were included in a web-based annotation tool (CAAT-Box platform) that allowed determination of the most probable start codons and comparison of coding sequences with those present in other mycoplasma genomes

(Molligen: http://cbi.labri.fr/outils/molligen/) or protein databases (Swiss-Prot: http://expasy.org/sprot/).

2.3. PCR labelling of the DNA probe

A 971 bp long DNA probe was prepared using primers ISMmy2-probe-F/R, which were selected based on the alignment of ISMmy*2* insertion sequence copies found in the 95010-C1 sequence. Digoxigenin (DIG) PCR labelling was performed according to the manufacturer's instructions (Roche Diagnostics GmbH, Mannheim, Germany). The reaction volume was fixed to 100 µl including: 10 µl of 10 X Taq buffer (Roche Diagnostics), 30 µM dig11dUTP, 180 µM dTTP, 200 µM dATP, 200 µM dGTP, 200 µM dCTP (for a 1:6 ratio of dig11dUTP:dTTP, Roche Diagnostics), 0.4 µM of each primer ISMmy2-probe-F and ISMmy2-probe-R, 5.25 U of Taq long expand DNA polymerase (Roche Diagnostics) and 1 µl of a 20 ng/µl DNA template (DNA extracted from strain 95010-C1). The PCR conditions consisted of an initial denaturation step at 94°C for 2 min, followed by 30 cycles of 94 °C for 30 sec, 52 °C for 30 sec and 68 °C for 1 min.

2.4. Restriction enzyme analysis and Southern blot

*Hind*III was used for initial ISMmy*2* screening: 2.5 µg genomic DNA from 32 mycoplasma strains was digested with *Hind*III and 15 µl of digested product from each sample was loaded on a 0.7% agarose gel. Electrophoresis was conducted overnight at 18 V. Gels were then depurinated (0.25 M HCl, 15 min), denatured (0.5 M NaOH/ 1.5 M NaCl, 30 min) and neutralized (Tris-HCl 0.5M/ 1.5 M NaCl, 30 min). DNA was then transferred to a nylon membrane overnight at room temperature by a capillary transfer method using 20X SSC buffer. Wet membranes were placed, DNA face up, over a 2X SSC soaked Whatman (3 mm) paper and transfixed under a UV light for 3 min. The membranes were then rinsed in double distilled water and air dried.

Membranes were pre-hybridized for 2 hrs at 55°C and hybridized overnight in blocking solution 5X SSC, 1 % (v/v) blocking reagent (Roche), 0.1 % (w/v) N-lauroylsarcosine, and 0.02 % (w/v) SDS. Membranes were then incubated in hybridization buffer containing the diluted probe overnight at 50°C. Upon completion of hybridization, membranes were washed 15 min each in low (2X SSC, 0.1% SDS at room temperature) and high stringency solutions (0.5X SSC, 0.1% SDS at 65°C). After blocking, the membranes were incubated in anti-DIG antibodies, 1/8000 Fab conjugate for 30 min (Roche Diagnostics). Subsequent

to multiple washing in 1X PBS, the hybridized probe was detected using Hyper film-ECL and chemiluminescent detection reagent following standard procedures (Amersham, Biosciences, UK). Only for those species which gave a positive hybridization band on preliminary Southern blot, additional enzymes such as *Eco*RI, *Eco*RV, and *Hind*II were used for the verification of ISMmy*2* copy numbers.

2.5. PCR and sequencing

To confirm the results obtained by Southern blotting a PCR was conducted on the 32 strains with a single primer, IR-ISMmy2, the sequence of which was based on the 17 bp perfect inverted repeat at both extremities of the ISs. All amplified products were sequenced by Cogenics (Meylan France). The full sequence of the transposase from the positive strains was then obtained using sequencing primers listed on Table 2.

Alignment of the different ISMmy*2* copies present in MmmLC strain 95010-C1 as well as IS sequences found in the *M. mycoides* cluster and closely related species were performed using Vector-NTI, AlignX software program (Invitrogen).

3. Results and Discussion

3.1. Characterization of ISMmy*2*

During the annotation step of the genome of MmmLC strain 95010-C1, a coding sequence was found to have high localized similarities with two MmmSC PG1^T coding sequences: MSC_0698 (length=112 amino-acids (aa)), similarity= 81 %) and MSC_0699 (length=215aa, similarity= 93 %). It also showed some similarities with the full length of a coding sequence of *M. pulmonis* (length=416aa, similarity= 56 %). This sequence, MYPU_0180, was annotated as a transposase for insertion sequence element IS*1138*. Furthermore, inverted repeats were identified upstream and downstream this coding sequence.

The full sequence was considered as a prototype for a new insertion sequence and submitted to GeneBank (Acc. N°: DQ887910) and to the "IS-finder" web server for Insertion sequences [17] (http://www-is.biotoul.fr/). It was characterized by a total length of 1374 bp, the presence of 24 bp-long inverted repeats and the generation of 3 bp-long direct repeats at the insertion sites. It presented a putative ribosome binding site (position 66-71) preceding by 5 bp the ATG start of a 1278 bp-long ORF (426aa) (Fig 1a). This ORF possessed a classical DDE catalytic site motif [7] (Fig 2).

Accordingly, this IS was identified as belonging to the IS*3* family and the IS*150* group. As it differed significantly from all known IS, it was given a new name: ISMmy*2*.

3.2. Distribution of ISMmy*2 in MmmLC*

A BLAST analysis on the 95010-C1 sequence revealed that there were 6 copies inserted within this genome (Fig 3). The copies were named by adding a unique letter after ISMmy*2*, the first, (ISMmy*2*-A) being the closest to the replication origin. Compared to the first copy, the five additional copies were one base shorter. As the deletion occurred upstream of the ORF, this did not modify the actual length of the putative transposase that remained 426aa-long in all copies. Compared to ISMmy*2*-A, the lowest percentage of identity was observed in the ISMmy*2*-B copy with 97.7 % identity in the protein sequence, showing the high conservation of ISMmy*2* sequences within this genome.

Four IS copies displayed a classical trinucleotide direct repeat (DR) at their extremities (Table 3). The most frequent trinucleotide DR was "taa" (C and D) but two copies displayed different DR: "att" (A) and "ggt" (B). Interestingly, two copies (E and F) seemed to have exchanged their DR "att" and "cat". This may have resulted from homologous inter- or intra-molecular recombination between two IS copies with a different DR sequence, resulting in a hybrid element carrying one DR of each parent. It must be noted that ISMmy*2*-E and F are quite close to one another (9kb) and that they present two ISMmy*1* copies in between, which may have caused the recombination.

3.3. Presence of ISMmy*2 in related mycoplasma species*

The presence of ISMmy*2* in related mycoplasma species was evaluated by two techniques: Southern blotting with a DIG-labelled probe and PCR using a single primer corresponding to the conserved part of the inverted repeat found at the extremities of the IS element. Both techniques yielded identical results in terms of detection. Elements related to ISMmy*2* were detected in 6 out of 10 strains belonging to the subspecies MmmLC and *Mycoplasma mycoides* subsp. *capri* (Mmc), which are now considered a single entity [21], and in 1 out of 5 *Mycoplasma capricolum* subsp. *capricolum* (Mcc) strains (Table 1, Fig 4).

On the other hand, this IS was not detected by any of the two techniques either in MmmSC or in Mccp strains (Table 1). The absence of detectable ISMmy*2* in the SC biotype was quite surprising, given that this IS was found to share strong homologies with two

MmmSC CDS. In fact, a single DNA sequence homologous to ISMmy*2* was found between positions 801579 and 802903 of the PG1^T sequence (NC_005364.2). This homologous sequence seems to represent the remnant of an ISMmy*2* deprived of transposase activity by a mutation disrupting the transposase (Fig 1b). The sequence presently found in PG1^T may be the result of ensuing genetic drift by accumulation of base insertions, deletions and mutations, as well as deletion of the 49 terminal bases of the original IS. Our inability to detect this sequence by Southern hybridization may be explained by the presence of *Hind*III cutting sites within the PG1^T ISMmy*2*-remnant sequence (that do not exist in MmmLC ISMmy*2* copies) resulting in much smaller fragments. Our inability to detect these sequences by PCR was easily explained by the absence of the Right Inverted Repeat that is essential for PCR amplification with a single primer.

Interestingly, elements related to ISMmy*2* were detected in the related species *M. cottewii* and *M. yeatsii*. Southern blotting allowed evaluation of the number of IS copies found in PCR positive strains, which varied from 2 to 7 (Table 1). The consensus sequence of the transposase gene of these two species was obtained by sequencing of the PCR product obtained with primer IR-ISMmy2 allowed us to get. The total length of the coding sequence (426aa) was identical to ISMmy*2* and the percentage of identical amino-acid to ISMmy*2*-A was 86 % (Table 3). This would suggest the presence of a “variant” of ISMmy*2* in these two species. The lower percentage of amino acid-identity, 86 %, (compared to 96 % identity found within the *M. mycoides* cluster) was well correlated to the phylogenetic distance between *M. cottewii* and *M. yeatsii* and the *M. mycoides* cluster [10]. On the other hand, the variant ISMmy*2* transposase sequence was found to show identical sequences in these two species, reflecting their phylogenetic proximity or their common ecological niche that could have favoured horizontal gene transfer.

No ISMmy*2* homologues were found by BLAST analysis on the non redundant databases (NCBI/BLASTp, non redundant databases 04th October 2007) and notably in the *M. agalactia*e PG2 genome (NC_007633) nor in the *M. bovis* PG45 available sequences (The Institute of Genomic Research).

4. Conclusion

ISMmy*2* is the fourth Insertion Sequence described in the *Mycoplasma mycoides* cluster. The first three, IS*1296*, IS*1634* and ISMmy*1*, were described in MmmSC. Phylogenetic studies have already shown that MmmSC and MmmLC are very closely related and it is no surprise that some of these IS elements are shared by the two biotypes (IS*1296*, ISMmy*1*). ISMmy*2* is also found in MmmSC but as a "remnant" that has lost its ability to transpose. There is no doubt that the more mycoplasma strains are sequenced, the more IS-elements will be found. Already there seem to be additional IS elements that have been spotted in the *M. capricolum* California kid^T genome (MCAP_0187 and MCAP_0315 in NC_007633.1) but are not fully characterized yet.

ISMmy*2* has also been found in one *M. capricolum* strain. This may be explained by horizontal gene transfer between species sharing the same habitat in small ruminants. The presence of ISMmy*2*-like sequences in *M. cottewii* and *M. yeatsii* may also be explained by lateral gene transfer. Although the latter two species are mostly found in the external ear canal of goats [5] it is not uncommon to find MmmLC strains in the same habitat [18]. However, given that the transposase sequence of *M. cottewii* and *M. yeatsii* differs significantly from that of ISMmy*2*, the acquisition of this IS element in these two species must have been an ancient event.

These IS elements are not necessarily present in all the strains of one species. When present, their copy number may vary as well as their insertion site. This may therefore make them good targets for rapid molecular typing methods based on PCR or Southern Blotting [3] [12]. Some of these IS elements may have the capacity to alter the expression of adjacent genes when they transpose to a new genome location [13]. Their study may therefore provide clues to understanding the molecular basis for the virulence of some mycoplasma strains or species.

Acknowledgements

We are particularly grateful to Valerie Barbe at Genoscope Evry France for the sequencing and assembly of the MmmLC 95010-C1 genome sequence. Patricia Siguier curator of IS-Finder: http://www-is.biotoul.fr/ Université Paul SABATIE-31062 Toulouse cedex 9 FRANCE for critical review of our ISMmy*2* submission.

We are also deeply indebted to the French Embassy in Addis Ababa for supporting Woubit Salah PhD studies in France

References

[1] B. Bhugra, K. Dybvig, Identification and characterization of IS1138, a transposable element from Mycoplasma pulmonis that belongs to the IS3 family., Mol. Microbiol. 7 (1993) 577-584.

[2] J. M. Bové, Molecular features of mollicutes, Clinical infectious diseases. 17 (1993) S10-31.

[3] X. Cheng, J. Nicolet, F. Poumarat, J. Regalla, F. Thiaucourt, J. Frey, Insertion element IS*1296* in *Mycoplasma mycoides* subsp. *mycoides* small colony identifies a European clonal line distinct from African and Australian strains, Microbiology. 141 (Pt 12) (1995) 3221-3228.

[4] G. S. Cottew, A. Breard, A. J. DaMassa, H. Erno, R. H. Leach, P. C. Lefevre, A. W. Rodwell, G. R. Smith, Taxonomy of the *Mycoplasma mycoides* cluster, Isr J Med Sci. 23 (1987) 632-635.

[5] G. S. Cottew, F. R. Yeats, Mycoplasmas and mites in the ears of clinically normal goats, Aust. Vet. J. 59 (1982).

[6] J. Frey, X. Cheng, P. Kuhnert, J. Nicolet, Identification and characterization of IS*1296* in *Mycoplasma mycoides* subsp. *mycoides* SC and presence in related mycoplasmas, Gene. 160 (1995) 95-100.

[7] J. Kulkosky, K. S. Jones, R. A. Katz, J. P. Mack, K. M. Skalka, Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases., Mol.Cell Biol. 12 (1992) 2331-2338.

[8] J. Mahillon, C. Leonard, M. Chandler, IS elements as constituents of bacterial genomes, Research in Microbiology. 150 (1999) 675-687.

[9] J. Mahilon, M. Chandler, Insertion sequences, Microbiol. Mol. Rev. 62 (1998) 725-774.

[10] L. Manso-Silvan., X. Perrier, F. Thiaucourt, Phylogeny of the Mycoplasma mycoides cluster based on analysis of five conserved protein-coding sequences and possible implications for the taxonomy of the group. Int J Syst Evol Microbiol 57 (2007) 2247-2258.

[11] K. Oshima, H. Nishida, Phylogenetic Relationships Among Mycoplasmas Based on the Whole Genomic Information, Journal of Molecular Evolution. (2007).

[12] P. Pilo, B. Fleury, M. Marenda, J. Frey, E. M. Vilei, Prevalence and distribution of the insertion element ISMag1 in *Mycoplasma agalactiae*, Veterinary Microbiology. 92 (2003) 37-48.

[13] I. Podglajen, J. Breuil, E. Collatz, Insertion of a novel DNA sequence, 1S1186, upstream of the silent carbapenemase gene cfiA, promotes expression of carbapenem resistance in clinical isolates of Bacteroides fragilis, Mol.Microbiol. 12 (1994) 105-114.

[14] A. Provost, P. Perreau, A. Breard, C. Le Goff, J. L. Martel, G. S. Cottew, Contagious bovine pleuropneumonia, Rev. sci. tech. Off. int. Epiz. 6 (1987) 625-679.

[15] A. Provost, P. Perreau, A. Breard, C. Le Goff, J. L. Martel, G. S. Cottew, Contagious bovine pleuropneumonia, Rev. Sci. Tech. Off. Int. Epiz. 6 (1987) 625-679.

[16] S. Razin, D. Yogev, Y. Naot, Molecular Biology and Pathogenicity of Mycoplasmas, Microbiol. Mol. Biol. Rev. 62 (1998) 1094-1156.

[17] P. Siguier, J. Perochon, L. Lestrade, J. Mahillon, M. Chandler, IS-finder: the reference centre for bacterial insertion sequences, Nucl. Acids Res. 34 (2006) D32-36.

[18] F. Tardy, P. Mercier, M. Solsona, E. Saras, F. Poumarat, *Mycoplasma mycoides* subsp. *mycoides* biotype large colony isolates from healthy and diseased goats: Prevalence and typing, Veterinary Microbiology. 121 (2007) 268-277.

[19] F. Thiaucourt, G. Bolske, Contagious caprine pleuropneumonia and other pulmonary mycoplasmoses of sheep and goats, Rev Sci Tech. 15 (1996) 1397-1414.

[20] F. Thiaucourt, C. Guerin, V. Mady, P. C. Lefevre, [Diagnosis of caprine contagious pleuropneumonia: recent improvements], Rev Sci Tech. 11 (1992) 859-865.

[21] E. M. Vilei, B. M. Korczak, J. Frey, *Mycoplasma mycoides* subsp. *capri* and *Mycoplasma mycoides* subsp. *mycoides* LC can be grouped into a single subspecies, Vet Res. 37 (2006) 779-790.

[22] E. M. Vilei, J. Nicolet, J. Frey, IS*1634*, a novel insertion element creating long, variable-length direct repeats which is specific for *Mycoplasma mycoides* subsp. *mycoides* small-colony type, J Bacteriol. 181 (1999) 1319-1323.

[23] W. G. Weisburg, J. G. Tully, D. L. Rose, J. P. Petzel, H. Oyaizu, D. Yang, L. Mandelco, J. Sechrest, T. G. Lawrence, J. Van Etten, et al., A phylogenetic analysis of the mycoplasmas: basis for their classification, J Bacteriol. 171 (1989) 6455-6467.

[24] J. Westberg, A. Persson, A. Holmberg, A. Goesmann, J. Lundeberg, K.-E. Johansson, B. Pettersson, M. Uhlen, The Genome Sequence of *Mycoplasma mycoides* subsp. *mycoides* SC Type Strain PG1T, the Causative Agent of Contagious Bovine Pleuropneumonia (CBPP), Genome Res. 14 (2004) 221-227.

[25] J. Westberg, A. Persson, B. Pettersson, M. Uhlen, K. E. Johansson, ISMmy1, a novel insertion sequence of *Mycoplasma mycoides* subsp. *mycoides* small colony type, FEMS Microbiol Lett. 208 (2002) 207-213.

Figures and Tables captions

Table 1. Mycoplasma strains used to study the distribution of the insertion sequence element ISMmy*2*.

Strain	Year	Origin	PCR	Southern blots			
				***EcoR*V**	***EcoR*I**	***Hind*III**	***Hind*II**
M. mycoides subsp. *mycoides* LC				**Apparent copy numbers**			
55507-1	1998	Germany	+	2	2	2	2
95010-C1	1995	France	+	6	6	6	6
Y-goat[R]	1956	Australia	+	6 ?	6	6 ?	6
2002-054 (VP9L)	<2002	India	-			-	
8756-C13	<1987	USA	-			-	
Kombolcha	1975	Ethiopia	-			-	
M. mycoides subsp *capri*							
WK354	1980	Switzerland	+	7	7	7	7
N108	1977	Nigeria	+	6	7		7
L	1975	France	+	5	5	5	5
PG3[T]	1950	Turkey	-			-	
M. capricolum subsp *capricolum*							
8086-1	1980	France	+	5	6	6	6
2002-053 (VP28L)	<2002	India	-			-	
California kid[T]	1955	USA	-			-	
96038	<1996	Greece	-			-	
90122 (C1547)	1990	Ivory Coast	-			-	
M. mycoides subsp. *mycoides* SC							
T1/44	1954	Tanzania	-			-	
94111		Rwanda	-			-	
8740-Rita	1987	Cameroon	-			-	
PG1[T]	Unknown	Unknown	-			-	
M. capricolum subsp *capripneumoniae*							
F38[T]	1976	Kenya	-			-	
GL102 (Gabes, 102 passage)	1981	Tunisia	-			-	
Gabes	1981	Tunisia	-			-	
95043	1995	Niger	-			-	
M. bovine group 7 of Leach							
PG50[R]		Australia	-			-	
D424	<1990	Germany	-			-	
9733	1993	India	-			-	
M. putrefaciens							
KS1[T]	1954	USA	-			-	

Tours 2	1972	France	-			-	
M. yeatsii							
GIH^{T}	1981	Australia	+	4	4	4	
M. cottewii							
VIS^{T}	1981	Australia	+	7	7	6	6
M. auris							
UIA^{T}	1981	Australia	-			-	

Table 2 Primers used for ISMmy*2* amplification and sequencing

Name	Position[a]	5'-3' sequence	
ISMmy2-probe-F	78-100	GAAAAGAGGAAAACAGCTAAGTG	PCR-probe
ISMmy2-probe-R	1024-1048	CTGTATGAATAATAAAGTCTTTAGG	PCR-probe
IR-ISMmy2	8-24 & 1351-1367	GGACAAAATTATTAGAC	PCR
ISMmy2-F1	942-971	AACAAARAAARTTGTTAGYTATAATTTATC	Sequencing
ISMmy2-R1	174-202	CAGAATTTCTAATTTTGCTATAAATATAA	Sequencing
ISMmy2-F	22-44	GACTCCCTAGGATAGACCCCCTA	Sequencing
ISMmy2-R	1322-1347	AAAGTTATGTTTATTATACACACCTC	Sequencing

[a] Nucleotide positions within the ISMmy*2* sequence (Acc. N°: DQ887910)

Table 3 Homologies between copy ISMmy2-A, other copies within MmmLC strain 95010 genome and with ISMmy*2* or its variant in closely related mycoplasma species. Asterisks indicate junctions between the insertion sequence and target site sequence

IS element			Coding region			
MmmLC strain 95010-C1	**Length**	**%bp identity**	**Nb ORF**	**Length**	**% identity**	**Insertion sites**
ISMmy2-A	1374	100	1	426	100	aacattaatt * gagagtagga...tcctggtttc * attacttgct
ISMmy2-B	1373	98	1	426	97.7	tccgggggt * aagagtagga...tcccggtttc * ggtctccgga
ISMmy2-C	1373	98.6	1	426	99.1	gaacatataa * gagagtagga...tcccggtttc * taatgcctaa
ISMmy2-D	1373	98.6			99.1	gagtgcataa * aagagtagga...tcccggtttc * taagcactct
ISMmy2-E	1373	98.8			99.1	aggaaaataa * gagagtagga...tcccggtttc * cattggatct
ISMmy2-F	1373	98.6			99.1	atataatcat * gagagtagga...tcccggtttc * taatgctgtt
M. mycoides cluster				426	96-99	
M. cottewii				426	85.9	
M. yeatsii				426	86.2	
MYPU_KD735-IS1138B	1288	58.5	1	402	38.3	

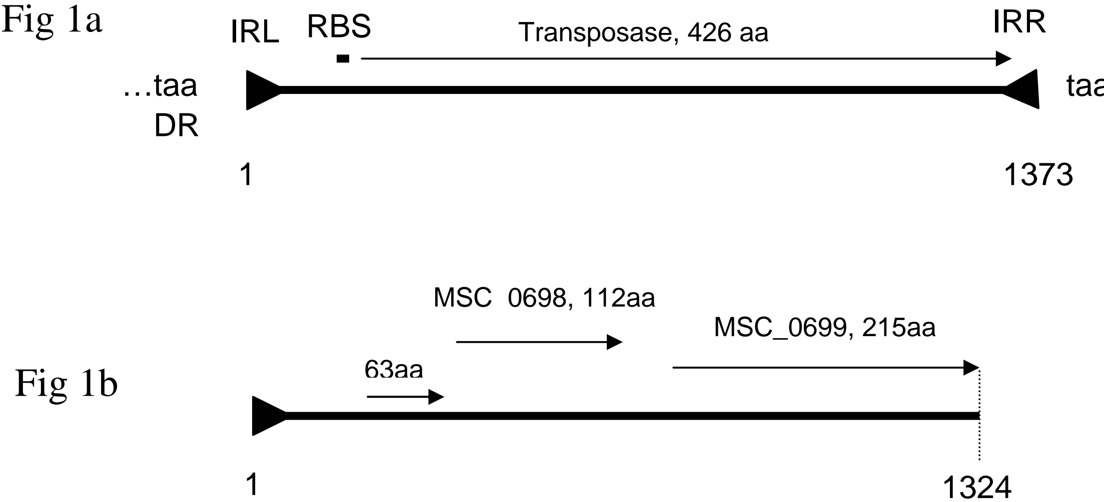


Fig. 1 Schematic representation of ISMmy*2* structure and organization.

In MmmLC, ISMmy*2* is 1373 bp long (Fig 1a). There is a single open reading frame coding for a 426aa long putative transposase with an upstream ribosome binding site (RBS). The IS is flanked by two 24 bp inverted repeats (IRL and IRR) and it generated trinucleotide direct repeats upon insertion, here "taa" in the case of ISMmy*2* C and D copies. In the PG1^{T} sequence (Fig 1b) a single remnant of an ISMmy*2* copy has been identified. It lacks the 5' end with the IRR and the former transposase CDS has been splitted in three because of frameshift mutations.

```
                  D1                   D2                               E-K
ISmmy2        LqnDIya   (64)   iHTDhGaaYsS     (35)   SriGnslDNreaEyFfsiLK

IS3 family    w--DiTy (58-60)  -HsDrGs-y-s     (35)   s--G---dN---Esf---lK
              l  l               t q
              v  v
```

Fig.2 Alignment of the putative transposase region containing the DDE-K motif of ISMmy*2* and consensus sequence of the IS3 family.

For the IS*3* family, amino acids forming part of the conserved motif are shown as large bold letters. Capital letters indicate conservation within a family, and lowercase letters indicate that the particular amino acid is predominant. The numbers in parentheses show the distance in amino acids between the amino acids of the conserved motif. (http://www-is.biotoul.fr/). For ISMmy*2*, amino acids that are identical to predominant or conserved amino acids of the IS3 family are shown in capital.

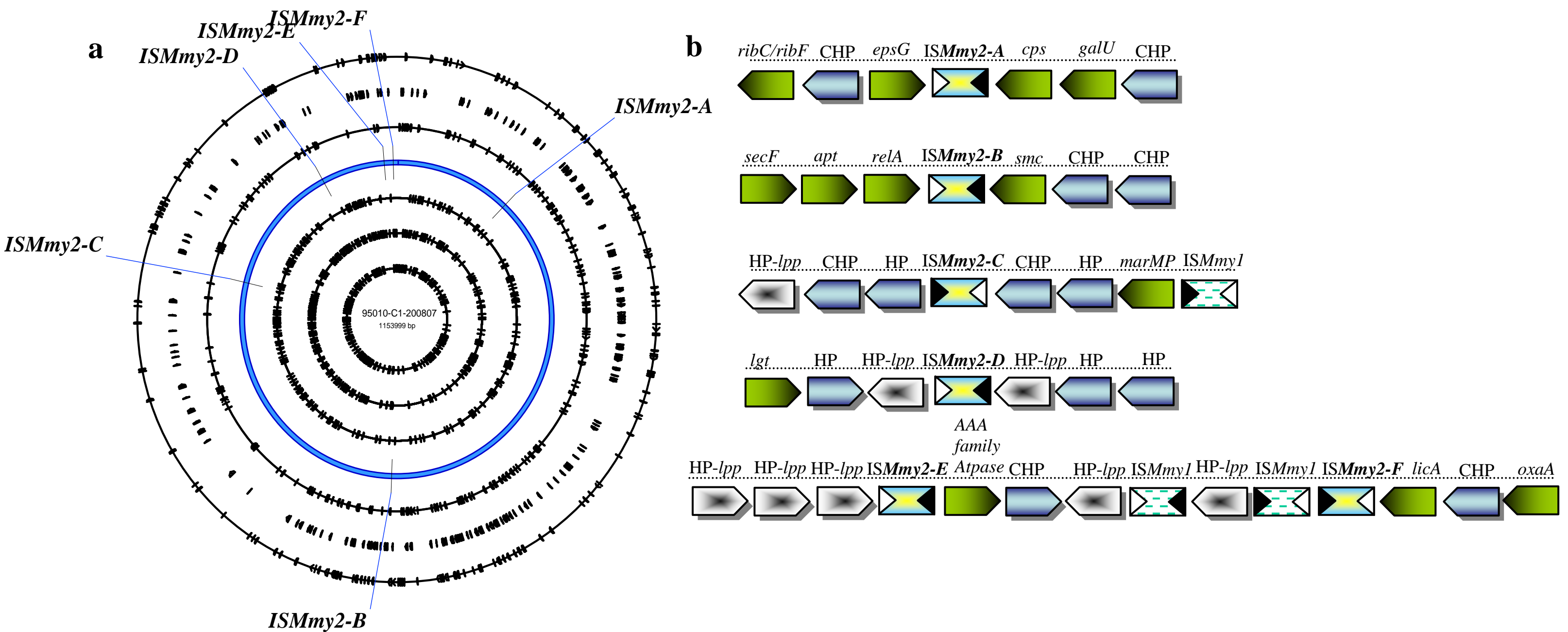


Fig.3 Genomic map of MmmLC strain 95010-C1 showing the positions of insertion sequences ISMmy2-A to -F (a), and diagram showing each target region, comprising three genes upstream and downstream from the ISMmy*2* copy (b).

Gene abbreviation: *ribC/ribF* (Riboflavin kinase / FAD synthetase, CHP (conserved hypothetical protein), *epsG (*Glycosyltransferase (EC 2.4.1.-), *cps (*Glycosyltransferase), *galU* (UTP-glucose-1-phosphate uridylyltransferase (EC 2.7.7.9)), *secF* (Conserved hypothetical protein, predicted transmembrane protein, probable protein-export membrane protein SecF), *apt* (Adenine phosphoribosyltransferase (EC 2.4.2.7)), *relA* (GTP pyrophosphokinase (EC 2.7.6.5), *smc* (Chromosome segregation ATPase), HP-*lpp* (Hypothetical protein predicted lipoprotein), HP (hypothetical protein), *marMP* (Putative C5 methylase), *lgt* (Prolipoprotein diacylglyceryl transferase (EC 2.4.99.-), *AAA family Atpase licA* (Pts system, lichenan-specific IIa component), *oxaA (*Membrane protein oxaA) Atpase licA (Pts system, lichenan-specific IIa component), oxaA (Membrane protein oxaA)

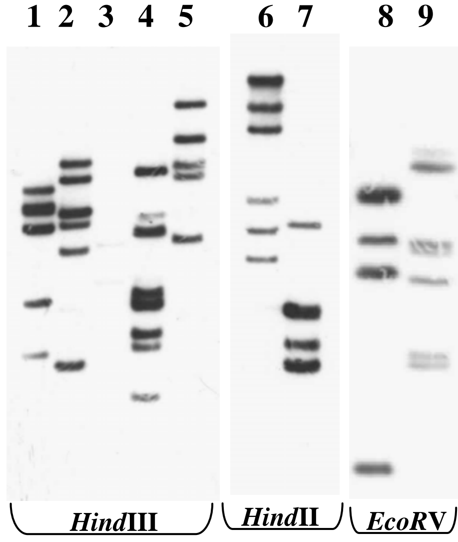
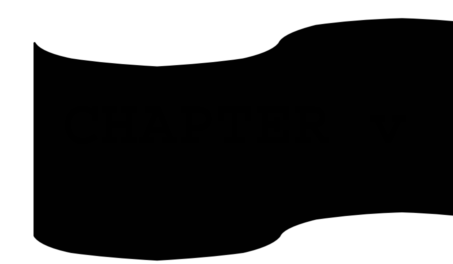


Fig. 4 Identification of multiple copies of ISMmy*2* in different mycoplasma strains.

Southern blot hybridization profiles of genomic DNA digested with *Hind*III, *Hind*II and *EcoR*V, as indicated. Lanes 1-3: *Mycoplasma mycoides* subsp. *mycoides* LC strains 95010-C1, Y-goatR and 55507-1; lanes 4-6: *Mycoplasma mycoides* subsp. *capri* strains WK354, L and N108; lanes 8 & 9: strains GIHT and VIST of *M. yeatsii* and *M. cottewii* respectively.



Transformation of MmmLC and expression of H-protein from peste des petits ruminant’s virus structural proteins

1 INTRODUCTION

Mycoplasmas have already been shown to be able to express heterologous genes (Duret et al., 2003; Dybvig et al., 2000; Knudtson and Minion, 1994). This was notably the case of Mcc, a member of the *M. mycoides* cluster (Janis et al., 2005). The main objective of this study was to assess the suitability of using mycoplasmas as vectors for the expression of genes of veterinary interest, with the final aim of developing multivalent recombinant vaccines.

Among ruminants, *Morbilliviruses* and mycoplasmas are key pathogens that are often associated and for which combined vaccination campaigns are often implemented. In bovines, vaccination campaigns against Rinderpest and CBPP have allowed the near eradication of the first and a control of the latter. This is also the case for small ruminants, where Peste des Petits Ruminants (PPR) and CCPP are considered among the major infectious diseases, although CCPP is often under-diagnosed in the field. However, goats can also be infected by a number of pathogenic mycoplasma species including MmmLC / Mmc. Many countries, especially in Asia, have tried to develop mycoplasma vaccines for MmmLC and related species.

Efficient homologous vaccines have already been developed for PPR (Diallo et al., 1989). However, the actual trend in vaccinology for PPR is to construct recombinant pox viruses bearing PPR virus (PPRV) antigens (Berhe et al., 2003; Diallo et al., 2002). The rationale for this approach is that it allows the production of multivalent vaccines (pox viruses can harbour more than one heterologous gene) that would be marked and thermo-tolerant. This approach may not be suitable for the development of mycoplasma vaccines as long as the protective immunogens have not been elucidated. So far, it is difficult to imagine that a single mycoplasma antigen, presented by a viral vector, may elicit a satisfactory protection. On the other hand, it is well known that whole mycoplasma cells may be used to elicit a good protection for the homologous disease through the use of either attenuated strains (i.e.: T1/44 strain for CBPP control) or inactivated cultures that are adjuvated (i.e.: CCPP vaccines). We hypothesized that MmmLC may be used as a vector for the expression of a PPRV antigen with the ultimate goal to obtain a multivalent vaccine.

1.1 Peste des petits ruminants (PPR)

PPR was first described in 1942 in Ivory Coast by Gargadennec and Lalanne, who identified a disease in goats and sheep that was similar to Rinderpest but which was not transmitted to in-contact cattle. This observation led to the name “peste des petits ruminants” (small ruminant plague) (Pastoret, 2006). In 1956, Mornet and collaborators showed that Rinderpest and PPR viruses were antigenically closely related and PPR virus was then classified as the fourth member of the Morbilliviruses together with Rinderpest, measles and distemper viruses (Bourdin and Laurent-Vautier, 1967).

PPR is known to be present in a broad belt of Sub-Saharan Africa, Arabia, the Middle East and southern Asia. In recent years major outbreaks in Turkey and India (Ozkul et al., 2002) and Tajikistan (Kwiatek et al., 2007) have indicated a marked rise in the global incidence of PPR.

PPR is an acute to sub acute viral disease of goats and sheep caused by PPR-virus. Goats are reported to be more susceptible than sheep but outbreaks have occurred with both species being equally affected and even with sheep being more affected than goats (Nanda et al., 1996). The mortality rate in susceptible goats ranges from 10 to 90 % and varies with the degree of innate resistance, body condition, age, virulence of the virus involved and occurrence of complications resulting from secondary bacterial and parasitic infections.

Depending on the clinical outcome, per acute, acute, sub acute and subclinical forms can be distinguished. The per acute form of the disease starts often after a short incubation period of 2 days with a sudden high rise in body temperature up to 40-42°C accompanied by serious oculo-nasal discharges, depression, dyspnoea, anorexia and constipation. The oral mucous membrane becomes congested and occasionally eroded. Affected animals develop profuse watery diarrhoea and die within 4-6 days after the onset of fever. The acute form represents the most common course of the disease. It is characterised by an incubation period of 3-4 days followed by a sudden rise in temperature, serous ocular and nasal discharges. Diarrhoea starts 2-3 days after the onset of pyrexia and is accompanied by sever dehydration, emaciation and prostration. The epithelium of the oral and nasal mucous membranes displays numerous, partially coalescing pin-point greyish necrotic foci, which leave partially demarcated hyperaemic, but non-haemorrhagic erosions after sloughing off the necrotic debris. The oral lesions are may be accompanied by profuse

salivation. Most goats die within 10-12 days after the onset of pyrexia. In sub acute forms less severe illness is observed after an incubation period of 6 days and low-grade fever. The clinical disease of PPR may be complex due to the involvement of many different secondary pathogens, frequently bacterial complications such as *Pasturella* spp. or *Mycoplasma* spp. For clinical differential diagnosis, the following diseases have to be considered: CCPP, bluetongue, contagious ecthyma, caprine and ovine poxvirus infections, foot and mouth disease and Nairobi sheep disease (Wholsein and Saliki, 2006).

1.2 Morbilliviruses

Members of the genus Morbilliviruses are classified under the family Paramyxoviridae, and are responsible for some of the most devastating diseases of humans and animals. These include measles virus (MV) affecting human beings, Rinderpest virus (RPV), PPRV, canine distemper virus (CDV) and phocine distemper virus (PDV) the latter affecting aquatic mammals. CDV infects many carnivore species including domestic dogs, mink and ferrets and can have serious consequences when endangered wildlife species are threatened (Banyard et al., 2006).

1.2.1 Structure of Morbilliviruses

Members of the Paramyxoviridae are indistinguishable by electron microscopy, where the virions are seen as pleomorphic particles with a lipid envelop enclosing a ribonucleoprotein (RNP) core. This RNP core contains the genome, a single strand of negative polarity RNA, encapsulated by a nucleocapsid protein. All *Morbilliviruses* share the same genome organization although their RNA length differ slightly, each being just under 16 Kbp. Their genomes are organized into six contiguous, non-overlapping, transcriptional units which encode six structural proteins, namely the nucleocapsid (N), the phosphoprotein (P), the matrix (M), the fusion (F), the haemagglutinin (H), and the large polymerase (L) protein, the latter being the viral RNA dependent RNA polymerase (RdRp). The three viral structural proteins N, P and L are internal polypeptides complexed with viral genome to form the nucleocapsid, while M, F and H forms the virus envelop (Figure V. 1a, Sibylle and Volker ter, 2002). The morbilliviruses produce two glycoproteins which are embedded in the viral envelop and protrude as spikes, the F and H-protein spikes. The F protein enables the virus to penetrate the cell by mediating the fusion of the viral and cellular membranes at the cell surface (Moll et al., 2002). The H

protein enables the virus to bind to the cell receptor, the first step in the process of infection (Plemper et al., 2000).

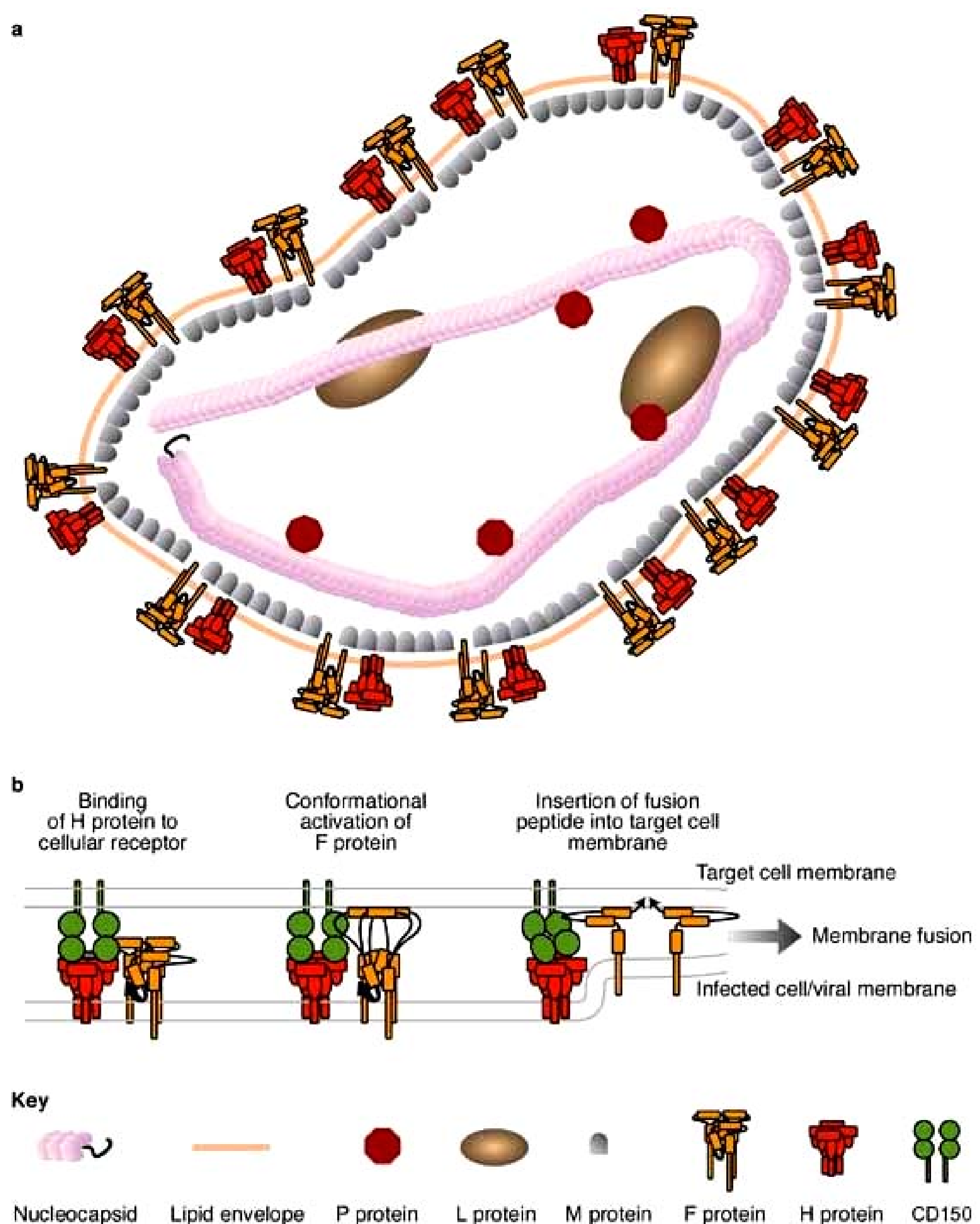


Figure V. 1 Schematic representation of the structure of *Morbillivirus*.

This illustration represents the two surface proteins, haemagglutinin (H) and fusion protein (F), together with the nucleocapsid protein (NP), phosphoprotein (P), matrix protein (M), and polymerase protein (L); This schematic representation was designed to show measles virus particle and mechanism of membrane fusion; adapted from expert review in Molecular medicine©2002 Cambridge University Press.

Interaction of the H-protein with cellular receptor (CD150) triggers a conformational change with F-protein leading to insertion of its N-terminus into the target cell membrane arrowheads in Figure V. 1b (Sibylle and Volker ter, 2002). Both H and F glycoproteins are the major protective immunogens and are responsible for the induction of neutralizing antibodies (Yilma et al., 1988). Because of the immunogenic nature of both F and H proteins, they have been used as effective recombinant vaccines using attenuated Capripox virus as expression vector (Romero et al., 1994a; Romero et al., 1994b). The H and F proteins of several morbilliviruses have been expressed in various vector systems and have been used as effective sub-unit vaccines (Diallo et al., 2007).

1.2.2 H-protein

The H-protein from PPRV has 609 amino acid residues and uses a terminal UGA (W) as a stop codon (Barrett et al., 2006). Structurally H protein has only one 23 aa-long hydrophobic domain near the N-terminus (position 35-58), which acts as a signal peptide and is used to anchor the protein to the target cell membrane. The 23 aa-long transmembrane helix of the H protein is remarkably similar to the 22 aa-long transmembrane helix membrane proteins found in MmmLC, as evidenced during genome sequence annotation of MmmLC strain 95010-C1. The N-terminal 34 aa remain in the cytoplasmic side of the membrane, while the C-terminus is extruded to the outside. This arrangement, with an N-terminal anchoring and with an external C-terminal domain classifies H as a type II glycoprotein (Plemper et al., 2000). More recent studies have shown that PPRV-H have haemagglutinin-neuraminidase (HN) activity (Seth and Shaila, 2001), constituting the only morbillivirus H-protein that resembles the HN protein of other paramyxovirus genera (Barrett et al., 2006).

The ultimate aim of this study was to provide a model for the production of a recombinant vaccine to protect animals from both MmmLC infections and PPR. The first objective was to evaluate the ability of MmmLC strain 95010-C1 to be transformed and express a heterologous gene. For this aim, the H-gene from PPRV vaccine strain Nig75/1 was chosen as a heterologous gene. This study required the choice of a plasmid vector and of appropriate promoter sequences for mycoplasma transformation and heterologous gene transcription respectively.

2 MATERIALS AND METHODS

2.1 Heterologous gene

A haemagglutinin gene (H-gene) from peste des petits ruminants virus (PPRV) was used as a heterologous gene of interest for this study.

2.1.1 Codon usage analysis

Graphical codon usage analyser a web based software (gcua v.1) http://gcua.schoedl.de/ http://www.kazusa.or.jp/codon/ (Fuhrmann et al., 2004) was used for the analysis of codon usage in H-gene. Codon usage of H-gene was then compared with the available mycoplasma data, the genome sequence of *M. genitallium*. Sequences from previously expressed heterologous genes such as *tetM* and *lacZ* were also analysed simultaneously.

2.2 Promoter sequence signal analysis

Promoter sequences of CDSs of IPF140, IPF221, IPF453, IPF989 and IPF1004 from the genome sequence annotation of strain 95010-C1 and promoter sequence of spiralin gene (Acc no. AF012877), were analysed for promoter signals using bacterial promoter prediction program (BPROM) (http://www.softberry.com). Based on the result of the analysis primers were designed for the amplification of promoter sequence from the genome sequence of 95010-C1.

2.3 Plasmids

Plasmid pCJ6 (Annex 2), containig the spiraline promoter (Janis et al., 2005), was used for amplification of the spiralin promoter, awaiting pMYCO1 availability. The *oriC* plasmid pMYCO1, containing the *ori*C of MmmLC strain Y-goat^{T} (Lartigue et al., 2003), was used as a cloning vector. Plasmid pH78-2 (Annex 1, pBacPAK9 containing PPRV-H of Nig75/1) (Diallo et al., 2002) was used for H-gene amplification. This vector's sequence was used for the design of sequencing primers. Additionally, plasmid pCR®2.1-TOPO® (Invitrogen, USA) was used as an intermediate cloning vector. Table V.1 provides the lists of primers used in this study.

2.4 Choice of restriction enzymes for cloning

The choice of restriction enzymes for cloning purposes was made using Vector-NTI Suite software (Invitrogen, USA).

Table V. 1 List of primers used for amplification, cloning and sequencing of promoters and foreign gene

Primer Name	Sequence (5'-3')	Origin	REA site
prom-CHP-F	TACTGGTTCAAGACGTCTTGGAAATCCAAGTAAT	95010-C1	*Aat* II
prom-CHP-R	CAATAGGTATTTACGCGTTTATAAATTCTCC		*Mlu* I
prom-SRm-F	TAAAACAATGACGTCTAAAGCTTTAGTG	pCJ6	*Aat* II
prom-SRm-R	GTGATTTTACGCGTTTCAAAGG		*Mlu* I
HF	AATACGGATACGCGTATGTCCGCACAAAGG	pH78-2	*Mlu* I
m2-HR	AATGGTTGCAGGCCGGCTTAGACTGGATTACATG		*NgoM*IV
HP73R	TTGACCACAGAGATCAGAGG	pH78-2	
Pg7848F	CGTGTCCTCAGTGTTTACCGTAGTCGAAG	pH78-2	
HP112R	GGTGATGAAGTCGGCATCAG	pH78-2	
insert-pmyco-F	ATGTTGAATACTCATACTCTTCC	pMYCO1	
Insert-pmyco-R	GAGCAAGAATGGCTAGGATCCCC		
Mut-F	GGAGAATTTATAAACGCGTATGTCCGCACAAAGGGAAAGGATC		*Mlu* I
Mut-R	GATCCTTTCCCTTTGTGCGGACATACGCGTTTATAAATTCTCC		
Seq-H-R1	AGTCTCAGATGGGGTGCATAGGG	pH78-2	
Seq-H-R2	GACTGGGTAGAAGTAAGATGATGAG	pH78-2	
Seq-H-R3	GTTGAGGAACTTAATCTTATCGGAG	pH78-2	

2.5 PCR conditions

Pfu DNA polymerase (Stratagene) that has a proof reading activitiy was used for the PCR amplifications of H-gene and promoter sequences. The reaction volume was fixed at 50 µl with: 5 µl of 10Xpfu buffer, 1.5 mM $MgCl_2$, 0.3 mM dATP and dTTP, 0.15 mM dCTP and dGTP (dNTPs from Roche), 100 ng/µl DNA template, 100 ng/µl of each primer and 2.5 U/µl of pfu DNA polymerase. The PCR conditions for both promoter amplifications were consisted of 40 cycles of 94°C for 45 sec, 55°C for 45 sec and 72°C for 1:50 min. For the amplification of H-gene the same PCR conditions were used excepting the elongation time which was 3 min.

2.6 Cloning procedure

The summary of steps followed for the cloning strategy is shown in Figure V. 4. The mutagenesis reaction in the cloning step was performed by Quick Change II XL (Stratagene), the reaction mix consisted of 5 µl of 10X pfu buffer, 1 µl of 10 ng plasmid template, and 125 ng of each primers (Mut-F / R), 1 µl dNTP (Stratagene), 3 µl of Q-solution (Stratagene), 2.5 U of pfu DNA polymerase (Stratagene) and H_2O to a final volume of 50 µl. The PCR condition consisted in an initial denaturation step of 95 °C for 1 min followed by a total 18 cycles of 95°C for 50 sec, 60°C for 50 sec and 68 °C for 9 min. PCR product was *Dpn*I digested and used to transform competent cells.

2.7 Plasmid propagation and amplification

For plasmid propagation and amplification two chemically competent *E. coli* strains Top10 and *Stbl*3 were used. In parenthesis is the genotype of each competent cells; Top10 (F- *mcr*A Δ (*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ (*araleu*) 7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G) and *Stbl*3 (F- *mcr*B *mrr hs*dS20 (r_B^-, m_B^-) *rec*A13 *sup*E44 *ar*a-14 *gal*K2 *lac*Y1 *pro*A2 *rps*L20 *(Str*^{r}) *xy*l-5 λ^- *leu mtl*-1. All *E. coli* strains were grown in Luria-Bertani medium. Antibiotics kanamycine 50 µg ml^{-1} was added for the selection of plasmid pCR®2.1-TOPO® (Sambrook et al., 1989).

2.7.1 Plasmid Mini-preps and Maxi-preps

QIAprep Spin Miniprep Kit and QIAGEN Plasmid Maxi Kit (QIAGEN) were used for mini-prep and for maxi-prep of plasmids. The same procedures were followed as described by the manufacturers' instruction booklet.

2.8 Mycoplasma transformation

Mycoplasma cells from strain 95010-C1 were cultivated in modified Hayflick media containing sodium pyruvate and glucose in a high security L_3 laboratory. Transformation of strain 95010-C1 was made by the polyethylen glycol (PEG) mediated procedure similar to that used for the transformation of gram positive bacterial protoplasts as recommended by (King and Dybvig, 1991). Exponential phase 5 ml cultures of MmmLC strain 95010-C1 were centrifuged at 10000 g for 10 min at 4 °C. The pellets were then resuspended in 250 µl of 0.1 M $CaCl_2$, and held on ice for 30 min. During the incubation time, 10 µg of yeast tRNA (Invitrogen) and 10 µg of plasmid were mixed and transferred to the cell suspension. Immediately after 2 ml of 70 % PEG solution (70 gm PEG 8000, 10 ml of

0.01 M trisHCl, pH 6.5 heated to 70-80°C and adjusted to 100 ml with double distilled water) was added to the cells and incubated for 2 min at room temperature. After 2 min of incubations 10 ml of S/T buffer (85.56 g Sucrose mixed with 5 ml of 1 M TrisHCl to a final volume of 500 ml double distilled water, pH 6.5) was added and the contents were mixed gently. This solution then centrifuged for 12 min at 12,000 g and 20 °C, this step was repeated to completely eliminate PEG from the cellular mix. Mycoplasma pellets free of PEG were then resuspended in 300 µl of modified Hayflick medium and were incubated for 2 hrs at 37°C. After 2 hrs of incubation cultures were diluted from 10^{-1} - 10^{-4}, and 200 µl of these dilutions and the non diluted culture were plated on modified Hayflick medium containing 2 µg/ml tetracycline. For the calculation of viable mycoplasma cells, 20 µl of these dilutions were plated on modified Hayflick media without tetracycline.

2.9 Immunoblotting

In order to screen the presence or absence of H-protein (immunodominant), 1 µl aliquot of concentrated culture dropped on a nitrocellulose membrane. Membrane then dried for a few minutes then blocked for 1 hr in 1X blocking solution (10X Blocking Solution, Sigma). Membrane incubated for additional 1 hr at room temperature in a goat hyper-immune anti-H-protein diluted 1: 50 in 1X blocking solution. Washed three times for 10 min, it was then incubated for 30 min in 1: 200 ant-goats IgG conjugated to horse radish peroxidase. Membrane washed three times for 5 min in washing buffer 1 (0.1 M Tris-HCl pH 7.4, 0.15 M NaCl, 0.3 % Tween 20) and two times in washing buffer 2 (0.1 M Tris-HCl pH 9.5, 0.1 M NaCl). Membrane revealed using DAB (3-3' diamino-benzidine, Sigma) colorimetric detection for 2 min and rinsed in double distilled water.

3 RESULTS AND DISCUSSION

The PPR-H gene was chosen as heterologous gene to be expressed in the sequenced MmmLC strain 95010-C1. This choice was not only based on the immunogenic properties of the encoded H-protein but, it was also supported by its codon usage. The replicative plasmid pMYCO1 (Lartigue et al., 2003), which has already been used successfully for heterologous gene expression in mycoplasmas (Janis et al., 2005), was chosen as a cloning vector. In the latter studies, transcription was controlled by the spiralin promoter, which was used as a control for this work. Another promoter candidate was chosen based on promoter signalling scores obtained using dedicated software for bacterial promoter signal analysis.

3.1 Codon usage in the PPR-H gene

The codon usage of a 1830 bp-long nucleotide sequence of the H-gene was compared with the codon usage of *M. genitalium* genome using the graphical codon usage analysis (gcua v.1) software. The analysis resulted in a mean difference of 17.05 %. However, the H gene had 47 % GC content, which is higher than the average found on mycoplasma sequences. Codon usage analysis of the 1932 bp gene *tetM* (GC % = 36 %) and *lacZ* (GC % = 56 %), which have already been expressed in mycoplasma showed a mean difference of 14.95 % and 36.56 % respectively as compared to the *M. genitalium* genome. The mean codon usage difference of H-gene was lower than that of other previously expressed genes (i.e.: *lacZ* expression in Mcc, Janis et al., 2005).

3.2 Choice of promoter sequences

From the genome sequence annotation of MmmLC strain 95010-C1 five promoter sequences of CDSs from IPF140, IPF221, IPF453, IPF989, IPF1004 were analysed (Table V. 2). These CDSs encode: conserved hypothetical protein (CHP), nitrogen fixation protein, oligopeptide ABC transporter-permease component, 30S ribosomal protein S13 and leucyl-tRNA synthetase, respectively. Upstream intergenic sequences including the putative ribosomal binding site (RBS, Shine-Dalgarno sequence) were analysed for promoter sequence signal using bacterial promoter prediction program (BPROM) http://www.softberry.com/berry. Based on score results, CHP promoter sequence from locus MLC_3560 of MmmLC strain 95010-C1 was chosen for our expression vector construct. The CHP of locus MLC_3560 is a protein of unknown function that is also found with 100 % homology in the genome sequence of MmmLC strain GM12b. A

simultaneous analysis was done for promoter spiralin which was destined to be used as a second promoter for H-gene. The mycoplasmal -10 region (Pribnow box) and, to a lesser extent, the -35 region resembled eubacterial promoter consensus sequence recognized by the RNA polymerase (Razin et al., 1998).

Table V. 2 BPROM promoter sequence analysis of genes from the genome sequence of MmmLC

IPF & Locus Tag	CDS	Upstream sequence length analyzed	No. of predicted Promoters	promoters (position from ORF)	Promoter Sequence -	score
140 MLC_3560	Conserved hypothetical protein	293 bp	1	-10 box (42)	tggtataat	94
				-35 box (63)	ttgtat	42
221 MLC_4800	Nitrogen fixation protein NifS	515 bp	2	-10 box (113)	atttaaaat	74
				-35 box (137)	atacca	8
				-10 box (437)	taataaaat	61
				-35 box (464)	ttgtaa	47
453 MLC_1560	Oligopeptide ABC transporter, permease component	527 bp	2	-10 box (169)	tataaaaat	36
				-35 box (329)	tttaaa	41
				-10 box (468)	ttcaattat	31
				-35 box (490)	ttaaaa	37
989 MLC_6730	30S ribosomal protein S13	500 bp	2	-10 box (36)	aaggataat	39
				-35 box (63)	ttactc	8
				-10 box (446)	gtgcattat	44
				-35 box (471)	ctgaca	31
1004 MLC_6610	Leucyl-tRNA synthetase	524 bp	2	-10 box (92)	ttttatact	76
				-35box (116)	ttgaca	66
				-10 box (458)	atgaaaaat	39
				-35 box (481)	tttaga	24

For promoter sequence CHP (locus MLC_3560) the BPROM software predicted a single promoter of -10 box at position 42 (TGGTATAAT) and -35 box at position 63 (TTGTAT) from the ORF start with scores corresponding to 94 and 42 respectively (Figure V. 2). The 245 bp sequence upstream from the spiralin gene (Acc no. AF012877) was also analysed. A single promoter sequence was predicted with -10 box at positions 168 (ATGTACAAT) and the -35 box at position 190 (TTGTTT) from the ORF with scores corresponding to 66 and 40 respectively (Figure V. 3). Both -10 and -35 boxes from the CHP promoter gave higher scores than the promoter sequence of the spiralin gene.

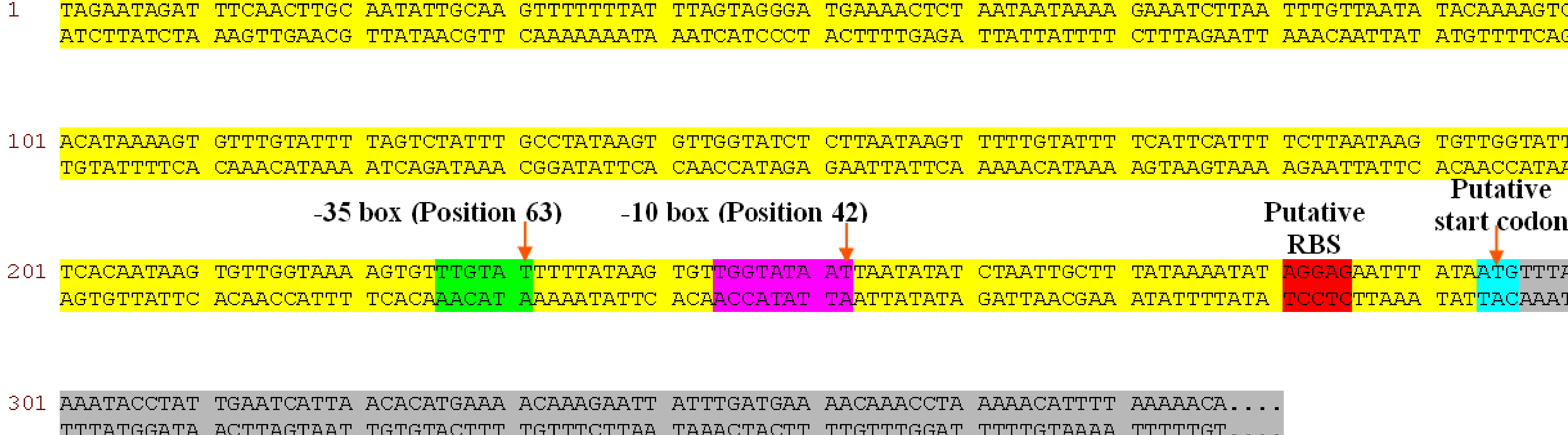


Figure V. 2 Promoter sequence of CHP of locus tag MLC_3560 from the genome sequence of MmmLC strain 95010-C1

Sequence region highlighted in yellow is the region analysed for promoter sequence signal using bacterial promoter prediction program (BPROM). The putative start codon is highlighted in blue down stream highlighted in grey is the sequence of CHP gene. 8 nucleotides upstream of the start codon is putative ribosomal binding site (RBS). The -10 and -35 boxes are located at position 42 and 63 respectively from the ORF.

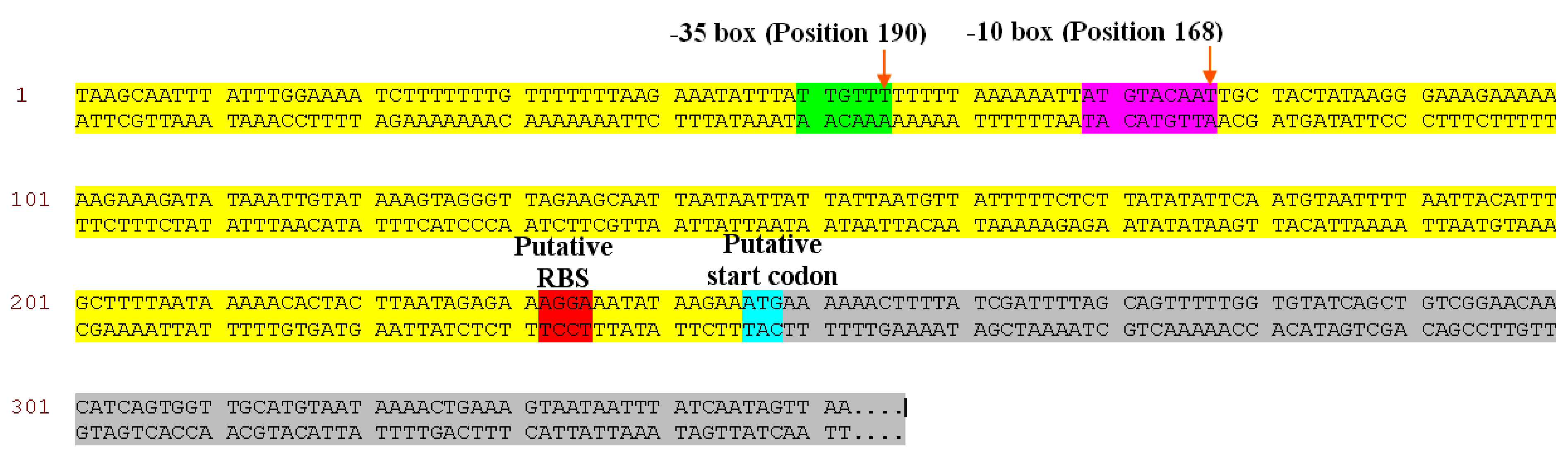


Figure V. 3 Promoter sequence of spiralin gene Acc no. AF012877.

Sequence region highlighted in yellow is the region analysed for promoter sequence signal using bacterial promoter prediction program (BPROM). The putative start codon is highlighted in blue down stream highlighted in grey is the sequence of spiralin gene. 10 nucleotides upstream of the start codon is the putative ribosomal binding site (RBS). The -10 and -35 boxes are located at positions 168 and 190 respectively.

3.3 Choice of restriction enzymes for cloning purposes

Sequences from promoters, H-gene and plasmid pMYCO1 were analysed for restriction enzyme sites using Vector NTI-Suite software. Three enzymes (*Aat*II, *Mlu*I and *NgoM*IV) were considered for cloning purposes. Each of these three enzymes generates protruding termini, which allow directional cloning (Sambrook et al., 1989). Sequences from promoters, H-gene and pMYCO1 were primarily analyzed for common non-cutting enzymes. Enzyme *Mlu*I was chosen among the few non-cutting enzymes. Enzymes *Aat*II and *NgoM*IV, which linearized plasmid pMYCO1 (have a single restriction recognition sites) and did not display a restriction recognition site in either the promoter sequences or the H-gene, were also selected for the cloning strategy (Figure V. 4). *In-silico* double enzyme digestion of pMYCO1 by *Aat*II and *NgoM*IV removes a truncated portion of the F1 origin of replication, present in pMYCO1, without any detrimental effect on this plasmid. The origin of DNA replication of bacteriophage, F1 origin is used to produce single-stranded DNA (Dotto et al., 1982), which was not a feature required in this study.

Primers containing enzyme recognition sites at their extrimities for *Aat*II, *Mlu*I and *NgoM*IV were therefore designed inorder to amplify promoters and H gene (Table V. 1). For both promoters, forward primers were designed containing an *Aat*II restriction site, whereas reverse primers were designed containing an *Mlu*I restriction site. The promoter sequence amplicon was meant to possess the putative Shine-Dalgarno sequences for efficent H-gene translation. For H-gene amplification, the forward primer was designed with an *Mlu*I restriction site and included 12 bp downstream from the ATG start codon, whereas the reverse primer contained the restriction site *NgoM*IV and a site mutagenesis reverting the stop codon TGA to TAA (Table V. 1). In mycoplasma, the universal stop codon TGA codes for tryptophan, a feature shared by mitochondria (Razin et al., 1998).

These primers were then used to amplify the spiralin (prom-SRm F/ R) and CHP (prom-CHP F/ R) promoters and the H-gene (HF/ m2-HR) (Figure V. 5).

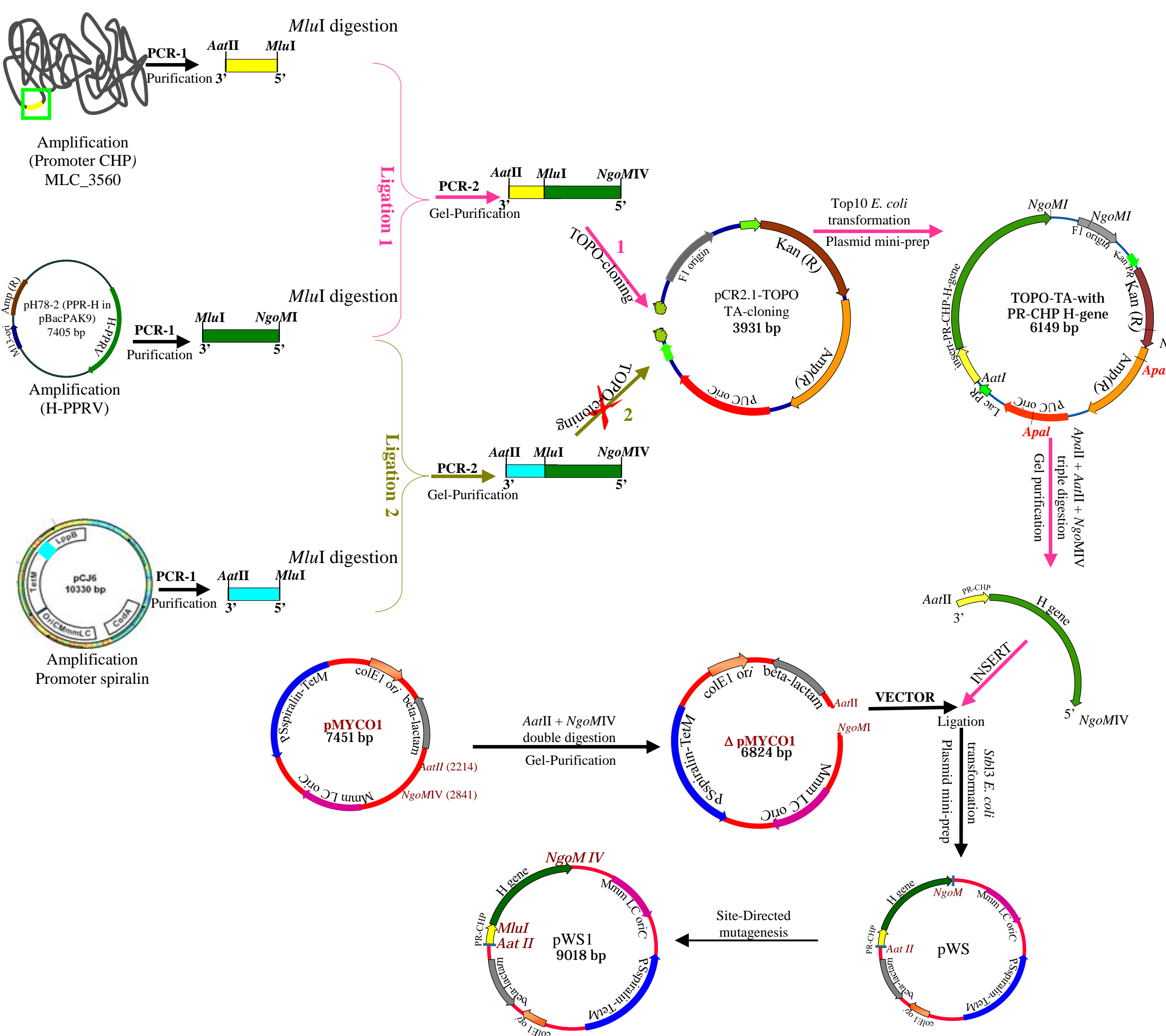


Figure V. 4 Flow diagram representing the cloning strategy used for final plasmid pWS1 construction.

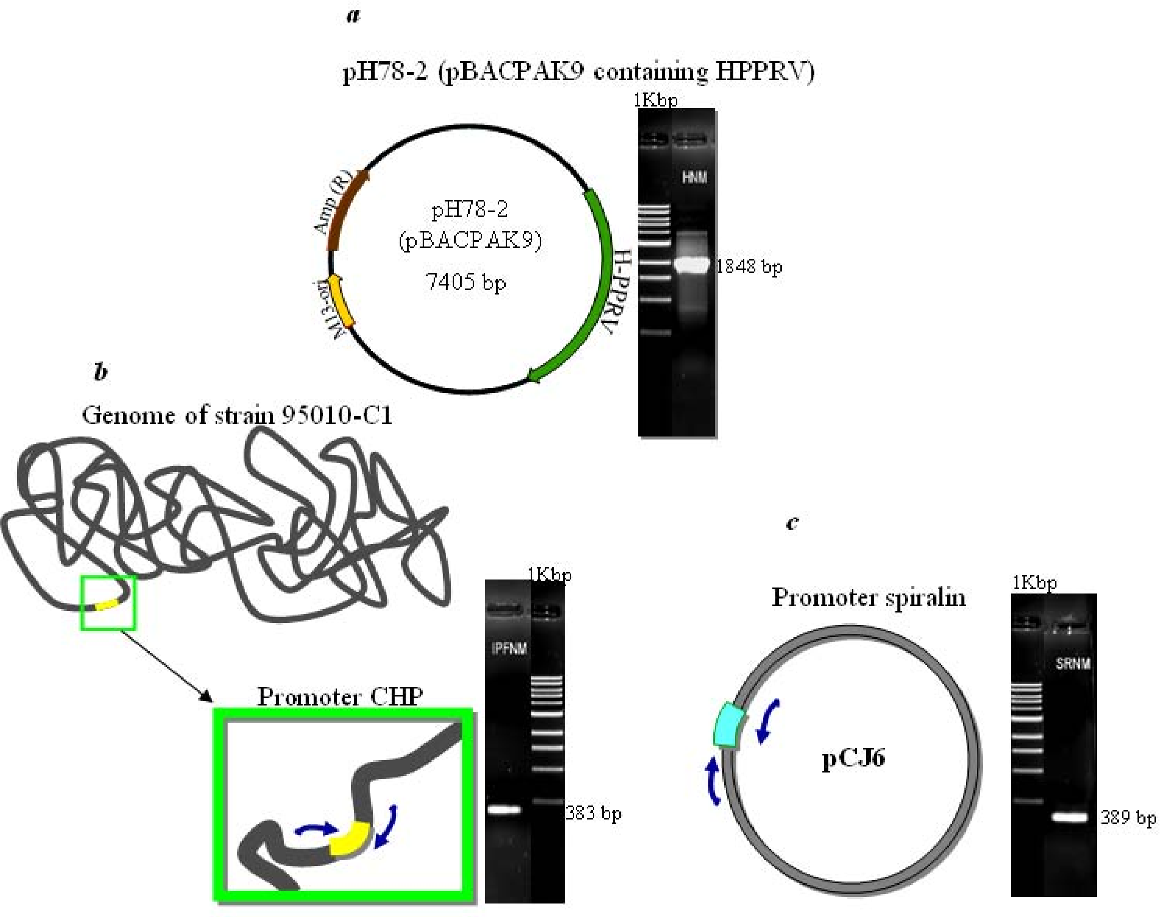


Figure V. 5 PCR amplifications of H-gene and promoter sequences.
H-gene from plasmid pBacPAK9 ***(a)***; promoter CHP of locus tag MLC_3560 from the genome of MmmLC strain 95010-C1 ***(b)***; and promoter spiralin from plasmid pCJ6 adapted from Janis et al. (2005) ***(c)***.

3.4 Promoter and H-gene ligations in an intermediate vector

Amplicons obtained from the H-gene and the two promoters were digested using restriction enzyme *Mlu*I. The digestion products of the promoter sequences were T4 DNA ligated separately with the digestion product obtained from H-gene. The two ligation products then re-amplified using the forward primer either of the two promoters and the reverse primer from the H-gene (PCR-2 in Figure V. 4). The PCR products giving the right

fragment size of about 2.2 Kbp were gel purified and used for cloning (Invitrogen gel purification Kit) (Figure V. 6).

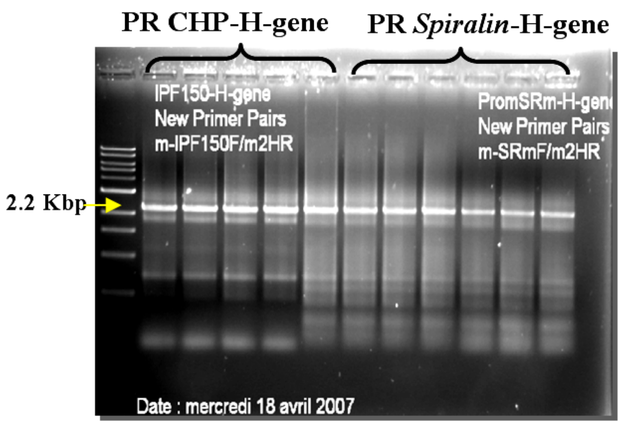


Figure V. 6 PCR amplification of ligation product of promoter sequences with H-gene. Fragment PR-CHP-H-gene (amplified by prom-CHP-F/ m2HR) and PR-spiralin-H-gene (amplified by prom-SRmF/ m2HR) at 2.2 Kbp, were gel extracted for further cloning.

Gel purified fragments PR-CHP-H-gene and PR-spiralin-H-gene were cloned in pCR®2.1-TOPO® vector (Invitrogen, USA). Ligation products were then used to transform *E. coli* Top10 competent cells with blue and white screening (Annex 4). White colonies were PCR tested using universal primers M13 and M13R for the right insert size. Out of 15 colonies analysed for both PR-CHP-H-gene and PR-spiralin-H-gene inserts, only two colonies from PR-CHP-H-gene were found to contain the right fragment size (Figure V. 7). None of the colonies tested for insert PR-spiralin-H-gene gave the right fragment size. Following two failed attempts the plasmid construct containing the spiralin promoter sequence was abandoned.

A plasmid miniprep (Qiagen, kit) was performed from the two positive colonies containing PR-CHP-H-gene. The PR-CHP-H-gene fragment was then excised from the pCR®2.1-TOPO® vector via triple enzymatic digestion of the plasmid using enzymes *Apa*lI, *Aat*II and *NgoM*IV. Enzymes *Aat*II and *NgoM*IV were used to separate PR-CHP-H-gene fragment. The third enzyme *Apa*lI was used to avoid the generation of DNA fragment of the vector having the same size as that of PR-CHP-H-gene. Finally, fragment PR-CHP-H-gene of around 2.2 Kbp was gel purified and used for plasmid construction (Figure V. 7).

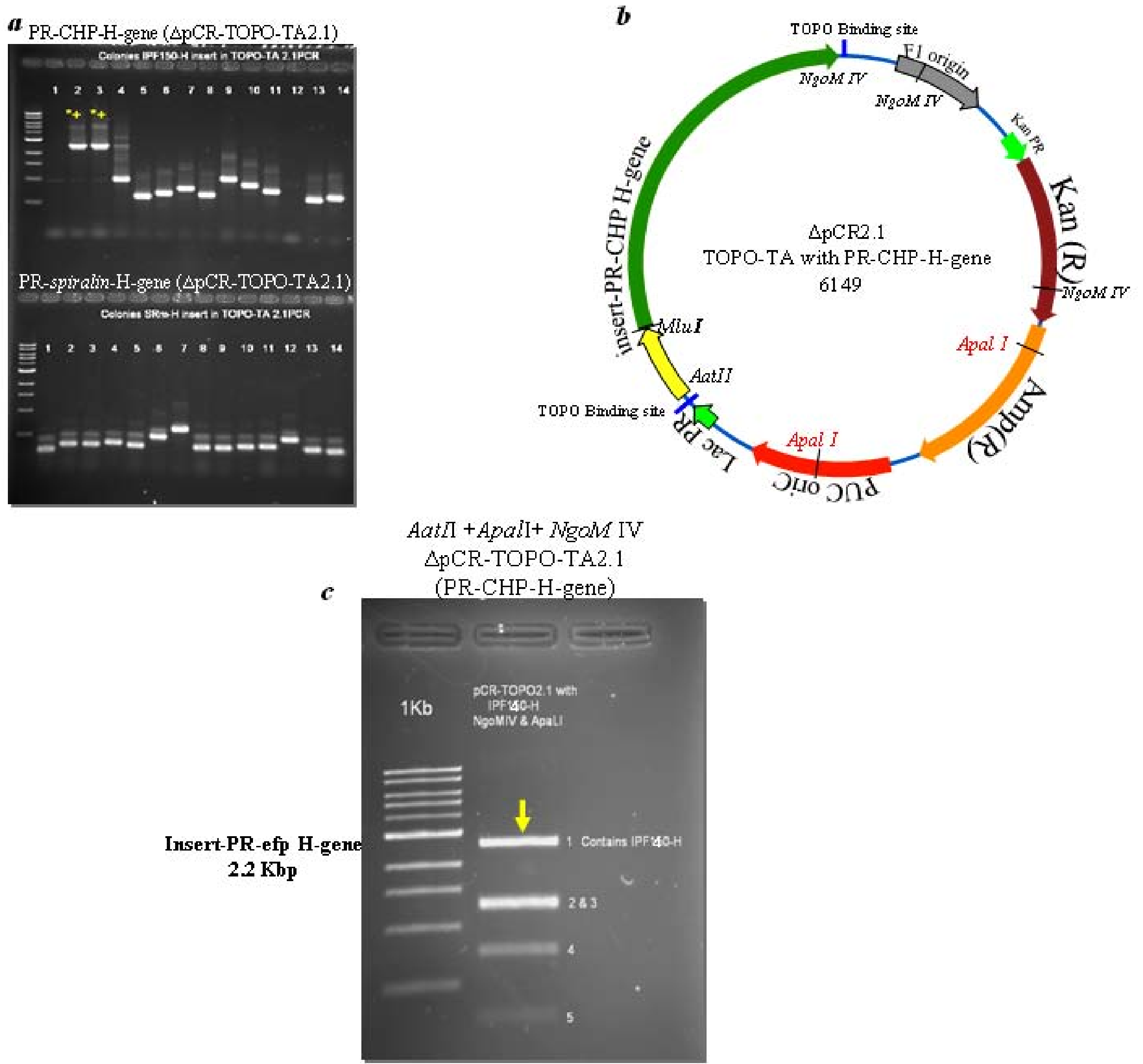


Figure V. 7 pCR®2.1-TOPO® vector containing inserts PR-CHP-H-gene and PR-spiralin-H-gene. PCR tests of white colonies revealed only colonies 2 and 3 from PR-CHP-H-gene inserts above gave the right fragment size, none of the PR-spiralin-H-gene gave the right fragment size ***(a)***; structure of ΔpCR-TOPO-TA2.1 containing PR-CHP-H-gene from the two positive colonies ***(b)***; triple enzymatic digestion of ΔpCR-TOPO-TA2.1 freeing the 2.2 Kbp PR-CHP-H-gene fragment ***(c)***.

3.5 Preparation of pMYCO1 for cloning

The 7451 bp plasmid pMYCO1, which contains the tetracycline resistance gene *tetM* as selective marker and the *oriC* of MmmLC strain Y-goat^{T}, has been described previously as one of the recent artifical *oriC* vectors for use in the *M. mycoides* cluster (Lartigue et al., 2003). Before inserting the PR-CHP-H-gene fragment, pMYCO1 plasmid was double digested by restriction enzymes *Aat*II and *NgoM*IV. The double digest gave two fragments of 6824 and 627 bp (Figure V. 8). The smaller DNA fragment contains a truncated portion of the F1origin and was excluded from final plasmid construction. The 6824 bp DNA fragment named ΔpMYCO1 was gel-purified and used as a vector for insertion of the PR-CHP-H-gene.

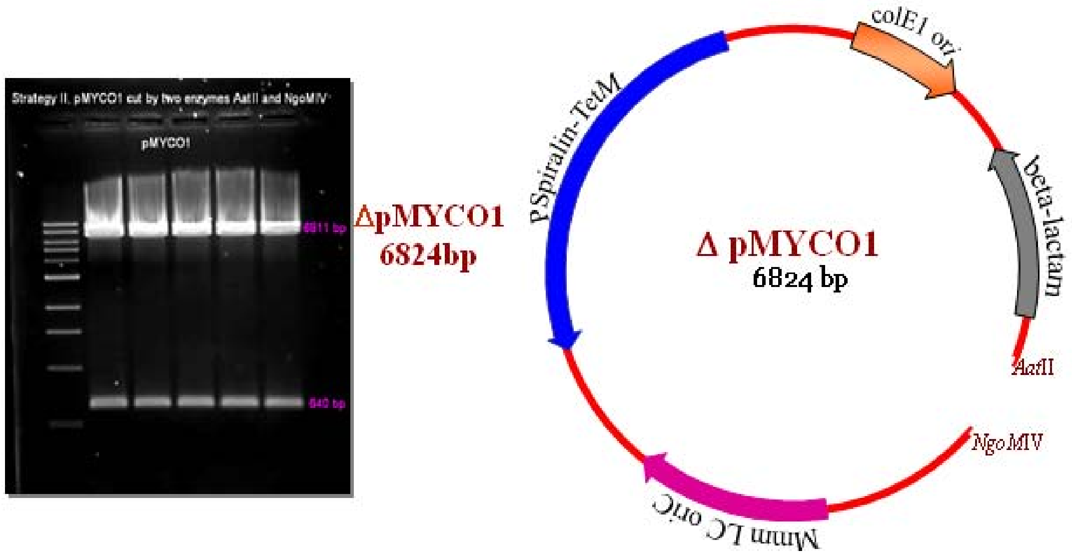


Figure V. 8 Double digest of pMYCO1 by enzymes *Aat*II and *NgoM*IV.

This has removed a 627 bp fragment containing a truncated portion of F1 origin, modified plasmid ΔpMYCO1 was then used for further cloning.

3.6 PR-CHP-H-gene construct in ΔpMYCO1

Plasmid pWS was obtained by inserting the ~2219 bp PR-CHP-H-gene fragment by directional cloning using the *Aat*II and *NgoM*IV restriction sites of plasmid ΔpMYCO1. Ligation produts of vector ΔpMYCO1 and insert PR-CHP-H-gene were used to transform *Stbl*3 competent cells (Invitrogen) under tetracyclin selective pressure.

PCR was used to verify the presence of the PR-CHP-H-gene insert. Two couples of primers were used for this analysis (Table V. 1): 1) primers designed on vector sequences flanking the insert (insert-pmycoF and m-insert-pmyco-R) and 2) a forward primer from the H-gene (Pg7848F) and the same reverse primer from the vector (m-insert-pmyco-R). Positive PCR results were obtained with both primer pairs. PCR using the first primer pair resulted in amplification a 2700 bp product and a 1827 bp fragment was obtained using the second primer pair (1 & 2 in Figure V. 9).

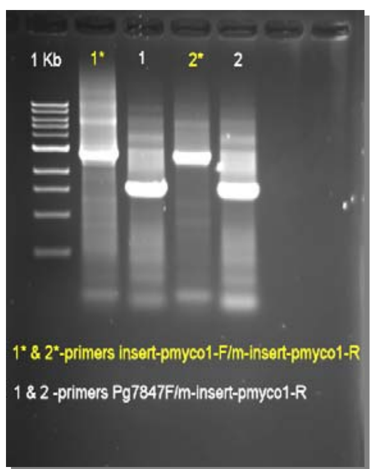


Figure V. 9 PCR amplification of pWS from *E. coli* transformed colonies.
1 & 2 in yellow using insert-pmycoF/m-insert-pmyco-R and 1 & 2 in white using a forward primer from the H-gene (Pg7848F) and the reverse primer (m-insert-pmyco-R); Bands of ~2700 bp and 1827 bp were obtained respectively.

3.7 Validation of the PR-CHP-H-gene insertion sequence in pWS

The 2700 bp PCR fragment containing the PR-CHP-H-gene insert was sent to Cogenics (Meylan, France) for DNA sequencing. Primers insert-pmycoF, Pg7848F, HP73R, HP112R, Seq-H-R2, Seq-H-R1, Seq-H-R3 and m-insert-pmyco-R (Table V.1) were used to obtain the entire 2700 bp sequence.

Awaiting sequencing results, a cloning simulation was conducted using Vector NTI-Suite software. A simulation was obtained by inserting *in-silico* the PR-CHP-H-gene sequence fragment into the *Aat*II and *Ngo*MIV sites of plasmid ΔpMYCO1 (Figure V. 10).

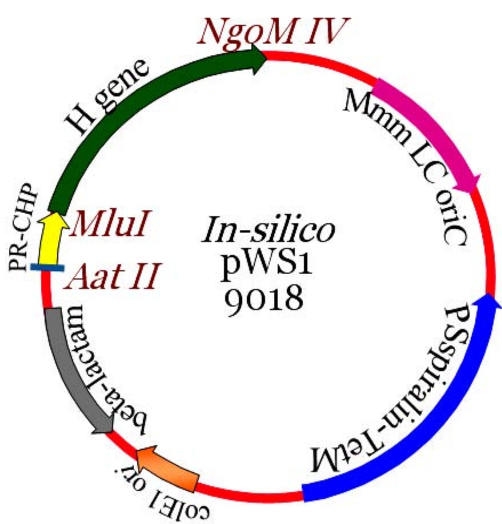


Figure V. 10 *In-silico* pWS construct containing PR-CHP-H-gene sequence.

Once the entire sequence was obtained, the *in-silico* construct was used for sequence alignment and verification of the insert PR-CHP-H-gene sequence, as actually present in plasmid pWS.

Sequence alignment (AlignX Vector NTI Suite, Invitrogen) of 2700 bp from pWS with the *in-silico* construct revealed four sites presenting mutations in pWS (Figure V. 11). Mutation site 1 was localized at the junction between the PR-CHP sequence and the H-gene sequence, presenting the insertion of three consecutive bases before the initiation codon and one base right after the H-gene start codon (Figure V. 11). Mutaion sites 2, 3 &

4 present single nucleotide substitutions within the H-gene sequence (Figure V. 11). The ‘TCG’insertion at mutation site 1 resulted in disappearance of the *Mlu*I restriction recognition site. In this same site, the insertion of a ‘G’ after the H-gene start codon resulted in a frameshift mutation and, therefore, truncation of the H-gene sequence. This explained the inability to re-linearize plasmid pWS by *Mlu*I restriction digestion that had been observed *in vitro*. On the other hand, both *Aat*II and *NgoM*IV were able to linearize pWS (data not shown).

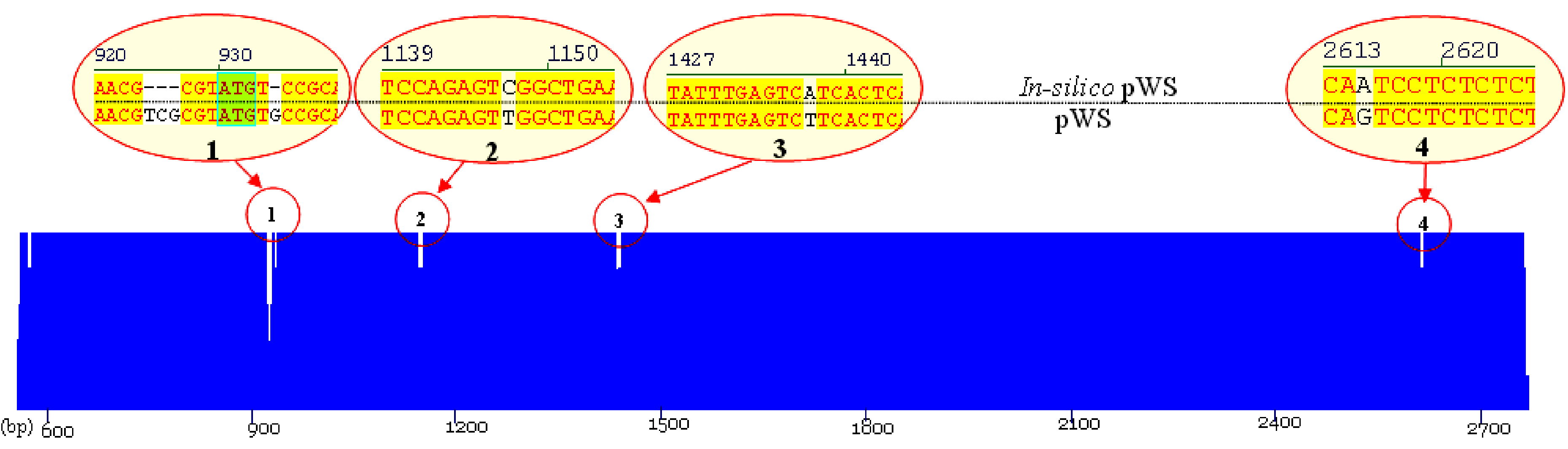


Figure V. 11 Sequence alignment of the actual and the *in-silico* generated pWS sequences.

Four mutated areas were observed on pWS: mutation site 1 contained an additional 'TCG' sequence before and an additional 'G' after the H-gene start codon (highlighted in green). Mutations at regions 2, 3 and 4 consisted in single nucleotide substitutions of 'C' to 'T', 'A' to 'T' and 'A' to 'G' respectively.

3.8 Obtention of the final construct, pWS1

Plasmid pWS1 was obtained after site directed mutagenesis of the additional 'TCG'and 'G' nucleotides observed in the mutation region 1 of plasmid pWS. Complementary primers correcting these mutations Mut-F and Mut-R (Table V.1) were use for the mutagenesis strategy. The resulting plasmid, pWS1, was used to transform *Stbl*3 competent cells under tetracycline selection pressure.

Plasmid mini-preps from ten colonies were subjected to linearization by *Mlu*I digestion, as well as double digestion with *Aat*II and *NgoM*IV, which results in separated insert and vector sequences. The plasmids from the ten colonies analyzed were linearized upon *Mlu*I digestion (Figure V.12). A PCR amplification of PR-CHP-H-gene insert from plasmid pWS1 was sequenced for verification.

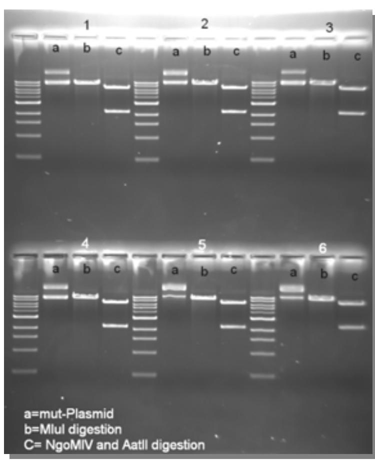


Figure V. 12 Site directed mutagenesis of pWS yielding plasmid pWS1.
Lane (a) intact plasmid revealing circular and supercoiled forms, lane (b) pWS1 linearized by *Mlu*I restriction digestion, lane (c) double restriction digestion of pWS1 with *Aat*II and *NgoM*IV separating insert from vector.

3.9 Verification of the H-gene sequence in plasmid pWS1

Exactly as it was performed for validation of the previous construct pWS, the pWS1 insert sequence was aligned with the *in-silico* construct sequence. Alignment analysis revealed that the two mutations observed in mutation site 1 had been successfully corrected. The single nucleotide mutations observed in sites 2, 3 and 4 were not corrected.

The final H-gene sequence in plasmid pWS1 was translated to amino acid sequence for alignment with the PPRV-H gene sequence. Sequence alignment revealed that the nucleotide substitution present on site 3 was actually a silent mutation, as it did not change the amino acid residue. Nucleotide substitutions on the sites 2 and 4, however, brought codon modifications resulting in a change of the amino acid residues (Figure V.13). In the amino acid sequence these modified residues corresponded to position 73, where arginine (R, a basic amino acid) was changed to tryptophan (W, a hydrophobic amino acid), and to position 562, asparagine (N) was changed to serine (S), both being polar amino acids.

These two amino acid mutations are not located at any of the critical H-protein residues interacting with the signalling lymphocyte activation molecule (SLAM). Regions for SLAM interaction are conserved among morbilliviruses, with conserved residues located at positions 505, 507, 521-523, 525-527, 529-533, 536-537, 547-548 and 552-554 (Tatsuo and Yanagi, 2002). The conservation of T-cell specific epitopes in H-gene subunit vaccines has been shown to be very important in order to generate a protective immune response (Cosby et al., 2006).

Plasmid pWS1 was further propagated in *Stbl*3 and a maxi-prep (Maxi-prep-kit, Qiagen) was done for mycoplasma transformation.

```
H-amino acid(PPRV) (1)   MSAQRERINAFYKDNLHNKTHRVILDRERLTIERPYILLGVLLVMFLSLIGLLAIAGIRL
H-amino acid(pWS1) (1)   MSAQRERINAFYKDNLHNKTHRVILDRERLTIERPYILLGVLLVMFLSLIGLLAIAGIRL
                                     73
H-amino acid(PPRV)(61)   HRATVGTAEIQSRLNTNIELTESIDHQTKDVLTPLFKIIGDEVGIRIPQKFSDLVKFISD
H-amino acid(pWS1)(61)   HRATVGTAEIQSWLNTNIELTESIDHQTKDVLTPLFKIIGDEVGIRIPQKFSDLVKFISD

H-amino acid(PPRV)(121)  KIKFLNPDREYDFRDLRWCMNPPERVKINFDQFCEYKAAVKSVEHIFESSLNRSERLRLL
H-amino acid(pWS1)(121)  KIKFLNPDREYDFRDLRWCMNPPERVKINFDQFCEYKAAVKSVEHIFESSLNRSERLRLL

H-amino acid(PPRV)(181)  TLGPGTGCLGRTVTRAQFSELTLTLMDLDLEIKHNVSSVFTVVEEGLFGRTYTVWRSDTG
H-amino acid(pWS1)(181)  TLGPGTGCLGRTVTRAQFSELTLTLMDLDLEIKHNVSSVFTVVEEGLFGRTYTVWRSDTG

H-amino acid(PPRV)(241)  KPSTSPGIGHFLRVFEIGLVRDLELGAPIFHMTNYLTVNMSDDYRSCLLAVGELKLTALC
H-amino acid(pWS1)(241)  KPSTSPGIGHFLRVFEIGLVRDLELGAPIFHMTNYLTVNMSDDYRSCLLAVGELKLTALC

H-amino acid(PPRV)(301)  TPSETVTLSESGVPKREPLVVVILNLAGPTLGGELYSVLPTTDPTVEKLYLSSHRGIIKD
H-amino acid(pWS1)(301)  TPSETVTLSESGVPKREPLVVVILNLAGPTLGGELYSVLPTTDPTVEKLYLSSHRGIIKD

H-amino acid(PPRV)(361)  NEANWVVPSTDVRDLQNKGECLVEACKTRPPSFCNGTGIGPWSEGRIPAYGVIRVSLDLA
H-amino acid(pWS1)(361)  NEANWVVPSTDVRDLQNKGECLVEACKTRPPSFCNGTGIGPWSEGRIPAYGVIRVSLDLA

H-amino acid(PPRV)(421)  SDPGVVITSVFGPLIPHLSGMDLYNNPFSRAAWLAVPPYEQSFLGMINTIGFPDRAEVMP
H-amino acid(pWS1)(421)  SDPGVVITSVFGPLIPHLSGMDLYNNPFSRAAWLAVPPYEQSFLGMINTIGFPDRAEVMP

H-amino acid(PPRV)(481)  HILTTEIRGPRGRCHVPIELSSRIDDDIKIGSNMVVLPTKDLRYITATYDVSRSEHAIVY
H-amino acid(pWS1)(481)  HILTTEIRGPRGRCHVPIELSSRIDDDIKIGSNMVVLPTKDLRYITATYDVSRSEHAIVY
                                                562
H-amino acid(PPRV)(541)  YIYDTGRSSSYFYPVRLNFRGNPLSLRIECFPWYHKVWCYHDCLIYNTITNEEVHTRGLT
H-amino acid(pWS1)(541)  YIYDTGRSSSYFYPVRLNFRGSPLSLRIECFPWYHKVWCYHDCLIYNTITNEEVHTRGLT

H-amino acid(PPRV)(601)  GIEVTCNPV
H-amino acid(pWS1)(601)  GIEVTCNPV
```

Figure V. 13 Amino acid sequence alignment of H-gene from PPRV and H-gene from plasmid pWS1.

Amino acid sequences at positions 73 and 562 were mutated to 'W' and 'S' respectively in H-gene from plasmid pWS1.

3.10 Mycoplasma transformation using pWS1

MmmLC strain 95010-C1 has been successfully transformed using pWS1. Unlike the fast growth nature of MmmLC, transformed MmmLC mycoplasmas cells grew very slowly on tetracycline (2µg/ml) selective media. Colonies of small size were observed after 7 days of incubation; the apparition of small, slow growing colonies continued until 30 days of incubation (Figure V. 14).

Figure V. 14 pWS1-transformed mycoplasma colonies after 14 days of incubation.
Growth of transformants colonies on modified Hayflick media supplemented with 2 µg/ml tetracycline

PEG mediated transformation has been described for several mollicutes, including *A. laidlawii*, *M. pulmonis*, *M. mycoides*, *M. capricolum*, *M. gallisepticum*, *M. arthritidis*, and *S. citri* (Cordova et al., 2002; Jarhede et al., 1995; King and Dybvig, 1994b; Renaudin et al., 1995). Although mycoplasmas are not easy targets for transformation and genetic manipulation, a combination of chemical treatments with $CaCl_2$ and PEG has even allowed effective genome transfer between two species of the *M. mycoides* cluster (Lartigue et al., 2007), which had never been achieved for any other species.

3.10.1 Transformation efficiency

The number of transformants colony forming units was calculated using the following formula:

No. transformants cfu = cfu x dil. ratio x transformant vol/plated vol

The total number of transformants obtained in a transformation experiment is calculated by multiplying the number of bacterial colonies growing on selective media by the dilution ratio and by the ratio between the original and the plated volumes.

In our experiment, twelve colonies were counted on plates inoculated with the 10^{-3} dilution, with the original transformation volume being 300 µl and the plated transformation volume being 200 µl. The number of transformants cfu was estimated as 1.8×10^{4} transformant cfu.

For the calculation of transformation efficiency the µg of plasmid DNA used for the transformation were not taken into consideration since the plasmid concentration had not been optimized in this study. The total mycoplasma cell count was performed on non-selective media, resulting in 75 colonies at 10^{-5} dilution. The transformation efficiency was then calculated as the ratio between the number of transformed mycoplasma cells and the total mycoplasma cell count. The transformation efficiency was estimated as 1.6×10^{-4} transformant cfu/total cfu; corresponding approximately 18 transformants in 112,500 mycoplasma cells.

This transformation efficiency is comparable to previous transformation efficiencies obtained for MmmLC strains. When transforming MmmLC strain GM12 the highest efficiency was of 8×10^{-4} transformant cfu/total cfu, which was obtained by using plasmid pIKΔ (5.9 Kbp, King and Dybvig, 1994a). These authors did neither consider the µg of plasmid used for the transformation in their calculation. When considering in the calculations the 10 µg of plasmid pMYCO1 (7.4 Kbp) used for transformation of MmmLC strain Y-goat^{T}, transformation efficiency was of 6×10^{-5} transformant cfu/total cfu/µg of plasmid DNA (Lartigue et al., 2003). Similarly, the transformation efficiencies of strain Y-goat^{T} were 4.5×10^{-6} and 5.8×10^{-7}, respectively using plasmids pMYCO1/IS*1296* and pMYCO1/lppB of both 8.7 Kbp (Lorenzon, personal communication). Discrepancies in

transformation efficiencies may be attributed to variable transformation abilities between strains, as well as to differences in the methodology used. Here, MmmLC strain 95010-C1 exhibited an increased transformation capacity.

3.10.2 Stability of pWS1 in MmmLC cells

After two and three passage on 4 µg/ml tetracycline selective media, mycoplasma colonies were tested for plasmid presence. A PCR was conducted on randomly selected colonies using primers flanking the insert PR-CHP-H-gene (insert-pmyco-F and m-insert-pmyco-R, Table V.1). Control PCRs using pWS1 plasmid and non-transformed MmmLC strain 95010-C1, were conducted simultaneously. PCR results revealed the presence of this plasmid in all the clones analyzed, which showed the expected fragment size of 2700 bp (Figure V. 15). This implied that pWS1 was stably maintained and replicate in MmmLC strain 95010-C1.

Figure V. 15 Amplification of plasmid pWS1 from transformed mycoplasma cells.
Primers insert-pmyco-F and m-insert-pmyco-R primers amplified 2700 bp expected fragment

3.11 Sequencing of the PR-CHP-H-gene fragment from transformed mycoplasma

The PCR product obtained from pWS1 transformed mycoplasma cells using insert-pmyco-F and m-insert-pmyco-R primers was sequenced as previously described. Sequence alignment revealed no further sequence modifications. In order to elicit protective immunity, the critical amino acids for SLAM interaction in the H-gene should remain

conserved. Apparently, the nucleotide sequences have not experienced any modification at least after three passages in mycoplasma culture.

3.12 H-protein expression analysis

Transformed mycoplasma colonies and those colonies which gave positive results on PCR were analyzed by dot-immunoblotting. As negative and positive controls, non-transformed MmmLC strain 95010-C1 and PPRV were included. Although variable intensities were observed from 13 transformant clones, a background similar to some of the transformed colonies was observed on the non-transformed MmmLC strain 95010-C1. It was therefore difficult to conclude from this test (Figure V. 16).

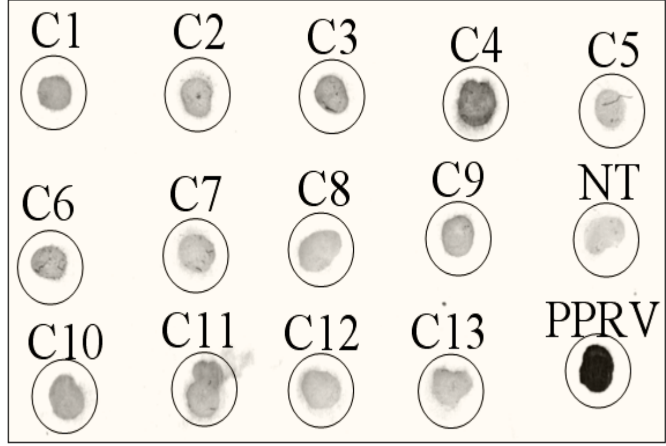


Figure V. 16 Dot-immunoblotting test of transformed mycoplasma cells C1-C13.
NT-non-transformed MmmLC strain 95010-C1 and PPRV (whole protein from peste des petits ruminant virus).

Although sequencing of the H-gene from transformed mycoplasma cells has showed that the gene is not truncated, H protein expression has not been demonstrated. Validation of protein expression constitutes a crucial step in the development of a subunit vaccine. This may be performed by using the western blot technique. This study should then be followed by an assessment of the localization of the H-protein in the transformed mycoplasma cells, which may be achieved by electron microscopy using gold-labelled anti-H specific antibodies.

4 CONCLUSION AND PERSPECTIVES

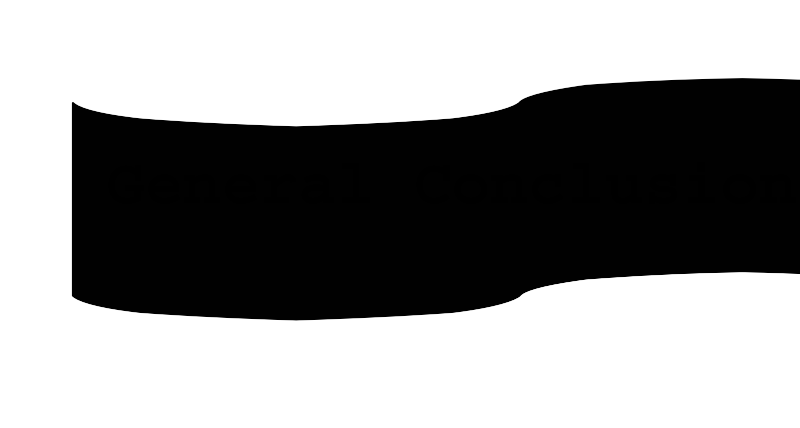
This study was one preliminary step in a wider study that would then involve various points. MmmLC strain 95010-C1 has been successfully transformed using plasmid pWS1. The result of immunoblotting test is not conclusive to say whether H-gene is expressed or not. After verification of H-protein expression in MmmLC, the following perspectives should be contemplated:

- To obtain an expression vector devoid of antibiotic marker.
- To control that goats are immunized with the vector, as a live vaccine or as inactivated antigen develop some immunity.
- To check if it confers protective immunity towards both PPR and MmmLC / Mmc infections.

Possible problems in foreign gene expression in mycoplasma could be difference in promoter sequence, codon usage, mRNA instability, inefficient translation, protein folding problems and rapid degradation. Most importantly however, problems with expression of foreign genes are thought to reside at the level of transcription and translation initiations (Jarhede and Wieslander, 2003). Initially, we attempted to use both promoters from spiralin and CHP genes for expression of H-gene. However, due to time limitations we were not able to repeat the experiment using the spiralin promoter as an alternative promoter. The spiralin promoter has already been successfully used for the expression of foreign genes such as *tetM,* ß-galactosidase (Janis et al., 2005; Lartigue et al., 2003) and green fluorescent protein (GFP) (Duret et al., 2003). In mycoplasma, the promoter sequence of the *vsa* gene has been effectively used in the expression of the chloramphenicol resistance gene *cat* from *E. coli* in *M. pulmonis* and *M. arthritidis* (Dybvig et al., 2000). We assume the promoter sequence of the CHP gene to be a strong promoter able to transcribe the H-gene in mycoplasma. Although the function of the CHP protein is not understood as yet, this protein is highly conserved in the genome of both MmmLC strains sequenced 95010-C1 and GM2b. We have observed that the promoter from CHP gave a higher score for both -10 and -35 boxes and possessed the consensus sequence 'AGGAG', known for ribosomal binding site. Inclusion of an extended region of a promoter sequence including Shine-Dalgarno sequence has been shown to promote foreign gene expression (Halbedel et al., 2007).

It is observed that the speed of growth of transformed MmmLC colonies has been extremely reduced. This may be due esteric interference of cell duplication as a result of incorporation of H-protein onto the mycoplasma cell surface. However, this hypothesis has yet to be investigated. Alternatively, this may be due to the load ensuing from addition of a 9 Kbp extra chromosomal element, particularly considering that MmmLC strain 95010-C1 already contains a natural plasmid of 1.8 Kbp (Chapter II, annotation data of whole genome sequence).

Previously H gene of PPRV was inserted in to the genome of attenuated Capri pox virus, the recombinant Capri pox-PPR haemagglutinin have protected goats against the virulent PPRV. The inclusion of H gene in the genome have not reverted attenuated Capri pox virus to a pathogen form, H-gene by itself as a subunit entity have not shown to induce activities related to cellular attachment.


General

GENERAL CONCLUSION

High throughput sequencing technologies and new bioinformatic tools have literally revolutionized modern microbiology. Phenotypic characterization of bacteria may now seem as nineteenth century tools. A technique such as iso-enzyme profiling has long ago been replaced by multilocus sequencing of all genes coding for the enzymes. Coding sequences can then be compared and the detection of synonymous mutations has further increased the resolving power of the technique.

Many mycoplasma genomes have been sequenced so far and many more are underway. This can be explained by the medical interest of these micro-organisms but also because their small genome makes them good examples for the application of new technologies or new tools. The recent "chromosomal transplant" of the genome of MmmLC into Mcc cells has even been reported in French newspaper like "Le Monde" (Lartigue et al., 2007). Mycoplasmas may also be used in the near future as a model for the development of the first synthetic cell containing minimal set of genes, *M. laboratorium* (Riech, 2000). This may pave the way for the development of new industrial applications.

In our case we have used a whole genome sequence of MmmLC for the development of new diagnostic tools through comparative genomics with already published mycoplasma sequences. In fact the delivery of a complete circularized genome of MmmLC has been delayed because of the presence of duplicated DNA fragments. Difficulty encountered by the Genoscope was linked to presence of a 30 Kbp and 28 Kbp long integrative and conjugative elements (ICEs). Unfortunately the resolving power of the PFGE we had performed was certainly not sufficient to guarantee precise chromosome size estimation. The accurate assembly therefore relied more on the ability of the bioinformaticians that analyzed the sequences generated by the various genome banks.

MmmLC makes no exception to the apparent rule and its genome sequence revealed that it contains insertion sequences of various kinds. A total of twenty three copies of IS elements were identified. The most frequent were IS*1296*, ISMmy*1* and ISMmy*2* while two others, ISMmy*3* and an uncharacterized one, were in smaller copy numbers. These Insertion Sequences are posing some interesting questions on genome plasticity linked to horizontal gene transfer. It must be noted, for example, that IS*1634* is the most frequent IS in MmmSC although it does not seem to be present in MmmLC, its closest neighbour from a

phylogenetic point of view. By contrast IS*1634* variant is present in *M. bovis* that shares the same habitat as MmmSC, the bovine lungs. Similarly we have shown here that there was an identical ISMmy*2* variant in *M. cottewii* and *M. yeatsii* that colonize the ear canals of goats. What remains to be examined is if IS-elements or other mobile genetic elements such as ICE may have any impact on mycoplasma pathogenicity. In theory this could be the case if these elements are favouring horizontal gene transfer of virulence associated genes or if they alter the expression of endogenous virulence associated genes. From a practical point of view IS elements have been used to develop molecular epidemiology tools for mycoplasmas by Southern blotting techniques. The new IS that we have characterized in MmmLC may be useful for such purpose as it varies in copy numbers from one strain to the other. However we did not develop such a tool as another one was developed in our laboratory at the same time. This tool is based on the Multi Locus Sequence Typing (MLST) technique, which is again relying on sequencing, and is much more precise and user-friendly than Southern blotting. In any case the unravelling of copy numbers and target insertion sites of IS-elements are prerequisites for the correct determination of chromosome full sequence.

Comparative genomics was the basis for the development of new diagnostic tools for some members of the "*M. mycoides* cluster". Our initial target was the remnant of the "ADI pathway" operon that was identified in all members of the *M. mycoides* cluster. It seemed that the arrangement of the various genes was differing from one species to the other and therefore we had concluded that we could design specific PCR primers for each of the species. This strategy was successful at first as we were able to develop specific PCR for *M. capricolum* subsp. *capripneumoniae* and for *M. putrefaciens*. These PCR were developed on the basis of much localized sequence polymorphism within the *arc*B and *arc*D genes. However we failed to develop specific PCR tools for the diagnosis of Mcc and MmmLC with the same strategy. The disclosure of the strain 95010-C1 MmmLC sequence data helped us to understand why. When we compared the ADI operon sequence in that strain with that of the type strainY-goat^{T}, obtained previously, we have observed that an important polymorphism existed not only in terms of gene sequences but also on "operon" structure. One of the reasons for this polymorphism could be that this operon is no more functional and therefore prone to the accumulation of any genetic event, from deletion or rearrangement of large DNA fragments to single nucleotide mutations. Alternatively, we developed a PCR for the *M. mycoides* cluster based on the amplification of the *glk* gene

which is well conserved within this group of species but also seems quite specific compared to other bacteria *glk* sequences.

In the near future, comparative genomics made on full genomes for more than one strain for each member of the *M. mycoides* cluster may reveal the extent of polymorphism that exist within mycoplasma chromosomes. This will encompass gene repertoires, chromosomal re-arrangements, lateral gene transfer or sequence variations. It will then be easier to find adequate DNA targets for any particular need. Quantitative PCR (QPCR) is on the verge of replacing classical PCR as a rapid diagnostic tool for many infectious diseases. Already the PCR primers that we had designed for the specific detection of Mccp have been successfully used for the development of a QPCR. QPCR has three main advantages compared to classical PCR: an improved sensitivity (about 2 logs), the detection of the amplified product without opening the tubes (hence reducing contamination risks) and the ability to quantify the amount of specific target in the sample. There is no doubt that this technique will soon replace classical PCR even in developing countries.

Using mycoplasmas as expression vectors is a relatively new concept. This possibility has arisen following the development of *ori*C vectors that can be used for functional genomic studies. These *ori*C vectors were developed primarily for the targeted disruption of genes to infer their functionality but it was also used for the expression of heterologous genes a promoterless *lac*Z gene and a gene encoding for spiralin, in Mcc. In the latter case, however the heterologous gene was originating from a close relative of the recipient cell as *Spiroplasma. citri* belongs to the same phylogenetic group. However the above finding has revealed more distant genes with despite evident codon usage differences can also be transcribed and translated. More recently promoterless *lac*Z gene encoding β-galactosidase from *E. coli* has also been used as a tool to analyse the activity of promoter fragments from *M. pneumoniae* (Halbedel and Stulke, 2006). In our case we wished to express genes of veterinary interest and we chose the H gene of “peste des petits ruminants” virus with the ultimate goal to develop multivalent vaccines. Again this is a new concept as the actual trend in veterinary vaccine is to use other expression vectors such as pox viruses. These pox viruses may offer multiple advantages as they are thermo stable and have a large genome that may accommodate more than one heterologous gene including genes coding for immunomodulators. However, in the case of mycoplasmas such pox vectors may not be

suitable firstly because the immune mechanisms at stake for the protection against mycoplasmas is still not well understood, secondly because it is not known if pox vectors can elicit the proper type of immune response and thirdly because it is not known which, and if, a single mycoplasma antigen would be able to trigger a protective response. For those reasons we wished to use an alternative approach and express viral antigens in mycoplasmas. Lack of time did not allow us to prove if this H protein was really expressed in MmmLC; further work is in progress to check this by western blotting experiment. However we were lacking the appropriate monoclonal antibodies or hyper-immune sera to do so. If this expression is evidenced there will be a number of points that will have to be checked. The first will be the quantification of the expression as well as assessment of the position of the expressed H protein. It will be notably interesting to check if this protein is located at the surface of the mycoplasma. This is theoretically possible. Again if the H protein is expressed by MmmLC it will be interesting to check if the virulence of that organism has been modified. On the one hand it could be diminished as the growth rate of the transformed MmmLC has been dramatically reduced. However the H protein may have conferred additional properties to the transformed mycoplasma that could have acquired the ability to bind to specific cells. Although previous works on subunit vaccines of H-protein have shown H-protein by itself do not induce these features unless accompanied by F-protein the other structural protein. The most important point in the subunit H vaccine is the conservation of T-cell recognition epitopes, which is the apparent situation in this study. Finally it will be interesting to check if animals have been able to develop an immune response both against MmmLC but also to the H protein and check if this response may elicit some protection against the homologous disease but also against PPR.

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990, Basic local alignment search tool. J Mol Biol 215, 403-410.

Arif, A., Schulz, J., Thiaucourt, F., Taha, A., Hammer, S., 2007, Contagious caprine pleuropneumonia outbreak in captive wild ungulates at Al Wabra Wildlife Preservation, State of Qatar. J Zoo Wildl Med 38, 93-96.

Banyard, A.C., Rima, B.K., Barrett, T., 2006, The morbilliviruses, In: Barrett, T., Pastoret, P.P., Taylor, W.P. (Eds.) Rinderpest and Peste des Petits Ruminants: Virus Plagues of large and small ruminants. Academic Press, London, pp. 13-30.

Barile, M.F., 1983, Arginine Hydrolysis, In: Razin, S., Tully, J.G. (Eds.) Methods in Mycoplasmology Volume I, Mycoplasma Characterization. Academic Press, London, pp. 345-349.

Barile, M.F., Schimke, R.T., Riggs, D.B., 1966, Presence of the Arginine Dihydrolase Pathway in Mycoplasma. J Bacteriol 91, 189-192.

Barre, A., de Daruvar, A., Blanchard, A., 2004, MolliGen, a database dedicated to the comparative genomics of Mollicutes. Nucleic Acids Res 32, D307-310.

Barrett, T., Banyard, A.C., Diallo, A., 2006, Molecular biologuy of the Morbilliviruses, In: Barrett, T., Pastoret, P.P., Taylor, W.P. (Eds.) Rinderpest and Peste des Petits Ruminants: Virus Plagues of large and small ruminants. Academic Press, London, pp. 31-67.

Bascunana, C.R., Mattsson, J.G., Bolske, G., Johansson, K.E., 1994, Characterization of the 16S rRNA genes from Mycoplasma sp. strain F38 and development of an identification system based on PCR. J Bacteriol 176, 2577-2586.

Bashiruddin, J.B., Taylor, T.K., Gould, A.R., 1994, A PCR-based test for the specific identification of Mycoplasma mycoides subspecies mycoides SC. J Vet Diagn Invest 6, 428-434.

Bergemann, A.D., Whitley, J.C., Finch, L.R., 1989, Homology of mycoplasma plasmid pADB201 and staphylococcal plasmid pE194. J Bacteriol 171, 593-595.

Bergemann, A.D., Whitley, J.C., Finch, L.R., 1990, Taxonomic significance of differences in DNA methylation within the M. mycoides cluster detected with restriction endonucleases MboI and DpnI. Letters in applied Microbiology 11, 48-51.

Bergonier, D., Berthelot, X., Poumarat, F., 1997, Contagious agalactia of small ruminants: current knowledge concerning epidemiology, diagnosis and control. Rev Sci Tech 16, 848-873.

Bergonier, D., Thiaucourt, F., 2003, Agalactie contagious des petits ruminants, In: Principales maladies infectieuses et parasitaires du bétail Europe et régions chaudes. Lavoisier, Paris, pp. 809-825.

Berhe, G., Minet, C., Le Goff, C., Barrett, T., Ngangnou, A., Grillet, C., Libeau, G., Fleming, M., Black, D.N., Diallo, A., 2003, Development of a dual recombinant vaccine to protect small ruminants against peste-des-petits-ruminants virus and capripoxvirus infections. J Virol 77, 1571-1577.

Bhugra, B., 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.

Bischof, D.F., Vilei, E.M., Frey, J., 2006, Genomic differences between type strain PG1 and field strains of Mycoplasma mycoides subsp. mycoides small-colony type. Genomics 88, 633-641.

Blank, W.A., Stemke, G.W., 2000, A physical and genetic map of the Mycoplasma hyopneumoniae strain J genome. Can J Microbiol 46, 832-840.

Bolske, G., 1988, Survey of Mycoplasma infections in cell cultures and a comparison of detection methods. Zentralbl Bakteriol Mikrobiol Hyg [A] 269, 331-340.

Bolske, G., Mattsson, J.G., Bascunana, C.R., Bergstrom, K., Wesonga, H., Johansson, K.E., 1996, Diagnosis of contagious caprine pleuropneumonia by detection and identification of Mycoplasma capricolum subsp. capripneumoniae by PCR and restriction enzyme analysis. J Clin Microbiol 34, 785-791.

Bonnet, F., Saillard, C., Bove, J.M., Leach, R.H., Rose, D.L., Cottew, G.S., Tully, J.G., 1993, DNA relatedness between field isolates of Mycoplasma F38 group, the agent of contagious caprine pleuropneumonia, and strains of Mycoplasma capricolum. Int J Syst Bacteriol 43, 597-602.

Bourdin, P., Laurent-Vautier, A., 1967, Note sur la structure de virus de la peste des petits ruminants. Rev Elev Med Vet Pays Trop 20, 383-386.

Bové, J.M., 1993, Molecular Features of Mollicutes. Clinical infectious disease 17, 10-31.

Brocchi, E., Gamba, D., Poumarat, F., Martel, J.L., De Simone, F., 1993, Improvements in the diagnosis of contagious bovine pleuropneumonia through the use of monoclonal antibodies. Rev Sci Tech 12, 559-570.

Bruce, A., Alexander, J., Julian, L., Martin, R., Keith, R., Peter, W., 2002, Molecular Biology of the Cell, 4th Edition. Garland Science, New York and London.

Buttery, S.H., Lloyd, L.C., Titchen, D.A., 1976, Acute respiratory, circulatory and pathological changes in the calf after intravenous injections of the galactan from Mycoplasma mycoides subsp. mycoides. J Med Microbiol 9, 379-391.

Calcutt, M.J., Lavrrar, J.L., Wise, K.S., 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.

Chen, K.-W., Lo, H.-J., Lin, Y.-H., Li, S.-Y., 2005, Comparison of four molecular typing methods to assess genetic relatedness of Candida albicans clinical isolates in Taiwan. J Med Microbiol 54, 249-258.

Cheng, X., Nicolet, J., Poumarat, F., Regalla, J., Thiaucourt, F., Frey, J., 1995, Insertion element IS1296 in Mycoplasma mycoides subsp. mycoides small colony identifies a European clonal line distinct from African and Australian strains. Microbiology 141 (Pt 12), 3221-3228.

Christiansen, C., Erno, H., 1990, RFLP in rRNA genes of Mycoplasma capricolum, the caprine F38-like group and the bovineserogroup 7. Zentralbl Bakteriol Hyg Suppl 20, 479-488.

Citti, C., Browning, G.F., Rosengarten, R., 2005, Phenotypic diversity and cell invasion in host subversion by pathogenic mycoplasmas, In: Blanchard, A., Browning, G. (Eds.) Mycoplasmas, Molecular biology pathogenicity and strategies for control. Horizon bioscience, Norfolk, UK, pp. 439-483.

Cordova, C.M., Lartigue, C., Sirand-Pugnet, P., Renaudin, J., Cunha, R.A., Blanchard, A., 2002, Identification of the origin of replication of the Mycoplasma pulmonis chromosome and its use in oriC replicative plasmids. J Bacteriol 184, 5426-5435.

Cosby, S.L., Chieko, K., Kazuya, Y., 2006, Immunology of rinderpest an immunosuppression but a lifelong vaccine protection, In: Barrett, T., Pastoret, P.P., Taylor, W.P. (Eds.) Rinderpest and Peste des Petits Ruminants: Virus Plagues of large and small ruminants. Academic Press, London, pp. 198-221.

Cottew, G.S., 1979, Pathogenicity of the subspecies mycoides of Mycoplasma mycoides for cattle, sheep and goats. Zentralbl Bakteriol [Orig A] 245, 164-170.

Cottew, G.S., Breard, A., DaMassa, A.J., Erno, H., Leach, R.H., Lefevre, P.C., Rodwell, A.W., Smith, G.R., 1987, Taxonomy of the Mycoplasma mycoides cluster. Isr J Med Sci 23, 632-635.

Cottew, G.S., Yeats, F.R., 1978, Subdivision of Mycoplasma mycoides subsp. mycoides from cattle and goats into two types. Aust Vet J 54, 293-296.

DaMassa, A.J., Tully, J.G., Rose, D.L., Pitcher, D., Leach, R.H., Cottew, G.S., 1994, Mycoplasma auris sp. nov., Mycoplasma cottewii sp. nov., and Mycoplasma yeatsii sp. nov., new sterol-requiring mollicutes from the external ear canals of goats. Int J Syst Bacteriol 44, 479-484.

DaMassa, A.J., Wakenell, P.S., Brooks, D.L., 1992, Mycoplasmas of goats and sheep. J Vet Diagn Invest 4, 101-113.

De la Fe, C., Assuncao, P., Antunes, T., Rosales, R.S., Poveda, J.B., 2005, Microbiological survey for Mycoplasma spp. in a contagious agalactia endemic area. Vet J 170, 257-259.

De la Fe, C., Gutierrez, A., Poveda, J.B., Assuncao, P., Ramirez, A.S., Fabelo, F., 2007, First isolation of Mycoplasma capricolum subsp. capricolum, one of the causal agents of caprine contagious agalactia, on the island of Lanzarote (Spain). Vet J 173, 440-442.

Dedieu, L., Breard, A., Lefevre, P.C., 1992, Development of a species-specific DNA probe for Mycoplasma capricolum. Vet Microbiol 32, 189-197.

Dedieu, L., Mady, V., Lefevre, P.C., 1994, Development of a selective polymerase chain reaction assay for the detection of Mycoplasma mycoides subsp. Mycoides S.C. (contagious bovine pleuropneumonia agent). Vet Microbiol 42, 327-339.

D'Hooghe, I., Vander Wauven, C., Michiels, J., Tricot, C., de Wilde, P., Vanderleyden, J., Stalon, V., 1997, The arginine deiminase pathway in Rhizobium etli: DNA sequence analysis and functional study of the arcABC genes. J Bacteriol 179, 7403-7409.

Diallo, A., Minet, C., Berhe, G., Le Goff, C., Black, D.N., Fleming, M., Barrett, T., Grillet, C., Libeau, G., 2002, Goat immune response to capripox vaccine expressing the hemagglutinin protein of peste des petits ruminants. Ann N Y Acad Sci 969, 88-91.

Diallo, A., Minet, C., Le Goff, C., Berhe, G., Albina, E., Libeau, G., Barrett, T., 2007, The threat of peste des petits ruminants: progress in vaccine development for disease control. Vaccine 25, 5591-5597.

Diallo, A., Taylor, W.P., Lefevre, P.C., Provost, A., 1989, Atténuation d'une souche de virus de la peste des petitis ruminants: candidat pour un vaccin homologue. Rev Elev Med Vet Pays Trop 42, 1479-1487.

Dotto, G.P., Horiuchi, K., Jakes, K.S., Zinder, N.D., 1982, Replication origin of bacteriophage f1. Two signals required for its function. J Mol Biol 162, 335-343.

Duret, S., Berho, N., Danet, J.L., Garnier, M., Renaudin, J., 2003, Spiralin is not essential for helicity, motility, or pathogenicity but is required for efficient transmission of Spiroplasma citri by its leafhopper vector Circulifer haematoceps. Appl Environ Microbiol 69, 6225-6234.

Dybvig, K., French, C.T., Voelker, L.L., 2000, Construction and use of derivatives of transposon Tn4001 that function in Mycoplasma pulmonis and Mycoplasma arthritidis. J Bacteriol 182, 4343-4347.

Dybvig, K., Voelker, L.L., 1996, Molecular biology of mycoplasmas. Annu Rev Microbiol 50, 25-57.

Edward, D.G.f., Freundt, E.A., 1956, The classification and nomenclature of organisms of the pleuropneumonia group. J. Gen. Microbiol. 14, 197-207.

Field, D., Hood, D., Moxon, R., 1999, Contribution of genomics to bacterial pathogenesis. Current opinion in Genetics & Development 9, 700-703.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M., et al., 1995, Whole-

genome random sequencing and assembly of Haemophilus influenzae Rd. Science 269, 496-512.
Frangeul, L., Glaser, P., Rusniok, C., Buchrieser, C., Duchaud, E., Dehoux, P., Kunst, F., 2004, CAAT-Box, Contigs-Assembly and Annotation Tool-Box for genome sequencing projects. Bioinformatics 20, 790-797.
Frangeul, L., Nelson, K.E., Buchrieser, C., Danchin, A., Glaser, P., Kunst, F., 1999, Cloning and assembly strategies in microbial genome projects. Microbiology 145 (Pt 10), 2625-2634.
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., Kelley, J.M., Fritchman, R.D., Weidman, J.F., Small, K.V., Sandusky, M., Fuhrmann, J., Nguyen, D., Utterback, T.R., Saudek, D.M., Phillips, C.A., Merrick, J.M., Tomb, J.F., Dougherty, B.A., Bott, K.F., Hu, P.C., Lucier, T.S., Peterson, S.N., Smith, H.O., Hutchison, C.A., 3rd, Venter, J.C., 1995, The minimal gene complement of Mycoplasma genitalium. Science 270, 397-403.
Frey, J., 2003, Insertion Sequence Analysis, In: Miles, R.A.J., Nicholas, R.A. (Eds.) Mycoplasma protocols. Humana Press Inc., Totowa, NJ, pp. 197-205.
Frey, J., Cheng, X., Kuhnert, P., Nicolet, J., 1995, Identification and characterization of IS1296 in Mycoplasma mycoides subsp. mycoides SC and presence in related mycoplasmas. Gene 160, 95-100.
Frey, J., Cheng, X., Monnerat, M.P., Abdo, E.M., Krawinkler, M., Bolske, G., Nicolet, J., 1998, Genetic and serological analysis of the immunogenic 67-kDa lipoprotein of Mycoplasma sp. bovine group 7. Res Microbiol 149, 55-64.
Fuhrmann, M., Hausherr, A., Ferbitz, L., Schödl, T., Heitzer, M., Hegemann, P., 2004, Monitoring dynamic expression of nuclear genes in Chlamydomonas reinhardtii by using a synthetic luciferase reporter gene. http://dx.doi.org/10.1007/s11103-005-2150-1. Plant Mol Biol 55(6), 869-881.
Galas, D.J., Chandler, M., 1981, On the molecular mechanisms of transposition. Proc Natl Acad Sci U S A 78, 4858-4862.
Gamper, M., Zimmermann, A., Haas, D., 1991, Anaerobic regulation of transcription initiation in the arcDABC operon of Pseudomonas aeruginosa. J Bacteriol 173, 4742-4750.
Gasparich, G.E., Whitcomb, R.F., Dodge, D., French, F.E., Glass, J., Williamson, D.L., 2004, The genus Spiroplasma and its non-helical descendants: phylogenetic classification, correlation with phenotype and roots of the Mycoplasma mycoides clade. Int J Syst Evol Microbiol 54, 893-918.
Gil, M.C., Hermoso de Mendoza, M., Rey, J., Alonso, J.M., Poveda, J.B., Hermoso de Mendoza, J., 1999, Aetiology of caprine contagious agalactia syndrome in Extremadura, Spain. Vet Rec 144, 24-25.
Glasner, J.D., Rusch, M., Liss, P., Plunkett, G., 3rd, Cabot, E.L., Darling, A., Anderson, B.D., Infield-Harm, P., Gilson, M.C., Perna, N.T., 2006, ASAP: a resource for annotating, curating, comparing, and disseminating genomic data. Nucleic Acids Res 34, D41-45.
Glass, J.I., Assad-Garcia, N., Alperovich, N., Yooseph, S., Lewis, M.R., Maruf, M., A., C., Hutchison, I., Smith, H.O., Venter, J.C., 2006, Essential genes of a minimal bacterium. PNAS 103, 425 - 430.
Gregory, S.G., Howell, G.R., Bentley, D.R., 1997, Genome mapping by fluorescent fingerprinting. Genome Res 7, 1162-1168.

Gurrieri, S., Smith, S.B., Bustamante, C., 1999, Trapping of megabase-sized DNA molecules during agarose gel electrophoresis. Proc Natl Acad Sci U S A 96, 453-458.
Halbedel, S., Eilers, H., Jonas, B., Busse, J., Hecker, M., Engelmann, S., Stulke, J., 2007, Transcription in Mycoplasma pneumoniae: Analysis of the Promoters of the ackA and ldh Genes. J Mol Biol 371, 596-607.
Halbedel, S., Stulke, J., 2006, Probing in vivo promoter activities in Mycoplasma pneumoniae: a system for generation of single-copy reporter constructs. Appl Environ Microbiol 72, 1696-1699.
Heldtander, M., Pettersson, B., Tully, J.G., Johansson, K.E., 1998, Sequences of the 16S rRNA genes and phylogeny of the goat mycoplasmas Mycoplasma adleri, Mycoplasma auris, Mycoplasma cottewii and Mycoplasma yeatsii. Int J Syst Bacteriol 48 Pt 1, 263-268.
Hollis, R.J., Bruce, J.L., Fritschel, S.J., Pfaller, M.A., 1999, Comparative evaluation of an automated ribotyping instrument versus pulsed-field gel electrophoresis for epidemiological investigation of clinical isolates of bacteria. Diagn Microbiol Infect Dis 34, 263-268.
Hotzel, H., Sachse, K., Pfutzner, H., 1996, A PCR scheme for differentiation of organisms belonging to the Mycoplasma mycoides cluster. Vet Microbiol 49, 31-43.
Howard, C.J., Taylor, G., 1985, Immune responses to mycoplasma infections of the respiratory tract. Vet Immunol Immunopathol 10, 3-32.
Hum, S., Kessell, A., Djordjevic, S., Rheinberger, R., Hornitzky, M., Forbes, W., Gonsalves, J., 2000, Mastitis, polyarthritis and abortion caused by Mycoplasma species bovine group 7 in dairy cattle. Aust Vet J 78, 744-750.
Isono, K., McIninch, J.D., Borodovsky, M., 1994, Characteristic features of the nucleotide sequences of yeast mitochondrial ribosomal protein genes as analyzed by computer program GeneMark. DNA Res 1, 263-269.
Janis, C., Lartigue, C., Frey, J., Wroblewski, H., Thiaucourt, F., Blanchard, A., Sirand-Pugnet, P., 2005, Versatile use of oriC plasmids for functional genomics of Mycoplasma capricolum subsp. capricolum. Appl Environ Microbiol 71, 2888-2893.
Jarhede, T.K., Le Henaff, M., Wieslander, A., 1995, Expression of foreign genes and selection of promoter sequences in Acholeplasma laidlawii. Microbiology 141 (Pt 9), 2071-2079.
Jarhede, T.K., Wieslander, A., 2003, Expression of foreign genes in Acholeplasma laidlawii, In: Miles, R.A.J., Nicholas, R.A. (Eds.) Mycoplasma protocols. Humana Press Inc., Totowa, NJ, pp. 247-257.
King, K.W., Dybvig, K., 1991, Plasmid transformation of Mycoplasma mycoides subspecies mycoides is promoted by high concentrations of polyethylene glycol. Plasmid 26, 108-115.
King, K.W., Dybvig, K., 1992, Nucleotide sequence of Mycoplasma mycoides subspecies Mycoides plasmid pKMK1. Plasmid 28, 86-91.
King, K.W., Dybvig, K., 1994a, Mycoplasmal cloning vectors derived from plasmid pKMK1. Plasmid 31, 49-59.
King, K.W., Dybvig, K., 1994b, Transformation of Mycoplasma capricolum and examination of DNA restriction modification in M. capricolum and Mycoplasma mycoides subsp. mycoides. Plasmid 31, 308-311.
Knudtson, K.L., Minion, F.C., 1994, Use of lac gene fusions in the analysis of Acholeplasma upstream gene regulatory sequences. J Bacteriol 176, 2763-2766.

Kusiluka, L.J., Ojeniyi, B., Friis, N.F., Kokotovic, B., Ahrens, P., 2001a, Molecular analysis of field strains of Mycoplasma capricolum subspecies capripneumoniae and Mycoplasma mycoides subspecies mycoides, small colony type isolated from goats in Tanzania. Vet Microbiol 82, 27-37.

Kusiluka, L.J., Ojeniyi, B., Friis, N.F., Kokotovic, B., Ahrens, P., 2001b, Molecular epidemiology of contagious bovine pleuropneumonia in Tanzania based on amplified fragment length polymorphism and pulsed-field gel electrophoresis analysis. J Vet Med B Infect Dis Vet Public Health 48, 303-312.

Kwiatek, O., Minet, C., Grillet, C., Hurard, C., Carlsson, E., Karimov, B., Albina, E., Diallo, A., Libeau, G., 2007, Peste des petits ruminants (PPR) outbreak in Tajikistan. J Comp Pathol 136, 111-119.

Lacks, S., Greenberg, B., 1977, Complementary specificity of restriction endonucleases of Diplococcus pneumoniae with respect to DNA methylation. J. Mol. Biol. 114, 153-168.

Lartigue, C., Blanchard, A., Renaudin, J., Thiaucourt, F., Sirand-Pugnet, P., 2003, Host specificity of mollicutes oriC plasmids: functional analysis of replication origin. Nucleic Acids Res 31, 6610-6618.

Lartigue, C., Glass, J.I., Alperovich, N., Pieper, R., Parmar, P.P., Hutchison, C.A., 3rd, Smith, H.O., Venter, J.C., 2007, Genome transplantation in bacteria: changing one species to another. Science 317, 632-638.

Le Grand, D., Saras, E., Blond, D., Solsona, M., Poumarat, F., 2004, Assessment of PCR for routine identification of species of the Mycoplasma mycoides cluster in ruminants. Vet Res 35, 635-649.

Leach, R.H., Erno, H., MacOwan, K.J., 1993, Proposal for designation of F38-type caprine mycoplasmas as Mycoplasma capricolum subsp. capripneumoniae subsp. nov. and consequent obligatory relegation of strains currently classified as M. capricolum (Tully, Barile, Edward, Theodore, and Erno 1974) to an additional new subspecies, M. capricolum subsp. capricolum subsp. nov. Int J Syst Bacteriol 43, 603-605.

Lewin, B., 2002, Genes VII, Oxford Edition. Oxford University Press Inc., New York.

MacOwan, K.J., Minette, J.E., 1976, A mycoplasma from acute contagious caprine pleuropneumonia in Kenya. Trop Anim Health Prod 8, 91-95.

Madanat, A., d., Z., Pospišil, z., 2001, A review: Contagious agalactia of sheep and goats. Acta vet. Brno 70, 403-412.

Maghnouj, A., de Sousa Cabral, T.F., Stalon, V., Vander Wauven, C., 1998, The arcABDC gene cluster, encoding the arginine deiminase pathway of Bacillus licheniformis, and its activation by the arginine repressor argR. J Bacteriol 180, 6468-6475.

Mahillon, J., Chandler, M., 1998, Insertion sequences. Microbiol Mol Biol Rev 62, 725-774.

Mahillon, J., Leonard, C., Chandler, M., 1999, IS elements as constituents of bacterial genomes. Res Microbiol 150, 675-687.

Maniloff, J., 1992, Phylogeny of mycoplasmas, In: Maniloff, J. (Ed.) Mycoplasmas : molecular biology and pathogenesis.New York, pp. 549-559.

Maniloff, J., 1996, The minimal cell genome: "on being the right size". Proc Natl Acad Sci U S A 93, 10004-10006.

Manso-Silvan, L., Perrier, X., Thiaucourt, F., 2007, Phylogeny of the Mycoplasma mycoides cluster based on analysis of five conserved protein-coding sequences and possible implications for the taxonomy of the group. Int J Syst Evol Microbiol 57, 2247-2258.

March, J.B., Clark, J., Brodlie, M., 2000, Characterization of strains of Mycoplasma mycoides subsp. mycoides small colony type isolated from recent outbreaks of

contagious bovine pleuropneumonia in Botswana and Tanzania: evidence for a new biotype. J Clin Microbiol 38, 1419-1425.
Marois, C., Dufour-Gesbert, F., Kempf, I., 2001, Comparison of pulsed-field gel electrophoresis with random amplified polymorphic DNA for typing of Mycoplasma synoviae. Vet Microbiol 79, 1-9.
Maxam, A.M., Gilbert, W., 1977, A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA 74 (2), 144-148.
McAuliffe, L., Kokotovic, B., Ayling, R.D., Nicholas, R.A., 2004, Molecular epidemiological analysis of Mycoplasma bovis isolates from the United Kingdom shows two genetically distinct clusters. J Clin Microbiol 42, 4556-4565.
McNeil, L.K., Reich, C., Aziz, R.K., Bartels, D., Cohoon, M., Disz, T., Edwards, R.A., Gerdes, S., Hwang, K., Kubal, M., Margaryan, G.R., Meyer, F., Mihalo, W., Olsen, G.J., Olson, R., Osterman, A., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Rodionov, D.A., Shi, X., Vassieva, O., Vonstein, V., Zagnitko, O., Xia, F., Zinner, J., Overbeek, R., Stevens, R., 2007, The National Microbial Pathogen Database Resource (NMPDR): a genomics platform based on subsystem annotation. Nucleic Acids Res 35, D347-353.
Miles, K., McAuliffe, L., Persson, A., Ayling, R.D., Nicholas, R.A., 2005, Insertion sequence profiling of UK Mycoplasma bovis field isolates. Vet Microbiol 107, 301-306.
Miserez, R., Pilloud, T., Cheng, X., Nicolet, J., Griot, C., Frey, J., 1997, Development of a sensitive nested PCR method for the specific detection of Mycoplasma mycoides subsp. mycoides SC. Mol Cell Probes 11, 103-111.
Moll, M., Klenk, H.D., Maisner, A., 2002, Importance of the cytoplasmic tails of the measles virus glycoproteins for fusogenic activity and the generation of recombinant measles viruses. J Virol 76, 7174-7186.
Monnerat, M.P., Thiaucourt, F., Nicolet, J., Frey, J., 1999a, Comparative analysis of the lppA locus in Mycoplasma capricolum subsp. capricolum and Mycoplasma capricolum subsp. capripneumoniae. Vet Microbiol 69, 157-172.
Monnerat, M.P., Thiaucourt, F., Poveda, J.B., Nicolet, J., Frey, J., 1999b, Genetic and serological analysis of lipoprotein LppA in Mycoplasma mycoides subsp. mycoides LC and Mycoplasma mycoides subsp. capri. Clin Diagn Lab Immunol 6, 224-230.
Nanda, Y.P., Chatterjee, A., Purohit, A.K., Diallo, A., Innui, K., Sharma, R.N., Libeau, G., Thevasagayam, J.A., Bruning, A., Kitching, R.P., Anderson, J., Barrett, T., Taylor, W.P., 1996, The isolation of peste des petits ruminants virus from northern India. Vet Microbiol 51, 207-216.
Nicholas, R.A., Santini, F.G., Clark, K.M., Palmer, N.M., De Santis, P., Bashiruddin, J.B., 1996, A comparison of serological tests and gross lung pathology for detecting contagious bovine pleuropneumonia in two groups of Italian cattle. Vet Rec 139, 89-93.
Oliveira, M.F.S., Mendes, D.Q., Ferrari, L.I., VAsconcelos, A.T.R., 2004, Ribosome binding site recognition using neural network. Genetics and Molecular biology 27, 644-650.
Overbeek, R., Begley, T., Butler, R.M., Choudhuri, J.V., Chuang, H.Y., Cohoon, M., de Crecy-Lagard, V., Diaz, N., Disz, T., Edwards, R., Fonstein, M., Frank, E.D., Gerdes, S., Glass, E.M., Goesmann, A., Hanson, A., Iwata-Reuyl, D., Jensen, R., Jamshidi, N., Krause, L., Kubal, M., Larsen, N., Linke, B., McHardy, A.C., Meyer, F., Neuweger, H., Olsen, G., Olson, R., Osterman, A., Portnoy, V., Pusch, G.D., Rodionov, D.A., Ruckert, C., Steiner, J., Stevens, R., Thiele, I., Vassieva, O., Ye, Y., Zagnitko, O., Vonstein, V., 2005, The subsystems approach to genome

annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res 33, 5691-5702.
Ozdemir, U., Loria, G.R., Godinho, K.S., Samson, R., Rowan, T.G., Churchward, C., Ayling, R.D., Nicholas, R.A., 2006, Effect of danofloxacin (Advocin A180) on goats affected with contagious caprine pleuropneumonia. Trop Anim Health Prod 38, 533-540.
Ozkul, A., Akca, Y., Alkan, F., Barrett, T., Karaoglu, T., Dagalp, S.B., Anderson, J., Yesilbag, K., Cokcaliskan, C., Gencay, A., Burgu, I., 2002, Prevalence, distribution, and host range of Peste des petits ruminants virus, Turkey. Emerg Infect Dis 8, 708-712.
Pastoret, P.P., 2006, Rienderpest: a general introductio,, In: Barrett, T., Pastoret, P.P., Taylor, W.P. (Eds.) Rinderpest and Peste des Petits Ruminants: Virus Plagues of large and small ruminants. Academic press, London, pp. 1-12.
Perreau, P., Bind, J.L., 1981, Infection naturelle du veau par Mycoplasma mycoides subsp. mycoides (biotype chèvre). Bull. Acad. Vet. Fr. 54, 491-496.
Persson, A., Jacobsson, K., Frykberg, L., Johansson, K.E., Poumarat, F., 2002, Variable surface protein Vmm of Mycoplasma mycoides subsp. mycoides small colony type. J Bacteriol 184, 3712-3722.
Persson, A., Pettersson, B., Bolske, G., Johansson, K.E., 1999, Diagnosis of contagious bovine pleuropneumonia by PCR-laser- induced fluorescence and PCR-restriction endonuclease analysis based on the 16S rRNA genes of Mycoplasma mycoides subsp. mycoides SC. J Clin Microbiol 37, 3815-3821.
Peterson, S.N., Bailey, C.C., Jensen, J.S., Borre, M.B., King, E.S., Bott, K.F., Hutchison, C.A., 3rd, 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.
Pettersson, B., Leitner, T., Ronaghi, M., Bolske, G., Uhlen, M., Johansson, K.E., 1996, Phylogeny of the Mycoplasma mycoides cluster as determined by sequence analysis of the 16S rRNA genes from the two rRNA operons. J Bacteriol 178, 4131-4142.
Peyraud, A., Woubit, S., Poveda, J.B., De la Fe, C., Mercier, P., Thiaucourt, F., 2003, A specific PCR for the detection of Mycoplasma putrefaciens, one of the agents of the contagious agalactia syndrome of goats. Mol Cell Probes 17, 289-294.
Pilo, P., Vilei, E.M., Peterhans, E., Bonvin-Klotz, L., Stoffel, M.H., Dobbelaere, D., Frey, J., 2005, A metabolic enzyme as a primary virulence factor of Mycoplasma mycoides subsp. mycoides small colony. J Bacteriol 187, 6824-6831.
Plemper, R.K., Hammond, A.L., Cattaneo, R., 2000, Characterization of a region of the measles virus hemagglutinin sufficient for its dimerization. J Virol 74, 6485-6493.
Pyle, L.E., Taylor, T., Finch, L.R., 1990, Genomic maps of some strains within the Mycoplasma mycoides Cluster. Journal of Bacteriology 172, 7265-7268.
Razin, S., 1997, The minimal cellular genome of mycoplasma. Indian J Biochem Biophys 34, 124-130.
Razin, S., Jacobs, E., 1992, Mycoplasma Adhesion. J Gen Microbiol 138, 407-422.
Razin, S., Yogev, D., Naot, Y., 1998, Molecular biology and pathogenicity of mycoplasmas. Microbiol Mol Biol Rev 62, 1094-1156.
Renaudin, J., Marais, A., Verdin, E., Duret, S., Foissac, X., Laigret, F., Bove, J.M., 1995, Integrative and free Spiroplasma citri oriC plasmids: expression of the Spiroplasma phoeniceum spiralin in Spiroplasma citri. J Bacteriol 177, 2870-2877.
Riech, K.A., 2000, The search for essential genes. Res Microbiol 151, 319-324.

Rodriguez, J.L., Ermel, R.W., Kenny, T.P., Brooks, D.L., DaMassa, A.J., 1997, Polymerase chain reaction and restriction endonuclease digestion for selected members of the "Mycoplasma mycoides cluster" and Mycoplasma putrefaciens. J Vet Diagn Invest 9, 186-190.

Romero, C.H., Barrett, T., Chamberlain, R.W., Kitching, R.P., Fleming, M., Black, D.N., 1994a, Recombinant capripoxvirus expressing the hemagglutinin protein gene of rinderpest virus: protection of cattle against rinderpest and lumpy skin disease viruses. Virology 204, 425-429.

Romero, C.H., Barrett, T., Kitching, R.P., Carn, V.M., Black, D.N., 1994b, Protection of cattle against rinderpest and lumpy skin disease with a recombinant capripoxvirus expressing the fusion protein gene of rinderpest virus. Vet Rec 135, 152-154.

Romualdi, A., Siddiqui, R., Glockner, G., Lehmann, R., Suhnel, J., 2005, GenColors: accelerated comparative analysis and annotation of prokaryotic genomes at various stages of completeness. Bioinformatics 21, 3669-3671.

Ronaghi, M., Uhlen, M., Nyren, P., 1998, A sequencing method based on real-time pyrophosphate. Science 281, 363, 365.

Rosendal, S., 1993, Mycoplasma, In: Gyles Carlton L , Thoen Charles O (Eds.) Pathogenesis of bacterial infections in Animals. Iowa state university press, Ames, Iowa.

Rosengarten, R., Behrens, A., Stetefeld, A., Heller, M., Ahrens, M., Sachse, K., Yogev, D., Kirchhoff, H., 1994, Antigen heterogeneity among isolates of Mycoplasma bovis is generated by high-frequency variation of diverse membrane surface proteins. Infect Immun 62, 5066-5074.

Rosengarten, R., Yogev, D., 1996, Variant colony surface antigenic phenotypes within mycoplasma strain populations: implications for species identification and strain standardization. J Clin Microbiol 34, 149-158.

Rottem, S., 2003, Interaction of Mycoplasmas with host cells. Physiological Reviews 83, 417-432.

Rurangirwa, F.R., Shompole, P.S., Wambugu, A.N., McGuire, T.C., 2000, Monoclonal antibody differentiation of Mycoplasma mycoides subsp. mycoides small-colony strains causing contagious bovine pleuropneumonia from less important large-colony strains. Clin Diagn Lab Immunol 7, 519-521.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989, Molecular Cloning: a Laboratory Manual, 2nd Edition. Cold Spring Harbor, NY.

Sanger, F., Nicklen, S., Coulson, A.R., 1977, DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463-5467.

Schimke, R.T., Barile, M.F., 1963, Arginine Metabolism in Pleuropneumonia-Like Organisms Isolated from Mammalian Cell Culture. J Bacteriol 86, 195-206.

Selengut, J.D., Haft, D.H., Davidsen, T., Ganapathy, A., Gwinn-Giglio, M., Nelson, W.C., Richter, A.R., White, O., 2007, TIGRFAMs and Genome Properties: tools for the assignment of molecular function and biological process in prokaryotic genomes. Nucleic Acids Res 35, D260-264.

Seth, S., Shaila, M.S., 2001, The hemagglutinin-neuraminidase protein of peste des petits ruminants virus is biologically active when transiently expressed in mammalian cells. Virus Res 75, 169-177.

Sibylle Schneider-Schaulies, Volker ter Meulen, 2002, Measles virus and immunomodulation: molecular bases and perspectives. Exp. Rev. Mol. Med. 30 May, http://www.expertreviews.org/02004696h.htm

Siguier, P., Filee, J., Chandler, M., 2006, Insertion sequences in prokaryotic genomes. Curr Opin Microbiol 9, 526-531.

Sirand-Pugnet, P., Lartigue, C., Marenda, M., Jacob, D., Barre, A., Barbe, V., Schenowitz, C., Mangenot, S., Couloux, A., Segurens, B., de Daruvar, A., Blanchard, A., Citti, C., 2007, Being pathogenic, plastic, and sexual while living with a nearly minimal bacterial genome. PLoS Genet 3, e75.

Southern, E.M., 1975, Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol. 98, 503-517.

Spano, G., Chieppa, G., Beneduce, L., Massa, S., 2004, Expression analysis of putative arcA, arcB and arcC genes partially cloned from Lactobacillus plantarum isolated from wine. J Appl Microbiol 96, 185-193.

Tardy, F., Mercier, P., Solsona, M., Saras, E., Poumarat, F., 2007, Mycoplasma mycoides subsp. mycoides biotype large colony isolates from healthy and diseased goats: prevalence and typing. Vet Microbiol 121, 268-277.

Tatsuo, H., Yanagi, Y., 2002, The morbillivirus receptor SLAM (CD150). Microbiol Immunol 46, 135-142.

Taylor, T.K., Bashiruddin, J.B., Gould, A.R., 1992a, Application of a diagnostic DNA probe for the differentiation of the two types of Mycoplasma mycoides subspecies mycoides. Res Vet Sci 53, 154-159.

Taylor, T.K., Bashiruddin, J.B., Gould, A.R., 1992b, Relationships between members of the Mycoplasma mycoides cluster as shown by DNA probes and sequence analysis. Int J Syst Bacteriol 42, 593-601.

Tcherepanov, V., Ehlers, A., Upton, C., 2006, Genome Annotation Transfer Utility (GATU): rapid annotation of viral genomes using a closely related reference genome. BMC Genomics 7, 150.

Thiaucourt, F., 2002, Contagious bovine pleuropneumonia and vaccine strain T1/44. Vet Rec 151, 156.

Thiaucourt, F., Bolske, G., 1996, Contagious caprine pleuropneumonia and other pulmonary mycoplasmoses of sheep and goats. Rev Sci Tech 15, 1397-1414.

Thiaucourt, F., Bolske, G., Libeau, G., Le Goff, C., Lefevre, P.C., 1994, The use of monoclonal antibodies in the diagnosis of contagious caprine pleuropneumonia (CCPP). Vet Microbiol 41, 191-203.

Thiaucourt, F., Guerin, C., Mady, V., Lefevre, P.C., 1992, [Diagnosis of caprine contagious pleuropneumonia: recent improvements]. Rev Sci Tech 11, 859-865.

Thiaucourt, F., Lorenzon, S., David, A., Breard, A., 2000, Phylogeny of the Mycoplasma mycoides cluster as shown by sequencing of a putative membrane protein gene. Vet Microbiol 72, 251-268.

Thomas, A., Linden, A., Mainil, J., Bischof, D.F., Frey, J., Vilei, E.M., 2005, Mycoplasma bovis shares insertion sequences with Mycoplasma agalactiae and Mycoplasma mycoides subsp. mycoides SC: Evolutionary and developmental aspects. FEMS Microbiol Lett 245, 249-255.

Trichard, C.J., Jordaan, P., Prozesky, L., Jacobsz, E.P., Henton, M.M., 1993, The identification of Mycoplasma mycoides mycoides LC as the aetiological agent of balanoposthitis and vulvovaginitis in sheep in South Africa. Onderstepoort J Vet Res 60, 29-37.

Turlan, C., Chandler, M., 1995, IS1-mediated intramolecular rearrangements: formation of excised transposon circles and replicative deletions. Embo J 14, 5410-5421.

Vilei, E.M., Korczak, B.M., Frey, J., 2006, Mycoplasma mycoides subsp. capri and Mycoplasma mycoides subsp. mycoides LC can be grouped into a single subspecies. Vet Res 37, 779-790.

Vilei, E.M., Nicolet, J., Frey, J., 1999, IS1634, a novel insertion element creating long, variable-length direct repeats which is specific for Mycoplasma mycoides subsp. mycoides small-colony type. J Bacteriol 181, 1319-1323.
Viswanath, L., Lu, Y., Fox, G.E., 2007, Genome display tool: visualizing features in complex data sets. Source Code Biol Med 2, 1.
Waite, E.R., March, J.B., 2001, Effect of HEPES buffer systems upon the pH, growth and survival of Mycoplasma mycoides subsp. mycoides small colony (MmmSC) vaccine cultures. FEMS Microbiol Lett 201, 291-294.
Waite, E.R., March, J.B., 2002, Capsular polysaccharide conjugate vaccines against contagious bovine pleuropneumonia: Immune responses and protection in mice. J Comp Pathol 126, 171-182.
Weisburg, W.G., Tully, J.G., Rose, D.L., Petzel, J.P., Oyaizu, H., Yang, D., Mandelco, L., Sechrest, J., Lawrence, T.G., Van Etten, J., et al., 1989, A phylogenetic analysis of the mycoplasmas: basis for their classification. J Bacteriol 171, 6455-6467.
Westberg, J., 2003. Genome Sequencing Analysisi of pathogenecity factors in hereditary and bacterial disseases. Royal Institute of Technology, Department of Biotechnology, Stockholm.
Westberg, J., Persson, A., Holmberg, A., Goesmann, A., Lundeberg, J., Johansson, K.E., Pettersson, B., Uhlen, M., 2004, The genome sequence of Mycoplasma mycoides subsp. mycoides SC type strain PG1T, the causative agent of contagious bovine pleuropneumonia (CBPP). Genome Res 14, 221-227.
Westberg, J., Persson, A., Pettersson, B., Uhlen, M., Johansson, K.E., 2002, ISMmy1, a novel insertion sequence of Mycoplasma mycoides subsp. mycoides small colony type. FEMS Microbiol Lett 208, 207-213.
Wholsein, P., Saliki, J., 2006, Rinderpest ans peste des petits ruminants-the diseases: clinical signs ans pathology, In: Barrett, T., Pastoret, P.P., Taylor, W.P. (Eds.) Rinderpest and Peste des Petits Ruminants: Virus Plagues of large and small ruminants. Academic press, London, pp. 68-85.
Wise, K.S., Foecking, M.F., Roske, K., Lee, Y.J., Lee, Y.M., Madan, A., Calcutt, M.J., 2006, Distinctive repertoire of contingency genes conferring mutation- based phase variation and combinatorial expression of surface lipoproteins in Mycoplasma capricolum subsp. capricolum of the Mycoplasma mycoides phylogenetic cluster. J Bacteriol 188, 4926-4941.
Woubit, S., Lorenzon, S., Peyraud, A., Manso-Silvan, L., Thiaucourt, F., 2004, A specific PCR for the identification of Mycoplasma capricolum subsp. capripneumoniae, the causative agent of contagious caprine pleuropneumonia (CCPP). Vet Microbiol 104, 125-132.
Woubit, S., Manso-Silvan, L., Lorenzon, S., Gaurivaud, P., Poumarat, F., Pellet, M.P., Singh, V.P., Thiaucourt, F., 2007, A PCR for the detection of mycoplasmas belonging to the Mycoplasma mycoides cluster: Application to the diagnosis of contagious agalactia. Mol Cell Probes 21, 391-399.
Wrestler, J.C., Lipes, B.D., Birren, B.W., Lai, E., 1996, Pulsed-field gel electrophoresis. Methods Enzymol 270, 255-272.
Ye, F., Laigret, F., Whitley, J.C., Citti, C., Finch, L.R., Carle, P., Renaudin, J., Bove, J.M., 1992, A physical and genetic map of the Spiroplasma citri genome. Nucleic Acids Res 20, 1559-1565.
Yilma, T., Hsu, D., Jones, L., Owens, S., Grubman, M., Mebus, C., Yamanaka, M., Dale, B., 1988, Protection of cattle against rinderpest with vaccinia virus recombinants expressing the HA or F gene. Science 242, 1058-1061.

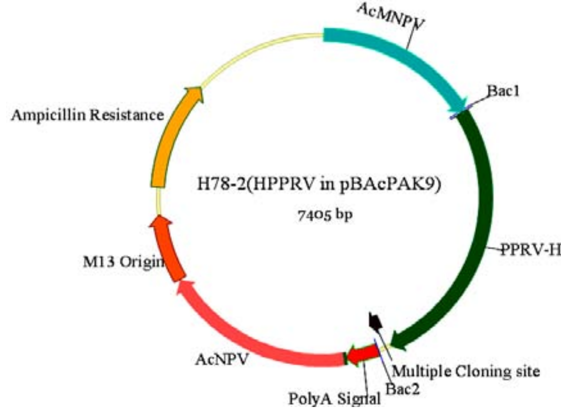
Zheng, J., McIntosh, M.A., 1995, Characterization of IS1221 from Mycoplasma hyorhinis: expression of its putative transposase in Escherichia coli incorporates a ribosomal frameshift mechanism. Mol Microbiol 16, 669-685.

Zhu, H., Hu, G.Q., Yang, Y.F., Wang, J., She, Z.S., 2007, MED: a new non-supervised gene prediction algorithm for bacterial and archaeal genomes. BMC Bioinformatics 8, 97.

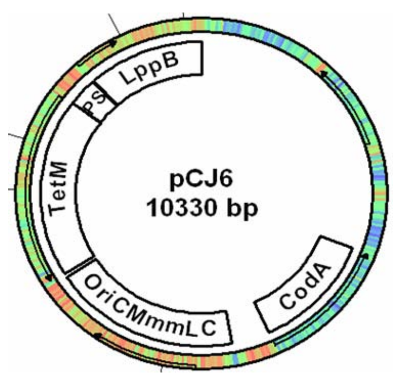
Zuniga, M., Perez, G., Gonzalez-Candelas, F., 2002, Evolution of arginine deiminase (ADI) pathway genes. Mol Phylogenet Evol 25, 429-444.

ANNEXES

Annex-1. Plasmid H78-2 containing PPR-H



Annex-2. Plasmid pCJ6



Annex-3. 20X SSC solution (3 M NaCl, 0.3 M sodium citrate, Adjusted pH to 7.0 (20°C) and autoclaved.

Annex-4. Blue and white screening

The molecular mechanism for blue white screening is based on the Lac operon. The vector (e.g. pBluescript) contains the *lac*Z gene with an internal multiple cloning site (MCS). The MCS can be cleaved by different restriction enzymes so that the foreign DNA can be inserted within *Lac*Z gene, thus disrupting the activity of the β-galactosidase when the protein is expressed. The chemical required for this screen is X-gal, a colorless modified galactose sugar that is metabolized by β-galactosidase the products are a bright blue, and thus functions as an indicator, and Isopropyl β-D-1-thiogalactopyranoside(IPTG), which functions as the inducer of the Lac operon in the absence of lactose. The hydrolysis of colourless X-gal by the β-galactosidase causes the characteristic blue color in the colonies; it shows that the colonies contain unligated vector. White colonies indicate insertion of foreign DNA and loss of the cells' ability to hydrolyse the marker.